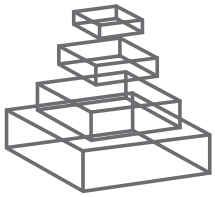
frontiers
RESEARCH TOPICS

MAPPING THE CONNECTOME: MULTI-LEVEL ANALYSIS OF BRAIN CONNECTIVITY

Topic Editors
Trygve B. Leergaard, Olaf Sporns and Claus Hilgetag

frontiers in
NEUROINFORMATICS



frontiers

Cover image provided by Ibbl sarl, Lausanne CH

ISSN 1664-8714
ISBN 978-2-88919-107-9
DOI 10.3389/978-2-88919-107-9

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.
All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

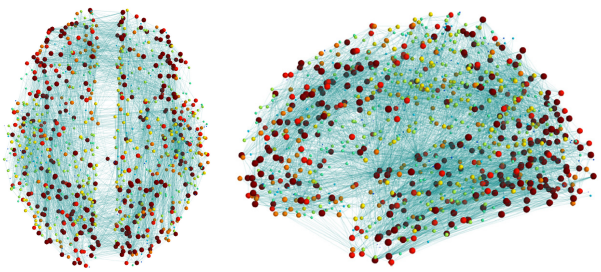
Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.
By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!
Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

MAPPING THE CONNECTOME: MULTI-LEVEL ANALYSIS OF BRAIN CONNECTIVITY

Topic Editors:
Trygve B. Leergaard, University of Oslo, Norway
Olaf Sporns, Indiana University, USA
Claus Hilgetag, Jacobs University Bremen, Germany



Brain Connectivity Hairball visualization of connectivity nodes based on diffusion MRI data. See Gerhard et al. (2011). The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Front. Neuroinform.* 5:3. doi: 10.3389/fninf.2011.00003 for further details.

The brain contains vast numbers of interconnected neurons that constitute functional networks. Structural descriptions of neuronal network elements and connections make up 'the connectome' of the brain, and are important for understanding normal brain function and disease-related dysfunction. A long-standing ambition of the neuroscience community is to achieve complete connectome maps for the human brain as well as primate and rodent brains. Currently, a wide repertoire of experimental tools is available for neural connectivity mapping at multiple levels of scale, from tracing of major pathways and trajectories, mapping of axonal distribution patterns, to the identification of the molecular properties of individual synapses. But, despite numerous connectivity studies through many decades, we are still far from achieving comprehensive descriptions of the connectome. There is increasing awareness that new neuroinformatics tools and strategies are needed to achieve the goal of compiling the brain's connectome, and that any such effort will require systematic, large-scale approaches. Initial attempts involving systematic literature mining have yielded promising results, but more coordinated efforts are needed to collect, organize and disseminate connectome data sets. To this end, there is an urgent need to develop and identify neuroinformatics approaches that allow different levels of connectivity data to be described, integrated, compared, and shared within the broader

neuroscience community. With this Research Topic, we aim to bring together different levels of connectivity analysis (from MRI-based methods, through axonal tracing techniques, to detailed EM-level synapse reconstructions), to elucidate neuroinformatics-related challenges at the level of data management, data comparison and analysis, and use of connectome data for neurocomputational models. We invite contributions related to all aspects of brain connectomics, with particular focus on state-of-the art tools for mapping connectivity, data sharing and comparison, and integration across different levels of mapping.

This Research Topic of Frontiers in Neuroinformatics is dedicated to the memory of Rolf Kötter, a pioneer in the field of brain connectomics.

Table of Contents

06 ***Mapping the Connectome: Multi-Level Analysis of Brain Connectivity***
Trygve B. Leergaard, Claus C. Hilgetag and Olaf Sporns

Advances in Multi-Level Connectivity Mapping

12 ***Automated Probabilistic Reconstruction of White-Matter Pathways in Health and Disease Using an Atlas of the Underlying Anatomy***
Anastasia Yendiki, Patricia Panneck, Priti Srinivasan, Allison Stevens, Lilla Zöllei, Jean Augustinack, Ruopeng Wang, David Salat, Stefan Ehrlich, Tim Behrens, Saad Jbabdi, Randy Gollub and Bruce Fischl

24 ***On the Influence of Amplitude on the Connectivity Between Phases***
Andreas Daffertshofer and Bernadette C. M. van Wijk

36 ***Correspondence Between Structure and Function in the Human Brain at Rest***
Judith M. Segall, Elena A. Allen, Rex E. Jung, Erik B. Erhardt, Sunil K. Arja, Kent Kiehl and Vince D. Calhoun

53 ***Microstructural Analysis of Human White Matter Architecture Using Polarized Light Imaging: Views from Neuroanatomy***
Hubertus Axer, Sindy Beck, Markus Axer, Friederike Schuchardt, Jörn Heepe, Anja Flücken, Martina Axer, Andreas Prescher and Otto W. Witte

65 ***High-Resolution Fiber Tract Reconstruction in the Human Brain by Means of Three-Dimensional Polarized Light Imaging***
Markus Axer, David Grässel, Melanie Kleiner, Jürgen Dammers, Timo Dickscheid, Julia Reckfort, Tim Hütz, Björn Eiben, Uwe Pietrzyk, Karl Zilles and Katrin Amunts

78 ***The Importance of Combining MRI and Large-Scale Digital Histology in Neuroimaging Studies of Brain Connectivity and Disease***
Jacopo Annese

84 ***Multiscale Exploration of Mouse Brain Microstructures Using the Knife-Edge Scanning Microscope Brain Atlas***
Ji Ryang Chung, Chul Sung, David Mayerich, Jaerock Kwon, Daniel E. Miller, Todd Huffman, John Keyser, Louise C. Abbott and Yoonsuck Choe

101 ***Tracing Activity Across the Whole Brain Neural Network with Optogenetic Functional Magnetic Resonance Imaging***
Jin Hyung Lee

Integrative Efforts: Assembling Connectomes

108 ***Informatics and Data Mining Tools and Strategies for the Human Connectome Project***
Daniel S. Marcus, John Harwell, Timothy Olsen, Michael Hodge, Matthew F. Glasser, Fred Prior, Mark Jenkinson, Timothy Laumann, Sandra W. Curtiss and David C. Van Essen

120 The Connectome Viewer Toolkit: An Open Source Framework to Manage, Analyze, and Visualize Connectomes
Stephan Gerhard, Alessandro Daducci, Alia Lemkaddem, Reto Meuli, Jean-Philippe Thiran and Patric Hagmann

135 Potential Synaptic Connectivity of Different Neurons onto Pyramidal Cells in a 3D Reconstruction of the Rat Hippocampus
Deepak Ropireddy and Giorgio A. Ascoli

148 The Retrosplenial Cortex: Intrinsic Connectivity and Connections with the (Para)Hippocampal Region in the Rat. An Interactive Connectome
Jørgen Sugar, Menno P. Witter, Niels M. van Strien and Natalie L. M. Cappaert

161 Knowledge Synthesis with Maps of Neural Connectivity
Marcelo Tallis, Richard Thompson, Thomas A. Russ and Gully A. P. C. Burns

171 Combining Collation and Annotation Efforts Toward Completion of the Rat and Mouse Connectomes in BAMS
Mihail Bota, Hong-Wei Dong and Larry W. Swanson

Applied Connectomics: Network Analysis and Modeling

181 Hierarchical Information-Based Clustering for Connectivity-Based Cortex Parcellation
Nico S. Gorbach, Christoph Schütte, Corina Melzer, Mathias Goldau, Olivia Sujazow, Jenia Jitsev, Tania Douglas and Marc Tittgemeyer

194 Integrating Temporal and Spatial Scales: Human Structural Network Motifs Across Age and Region of Interest Size
Christoph Echtermeyer, Cheol E. Han, Anna Rotarska-Jagiela, Harald Mohr, Peter J. Uhlhaas and Marcus Kaiser

208 Modeling the Connectome of a Simple Spinal Cord
Roman Borisyuk, Abul Kalam al Azad, Deborah Conte, Alan Roberts and Stephen R. Soffe

219 Meeting the Memory Challenges of Brain-Scale Network Simulation
Susanne Kunkel, Tobias C. Potjans, Jochen M. Eppler, Hans Ekkehard Plesser, Abigail Morrison and Markus Diesmann

234 Large-Scale Analysis of Gene Expression and Connectivity in the Rodent Brain: Insights Through Data Integration
Leon French, Powell Patrick Cheng Tan and Paul Pavlidis

245 CoCoMac 2.0 and the Future of Tract-Tracing Databases
Rembrandt Bakker, Thomas Wachtler and Markus Diesmann

frontiers in
NEUROINFORMATICS

EDITORIAL
published: 01 May 2012
doi: 10.3389/fninf.2012.00014

Mapping the connectome: multi-level analysis of brain connectivity

Trygve B. Leergaard[1]*, Claus C. Hilgetag[2,3] and Olaf Sporns[4]

[1] *Centre for Molecular Biology and Neuroscience, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway*
[2] *Department of Computational Neuroscience, University Medical Center Eppendorf, Hamburg University, Hamburg, Germany*
[3] *Department of Health Sciences, Boston University, Boston, MA, USA*
[4] *Department of Psychological and Brain Sciences, Programs in Neuroscience and Cognitive Science, Indiana University, Bloomington, IN, USA*
**Correspondence: t.b.leergaard@medisin.uio.no*

BACKGROUND AND SCOPE

The brain contains vast numbers of interconnected neurons that constitute anatomical and functional networks. Structural descriptions of neuronal network elements and connections make up the "connectome" of the brain (Hagmann, 2005; Sporns et al., 2005; Sporns, 2011), and are important for understanding normal brain function and disease-related dysfunction. A long-standing ambition of the neuroscience community has been to achieve complete connectome maps for the human brain as well as the brains of non-human primates, rodents, and other species (Bohland et al., 2009; Hagmann et al., 2010; Van Essen and Ugurbil, 2012). A wide repertoire of experimental tools is currently available to map neural connectivity at multiple levels, from the tracing of mesoscopic axonal connections and the delineation of white matter tracts (Saleem et al., 2002; Van der Linden et al., 2002; Sporns et al., 2005; Schmahmann et al., 2007; Hagmann et al., 2010), the mapping of neurons organized into functional circuits (Geerling and Loewy, 2006; Ohara et al., 2009; Thompson and Swanson, 2010; Ugolini, 2011), to the identification of cellular-level connections, and the molecular properties of individual synapses (Harris et al., 2003; Arellano et al., 2007; Staiger et al., 2009; Micheva et al., 2010; Wouterlood et al., 2011). But despite the numerous connectivity studies conducted through many decades we are still far from achieving comprehensive descriptions of the connectome across all these levels. There is increasing awareness that new neuroinformatics tools and strategies are needed to achieve the goal of compiling the brain's connectome, and that any such effort will require systematic, large-scale approaches (Bohland et al., 2009; Akil et al., 2011; Zakiewicz et al., 2011; Van Essen and Ugurbil, 2012).

Systematic literature mining to compile and share complete overview of known connections in the macaque brain was pioneered by Rolf Kötter and co-workers (Stephan et al., 2001, 2010). While yielding promising results (Kötter, 2004; Bota et al., 2005; van Strien et al., 2009), more coordinated efforts are needed to collect, organize, and disseminate connectome data sets. To this end, there is an urgent need to develop and identify neuroinformatics approaches that allow different levels of connectivity data to be described, integrated, compared, and shared within the broader neuroscience community.

This Research Topic of Frontiers in Neuroinformatics, dedicated to the memory of Rolf Kötter (1961–2010) and his pioneering work in the field of brain connectomics, comprises contributions that elucidate different levels of connectivity analysis (from MRI-based methods, through axonal tracing techniques, to mapping of functional connectivity in relation to detailed 3-D reconstructions of individual neurons), and point to several recent methodological advances and neuroinformatics-related challenges at the level of data mining, management, and integration. In this Editorial, we review the advances represented in these reports, and discuss some of the grand challenges in this emerging field.

ADVANCES IN MULTI-LEVEL CONNECTIVITY MAPPING

Sophisticated neuroimaging techniques have opened up new possibilities to infer structural and functional connectivity at a macroscopic scale. Through measurement of oriented water diffusion restricted by cellular elements in the brain, non-invasive methods based on diffusion magnetic resonance imaging (dMRI, **Figures 1A,B**) play a key role in current neuroanatomical efforts to explore the human connectome (Hagmann et al., 2010; Van Essen and Ugurbil, 2012). The different dMRI tractography methods proposed so far still require time-consuming manual intervention and supervision that may compromise reliability. To overcome this problem, Yendiki et al. (2011)[1] present a method for automated probabilistic reconstruction of white matter pathways that incorporates *a priori* anatomical knowledge, and demonstrate automatic tractography analyses in schizophrenia patients and healthy subjects (**Figure 1B**). The ability to perform dMRI tractography without manual intervention will greatly facilitate studies with very large populations, which will be essential for establishing a connectome for the human brain (Marcus et al., 2011) as well as for improving early diagnostic imaging in brain disease.

Estimates of "functional networks" described on the basis of statistical associations derived from time series data (neuronal recordings) represent another important category of approaches to define the human brain connectome. The relationship of anatomical to functional networks is explored by Daffertshofer and van Wijk (2011). Using computational modeling of large-scale neural networks these authors argue that patterns of synchronization should be analyzed in the context of changes in local amplitude to improve prediction of brain dynamics from structure. In a related paper, Segall et al. (2012) also employ statistical methods and independent component analysis to describe spatial correspondences between gray matter density measurements and resting state functional MRI signal fluctuations recorded from a very large group of healthy subjects. But while associations between several structural and functional features can be observed (Segall et al., 2012), the anatomical substrates underlying such indirect *in vivo* measurements remain obscure and require further investigation.

[1]Research Topic references are indicated by underlined text.

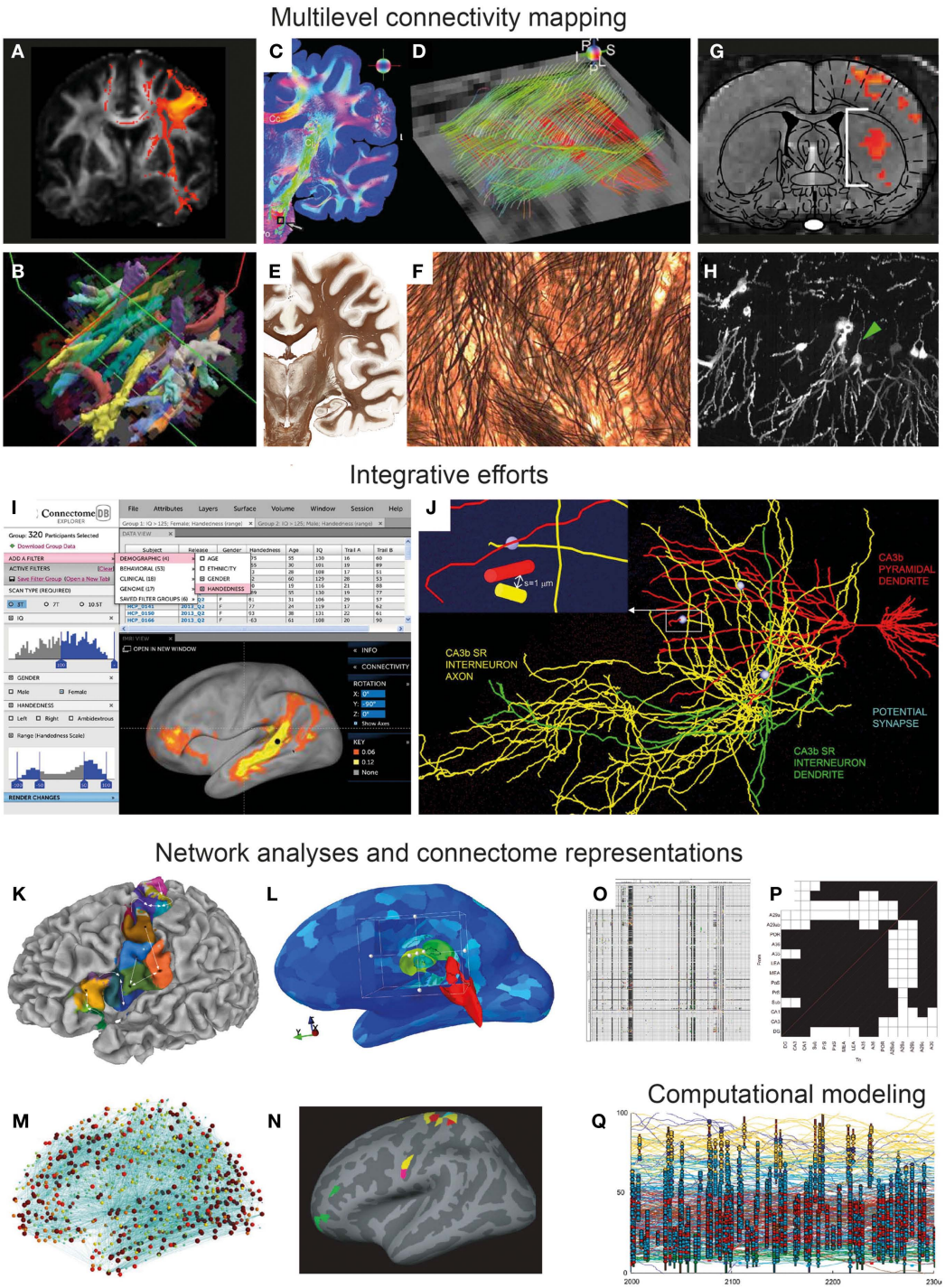


FIGURE 1 | From multi-modal connectivity data to integrated connectomes. Image panels illustrating recent progress, selected and modified from contributions to the Research Topic "Mapping the Connectome" as specified below. **(A–F)** Connectivity mapping methods in the human brain span from indirect *in vivo* assessment of fiber tracts from diffusion MRI data [**(A,B)** Gorbach et al., 2011; Yendiki et al., 2011] to *ex vivo* mapping of detailed fiber architectures [**(C,E)** Axer et al., 2011b; **(E,F)** Annese, 2012]. **(G–H)** Novel experimental methods in animal models include combined optogenetic and functional MRI mapping of specific connections **(G)** (Lee, 2011) and high-resolution histological imaging [**(H)** Chung et al., 2011]. **(I–J)** Examples of recent efforts to accumulate, integrate, and share connectivity data, represented by the graphical user interface of the Human Connectome Project [**(I)** Marcus et al., 2011], and a data mining effort combining 3-D reconstructions of hippocampal neurons to explore potential synaptic connections [**(J)** Ropireddy and Ascoli, 2011]. **(K–N)** Recent sophisticated approaches to network analysis, based on connectivity-based cortical parcelation [**(K)** Gorbach et al., 2011], different connectivity visualizations using open-source tools [**(L,M)** Gerhard et al., 2011], and identification of structural network motifs [**(N)** Echtermeyer et al., 2011]. **(O–P)** Updated connectome matrix representations from large-scale data mining efforts for the whole rat brain [**(O)** Bota et al., 2012] and for the rat hippocampal region [**(P)** Sugar et al., 2011]. **(Q)** Visualization of computational model of a tadpole spinal cord connectome (Borisyuk et al., 2011).

In this respect, novel methodologies for characterizing microstructural aspects of the human and primate brains may provide new ways to bridge the gap between *post mortem* microscopic and *in vivo* macroscopic and functional measures reflecting neural connectivity. Polarized light imaging (PLI) of histological sections allows quantitative analysis of fiber orientations with very-high spatial resolution. Axer et al. (2011a) demonstrate automated 3-D reconstruction of fiber orientations across multiple histological sections in the human brain stem, yielding highly resolved datasets that are useful complements for both conventional histological stains and DTI data (**Figures 1C,D**). Using the same approach, Axer et al. (2011b) show how 3-D PLI derived fiber orientation vectors can subsequently be used as a basis for high-resolution tractography of fiber tracts, potentially suitable for bridging microscopic and macroscopic connectome representations. The importance of correlating various non-invasive MRI derived measurements to cellular-level morphological data is also emphasized by Annese (2012), presenting the perspective that whole-brain histological maps (**Figures 1E,F**) created using large-scale digital microscopy spanning several histological modalities will support the analysis and interpretation of MRI-based connectivity studies. The potential of high-throughput and very-high-resolution histological methods for creating multiscale representations of brain data is further demonstrated by Chung et al. (2011), who use knife-edge scanning microscopy to section and reconstruct microscopic brain data at sub-micrometer resolution (**Figure 1H**), within volumes that can span an entire mouse brain. The inherently multiscale nature of the acquired data sets demand sophisticated visualization and analysis tools for integrating cellular to systems scales, a challenge that is addressed with the introduction of a web-based neuroinformatics platform (Chung et al., 2011). It will be a further, substantial challenge to extract actual connectivity information from these histological representations.

Recent advances in optogenetic functional MRI (ofMRI) allow non-invasive, selective mapping of brain circuit elements that are triggered on the basis of genetic markers, anatomic location, or axonal projection target. In a perspective article, Lee (2011) outlines the potential of ofMRI to enhance system level mapping and our understanding of neural circuits. The combination of the unique ability of optogenetics to selectively control cellular activity and of the mapping of hemodynamic responses with fMRI, opens exiting new possibilities for *in vivo* functional circuit analysis in animals (**Figure 1G**), as well as diverse models of neurological disease. The approach potentially has considerable advantages over classical axonal tracing methods, by allowing the investigation of network dynamics and longitudinal investigations of development and aging, as well as monitoring of neural network changes occurring in disease models.

INTEGRATIVE EFFORTS: ASSEMBLING CONNECTOMES

The methodological diversity of current (functional and structural) approaches to mapping connectomes represents a major challenge to the field, and comparison of data across scales, modalities, and species remains a formidable problem (Sporns, 2011). It is therefore a long-standing ambition of the neuroinformatics community to provide new tools and approaches for integrating neuroscience data (Akil et al., 2011; Van Essen and Ugurbil, 2012). Several of the contributions included in the "Mapping the Connectome" Research Topic present novel ways to visualize and assemble connectome data.

In their technology report, Marcus et al. (2011) of the Human Connectome Project consortium centered at Washington University and the University of Minnesota[2] present their plans to non-invasively assess the long range connections of the human brain using diffusion MRI data from large numbers of subjects, supplemented by structural MRI, genetic, and behavioral data. An important first step for this ambitious project is to develop an informatics platform for storing, visualizing, and analyzing massive amounts of data (**Figure 1I**). A key element will be to share the open-source platform and associated tools with the community, to allow further hypothesis-driven analysis and data mining in this multi-modal pool of data. A particular challenge for future data integration and knowledge synthesis will be to achieve interoperability with other neuroinformatics tools and data collections.

The challenge of integrating and visualizing human connectome data of different modalities has been addressed by Gerhard et al. (2011). Their flexible open-source Connectome Viewer Toolkit[3] allows integrated storage and viewing of different types of imaging data and structured metadata. The toolkit is based on a container file format, and linked to a suite of existing tools libraries allowing data mining, viewing, and comparison. The authors present several example analyses using diffusion MRI data that are integrated, manipulated, and analyzed (**Figures 1L,M**). This freely available open-source toolkit stands out as a powerful workbench platform for future integration of human connectome data coming from multi-modal neuroimaging.

Another approach to combine and integrate various neuroanatomical data is to use a common reference space (Bjaalie, 2002; Hawrylycz et al., 2011). In their contribution to this Research Topic, Ropireddy and Ascoli (2011) demonstrate how potential connections (defined by close apposition of axonal and dendritic segments) can be inferred from a collection of detailed 3-D reconstructions of neuronal extensions that were accumulated in a common 3-D reconstruction of the rat hippocampus on the basis of positional values (**Figure 1J**). The major innovation lies in estimating this aspect of neuronal connectivity at high-resolution not only over a narrow field of view, but across a macroscopic brain region, the rat hippocampus. The paper reports on anatomical differences across different locations in the hippocampus and points to possible generalizations of the approach to other regions of the brain.

Axonal connections have been investigated experimentally in animal models for many decades, yielding valuable data reported in a huge number of publications. While the textual format of such reports make it difficult to extract formal knowledge and compare connectivity data, this pool of laboriously collected experimental results remains important to utilize. Rolf Kötter's pioneering work, particularly the creation of the CoCoMac database[4], is a major milestone on the way to creating comprehensive and accessible online repositories for connectome data (Stephan et al., 2001, 2010). The availability of this repository of primate cortical connections

[2]http://www.humanconnectome.org/

[3]http://www.cmtk.org/

[4]www.cocomac.org

has spurred a large number of projects in the analysis of complex brain connectivity (e.g., Modha and Singh, 2010) as well as the computational modeling of associated brain dynamics (e.g., Cabral et al., 2011; Deco and Jirsa, 2012).

Several groups have invested in further ambitious data mining projects to aggregate connectome data from legacy data (Sugar et al., 2011; Tallis et al., 2011; Bota et al., 2012). Over several years, Bota and co-workers have developed a publicly available neuroinformatics system, called the brain architecture knowledge management system (BAMS), which currently contains >50,000 connectivity reports from the rat brain (**Figure 1O**). The BAMS system[5] provides a valuable framework in which curated connectivity data can be stored and retrieved. In their review article, Bota et al. (2012) provide an update on recently added data and functionality, and discuss general methodology and strategy for producing global connection matrices.

Building upon their earlier work in the hippocampus (van Strien et al., 2009), Sugar et al. (2011) present an interactive connectome of hippocampal and parahippocampal connections. The authors have extracted ~2600 descriptions of hippocampal connections from 226 published reports, and assembled them into a versatile, searchable application providing a comprehensive description of all known network elements in this region (**Figure 1P**). Overcoming a range of challenges related to level of detail, incongruent and incomplete reports, and diverse use of nomenclatures, Sugar et al. provide the most comprehensive connectome description so far for a specific rat brain region. The efforts of these hippocampal experts present an excellent case study for how connectivity data should be assembled for other brain regions.

The process of extracting integrated knowledge representations from connectivity data requires a framework for standardized data descriptions, such as a common atlas space and consistent terminologies for neuroanatomical entities. Tallis et al. (2011) have combined two pre-existing rat brain atlas systems (Dashti et al., 2001; Burns et al., 2006) with a data management system (Russ et al., 2011) into a software system for synthesizing knowledge based on neural connectivity data. This system provides access to experimental tract-tracing data mapped onto atlas plates (Swanson, 2004) coupled to a semantic framework, and permits analyses and interpretation of connectivity patterns based on spatial and semantic views.

Taken together, these different attempts at assembling connectomes from existing and new data illustrate novel possibilities of gaining new knowledge through data systems that allow systematic integration and comparison of data based on standard nomenclatures, semantics, and spatial frameworks. To fulfill the long-term ambition of having reasonably complete connectomes we must also address the challenge of data exchange and interoperability across database systems of this kind.

APPLIED CONNECTOMICS: NETWORK ANALYSIS AND MODELING

While new and sophisticated approaches to connectome mapping and data integration will be essential for the further advancement of this field, it is important to also investigate how accumulated connectome data can be utilized to further our understanding of relationships between brain structure and function. The ways in which connectome data are used will to a large extent also dictate priorities for data collection and assembly. In this Research Topic, five contributions illustrate how connectomics can be applied for network analysis and modeling.

A fundamental principle of brain organization is that functional systems are formed by specific anatomical connections. Knowledge about the specific hodological organization of different brain regions may thus predict the various functional properties of such regions. Gorbach et al. (2011) explore relationships between the functional and connectional "fingerprints" of cerebrocortical areas in the human brain, by using hierarchical information-based clustering of MRI-based connectivity measures. They propose an automated hierarchical parcelation approach to identify cortical subunits that are consistent with cytoarchitectonic maps and previous connectivity-based parcelation schemes (**Figure 1K**).

Echtermeyer et al. (2011) examine changes in network structure that appear when the same brain is mapped at different spatial resolutions. Their emphasis is on network motifs and their composition and spatial distribution (**Figure 1N**). The study highlights the important roles of spatial scale and resolution for drawing conclusions based on network analysis.

Computational modeling represents an increasingly important approach aiming at combining available data on anatomical connectivity with a virtual exploration of functional properties that emerge from the interaction of structural coupling and neural dynamics. Using a model of a simple spinal cord system, the developing spinal cord of the *Xenopus*, Borisyuk et al. (2011) introduce a new approach toward characterizing connectomes by constructing the network on the basis of known developmental processes of neuronal and axonal growth. The resulting network (**Figure 1Q**) is then studied with a number of visualization and topological analysis tools, revealing relationships between sets of simple developmental rules and topological regularities. To meet the considerable computational challenges of simulating complex neural network models, Kunkel et al. (2011) have developed strategies for adapting network representations to reduce the memory consumption for simulation software. Such efforts are highly relevant for future up-scaling of computational modeling efforts, which will be necessary to incorporate more comprehensive structural and functional connectome data.

Finally, French et al. (2011) utilize data sets on rodent brain connectivity and gene expression patterns to explore possible inter-relations, and identify several aspects of connectivity and gene expression that are indeed correlated. This work highlights an integrative aspect of connectomics that links connectivity data with other non-connectomic data sources, e.g., from genomics and proteomics. The confluence of connectomics and genomics will likely be a major growth area in the not-too-distant future.

PROGRESS MADE AND MAIN CHALLENGES

The papers presented in this Research Topic demonstrate approaches to mapping, integrating, and utilizing connectivity data through structured neuroinformatics, in the spirit of the "Kötter School of Neuroinformatics." The multiple modalities and levels of investigation represented in this collection of Research Topic papers illustrate the need for concerted and sustained efforts by several research communities to arrive at reasonably comprehensive connectome maps

[5]http://brancusi.usc.edu/bkms

linking brain regions and neurons. The complexity and multi-faceted nature of brain connectivity clearly calls for cooperation, collaboration, and mutual understanding of methods, problems, and results.

A fundamental challenge for the establishment of a common description of brain connectivity will be to cross-validate the different methodologies. Some important efforts have already been made at different scales (Schmahmann et al., 2007; Bock et al., 2011; Briggman et al., 2011), but further work is needed to link methods such as structural and functional MRI (Segall et al., 2012), 3-D-PLI and tractography (Axer et al., 2011b), and optogenetics and fMRI (Lee, 2011). Since studies of the human connectome continue to rely on indirect connectivity measures, such as DTI and related imaging methods, we will need parallel experimental efforts in animal models and robust statistical methods to demonstrate the validity of connectivity data obtained by non-invasive means. This need for validation also underlines the continuing value of conventional connectivity data that are derived by tract-tracing in animal models and are compiled in databases such as CoCoMac or BAMS. A number of strategies may be imagined to further expand and enhance such compilations. For example, computational text mining approaches may develop to a level where they can be employed for the automated gathering of connectivity information from the published research literature[6]. Further, curated databases may collaborate more closely with field experts, in order to preserve the multi-faceted neuroanatomical knowledge acquired in many experimental labs over several decades (where knowledge is currently threatened by the upcoming retirement of many of the "traditional" anatomists), or new experimental efforts may be started for the systematic, "industrial" brain connectivity gathering in rodent models (Bohland et al., 2009; Zakiewicz et al., 2011).

The papers collected in this Research Topic offer a snapshot of diverse approaches pursued at the inception of a wide array of connectome studies across scales and species. The end goal of a comprehensive understanding of the network structure of complex nervous systems will require the integration of data and methods for mapping connectivity from neurons to systems. We have an exciting time ahead of us.

[6]www.textpresso.org/neuroscience/

REFERENCES

Akil, H., Martone, M. E., and Van Essen, D. C. (2011). Challenges and opportunities in mining neuroscience data. *Science* 331, 708–712.

Annese, J. (2012). The importance of combining MRI and large-scale digital histology in neuroimaging studies of brain connectivity and disease. *Front. Neuroinform.* 6:13. doi: 10.3389/fninf.2012.00013

Arellano, J. I., Benavides-Piccione, R., Defelipe, J., and Yuste, R. (2007). Ultrastructure of dendritic spines: correlation between synaptic and spine morphologies. *Front. Neurosci.* 1:1. doi: 10.3389/neuro.01/1.1.010.2007

Axer, H., Beck, S., Axer, M., Schuchardt, F., Heepe, J., Flucken, A., Axer, M., Prescher, A., and Witte, O. W. (2011a). Microstructural analysis of human white matter architecture using polarized light imaging: views from neuroanatomy. *Front. Neuroinform.* 5:28. doi: 10.3389/fninf.2011.00028

Axer, M., Grässel, D., Kleiner, M., Dammers, J., Dickscheid, T., Reckfort, J., Hütz, T., Eiben, B., Pietrzyk, U., Zilles, K., and Amunts, K. (2011b). High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging. *Front. Neuroinform.* 5:34. doi: 10.3389/fninf.2011.00034

Bjaalie, J. G. (2002). Localization in the brain: new solutions emerging. *Nat. Neurosci. Rev.* 3, 322–325.

Bock, D. D., Lee, W. C., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S., and Reid, R. C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature* 471, 177–182.

Bohland, J. W., Wu, C., Barbas, H., Bokil, H., Bota, M., Breiter, H. C., Cline, H. T., Doyle, J. C., Freed, P. J., Greenspan, R. J., Haber, S. N., Hawrylycz, M., Herrera, D. G., Hilgetag, C. C., Huang, Z. J., Jones, A., Jones, E. G., Karten, H. J., Kleinfeld, D., Kötter, R., Lester, H. A., Lin, J. M., Mensh, B. D., Mikula, S., Panksepp, J., Price, J. L., Safdieh, J., Saper, C. B., Schiff, N. D., Schmahmann, J. D., Stillman, B. W., Svoboda, K., Swanson, L. W., Toga, A. W., Van Essen, D. C., Watson, J. D., and Mitra, P. P. (2009). A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLoS Comput. Biol.* 5, e1000334. doi: 10.1371/journal.pcbi.1000334

Borisyuk, R., Al Azad, A. K., Conte, D., Roberts, A., and Soffe, S. R. (2011). Modeling the connectome of a simple spinal cord. *Front. Neuroinform.* 5:20. doi: 10.3389/fninf.2011.00020

Bota, M., Dong, H. W., and Swanson, L. W. (2005). Brain architecture management system. *Neuroinformatics* 3, 15–48.

Bota, M., Dong, H. W., and Swanson, L. W. (2012). Combining collation and annotation efforts toward completion of the rat and mouse connectomes in BAMS. *Front. Neuroinform.* 6:2. doi: 10.3389/fninf.2012.00002

Briggman, K. L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature* 471, 183–188.

Burns, G. A., Cheng, W. C., Thompson, R. H., and Swanson, L. W. (2006). The NeuARt II system: a viewing tool for neuroanatomical data based on published neuroanatomical atlases. *BMC Bioinformatics* 7, 531. doi: 10.1186/1471-2105-7-531

Cabral, J., Hugues, E., Sporns, O., and Deco, G. (2011). Role of local network oscillations in resting-state functional connectivity. *Neuroimage* 57, 130–139.

Chung, J. R., Sung, C., Mayerich, D., Kwon, J., Miller, D. E., Huffman, T., Keyser, J., Abbott, L. C., and Choe, Y. (2011). Multiscale exploration of mouse brain microstructures using the knife-edge scanning microscope brain atlas. *Front. Neuroinform.* 5:29. doi: 10.3389/fninf.2011.00029

Daffertshofer, A., and van Wijk, B. C. (2011). On the influence of amplitude on the connectivity between phases. *Front. Neuroinform.* 5:6. doi: 10.3389/fninf.2011.00006

Dashti, A., Burns, G. A., Simmons, D., Swanson, L. W., Ghandeharizadeh, S., Shahabi, C., Stone, J., and Jia, S. (2001). "The neuroanatomical rat brain viewer (NeuART)," in *Computing the Brain: A Guide to Neuroinformatics*, eds J. S. Grethe and M. A. Arbib (San Diego: Academic Press), 189–202.

Deco, G., and Jirsa, V. K. (2012). Ongoing cortical activity at rest: criticality, multistability, and ghost attractors. *J. Neurosci.* 32, 3366–3375.

Echtermeyer, C., Han, C. E., Rotarska-Jagiela, A., Mohr, H., Uhlhaas, P. J., and Kaiser, M. (2011). Integrating temporal and spatial scales: human structural network motifs across age and region of interest size. *Front. Neuroinform.* 5:10. doi: 10.3389/fninf.2011.00010

French, L., Tan, P. P., and Pavlidis, P. (2011). Large-scale analysis of gene expression and connectivity in the rodent brain: insights through data integration. *Front. Neuroinform.* 5:12. doi: 10.3389/fninf.2011.00012

Geerling, J. C., and Loewy, A. D. (2006). Aldosterone-sensitive neurons in the nucleus of the solitary tract: bidirectional connections with the central nucleus of the amygdala. *J. Comp Neurol.* 497, 646–657.

Gerhard, S., Daducci, A., Lemkaddem, A., Meuli, R., Thiran, J. P., and Hagmann, P. (2011). The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Front. Neuroinform.* 5:3. doi: 10.3389/fninf.2011.00003

Gorbach, N. S., Schutte, C., Melzer, C., Goldau, M., Sujazow, O., Jitsev, J., Douglas, T., and Tittgemeyer, M. (2011). Hierarchical information-based clustering for connectivity-based cortex parcellation. *Front. Neuroinform.* 5:18. doi: 10.3389/fninf.2011.00018

Hagmann, P. (2005). *From Diffusion MRI to Brain Connectomics*. Ecole Polytechnique Fédérale de Lausanne (Lausanne: EPFL), 1.

Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Grant, P. E., Wedeen, V., Meuli, R., Thiran, J. P., Honey, C. J., and Sporns, O. (2010). MR connectomics: principles and challenges. *J. Neurosci. Methods* 194, 34–45.

Harris, K. M., Fiala, J. C., and Ostroff, L. (2003). Structural changes at dendritic spine synapses during long-term potentiation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 745–748.

Hawrylycz, M., Baldock, R. A., Burger, A., Hashikawa, T., Johnson, G. A., Martone, M., Ng, L., Lau, C., Larsen, S. D., Nissanov, J., Puelles, L., Ruffins, S., Verbeek, F., Zaslavsky, I., and Boline, J. (2011). Digital atlasing and standardization in the mouse brain. *PLoS Comput. Biol.* 7, e1001065. doi: 10.1371/journal.pcbi.1001065

Kötter, R. (2004). Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* 2, 127–144.

Kunkel, S., Potjans, T. C., Eppler, J. M., Plesser, H. E., Morrison, A., and Diesmann, M. (2011). Meeting the memory challenges of brain-scale network simulation. *Front. Neuroinform.* 5:35. doi: 10.3389/fninf.2011.00035

Lee, J. H. (2011). Tracing activity across the whole brain neural network with optogenetic functional magnetic resonance imaging. *Front. Neuroinform.* 5, 21. doi: 10.3389/fninf.2011.00021

Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S. W., and Van Essen, D. C. (2011). Informatics and data mining tools and strategies for the human connectome project. *Front. Neuroinform.* 5:4. doi: 10.3389/fninf.2011.00004

Micheva, K. D., Busse, B., Weiler, N. C., O'Rourke, N., and Smith, S. J. (2010). Single-synapse analysis of a diverse synapse population: proteomic imaging methods and markers. *Neuron* 68, 639–653.

Modha, D. S., and Singh, R. (2010). Network architecture of the long-distance pathways in the macaque brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13485–13490.

Ohara, S., Inoue, K., Witter, M. P., and Iijima, T. (2009). Untangling neural networks with dual retrograde transsynaptic viral infection. *Front. Neurosci.* 3:3. doi: 10.3389/neuro.01.032.2009

Ropireddy, D., and Ascoli, G. A. (2011). Potential synaptic connectivity of different neurons onto pyramidal cells in a 3D reconstruction of the rat hippocampus. *Front. Neuroinform.* 5:5. doi: 10.3389/fninf.2011.00005

Russ, T. A., Ramakrishnan, C., Hovy, E. H., Bota, M., and Burns, G. A. (2011). Knowledge engineering tools for reasoning with scientific observations and interpretations: a neural connectivity use case. *BMC Bioinformatics* 12, 351. doi: 10.1186/1471-2105-12-351

Saleem, K. S., Pauls, J. M., Augath, M., Trinath, T., Prause, B. A., Hashikawa, T., and Logothetis, N. K. (2002). Magnetic resonance imaging of neuronal connections in the macaque monkey. *Neuron* 34, 685–700.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.

Segall, J. M., Allen, E. A., Jung, R. E., Erhardt, E. B., Arja, S. K., Kiehl, K., and Calhoun, V. D. (2012). Correspondence between structure and function in the human brain at rest. *Front. Neuroinform.* 6:10. doi: 10.3389/fninf.2012.00010

Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125.

Sporns, O., Tononi, G., and Kotter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042

Staiger, J. F., Zuschratter, W., Luhmann, H. J., and Schubert, D. (2009). Local circuits targeting parvalbumin-containing interneurons in layer IV of rat barrel cortex. *Brain Struct. Funct.* 214, 1–13.

Stephan, K. E., Kamper, L., Bozkurt, A., Burns, G. A., Young, M. P., and Kotter, R. (2001). Advanced database methodology for the collation of connectivity data on the macaque brain (CoCoMac). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1159–1186.

Stephan, K. E., McIntosch, A. R., and Hilgetag, C. C. (2010). In memoriam: Rolf Kötter (1961-2010). *PLoS Comput. Biol.* 6, e1000965. doi: 10.1371/journal.pcbi.1000965.

Sugar, J., Witter, M. P., van Strien, N. M., and Cappaert, N. L. (2011). The retrosplenial cortex: intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome. *Front. Neuroinform.* 5:7. doi: 10.3389/fninf.2011.00007

Swanson, L. W. (2004). *Brain Maps: Structure of the Rat Brain.* Amsterdam: Elsevier Science B.V.

Tallis, M., Thompson, R., Russ, T. A., and Burns, G. A. (2011). Knowledge synthesis with maps of neural connectivity. *Front. Neuroinform.* 5:24. doi: 10.3389/fninf.2011.00024

Thompson, R. H., and Swanson, L. W. (2010). Hypothesis-driven structural connectivity analysis supports network over hierarchical model of brain architecture. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15235–15239.

Ugolini, G. (2011). Rabies virus as a transneuronal tracer of neuronal connections. *Adv. Virus Res.* 79, 165–202.

Van der Linden, A., Verhoye, M., Van Meir, V., Tindemans, I., Eens, M., Absil, P., and Balthazart, J. (2002). In vivo manganese-enhanced magnetic resonance imaging reveals connections and functional properties of the songbird vocal control system. *Neuroscience* 112, 467–474.

Van Essen, D. C., and Ugurbil, K. (2012). The future of the human connectome. *Neuroimage*. doi: 10.1016/j.neuroimage.2012.01.032. [Epub ahead of print].

van Strien, N. M., Cappaert, N. L., and Witter, M. P. (2009). The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network. *Nat. Rev. Neurosci.* 10, 272–282.

Wouterlood, F. G., Hartig, W., Groenewegen, H. J., and Voorn, P. (2011). Density gradients of vesicular glutamate- and GABA transporter immunoreactive boutons in calbindin- and mu-opiod receptor defined compartments in the rat striatum. *J. Comp. Neurol.* doi: 10.1002/cne.23031. [Epub ahead of print].

Yendiki, A., Panneck, P., Srinivasan, P., Stevens, A., Zöllei, L., Augustinack, J., Wang, R., Salat, D., Ehrlich, S., Behrens, T., Jbabdi, S., Gollub, R., and Fischl1, B. (2011). Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front. Neuroinform.* 5:23. doi: 10.3389/fninf.2011.00023

Zakiewicz, I. M., van Dongen, Y. C., Leergaard, T. B., and Bjaalie, J. G. (2011). Workflow and atlas system for brain-wide mapping of axonal connectivity in rat. *PLoS ONE* 6, e22669. doi: 10.1371/journal.pone.0022669

Received: 02 April 2012; accepted: 03 April 2012; published online: 01 May 2012.
*Citation: Leergaard TB, Hilgetag CC and Sporns O (2012) Mapping the connectome: multi-level analysis of brain connectivity. Front. Neuroinform. **6**:14. doi: 10.3389/fninf.2012.00014*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 14 October 2011
doi: 10.3389/fninf.2011.00023

Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy

Anastasia Yendiki[1]*, Patricia Panneck[2], Priti Srinivasan[1], Allison Stevens[1], Lilla Zöllei[1], Jean Augustinack[1], Ruopeng Wang[1], David Salat[1], Stefan Ehrlich[3,4], Tim Behrens[5], Saad Jbabdi[5], Randy Gollub[1,4] and Bruce Fischl[1,6]

[1] *Department of Radiology, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*
[2] *Charité – Universitätsmedizin, Berlin, Germany*
[3] *Translational Developmental Neuroscience Section, Department of Child and Adolescent Psychiatry, University Hospital Carl Gustav Carus, Dresden University of Technology, Dresden, Germany*
[4] *Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*
[5] *Department of Clinical Neurology, Centre for Functional MRI of the Brain, University of Oxford, Oxford, UK*
[6] *Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA*

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Simon B. Eickhoff, Institut for Medicine, Germany
Marc Tittgemeyer, Max-Planck-Institute for Neurological Research, Germany

****Correspondence:***
Anastasia Yendiki, Martinos Center for Biomedical Imaging, 149 13th Street Suite 2301, Charlestown, MA 02129, USA.
e-mail: ayendiki@nmr.mgh.harvard.edu

We have developed a method for automated probabilistic reconstruction of a set of major white-matter pathways from diffusion-weighted MR images. Our method is called TRACULA (TRActs Constrained by UnderLying Anatomy) and utilizes prior information on the anatomy of the pathways from a set of training subjects. By incorporating this prior knowledge in the reconstruction procedure, our method obviates the need for manual interaction with the tract solutions at a later stage and thus facilitates the application of tractography to large studies. In this paper we illustrate the application of the method on data from a schizophrenia study and investigate whether the inclusion of both patients and healthy subjects in the training set affects our ability to reconstruct the pathways reliably. We show that, since our method does not constrain the exact spatial location or shape of the pathways but only their trajectory relative to the surrounding anatomical structures, a set a of healthy training subjects can be used to reconstruct the pathways accurately in patients as well as in controls.

Keywords: tractography, diffusion MRI, white matter

1. INTRODUCTION

Diffusion MRI has become an important tool in the study of a wide range of diseases affecting the brain, as it allows us to probe the shape and integrity of the white-matter pathways that connect functionally related cortical and subcortical regions. Although it is possible to compare diffusion measures between populations on a voxel-by-voxel basis, more specific hypotheses on disease progression can be tested if aggregate measures can be computed for specific pathways that are known or assumed to serve different brain networks.

Several diffusion tractography methods have been proposed over the years to reconstruct white-matter pathways. Most early methods were deterministic and followed the *streamline* approach, which modeled a path as a one-dimensional curve. The curve was grown from a starting point by taking steps in directions that were determined by the diffusion orientation in the underlying voxels (Conturo et al., 1999; Mori et al., 1999; Basser et al., 2000; Poupon et al., 2000; Lazar et al., 2003). Other deterministic methods were *volumetric*, modeling the path as a volume, and allowing it to grow in three dimensions (Jones et al., 1999; O'Donnell et al., 2002; Parker et al., 2002; Jackowski et al., 2005; Pichon et al., 2005). Both streamline and volumetric approaches were *local*, in the sense that the algorithm considered the image data at a single location to determine how to grow the path at each step. Statistical extensions to local streamline tractography were introduced to model uncertainty in the image data by drawing samples from an assumed local distribution of diffusion directions at each voxel (Behrens et al., 2003; Hagmann et al., 2003; Cook et al., 2005; Parker and Alexander, 2005; Friman et al., 2006) or by boot-strapping (Jones and Pierpaoli, 2005; Lazar and Alexander, 2005).

Local tractography algorithms, whether deterministic or probabilistic, are best suited for exploring all possible connections from one brain region, which is used as the tractography seed, to any other region. However, if the goal is to isolate specific white-matter pathways, the required post-processing of the streamlines poses various challenges. Typically a user with substantial neuroanatomical expertise needs to interact manually with the data on a pathway-by-pathway and subject-by-subject basis. For example, thresholds on the curvature of each pathway need to be adjusted by trial-and-error and regions that each pathway does or does not intersect need to be defined. This makes tractography studies time-consuming and compromises their robustness and reliability. Even if an automated method is used to cluster the streamlines into larger bundles *a posteriori* (O'Donnell and Westin, 2007; Maddah et al., 2008; Wassermann et al., 2010), the results are largely dependent on the quality of the original streamlines. For example streamline tractography might miss a sparser pathway if

it is dominated by other, denser pathways that intersect it or that originate in the same region.

Global tractography methods were suggested as an alternative approach to address the problem of identifying specific white-matter pathways (Fletcher et al., 2007; Jbabdi et al., 2007; Melonakos et al., 2007). The global approach defines both end regions where the pathway is thought to terminate and searches the space of all possible connections between these two regions for the connection that best fits the data. Thus the entire pathway is estimated at once, rather than step-by-step. The solution is symmetric with respect to the two end regions, instead of treating one as the "seed" and the other as the "target." Since global optimization integrates along the length of the pathway, it is less sensitive to localized regions of high uncertainty (e.g., pathway crossings) than the streamline approach. A challenge with global tractography is the size of the solution space, which consists of all possible connections between two regions. Although the pathway is typically parameterized in some way to contain the size of this space, searching through it remains cumbersome and sensitive to initialization, especially for large end regions.

To address these issues, we have developed TRACULA (TRActs Constrained by UnderLying Anatomy), a method for automated reconstruction of major white-matter pathways that is based on the global probabilistic approach of Jbabdi et al. (2007) and utilizes prior information on the anatomy of the pathways from a set of training subjects. Once the pathways have been labeled manually in the training set, their trajectories are combined with an automatic anatomical segmentation of the same subject (Dale et al., 1999; Fischl et al., 1999a,b, 2002, 2004a,b; Fischl and Dale, 2000) to derive a description of the pathways in terms of the structures that they intersect and neighbor. The knowledge on path anatomy that is extracted from the training set is then used to initialize a global probabilistic tractography algorithm and also to constrain its search space by penalizing connections that do not match our prior anatomical knowledge. This allows the algorithm to reconstruct the pathways reliably in a novel subject with no manual intervention, facilitating the analysis of large data sets.

An important question regarding our method is whether a training set consisting entirely of healthy subjects can be used to reconstruct pathways in a diseased population. As a test case, we applied our method to a schizophrenia study, using training sets with different proportions of patients and healthy controls. Several studies of schizophrenia using diffusion MRI have been published to date (see, e.g., Kubicki et al., 2007 for a review). Although several early region-based studies showed anisotropy decreases in patients compared to controls (Lim et al., 1999; Foong et al., 2000; Agartz et al., 2001; Ardekani et al., 2003), others did not find such a decrease (Steel et al., 2001; Kubicki et al., 2002, 2003; Begré et al., 2003). Studies of how white-matter integrity relates to age in schizophrenia patients vs. controls have yielded contradictory results (Jones et al., 2006; Mori et al., 2007; Rosenberger et al., 2008; Voineskos et al., 2010). Although any discrepancies between studies are likely partly due to differences in data acquisition and variability in disease subtypes, part of the challenge has also been defining the regions of interest in a manner that is accurate and repeatable across subjects and studies. Thus investigators have been turning increasingly to tractography for better localization of the effects of schizophrenia in specific pathways (e.g., Buchsbaum et al., 2006; Price et al., 2008; Jeong et al., 2009; Kubicki et al., 2009, 2011; Oh et al., 2009; Skudlarski et al., 2010; Whitford et al., 2010). In this work we show that our automated method for reconstructing white-matter fascicles can be applied to data from schizophrenia patients, even if the training subjects are healthy. This development should allow automatic tractography analyses of even larger data sets to investigate subtle changes in specific fascicles, not only in schizophrenia but in a wide variety of neurological disorders, as well as brain development and aging.

2. MATERIALS AND METHODS

2.1. IMAGE DATA

We used image data from 34 schizophrenia patients (ages 37 ± 10, 9 female) and 33 healthy controls (ages 42 ± 10, 14 female). The data was all collected at MGH as part of a multi-site MIND Clinical Imaging Consortium (Magnotta et al., 2008; Roffman et al., 2008; Ehrlich et al., 2010; White et al., 2011). Patients had to meet DSM-IV diagnostic criteria for schizophrenia. Information on their average duration of illness, symptoms, and antipsychotic medication history is given in **Table 1**. More details on the multi-site patient population that this data set is part of can be found in Ehrlich et al. (2010). Healthy controls had no history of psychiatric diagnosis and were matched to the patient cohort for age, gender, and parental education. Exclusion criteria for both patients and controls were IQ lower than 70 based on a standardized IQ test, history of a head injury resulting in prolonged loss of consciousness, neurosurgical procedure, neurological disease, history of skull fracture, severe or disabling medical conditions, or any contraindication for MRI scanning. All subjects spoke English as their native language. They provided informed consent to participate in the study in accordance with MGH Internal Review Board regulations.

The subjects were scanned in a 1.5 T Siemens scanner with an 8-channel head coil. Diffusion-weighted images were acquired with axial in-plane isotropic resolution 2 mm, slice thickness 2 mm, $128 \times 128 \times 60$ image matrix, TR= 8900 ms, TE = 80 ms, NEX = 1, BW = 1860 Hz/pixel, GRAPPA acceleration factor 2. The series included images acquired with diffusion weighting along 60 non-collinear directions ($b = 700\ \mathrm{sm}^{-2}$), and 10 images acquired without diffusion weighting ($b = 0$). T_1-weighted images were acquired in the same session with an oblique axial GRE sequence, in-plane isotropic resolution 0.625 mm, slice thickness 1.5 mm, $256 \times 256 \times 144$ image matrix, TR = 12 ms, TE = 4.76 ms, FA = 20, NEX = 3, BW = 110 Hz/pixel.

2.2. IMAGE PREPROCESSING

We used a standard method, available in FSL[1], for mitigating distortions induced by eddy currents and motion by registering the diffusion-weighted to the $b = 0$ images. For each subject, we registered the $b = 0$ image to the T_1-weighted image by an affine registration method that seeks to maximize the intensity contrast of the $b = 0$ image across the cortical gray/white boundary, which

[1] http://www.fmrib.ox.ac.uk/fsl

Table 1 | Details on symptoms and medication history of schizophrenia patient population.

	Mean and standard deviation
Duration of illness [years]	14 ± 12
Positive symptom composite score	4.82 ± 3.29
Negative symptom composite score	7.03 ± 4.48
Disorganized symptom composite score	1.26 ± 1.68
Lifetime antipsychotic exposure [CPZ dose years]	78.80 ± 180.16
Current antipsychotic dose [CPZ units]	539.53 ± 497.79

Positive symptom composite score: Sum of the global rating of severity of delusions and the global rating of severity of hallucinations from the scale for the assessment of positive symptoms SAPS; (Andreasen, 1984). Negative symptom composite score: Sum of the values from the global rating of affective flattening, the global rating of alogia, the global rating of avolition-apathy, and the global rating of anhedonia-asociality from the scale for the assessment of negative symptoms SANS; (Andreasen, 1983). Disorganized symptom composite score: Sum of the values from the global rating of severity of bizarre behavior and the global rating of positive formal thought disorder from the scale for the assessment of positive symptoms SAPS; (Andreasen, 1984). Cumulative antipsychotic drug exposures: Given in chlorpromazine dose years, where 1 dose year = 100 chlorpromazine equivalents per day for 1 year. Current antipsychotic drug dose: Given in chlorpromazine units. Antipsychotic history was collected as part of the psychiatric assessment using the PSYCH instrument (Andreasen, 1987). Cumulative and current antipsychotic exposure was calculated using the chlorpromazine (CPZ) conversion factors of Andreasen et al. (2010).

is obtained from the T_1 scan (Greve and Fischl, 2009). We registered each individual's T_1-weighted image to the 1 mm-resolution MNI-152 atlas (Talairach and Tournoux, 1988), using affine registration (Jenkinson et al., 2002). We used the automated tools in FreeSurfer[2] to obtain a cortical parcellation and subcortical segmentation for each subject from its T_1-weighted image (Dale et al., 1999; Fischl et al., 1999a,b, 2002, 2004a,b; Fischl and Dale, 2000).

2.3. MANUAL LABELING

Our automated tractography method relies on prior anatomical information derived from a set of training subjects. We obtained this training data by labeling a set of major white-matter pathways manually in each subject from our cohort. The manual labeling was performed on the eddy-current corrected diffusion images in Trackvis[3]. Conventional deterministic streamline tractography was performed on the whole brain using the FACT method (Mori et al., 1999). Then an expert interacted with the streamlines in Trackvis to isolate the ones belonging to specific white-matter pathways. For each pathway the expert drew at least two regions of interest (ROIs) in anatomical locations that the pathway is known to traverse. We followed an established protocol for identifying these locations and drawing the ROIs (Wakana et al., 2007). Additional ROIs were placed as needed to eliminate streamlines that did not belong to the pathway of interest or to cut streamlines where they merged erroneously with other pathways. Most ROIs were hand-drawn on single slices of the individual's fractional anisotropy (FA) map, except for the end ROIs for the CST, which came from the FreeSurfer anatomical segmentation.

This was done for all the pathways listed in Wakana et al. (2007) except for the inferior fronto-occipital fasciculus, which we chose not to label due to the controversy surrounding its existence as a separate fascicle (Schmahmann and Pandya, 2007). The pathways that we did label were:

- Corticospinal tract (CST)
- Inferior longitudinal fasciculus (ILF)
- Uncinate fasciculus (UNC)
- Anterior thalamic radiation (ATR)
- Cingulum – cingulate gyrus (supracallosal) bundle (CCG)
- Cingulum – angular (infracallosal) bundle (CAB)
- Superior longitudinal fasciculus – parietal bundle (SLFP)
- Superior longitudinal fasciculus – temporal bundle (SLFT)
- Corpus callosum – forceps major (FMAJ)
- Corpus callosum – forceps minor (FMIN)

Based on the subdivision of the SLF that has been suggested in the literature (Makris et al., 2005), the SLFP and SLFT above correspond most closely to SLF III and the arcuate fasciculus, respectively. Except for FMAJ and FMIN, which are interhemispheric connections, all other pathways were labeled on the left and right hemisphere. Thus we ended up with a total of 18 pathways per subject. **Figure 1** shows an example of a full set of manually labeled pathways and all the ROIs that were drawn for the labeling.

We assessed the intra- and inter-rater reliability of the manual labeling method in the left and right uncinate. The uncinate was labeled twice by rater 1 and once each by raters 2 and 3 in 10 healthy subjects. Intra-rater reliability was quantified as the modified Hausdorff distance between the two labels of the same pathway produced by rater 1. Inter-rater reliability was quantified as the modified Hausdorff distance between labels of the same pathway produced by raters 1 and 2 or raters 1 and 3. We define the modified Hausdorff distance between two labels as the minimum distance of each point on one label from the other label, averaged over all points on the two labels. The means and standard errors of the distances over the 10 subjects are shown in **Figure 2**.

2.4. AUTOMATED TRACTOGRAPHY

Our method for automated reconstruction of white-matter pathways is based on the Bayesian framework for global tractography proposed in Jbabdi et al. (2007). In this framework the unknown pathway $\boldsymbol{\mathcal{F}}$ in any new test subject is estimated from the diffusion-weighted images $\boldsymbol{Y}$ of that subject via the posterior probability distribution of $\boldsymbol{\mathcal{F}}$ given $\boldsymbol{Y}$,

$$p(\boldsymbol{\mathcal{F}}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\boldsymbol{\mathcal{F}})p(\boldsymbol{\mathcal{F}}). \quad (1)$$

We can think of the likelihood $p(\boldsymbol{Y}|\boldsymbol{\mathcal{F}})$ as the variability in the measured data given the shape of the pathway in the specific subject and the prior distribution $p(\boldsymbol{\mathcal{F}})$ as the variability in the pathway shape from subject to subject. Therefore the likelihood represents uncertainty in the data due to measurement noise and the prior represents uncertainty due to individual anatomical variation.

[2]http://surfer.nmr.mgh.harvard.edu
[3]http://trackvis.org

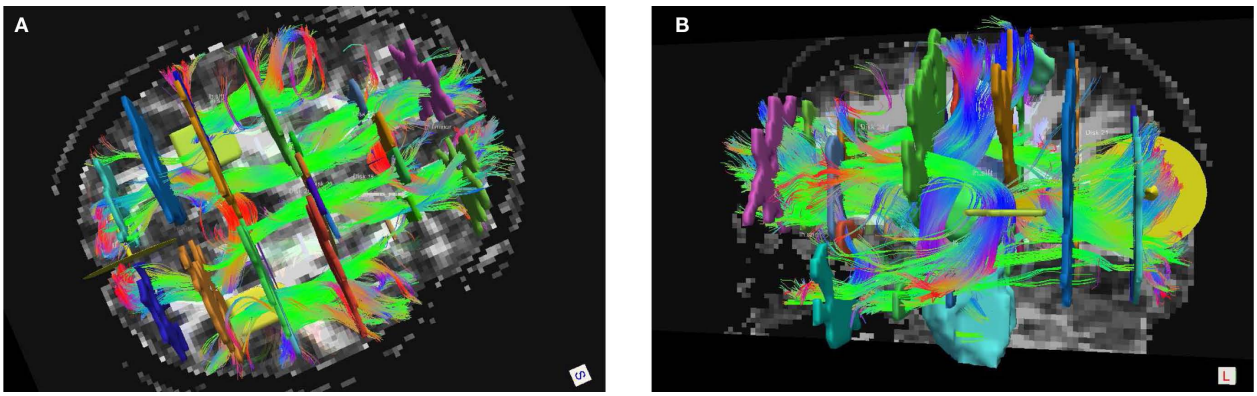


FIGURE 1 | Manual labeling of all 18 pathways in a single subject, shown in Trackvis with the disks and manually drawn ROIs that were used to define the pathways. An axial **(A)** and sagittal **(B)** view is shown, with some ROIs hidden in each case to provide an unobstructed view.

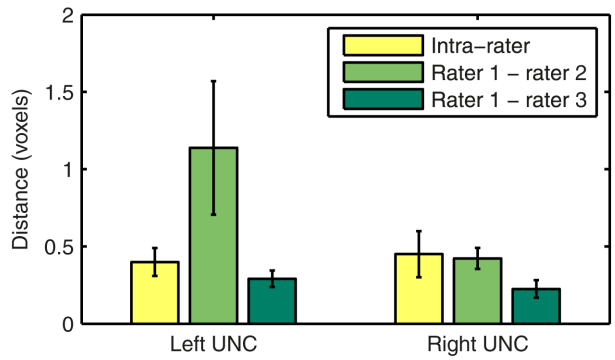


FIGURE 2 | Intra- and inter-rater reliability of the manual labeling of the left and right uncinate. Reliability is quantified as the modified Hausdorff distance between two labels of the same pathway produced by the same rater or by two different raters.

In our approach we use the same formulation for the likelihood $p(\boldsymbol{Y}|\boldsymbol{\mathcal{F}})$ as Jbabdi et al. (2007), which assumes Gaussian noise and uses the "ball-and-stick" model of diffusion (Behrens et al., 2003). This model allows for multiple compartments of anisotropic diffusion and one compartment of isotropic diffusion per voxel, expressing the diffusion image data at that voxel as a function of the volumes and orientations of these compartments (Behrens et al., 2007). We used the bedpostx tool in FSL to estimate the distributions of the ball-and-stick model parameters at each voxel from the diffusion data, assuming up to two anisotropic compartments per voxel.

Our departure from Jbabdi et al. (2007) is that, instead of assuming equal prior probability for all possible paths connecting two regions of interest, we use a prior of the form:

$$p(\boldsymbol{\mathcal{F}}) = p\left(\boldsymbol{\mathcal{F}}|\boldsymbol{A}, \{\boldsymbol{\mathcal{F}}_k\}_{k=1}^{N_t}, \{\boldsymbol{A}_K\}_{k=1}^{N_t}\right), \quad (2)$$

where $\boldsymbol{A}$ is the anatomical segmentation map of the test subject, $\boldsymbol{\mathcal{F}}_k$, $k=1,\ldots,N_t$, is the pathway of interest in each of the N_t training subjects, and $\boldsymbol{A}_k$, $k=1,\ldots,N_t$ the anatomical segmentation map of each training subject. Thus we allow our prior knowledge on the anatomy of the pathway in the training subjects to inform our belief on the anatomy of the pathway in the test subject.

Specifically, the information that we glean from the training set is which anatomical regions the pathway intersects and neighbors along its trajectory. For each training subject, the anatomical segmentation map $\boldsymbol{A}_k$, obtained from the T_1 image using FreeSurfer, and the pathway $\boldsymbol{\mathcal{F}}_k$, obtained from the manual labeling of streamlines in Trackvis, are coregistered using the intra- and inter-subject registration methods described in section 2.2. Once they have been mapped to the common space (here MNI space), the streamlines from the manual labeling of the training subjects are divided into N_s segments along their arc length. The number of segments N_s is chosen separately for each of the 18 pathways so that every training streamline has at least 3 voxels in each segment. For each segment $i=1,\ldots,N_s$ along the streamlines of $\boldsymbol{\mathcal{F}}_k$ we compute histograms of how often each segmentation label a occurs in the anatomical segmentation map $\boldsymbol{A}_k$ at the voxels that the streamlines traverse, or at the nearest neighboring structures of the streamlines in the left, right, anterior, posterior, superior, and inferior directions. This yields estimates of the *a priori* probability $p_i^0(a)$ that a voxel in the pathway's i-th segment intersects a segmentation label a, and of the *a priori* probabilities $p_i^L(a)$, $p_i^R(a)$, $p_i^A(a)$, $p_i^P(a)$, $p_i^I(a)$, $p_i^S(a)$ that the nearest neighboring segmentation label to a voxel in the pathway's i-th segment in the left, right, anterior, posterior, superior, and inferior directions, respectively, is label a. These probabilities form a statistical framework for introducing into the tractography algorithm the same type of anatomical knowledge that an expert would use to label the pathways manually.

The prior probability $p(\boldsymbol{\mathcal{F}}|\boldsymbol{A}, \{\boldsymbol{\mathcal{F}}_k\}_{k=1}^{N_t}, \{\boldsymbol{A}_k\}_{k=1}^{N_t})$ of a path in the test subject given the training data is computed by splitting the path into the same number of segments N_v as the training paths. Let the test path $\boldsymbol{\mathcal{F}}$ go through N_v voxels in the common space and $i(j)$, $j=1,\ldots,N_v$ be the segment along the path that the j-th voxel belongs to, where $i \in \{1,\ldots,N_s\}$. From the test subject's segmentation map $\boldsymbol{A}$ we obtain the segmentation label a_j^0 that the j-th voxel intersects and its nearest neighboring segmentation labels a_j^L, a_j^R, a_j^A, a_j^P, a_j^I, a_j^S in the left, right, anterior, posterior, superior, and inferior directions, respectively. The prior probability of the

path is assumed to be the product of the prior probabilities of each voxel along the path:

$$p\left(\boldsymbol{\mathcal{F}}|\boldsymbol{A},\{\boldsymbol{\mathcal{F}}_k\}_{k=1}^{n},\{\boldsymbol{A}_k\}_{k=1}^{n}\right)=\prod_{j=1}^{N_v}p_{i(j)}^{0}\left(a_j^{0}\right)p_{i(j)}^{L}\left(a_j^{L}\right)p_{i(j)}^{R}\left(a_j^{R}\right)$$
$$\times\, p_{i(j)}^{A}\left(a_j^{A}\right)p_{i(j)}^{P}\left(a_j^{P}\right)p_{i(j)}^{I}\left(a_j^{I}\right)p_{i(j)}^{S}\left(a_j^{S}\right) \quad (3)$$

We estimate the posterior distribution $p(\boldsymbol{\mathcal{F}} \mid \boldsymbol{Y})$ for the test subject via a Markov Chain Monte Carlo (MCMC) algorithm. The pathway $\boldsymbol{\mathcal{F}}$ is modeled as a cubic spline with a fixed number of control points. For the results shown here we used 5 control points to model all pathways. Further investigation is needed to determine the optimal number of control points, as it is possible that increasing it could yield better results for higher-curvature paths such as the corpus callosum. In addition to the estimation of anatomical priors, the training data is also used to derive the initialization of the control points and the end ROIs that are used to constrain the two end points of each pathway. The initialization is obtained by fitting a spline to the median of the training set of streamlines. The end ROIs are obtained by dilating the end points of the training streamlines and finding their intersection with the cortex of the test subject.

The MCMC algorithm generates samples from the posterior distribution $p(\boldsymbol{\mathcal{F}} \mid \boldsymbol{Y})$ of the path by perturbing the control points, thus changing the shape of the spline, and computing the likelihood and prior probability of the new spline. The likelihood expresses how well the spline fits the diffusion data, that is, how closely the orientation of the spline at each voxel that the spline goes through matches the orientation of the anisotropic diffusion compartments of the ball-and-stick model at the same voxel. The prior expresses how well the spline fits the training set, that is, how well the anatomical regions that the spline goes through or passes next to in the test subject match those found in the training subjects. In each iteration of the algorithm the control points are perturbed in random order. If the perturbed control point is one of the two end points of the path and the perturbation has placed it outside the end ROI obtained from the training set, it is rejected. Otherwise, every time a control point is perturbed, the likelihood and prior distribution is computed for every voxel along the spline. The control point perturbation and likelihood computation is performed in the native diffusion space, so that the DWI data itself does not need to be mapped to another space. However, the anatomical prior computation requires that each voxel on the spline is mapped to the common coordinate system where all training subjects and the corresponding anatomical segmentations have been normalized (here MNI space). The likelihood and prior are integrated over all voxels along the spline to compute its posterior probability, which is then compared to the posterior probability of the spline from the previous iteration to determine if the new spline will be accepted or rejected. A number of "burn-in" iterations (200 in this experiment) are performed in the beginning of the algorithm. The splines sampled during the burn-in period are discarded to ensure that the spline is initialized close to the center of the distribution. Then the main set of iterations (5000 in this experiment) are run. The splines that are sampled and accepted during this set of iterations are summed to yield an estimate of the posterior distribution of the pathway in the test subject. The optimal number of burn-in and sampling iterations that are needed to ensure convergence for each pathway is a topic for future investigation.

For this paper we investigated how the inclusion of both patients and healthy subjects in the training set affected our ability to reconstruct the pathways reliably in either population. We tested our method by a leave-N-out approach, where we performed automated tractography in each subject using the manual labels from a set of the remaining subjects as training data. We repeated this three times per subject, each time using a different combination of 30 training subjects: (i) all 30 healthy, (ii) 15 healthy and 15 diseased, and (iii) all 30 diseased. Training subjects were selected randomly from the healthy and diseased groups. We also performed the reconstruction using the healthy training data only to initialize the algorithm but not in the pathway prior. This was equivalent to assuming a uniform prior probability and relying on the likelihood term alone for estimating the pathway posterior. We compared the results by computing the distance of an automatically reconstructed pathway in the test subject to the respective manually labeled pathway for the same subject, which had been excluded from the training set.

3. RESULTS

Our method yields volumetric distributions of the pathways. As an example, **Figure 3** shows the estimated pathway distributions in three healthy controls and three schizophrenia patients, displayed as isosurfaces at 20% of the maximum value of each distribution.

Figure 4 shows plots of the distances between the manually labeled pathways and automatically reconstructed pathways that were estimated with different sets of training subjects. In each case we computed a modified Hausdorff distance between the automatically reconstructed pathway and the manually defined pathway. Before computing the modified Hausdorff distance, the distribution estimates were thresholded by masking out all values below 20% of the maximum. Thus the comparison is based on the center of the distribution and not its tails, as we expect the center and not the tails to overlap with the manual labels. In all cases, the paths reconstructed with the anatomical priors were closer to the manual labels than the ones reconstructed without prior information. The priors reduced both the mean distance and the variance of the distance from the manual labels, thus improving accuracy and robustness.

Changing the make-up of our training set did not affect this result significantly. In particular our results indicate that it is possible to reconstruct the pathways accurately in the entire study cohort using a training set consisting of healthy subjects only. We computed uncorrected p-values from T-tests on the difference in the modified Hausdorff distances between the case where a healthy-only training set was used and the other three cases. The differences between the cases using a healthy-only training set and no anatomical priors were significant ($p < 0.01$) for all pathways except for the left ILF. The differences between the cases using a healthy-only and a mixed or patient-only training set were not significant.

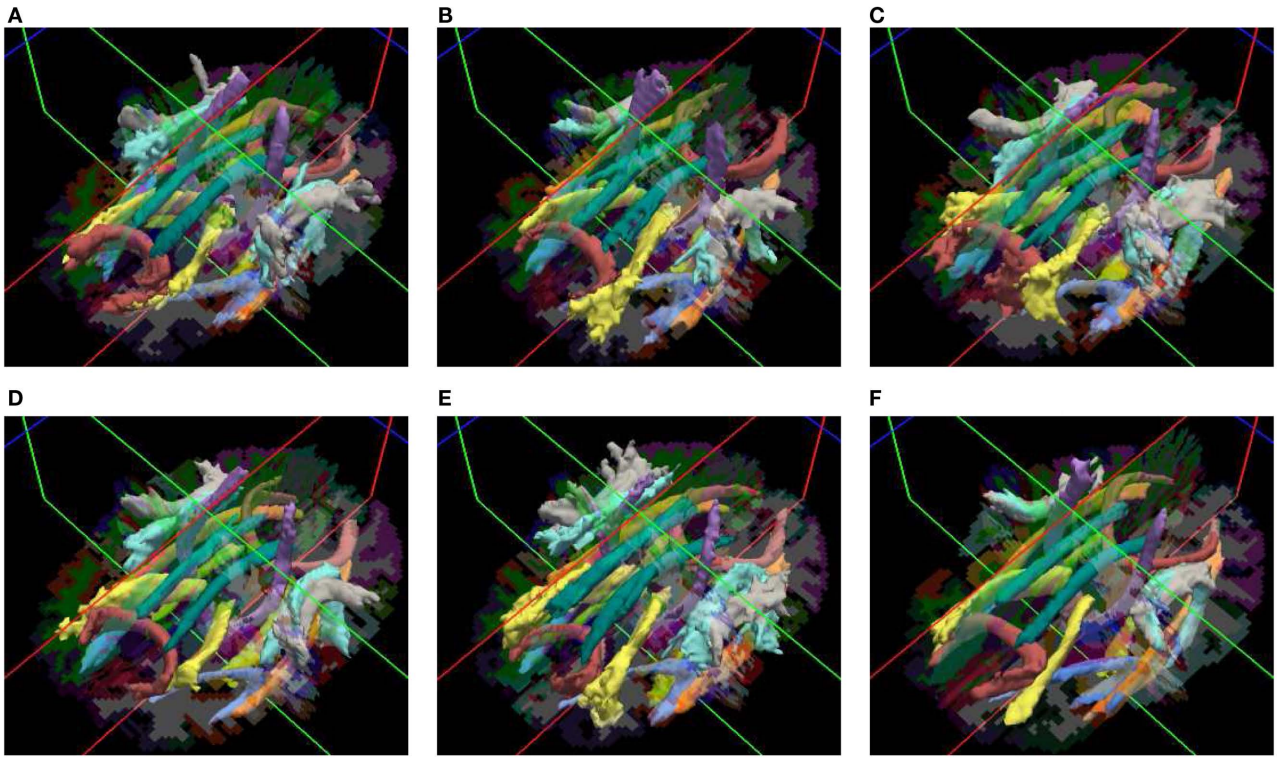


FIGURE 3 | Pathways reconstructed automatically with TRACULA in three healthy controls (A–C) and three schizophrenia patients (D–F). The posterior distribution of each pathway, estimated using 30 healthy subjects as the training set and thresholded at 20% of its maximum, is displayed as an isosurface over each individual's FreeSurfer segmentation.

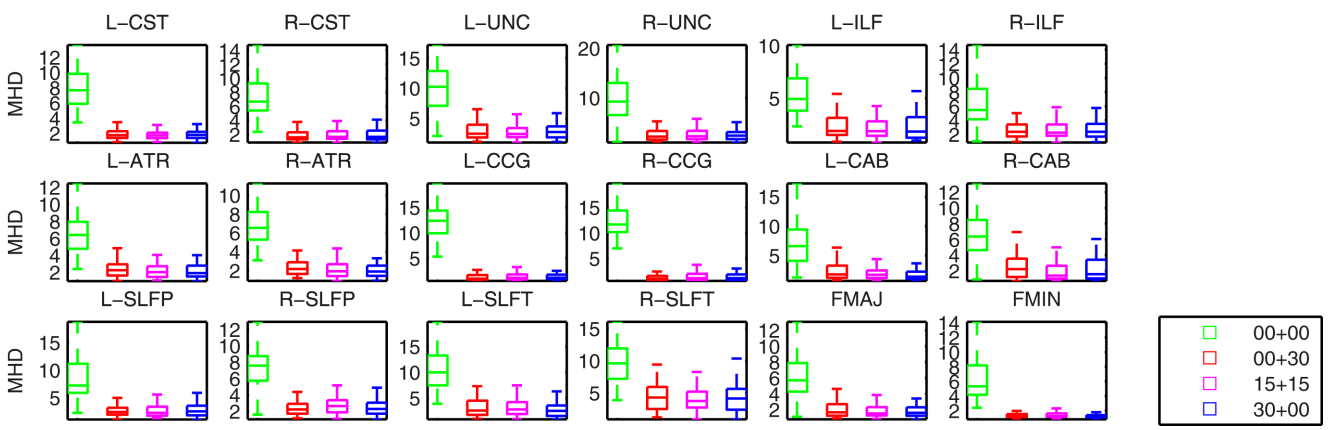


FIGURE 4 | Modified Hausdorff distances (MHD) between the manually labeled pathways and automatically reconstructed pathways. Pathway posteriors were estimated without an anatomical prior (green) and with anatomical priors derived from 3 different sets of training subjects: 30 healthy training subjects (red), 15 healthy and 15 patients (magenta), and 30 patients (blue).

Figure 5 illustrates the amount of anatomical variability that is captured by the priors for an example training set consisting of 30 healthy subjects. To quantify this variability we found the maximum fraction of training samples that corresponded to a single label, i.e., the fraction of training samples corresponding to the most commonly encountered label. The higher this fraction, the lower the variability in the segmentation labels encountered across the samples. We averaged the fraction over all segments along the length of a pathway. We calculated this average maximum fraction for the labels that intersect the pathway (f_0) and for the nearest neighboring labels in the left, right, anterior, posterior, superior, and inferior directions (f_L, f_R, f_A, f_P, f_S, and f_I, respectively). The

segmentation labels intersected by the pathways were the least variable (f_0 was 0.8 on average), as these are most commonly white matter. The fractions for the neighboring segmentation labels were 0.5 on average. The left and right CAB were the pathways with the most anatomical variability.

As an illustration of the utility of the tract-based approach in localizing white-matter degeneration in schizophrenia, and to confirm that our data set is consistent with prior findings from the literature, **Figure 6** shows group averages and corresponding standard errors of the average fractional anisotropy (FA) in each pathway. Based on uncorrected p-values from a T-test on the difference between groups, we found average FA to be significantly lower in patients compared to controls in the left UNC ($p = 0.005$), left ATR ($p = 0.019$), left CCG ($p = 0.011$), left CAB ($p = 0.006$), right SLFP ($p = 0.033$), FMAJ ($p = 0.00005$), and FMIN ($p = 0.034$), with trend toward significance in the FA reductions that were observed in the right UNC ($p = 0.067$), left ILF ($p = 0.059$), right ATR ($p = 0.061$), and right CCG ($p = 0.088$). The average FA over the entire white matter (based on white-matter masks obtained from the FreeSurfer segmentations) was also significantly reduced in patients vs. controls ($p = 0.006$). We performed similar tests on the average radial diffusivity (RD)

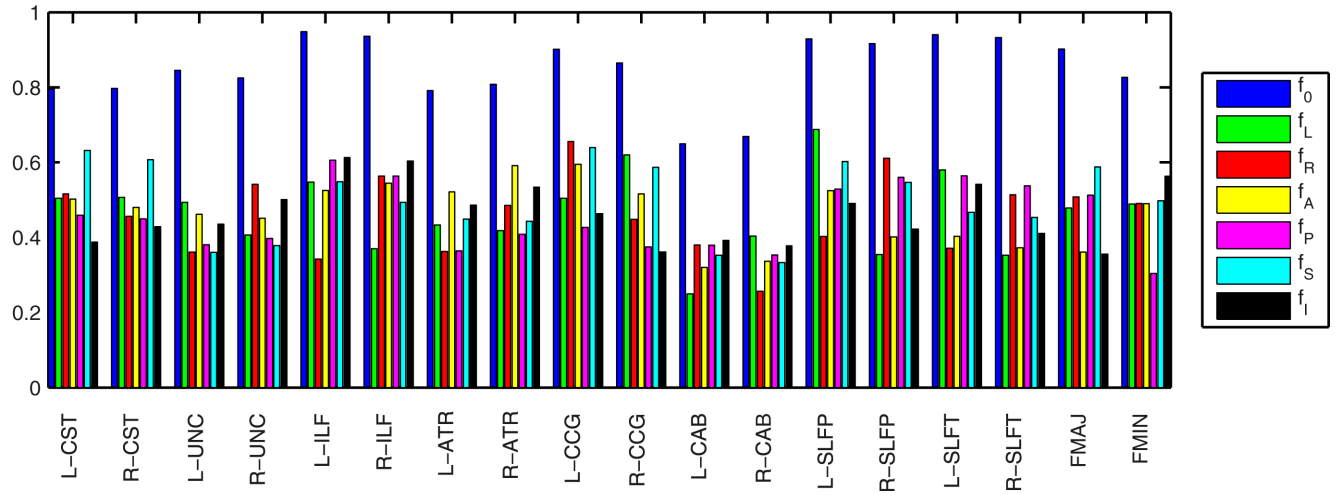


FIGURE 5 | Anatomical variability captured by the priors in an example training set of 30 healthy subjects. The plots show the fraction of training samples corresponding to the most commonly encountered label, averaged over all segments along the length of each pathway. This was computed for the labels that intersect the pathway (f_0) and for the nearest neighboring labels in the left, right, anterior, posterior, superior, and inferior directions (f_L, f_R, f_A, f_P, f_S, and f_I, respectively). The higher the fraction, the lower the variability of the labels across training samples.

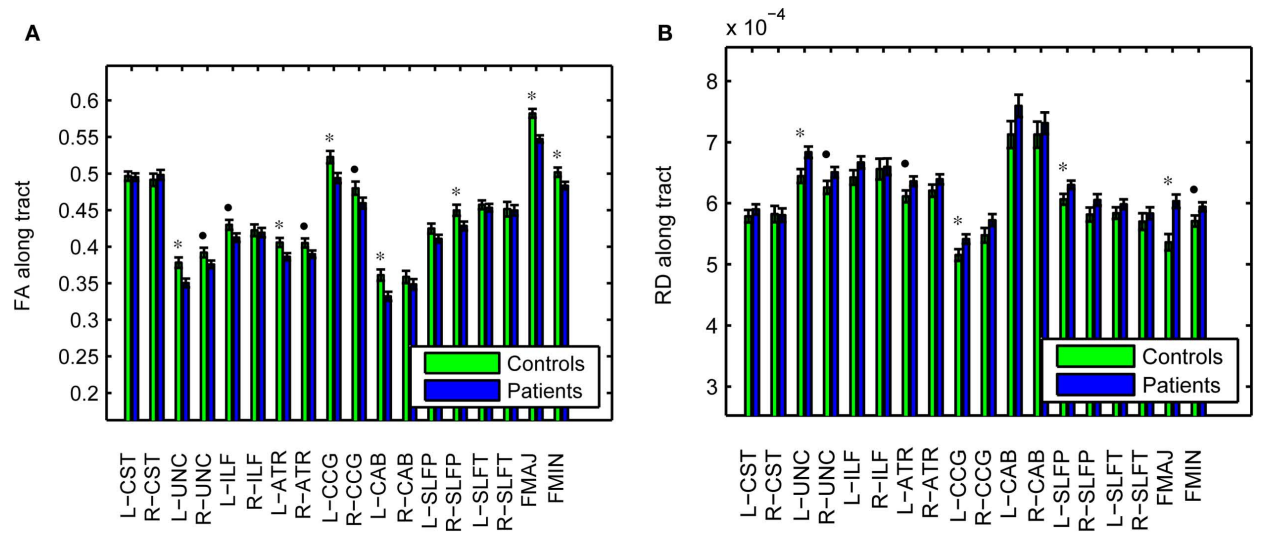


FIGURE 6 | Fractional anisotropy (A) and radial diffusivity (B), averaged over individual pathways in schizophrenia patients and healthy controls. $^*p < 0.05$, $^\bullet p < 0.1$.

and axial diffusivity (AD) in each pathway. Pathways with significantly reduced FA also exhibited increased RD in the patients compared to the controls. More specifically, significant increases of RD in patients were found in the left UNC ($p = 0.010$), left CCG ($p = 0.042$), left SLFP ($p = 0.044$), and FMAJ ($p = 0.0003$), with trend toward significance in the right UNC ($p = 0.080$), left ATR ($p = 0.057$), and FMIN ($p = 0.054$). The average RD over the whole white matter was also higher in the patients ($p = 0.014$). In contrast we found no significant changes in the average AD over individual pathways or over the entire white matter.

4. DISCUSSION

We have evaluated TRACULA, a method for automated global probabilistic tractography, on a population of schizophrenia patients and healthy controls. Our method yields volumetric distributions of major pathways in a novel subject without the need for manual intervention, thus facilitating clinical studies where large populations need to be analyzed to detect subtle changes in white-matter integrity. Our experiments showed that this approach produced results very close to those of conventional, manually assisted tractography, but without the manual editing. Further investigation is needed to determine the optimal number of subjects that should be included in the training set.

Including patients in the training set did not improve the accuracy of our results. That is, despite the relative clinical heterogeneity of our patients (**Table 1**), we were able to reconstruct pathways in this population using only healthy training subjects without a decrease in accuracy. This is not entirely surprising as our method does not constrain the exact spatial location or shape of the pathways and is thus impervious to changes in these features between populations. Our priors use only the trajectory of the training paths relative to the surrounding anatomical structures. As long as the disease that we are studying does not cause a radical reorganization of the brain and rerouting of white-matter connections, healthy training subjects could be used to reconstruct the pathways accurately in patients as well as in controls.

Several aspects of the anatomical prior computation can have an impact on the validity of our method. These include the accuracy of the automated segmentation of the T_1-weighted images, the registration of each individual's T_1-weighted and diffusion-weighted images, and the registration across individuals. The accuracy of the automated anatomical segmentation has been addressed elsewhere (Fischl et al., 2002, 2004b). The intra-subject registration method that we used here benefits from information on the gray/white-matter boundary to improve the alignment of the diffusion and T_1-weighted image. However, this alignment remains a difficult problem, most notably due to susceptibility artifacts that cause distortions in the DWIs. Thus care should be taken to minimize such distortions to improve the accuracy of the reconstructed pathways. Nevertheless it is worth noting that the range of misregistration between the T_1-weighted and DW images across the training subjects will be reflected in the anatomical priors as blurring. If any potential misregistration in the test subject is within the range present in the training set, this misregistration should be less of a problem for pathway reconstruction.

The inter-subject alignment for this study was performed by registering the subjects' T_1-weighted images to the MNI template. However, recent work from our group has shown that aligning the T_1-weighted images to each other by a combined volume- and surface-based non-linear registration can lead to improved inter-subject alignment of streamlines from deterministic tractography, when compared to affine registration (Zöllei et al., 2010). We are currently investigating the incorporation of this common coordinate system in our tractography framework to replace the MNI template. We expect that improved spatial normalization will be particularly beneficial for the initialization of the control points and for the definition of the end ROIs, as these aspects of the algorithm rely on good spatial correspondence between the training subjects and the test subject. Beyond that, however, we expect that our tractography method would be less sensitive to small misregistrations between subjects than, for example, a voxel-based comparison, since our priors use information on the surrounding anatomical structures of the pathways and not on their exact spatial location.

In the experiments presented here we evaluated the accuracy of the automated tractography by comparing it to the respective manual labels. Of course, the manual labels cannot be considered ground truth, as they are limited by the inability of the deterministic streamline tractography to reach some parts of certain pathways. For example, the more lateral terminations of the CST in the motor cortex, e.g., those corresponding to the hand region, are more challenging to trace than the more medial ones due to intersecting pathways. Similarly the frontal terminations of the SLF are longer and thus more challenging to trace than the prefrontal and premotor ones. Using a high angular-resolution model (Q-ball) instead of the tensor model to obtain the streamlines used for labeling did not yield improvements, since our data acquisition ($b = 700\ \mathrm{sm}^{-2}$, 60 directions) was suboptimal for this purpose.

However, we expect the global probabilistic approach to explore areas of lower anisotropy and tract crossings that are unreachable by deterministic tensor tractography, as long as these areas lie within the same anatomical neighborhood as the training streamlines. One reason for this is that the multi-fiber ball-and-stick model can model more than one tract orientation per voxel. Another reason is that global tractography integrates along the length of the path and would be less sensitive to a low-anisotropy crossing somewhere on that trajectory that could cause streamline tractography to terminate prematurely. Ultimately the availability of high-quality training data will be very beneficial to our method and each tractography approach, manually assisted or automatic, should be validated further by comparing it to tracer studies.

The data likelihood model that is used by our method assumes a Gaussian distribution for the DWI intensity values. This is a good approximation for magnitude images when the SNR is sufficiently high but breaks down at low SNR. To test the Gaussianity of the noise in our data, we used the DWI values in each voxel in the ventricles, where the intensity is independent of gradient direction due to isotropic diffusion. For each of these voxels we used the 60 DWI values available from the 60-direction data to estimate the SNR and test for Gaussianity using a Kolmogorov-Smirnov test. A total of 147781 voxels were tested over all subjects. The null

hypothesis of Gaussianity was rejected in only 0.1% of these tests. The average SNR was 5.5.

The data set that we chose to both train and test our method in this work was acquired with the standard DWI sequence that is used routinely to collect data for research studies at MGH. This included using the default choices for *b*-value, gradient directions, and spatial resolution. It will be important to evaluate our method further on data acquired with different acquisition parameters. Beyond the quality of the test data, the quality of the training data is crucial to our method, since the accuracy of the reconstructed pathways is strongly dependent on the accuracy of the prior information used by the algorithm. In the future, as improved acquisition methods and hardware become available, training data of higher quality can be collected and used to increase automated reconstruction accuracy in data sets of routine quality.

Tractography can be used to qualify white-matter differences between populations in much greater detail than it is possible with a voxel-based or ROI-based approach. Local tractography can handle exploratory analyses, where the anatomy of a connection is not known or the connection may not be present in all subjects. Global tractography is geared toward the reconstruction of a known connection between two end regions. A feature of the global approach is that, by constraining both end points of the pathway, it provides us with a straightforward way to parameterize the pathway by arc length. With such a parameterization one could localize effects further by comparing diffusion measures, such as FA, not only in terms of their averages over a pathway, but also as a function of position along the length of the pathway. With global tractography, in particular, we estimate the posterior distribution of each pathway, from which it is straightforward to calculate the posterior mean or maximum *a posteriori* pathway for each subject and compare FA or other measures at different locations along the arc length. Since differences may be more pronounced in a particular portion of a pathway, e.g., due to greater disorganization of connections in that portion or more crossings with another pathway, such analyses may be helpful for further interpretation of population differences.

To illustrate the validity of the data set used here, we have also presented results from a tract-based comparison of FA between the schizophrenia patients and matched controls in our cohort. A superset of this cohort, including data acquired at three additional sites, was studied previously with an ROI-based approach. FA was found to be lower in patients than controls when averaged over large regions (whole brain, frontal, parietal, occipital, and temporal lobes) (White et al., 2011). We were able to replicate this result in this much smaller data set and show significant FA reductions localized in specific pathways, as seen in **Figure 6**. Our results are consistent with prior studies on white-matter integrity in schizophrenia that have sought to localize effects in specific fascicles. In agreement to what we have found, anterior thalamic radiations (Buchsbaum et al., 2006; Oh et al., 2009), cingulum (Kubicki et al., 2003, 2005; Manoach et al., 2007; Mori et al., 2007; Nestor et al., 2007), corpus callosum (Foong et al., 2000; Agartz et al., 2001; Hubl et al., 2004; Kubicki et al., 2005; Douaud et al., 2007a; Whitford et al., 2010), inferior longitudinal fasciculus (Hubl et al., 2004; Jeong et al., 2009), superior longitudinal fasciculus (Hubl et al., 2004; Kubicki et al., 2005; Jones et al., 2006; Karlsgodt et al., 2008; Jeong et al., 2009), and uncinate (Kubicki et al., 2002; Burns et al., 2003; Mori et al., 2007; Price et al., 2008; Szeszko et al., 2008; Voineskos et al., 2010) are major sites where alterations have been reported.

Common limitations of diffusion MRI studies, including our own, are our inability to determine the exact biological causes of diffusion anisotropy changes, our difficulty in distinguishing the effects of the disease from those of medication, and the potentially increased subject motion in patients as compared to controls. Histological studies have shown several changes in the white matter of schizophrenia patients when compared to healthy subjects, including differences in myelination and neuronal arborization patterns (Davis et al., 2003; Flynn et al., 2003). Distinguishing between potential neurobiological causes based on FA changes alone is not possible. However, in combination with other measures extracted from DWIs, such as mean, radial, and axial diffusivity (Kubicki et al., 2003; Douaud et al., 2007b; Whitford et al., 2010), length of tractography streamlines (Buchsbaum et al., 2006), or even measures from magnetization transfer imaging (Kubicki et al., 2005), these findings have been hypothesized to support either demyelination or geometric disorganization as their underlying etiology. Similarly to Whitford et al. (2010), our results show increased radial but unchanged axial diffusivity in the patients, which has been interpreted as evidence of myelin abnormalities (Song et al., 2002).

Whichever the biological cause of changes in white-matter integrity measures derived from diffusion MRI, several studies have found these changes to be associated with cognitive deficits in schizophrenia patients. This includes associations with performance in attention and memory tasks (Kubicki et al., 2002, 2003, 2009, 2011; Nestor et al., 2007; Karlsgodt et al., 2008; Szeszko et al., 2008), with fMRI activation in working memory-related areas (Schlösser et al., 2007), and with fMRI time course correlations within the semantic network (Jeong et al., 2009). Such findings illustrate the potential of diffusion MRI to improve our understanding of the mechanisms of schizophrenia but they also underline the need for extracting diffusion measures specific to each affected network. Our tractography method allows the automatic extraction of such measures and can thus facilitate pathway-specific studies on larger populations than what has been possible with manually assisted tractography.

CONCLUSION

We have developed TRACULA, a method for automated tractography that uses prior information on the anatomy of white-matter pathways from a set of training subjects. We have evaluated the accuracy of the method on a population of schizophrenia patients and healthy volunteers, to determine how it is affected by the inclusion of patients in the training set. We have found that a training set consisting entirely of healthy subjects could be used to reconstruct white-matter pathways in both patients and healthy controls without compromising accuracy. Of course this conclusion cannot be generalized to every clinical population and further evaluation on other populations is warranted to determine if the training set should be tailored to specific studies.

TRACULA is available for download as part of FreeSurfer 5.1.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gary Zhang of University College London for valuable discussions on white-matter labeling that have informed this work. Funding was provided in part by NIH/NIBIB K99/R00 *Pathway to Independence* award EB008129. Additional support was provided by the NIH Blueprint for Neuroscience Research (U01-MH093765, part of the multi-institutional Human Connectome Project), the National Center for Research Resources (P41-RR14075, and the NCRR BIRN Morphometric Project BIRN002, U24-RR021382), the National Institute for Biomedical Imaging and Bioengineering (R01-EB006758), the National Institute on Aging (R01-AG022381), the National Center for Alternative Medicine (RC1-AT005728), the National Institute for Neurological Disorders and Stroke (R01-NS052585, R21-NS072652, R01-NS070963), and the Autism & Dyslexia Project funded by the Ellison Medical Foundation. This work was made possible by the resources provided by Shared Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043. The imaging data and demographic information was collected and shared by the MGH team of the MIND Clinical Imaging Consortium, funded by the U.S. Department of Energy DE-FG02-99ER62764. Tim Behrens and Saad Jbabdi were supported by the Wellcome Trust (WT088312AIA), the UK Medical Research Council (G0800578), the EU CONNECT project, and the Human Connectome Project (1U54MH091657-01) from the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research. The project CONNECT acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 238292.

REFERENCES

Agartz, I., Andersson, J. L., and Skare, S. (2001). Abnormal brain white matter in schizophrenia: a diffusion tensor imaging study. *Neuroreport* 12, 2251–2254.

Andreasen, N. C. (1983). *Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City, IA: University of Iowa.

Andreasen, N. C. (1984). *Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City, IA: University of Iowa.

Andreasen, N. C. (1987). *Psychiatric Symptoms You Currently Have – Baseline (PSYCH-BASE)*. Iowa City, IA: University of Iowa.

Andreasen, N. C., Pressler, M., Nopoulos, P., Miller, D., and Ho, B.-C. (2010). Antipsychotic dose equivalents and dose-years: a standardized method for comparing exposure to different drugs. *Biol. Psychiatry* 67, 255–262.

Ardekani, B. A., Nierenberg, J., Hoptman, M. J., Javitt, D. C., and Lim, K. O. (2003). MRI study of white matter diffusion anisotropy in schizophrenia. *Neuroreport* 14, 2025–2029.

Basser, P. J., Pajevic, S., Pierpaoli, C., Duda, J., and Aldroubi, A. (2000). In vivo fiber tractography using DT-MRI data. *Magn. Reson. Med.* 44, 625–632.

Begré, S., Federspiel, A., Kiefer, C., Schroth, G., Dierks, T., and Strik, W. K. (2003). Reduced hippocampal anisotropy related to anteriorization of alpha EEG in schizophrenia. *Neuroreport* 14, 739–742.

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34, 144–155.

Behrens, T. E. J., Woolrich, M. W., Jenkinson, M., Johansen-Berg, H., Nunes, R. G., Clare, S., Matthews, P. M., Brady, J. M., and Smith, S. M. (2003). Characterization and propagation of uncertainty in diffusion-weighted MR imaging. *Magn. Reson. Med.* 50, 1077–1088.

Buchsbaum, M. S., Schoenknecht, P., Torosjan, Y., Newmark, R., Chu, K.-W., Mitelman, S., Brickman, A. M., Shihabuddin, L., Haznedar, M. M., Hazlett, E. A., Ahmed, S., and Tang, C. (2006). Diffusion tensor imaging of frontal lobe white matter tracts in schizophrenia. *Ann. Gen. Psychiatry* 5, 19.

Burns, J., Job, D., Bastin, M. E., Whalley, H., Macgillivray, T., Johnstone, E. C., and Lawrie, S. M. (2003). Structural disconnectivity in schizophrenia: a diffusion tensor magnetic resonance imaging study. *Br. J. Psychiatry* 182, 439–443.

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H., and Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10422–10427.

Cook, P. A., Zhang, H., Avants, B. B., Yushkevich, P., Alexander, D. C., Gee, J. C., Ciccarelli, O., and Thompson, A. J. (2005). An automated approach to connectivity-based partitioning of brain structures. *Med. Image Comput. Comput. Assist. Interv.* 8(Pt 1), 164–171.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. segmentation and surface reconstruction. *Neuroimage* 9, 179–194.

Davis, K. L., Stewart, D. G., Friedman, J. I., Buchsbaum, M., Harvey, P. D., Hof, P. R., Buxbaum, J., and Haroutunian, V. (2003). White matter changes in schizophrenia: evidence for myelin-related dysfunction. *Arch. Gen. Psychiatry* 60, 443–456.

Douaud, G., Smith, S., Jenkinson, M., Behrens, T., Johansen-Berg, H., Vickers, J., James, S., Voets, N., Watkins, K., Matthews, P. M., and James, A. (2007a). Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. *Brain* 130(Pt 9), 2375–2386.

Douaud, G., Smith, S., Jenkinson, M., Behrens, T., Johansen-Berg, H., Vickers, J., James, S., Voets, N., Watkins, K., Matthews, P. M., and James, A. (2007b). Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. *Brain* 130(Pt 9), 2375–2386.

Ehrlich, S., Morrow, E. M., Roffman, J. L., Wallace, S. R., Naylor, M., Bockholt, H. J., Lundquist, A., Yendiki, A., Ho, B.-C., White, T., Manoach, D. S., Clark, V. P., Calhoun, V. D., Gollub, R. L., and Holt, D. J. (2010). The COMT Val108/158Met polymorphism and medial temporal lobe volumetry in patients with schizophrenia and healthy adults. *Neuroimage* 53, 992–1000.

Fischl, B., and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11050–11055.

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., and Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.

Fischl, B., Salat, D. H., van der Kouwe, A. J. W., Makris, N., Ségonne, F., Quinn, B. T., and Dale, A. M. (2004a). Sequence-independent segmentation of magnetic resonance images. *Neuroimage* 23(Suppl. 1), S69–S84.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. (2004b). Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.

Fischl, B., Sereno, M. I., and Dale, A. M. (1999a). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 9, 195–207.

Fischl, B., Sereno, M. I., Tootell, R. B., and Dale, A. M. (1999b). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum. Brain Mapp.* 8, 272–284.

Fletcher, P. T., Tao, R., Jeong, W.-K., and Whitaker, R. T. (2007). A volumetric approach to quantifying region-to-region white matter connectivity in diffusion tensor MRI. *Inf. Process. Med. Imaging* 20, 346–358.

Flynn, S. W., Lang, D. J., Mackay, A. L., Goghari, V., Vavasour, I. M., Whittall, K. P., Smith, G. N., Arango, V., Mann, J. J., Dwork, A. J., Falkai, P., and Honer, W. G. (2003). Abnormalities of myelination in schizophrenia detected in vivo with MRI, and post-mortem with analysis of oligodendrocyte proteins. *Mol. Psychiatry* 8, 811–820.

Foong, J., Maier, M., Clark, C. A., Barker, G. J., Miller, D. H., and Ron, M. A. (2000). Neuropathological abnormalities of the corpus callosum in schizophrenia: a diffusion tensor imaging study. *J. Neurol. Neurosurg. Psychiatr.* 68, 242–244.

Friman, O., Farnebäck, G., and Westin, C.-F. (2006). A Bayesian approach for stochastic white matter tractography. *IEEE Trans. Med. Imaging* 25, 965–978.

Greve, D. N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* 48, 63–72.

Hagmann, P., Thiran, J.-P., Jonasson, L., Vandergheynst, P., Clarke, S., Maeder, P., and Meuli, R. (2003). DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection. *Neuroimage* 19, 545–554.

Hubl, D., Koenig, T., Strik, W., Federspiel, A., Kreis, R., Boesch, C., Maier, S. E., Schroth, G., Lovblad, K., and Dierks, T. (2004). Pathways that make voices: white matter changes in auditory hallucinations. *Arch. Gen. Psychiatry* 61, 658–668.

Jackowski, M., Kao, C. Y., Qiu, M., Constable, R. T., and Staib, L. H. (2005). White matter tractography by anisotropic wavefront evolution and diffusion tensor imaging. *Med. Image Anal.* 9, 427–440.

Jbabdi, S., Woolrich, M. W., Andersson, J. L. R., and Behrens, T. E. J. (2007). A Bayesian framework for global tractography. *Neuroimage* 37, 116–129.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841.

Jeong, B., Wible, C. G., ichiro Hashimoto, R., and Kubicki, M. (2009). Functional and anatomical connectivity abnormalities in left inferior frontal gyrus in schizophrenia. *Hum. Brain Mapp.* 30, 4138–4151.

Jones, D. K., Catani, M., Pierpaoli, C., Reeves, S. J., Shergill, S. S., O'Sullivan, M., Golesworthy, P., Mcguire, P., Horsfield, M. A., Simmons, A., Williams, S. C., and Howard, R. J. (2006). Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. *Hum. Brain Mapp.* 27, 230–238.

Jones, D. K., and Pierpaoli, C. (2005). Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. *Magn. Reson. Med.* 53, 1143–1149.

Jones, D. K., Simmons, A., Williams, S. C., and Horsfield, M. A. (1999). Non-invasive assessment of axonal fiber connectivity in the human brain via diffusion tensor MRI. *Magn. Reson. Med.* 42, 37–41.

Karlsgodt, K. H., van Erp, T. G. M., Poldrack, R. A., Bearden, C. E., Nuechterlein, K. H., and Cannon, T. D. (2008). Diffusion tensor imaging of the superior longitudinal fasciculus and working memory in recent-onset schizophrenia. *Biol. Psychiatry* 63, 512–518.

Kubicki, M., Alvarado, J. L., Westin, C.-F., Tate, D. F., Markant, D., Terry, D. P., Whitford, T. J., Siebenthal, J. D., Bouix, S., McCarley, R. W., Kikinis, R., and Shenton, M. E. (2011). Stochastic tractography study of inferior frontal gyrus anatomical connectivity in schizophrenia. *Neuroimage* 55, 1657–1664.

Kubicki, M., McCarley, R., Westin, C.-F., Park, H.-J., Maier, S., Kikinis, R., Jolesz, F. A., and Shenton, M. E. (2007). A review of diffusion tensor imaging studies in schizophrenia. *J. Psychiatr. Res.* 41, 15–30.

Kubicki, M., Niznikiewicz, M., Connor, E., Nestor, P., Bouix, S., Dreusicke, M., Kikinis, R., McCarley, R., and Shenton, M. (2009). Relationship between white matter integrity, attention, and memory in schizophrenia: a diffusion tensor imaging study. *Brain Imaging Behav.* 3, 191–201.

Kubicki, M., Park, H., Westin, C. F., Nestor, P. G., Mulkern, R. V., Maier, S. E., Niznikiewicz, M., Connor, E. E., Levitt, J. J., Frumin, M., Kikinis, R., Jolesz, F. A., McCarley, R. W., and Shenton, M. E. (2005). DTI and MTR abnormalities in schizophrenia: analysis of white matter integrity. *Neuroimage* 26, 1109–1118.

Kubicki, M., Westin, C.-F., Maier, S. E., Frumin, M., Nestor, P. G., Salisbury, D. F., Kikinis, R., Jolesz, F. A., McCarley, R. W., and Shenton, M. E. (2002). Uncinate fasciculus findings in schizophrenia: a magnetic resonance diffusion tensor imaging study. *Am. J. Psychiatry* 159, 813–820.

Kubicki, M., Westin, C.-F., Nestor, P. G., Wible, C. G., Frumin, M., Maier, S. E., Kikinis, R., Jolesz, F. A., McCarley, R. W., and Shenton, M. E. (2003). Cingulate fasciculus integrity disruption in schizophrenia: a magnetic resonance diffusion tensor imaging study. *Biol. Psychiatry* 54, 1171–1180.

Lazar, M., and Alexander, A. L. (2005). Bootstrap white matter tractography (BOOT-TRAC). *Neuroimage* 24, 524–532.

Lazar, M., Weinstein, D. M., Tsuruda, J. S., Hasan, K. M., Arfanakis, K., Meyerand, M. E., Badie, B., Rowley, H. A., Haughton, V., Field, A., and Alexander, A. L. (2003). White matter tractography using diffusion tensor deflection. *Hum. Brain Mapp.* 18, 306–321.

Lim, K. O., Hedehus, M., Moseley, M., de Crespigny, A., Sullivan, E. V., and Pfefferbaum, A. (1999). Compromised white matter tract integrity in schizophrenia inferred from diffusion tensor imaging. *Arch. Gen. Psychiatry* 56, 367–374.

Maddah, M., Grimson, W. E. L., Warfield, S. K., and Wells, W. M. (2008). A unified framework for clustering and quantitative analysis of white matter fiber tracts. *Med. Image Anal.* 12, 191–202.

Magnotta, V. A., Adix, M. L., Caprahan, A., Lim, K., Gollub, R., and Andreasen, N. C. (2008). Investigating connectivity between the cerebellum and thalamus in schizophrenia using diffusion tensor tractography: a pilot study. *Psychiatry Res.* 163, 193–200.

Makris, N., Kennedy, D. N., McInerney, S., Sorensen, A. G., Wang, R., Caviness, V. S., and Pandya, D. N. (2005). Segmentation of subcomponents within the superior longitudinal fascicle in humans: a quantitative, in vivo, DT-MRI study. *Cereb. Cortex* 15, 854–869.

Manoach, D. S., Ketwaroo, G. A., Polli, F. E., Thakkar, K. N., Barton, J. J. S., Goff, D. C., Fischl, B., Vangel, M., and Tuch, D. S. (2007). Reduced microstructural integrity of the white matter underlying anterior cingulate cortex is associated with increased saccadic latency in schizophrenia. *Neuroimage* 37, 599–610.

Melonakos, J., Mohan, V., Niethammer, M., Smith, K., Kubicki, M., and Tannenbaum, A. (2007). Finsler tractography for white matter connectivity analysis of the cingulum bundle. *Med. Image Comput. Comput. Assist. Interv.* 10(Pt 1), 36–43.

Mori, S., Crain, B. J., Chacko, V. P., and van Zijl, P. C. (1999). Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* 45, 265–269.

Mori, T., Ohnishi, T., Hashimoto, R., Nemoto, K., Moriguchi, Y., Noguchi, H., Nakabayashi, T., Hori, H., Harada, S., Saitoh, O., Matsuda, H., and Kunugi, H. (2007). Progressive changes of white matter integrity in schizophrenia revealed by diffusion tensor imaging. *Psychiatry Res.* 154, 133–145.

Nestor, P. G., Kubicki, M., Spencer, K. M., Niznikiewicz, M., McCarley, R. W., and Shenton, M. E. (2007). Attentional networks and cingulum bundle in chronic schizophrenia. *Schizophr. Res.* 90, 308–315.

O'Donnell, L., Haker, S., and Westin, C. F. (2002). New approaches to estimation of white matter connectivity in diffusion tensor MRI: Elliptic PDEs and geodesics in a tensor-warped space. *Med. Image Comput. Comput. Assist. Interv.* 5, 459–466.

O'Donnell, L. J., and Westin, C.-F. (2007). Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Trans. Med. Imaging* 26, 1562–1575.

Oh, J. S., Kubicki, M., Rosenberger, G., Bouix, S., Levitt, J. J., McCarley, R. W., Westin, C.-F., and Shenton, M. E. (2009). Thalamo-frontal white matter alterations in chronic schizophrenia: a quantitative diffusion tractography study. *Hum. Brain Mapp.* 30, 3812–3825.

Parker, G. J. M., and Alexander, D. C. (2005). Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 893–902.

Parker, G. J. M., Wheeler-Kingshott, C. A. M., and Barker, G. J. (2002). Estimating distributed anatomical connectivity using fast marching methods and diffusion tensor imaging. *IEEE Trans. Med. Imaging* 21, 505–512.

Pichon, E., Westin, C.-F., and Tannenbaum, A. R. (2005). A Hamilton-Jacobi-Bellman approach to high angular resolution diffusion tractography. *Med. Image Comput. Comput. Assist. Interv.* 8(Pt 1), 180–187.

Poupon, C., Clark, C. A., Frouin, V., Régis, J., Bloch, I., Bihan, D. L., and Mangin, J. (2000). Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. *Neuroimage* 12, 184–195.

Price, G., Cercignani, M., Parker, G. J. M., Altmann, D. R., Barnes, T. R. E., Barker, G. J., Joyce, E. M., and Ron, M. A. (2008). White matter tracts in first-episode psychosis: a DTI tractography study of the uncinate fasciculus. *Neuroimage* 39, 949–955.

Roffman, J. L., Gollub, R. L., Calhoun, V. D., Wassink, T. H., Weiss, A. P., Ho, B. C., White, T., Clark, V. P., Fries, J., Andreasen, N. C., Goff, D. C., and Manoach, D. S. (2008). MTHFR 677C → T genotype disrupts prefrontal function in schizophrenia through an interaction with COMT 158Val → Met. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17573–17578.

Rosenberger, G., Kubicki, M., Nestor, P. G., Connor, E., Bushell, G. B., Markant, D., Niznikiewicz, M., Westin, C.-F., Kikinis, R., Saykin, A. J., McCarley, R. W., and Shenton, M. E. (2008). Age-related deficits in fronto-temporal connections in schizophrenia: a diffusion tensor imaging study. *Schizophr. Res.* 102, 181–188.

Schlösser, R. G. M., Nenadic, I., Wagner, G., Güllmar, D., von Consbruch, K., Köhler, S., Schultz, C. C., Koch, K., Fitzek, C., Matthews, P. M., Reichenbach, J. R., and Sauer, H. (2007). White matter abnormalities and brain activation in schizophrenia: a combined DTI and Fmri study. *Schizophr. Res.* 89, 1–11.

Schmahmann, J. D., and Pandya, D. N. (2007). The complex history of the fronto-occipital fasciculus. *J. Hist. Neurosci.* 16, 362–377.

Skudlarski, P., Jagannathan, K., Anderson, K., Stevens, M. C., Calhoun, V. D., Skudlarska, B. A., and Pearlson, G. (2010). Brain connectivity is not only lower but different in schizophrenia: a combined anatomical and functional approach. *Biol. Psychiatry* 68, 61–69.

Song, S.-K., Sun, S.-W., Ramsbottom, M. J., Chang, C., Russell, J., and Cross, A. H. (2002). Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. *Neuroimage* 17, 1429–1436.

Steel, R. M., Bastin, M. E., McConnell, S., Marshall, I., Cunningham-Owens, D. G., Lawrie, S. M., Johnstone, E. C., and Best, J. J. (2001). Diffusion tensor imaging (DTI) and proton magnetic resonance spectroscopy (1H MRS) in schizophrenic subjects and normal controls. *Psychiatry Res.* 106, 161–170.

Szeszko, P. R., Robinson, D. G., Ashtari, M., Vogel, J., Betensky, J., Sevy, S., Ardekani, B. A., Lencz, T., Malhotra, A. K., McCormack, J., Miller, R., Lim, K. O., Gunduz-Bruce, H., Kane, J. M., and Bilder, R. M. (2008). Clinical and neuropsychological correlates of white matter abnormalities in recent onset schizophrenia. *Neuropsychopharmacology* 33, 976–984.

Talairach, J., and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain.* New York, NY: Thieme Medical Publishers.

Voineskos, A. N., Lobaugh, N. J., Bouix, S., Rajji, T. K., Miranda, D., Kennedy, J. L., Mulsant, B. H., Pollock, B. G., and Shenton, M. E. (2010). Diffusion tensor tractography findings in schizophrenia across the adult lifespan. *Brain* 133(Pt 5), 1494–1504.

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., Dubey, P., Blitz, A., van Zijl, P., and Mori, S. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 36, 630–644.

Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., and Deriche, R. (2010). Unsupervised white matter fiber clustering and tract probability map generation: applications of a gaussian process framework for white matter fibers. *Neuroimage* 51, 228–241.

White, T., Magnotta, V. A., Bockholt, H. J., Williams, S., Wallace, S., Ehrlich, S., Mueller, B. A., Ho, B.-C., Jung, R. E., Clark, V. P., Lauriello, J., Bustillo, J. R., Schulz, S. C., Gollub, R. L., Andreasen, N. C., Calhoun, V. D., and Lim, K. O. (2011). Global white matter abnormalities in schizophrenia: a multisite diffusion tensor imaging study. *Schizophr. Bull.* 37, 222–232.

Whitford, T. J., Kubicki, M., Schneiderman, J. S., O'Donnell, L. J., King, R., Alvarado, J. L., Khan, U., Markant, D., Nestor, P. G., Niznikiewicz, M., McCarley, R. W., Westin, C.-F., and Shenton, M. E. (2010). Corpus callosum abnormalities and their association with psychotic symptoms in patients with schizophrenia. *Biol. Psychiatry* 68, 70–77.

Zöllei, L., Stevens, A., Huber, K., Kakunoori, S., and Fischl, B. (2010). Improved tractography alignment using combined volumetric and surface registration. *Neuroimage* 51, 206–213.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 March 2011; accepted: 23 September 2011; published online: 14 October 2011.

Citation: Yendiki A, Panneck P, Srinivasan P, Stevens A, Zöllei L, Augustinack J, Wang R, Salat D, Ehrlich S, Behrens T, Jbabdi S, Gollub R and Fischl B (2011) Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. Front. Neuroinform. ***5****:23. doi: 10.3389/fninf.2011.00023*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 15 July 2011
doi: 10.3389/fninf.2011.00006

On the influence of amplitude on the connectivity between phases

Andreas Daffertshofer* and Bernadette C. M. van Wijk

Research Institute MOVE, VU University Amsterdam, Amsterdam, Netherlands

Edited by:
Olaf Sporns, Indiana University, USA

Reviewed by:
Joana R. B. Cabral, Universitat Pompeu Fabra, Spain
Juergen Kurths, Humboldt Universität, Germany

****Correspondence:***
Andreas Daffertshofer, Research Institute MOVE, VU University Amsterdam, Van der Boechorststraat 9, 1081 BT Amsterdam, Netherlands.
e-mail: a.daffertshofer@vu.nl

In recent studies, functional connectivities have been reported to display characteristics of complex networks that have been suggested to concur with those of the underlying structural, i.e., anatomical, networks. Do functional networks always agree with structural ones? In all generality, this question can be answered with "no": for instance, a fully synchronized state would imply isotropic homogeneous functional connections irrespective of the "real" underlying structure. A proper inference of structure from function and *vice versa* requires more than a sole focus on phase synchronization. We show that functional connectivity critically depends on amplitude variations, which implies that, in general, phase patterns should be analyzed in conjunction with the corresponding amplitude. We discuss this issue by comparing the phase synchronization patterns of interconnected Wilson–Cowan models vis-à-vis Kuramoto networks of phase oscillators. For the interconnected Wilson–Cowan models we derive analytically how connectivity between phases explicitly depends on the generating oscillators' amplitudes. In consequence, the link between neurophysiological studies and computational models always requires the incorporation of the amplitude dynamics. Supplementing synchronization characteristics by amplitude patterns, as captured by, e.g., spectral power in M/EEG recordings, will certainly aid our understanding of the relation between structural and functional organizations in neural networks at large.

Keywords: connectivity, phase synchronization, Kuramoto network, Wilson–Cowan model, amplitude dependency

INTRODUCTION

The interplay between structural and functional brain networks has become a popular topic of research in recent years. It is currently believed that the topologies of structural and functional networks in various empirical systems may disagree (Sporns and Kötter, 2004) but systematic analyses tackling this issue are few and far between. In a combined neural mass and graph theoretical model of electroencephalographic signals, it was found that patterns of functional connectivity are influenced by – but not identical to – those of the corresponding structural level (Ponten et al., 2010). In this and many other studies, functional connectivity has been defined through the synchronization between activities at different nodes.

Neurons synchronize their firing pattern in accordance with different behavioral states. On a larger scale, synchronous activities are considered to stem from meso-scale neural populations that oscillate at certain frequencies with certain amplitudes. That is, oscillatory activity may yield synchronization characteristics within a neural population or between populations (Salenius and Hari, 2003). The amplitude of a single oscillatory neural population reflects the degree of synchronization of its neurons, that is, it measures local synchrony. By contrast, synchronization between two or more oscillatory neural populations is typically defined by their (relative) phase variance. Changes in instantaneous phase locking or coherence reflect changes in more global, distributed synchronization, i.e., between ensembles or between areas. In fact, synchronized activity across neural networks is believed to offer an effective mechanism for information transfer, especially when discriminating between frequency and phase-locked activity (Baker et al., 1999; Mima and Hallett, 1999; Salinas and Sejnowski, 2001; Fries, 2005; Womelsdorf et al., 2007). It is usually assumed that amplitude or power variations take place on long time scales when compared to the phase dynamics and are therefore considered negligible. The coupling that does, or does not, yield synchrony between oscillators hence exclusively depends on the phase. Here we ask whether this assumption is valid, and by this, tackle if a sole focus on phase really covers all functional characteristics of networks. In the present study we describe the dynamics of neural populations at every node as a neural mass model (Wilson and Cowan, 1972; Lopes Da Silva et al., 1974, 1976; Freeman, 1975; Lopes Da Silva, 1991; Jansen and Rit, 1995; Deco et al., 2008) that can behave like weakly coupled self-sustained non-linear oscillators. This description generally allows for deducing the corresponding phase dynamics (Schuster and Wagner, 1990a,b; Aoyagi, 1995; Tass, 1999) and, by this, to investigate how amplitude affects the phase dynamics in neural networks. The phase dynamics is indeed influenced by the amplitudes of the individual oscillators as we show analytically.

In a nutshell, we start off with a network of N Wilson–Cowan neural mass models (Wilson and Cowan, 1972) that are each located at network nodes $k = 1, 2, \ldots, N$ and linked solely through excitatory connections. Every model displays self-sustained oscillations with slightly different natural frequencies. Given a certain structural connectivity between the oscillators denoted by C_{kl}, we discuss how the connectivity D_{kl} between phases explicitly depends on the oscillators amplitudes R_k. The expression $D_{kl} \propto (R_l/R_k)C_{kl}$ can be derived analytically

by characterizing every oscillator via its amplitude and phase and formulating for the latter the dynamics in terms of a Kuramoto network (Kuramoto, 1984; Strogatz, 2000; Acebron et al., 2005).

The discussed structural connectivities differ qualitatively in their topology. In detail, we consider the fully connected isotropic network, a network with small-world topology generated by the Watts–Strogatz model (Watts and Strogatz, 1998), and an anatomical network reported by Hagmann et al. (2008). Capitalizing on the derived analytical expression for D_{kl}, we show how the amplitude dependency can alter the topology of connectivity in the network of Wilson–Cowan oscillators when reducing them to the Kuramoto-like network of mere phase oscillators. The connectivity at the level of phase dynamics, D_{kl}, largely prescribes the functional connectivity as quantified by the resulting synchronization patters. We illustrate this numerically using the aforementioned network topologies that are known to influence synchronizability (Watts and Strogatz, 1998; Barahona and Pecora, 2002; Achard and Bullmore, 2007; Brede, 2008).

MATERIALS AND METHODS

NETWORK MODELS

To understand the qualitative relationship between macroscopically defined functional networks and the (underlying) structural connectivity, modeling local populations of neurons in terms of averaged properties like their mean voltage and/or firing rates appears very efficient. This mean-field-like approach has a long tradition and is typically referred to as neural mass modeling (Wilson and Cowan, 1972; Lopes Da Silva et al., 1974, 1976; Freeman, 1975; Lopes Da Silva, 1991; Jansen and Rit, 1995; Deco et al., 2008). Neural mass models have been used to study the origin of alpha rhythm, evoked potentials, pathological brain rhythms, and the transition between normal and epileptic activity (Lopes Da Silva et al., 1974; Jansen and Rit, 1995; Stam et al., 1999a,b; Valdes et al., 1999; David et al., 2005). Several studies considered small networks of two or three interconnected neural mass models (Van Rotterdam et al., 1982; Schuster and Wagner, 1990a,b; Wendling et al., 2001; David and Friston, 2003; Ursino et al., 2007) as well as larger networks of interconnected models (Sotero et al., 2007; Ponten et al., 2010).

Here we chose for Wilson–Cowan as seminal neural mass model because it can readily be derived from microscopic descriptions like integrate-and-fire neurons, but also from more general models like Haken (2002) pulse-coupled neurons. By the same token, the Wilson–Cowan model provides a comprehensive link toward an even more macroscopic description as its continuum limit resembles by now well-established neural field equations (Jirsa and Haken, 1996). That is, Wilson–Cowan units may be viewed as an intermediate but in some sense generic description of densely connected neural populations.

Network of Wilson–Cowan models

As said, we are going to put individual Wilson–Cowan models at every node k of the network under study. Every model contains distinct populations of excitatory and inhibitory neurons that are described by their firing rates. If e_n denotes the firing rate of an excitatory neuron and i_n the firing rate of an inhibitory neuron, then a neural mass description can be obtained by averaging over the neural population in terms of $E = \frac{1}{N_e}\sum_{n=1}^{N_e} e_n$ and, $I = \frac{1}{N_i}\sum_{n=1}^{N_i} i_n$ where N_e and N_i are the numbers of excitatory and inhibitory neurons. By this averaging, E and I represent the mean firing rates of all excitatory and inhibitory neurons, respectively, of the neural population in question, i.e., that at node k.

Within that population, every neuron receives input from all other neurons of the population. Furthermore, the excitatory units individually receive constant external inputs p_n, whose average is given by $P = \frac{1}{N_e}\sum_{n=1}^{N_e} p_n$. The sum of all inputs is (instantaneously) integrated in time when it exceeds some threshold θ. This thresholding is realized by means of a sigmoid function S. Without loss of generality we here chose $S[x] = (1 + e^{-x})^{-1}$; we note that, in general, the thresholds may differ between excitatory and inhibitory units[1]. In consequence, the mean firing rates of the neural populations can be cast in the following dynamical system

$$\frac{d}{dt}E = -E + S\left[a_E\left(c_{EE}E - c_{IE}I - \theta_E + P\right)\right]$$
$$\frac{d}{dt}I = -I + S\left[a_I\left(c_{EI}E - c_{II}I - \theta_I\right)\right]$$

The characteristics of this dynamical system range from a mere fixed-point relaxation to limit cycle oscillations (self-sustained oscillations) depending on parameter settings (Wilson and Cowan, 1972), in particular on the choice of the external input P. That input is usually chosen at random. In the current study, we restrict all parameter values to the regime within which the dynamics displays self-sustained oscillations; see Appendix.

To combine Wilson–Cowan models in a network, different populations are now connected via their excitatory units by virtue of the sum of all E_l in the dynamics of E_k (see **Figure 1**). The dynamics at node k then becomes

$$\frac{d}{dt}E_k = -E_k + S\left[a_E\left(c_{EE}E_k - c_{IE}I_k - \theta_E + P_k + \frac{\eta}{N}\sum_{l=1}^{N} C_{kl}E_l\right)\right]$$
$$\frac{d}{dt}I_k = -I_k + S\left[a_I\left(c_{EI}E_k - c_{II}I_k - \theta_I\right)\right] \quad (1)$$

In words, all Wilson–Cowan oscillators, located at nodes l in the network drive the change of the firing rate of the excitatory units E_k. The connectivity is given by the real-valued matrix C_{kl} that has vanishing diagonal elements, i.e., $C_{kk} = 0$. That connectivity matrix is scaled via the overall coupling strength η. It is important to note that the C_{kl} connectivity matrix is here always identified as the structural connectivity.

As the different Wilson–Cowan models display self-sustained oscillations, it seems obvious to describe them using their amplitude and phase dynamics. The required transforms and approximations are summarized in the Appendix and the outcomes reveal a phase dynamics similar to the seminal Kuramoto network of phase

[1]At the individual neuron level, the dynamics reads:

$$\frac{d}{dt}e_n = -e_n + S\left[a_e\left(\frac{1}{N_e}\sum_{m=1}^{N_e} u_{mn}e_m - \frac{1}{N_i}\sum_{m=1}^{N_i} v_{mn}i_m - \theta_n^e + p_n\right)\right]$$
$$\frac{d}{dt}i_n = -i_n + S\left[a_i\left(\frac{1}{N_e}\sum_{m=1}^{N_e} w_{mn}e_m - \frac{1}{N_i}\sum_{m=1}^{N_i} z_{mn}i_m - \theta_n^i\right)\right]$$

where u, v, w, and z are positive constants representing coupling matrices within the local neural population – see, e.g., Schuster and Wagner (1990a,b) for details.

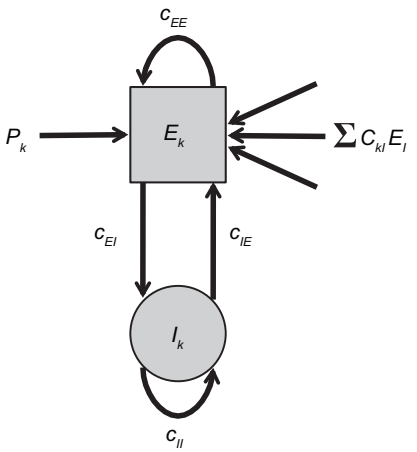


FIGURE 1 | Network of Wilson–Cowan models. At each node k a neural population containing excitatory and inhibitory units (E_k and I_k, respectively) yields self-sustained oscillations. Other nodes are connected to the excitatory unit by means of $\Sigma C_{kl}E_l$. Note that this (mean-field) coupling is scaled by a scalar η – see Eq. 1 for details.

oscillators. The Kuramoto model and its link to the here-discussed network of Wilson–Cowan models will be briefly sketched in the following two sub-sections.

Kuramoto network of phase oscillators

The collective behavior of a network of oscillators, whose states are captured by a single scalar phase φ_k each, can, in first approximation, be represented by the set of N coupled differential Eq.

$$\frac{d}{dt}\varphi_k = \omega_k + \frac{\eta}{N}\sum_{i=1}^{N} D_{kl}\sin\left(\varphi_l - \varphi_k\right) \tag{2}$$

That is, the k-th oscillator, with natural frequency ω_k, adjusts its phase according to input from other oscillators through a pair-wise phase interaction function $\sin(\varphi_l - \varphi_k)$. The connectivity matrix D_{kl} is again scaled by an overall coupling strength, η. As will be sketched below, η serves as a bifurcation parameter in that small values of η yield a network behavior that essentially agrees with the entirely uncoupled case (i.e., the phases are not synchronized), whereas η larger than a certain critical value η_c causes the phases to synchronize. The frequencies ω_k are distributed according to a specified probability density usually taken to be a symmetric, unimodal distribution (e.g., Lorentzian or Gaussian distributions) with mean ω_0. Although the sinusoidal interaction function is an approximation, it still permits a variety of highly non-trivial solutions. As such the model (2) can be viewed as the canonical form for synchronization in extended, oscillatory media. We note that the connectivity matrix D_{kl} represents also a structural connectivity that does not necessarily agree with that of the Wilson–Cowan model – see below.

Strictly speaking the system (2) does not represent the Kuramoto model in its original form as there the coupling between nodes k and l was considered isotropic and homogeneous, i.e., $D_{kl} = 1$ for all connections, by which the model reduces to

$$\frac{d}{dt}\varphi_k = \omega_k + \frac{\eta}{N}\sum_{l=1}^{N}\sin\left(\varphi_l - \varphi_k\right)$$

For the sake of legibility, however, we here refer to (2) also as the Kuramoto model.

As mentioned above, the effect of increasing η in the isotropic case is to increase the phase synchrony amongst the oscillators. Suppose the coupling is weak (i.e., smaller than the critical value, or $\eta \ll \eta_c$, then the oscillators' phases disperse, whereas for strong coupling $\eta \gg \eta_c$ the oscillators become synchronous, i.e., the phases are locked at fixed differences. In the intermediate case $\eta \approx \eta_c$, clusters of synchronous oscillators may emerge. However, many other oscillators, whose natural frequencies are at the tails of the distribution, are not locked into a cluster. In other words, as η increases, the interaction functions overcome the dispersion of natural frequencies ω_n resulting in a transition from incoherence, to partial and then full synchronization (Acebron et al., 2005; Breakspear et al., 2010).

Linking neural mass models to phase oscillators

When deriving the Kuramoto network from the Wilson–Cowan oscillator network, the major ingredient is to average every oscillator over one cycle when assuming that its amplitude and phase change slowly as compared to the oscillator's frequency. That is, time-dependent amplitude and phase are fixed, the system is integrated over one period to remove all harmonic oscillations, and, subsequently, amplitude and phase are again considered to be time-dependent (Guckenheimer and Holmes, 1990) – this procedure is also referred to as a combination of rotating wave approximation and slowly varying amplitude approximation (Haken, 1974). As shown in more detail in the Appendix, the phase dynamics of the system (1) can in this way be approximated as

$$\begin{aligned}\frac{d}{dt}\varphi_k = \omega_k &+ \frac{\eta}{2N}\sum_{l=1}^{N} C_{kl}a_E S'\left[\chi_{E,k}^{(0)}\right]\frac{R_l}{R_k}\sin\left(\varphi_l - \varphi_k\right)\\ &+ \frac{\eta}{16N}\sum_{l=1}^{N} C_{kl}a_E^3 S'''\left[\chi_{E,k}^{(0)}\right]R_k R_l\left(\left(c_{EE}^2 + 3c_{IE}^2\right)\sin\left(\varphi_l - \varphi_k\right)\right.\\ &\left. + 2c_{EE}c_{IE}\cos\left(\varphi_l - \varphi_k\right)\right)\end{aligned}$$

with S' and S''' referring to the first and third derivative of the sigmoid function S. The parameter $\chi_{E,k}^{(0)}$ is given by

$$\chi_{E,k}^{(0)} = a_E\left(c_{EE}E_k^{(0)} - c_{IE}I_k^{(0)} - \theta_E + P_k + \frac{\eta}{N}\sum_{l=1}^{N} C_{kl}E_l^{(0)}\right)$$

with $(E_k^{(0)}, I_k^{(0)})$ defining the unstable node within the limit cycle of the Wilson–Cowan model (1) and at network node k. For more details including the definition of the natural frequency we refer to the Appendix. Considering the case that the amplitudes R_k are reasonably small, this phase dynamics can be further simplified to

$$\frac{d}{dt}\varphi_k \approx \omega_k + \frac{\eta}{2N}\sum_{l=1}^{N} a_E S'\left[\chi_{E,k}^{(0)}\right]\frac{R_l}{R_k}C_{kl}\sin\left(\varphi_l - \varphi_k\right)$$

which does resemble a Kuramoto network. In fact, by comparing this form with the dynamics (2) we find

$$D_{kl} = \frac{1}{2}a_E S'\left[\chi_{E,k}^{(0)}\right]\frac{R_l}{R_k}C_{kl} \tag{3}$$

In sum, the phase dynamics can, in good approximation, be cast into the form of a Kuramoto network provided the connectivity matrix is corrected by means of (3). This correction yields a non-trivial amplitude dependence of the connectivity at the level of the phase dynamics. Since S is a sigmoid function, S' becomes bell-shaped implying a change in connectivity D_{kl} whenever the parameter $\chi^{(0)}_{E,k}$ is altered, e.g., by shifting the center of the Wilson–Cowan limit cycle at node k and/or l. This probably more global dependence is supplemented by the here more important node-by-node dependence. When the amplitudes R_k differ per node, the ratio R_l/R_k in (3) directly affects the value of D_{kl}, which can, strictly speaking, be entirely independent on the choice of the connectivity matrix C_{kl}. Put differently, the structural connectivity at the neural mass level does not necessarily agree with the structural connectivity at the phase dynamics level.

Given our interest in amplitude dependency, we finally add a note about "large" amplitudes. In line with the Appendix Eq. A.7 including larger amplitudes yields a slight modification of the phase dynamics that we here abbreviate as

$$\frac{d}{dt}\varphi_k = \omega_k + \frac{\eta}{N}\sum_{l=1}^{N}\tilde{D}_{kl}\sin\left(\varphi_l - \varphi_k - \alpha_{kl}\right)$$

Interestingly, the presence of large amplitudes yields, apart from slightly different coupling coefficients $\tilde{D}_{kl}$, phase shifts α_{kl} that translate to finite transmission delays. Prior studies that incorporate transmission delays into phase oscillators have revealed elaborate synchronization behaviors (Zanette, 2000; Jeong et al., 2002). The more complex dynamics due to α suggests the notion of frustration, whereby the interaction functions require some finite phase offset in order to vanish (Acebron et al., 2005). For a more detailed discussion we refer to a recent review by Breakspear et al. (2010). Note that for our analytical estimates we always consider the case in which Eq. (2) and (3) apply to good approximation.

SIMULATIONS

More recently, several research groups started investigating the relationship between structural and functional connectivity, suggesting that functional connectivity may indeed resemble aspects of structural connectivity, at least to some extent (Lebeau and Whittington, 2005; Ingram et al., 2006; Honey et al., 2007, 2009, 2010; Voss and Schiff, 2009; D'angelo et al., 2010). In most studies, a fixed structural architecture was implemented based on, for instance, the cortical structure of the cat (Zhou et al., 2007), or the macaque neo-cortex (Honey et al., 2007). Yet it is unclear how variations in the network properties at the structural level or fixed network properties with variations by means of (node-dependent) amplitudes may affect the synchronization strength and more global network characteristics at the functional level.

Synchronization was quantified via the phase locking index or the phase uniformity ρ, defined as (Mardia and Jupp, 2000)

$$\rho = \frac{1}{N}\left|\sum_{k=1}^{N} e^{i\varphi_k}\right|$$

This index agrees with the so-called Kuramoto order parameter and reflects the degree of divergence of the different phases in the network (not the relative phases). By varying the overall coupling strength η we induced qualitative differences in synchronization as the order parameter was expected to undergo well-defined bifurcations from an unlocked state to in-phase locking. We simulated both the network of Wilson–Cowan oscillators as well as the Kuramoto network. For the Wilson–Cowan model, we defined the phase as the quadrant-corrected inverse tangents of the ratio of excitatory and inhibitory units at node k, i.e., $\varphi_k = \arctan(E_k/I_k)$ – this phase largely agreed with the Hilbert-phase of E_k because of the smoothness of the Wilson–Cowan limit cycle. For the Kuramoto network, the phase was, of course, the state variable under study, which did not require any further definition. In all simulations the primary outcome variable in all simulations was, hence, $\rho(\eta)$ for different network types and, in the case of the Wilson–Cowan network, distinct ranges of input values P_k as will be explained below in all detail. In addition, we computed the phase locking index of the pair-wise relative phases between nodes which served as definition of the functional networks. The precise transform of the Kuramoto network dynamics to the dynamics of relative phases is beyond the scope of the current paper.

To study potentially "erroneous" simulations of the phase dynamics – and thus possible "misinterpretations" of structural connectivity when solely looking at functional networks defined via phase synchrony – we ignored for the Kuramoto network the amplitude dependency (3) of the connectivity matrix and simply identified D_{kl} by C_{kl}. We further accelerated numerical simulations by adding some small dynamic noise (Stratonovich, 1963; Risken, 1989; Daffertshofer, 1998), so-called Langevin forces $\Gamma_k(t)$, in the form of mean-centered Gaussian white noise. The simulated dynamics hence looked like

$$\begin{aligned}\frac{d}{dt}E_k &= -E_k + S\left[a_E\left(c_{EE}E_k - c_{IE}I_k - \theta_E + P_k + \frac{\eta}{N}\sum_{l=1}^{N}C_{kl}E_l + \sqrt{2\varepsilon}\,\Gamma_k\right)\right]\\ \frac{d}{dt}I_k &= -I_k + S\left[a_I\left(c_{EI}E_k - c_{II}I_k - \theta_I\right)\right]\end{aligned} \tag{4}$$

and

$$\frac{d}{dt}\varphi_k = \omega_k + \frac{\eta}{N}\sum_{l=1}^{N}C_{kl}\sin\left(\varphi_l - \varphi_k\right) + \sqrt{2\varepsilon}\,\Gamma_k \tag{5}$$

Recall that the connectivity in (5) differs from (2) by means of $D_{kl} \to C_{kl}$.

Throughout simulations we fixed parameter settings as: $a_E = 1.2$, $a_I = 2$, $c_{EE} = 5$, $c_{II} = 1$, $c_{IE} = 6$, $c_{EI} = 10$, $\theta_E = 2$, $\theta_I = 3.5$. The strength of the dynamical noise was always considered very small (it only served to accelerate numerics and not to discuss impact of stochastic forces). It was set to $\varepsilon = 10^{-4}$ for all simulations of the Wilson–Cowan network (4) and to $\varepsilon = 10^{-2}$ for the Kuramoto network (5). Simulations were realized using a simple Euler-forward scheme with step-size 10^{-2}. Per run a total number of 10^5 samples were simulated. For each network, simulations were repeated with 10 different realizations of constant but random inputs P_k (Wilson–Cowan oscillators) or constant but random natural frequencies ω_n (Kuramoto oscillators). In addition, for the small-world network, new C_{kl} matrices were generated with different rewiring pattern for each realization. Each of these 10 realizations was again repeated five times with different initial values of E_k and I_k, or φ_k. The resulting $\rho(\eta)$ values were computed over the final 100 samples of every run

and averaged over all simulations. Primary outcome variable was, hence, $\rho(\eta)$ for simulations of (4) and (5) using the three different network types, and in the case of the Wilson–Cowan network (4), using altered input-distributions to set P_k.

For the Kuramoto model the natural frequencies ω_n were randomly drawn from a Cauchy–Lorentz distribution with width $\gamma = 0.5$ and initial φ_k values at time $t = 0$ were drawn from a uniform distribution over the interval $[0, 2\pi)$. The initial E_k and I_k values for the Wilson–Cowan oscillators were uniformly chosen from the interval $[0, 1]$.

By default, the constant input values P_k were drawn from a uniform distribution with $-0.25 \le P_k \le 0.25$ for every node k. In order to tackle amplitude effects, however, we looked also at the case in which (selected) nodes displayed oscillations with clearly different amplitudes than others. For this we selected four different intervals from which P_k was drawn: $-0.25 \le P_k \le -0.20$; $0.20 \le P_k \le 0.25$; $-0.8 \le P_k \le -0.7$; and $0.7 \le P_k \le 0.8$. Simulations were performed using either a single interval or a combination of two intervals for which the first 50% of the nodes were assigned a P_k from the first interval and the second 50% from the second interval. These combinations of intervals were between similar ranges, hence: $-0.25 \le P_k \le_k -0.20$ with $0.20 \le P_k \le 0.25$ and $-0.8 \le P_k \le -0.7$ with $0.7 \le P_k \le 0.8$. As shown in the final part of the Appendix the stationary amplitude at network node k either vanishes, i.e., $R_{k,\text{stationary}} = 0$ or it obeys the form

$$R_{k,\text{stationary}} = \sqrt{8\frac{2 - a_E c_{EE} S'\left[\chi^{(0)}_{E,k}\right] + a_I c_{II} S'\left[\chi^{(0)}_{I,k}\right]}{a_E^3 c_{EE}\left(c_{EE}^2 + c_{IE}^2\right) S'''\left[\chi^{(0)}_{E,k}\right] - a_I^3 c_{II}\left(c_{II}^2 + c_{EI}^2\right) S'''\left[\chi^{(0)}_{I,k}\right]}}$$

that by virtue of $\chi^{(0)}_{E,k}$ explicitly depends on the input P_k. Given this dependency, varying the input P_k systematically could be used to create different scenarios of amplitude effects, which – in particular if selected nodes received significantly different input than others – potentially caused pronounced, qualitative differences between C_{kl} and D_{kl}. In these cases, the simulations of (4) and (5) were expected to disagree.

The connectivity matrices C_{kl} were chosen as either a fully connected isotropic network, as a network with small-world topology generated by the Watts–Strogatz model (Watts and Strogatz, 1998), or via an anatomical network reported by Hagmann et al. (2008). For all the connectivities we estimated the functional networks via phase locking between nodes.

Fully connected homogeneous network

The original Kuramoto network comprises a fully connected homogeneous network – see above. C_{kl} in this case consists of an $N \times N$ matrix containing ones everywhere except for the diagonal, where all values were set to zero, i.e., we did not allow for self-connections. We note that discarding diagonal elements is, strictly speaking, not necessary for the phase dynamics (2) or (6) as the coupling via the sine of relative phase vanishes, i.e., by construction (or symmetry) there are no self-connections. This argument, however, does not apply for the network (1) or (5), hence we always set $C_{kk} = 0$.

Although the Kuramoto network is usually studied for large size networks, we chose a network of 66 nodes in order to make a better comparison with the Hagmann dataset.

Small-world network

The model for generating small-world networks employed here was introduced by Watts and Strogatz (1998) to generate graphs with high clustering and low path length (high efficiency). Starting from an ordered network on a ring lattice where nodes are only connected to a small number of direct neighbors, connections are subsequently rewired to a random (distant) node with certain probability. The introduction of a few random connections in an ordered network drastically increases the synchronizability of the network (Watts and Strogatz, 1998; Barahona and Pecora, 2002; Motter et al., 2005; Zhou and Kurths, 2006; Stam and Reijneveld, 2007; Wu et al., 2008; Chen et al., 2009). We used a network with an average degree of 10 and a rewiring probability of 0.2. An example of a C_{kl} matrix is given in **Figure 5** below.

Hagmann network

Empirical networks are unlikely to have an organization that can be exactly described by one of the theoretical network models. To study a network that more realistically represents anatomical connections in the human brain we repeated our simulations on a network that was based on axonal pathways obtained by diffusion spectrum imaging. This dataset has been used to identify the so-called "structural core" of anatomical connections in the human cerebral cortex as described by Hagmann et al. (2008), which is accessible via http://www.connectomeviewer.org/viewer/datasets. To reduce the size of the network and, by this, accelerate simulation time, the original 998 regions were assigned to a 66-node parcellation scheme and averaged over all five subjects as was also done in the original study (Hagmann et al., 2008). The resulting weighted, undirected network was subsequently thresholded to obtain a binary network with an average degree of 10. This network served as our connectivity matrix C_{kl}; see **Figure 2** and also **Figure 6** below.

RESULTS

The changes in synchronization ρ as a function of overall coupling strength η are summarized in **Figure 3**. First thing to notice is that, for a critical η, the Wilson–Cowan model shows a brisk increase in ρ after which maximal synchronization is reached. Increasing η again after a critical value breaks down the synchronization as the individual Wilson–Cowan oscillators leave the stable limit cycle regime when their inputs exceed a certain value (Schuster and Wagner, 1990a). That means, the neural masses at the different nodes stop oscillating altogether if coupling is too strong. Of course, this does not apply for the Kuramoto model since, by construction, the phases keep oscillating. In consequence, ρ keeps increasing with η and reaches asymptotically maximum synchronization (see bottom row's panels in **Figure 3**).

The different choices of P_k intervals result in altered synchronization curves. This was most apparent for the $[-0.8,\ldots,-0.7]$ and $[0.7,\ldots,0.8]$ intervals (blue solid lines in **Figure 3**, third row's panels) when amplitudes lie furthest apart. In general, different P_k intervals caused a shift in critical η, with networks with larger amplitudes reaching maximum synchronization for lower coupling strength than oscillators with smaller amplitude. Interestingly, the cases with bimodal amplitude distributions (dashed lines) were

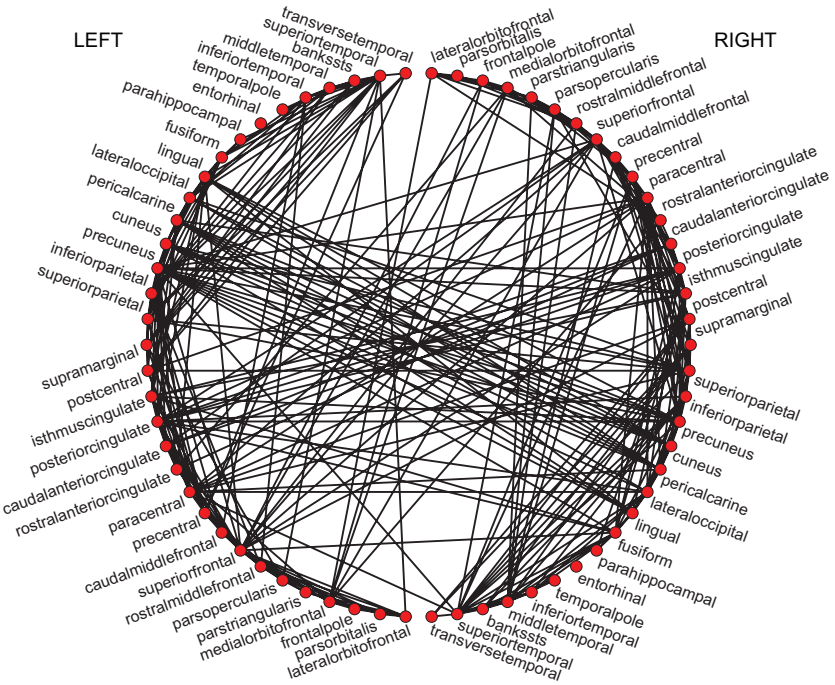


FIGURE 2 | Plot of the Hagmann network (Hagmann et al., 2008). The original 998 regions were assigned to a 66-node parcellation scheme. For the sake of visualization, all 66 nodes are located on the circle; see text for more details.

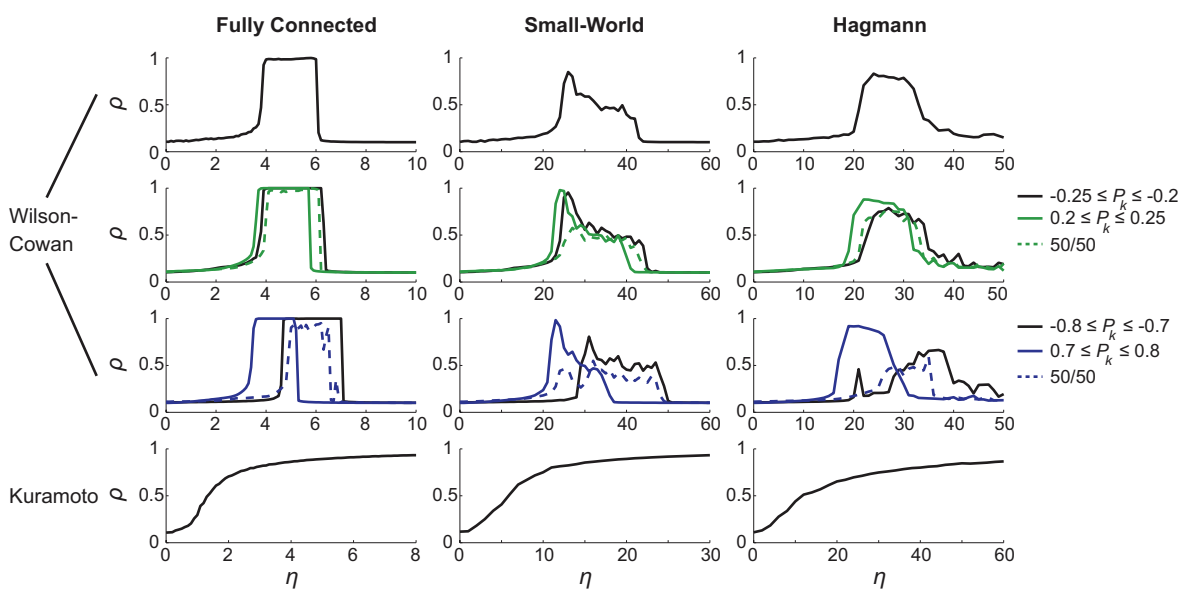


FIGURE 3 | The synchronization ρ as a function of overall coupling strength η for the network of Wilson–Cowan oscillators [Eq. 4; three upper rows] and for the Kuramoto network [Eq. 5; bottom row]. For the upper row, P_k values were drawn from the interval [–0.25,…,0.25], for the second and third row from the indicated intervals (see right-hand side).

less synchronizable than their unimodal counterparts. An example of this phenomenon is the case of a fully connected network that reaches global synchronization for each of the [−0.8,…,−0.7] and [0.7,…,0.8] P_k intervals separately but appears unable to fully synchronize for a combination of the two. A closer look at the functional synchronization patterns between individual nodes of this network revealed that two distinct clusters emerged corresponding to the bimodal inputs and thus amplitude distribution (**Figure 4**).

If the structural connectivity is isotropic, then amplitude distribution largely (if not fully) prescribes the functional connectivity pattern that thus clearly disagrees with the structural connectivity. In consequence, the current example revealed two strongly synchronized local clusters but the large difference between the input intervals prevented them from synchronizing with one another. It is important to note that, if amplitude effects were not taken into account, a full synchronization of the network would have been found.

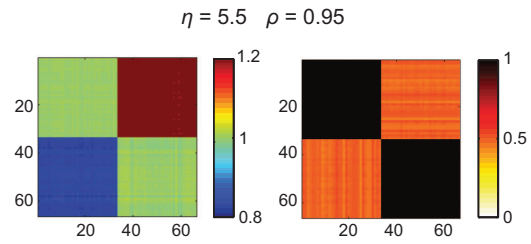


FIGURE 4 | Functional connectivity between the nodes of the fully connected (isotropic and homogeneous) network. $C_{kl} = 1$ except for its diagonal elements, $C_{kk} = 0$. Input values P_k for the first 33 nodes were drawn from the interval [–0.8,…,–0.7] and the second 33 nodes from the interval [0.7,…,0.8]. Due to this bimodal input distribution, the amplitudes at the nodes and hence their ratios R_l/R_k differ between pairs of nodes (left panel), which yields – by virtue of Eq. (3) – a change in the functional connectivity (right panel). Here it can be seen that the phases of nodes within the same P_k range are fully synchronized but fail to synchronize between clusters; coupling strength was set to $\eta = 5.5$ resulting in a global phase synchrony of $\rho = 0.95$.

With the current parameter settings no full global synchronization could be achieved in both the small-world and the Hagmann network. However, partial synchronization patterns could be observed that did not correspond with the structural connectivity but also not with the distribution of amplitudes (**Figures 5 and 6**). These patterns rapidly emerged and disappeared with varying η. Although the match with the amplitude distribution was not as clear-cut as in the case of the fully connected network (**Figure 4**), a similar clustering could be observed, by which the functional connectivities turned out to differ not only quantitatively but also qualitatively from the underlying structural connectivity – a fact that would be missed if relying on a description of sole phase oscillators that show such partial synchronization patterns only in close vicinity of the critical coupling strength.

DISCUSSION AND CONCLUSION

The introduction of network analysis to neuroscience has paved new ways for the study of neural network organizations. Particular focus has been on the search for complex networks since many of these networks – especially in the neuroinformatics context – are known

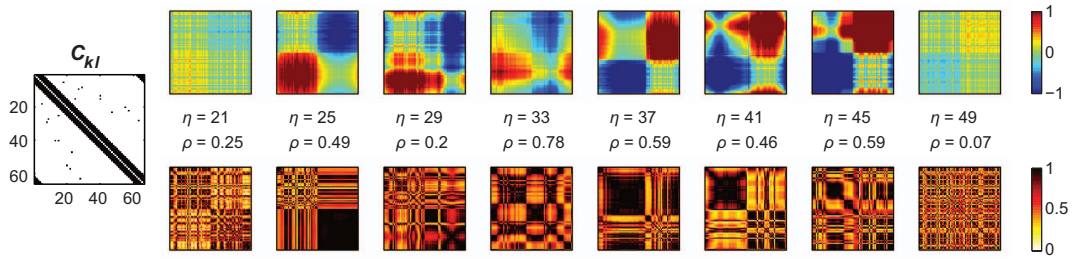


FIGURE 5 | Amplitude ratio distribution [log(R_l/R_k); upper row] and functional connectivity (lower row) of the small-world network that differs from the underlying structural connectivity (C_{kl}, most left panel). As in **Figure 4** we here chose the inputs from a bimodal distribution, here the intervals [–0.8,…,–0.7] and [0.7,…,0.8], which causes the amplitudes to differ between the first and second half of the nodes and, consequently, the functional connectivity matrix to disagree with C_{kl}. Similar to the homogeneous case in **Figure 4** the ratios R/R_k largely prescribe the functional connectivity, here, however, mixed with the small-world structure C_{kl} – see, e.g., the pronounced synchrony along the diagonal that is absent in the patterns of amplitude ratios. In addition, dependent on the overall coupling strength η, pronounced clusters of synchronized nodes appear – the corresponding coupling values, $\eta = 25$–45, agree with the synchronization regime of this small-world network displayed in **Figure 3**, middle column, third row, blue dashed line.

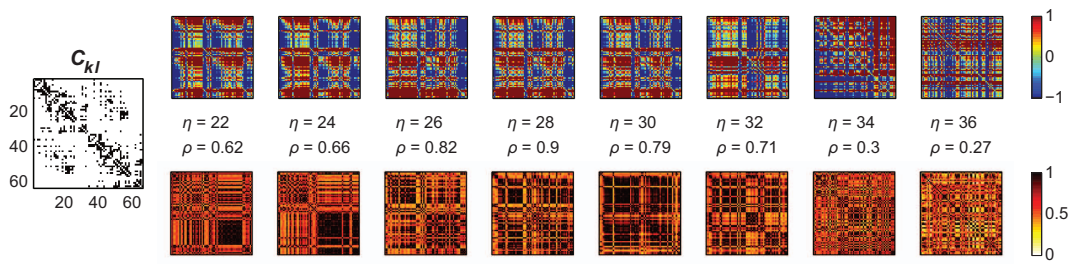


FIGURE 6 | Amplitude ratios [upper row; log(R_l/R_k)] and functional connectivity (lower row) of the Hagmann network using again a bimodal distribution of P_k. The left most panel is again the structural network C_{kl}; we here used $P_k \in$ [–0.25,…,–0.20] and [0.20,…,0.25]. As in the small-world case, localized clusters of synchronized nodes emerge dependent on the overall coupling strength η. These clustered patterns apparently disagree with the underlying anatomical network (most left panel); cf. **Figure 3**, right column, second row, green dashed lines.

for their efficiency when transferring and integrating information from local, specialized brain areas, even when they are distant (Sporns and Zwi, 2004). Over the years, small-world structural networks have been found for *C. Elegans* (Watts and Strogatz, 1998), cat cortex, and macaque (visual) cortex (Sporns and Zwi, 2004). In humans, anatomical connectivity can be estimated *in vivo* indirectly via cross-correlation analysis of cortical thickness in structural MRI (He et al., 2007; Chen et al., 2008) and more directly using tractography based on diffusion tensor imaging and diffusion spectrum imaging (Iturria-Medina et al., 2007; Hagmann et al., 2008; Gong et al., 2009). Similar to the structural, i.e. anatomical connections, functional connections have also been found to display characteristics of complex networks, especially by looking at human functional networks in resting state, using either fMRI (Salvador et al., 2005; Achard et al., 2006; Van den Heuvel et al., 2008; Ferrarini et al., 2009) or M/EEG (Stam, 2004; Bassett et al., 2006; Stam and Reijneveld 2007; Bullmore and Sporns, 2009). Changes in these resting state networks appear to relate to neurological and psychiatric diseases like Alzheimer's disease (Stam et al., 2007, 2009; He et al., 2008; Supekar et al., 2008; de Haan et al., 2009), schizophrenia (Bassett et al., 2008; Liu et al., 2008; Rubinov et al., 2009), ADHD in children (Wang et al., 2009), (removal of) brain tumors (Bartolomei et al., 2006; Bosma et al., 2009), and during epileptic seizures (Kramer et al., 2008; Schindler et al., 2008; Ponten et al., 2009), but also to aging (Achard and Bullmore, 2007; Meunier et al., 2009; Micheloyannis et al., 2009) or to different sleep stages (Ferri et al., 2008; Dimitriadis et al., 2009), as well as during foot movements (De Vico Fallani et al., 2008) and finger tapping (Bassett et al., 2006).

Do these functional networks precisely match their underlying structural counterparts? In general, networks do not agree, especially when the functional networks are solely defined via (phase) synchronization patterns, which is common practice when studying electrophysiological signals, for instance, M/EEG. We have shown that, even if the local dynamics at every node of a network can be described as phase dynamics in the form of a Kuramoto network, the connectivity matrix at this level of phases does not necessarily agree with the connectivity at the level of neural mass models describing firing rates of local neural populations. The connectivity at the phase dynamics level has to be corrected by its amplitude dependency. This phase level is indeed closely related to the empirically assessed functional connectivity matrix as this, as said, is commonly defined through locking patterns of phases. If relying on Kuramoto-like approximations, the connectivity matrix has to be corrected via the relation (3) that may include non-trivial amplitude dependency. Especially, when the amplitudes differ from node to node, the connectivity at the level of phases can qualitatively differ from the structural connectivity at the level of neural mass or mean firing rates. That is, structural and functional connectivity may differ simply because of the latter's amplitude dependency.

In consequence, phase dynamics and, hence, synchrony patterns should always be analyzed in conjunction with the corresponding amplitude changes. Patterns of global synchrony (phase) may depend on local synchrony (amplitude). This may have profound impacts when linking, for instance, M/EEG studies to neural modeling. Amplitude there translates to (spectral) power, which typically differs between distinct behavioral states or due to pathology. Incorporating these amplitude changes will certainly help to understand how structural and functional network organizations in the cortex, in particular, and in the central nervous system, in general, may relate to one another.

ACKNOWLEDGMENTS

We thank the Netherlands Organisation for Scientific Research for financial support (NWO grant # 021-002-047).

REFERENCES

Acebron, J., Bonilla, L., Pérez Vicente, C., Ritort, F., and Spigler, R. (2005). The Kuramoto model: a simple paradigm for synchronization phenomena. *Rev. Mod. Phys.* 77, 137.

Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* 3, e174. doi: 10.1371/journal.pcbi.0030017

Achard, S., Salvador, R., Whitcher, B., Suckling, J., and Bullmore, E. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *J. Neurosci.* 26, 63.

Aoyagi, T. (1995). Network of neural oscillators for retrieving phase information. *Phys. Rev. Lett.* 74, 4075.

Baker, S. N., Kilner, J. M., Pinches, E. M., and Lemon, R. N. (1999). The role of synchrony and oscillations in the motor output. *Exp. Brain Res.* 128, 109.

Barahona, M., and Pecora, L. M. (2002). Synchronization in small-world systems. *Phys. Rev. Lett.* 89, 054101.

Bartolomei, F., Bosma, I., Klein, M., Baayen, J. C., Reijneveld, J. C., Postma, T. J., Heimans, J. J., van Dijk, B. W., de Munck, J. C., de Jongh, A., Cover, K. S., and Stam, C. J. (2006). Disturbed functional connectivity in brain tumour patients: evaluation by graph analysis of synchronization matrices. *Clin. Neurophysiol.* 117, 2039.

Bassett, D.S., Meyer-Lindenberg, A., Achard, S., Duke, T., and Bullmore, E. (2006). Adaptive reconfiguration of fractal small-world human brain functional networks. *Proc. Natl. Acad. Sci. U. S. A.* 103, 19518–19523.

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *J. Neurosci.* 28, 9239.

Bosma, I., Reijneveld, J. C., Klein, M., Douw, L., van Dijk, B. W., Heimans, J. J., and Stam, C. J. (2009). Disturbed functional brain networks and neurocognitive function in low-grade glioma patients: a graph theoretical analysis of resting-state MEG. *Nonlinear Biomed. Phys.* 3, 9.

Breakspear, M., Heitmann, S., and Daffertshofer, A. (2010). Generative models of cortical oscillations: neurobiological implications of the Kuramoto model. *Front. Hum. Neurosci.* 4:190. doi: 10.3389/fnhum.2010.00190

Brede, M. (2008). Locals vs. global synchronization in networks of non-identical Kuramoto oscillators. *Eur. Phys. J. B* 62, 87.

Bullmore, E. and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186.

Chen, M., Shang, Y., Zhou, C., Wu, Y., and Kurths, J. (2009). Enhanced synchronizability in scale-free networks. *Chaos* 19, 013105.

Chen, Z.J., He, Y., Rosa-Neto, P., Germann, J., and Evans, A. C. (2008). Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb. Cortex* 18, 2374.

Daffertshofer, A. (1998). Effects of noise on the phase dynamics of nonlinear oscillators. *Phys. Rev. E* 58, 327.

D'angelo, E., Mazzarello, P., Prestori, F., Mapelli, J., Solinas, S., Lombardo, P., Cesana, E., Gandolfi, D., and Congi, L. (2010). The cerebellar network: from structure to function and dynamics. *Brain Res. Rev.* 66, 5.

David, O., and Friston, K. J. (2003). A neural mass model for MEG/EEG: coupling and neuronal dynamics. *Neuroimage* 20, 1743.

David, O., Harrison, L., and Friston, K. J. (2005). Modeling event-related responses in the brain. *Neuroimage* 25, 756.

De Haan, W., Pijnenburg, Y. A., Strijers, R. L., van der Made, Y., van der Flier, W. M., Scheltens, P., and Stam, C. J. (2009). Functional neural network analysis in frontotemporal dementia and Alzheimer's disease using EEG and graph theory. *BMC Neurosci.* 10, 101. doi: 10.1186/1471-2202-10-101

De Vico Fallani, F., Astolfi, L., Cincotti, F., Mattia, D., Marciani, M. G., Tocci, A., Salinari, S., Witte, H., Hesse, W., Gao, S., Colosimo, A., and Babiloni, F. (2008). Cortical network dynamics during foot movements. *Neuroinformatics* 6, 23.

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: from spiking neu-

rons to neural masses and cortical fields. *PLoS Comput. Biol.* 4, E1000092. doi: 10.1371/journal.pcbi.1000092

Dimitriadis, S.I., Laskaris, N. A., Del Rio-Portilla, Y., and Koudounis, G. C. H. (2009). Characterizing dynamic functional connectivity across sleep stages from EEG. *Brain Topogr.* 22, 119.

Ferrarini, L., Veer, I. M., Baerends, E., van Tol, M. J., Renken, R. J., van der Wee, N. J., Veltman, D. J., Aleman, A., Zitman, F. G., Penninx, B. W., van Buchem, M. A., Reiber, J. H., Rombouts, S. A., and Milles, J. (2009). Hierarchical functional modularity in the resting-state human brain. *Hum. Brain Mapp.* 30, 2220.

Ferri, R., Rundo, F., Bruni, O., Terzano, M. G., and Stam, C. J. (2008). The functional connectivity of different EEG bands moves towards small-world network organization during sleep. *Clin. Neurophysiol.* 119, 2026.

Freeman, W. J. (1975). *Mass Action in the Nervous System*. New York: Academic Press.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, 474.

Gong, G., He, Y., Concha, L., Lebel, C., Gross, D.W., Evans, A.C., and Beaulieu, C. (2009). Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb. Cortex* 19, 524–536.

Guckenheimer, J., and Holmes, P. (1990). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6, e159. doi: 10.1371/journal.pbio.0060159

Haken, H. (1974). *Synergetics*. Berlin: Springer.

Haken, H. (2002). *Brain Dynamics*. Berlin: Springer.

He, Y., Chen, Z., and Evans, A. (2008). Structural insights into aberrant topological patterns of large-scale cortical networks in Alzheimer's Disease. *J. Neurosci.* 28, 4756.

He, Y., Chen, Z. J., and Evans, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb. Cortex* 17, 2407.

Honey, C. J., Kötter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10240.

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., and Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035.

Honey, C. J., Thivierge, J. P., and Sporns, O. (2010). Can structure predict function in the human brain? *Neuroimage* 52, 766.

Ingram, P. J., Stumpf, M. P., and Stark, J. (2006). Network motifs: structure does not determine function. *BMC Genomics* 7, 108. doi: 10.1186/1471-2164-7-108

Iturria-Medina, Y., Canales-Rodríguez, E.J., Melie-García, L., Valdés-Hernández, P.A., Martínez-Montes, E., Alemán-Gómez, Y., and Sánchez-Bornot, J.M. (2007). Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *Neuroimage* 36, 645–660.

Jansen, B. H., and Rit, V. G. (1995). Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biol. Cybern.* 73, 357.

Jeong, S. O., Ko, T. W., and Moon, H.T. (2002). Time-delayed spatial patterns in a two-dimensional array of coupled oscillators. *Phys. Rev. Lett.* 89, 154104.

Jirsa, V. K., and Haken, H. (1996). Field theory of electromagnetic brain activity. *Phys. Rev. Lett.* 77, 960.

Kramer, M. A., Kolaczyk, E. D., and Kirsch, H. E. (2008). Emergent network topology at seizure onset in humans. *Epilepsy Res.* 79, 173.

Kuramoto, Y. (1984). *Chemical Oscillations, Waves, and Turbulence*. New York: Springer.

Lebeau, F. E., and Whittington, M. A. (2005). Structure/function correlates of neuronal and network activity – an overview. *J. Physiol.* 562(Pt 1), 1.

Liu, Y., Liang, M., Zhou, Y., He, Y., Hao, Y., Song, M., Yu, C., Liu, H., Liu, Z., and Jiang, T. (2008). Disrupted small-world networks in schizophrenia. *Brain* 131, 945.

Lopes Da Silva, F. (1991). Neural mechanisms underlying brain waves: from neural membranes to networks. *Electroencephalogr. Clin. Neurophysiol.* 79, 81.

Lopes Da Silva, F. H., Smits, H., and Zetterberg, L. H. (1974). Model of brain rhythmic activity. The alpha-rhythm of the thalamus. *Kybernetik* 15, 27.

Lopes Da Silva, F. H., van Rotterdam, A., Barts, P., van Heusden, E., and Burr, W. (1976). Models of neuronal populations: the basic mechanisms of rhythmicity. *Prog. Brain Res.* 45, 281.

Mardia, K. V., and Jupp, P. E. (2000). *Directional Statistics*. Chichester, Uk: John Wiley and Sons.

Meunier, D., Achard, S., Morcom, A., and Bullmore, E. (2009). Age-related changes in modular organization of human brain functional networks. *Neuroimage* 44, 715.

Micheloyannis, S., Vourkas, M., Tsirka, V., Karakonstantaki, E., Kanatsouli, K., and Stam, C. J. (2009). The influence of ageing on complex brain networks: a graph theoretical analysis. *Hum. Brain Mapp.* 30, 200.

Mima, T., and Hallett, M. (1999). Corticomuscular coherence: a review. *J. Clin. Neurophysiol.* 16, 501.

Motter, A. E., Zhou, C., and Kurths, J. (2005). Network synchronization, diffusion, and the paradox of heterogeneity. *Phys. Rev. E* 71(Pt 2), 016116.

Ponten, S. C., Daffertshofer, A., Hillebrand, A., and Stam, C. J. (2010). The relationship between structural and functional connectivity: graph theoretical analysis of an EEG neural mass model. *Neuroimage* 52, 985.

Ponten, S.C., Douw, L., Bartolomei, F., Reijneveld, J. C, and Stam, C. J. (2009). Indications for network regularization during absence seizures: weighted and unweighted graph theoretical analyses. *Exp. Neurol.* 217, 197.

Risken, H. (1989). *The Fokker-Planck Equation*. Berlin: Springer.

Rubinov, M., Knock, S. A., Stam, C. J., Micheloyannis, S., Harris, A. W., Williams, L. M., and Breakspear, M. (2009). Small-world properties of nonlinear brain activity in schizophrenia. *Hum. Brain Mapp.* 30, 403.

Salenius, S., and Hari, R. (2003). Synchronous cortical oscillatory activity during motor action. *Curr. Opin. Neurobiol.* 13, 678.

Salinas, E., and Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* 2, 539.

Salvador, R., Veer, I. M., Baerends, E., van Tol, M. J., Renken, R. J., van der Wee, N. J., Veltman, D. J., Aleman, A., Zitman, F. G., Penninx, B. W., van Buchem, M. A., Reiber, J. H., Rombouts, S. A., and Milles, J. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cereb. Cortex* 15, 1332.

Schindler, K.A., Bialonski, S., Horstmann, M. T., Elger, C. E., and Lehnertz, K. (2008). Evolving functional network properties and synchronizability during human epileptic seizures. *Chaos* 18, 033119.

Schuster, H. G., and Wagner, P. (1990a). A model for neuronal oscillations in the visual cortex. 1. Mean-field theory and derivation of the phase equations. *Biol. Cybern.* 64, 77.

Schuster, H. G., and Wagner, P. (1990b). A model for neuronal oscillations in the visual cortex. 2. Phase description of the feature dependent synchronization. *Biol. Cybern.* 64, 83.

Sotero, R. C., Trujillo-Barreto, N. J., Iturria-Medina, Y., Carbonell, F., and Jimenez, J. C. (2007). Realistically coupled neural mass models can generate EEG rhythms. *Neural Comput.* 19, 478.

Sporns, O., and Kötter, R. (2004). Motifs in brain networks. *PLoS Biol.* 2, E369. doi: 10.1371/journal.pbio.0020369

Sporns, O., and Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics* 2, 145.

Stam, C., de Haan, W., Daffertshofer, A., Jones, B. F., Manshanden, I., van Cappellen van Walsum, A. M., Montez, T., Verbunt, J. P., de Munck, J. C., van Dijk, B. W., Berendse, H. W., and Scheltens, P. (2009). Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimers disease. *Brain* 132, 213.

Stam, C. J. (2004). Functional connectivity patterns of human magnetoencephalographic recordings: a "small-world" network? *Neurosci. Lett.* 355, 25.

Stam, C. J., Jones, B. F., Nolte, G., Breakspear, M., and Scheltens, P. (2007). Small-world networks and functional connectivity in Alzheimer's disease. *Cereb. Cortex* 17, 92.

Stam, C. J., and Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomed. Phys.* 1, 3.

Stam, C. J., Pijn, J. P., Suffczynski, P., and Lopes da Silva, F. H. (1999a). Dynamics of the human alpha rhythm: evidence for non-linearity? *Clin. Neurophysiol.* 110, 1801.

Stam, C. J., Vliegen, J. H., and Nicolai, J. (1999b). Investigation of the dynamics underlying periodic complexes in the EEG. *Biol. Cybern.* 80, 57.

Stratonovich, R. (1963). *Topics in the Theory of Random Noise*, Vol. I & II. New York, NY: Gordon and Breach.

Strogatz, S. H. (2000). From Kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D* 143, 1.

Supekar, K., Menon, V., Rubin, D., Musen, M., and Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS Comput. Biol.* 4, E1000100. doi: 10.1371/journal.pcbi.1000100

Tass, P. A. (1999). *Phase Resetting in Medicine and Biology*. Berlin: Springer.

Ursino, M., Zavaglia, M., Astolfi, L., and Babiloni, F. (2007). Use of a neural mass model for the analysis of effective connectivity among cortical regions based on high resolution EEG recordings. *Biol. Cybern.* 96, 351.

Valdes, P. A., Jimenez, J. C., Riera, J., Biscay, R., and Ozaki, T. (1999). Nonlinear EEG analysis based on a neural mass model. *Biol. Cybern.* 81, 415.

Van Den Heuvel, M. P., Stam, C. J., Boersma, M., and Hulshoff Pol, H. E. (2008). Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain. *Neuroimage* 43, 528.

Van Rotterdam, A., Lopes da Silva, F. H., van den Ende, J., Viergever, M. A., and Hermans, A. J. (1982). A model of the spatial-temporal characteristics of the alpha rhythm. *Bull. Math. Biol.* 44, 283.

Voss, H. U., and Schiff, N. D. (2009). MRI of neuronal network structure, function, and plasticity. *Prog. Brain Res.* 175, 483.

Wang, L., Zhu, C., He, Y., Zang, Y., Cao, Q., Zhang, H., Zhong, Q., and Wang, Y. (2009). Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Hum. Brain Mapp.* 30, 638.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature* 393, 440.

Wendling, F., Bartolomei, F., Bellanger, J. J., and Chauvel, P. (2001). Interpretation of interdependencies in epileptic signals using a macroscopic physiological model of the EEG. *Clin. Neurophysiol.* 112, 1201.

Wilson, H., R. and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys J.* 12, 1.

Womelsdorf, T., Schoffelen, J. M., Oostenveld, R., Singer, W., Desimone, R., Engel, A. K., and Fries, P. (2007). Modulation of neuronal interactions through neuronal synchronization. *Science* 316, 1609.

Wu, Y., Shang, Y., Chen, M., Zhou, C., and Kurths, J. (2008). Synchronization in small-world networks. *Chaos* 18, 037111.

Zanette, D. H. (2000). Propagating structures in globally coupled systems with time delays. *Phys. Rev. E* 62(Pt A), 3167.

Zhou, C., and Kurths, J. (2006). Dynamical weights and enhanced synchronization in adaptive complex networks. *Phys. Rev. Lett.* 96, 164102.

Zhou, C. S., Zemanová, L., Zamora-López, G., Hilgetag, C. C., and Kurths, J. (2007). Structure-function relationship in complex brain networks expressed by hierarchical synchronization. *New J. Phys.* 9, 1.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 May 2011; accepted: 20 June 2011; published online: 15 July 2011.
*Citation: Daffertshofer A and van Wijk BCM (2011) On the influence of amplitude on the connectivity between phases. Front. Neuroinform. **5**:6. doi: 10.3389/fninf.2011.00006*

APPENDIX

To show the link between the network of Wilson–Cowan models (1) and the Kuramoto network of phase oscillators (2) we adopt Schuster and Wagner's derivation (Schuster and Wagner, 1990). In contrast to their description of two coupled oscillators, however, we explicitly account for a network structure containing N nodes.

When deriving the Kuramoto network, the strategy is to consider the Wilson–Cowan model in the oscillatory regime, i.e., in the presence of a stable limit cycle (**Figure A1**), which is first "mean"-centered simplifying the expansion of the sigmoid function S. Then, the oscillator is averaged over one cycle when assuming that its amplitude and phase change slowly as compared to the oscillator's frequency. That is, time-dependent amplitude and phase are fixed, the system is integrated over one period to remove all harmonic oscillations, and, subsequently, amplitude and phase are again considered to be time-dependent (Guckenheimer and Holmes, 1990) – we note that this procedure is also referred to as a combination of rotating wave approximation and slowly varying amplitude approximation (Haken, 1974). The averaging immediately results in the oscillator network that, when assuming weak coupling and small amplitudes, resembles the Kuramoto network.

More explicitly, let $(E_k^{(0)}, I_k^{(0)})$ be a known solution, for which

$$\begin{aligned}\frac{d}{dt}E_k^{(0)} &= -E_k^{(0)} + S\left[a_E\left(c_{EE}E_k^{(0)} - c_{IE}I_k^{(0)} - \theta_E + P_k + \frac{\eta}{N}\sum_{l=1}^{N} C_{kl}E_l^{(0)}\right)\right]\\ \frac{d}{dt}I_k^{(0)} &= -I_k^{(0)} + S\left[a_I\left(c_{EI}E_k^{(0)} - c_{II}I_k^{(0)} - \theta_I\right)\right]\end{aligned} \tag{A.1}$$

holds. In principle this can be any solution but here we identify $(E_k^{(0)}, I_k^{(0)})$ with the unstable fixed-point (unstable node to be precise) within the stable limit cycle (see the intersection point of the nullclines in **Figure A1**). We investigate the deviation of this solution by means of

$$\begin{aligned}\frac{d}{dt}\left(E_k^{(0)} + \delta E_k\right) &= -\left(E_k^{(0)} + \delta E_k\right)\\ &\quad + S\left[\chi_{E,k}^{(0)} + a_E\left(c_{EE}\delta E_k - c_{IE}\delta I_k + \frac{\eta}{N}\sum_{l=1}^{N} C_{kl}\delta E_l\right)\right]\\ \frac{d}{dt}\left(I_k^{(0)} + \delta I_k\right) &= -\left(I_k^{(0)} + \delta I_k\right) + S\left[\chi_{I,k}^{(0)} + a_I\left(c_{EI}\delta E_k - c_{II}\delta I_k\right)\right]\end{aligned} \tag{A.2}$$

where we abbreviated

$$\begin{aligned}\chi_{E,k}^{(0)} &= a_E\left(c_{EE}E_k^{(0)} - c_{IE}I_k^{(0)} - \theta_E + P_k + \frac{\eta}{N}\sum_{l=1}^{N} C_{kl}E_l^{(0)}\right)\\ \chi_{I,k}^{(0)} &= a_I\left(c_{EI}E_k^{(0)} - c_{II}I_k^{(0)} - \theta_I\right)\end{aligned} \tag{A.3}$$

As said, this "mean"-centering allows for expanding the sigmoid function to the M-th order, $S[x_0 + \delta x] \approx S[x_0] + \sum_{n=1}^{M} \frac{1}{n!} S^{(n)}[x_0]\delta x^n$; here $S^{(n)}$ denotes the n-th derivative of S; see also **Figure A2**. Inserting this expansion into (A.2) yields the following system differential Eq.

$$\begin{aligned}\frac{d}{dt}\delta E_k &= -\delta E_k\\ &\quad + \sum_{n=1}^{M}\frac{1}{n!}S^{(n)}\left[\chi_{E,k}^{(0)}\right]\left(a_E\left(c_{EE}\delta E_k - c_{IE}\delta I_k + \frac{\eta}{N}\sum_{l=1}^{N} C_{kl}\delta E_l\right)\right)^n\\ \frac{d}{dt}\delta I_k &= -\delta I_k + \sum_{n=1}^{M}\frac{1}{n!}S^{(n)}\left[\chi_{I,k}^{(0)}\right]\left(a_E\left(c_{EI}\delta E_k - c_{II}\delta I_k\right)\right)^n\end{aligned} \tag{A.4}$$

Here the zero-th order $S^{(0)}$ cancels because of (A.1). The system (A.4) is weakly non-linear presuming M is small implying the presence of only low-order polynomial terms. Put differently, the sigmoid function S is evaluated close to its threshold. For the sake of simplicity we here use $M = 3$. Furthermore we set overall coupling strength η to be small, i.e., we drop all terms containing η^2 or higher orders in η. By this (A.4) can be reduced to

$$\begin{aligned}\frac{d}{dt}\delta E_k &= -\delta E_k\\ &\quad + \left(\sum_{n=1}^{M}\frac{1}{n!}S^{(n)}\left[\chi_{E,k}^{(0)}\right]\left(a_E\left(c_{EE}\delta E_k - c_{IE}\delta I_k\right)\right)^n\right)\left(1 + \frac{\eta}{N}\sum_{l=1}^{N} C_{kl}\delta E_l\right)\\ \frac{d}{dt}\delta I_k &= -\delta I_k + \sum_{n=1}^{M}\frac{1}{n!}S^{(n)}\left[\chi_{I,k}^{(0)}\right]\left(a_E\left(c_{EI}\delta E_k - c_{II}\delta I_k\right)\right)^n\end{aligned} \tag{A.5}$$

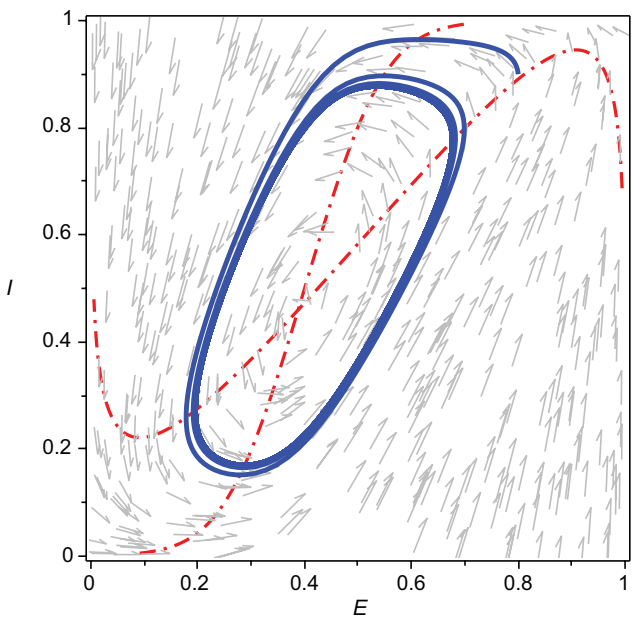


FIGURE A1 | Limit cycle oscillations of a single Wilson–Cowan oscillator,

$$\frac{d}{dt}E = -E + S\left[a_E\left(c_{EE}E - c_{IE}I - \theta_E + P\right)\right]$$
$$\frac{d}{dt}I = -I + S\left[a_I\left(c_{EI}E - c_{II}I - \theta_I\right)\right]$$

i.e., Eq. 1 with $N = 1$; the sigmoid function was set to $S[x] = (1 + e^{-x})^{-1}$. Dot-dashed lines represent the nullclines (temporal derivatives of E and I vanish). At the intersection of the nullclines is an unstable node. Parameter values: $a_E = 1.2$, $a_I = 2$, $c_{EE} = c_{II} = 10$, $c_{IE} = 6$, $c_{EI} = 1$, $\theta_E = 2$, $\theta_I = 3.5$, $P = 0.5$.

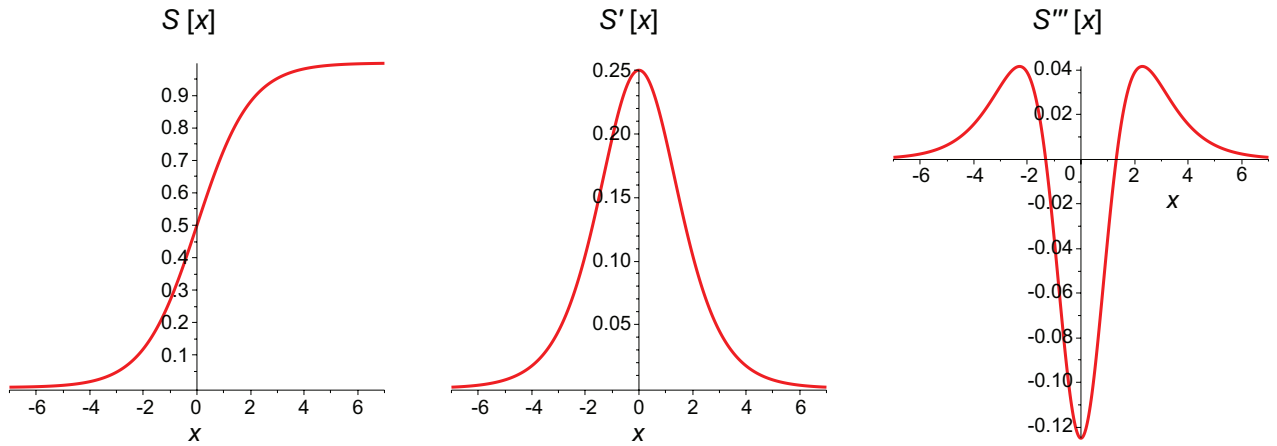


FIGURE A2 | Shape of the sigmoid function $S[x] = 1/(1 + e^{-x})$ and its first and third derivatives. In the vicinity of the threshold $x = 0$, the third derivative (right panel) is negative allowing for the existence of the stationary amplitude $R_{k,\text{stationary}}$ as given in (A.14) dependent on parameter settings.

which represents a network of weakly non-linear, self-sustained oscillators. Conventionally its characteristics are studied after transforming the system into polar coordinates

$$\delta E_k = R_k \cos\left(\Omega t + \varphi_k\right) \text{ and } \delta I_k = R_k \sin\left(\Omega t + \varphi_k\right) \tag{A.6}$$

where R_k and φ_k are the time-dependent amplitude and phase, respectively, of the network node k, and Ω is a yet unknown (mean) frequency. As said, we assume that amplitude and phase change slowly with respect to Ω, and average the system (A.5) over a cycle $t = [0 \ldots 2\pi/\Omega)$. This averaging yields the phase dynamics as

$$\begin{aligned}\frac{d}{dt}\varphi_k = \omega_k &+ \frac{\eta}{2N}\sum_{l=1}^{N} C_{kl} a_E S'\left[\chi_{E,k}^{(0)}\right]\frac{R_l}{R_k}\sin\left(\varphi_l - \varphi_k\right) \\ &+ \frac{\eta}{16N}\sum_{l=1}^{N} C_{kl} a_E^3 S'''\left[\chi_{E,k}^{(0)}\right] R_k R_l \left(\left(c_{EE}^2 + 3c_{IE}^2\right)\sin\left(\varphi_l - \varphi_k\right)\right. \\ &\left. + 2c_{EE}c_{IE}\cos\left(\varphi_l - \varphi_k\right)\right)\end{aligned} \tag{A.7}$$

with S' and S''' referring to the first and third derivative of the sigmoid function S, respectively (see **Figure A.2**), and the frequency being given by

$$\begin{aligned}\omega_k = -\Omega &+ \frac{1}{2}\left(a_E c_{IE} S'\left[\chi_{E,k}^{(0)}\right] + a_I c_{EI} S'\left[\chi_{I,k}^{(0)}\right]\right) \\ &+ \frac{1}{16}\left(a_E^3 c_{IE}\left(c_{EE}^2 + c_{IE}^2\right)S'''\left[\chi_{E,k}^{(0)}\right] + a_I^3 c_{EI}\left(c_{EI}^2 + c_{II}^2\right)S'''\left[\chi_{I,k}^{(0)}\right]\right)R_k^2\end{aligned} \tag{A.8}$$

As a last approximation, we consider the case in which all amplitudes R_k are sufficiently small so that their quadratic and higher orders can be ignored. We note that R_k are the amplitudes of the limit cycles describing $(\delta E_k, \delta I_k)$ which do not agree with the mean "activities" of the Wilson–Cowan oscillators as they are shifted by $(E_k^{(0)}, I_k^{(0)})$. Discarding these higher order terms finally leads to

$$\omega_k \approx -\Omega + \frac{1}{2}\left(a_E c_{IE} S'\left[\chi_{E,k}^{(0)}\right] + a_I c_{EI} S'\left[\chi_{I,k}^{(0)}\right]\right) \tag{A.9}$$

and

$$\frac{d}{dt}\varphi_k \approx \omega_k + \frac{\eta}{2N}\sum_{l=1}^{N} a_E S'\left[\chi_{E,k}^{(0)}\right]\frac{R_l}{R_k} C_{kl}\sin\left(\varphi_l - \varphi_k\right) \tag{A.10}$$

which is equivalent to (2).

For the sake of completeness we also list the natural frequencies Ω_k of the uncoupled and linearized Wilson–Cowan oscillators:

$$\begin{aligned}\Omega_k^2 = &\frac{1}{4} - \left(a_E c_{EE} S'\left[\chi_{E,k}^{(0)}\right] + a_I c_{II} S'\left[\chi_{I,k}^{(0)}\right]\right)^2 \\ &- a_E c_{IE} S'\left[\chi_{E,k}^{(0)}\right]\left(a_E c_{EE} S'\left[\chi_{E,k}^{(0)}\right] - 1\right)\end{aligned} \tag{A.11}$$

with which Ω in (A.9) can be defined via averaging over nodes, i.e., $\Omega = N^{-1}\sum_{k=1}^{N}\Omega_k$. Furthermore the amplitude dynamics corresponding to the phase Eq. (A.7) reads

$$\begin{aligned}\frac{d}{dt}R_k = &\frac{1}{2}\left(a_E c_{IE} S'\left[\chi_{E,k}^{(0)}\right] - a_I c_{EI} S'\left[\chi_{I,k}^{(0)}\right] - 2\right)R_k \\ &+ \frac{1}{16}\left(a_E^3 c_{EE}\left(c_{EE}^2 + c_{IE}^2\right)S'''\left[\chi_{E,k}^{(0)}\right]\right. \\ &\left. - a_I^3 c_{II}\left(c_{II}^2 + c_{EI}^2\right)S'''\left[\chi_{I,k}^{(0)}\right]\right)R_k^3 \\ &- \frac{\eta}{N}\sum_{l=1}^{N} C_{kl} a_E S'\left[\chi_{E,k}^{(0)}\right] R_l \cos\left(\varphi_l - \varphi_k\right) \\ &+ \frac{\eta}{16N}\sum_{l=1}^{N} C_{kl} a_E^3 S'''\left[\chi_{E,k}^{(0)}\right] R_k^2 R_l \left(\left(3c_{EE}^2 + c_{IE}^2\right)\cos\left(\varphi_l - \varphi_k\right)\right. \\ &\left. + 2c_{EE}c_{IE}\sin\left(\varphi_l - \varphi_k\right)\right)\end{aligned} \tag{A.12}$$

When ignoring all coupling terms (i.e., setting $\eta = 0$), this amplitude dynamics of such isolated Wilson–Cowan oscillators reduces to

$$\begin{aligned}\frac{d}{dt}R_k = &\frac{1}{2}\left(a_E c_{IE} S'\left[\chi_{E,k}^{(0)}\right] - a_I c_{EI} S'\left[\chi_{I,k}^{(0)}\right] - 2\right)R_k \\ &+ \frac{1}{16}\left(a_E^3 c_{EE}\left(c_{EE}^2 + c_{IE}^2\right)S'''\left[\chi_{E,k}^{(0)}\right] - a_I^3 c_{II}\left(c_{II}^2 + c_{EI}^2\right)S'''\left[\chi_{I,k}^{(0)}\right]\right)R_k^3\end{aligned} \tag{A.13}$$

Which has the stationary solutions $R_{k,\text{stationary}} = 0$ and

$$\begin{aligned}&R_{k,\text{stationary}} \\ &= \pm\sqrt{8\frac{2 - a_E c_{EE} S'\left[\chi_{E,k}^{(0)}\right] + a_I c_{II} S'\left[\chi_{I,k}^{(0)}\right]}{a_E^3 c_{EE}\left(c_{EE}^2 + c_{IE}^2\right)S'''\left[\chi_{E,k}^{(0)}\right] - a_I^3 c_{II}\left(c_{II}^2 + c_{EI}^2\right)S'''\left[\chi_{I,k}^{(0)}\right]}}\end{aligned} \tag{A.14}$$

provided the square-root exists; cf. **Figure A2**.

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 27 March 2012
doi: 10.3389/fninf.2012.00010

Correspondence between structure and function in the human brain at rest

Judith M. Segall[1,2]*, Elena A. Allen[1], Rex E. Jung[1,3], Erik B. Erhardt[1], Sunil K. Arja[1], Kent Kiehl[1] and Vince D. Calhoun[1,4]

[1] *Mind Research Network, Albuquerque, NM, USA*

[2] *Department of Family and Community Medicine, University of New Mexico, Albuquerque, NM, USA*

[3] *Department of Neurosurgery, University of New Mexico, Albuquerque, NM, USA*

[4] *Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA*

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Trygve B. Leergaard, University of Oslo, Norway
Thomas R. Knösche, Max Planck Institute for Human Cognitive and Brain Sciences, Germany

****Correspondence:***
Judith M. Segall, Mind Research Network, 1101 Yale Boulevard NE, Albuquerque, NM 87106, USA.
e-mail: jsegall@mrn.org

To further understanding of basic and complex cognitive functions, previous connectome research has identified functional and structural connections of the human brain. Functional connectivity is often measured by using resting-state functional magnetic resonance imaging (rs-fMRI) and is generally interpreted as an indirect measure of neuronal activity. Gray matter (GM) primarily consists of neuronal and glia cell bodies; therefore, it is surprising that the majority of connectome research has excluded GM measures. Therefore, we propose that by exploring where GM corresponds to function would aid in the understanding of both structural and functional connectivity and in turn the human connectome. A cohort of 603 healthy participants underwent structural and functional scanning on the same 3 T scanner at the Mind Research Network. To investigate the spatial correspondence between structure and function, spatial independent component analysis (ICA) was applied separately to both GM density (GMD) maps and to rs-fMRI data. ICA of GM delineates structural components based on the covariation of GMD regions among subjects. For the rs-fMRI data, ICA identified spatial patterns with common temporal features. These decomposed structural and functional components were then compared by spatial correlation. Basal ganglia components exhibited the highest structural to resting-state functional spatial correlation ($r = 0.59$). Cortical components generally show correspondence between a single structural component and several resting-state functional components. We also studied relationships between the weights of different structural components and identified the precuneus as a hub in GMD structural network correlations. In addition, we analyzed relationships between component weights, age, and gender; concluding that age has a significant effect on structural components.

Keywords: structural, functional, networks, source-based morphometry, independent component analysis, resting-state, gray matter density

INTRODUCTION

A central assumption of systems neuroscience is that the structure of the brain can predict and/or is related to functional connectivity. This belief is derived from basic human anatomy and biomechanics where the structure and form of body parts are directly related to their function. The structure–function relationship is found at different scales in nature, from the molecular composition of enzymes, the morphology of organometallics, to the collective behavior of ant colonies. For the past 20 years, the field of neuroimaging has demonstrated that function and behavior arise from specific regions in the brain. Structural adaptations in the cortex have been found in plasticity studies. For example, people who recently acquired the ability to juggle exhibit changes in gray matter (GM) volumes of the mid-temporal area (MT/V5) and intraparietal sulcus (Draganski et al., 2004), in professional female ballet dancers there are distinct differences in white and gray matter compared to controls (Hanggi et al., 2010), and in musicians, greater cortical thickness is found in superior temporal and dorsolateral frontal regions as well, as increased GM concentration (GMC) in aspects of the Heschl's gyrus (Bermudez et al., 2009) when compared to non-musicians. These structural alterations reflect subjects' specialized, and, in some cases, exceptional, functional abilities. However, recent developments in neuroimaging have shifted the structural–functional relationship away from distinct brain regions and toward distributed function, with the view that cognition is the result of the "dynamic interactions of distributed brain areas operating in a large-scale network" (Bressler and Menon, 2010). The concept that the human brain is a complex network of neurons linking physical structure to function (Power et al., 2010) is not new to cognitive science.

In his seminal paper, Sporns coined the term the human connectome, which is, "a comprehensive structural description of the network of elements and connections forming the human brain (Sporns et al., 2005)." Theoretically, by mapping the networks of the human brain, we will strengthen our understanding of how functional brain activity emerges from anatomical structure. This

knowledge will provide a more comprehensive model of cognition and produce new insights into how brain functions are affected if there are structural irregularities (Hagmann et al., 2010). Multiple studies of schizophrenia have already identified both structural network abnormalities (Bassett et al., 2008) and functional network connectivity differences (Lynall et al., 2010) when compared to health controls. The majority of current brain network studies tend to focus on one connectome modality, either structural or functional.

A large-scale functional network is defined as a "collection of interconnected brain areas that interact to perform circumscribed functions (Bressler and Menon, 2010)." Functional networks can be identified using electroencephalography (EEG), magnetoencephalography (MEG), and functional magnetic resonance imaging (fMRI). Recent functional network studies have used resting-state fMRI (rs-fMRI), which measures spontaneous, high-amplitude, (mostly) low-frequency (<0.1 Hz) blood-oxygen-level-dependent (BOLD) signal fluctuations in subjects who are at rest. Several different analysis approaches, including seed-based correlation maps (Biswal et al., 1995; Fox et al., 2005) and independent component analysis (ICA; Damoiseaux et al., 2006; Biswal et al., 2010; Allen et al., 2011) have identified from rs-fMRI a large number of functional networks. These networks appear to provide robust measures of the intrinsic functional activity of the brain (Miller et al., 2009) and have been identified both in resting data and data collected during a task (Calhoun et al., 2008). Because these intrinsic networks (INs) exhibit moderately high reliability (Shehzad et al., 2009; Zuo et al., 2010a), interrater and intermethod reliability (Franco et al., 2009; Zuo et al., 2010b), and consistency (Damoiseaux et al., 2006), they provide a framework for studying the functional architecture of the human connectome (Biswal et al., 2010; Allen et al., 2011) and are a key focus of this study.

Structural networks of the human brain have typically been constructed directly using various white matter (WM) connectivity measurements obtained from diffusion weighted imaging (DWI; Bassett and Bullmore, 2009) and constructed using graph theoretical techniques. Indeed, in much of the literature structural connectivity is obtained from diffusion imaging (Honey et al., 2010). DWI can quantify and identify structural connectivity by tracking WM bundle pathways that link to cortical regions (Guye et al., 2008). Structural networks have also been inferred indirectly from the inter-regional covariation of GM volume or cortical thickness and usually measured at the group level (Sporns, 2011). Using covariance measures for specific ROIs, Mechelli et al. (2005a) reported that the "gray matter densities (GMD) of different regions of the human cortex is coordinated within an individual." Inter-regional covariation of GM volume has also shown differences in network organization between healthy participants and those with schizophrenia (Bassett and Bullmore, 2009). Other studies have examined cortical thickness to construct GM structural networks, for example He et al. (2007) used the inter-regional correlation of cortical thickness measurements to construct structural networks. Additionally, modularity analysis of the relationships between structural cortical networks identified modules similar to known functional domains, such as sensorimotor, visual, auditory/language, strategic/executive, and mnemonic processing (Chen et al., 2008).

These prior studies highlight efforts to separately explore the connections in structural networks or in functional networks. The majority of studies that incorporate both structural and functional imaging to investigate the human brain connectome tend to use rs-fMRI and WM analysis. Several papers have recently reviewed these studies (Rykhlevskaia et al., 2008; Bassett and Bullmore, 2009; Damoiseaux and Greicius, 2009; Honey et al., 2010; Sporns, 2011) and in general, concluded that when structural connectivity is high, functional connectivity tends to be high as well (Koch et al., 2002).

When comparing the relationship between anatomic structure and functional connectivity only a few studies have used GM (Calhoun et al., 2006; Seeley et al., 2009; Michael et al., 2010; Supekar et al., 2010). This is somewhat surprising considering that fMRI is generally interpreted as an indirect measure of neuronal activity and GM primarily consists of neuronal and glia cell bodies (Logothetis, 2002). We propose that exploring where GM corresponds to function would aid in the understanding of both structural and functional connectivity and in turn the human connectome. We will not discuss the current debate about the exact origins of the BOLD response measured through fMRI; however, the relationship between the BOLD signal and the underlying neuronal activation is an area of active debate and should be interpreted carefully when making direct inferences between neuronal activity (Ekstrom, 2010).

This study has three primary aims, centered on the use of GM to assess structural–functional spatial relationships of the human brain. The first aim is to identify GM structural components using GMD measurements and its variation among a large cohort of healthy individuals ($n = 603$). This will be ascertained with source-based morphometry (SBM; Xu et al., 2009), which applies spatial ICA to find patterns of GMD with common covariation among subjects. In this paper, the term structural components will refer to the components revealed by SBM. Association between age, gender, and the structural components sources will be further assessed. We expect to see a general reduction in GMD as age increases, particularly in regions, such as the parietal and frontal cortex, where reductions in GMD with age have been reported to previously (Sowell et al., 2003).

The second aim is to compare structural components with previously determined rs-functional components and determine their spatial similarity. This will be done by spatially correlating structural components from the SBM analysis with functional components from a group ICA (GICA) of rs-fMRI previously reported by Allen et al. (2011). Both of the structural and functional components were obtained from the same set of subjects. Our hypothesis is that there will be correspondence between structural and functional components, particularly in the precuneus and posterior cingulate cortex (PCC) regions of the default mode network (DMN) given that the linkage between structure and function is particularly strong in these regions (Hagmann et al., 2008; Skudlarski et al., 2008; Honey et al., 2009; van den Heuvel et al., 2009). We also assume that there will not be a complete direct correspondence between structure and function; consequently, we expect to find fewer structural components than functional networks. This is because the presence of functional connectivity has been observed when there are no

supportive structural connections, at least as reflected in DTI (Greicius et al., 2009).

The third and final aim is an exploratory investigation of the relationships between the different structural components and for the purpose of this paper, is referred to as structural network correlations (SNC). Specifically, SNC will be performed by investigating the correlations among the structural component loading parameters. Networks and components are often used interchangeably in the literature, although the definitions of networks and components are not always consistent. Therefore, Erhardt et al. (2011) suggested to always define the term network when it is used. Consequently, our usage of networks (SNC) and components are defined in the Section "Materials and Methods."

Our results, using a novel approach that utilizes regions of covariation, generally show correspondence between structure and function and further elucidate the relationship between function at rest and GM. Additionally, our results corroborate with previous findings on the effects of age on GMD. Lastly, from our findings we suggest structural and functional regions that warrant further investigations.

MATERIALS AND METHODS

Full details on participants, data collection, and image processing can be found in Allen et al. (2011). For completeness, we briefly provide pertinent information here.

PARTICIPANTS

This analysis combined existing data from 603 healthy participants scanned on the same scanner and spread across 34 studies and 18 principal investigators at the Mind Research Network (MRN). Informed consent was obtained from all subjects according to institutional guidelines at the University of New Mexico (UNM) and all data were anonymized prior to group analysis. The cohort is nearly balanced on gender (305 females) with similar age distributions across genders. Because the sample is overwhelmingly right-handed (46 ambidextrous or left-handed individuals), handedness will not be considered in this study. The age range is 12–71 with a strong right skew (mean = 23.4; SD = 9.2), thus as in Allen et al. (2011) we use the normalizing transformation, log (age), to reduce the leverage of older subjects in correlation and regression analyses.

DATA COLLECTION

All MR images were collected on a 3-Tesla Siemens Trio scanner. High-resolution T1-weighted structural images were acquired with a 5-echo multi-echo MPRAGE sequence with TE = 1.64, 3.5, 5.36, 7.22, and 9.08 ms, TR = 2.53 s, TI = 1.2 s, flip angle = 7°, number of excitations = 1, slice thickness = 1 mm, field of view = 256 mm, resolution = 256 × 256. T2*-weighted functional images were acquired using a gradient-echo EPI sequence with TE = 29 ms, TR = 2 s, flip angle = 75°, slice thickness = 3.5 mm, slice gap = 1.05 mm, field of view 240 mm, matrix size = 64 × 64, voxel size = 3.75 mm × 3.75 mm × 4.55 mm. Resting-state scans were a minimum of 5 min, 4 s in duration (truncated to 152 volumes for all subjects). Participants were instructed to keep their eyes open during the scan and stare passively at a presented fixation cross.

IMAGE PREPROCESSING

The structural data the T1 images were preprocessed through an automated pipeline developed at MRN (Bockholt et al., 2010). First the images were resliced to 2 mm × 2 mm × 2 mm voxels. Tissue classification, bias correction, image registration, and spatial normalization were automatically performed using voxel-based morphometry (VBM) in SPM5[1], wherein the above steps are integrated into a unified model (Ashburner and Friston, 2005). Unmodulated GM segmentations, which produce an estimation of local GMD, were smoothed using a Gaussian kernel with a full-width at half-maximum (FWHM) of 10 mm (**Figure 1**; step 1, left side). The smoothed GMD images were then correlated to an *a priori* GM template to access segmentation outliers. Those GMD images that were not highly correlated to the Montreal Neurological Institute (MNI) template in SPM5 where manually adjusted to the AC–PC line and rerun through our automated pipeline, where they were segmented and smoothed again. GMD is the probability distribution of the GM proportion of a voxel and the term is synonymous with GMC, whereby concentration and density are used interchangeably in neuroimaging literature. The relative density or concentration of GM, from non-modulated VBM, is the proportion of GM relative to other tissue types (WM, cerebrospinal fluid) within a region (Mechelli et al., 2005b). As

[1]http://www.fil.ion.ucl.ac.uk/spm/software/spm5

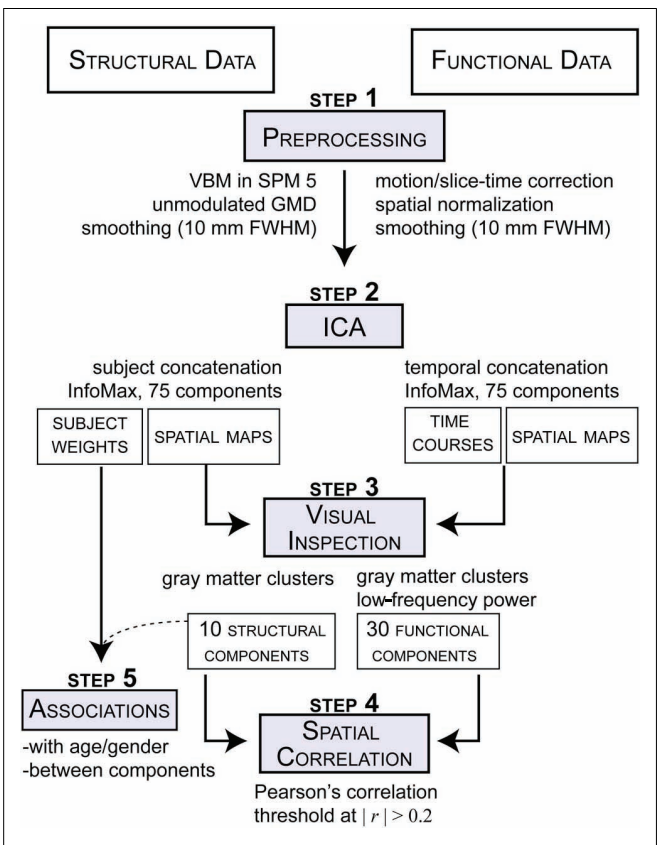


FIGURE 1 | Schematic of preprocessing and analyses for both structural GMD images and rs-functional images.

a caveat, GMD is sensitive in detecting some local GM structural properties, but lacks some specificity for particular structural properties.

The fMRI data underwent typical preprocessing of realignment, slice-timing correction, spatial normalization to a template in standard MNI space, reslicing (3 mm × 3 mm × 3 mm voxels) and spatial smoothing (FWHM = 10 mm; **Figure 1**; step 1, right side). The spatial maps were evaluated for outliers and if possible were corrected using the same realignment procedure that was used for the GM segmentations. Subsequent to automated preprocessing, the data were intensity normalized to improve the accuracy and test–retest reliability of the ICA output.

SOURCE-BASED MORPHOMETRY

Source-based morphometry is a multivariate analysis, similar to VBM, used to examine the relationships between GMD regions (see Xu et al., 2009 for further details). GMD images from each subject were flattened into row vectors and stacked to form the subjects-by-voxel matrix upon which spatial ICA was applied (Calhoun et al., 2001). ICA linearly decomposed the GMD matrix into a mixing matrix (subjects-by-components) that represents the relative strength (weight) of components for each subject and the source matrix (voxels-by-components) that represents the maximally spatially independent GMD sources. ICA was performed with the GIFT toolbox[2] using the infomax algorithm (**Figure 1**; step 2, left side). We evaluated GM maps decomposed at several different model orders (number of components). The model orders investigated were 20, 40, 60, 75, 80, and 100. We found similar components at the different model orders and ultimately used the high model order of 75 components to match the number of components used in the rs-fMRI analysis, as discussed in the next section. Briefly, model orders 60 and 80 yielded comparable components as the model order of 75, which was validated by correlational analyses and visual inspection. For the purpose of this paper, sources of GM covariation obtained from this the SBM analysis will be referred to in this paper as structural components.

GROUP INDEPENDENT COMPONENT ANALYSIS OF fMRI DATA

Resting-state data were decomposed into components using spatial ICA to identify temporally coherent networks and their associated time courses by estimating maximally independent spatial sources from their linearly mixed fMRI signals. For this study, spatial sources obtained from the resting-state data will be referred to as rs-functional components. GICA was also performed using the GIFT toolbox with a model order ICA of 75 components (**Figure 1**; step 2, right side). This model order has been noted in the literature to yield refined components that correspond to known anatomical and functional segmentations (Kiviniemi et al., 2009; Smith et al., 2009; Abou-Elseoud et al., 2010; Ystad et al., 2010). See Allen et al. (2011) for a complete treatment of the GICA implementation. For the purpose of this study the functional components were resliced to 2 × 2 × 2 to match the dimensions of the structural components.

[2]http://mialab.mrn.org/software

FEATURE IDENTIFICATION

All 75 structural components were visually inspected by three reviewers and the GM composition of each component was evaluated (**Figure 1**; step 3, left side). We excluded structural components that had significant spatial overlap with ventricles, WM, large vasculature, and the brainstem, or were located at the boundaries between these regions and GM. These criteria were designed to exclude any component that were of possible mixed tissue sources, such that structural components for subsequent analysis only included GM. Of the 10 structural components that met the inclusion criteria, eight comparable components were identified in each of the model orders mentioned previously. The two remaining components were only not observed in the lower model orders of 20 and 40, but were found in the other model orders. For the rs-functional components, we followed guidelines similar to those used by Allen et al. (2011) to select a subset of functional components; however, we were slightly less stringent and included two additional subcortical and cerebellar components that were excluded from prior analyses (for further details on rs-functional component selection, please see Allen et al., 2011).

STATISTICAL ANALYSIS

To assess spatial correspondence between structural and rs-functional components, we calculated all pair-wise Pearson correlations between the selected structural and functional component spatial maps, yielding a n-by-m correlation matrix, where n is the number of selected structural components and m is the number of selected rs-functional components (**Figure 1**; step 4).

Structural component loading parameters, representing the contribution of each component to a given subject, were also used in additional association analyses (**Figure 1**; step 5). Pearson correlations were computed between the structural loading parameters and the log-transformed (age). Finally, all pair-wise correlations between the structural loading parameters of the selected structural components were computed (SNC). Here, we also used partial correlations to remove the possible effect of age on between-component associations.

RESULTS

STRUCTURAL–FUNCTIONAL COMPONENT CORRELATIONS

Of 75 structural components, $n = 10$ met the inclusion criteria; for rs-functional components, $m = 30$ were selected for analysis. Out of the 10-by-30 structural and rs-functional component comparisons, 24 structural–functional component pairs were above the determined correlation coefficient threshold of $|r| > 0.20$. Note that this threshold also conservatively represents a significance level of $p < 0.005$, corrected. Accounting for spatial smoothness in the spatial maps and assuming, as in Smith et al. (2009), roughly 500 degrees of freedom, a correlation of $r = 0.2$ has a p-value of 6×10^{-6}, which when Bonferroni correcting for 300 tests is 0.002. The 24 structural–functional component pairs are presented in order of decreasing correlation coefficient magnitude and divided into groups with similar spatial topography (see **Figures 2–4**).

BASAL GANGLIA COMPONENTS

Subcortical structures comprising the basal ganglia had the highest structural–rs-functional component correlations (0.59 and 053).

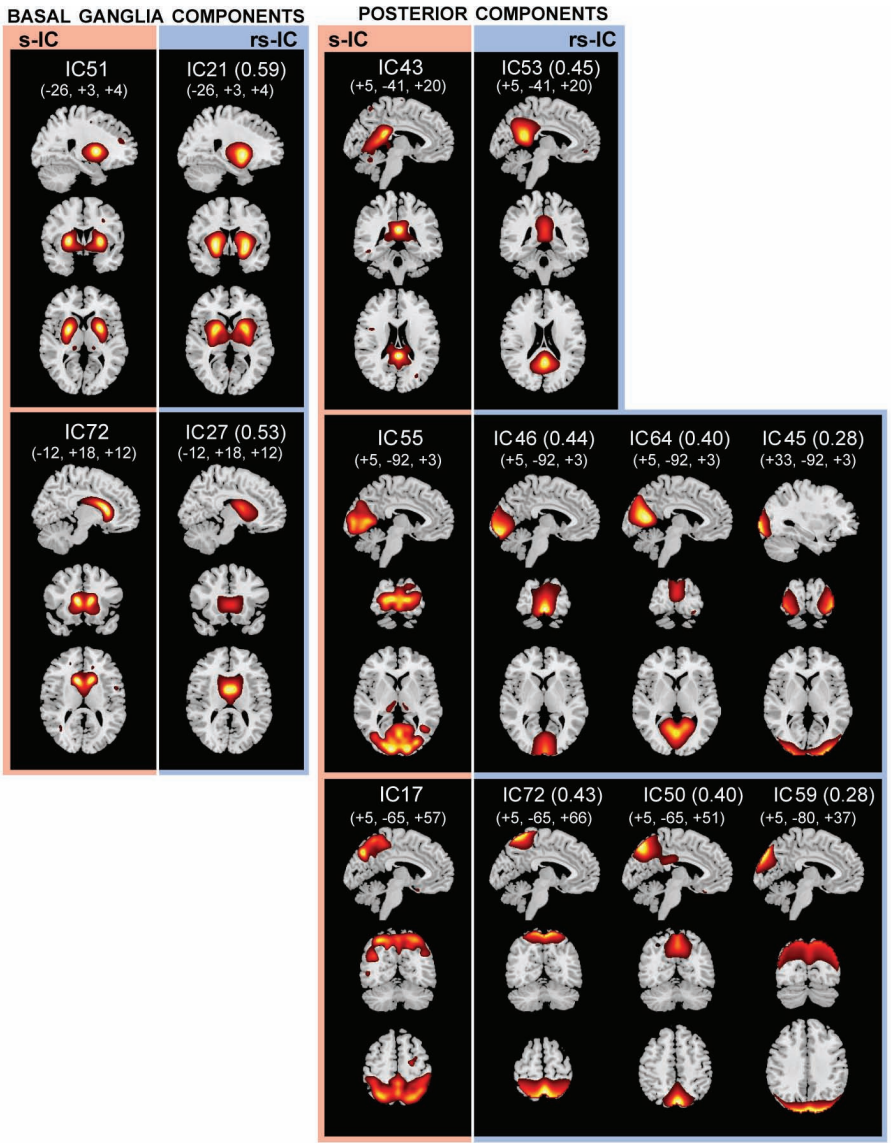


FIGURE 2 | The structural (sMRI) components (red) and corresponding rs-fMRI components (blue). The spatial correlation between component pairs is indicated adjacent to the functional component number. Both sMRI and fMRI aggregate components were converted to *z*-scores and thresholded at $Z > 2$. Structural components are displayed at the slices with peak activation, indicated as (*x*, *y*, *z*) coordinates in MNI space. When structural components are paired with a single functional component, the functional component is displayed at the same slices. If a structural component corresponds to several fMRI components, functional components may be displayed at different coordinates that best represent their activation.

Structural components, s-IC51 and s-IC72, were respectively comprised of the bilateral putamen and the bilateral caudate. They corresponded to rs-functional components, rs-IC21 and rs-IC27, which were primarily composed of the left and right putamen and the bilateral caudate, respectively (see **Figure 2**; **Table 1**). In both the structural and rs-functional components, there was only one component meeting the correlation threshold for each of these respective structures, i.e., the pairings were distinct and unique.

POSTERIOR COMPONENTS

Posterior components showed the second highest set of correlations (see **Figure 2**; **Table 1**). Component s-IC43, primarily comprised of the PCC, is correlated to rs-IC53, which is comprised of the PCC, the L/R angular gyri and the medial frontal gyrus (MFG). Component s-IC55, which contains voxels spanning much of the occipital cortex, is correlated to three rs-functional components. In order of correlation magnitude they are rs-IC46, rs-IC64, and rs-IC45, which represent aspects of the medial and lateral visual cortex. Component s-IC17, which peaks at the precuneus and extends laterally, is also correlated to three rs-functional components. In order of correlation magnitude these are rs-IC72 and rs-IC50 which also largely cover the precuneus, and rs-IC59, which represents more posterior activation over the bilateral cuneus.

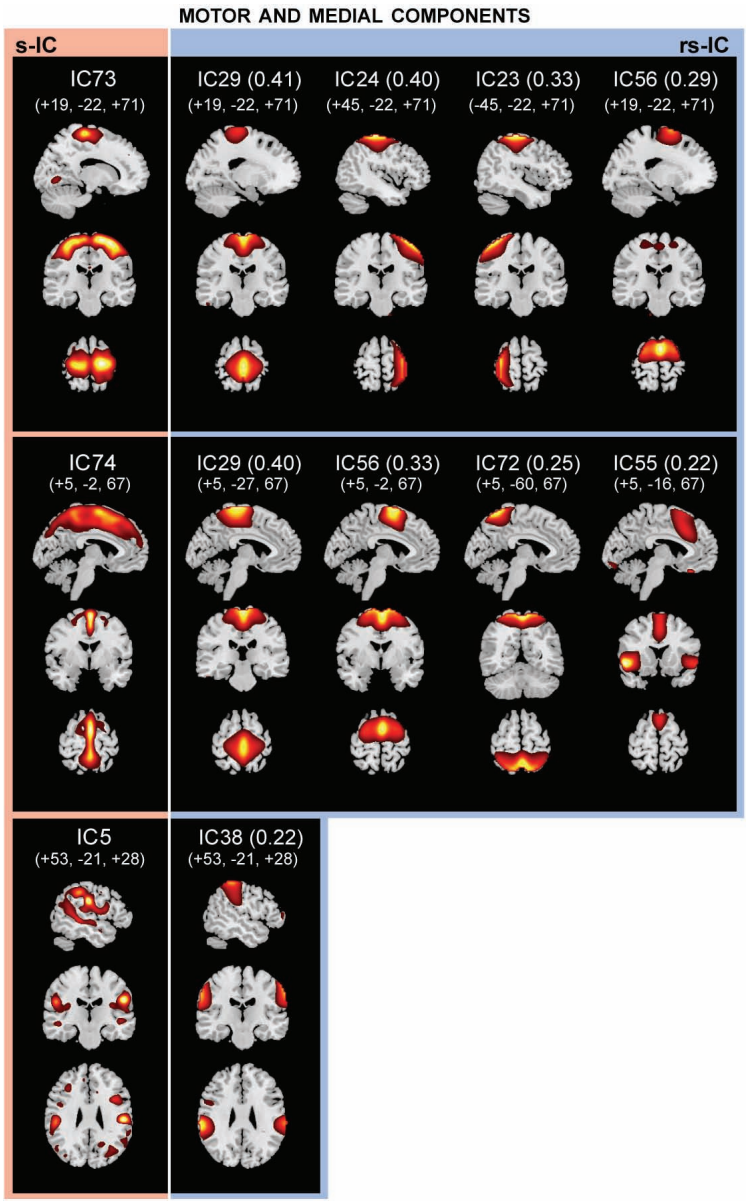


FIGURE 3 | The structural (sMRI) components (red) and corresponding rs-fMRI components (blue). The spatial correlation between component pairs is indicated adjacent to the functional component number. Both sMRI and fMRI aggregate components were converted to *z*-scores and thresholded at $Z > 2$. Structural components are displayed at the slices with peak activation, indicated as (*x*, *y*, *z*) coordinates in MNI space. When structural components are paired with a single functional component, the functional component is displayed at the same slices. If a structural component corresponds to several fMRI components, functional components may be displayed at different coordinates that best represent their activation.

MOTOR AND MEDIAL COMPONENTS

Notably, structural components determined to be motor and medial components are related to multiple functional components (**Figure 3**; **Table 2**). A large component, s-IC73, spanning the supplementary motor areas (SMA) and bilateral pre- and post-central gyri correlates to four rs-functional components. These are rs-IC29, with peaks at the bilateral paracentral lobule and left insula, rs-IC24 and rs-IC23, which represent lateralized aspects of the motor system, and rs-IC56, which is centered at the SMA. A second structural component, s-IC74, is also quite large and extends over much of the medial surface, particularly in the MFG. Component s-IC74 is correlated to three rs-functional components, two of which are also correlated to s-IC73 and one of which is correlated to s-IC17. In order of correlation magnitude these are rs-IC29, rs-IC56, rs-IC72, and rs-IC55, which represents the bilateral cingulate gyrus, the left and right insula, and the bilateral MFG. A third structural component, s-IC5, largely comprised of the bilateral supramarginal gyrus (SMG), is weakly correlated to rs-IC38, which represents activations over similar regions.

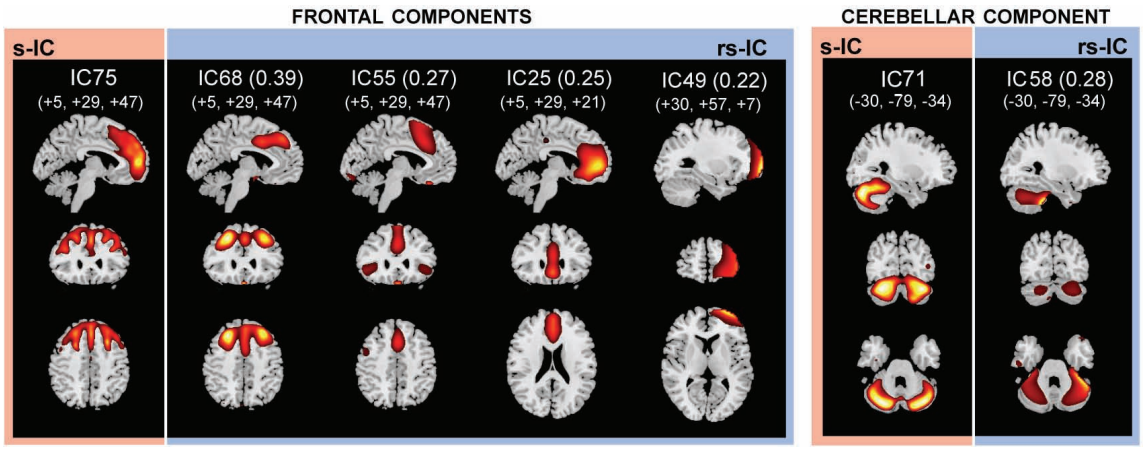


FIGURE 4 | The structural (sMRI) components (red) and corresponding rs-fMRI components (blue). The spatial correlation between component pairs is indicated adjacent to the functional component number. Both sMRI and fMRI aggregate components were converted to *z*-scores and thresholded at $Z > 2$. Structural components are displayed at the slices with peak activation, indicated as (*x*, *y*, *z*) coordinates in MNI space. When structural components are paired with a single functional component, the functional component is displayed at the same slices. If a structural component corresponds to several fMRI components, functional components may be displayed at different coordinates that best represent their activation.

FRONTAL AND CEREBELLAR COMPONENTS

The frontal component, s-IC75, which is primarily comprised of the left MFG and the right SFG is correlated to four rs-functional components. These rs-functional components in order of correlation magnitude are rs-IC68, rs-IC55, and rs-IC2, which are primarily comprised of the MFG, the cingulate gyrus, the insula, and the anterior cingulate cortex. Lastly, s-IC71 is correlated with functional component, rs-IC58. Both components largely represent cerebellar cortex (**Figure 4**; **Table 3**).

STRUCTURAL COMPONENTS AGE AND GENDER CORRELATIONS

Pearson correlations between log(age) and the structural component loading parameters are uniformly negative (**Figure 5A**). **Figure 5B** shows an example of this negative correlation, with a scatter-plot of the loading parameters for component 17 (precuneus) as a function of age for all 603 subjects. The trend predicts almost a 50% decrease in component weights from adolescence (∼12 years age) to the age of retirement (∼65–70 years). We found no significant correlations between gender and the structural loading parameters.

STRUCTURAL NETWORK CORRELATIONS

The cross-correlation matrix between the structural loading parameters is shown in the top half of the correlation matrix in **Figure 6A**, wherein the majority of correlation coefficients are above values of $r = 0.2$. Because we found associations between age and loading parameters for all components, we also performed a correlation analysis after adjusting for age (bottom half of correlation matrix; **Figure 6A**). Partialling out variance due to age weakened all the correlations; however, a few structural component loading parameter pairs stayed significantly correlated after age adjustment. An example of this is in **Figure 6B**, which shows the relationship between loading parameters of s-IC17 and s-IC73 before ($r = 0.68$) and after ($r = 0.48$) adjusting for age. **Figure 6C** shows an example of the relationship between components largely due to age: the correlation between s-IC5 and s-IC73 loading parameters falls from $r = 0.58$ to 0.24 after adjusting for age.

We can also create a graph of the correlations to elucidate the more complex relationships (beyond pair-wise) between structural components. Here, we used a conservative threshold of $r > 0.4$ to create a graph, as shown in **Figure 7**. Based on the original correlation values, this yields 13 "edges" between 6 component "nodes." Using the age-corrected correlations, we find a graph with four "edges" between four "nodes." In both the original and age-corrected correlations, component s-IC17 (the precuneus) is identified as a "hub," for it was the component with the greatest number of correlations.

DISCUSSION

Investigating GM structural networks is a crucial next step in mapping the correspondence between structure and function in the human brain. To our knowledge, this is the first study to link GM structure and function using spatial components, obtained from high model order spatial ICA, and from GM structural and rs-fMRI. We have also developed a framework for processing and analyzing GM structure and function in the same large cohort of healthy individuals. Our overall goal was to assess structural–functional relationships of the human brain and we found several GM structural components that spatially corresponded to rs-functional components. We used spatial component correspondence as our framework for investing how structure relates to function, but if we had a single functional parameter instead of group obtained functional components then we could have delved deeper into the covariation of structural–functional correspondence across individual subjects. An example of a single functional parameter for future investigation is the amplitude of low-frequency fluctuation (ALFF) of the rs-fMRI signal, because the literature suggests that ALFFs reflect the intensity of regional

Table 1 | Structural and corresponding rs-functional components representing basal ganglia and posterior components.

Structural components					|r|	Resting-state functional components				
	BA	Volume	max *z*	MNI coordinates			BA	Volume	max *t*	MNI coordinates
Basal ganglia components										
s-IC51					0.59	rs-IC21				
R putamen		203.88	16	25, 3, 1		R putamen		53.85	108.7	25, −1, 0
L putamen		187.38	15.6	−25, 1, 2		L putamen		52.11	108.7	−25, −3, 0
s-IC72					0.53	rs-IC27				
Bi caudate		511.88	16.2	0, 13, 11		Bi caudate		141.56		0, 5, 8
						R posterior cingulate gyrus		11.96		1, −38, 14
Posterior components										
s-IC43					0.45	rs-IC53				
Bi posterior cingulate cortex	23	420.38	18.7	1, −47, 15		Bi posterior cingulate	23	88.41	139.6	0, −52, 22
						L angular gyrus	39	12.30	71.5	−43, −69, 33
						R angular gyrus	39	7.19	59.8	47, −66, 32
						Bi medial frontal gyrus	10	2.26	50.7	−1, 45, −9
s-IC55					0.44	rs-IC46				
Bi calcarine gyrus	17, 18	825.38	7.7	0, −84, 4		Bi lingual gyrus	17, 18	135.33	87.3	1, −87, −2
						Bi middle cingulate cortex	31	8.52	34.1	1, −45, 32
s-IC55 (see above)					0.40	rs-IC64				
						Bi calcarine gyrus	17, 18	136.81	117.9	1, −71, 13
s-IC55 (see above)					0.28	rs-IC45				
						Bi calcarine gyrus	18	104.62	101.7	4, −92, −2
s-IC17					0.43	rs-IC72				
Bi precuneus	7	1191.63	7.5	0, −64, 49		Bi precuneus	7	121.59	105.2	0, −53, 61
						L superior frontal gyrus	9	4.11	35.8	−32, 38, 39
						R middle frontal gyrus	6	3.15	32.4	26, 0, 60
						L middle frontal gyrus	6	2.96	32.4	−23, 0, 63
						R superior frontal gyrus	9	1.96	30.3	33, 39, 35
s-IC17 (see above)					0.39	rs-IC50				
						Bi precuneus	7	107.48	102.5	1, −64, 43
s-IC17 (see above)					0.28	rs-IC59				
						Bi cuneus	19	114.04	113.7	2, −84, 28

Contains absolute correlation coefficient (|r|), s-IC and rs-IC number, Brodmann area (BA), volume of cluster in microliter, MNI coordinate of peak voxel in LPI convention (millimeter), and list of clusters that comprise each component.

Table 2 | Structural and corresponding rs-functional components representing motor and medial components.

Structural components					|r|	Resting-state functional				
	BA	Volume	max *z*	MNI coordinates			BA	Volume	max *t*	MNI coordinates
Motor and medial components										
s-IC73					0.41					
Bi supplementary motor area	6	1012.88	8	4, −21, 61		rs-IC29				
						Bi paracentral lobule	6	118.48	100.9	1, −28, 61
						L insula	13	1.63	40.1	−35, −24, 15
s-IC73 (see above)					0.40	rs-IC24				
						R precentral gyrus	4	143.78	83.3	37, −25, 53
						R middle temporal gyrus		6.11	36.1	50, −64, −2
						L cerebellum		3.67	26.1	−20, −53, −24
						L middle temporal gyrus		2.33	23.5	−61, −28, −8
						L middle temporal gyrus		1.63	24.5	−51, −70, 4
s-IC73 (see above)					0.33	rs-IC23				
						L precentral gyrus	4	134.19	86.8	−35, −27, 53
						R cerebellum		12.67	40.3	24, −52, −23
						R postcentral gyrus		9.33	35.6	44, −28, 56
						R inferior frontal gyrus	45	2.93	23.7	54, 29, 0
						R precuneus		2.81	26.7	8, −62, 32
s-IC73 (see above)					0.28	rs-IC56				
						Bi supplementary motor area	6	139.63	122.7	1, −3, 61
						R superior temporal gyrus	22	7.15	48.2	50, 8, 4
						L inferior frontal gyrus	44	5.52	42.7	−53, 5, 14
						R inferior parietal lobule	40	2.26	41.1	58, −29, 24
						L inferior parietal lobule	40	0.96	35.9	−58, −32, 23
s-IC74					0.39	rs-IC29 (see above)				
Bi medial frontal gyrus		771.38	9.6	0, −10, 57						
s-IC74 (see above)					0.33	rs-IC56 (see above)				

(Continued)

Table 2 | Continued

Structural components					|r|	Resting-state functional				
	BA	**Volume**	**max *z***	**MNI coordinates**			**BA**	**Volume**	**max *t***	**MNI coordinates**
Motor and medial components										
s-IC74 (see above)					0.25	rs-IC72				
						Bi precuneus	7	121.59	105.2	0, −53, 61
						L superior frontal gyrus	9	4.11	35.8	−32, 38, 39
						R middle frontal gyrus	6	3.15	32.4	26, 0, 60
						L middle frontal gyrus	6	2.96	32.4	−23, 0, 63
						R superior frontal gyrus	9	1.96	30.3	33, 39, 35
s-IC74 (see above)					0.21	rs-IC55				
						Bi cingulate gyrus	32	44.81	92.8	0, 22, 45
						L insula	47	24.81	103.1	−46, 15, −5
						R insula	47	12.26	80.8	45, 18, −6
						L middle frontal gyrus	10	8.04	65.4	−32, 53, 21
s-IC5					0.22	rs-IC38				
L supramarginal gyrus	2	183.13	7.64	−52, −26, 27		L supramarginal gyrus	2	51.00	110.5	−55, −34, 37
R supramarginal gyrus	2	341.38	11.3	53, −26, 32		R supramarginal gyrus	2	35.67	96.2	56, −32, 40
L inferior frontal gyrus	44	52.75	6.5	−33, 36, 17		L inferior frontal gyrus	44	7.67	58.6	−48, 5, 18
L middle temporal gyrus	37	60.75	9.32	−46, −55, 9		L middle temporal gyrus	37	4.74	54.4	−57, −60, −2
						Bi middle cingulate cortex	24	7.00	51.5	1, 7, 38

Contains absolute correlation coefficient (|r|), s-IC and rs-IC number, Brodmann area (BA), volume of cluster in microliter, MNI coordinate of peak voxel in LPI convention (millimeter), and list of clusters that comprise each component.

Table 3 | Structural and corresponding rs-functional components representing frontal and cerebellar components.

Structural components						Resting-state functional				
	BA	**Volume**	**max *z***	**MNI coordinates**	**|*r*|**		**BA**	**Volume**	**max *t***	**MNI coordinates**
Frontal components										
s-IC75					0.39	rs-IC68				
L middle frontal gyrus	8	465.25	4.75	−28, 46, 27		L middle frontal gyrus	8	55.19	95.2	−26, 26, 42
R superior medial gyrus		233.75	4.39	14, 47, 21		R middle frontal gyrus	8	44.81	87.9	26, 33, 41
						Bi middle cingulate cortex	32	16.67	67.6	0, 21, 40
s-IC75 (see above)					0.27	rs-IC55				
						Bi cingulate gyrus	32	44.81	92.8	0, 22, 45
						L insula	47	24.81	103.1	−46, 15, −5
						R insula	47	12.26	80.8	45, 18, −6
						L middle frontal gyrus	10	8.04	65.4	−32, 53, 21
s-IC75 (see above)					0.25	rs-IC25				
						Bi anterior cingulate cortex	32	115.78	114.5	0, 41, 4
						Bi middle cingulate cortex	31	13.26	53.6	1, −30, 41
						R inferior frontal gyrus		3.44	48.2	32, 22, −15
						R middle frontal gyrus	46	2.33	37.8	40, 43, 8
s-IC75 (see above)					0.22	rs-IC49				
						R middle frontal gyrus	10	61.52	84.3	31, 55, 7
						L pyramis		5.33	42.2	−39, −66, −44
						L middle frontal gyrus	10	2.37	33.4	−31, 52, 8
Cerebellar components										
s-IC71					0.28	rs-IC58				
Cerebellar vermis		1119.88	5.76	−1, −68, −29		Cerebellar vermis		167.48	94.2	7, −55, −37
						L middle frontal gyrus	8	12.07	33.7	−31, 44, 33

Contains absolute correlation coefficient (|r|), s-IC and rs-IC number, Brodmann area (BA), volume of cluster in microliter, MNI coordinate of peak voxel in LPI convention (millimeter), and list of clusters that comprise each component.

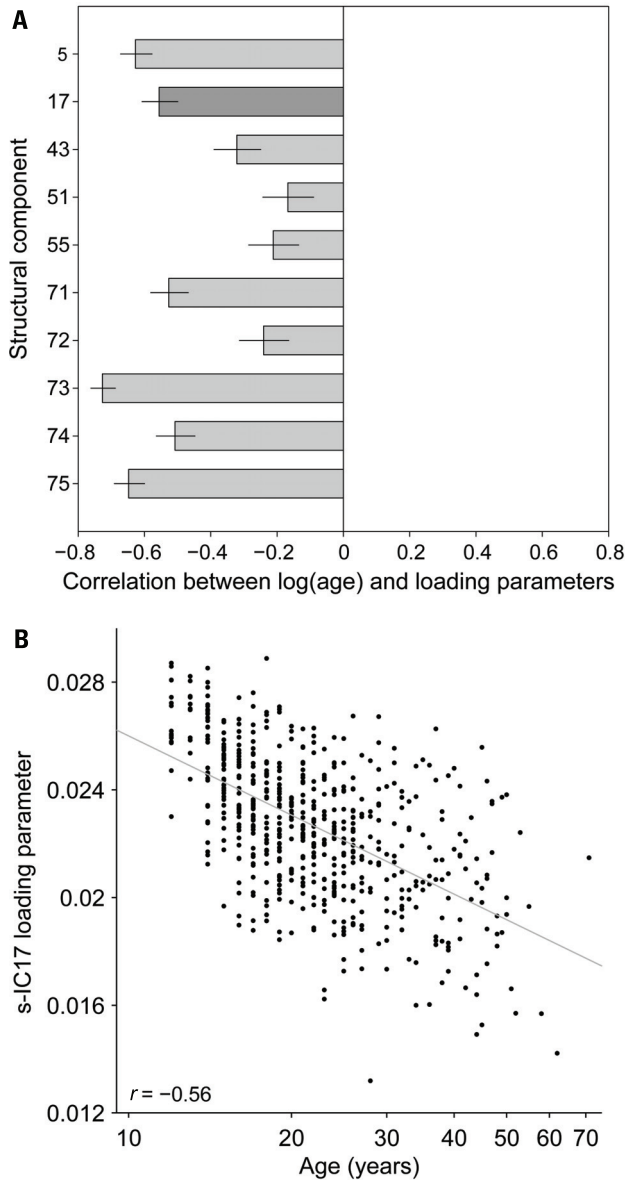


FIGURE 5 | (A) Structural components correlated to the log(age) for the 10 selected s-IC. Dark gray bar corresponds to s-IC17. In general, structural components are negatively correlated with age. Ninety-five percent confidence intervals are reported. **(B)** Correlation of the age (years) and loading parameter for all 603 subjects using s-IC17, which is primarily comprised of the precuneus, as an example. Where $r = -0.56$.

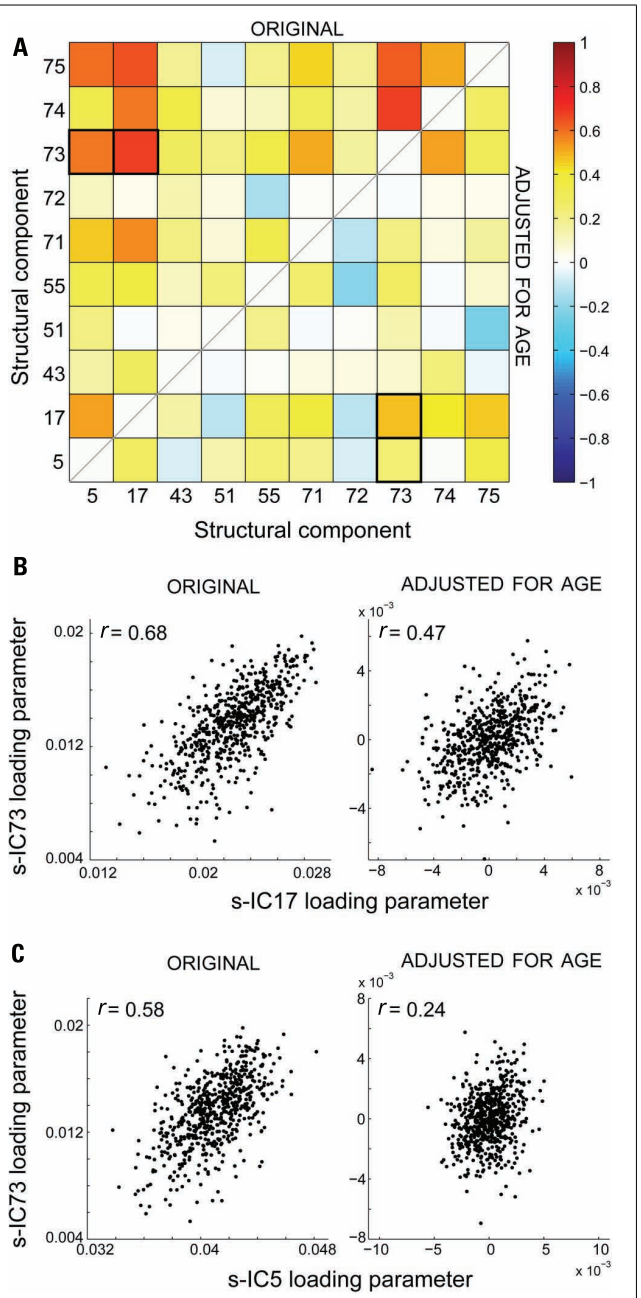


FIGURE 6 | (A) Structural network correlations (SNC) of the 10 selected structural components. The top half of the matrix is without age adjustment (original) and the bottom half is after age adjustment. **(B)** An example of the effect of adjusting for age, where the adjustment does not remove the significant correlation between the components for s-IC73 and s-IC17. **(C)** An example of the effect of adjusting for age, where the adjustment does remove the significant correlation between the components for s-IC73 and s-IC5.

spontaneous brain activity (Yang et al., 2007). Using ALFFs, GMD, and parallel ICA (Calhoun et al., 2009) would be a good next direction for future studies.

The basal ganglia components are the most spatially correlated structural–functional components. Additionally, it is one of only a few structural components in which the component was comprised of only one source, which indicates how different the GMD is in the basal ganglia compared to other regions in the brain. As seen in **Tables 1–3**, the majority of the components are composed of several GM sources. This component could have also been so clearly identified because the GMD of the basal ganglia is so different compared to the surround WM. This structure–function pairing is also unique because it is one of only four pairs (out of the 24 structural–functional pairs) that has a direct

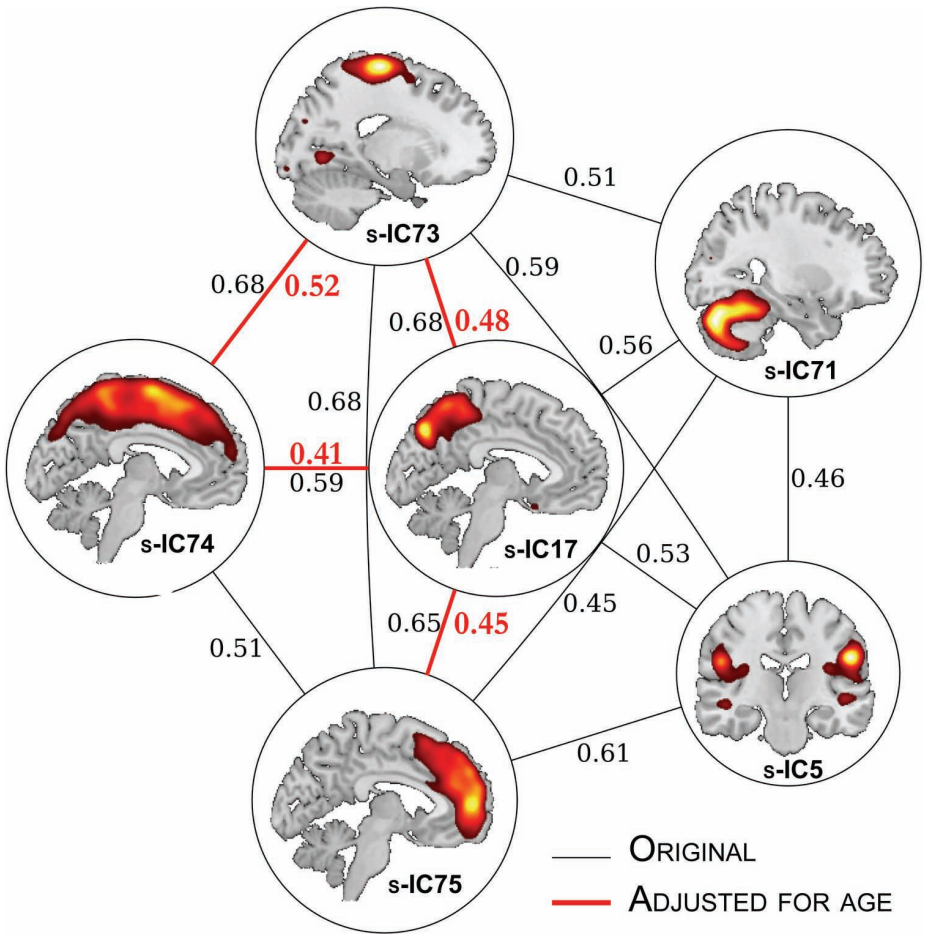


FIGURE 7 | Structural network correlations. The SNC results for component pairs that are $r < 0.4$. The edges in red refer to components pairs that survived age adjustment. The correlation coefficient values are adjacent to the edges.

one-to-one correspondence to a rs-functional component. Another structural–functional component pair where this is seen in the cerebellar component; however, the cerebellar correlation value, though significant, is the weakest of the entire structure–function analysis – nearly half the strength as in the basal ganglia components. In similar structural network studies the focus tends to be on the cortex, consequently, paying less attention to subcortical regions, such as the basal ganglia (Robinson et al., 2009; Bressler and Menon, 2010). Our finding further promotes the basal ganglia as a viable and interesting region that should be addressed in further structure–function studies.

Another directly correlated structure–functional pair is found in the components that are primarily comprised of the SMG. Results from lesion studies of aphasic stroke patients have found that the left SMG plays a role in acoustic-phonetic processing, which is an example of how structural abnormalities directly relate to functional processing (Caplan et al., 1995).

The remaining structural–functional pair with direct correspondence is found in the posterior component, specifically comprised of the PCC. This PCC is commonly seen at rest and is considered part of the DMN (Buckner et al., 2008). The DMN is a particular grouping of brain regions that are consistently found to be active during the resting-state (Raichle et al., 2001; Raichle and Snyder, 2007; Buckner et al., 2008). The precuneus, also considered part of the DMN, is seen in our structural–functional component correlations. Although, unlike the PCC, it is correlated to several functional components, which is consistent with the precuneus having many functional roles in addition to its role in the DMN. The precuneus exhibits functional connectivity in several highly integrated tasks, such as episodic memory retrieval, self-referential processing and visuo-spatial imagery (Cavanna and Trimble, 2006). The precuneus along with the PCC have also been identified as part of structural core in a graph study of diffusion imaging data (Hagmann et al., 2008). Prior structural connectivity studies have reflected the functional connectivity of the DMN (Greicius et al., 2009; Skudlarski et al., 2010) and as hypothesized, our study is another example of how structure corresponds to function in the DMN. Additionally, the DMN regions have shown a high degree of heritability (Glahn et al., 2010) and alterations in the DMN have been found in many neurological and psychiatric disorders (Garrity et al., 2007; Greicius, 2008; Paakki et al., 2010; Weng et al., 2010).

The remaining structural–functional component correlations are comprised of one structural components that corresponds to several functional components. Typically, the structural component is a broad region of GM, while the functional components are broken up into smaller and sometimes lateralized components. This type of structural–functional correspondence is found in the frontal, SMA, and visual regions. The structural components comprised of the frontal and SMA are correlated to some of the same functional components. These frontal networks are associated with strategic and executive functions (Duncan and Owen, 2000) and the SMA networks are primarily associated with sensorimotor/spatial functions. Unlike, the frontal and SMA components, the functional components comprising a visual network do not correspond to any other structural components, besides s-IC55, which is mainly composed of the calcarine gyrus.

Interestingly, we found similar spatial overlap between function and structure as Chen et al. (2008) which found that the cortex was organized into six topological modules. The lack of more direct correspondence between structure and function was not surprising to us. Recent studies have shown that structural changes in cortical thickness believed to be induced by activity are not found in the same regions where there are functional connectivity changes (Haier et al., 2009).

The field of human connectome research could benefit from examining the similarities that exist between patterns of GM covariation and functional connectivity in healthy individuals, for these patterns may be a foundation for future research on both healthy connectivity and changes associated with neurodegenerative disorders. A previous study found using GM volume measures and ICA on rs-fMRI that, "normal intrinsic connectivity and structural covariance patterns mirrored each other and reflected, with high fidelity, those regions that codegenerate in distinct human neurodegenerative syndromes (Seeley et al., 2009)." This study differed from ours in several ways, mainly that the GM regions selected were those associated with neurodegenerative syndromes. Indeed, both in Seeley et al. (2009) and in our study, GM structure is directly observed to be associated with function; however, the exact mechanism is still unclear. A study on the developing brain demonstrated that different GM structural covariance networks exist at various developmental stages and as children aged their GM covariance topology eventually resembled an adults intrinsic connectivity network (Zielinski et al., 2010). Zielinski et al. (2010) findings suggest that GM structural covariance networks may mature after functional coactivation. Perhaps aging and the effects of neurodegenerative syndromes are possible mechanisms as to how structural GM covariation occurs. The normal aging process in healthy controls could also explain our results of structure–function correlations.

As predicted age had a strong affect on the structural components, for decreases in GM volume are thought to be both from maturational and degenerative changes (Taki et al., 2011) and this study age range spanned from adolescence to later adulthood (12–71 years). After adjusting for age, the strength of our between-component correlations was not significantly mitigated for all of the structural component pairs. Future studies should evaluate structural components for distinct age groups to determine age specific structural covariance patterns. Surprisingly, we did not find significant associations with gender and structural loading parameters. There are inconsistencies in the literature regarding sex differences and GM (Sowell et al., 2007), which could be attributed to the differences in methods used to obtain GM measurements. Our study was comprised of a large age rage of males and females and that may be why we did not find a main effect of gender.

The SNC analysis revealed several structural components that were highly covariant and similar to regions that were found in other structural–functional correspondence studies. Specifically, the precuneus was linked to frontal, cerebral, parietal, and motor areas. This relationship with the precuneus and the other structural components survives after adjusting for age, which is a possible indicator that the precuneus might play a role of a structural "hub" in the SNC. The finding of a precuneus "hub" is consistent with a previous study, which through mapping structural cortico-cortical pathways identified the precuneus as one of the hubs of the structural core of the human cortex (Hagmann et al., 2008). The subcortical components were not as strongly correlated to the other cortical components, which could be caused by differences in types of GM.

The findings from our structural–functional analysis are relevant to human connectome research, for the correspondence between structural and functional covariations provides us with information about brain connectivity. A meta-analysis by Smith et al. (2009) demonstrated that covariation of functional networks, from over 1,600 functional neuroimaging studies, are similar to the functional networks found at rest. Therefore, INs can be identified from inter-subject covariation. Additionally, a recent ICA paper showed that estimates of functional activity can be accurately predicted from covariation analysis, which again provides us with information that functional covariation is directly related to connectivity (Calhoun and Allen, in press). This current paper is extending this work further, by identifying both structural covariation components and components where structural–functional covariations correspond. Regions where we have identified correspondence are perhaps highly dependent on structure. The regions where we have not shown structural–functional correspondence are also of interest. For instance, we did not find any structural or functional covariations in regions such as the prefrontal cortex, which are associated with higher cognitive functions (Jung and Haier, 2007).

There are several limitations to this study, the first being the investigation of structural networks using GMD. GMD only provides indirect measurements of structural networks; therefore, since we do not use DWI and directly evaluate WM tract involvement we are unable to make direct statements about anatomical connectivity (Stam, 2010). Yet, as identified in a primate study, the number of neurons differ depending if the GM is gyral or not (Hilgetag and Barbas, 2005). Therefore, it is no longer as straightforward to infer that the underlying anatomical connectivity corresponds to GM measures, such as cortical thickness. Future work should incorporate all three measures (fMRI, structural MRI, and DWI). The methods we used were to describe the relationship between structure and function at the group level, consequently the next step would be to investigate this

relationship at the individual level. To further substantiate our findings of structural–functional correspondence, future studies should also investigate structural and functional covariations between structural–functional pairs at the subject level to ensure the findings are replicable and reliable.

Additionally, we conducted a very constrained analysis that investigated a limited number of components. We decided to use a strict feature selection criteria to limit our analysis and results to components that were explicitly within GM. We do not find it surprising that we eliminated 65 components, for SBM is most likely demonstrating that the brain is comprised of complex morphometry that cannot be easily characterized by a voxel-wise map. The observed SBM findings of the separate components identified for the basal ganglia and in the cerebellum are interesting, for those regions have well known stereotypical anatomical architectures. These findings provide additional evidence that GMD covariation detects regions of the brain that share similar anatomical architectures. Our study was also comprised from healthy individuals over a wide age range and future studies could explore how these structural–functional relationship manifest at different age stages.

CONCLUSION

The use of non-invasive neuroimaging provides the ability to describe and find structural and functional networks, fostering opportunities to further understand the complexity of the human brain. Using the multivariate approach, SBM, we found GM structural relationship patterns and several areas in the brain where structure and function correspond. More importantly, we also demonstrated that GM structural components are directly associated with functional components.

ACKNOWLEDGMENTS

The authors would like to thank Sergey Plis and Eswar Damarju with their help generating figures and tables. This research was supported by NIH 1R01-EB006841, NIH 1R01-EB005846, NIH 2R01-EB000840, NIH 1 P20 RR021938-01, and DOE DE-FG02-08ER64581 (PI: Calhoun). Support for data collection by NIMH 1R01-MH072681-01 (PI: Kiehl); John Templeton Foundation Grant #12456 (PI: Jung); NRC Bilatgrunn (PI: Eichele); NIH 1P20 AA017068 (PI: Kodituwakku); NIH R21NS064464, NIH 1 R03 DA022435-01A1, and NIH 1 R03 DA024212-01A1 (PI: Mayer); NIH KO1-DA021632-02 (PI: Filbey).

REFERENCES

Abou-Elseoud, A., Starck, T., Remes, J., Nikkinen, J., Tervonen, O., and Kiviniemi, V. (2010). The effect of model order selection in group PICA. *Hum. Brain Mapp.* 31, 1207–1216.

Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., Havlicek, M., Rachakonda, S., Fries, J., Kalyanam, R., Michael, A. M., Caprihan, A., Turner, J. A., Eichele, T., Adelsheim, S., Bryan, A. D., Bustillo, J., Clark, V. P., Feldstein Ewing, S. W., Filbey, F., Ford, C. C., Hutchison, K., Jung, R. E., Kiehl, K. A., Kodituwakku, P., Komesu, Y. M., Mayer, A. R., Pearlson, G. D., Phillips, J. P., Sadek, J. R., Stevens, M., Teuscher, U., Thoma, R. J., and Calhoun, V. D. (2011). A baseline for the multivariate comparison of resting-state networks. *Front. Syst. Neurosci.* 5:2. doi: 10.3389/fnsys.2011.00002

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851.

Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., and Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *J. Neurosci.* 28, 9239–9248.

Bassett, D. S., and Bullmore, E. T. (2009). Human brain networks in health and disease. *Curr. Opin. Neurol.* 22, 340–347.

Bermudez, P., Lerch, J. P., Evans, A. C., and Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb. Cortex* 19, 1583–1596.

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34, 537–541.

Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A.-M., Ernst, M., Fair, D., Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kötter, R., Li, S.-J., Lin, C.-P., Lowe, M. J., Mackay, C., Madden, D. J., Madsen, K. H., Margulies, D. S., Mayberg, H. S., Mcmahon, K., Monk, C. S., Mostofsky, S. H., Nagel, B. J., Pekar, J. J., Peltier, S. J., Petersen, S. E., Riedl, V., Rombouts, S. A., Rypma, B., Schlaggar, B. L., Schmidt, S., Seidler, R. D., Siegle, G. J., Sorg, C., Teng, G.-J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X.-C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y.-F., Zhang, H.-Y., Castellanos, F. X., and Milham, M. P. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739.

Bockholt, H. J., Scully, M., Courtney, W., Rachakonda, S., Scott, A., Caprihan, A., Fries, J., Kalyanam, R., Segall, J. M., De La Garza, R., Lane, S., and Calhoun, V. D. (2010). Mining the mind research network: a novel framework for exploring large scale, heterogeneous translational neuroscience research data sources. *Front. Neuroinform.* 3:36. doi: 10.3389/neuro.11.036.2009

Bressler, S. L., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38.

Calhoun, V. D., Adali, T., Giuliani, N. R., Pekar, J. J., Kiehl, K. A., and Pearlson, G. D. (2006). Method for multimodal analysis of independent source differences in schizophrenia: combining gray matter structural and auditory oddball functional data. *Hum. Brain Mapp.* 27, 47–62.

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. J. (2001). A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151.

Calhoun, V. D., and Allen, E. A. (in press). Extracting intrinsic functional networks with feature-based group independent component analysis. *Psychometrika.*

Calhoun, V. D., Kiehl, K. A., and Pearlson, G. D. (2008). Modulation of temporally coherent brain networks estimated using ICA at rest and during cognitive tasks. *Hum. Brain Mapp.* 29, 828–838.

Calhoun, V. D., Liu, J., and Adali, T. (2009). A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* 45, S163–S172.

Caplan, D., Gow, D., and Makris, N. (1995). Analysis of lesions by MRI in stroke patients with acoustic-phonetic processing deficits. *Neurology* 45, 293–298.

Cavanna, A. E., and Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. *Brain* 129, 564–583.

Chen, Z. J., He, Y., Rosa-Neto, P., Germann, J., and Evans, A. C. (2008). Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. *Cereb. Cortex* 18, 2374–2381.

Damoiseaux, J. S., and Greicius, M. D. (2009). Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity. *Brain Struct. Funct.* 213, 525–533.

Damoiseaux, J. S., Rombouts, S. A., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., and Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13848–13853.

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., and May, A. (2004). Neuroplasticity: changes in grey matter induced by training. *Nature* 427, 311–312.

Duncan, J., and Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends Neurosci.* 23, 475–483.

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: the danger

in dissociation. *Brain Res. Rev.* 62, 233–244.

Erhardt, E. B., Allen, E. A., Damaraju, E., and Calhoun, V. D. (2011). On network derivation, classification, and visualization: a response to Habeck and Moeller. *Brain Connect.* 1, 1–19.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678.

Franco, A. R., Pritchard, A., Calhoun, V. D., and Mayer, A. R. (2009). Interrater and intermethod reliability of default mode network selection. *Hum. Brain Mapp.* 30, 2293–2303.

Garrity, A. G., Pearlson, G. D., Mckiernan, K., Lloyd, D., Kiehl, K. A., and Calhoun, V. D. (2007). Aberrant "default mode" functional connectivity in schizophrenia. *Am. J. Psychiatry* 164, 450–457.

Glahn, D. C., Winkler, A. M., Kochunov, P., Almasy, L., Duggirala, R., Carless, M. A., Curran, J. C., Olvera, R. L., Laird, A. R., Smith, S. M., Beckmann, C. F., Fox, P. T., and Blangero, J. (2010). Genetic control over the resting brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1223–1228.

Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. *Curr. Opin. Neurol.* 21, 424–430.

Greicius, M. D., Supekar, K., Menon, V., and Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* 19, 72–78.

Guye, M., Bartolomei, F., and Ranjeva, J. P. (2008). Imaging structural and functional connectivity: towards a unified definition of human brain organization? *Curr. Opin. Neurol.* 21, 393–403.

Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Ellen Grant, P., Wedeen, V., Meuli, R., Thiran, J. P., Honey, C. J., and Sporns, O. (2010). MR connectomics: principles and challenges. *J. Neurosci. Methods* 194, 34–45.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6, e159. doi: 10.1371/journal.pbio.0060159

Haier, R. J., Karama, S., Leyba, L., and Jung, R. E. (2009). MRI assessment of cortical thickness and functional activity changes in adolescent girls following three months of practice on a visual–spatial task. *BMC Res. Notes* 2, 174. doi: 10.1186/1756-0500-2-174

Hanggi, J., Koeneke, S., Bezzola, L., and Jancke, L. (2010). Structural neuroplasticity in the sensorimotor network of professional female ballet dancers. *Hum. Brain Mapp.* 31, 1196–1206.

He, Y., Chen, Z. J., and Evans, A. C. (2007). Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb. Cortex* 17, 2407–2419.

Hilgetag, C. C., and Barbas, H. (2005). Developmental mechanics of the primate cerebral cortex. *Anat. Embryol.* 210, 411–417.

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., and Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040.

Honey, C. J., Thivierge, J. P., and Sporns, O. (2010). Can structure predict function in the human brain? *Neuroimage* 52, 766–776.

Jung, R. E., and Haier, R. J. (2007). The Parieto-Frontal Integration Theory (P-FIT) of intelligence: converging neuroimaging evidence. *Behav. Brain Sci.* 30, 135–154; discussion 154–187.

Kiviniemi, V., Starck, T., Remes, J., Long, X., Nikkinen, J., Haapea, M., Veijola, J., Moilanen, I., Isohanni, M., Zang, Y. F., and Tervonen, O. (2009). Functional segmentation of the brain cortex using high model order group PICA. *Hum. Brain Mapp.* 30, 3865–3886.

Koch, M. A., Norris, D. G., and Hund-Georgiadis, M. (2002). An investigation of functional and anatomical connectivity using magnetic resonance imaging. *Neuroimage* 16, 241–250.

Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 357, 1003–1037.

Lynall, M. E., Bassett, D. S., Kerwin, R., Mckenna, P. J., Kitzbichler, M., Muller, U., and Bullmore, E. (2010). Functional connectivity and brain networks in schizophrenia. *J. Neurosci.* 30, 9477–9487.

Mechelli, A., Friston, K. J., Frackowiak, R. S., and Price, C. J. (2005a). Structural covariance in the human cortex. *J. Neurosci.* 25, 8303–8310.

Mechelli, A., Price, C. J., Friston, K. J., and Ashburner, J. (2005b). Voxel-based morphometry of the human brain: methods and applications. *Curr. Med. Imaging Rev.* 1, 105–113.

Michael, A. M., Baum, S. A., White, T., Demirci, O., Andreasen, N. C., Segall, J. M., Jung, R. E., Pearlson, G., Clark, V. P., Gollub, R. L., Schulz, S. C., Roffman, J. L., Lim, K. O., Ho, B. C., Bockholt, H. J., and Calhoun, V. D. (2010). Does function follow form?: methods to fuse structural and functional brain images show decreased linkage in schizophrenia. *Neuroimage* 49, 2626–2637.

Miller, K. J., Weaver, K. E., and Ojemann, J. G. (2009). Direct electrophysiological measurement of human default network areas. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12174–12177.

Paakki, J. J., Rahko, J., Long, X., Moilanen, I., Tervonen, O., Nikkinen, J., Starck, T., Remes, J., Hurtig, T., Haapsamo, H., Jussila, K., Kuusikko-Gauffin, S., Mattila, M. L., Zang, Y., and Kiviniemi, V. (2010). Alterations in regional homogeneity of resting-state brain activity in autism spectrum disorders. *Brain Res.* 1321, 169–179.

Power, J. D., Fair, D. A., Schlaggar, B. L., and Petersen, S. E. (2010). The development of human functional brain networks. *Neuron* 67, 735–748.

Raichle, M. E., Macleod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676–682.

Raichle, M. E., and Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage* 37, 1083–1090; discussion 1097–1099.

Robinson, S., Basso, G., Soldati, N., Sailer, U., Jovicich, J., Bruzzone, L., Kryspin-Exner, I., Bauer, H., and Moser, E. (2009). A resting state network in the motor control circuit of the basal ganglia. *BMC Neurosci.* 10, 137. doi: 10.1186/1471-2202-10-137

Rykhlevskaia, E., Gratton, G., and Fabiani, M. (2008). Combining structural and functional neuroimaging data for studying brain connectivity: a review. *Psychophysiology* 45, 173–187.

Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L., and Greicius, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron* 62, 42–52.

Shehzad, Z., Kelly, A. M., Reiss, P. T., Gee, D. G., Gotimer, K., Uddin, L. Q., Lee, S. H., Margulies, D. S., Roy, A. K., Biswal, B. B., Petkova, E., Castellanos, F. X., and Milham, M. P. (2009). The resting brain: unconstrained yet reliable. *Cereb. Cortex* 19, 2209–2229.

Skudlarski, P., Jagannathan, K., Anderson, K., Stevens, M. C., Calhoun, V. D., Skudlarska, B. A., and Pearlson, G. (2010). Brain connectivity is not only lower but different in schizophrenia: a combined anatomical and functional approach. *Biol. Psychiatry* 68, 61–69.

Skudlarski, P., Jagannathan, K., Calhoun, V. D., Hampson, M., Skudlarska, B. A., and Pearlson, G. (2008). Measuring brain connectivity: diffusion tensor imaging validates resting state temporal correlations. *Neuroimage* 43, 554–561.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., and Beckmann, C. F. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13040–13045.

Sowell, E. R., Peterson, B. S., Kan, E., Woods, R. P., Yoshii, J., Bansal, R., Xu, D., Zhu, H., Thompson, P. M., and Toga, A. W. (2007). Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cereb. Cortex* 17, 1550–1560.

Sowell, E. R., Peterson, B. S., Thompson, P. M., Welcome, S. E., Henkenius, A. L., and Toga, A. W. (2003). Mapping cortical change across the human life span. *Nat. Neurosci.* 6, 309–315.

Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125.

Sporns, O., Tononi, G., and Kotter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042

Stam, C. J. (2010). Characterization of anatomical and functional connectivity in the brain: a complex networks perspective. *Int. J. Psychophysiol.* 77, 186–194.

Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., and Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. *Neuroimage* 52, 290–301.

Taki, Y., Thyreau, B., Kinomura, S., Sato, K., Goto, R., Kawashima, R., and Fukuda, H. (2011). Correlations among brain gray matter volumes, age, gender, and hemisphere in healthy individuals. *PLoS ONE* 6, e22734. doi: 10.1371/journal.pone.0022734

van den Heuvel, M. P., Mandl, R. C., Kahn, R. S., and Hulshoff Pol, H. E. (2009). Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Hum. Brain Mapp.* 30, 3127–3141.

Weng, S. J., Wiggins, J. L., Peltier, S. J., Carrasco, M., Risi, S., Lord, C., and Monk, C. S. (2010). Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders. *Brain Res.* 1313, 202–214.

Xu, L., Groth, K. M., Pearlson, G., Schretlen, D. J., and Calhoun, V. D. (2009). Source-based morphometry: the use of independent component analysis to identify gray matter differences with application to schizophrenia. *Hum. Brain Mapp.* 30, 711–724.

Yang, H., Long, X. Y., Yang, Y., Yan, H., Zhu, C. Z., Zhou, X. P., Zang, Y. F., and Gong, Q. Y. (2007). Amplitude of low frequency fluctuation within visual areas revealed by resting-state functional MRI. *Neuroimage* 36, 144–152.

Ystad, M., Eichele, T., Lundervold, A. J., and Lundervold, A. (2010). Subcortical functional connectivity and verbal episodic memory in healthy elderly – a resting state fMRI study. *Neuroimage* 52, 379–388.

Zielinski, B. A., Gennatas, E. D., Zhou, J., and Seeley, W. W. (2010). Network-level structural covariance in the developing brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18191–18196.

Zuo, X. N., Di Martino, A., Kelly, C., Shehzad, Z. E., Gee, D. G., Klein, D. F., Castellanos, F. X., Biswal, B. B., and Milham, M. P. (2010a). The oscillating brain: complex and reliable. *Neuroimage* 49, 1432–1445.

Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., and Milham, M. P. (2010b). Reliable intrinsic connectivity networks: test–retest evaluation using ICA and dual regression approach. *Neuroimage* 49, 2163–2177.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 24 May 2011; accepted: 12 March 2012; published online: 27 March 2012.

Citation: Segall JM, Allen EA, Jung RE, Erhardt EB, Arja SK, Kiehl K and Calhoun VD (2012) Correspondence between structure and function in the human brain at rest. Front. Neuroinform. ***6****:10. doi: 10.3389/fninf.2012.00010*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 11 November 2011
doi: 10.3389/fninf.2011.00028

Microstructural analysis of human white matter architecture using polarized light imaging: views from neuroanatomy

Hubertus Axer[1]*, Sindy Beck[1], Markus Axer[2], Friederike Schuchardt[1], Jörn Heepe[1], Anja Flücken[1], Martina Axer[3], Andreas Prescher[4] and Otto W. Witte[1]

[1] *Hans Berger Department of Neurology, Jena University Hospital, Jena, Germany*
[2] *Institute of Neuroscience and Medicine, Research Center Jülich, Jülich, Germany*
[3] *Department of Psychiatry and Psychotherapy, Friedrich-Schiller-University, Jena, Germany*
[4] *Institute of Anatomy, Technical University Aachen, Aachen, Germany*

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Basilis Zikopoulos, Boston University, USA
Marc Tittgemeyer, Max-Planck-Institute for Neurological Research, Germany

****Correspondence:***
Hubertus Axer, Hans Berger Department of Neurology, Jena University Hospital, Friedrich-Schiller-University Jena, Erlanger Allee 101, D-07747 Jena, Germany.
e-mail: hubertus.axer@med.uni-jena.de

To date, there are several methods for mapping connectivity, ranging from the macroscopic to molecular scales. However, it is difficult to integrate this multiply-scaled data into one concept. Polarized light imaging (PLI) is a method to quantify fiber orientation in gross histological brain sections based on the birefringent properties of the myelin sheaths. The method is capable of imaging fiber orientation of larger-scale architectural patterns with higher detail than diffusion MRI of the human brain. PLI analyses light transmission through a gross histological section of a human brain under rotation of a polarization filter combination. Estimates of the angle of fiber direction and the angle of fiber inclination are automatically calculated at every point of the imaged section. Multiple sections can be assembled into a 3D volume. We describe the principles of PLI and present several studies of fiber anatomy as a synopsis of PLI: six brainstems were serially sectioned, imaged with PLI, and 3D reconstructed. Pyramidal tract and lemniscus medialis were segmented in the PLI datasets. PLI data from the internal capsule was related to results from confocal laser scanning microscopy, which is a method of smaller scale fiber anatomy. PLI fiber architecture of the extreme capsule was compared to macroscopical dissection, which represents a method of larger-scale anatomy. The microstructure of the anterior human cingulum bundle was analyzed in serial sections of six human brains. PLI can generate highly resolved 3D datasets of fiber orientation of the human brain and has high comparability to diffusion MR. To get additional information regarding axon structure and density, PLI can also be combined with classical histological stains. It brings the directional aspects of diffusion MRI into the range of histology and may represent a promising tool to close the gap between larger-scale diffusion orientation and microstructural histological analysis of connectivity.

Keywords: polarized light imaging, brainstem, pyramidal tract, internal capsule, cingulum, extreme capsule

INTRODUCTION

Of late, the discussion relating to exploring the human connectome to attain a comprehensive structural description of the overall connectivity in the human brain has gained increasing attention (Sporns et al., 2005; Sporns, 2011). Mapping the anatomical fiber pathways connecting the various regions of the human brain is the basis for comprehending its complex function. In this context the advance of MRI methods for mapping the human connectome has recently been discussed (Hagmann et al., 2010). The method is non-invasive and allows the *in vivo* study of the human brain with regard to anatomical connectivity (diffusion MRI) as well as functional interrelationships [functional neuroimaging, e.g., functional magnetic resonance imaging (fMRI), but also PET, etc.]. Although, several methods for mapping anatomical connectivity extending from the macroscopic to molecular scale levels are established it is difficult to integrate these multiply-scaled data into a single concept. This difficulty arises since different imaging methods use different coordinate systems. The assembly of microscopical slices into a 3D dataset is possible but the projection of these data into a 3D reference coordinate system of the human brain is not generally done. There is a need for a reference coordinate system which is applicable to a wide range of different imaging modalities. Moreover, each method only shows a selective view on the object, such as connectivity, nerve fiber architecture at a specific location in the brain, diameter of fibers, fiber density, as well as fiber orientation, and many more. The method used depends of the hypothesis to be proven. A further difficulty arises from the fact that the large living human brain generally cannot be studied using different methods in parallel and diffusion MRI and fMRI is not possible to be done or at least is hindered in the formalin fixed cadaver brain.

To talk about scales and structure it is indispensable to consider the dimensions of the anatomical structures to be imaged. A nerve fiber is composed of the axon plus its myelin sheath. As viewed under the electron microscope, the size of myelinated fibers in the human corpus callosum range from 0.2 to more than 10 μm in diameter, whilst the diameters of unmyelinated fibers span 0.1–1 μm (Aboitiz et al., 1992). Fiber density in the corpus callosum is between 300,000 and 400,000 per mm^2. The number of fibers in the corpus callosum is in the order of 10^8, whereas the number of cortico-cortical projections in one hemisphere is at least one magnitude higher (Schüz and Preißl, 1996). In the human pyramidal tract, 87.9% of fibers are below 4 μm, 10.7% range from 4 to 10 μm, and 1.4% of fibers are larger than 10 μm (Graf von Keyserlingk and Schramm, 1984). Fiber density in the pyramid of the medulla oblongata is about 11,000 fibers per mm^2.

Single nerve fibers are mostly collected in fiber bundles. For example, in the anterior limb of the internal capsule, the fiber bundles of the frontopontine tract are arranged in sheaths with a diameter of about 100–150 μm that intermingle with fibers from the anterior thalamic peduncle (Axer et al., 1999a). At this scale the detection of a single axon may not be critical. On the contrary, a lower resolution might be better suited to visualize the structure of larger fiber bundles. Classical fiber tracts, e.g., the pyramidal tract, or fasciculi, e.g., the arcuate fasciculus, are collections of fibers which have been anatomically defined mainly inspired from macroscopical dissection (Ludwig and Klingler, 1956). Terms like "pathway" or "stream" are not anatomical descriptions but are mainly used in the context of functional considerations of connectivity.

In the CNS, compact fiber bundles such as the pyramidal tracts at the level of the medulla oblongata are the exception. Generally, different fiber bundles intermingle with each other in the white matter. A very critical region is the subcortical white matter directly adjacent to the gray matter, where single nerve fibers intermingle diffusely. Nerve fibers are generally not arranged in bundles at this location and these regions are therefore difficult to be imaged with diffusion MRI, due to its limited resolution power.

Obviously, the resolution of the imaging system strongly influences the level of architectural scale to be visualized. In **Figure 1**, examples of the scale of various anatomical structures together with the resolution of different imaging modalities can be seen. For instance, electron microscopy, confocal laser scanning microscopy (CLSM), and classical light microscopy are used for visualization of every single axon, while MR will focus on larger architectural pattern, e.g., larger fiber tracts. *Ex vivo* MRI in mice (9.4 T) achieves a resolution in the range of 100 μm or less (Zhang et al., 2011), while *in vivo* diffusion MRI of the whole human brain will maximally reach a resolution of 0.5 mm (Finsterbusch, 2009). Diffusion MR allows the inference of 3D fiber orientation data in the human brain and, therefore, enables the tracking of fiber pathways in the brain based on various methods such as streamline tractography, probabilistic fiber tracking, and several others (Hagmann et al., 2010). Overall, the relatively low magnification of *in vivo* diffusion MRI of the human brain solely allows the analysis of larger-scale architectural patterns.

In contrast, using high resolution anatomical methods, e.g., electron microscopy, it is possible to visualize single axons, dendrites, dendritic spines, and synapses at the highest resolution (nm-range), even in 3D (Briggman and Denk, 2006). The recently described micro-optical sectioning tomography (MOST) is able to collect highly resolved 3D data of a complete Golgi-stained mouse brain with a voxel size of 0.33 μm × 0.33 μm × 1 μm (Li et al., 2010). Such methods, however, only cater for visualization of volumes that are smaller in size than the human brain due to the enormous amount of data they accumulate. In future, high performance computing might provide analysis tools and procedures to study specimen such as the human brain reasonably. Difficulties arise, therefore, in the transfer from human connectivity data from high-scaled anatomical investigations. The higher the magnification of the method is, the smaller is the field of view or the volume of view. Thus, there is a gap between analyzing methods of the human brain and highly resolved anatomical methods.

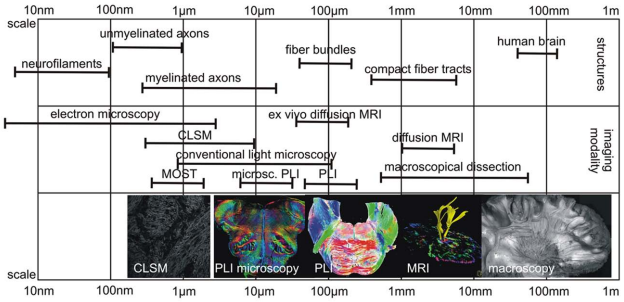


FIGURE 1 | Scales of anatomical structure and image modality. The image shows the different scales of anatomical structures of interest and the magnification of different image modalities, which can be used to study white matter architecture. Abbreviations: CLSM, confocal laser scanning microscopy; MOST, micro-optical sectioning tomography (Li et al., 2010).

Polarized light imaging (PLI), a method based on the birefringent properties of the myelin sheaths, is used to quantify fiber orientation in gross histological brain sections. We developed PLI more than a decade ago during the late 1990s (Axer et al., 2000, 2001). The idea was inspired from studies of electrical impedance measurements using impedance measurements in the human brain as an online verification of the position of the electrode during stereotaxic procedures. Impedance in the white matter is mainly influenced by the orientation of the nerve fibers in relation to the electrode (Axer et al., 1999b). Hence, an anatomical method was needed to construct a highly detailed map of the fiber orientation in the human brain in order to create a 3D atlas of fiber orientation as an anatomical correlate of impedance measurements (Axer et al., 2002a). Although the impedance principle for white matter was not followed up, the introduction of diffusion imaging to visualize fiber orientation in the living human brain has led to a growing interest in connectivity studies of the human brain (Catani et al., 2002; Jones, 2011).

This paper aims to describe PLI for imaging the human brain to investigate anatomical connectivity under the aspect of macroscopic and microscopic scales together with its relation to other methods. Herein, we present four different circumscript PLI studies, which show the relation of this new method to other imaging procedures such as diffusion tensor imaging (DTI), macroscopical

dissection, and CLSM. The following four studies have not been designed to determine the problem of scale in connectivity, rather each study demonstrates specific anatomical aspects relevant for PLI with respect to other imaging apertures and also enables the interpretation of the results in the context of functionality of fiber tracts:

1. Study of the cingulum bundle, a fiber bundle mainly involved in performing complex cognitive and neuropsychological tasks, is an example of the use of PLI and macroscopical dissection to obtain an anatomical description of an association fiber tract in the human brain.
2. Polarized light imaging analysis of the human brainstem shows the use of PLI in imaging compact fiber tracts in this study. PLI is applied for detection and segmentation as well as for morphometric analysis of these fiber tracts. Although, the analysis of the pyramidal tract is also undertaken with diffusion MRI such as for Wallerian degeneration after stroke (Grässel et al., 2010), smaller fiber tracts such as the lemniscus medialis and many others are difficult to be reliably imaged using MRI.
3. The internal capsule is a structure of white matter composed of diverse fiber bundles of varying orientation and differing dimensions. Thus, the internal capsule is a bottleneck of fibers trafficking signals from the cortex to the thalamus and the brainstem, and vice versa. Since it represents a clinically essential and complex structure, an imaging system with a larger magnification, such as PLI may be of benefit for investigating the internal capsule. Therefore, the use of CLSM is used for comparison to analyze very highly resolved nerve fiber architecture at different points of the internal capsule.
4. The extreme capsule constitutes a collection of diversely associated fiber pathways located between the claustrum and the insular cortex. Recently, interest has been sparked in the structure as fMRI and DTI studies revealed a functional involvement of these fiber pathways in language tasks (Saur et al., 2008). Thus, PLI was used to analyze fiber orientation in relation to macroscopical dissection of this brain region.

MATERIALS AND METHODS

PRINCIPLES OF PLI

Currently *polarized light* has a broad range of technological applications that cover diverse industrial and scientific fields, including optical communication, display technology, solid state physics, and biomedical optics (Brosseau, 1998). In many of these applications, polarized light is used to visualize the birefringence and, consequently, to demonstrate the spatial structure of the material under study. Generally speaking, a material is referred to as being birefringent if its molecular structure has anisotropic properties, i.e., when shape and orientation of individual molecules introduce directional dependencies. Such directional dependencies become detectable when linearly polarized light passes through a sample. For anisotropic media, the velocity of light v will vary depending on the direction of the advancing light wave and its state of polarization. This observation can be described by an elliptically shaped oblate surface, the refractive index ellipsoid or the indicatrix, characterizing the structural composition of the material by means of direction-dependent refractive indices n (= the ratio of light velocities in vacuum and in the material). The indicatrix is defined by three principal refractive indices along the principal axes of the medium. If two of these principal refractive indices are equal due to symmetry reasons, the medium is said to be uniaxial.

It has been well known for some time that the white matter of the brain has a distinct anisotropy and, thus, exhibits optical birefringence (Brodmann, 1903; Göthlin, 1913; Schmidt, 1924; Schmitt and Bear, 1935; Kretschmann, 1967; Wolman, 1970, 1975; Miklossy and Van der Loos, 1991). White matter basically consists of closely packed myelinated fibers. The myelin sheaths are multilayer membranes that are wrapped around the axons (Martenson, 1992). About 70–85% of the myelin sheaths are composed of radially ordered lipids (cholesterol, sphingolipids, and phospholipids) whilst only 15–30% consist of proteins [proteolipidprotein (PLP) and myelin basic protein (MBP)] arranged tangentially to the axon. Vidal et al. (1979) were able to attribute the birefringence of the white matter mainly to the radially arranged lipids of the myelin sheaths and observed an overall negative uniaxial birefringence. As a consequence, a myelinated fiber can be described by two principal refractive indices, parallel (n_{ff}) and orthogonal (n_{cc}) to the physical fiber axis. Given that the difference of n_{cc} and n_{ff} is small and the fiber is inclined by the angle α with respect to the front of the advancing light wave (Scheuner and Hutschenreuter, 1972; Larsen et al., 2007), the optical birefringence Δn of the fiber can be approximately expressed by:

$$\Delta n \approx (n_{cc} - n_{ff}) \cos^2 \alpha. \quad (1)$$

A beam of linearly polarized light will, on passing through a thin section of white matter, split up into two perpendicular components, the ordinary and the extraordinary ray. Due to different propagation velocities of the two rays, the beam in general becomes elliptically polarized with a phase shift δ and a difference in amplitude depending on the local fiber orientation. If the light travels along the fiber axes, the light polarization will remain unaltered. The induced phase shift is a function of the wavelength of the light source λ, the section thickness d, and the optical birefringence Δn:

$$\delta = 2\pi d \Delta n / \lambda. \quad (2)$$

An apparatus that enables the measurement of the described birefringence effects is referred to as polarimeter. The simplest setup of a polarimeter consists of two crossed linear polarizers with an object stage in between, a CCD camera, and a light source. The polarizers are rotatable (by discrete angles ρ) around the fixed sample. Since the optical system is exclusively composed of linear optical elements and the light beam that enters the sample can be considered to be fully polarized after passing the first polarizer, the application of Jones calculus (Jones, 1941) is permitted to describe the light transmittance I through the polarimeter:

$$I = I_0 \sin^2(2\rho - 2\varphi) \sin^2(\delta/2). \quad (3)$$

I_0 describes the intensity of the incident light, ρ is the azimuth of the transmission axis of the first linear polarizer, φ is the projection of the fiber into the polarizer's plane relative to the null position of the polarimeter ($\rho = 0^\circ$), and δ is the phase shift as specified in Eqs 1 and 2.

A more advanced polarimeter system houses an additional rotating quarter-wave retarder above the object stage, which leads to the following formula for the light transmittance I:

$$I = I_0/2(1 + \sin(2\rho - 2\varphi)\sin(\delta)). \quad (4)$$

In conclusion, these types of polarimeter systems sense both the fiber inclination α and the fiber direction φ. Therefore, by measuring the transmitted light intensity per image pixel at discrete rotation angles ρ and subsequent data analysis, the system provides direct access to the locally prevailing 3D fiber orientations. This imaging procedure is known as PLI.

PLI SYSTEM AND PREPARATION OF THE SAMPLES

The polarimeter system used herein (cf. Larsen et al., 2007) consists of a filter combination of a pair of horizontally mounted crossed polars and a removable quarter-wave plate (B&W Filter, Schneider, Bad Kreuznach, Germany). Diffuse light of five fluorescent tubes behind a diffusing screen (Osram GmbH, München, Germany, Universal White) is passed through the system from below. The transmitted light is imaged by a downward pointing CCD camera (Axiocam HR, Carl Zeiss, Göttingen, Germany, basic resolution of 1388 × 1040 pixel). This system allows the filters to be rotated while maintaining their relative orientation constant, and keeping the sample fixed on a stage. The images were acquired using the AxioVision software (Carl Zeiss, Göttingen, Germany). The magnification of the system was such as to result in pixels of dimension 64 µm × 64 µm over a thickness of 100 µm.

For each brain section, nine images separated by 10° rotations of the filters using the crossed polars only and nine images separated by 20° rotations of the filters using an introduced quarter-wave plate were acquired (**Figure 2**). Sinusoids were fitted to the two sequences of nine intensity values at each pixel to recover direction φ (in-plane orientation) and inclination α (out-of-plane orientation) information by means of Eqs 3 and 4. The relationship between inclination and maximum intensity of transmitted light is not linear (cf. Eqs 1–4), and it is clear that the accuracy with which the inclination can be estimated will be best for inclinations between 30° and 60°.

Sections of human cadaver brains fixed in 4% aqueous formalin solution for at least 3 weeks were used for PLI. The brains were taken from persons (without history of neurologic or psychiatric disease) who donated their body for anatomical study before death. All brains were collected from the body donor program of the Institute of Anatomy at the Technical University Aachen (RWTH). After cryoprotection, brain slabs of the region of interest were serially sectioned with a cryostat microtome (CM3050 S, Leica Microsystems, Bensheim, Germany) at a thickness of 100 µm. The sections were mounted with the aqueous mounting medium Aquatex™(Merck, Darmstadt, Germany) and coverslipped without staining.

Image processing was realized with scripts written in MATLAB 7.7.0 (MathWorks Inc., Natick, MA, USA) including the Image Processing Toolbox. The method used for 3D reconstruction has been described elsewhere in detail (Axer et al., 2002b) In short, rigid (isomorphic) transformations were computed on the serial sections of the brain. Maximum intensity maps (which represent the brightest intensity of transmitted light in each pixel derived from each polarization sequence using the crossed polars only) were used for this purpose because of their good anatomical contrast. Each image was translated and rotated with respect to its predecessor. In case of rotations, the angles of direction have to be adjusted according to the degree of rotation. The minimized Euclidean distance was used as criterion for the best fit between the succeeding slices (Axer et al., 2002b).

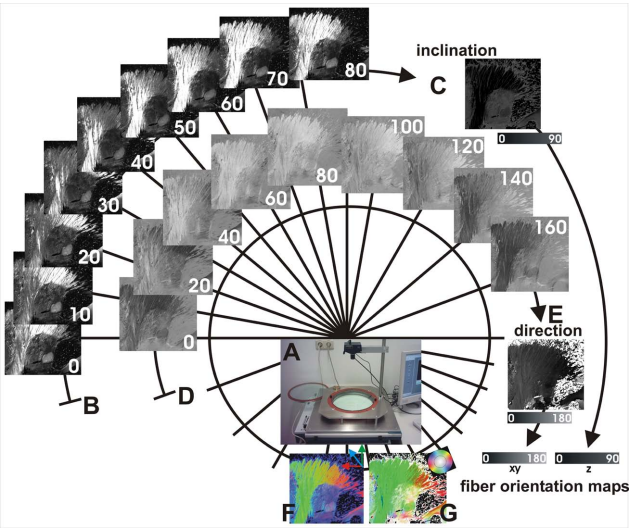


FIGURE 2 | Image acquisition and image processing of PLI. Equipment of the PLI system **(A)** consists of two rotatable perpendicularly oriented polarization filters and an insertable quarter-wave plate. For each section, nine images separated by 10° rotations of the filters using two crossed polars are acquired **(B)**, which are used to calculate angles of fiber inclination **(C)**. In addition, nine images separated by 20° rotations of the filters using an additionally introduced quarter-wave plate **(D)** are used to calculate angles of direction **(E)**. Angles of inclination and angles of direction in every pixel of the section define a vector representing the major 3D fiber orientation at that point. Fiber orientation maps (FOMs) can be visualized using different color schemes **(F,G)**.

Three dimensional Slicer 2.6 (www.slicer.org) was used for visualization of the 3D datasets and for segmentation of fiber tracts. It is an open-source, cross-platform application for visualizing and analyzing medical image data (Gering et al., 2001). SPSS 11.5.1 (SPSS Inc., Chicago, IL, USA) was used for statistical analysis.

RESULTS

CINGULUM BUNDLE

Six formalin fixed human cadaver brains were macroscopically dissected (four males, two females, median 79 years, range 72–88 years). A 1.5-cm thick slabs of the mediofrontal brain including the anterior cingulum bundle were cut in four separate pieces, serially sliced and analyzed using PLI. The target of this analysis was a parcellation of the anterior cingulum bundle according to fiber orientation (**Figure 3**). The anterior part of the cingulum is defined as the fiber bundles located above the corpus callosum and running into the frontal lobe.

The anterior cingulum bundle can topographically be subdivided into a supracallosal, pregenual, and subgenual part for

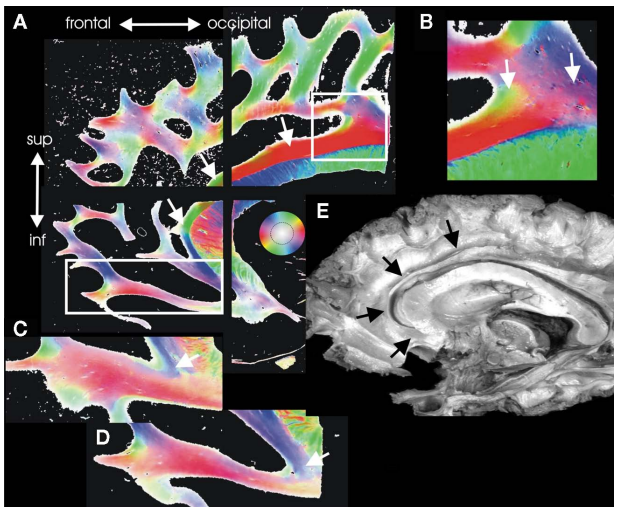


FIGURE 3 | Architectural parcellation of the anterior cingulum bundle. The cingulum bundle is marked by the white arrows in the Fiber Orientation Maps [FOMs **(A)**]. **(E)** Shows the cingulum bundle in a dissected human brain for comparison (black arrows). The anterior supracallosal and the pregenual part of the cingulum bundle is a single compact fiber bundle [white arrows in **(A)**]. In its midcingulate posterior part, the supracallosal cingulum bundle gets several inputs from fiber bundles coming from the adjacent white matter. This part is magnified in **(B)** and shows intermingling fiber bundles running into the cingulum bundle visualized in green and blue color (arrows). In the subgenual part, the fibers of the cingulum bundle curve and spread into the orbitofrontal regions [**(C,D)** arrows].

purposes of systematic anatomical description. The anterior part of the supracallosal cingulum bundle is a single compact fiber bundle without intermingling fiber bundles and without significant inputs. In its posterior part (midcingulate), the supracallosal cingulum bundle receives several inputs from fiber bundles originating from the adjacent white matter. Therefore, the midcingular segment gets inputs from fiber bundles oriented to motor and premotor cortices and parietal cortex.

In the pregenual part, the cingulum bundle is rather compact and curves around the genu of the corpus callosum. The fibers of the cingulum bundle in the subgenual part curve and spread into the orbitofrontal white matter. These fibers are diffusely intermingled with fibers deriving from the medial forebrain bundle and the forceps minor of the corpus callosum. A part of the cingulum further proceeds in the direction of the limen insulae.

HUMAN BRAINSTEM

Six human brainstems (three males and three female, median 81.5 years, range 65–92 years) were serially sectioned axially to the brainstem axis and every second slice was imaged using PLI. **Figure 4** shows sections through the medulla oblongata, pons, and mesencephalon. The major fiber tracts in the brainstem could be identified by means of PLI data, i.e., pyramidal tracts, lemniscus medialis, medial longitudinal fascicle, cerebellar peduncles, and others. Although PLI data of the larger fiber tracts show similarities to DTI results (**Figures 4A,B**), the advantage of PLI lies in the greater resolution which additionally allows for localization of smaller fiber bundles.

The serial fiber orientation maps were registered three-dimensionally and the volumes were imported into the Slicer software. Pyramidal tract and lemniscus medialis were manually segmented according to the fiber orientation maps (**Figure 5**). The volumes of these fiber tracts were estimated and related to the volume of the whole white matter of the brainstem in order to normalize according to individual brain size. **Figure 6A** shows the normalized measurements of the six brains.

The volumes of the pyramidal tracts showed a negative correlation to age (**Figure 6B**, left pyramidal tracts: Pearson correlation coefficient -0.930, $p = 0.007$, right pyramidal tracts: Pearson correlation coefficient -0.882, $p = 0.02$). In contrast, no such relationship could be found for the volumes of the medial lemniscus (**Figure 6C**) or the volume of the whole brainstem white matter.

INTERNAL CAPSULE

The internal capsule is a collection of intermingling fiber bundles of different fiber systems, i.e., the pyramidal tract, thalamic radiations, corticopontine fiber systems, and others. Using CLSM, it was demonstrated that the individual orientation of distinct fiber bundles could be used to describe a parcellation of the internal capsule (Axer and Keyserlingk, 2000). Right–left comparison of morphometric parameters of these bundles demonstrated that more and smaller bundles of the frontopontine fiber tract were located on the left-hand side than on the right-hand side of the anterior limb of the internal capsule (Axer et al., 1999a).

Confocal laser scanning microscopy allows information to be collected from well-defined optical sections through a fluorescent sample. This is done through sequential illumination focusing on one volume element of the specimen at a time (Wright et al., 1993). Stacks of serial optical sections allow a 3D reconstruction of the fiber architecture and provide good quality information regarding the orientation of the fibers at a high resolution. The myelin sheaths of the nerve fibers can be labeled with the fluorescent dye DiI and has been used to analyze fiber architecture in the internal capsule systematically (Axer and Keyserlingk, 2000).

At the level of the internal capsule, the different tracts do not consist of individual, separate tracts of fibers. In comparison to these earlier CLSM fiber studies of the internal capsule, we generated PLI sections through the internal capsule in the horizontal and the sagittal plane (**Figure 7**). The heterogeneity in fiber orientation is clearly visible in the fiber orientation maps. However, single fibers running perpendicular to the majority of fibers in the anterior limb could not be detected with PLI (**Figure 7E**). Nevertheless, there is a good correlation between confocal images of fiber architecture and the larger-scale architectural pattern of PLI.

EXTREME CAPSULE

Recently, Saur et al. (2008) identified anatomical pathways connecting brain regions activated during two prototypical language tasks by combining fMRI with a DTI-based tractography. Tractography showed that language networks interact by distinct ventral (via the extreme capsule) and dorsal (via the superior longitudinal fasciculus) association tracts as well as commissural fibers (Saur et al., 2010). In order to validate the ventral stream anatomically, a human brain was macroscopically dissected using the method of Klingler (1935), i.e., the brain is frozen down to -5 to -10°C and

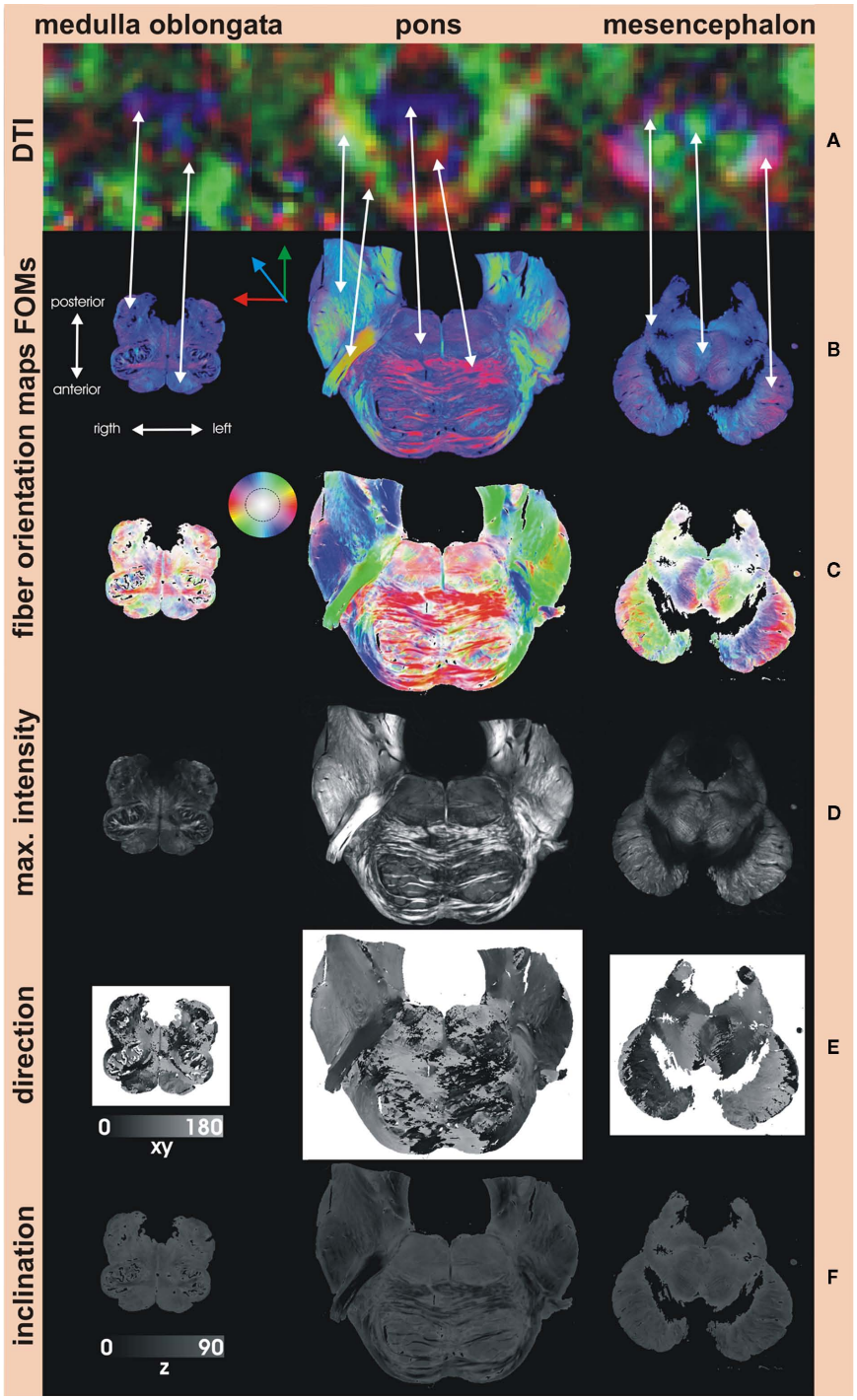


FIGURE 4 | Fiber Orientation Maps of the human brainstem. **(A)** DTI color maps (1.5T MRI with a voxel resolution of 2 mm × 2 mm × 2 mm) of comparable sections through the brainstem are able to show the larger fiber tracts which at least are at the scale of one voxel. However, the resolution is too low to distinguish smaller structures as seen in the medulla oblongata. Corresponding fibers are marked by the arrows. The advantage of PLI is its greater resolution, which also allows for visualization of smaller fiber bundles. Fiber orientation maps (FOMs) can be visualized using a color scheme similar to the DTI slices with the absolute X component of the vector shown in red, the absolute of the Y component in green, and the absolute of the Z component in blue **(B)**. Another color scheme **(C)** is more beneficial for PLI data and codes in-plane orientation in color and out-of-plane rotation in intensity (inset key). Maximum intensity maps **(D)** show the highest intensities of each polarization sequence and give relative good anatomical contrast similar to histological myelin stains. Angles of direction (in-plane) and angles of inclination (out-of-plane) generate direction **(E)** and inclination maps **(F)**. Note that the angle of inclination can be estimated between 0° and 90° only, not distinguishable from angles between 0° and (−90° (inclination ambiguity).

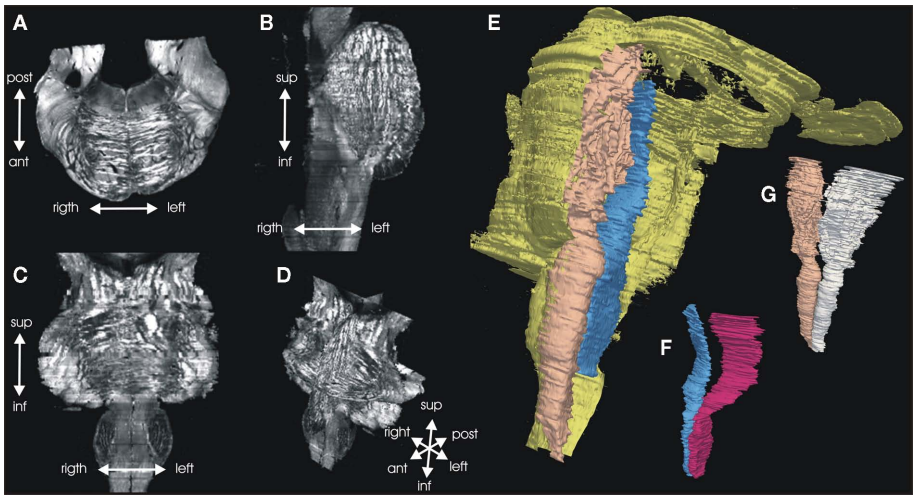


FIGURE 5 | Three dimension reconstruction of the brainstem. Serial FOMs are three-dimensionally reconstructed and a 3D data set is produced with a vector in each voxel representing the 3D fiber orientation. The example shows one original axial **(A)** section, and a calculated sagittal **(B)**, and horizontal **(C)** section through the 3D data set **(D)** shown as maximum intensity data. Pyramidal tract and lemniscus medialis were manually segmented, based on the orientation data. Volume models of the whole brainstem [**(E)** yellow], pyramidal tracts **(G)**, and lemnisci mediales **(F)** were computed.

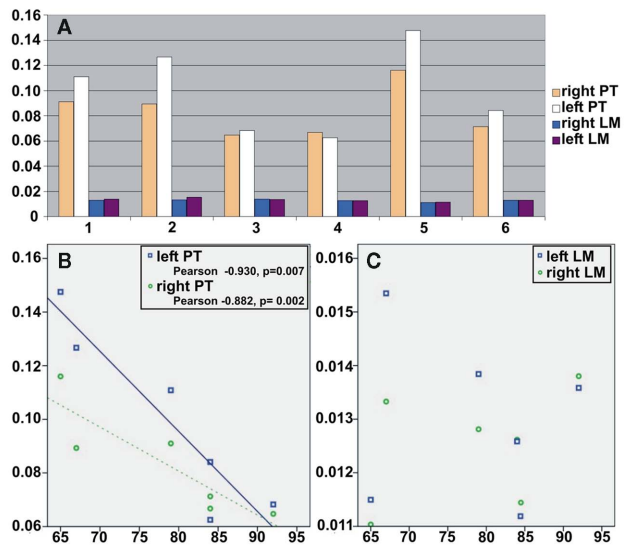


FIGURE 6 | Volumes of pyramidal tract and lemniscus medialis. The volumes of the pyramidal tracts and the lemnisci mediales were normalized to the whole white matter volume of the individual brainstem. The measurements of the analyzed six brainstems are shown in **(A)**. The normalized volumes (*y*-axis) of the pyramidal tracts revealed a negative correlation **(B)** to age (*x*-axis), while the normalized volumes of the medial lemniscus were not correlated to age **(C)**. Abbreviations: PT, pyramidal tract; LM, lemniscus medialis.

allowed to return to room temperature several times. The growing ice crystals lead to a slight separation of the nerve fibers from each other, allowing the fibers to be carefully dissected using fine forceps or blunt spatulas.

Macroscopically, distinct fiber systems could be differentiated which all channel in the depths of the limen insulae and proceed from the temporal to the frontal lobe: fibers from the extreme capsule to the frontal operculum, from the extreme capsule to the orbitofrontal cortex, and from the amygdala to the orbitofrontal cortex (the uncinate fasciculus; **Figure 8**). The uncinate fasciculus which is localized more deeply can be clearly distinguished from the other two fiber systems.

An additional brain (female, 85 years) was analyzed using PLI. For the investigation, a sagittal slab containing the limen insulae and the underlying white matter was dissected and serially sliced. Fiber orientation maps of these regions clearly reproduced the aforementioned fiber systems (**Figure 8**) as different fiber bundles. Therefore, we were able to anatomically reproduce the ventral stream system in the human brain using fiber dissection and PLI.

DISCUSSION

Studying the fiber anatomy of the human brainstem will not provide new insights into new fiber tracts or connectivity, but an analysis of these fiber tracts may be helpful for describing side-dependent asymmetries, and gender- or age-related differences. However, the low magnification of DTI (see also **Figures 4A,B**) in this region only allows evaluation of large fiber tracts, such as the pyramidal tract (Grässel et al., 2010).

Significantly, a negative relationship between age and the volumes of the pyramidal tracts (normalized to the whole white matter volume of the brainstem) was found in the current study. The result points to an age-related atrophy of the pyramidal tract system. In contrast, we found no age-related changes in the volume of the lemniscus medialis. Although, because of the small number of subjects analyzed in this study it is difficult to form reliable conclusions. Since it is obvious from **Figures 6A,B** that the three male subjects had larger pyramidal tract volumes and were younger than the female subjects, one might argue that these differences could be due to sex, age, both, or none of the above. Once again, the small sample size of three male and three female subjects, within a fairly

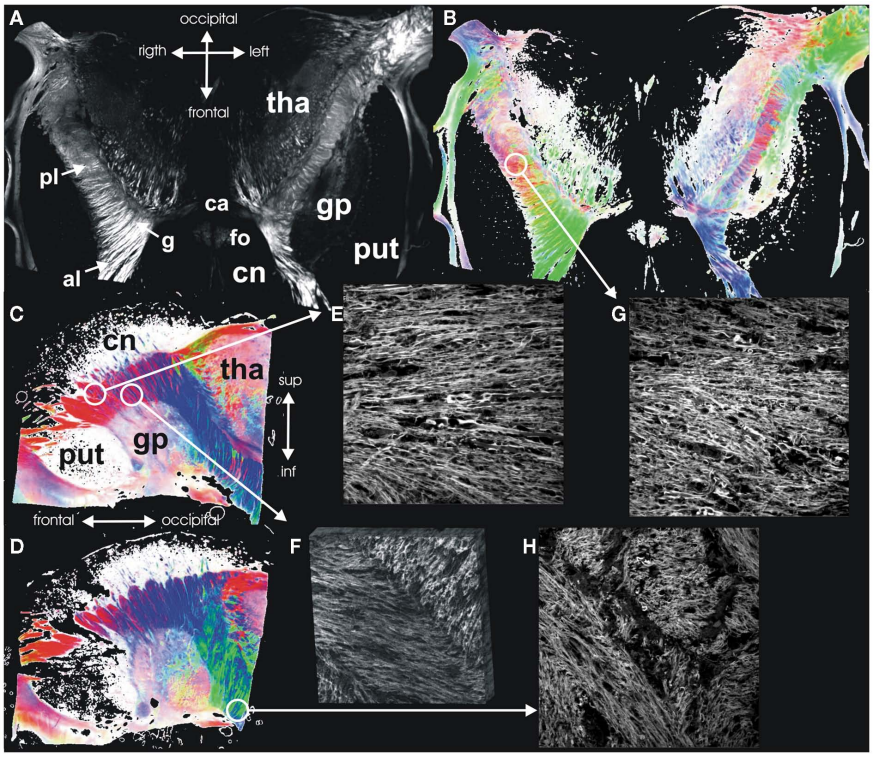


FIGURE 7 | Internal capsule. Maximum intensity map **(A)** and FOM **(B)** of a horizontal section through the internal capsule. **(C,D)** Show FOMs of sagittal sections through the internal capsule. Confocal images of different regions of the internal capsule demonstrate the fiber architecture at these points in high detail: **(E)** lateral anterior limb with parallel fibers and single nerve fiber crossing the internal capsule, **(F)** medial anterior limb with intermingling fiber bundles from the anterior thalamic radiation and the frontopontine fiber bundles (3D reconstruction), **(G)** posterior limb with intermingling fiber bundles from the pyramidal tract and superior thalamic radiation, **(H)** sublentiform part with larger intermingling compact bundles of fibers. Abbreviations: pl, posterior limb; g, genu; and al, anterior limb of the internal capsule; ca, anterior commissure; fo, fornix; gp, globus pallidus; pt, putamen; cn, caudate nucleus; tha, thalamus.

limited age-range is not enough for a proper comparison, such as for laterality. Therefore, one can only form inspired hypotheses on this subject.

Currently, studies of age-related changes of distinct fiber tracts are sparse, although the numbers will certainly increase with a rise in the systematic use of MRI in this field (Wozniak and Lim, 2006). Typical DTI changes of white matter architecture over age are a decline of fractional anisotropy (FA) and an increase of diffusivity in normal healthy adults with an anterior–posterior gradient (Sullivan and Pfefferbaum, 2006; Yoon et al., 2008). In a voxel-based morphometric study of 465 healthy subjects, Good et al. (2001) found that the global white matter did not decline with age, but local areas of relative accelerated loss and preservation were seen primarily in the posterior limb of the internal capsule. Kawaguchi et al. (2010) recently found a decrease of FA in DTI datasets in the posterior limb of the internal capsule, which was interpreted as an atrophy of pyramidal tract fibers over age.

Most studies show a decline in FA and increase in diffusivity of white matter with age, which is typically interpreted as atrophy. However, to date these alterations can not be interpreted, since they could be due to changes in a number of factors including the number or density of myelinated axons, the spacing between axons, the branching patterns or crossing of axons, or the thickness of the axons and their myelin sheaths, to name only a few. PLI can provide information regarding fiber orientation only, although, data with respect to the number, density, and thickness is not available, while the grade of myelination of nerve fibers may be estimated from PLI data.

The internal capsule is an anatomical location where different projection fiber systems are closely packed in a small volume, in particular, the pyramidal tract, but also the thalamo-cortical, and the corticopontine fiber systems (Axer and Keyserlingk, 2000). Hence, this structure is the target of many studies using DTI (Kawaguchi et al., 2010). In this context, it is of significance to keep the complexity of this structure in mind in order to draw reliable conclusions from the results of DTI studies.

Connectivity and integrity of association pathways such as the cingulum and the extreme capsule have, for instance, a high impact on cognitive functioning (Rudrauf et al., 2008). The classical concept of language circuitry has been the connection between the temporal sensory region (Wernicke's area) and the frontal motor region (Broca's area) via the arcuate fasciculus (e.g., see Geschwind, 1970). However, modern concepts describe the classical dorsal pathway along the arcuate fasciculus/superior longitudinal fasciculus that is activated during repetition and a second ventral pathway via the extreme capsule which is activated

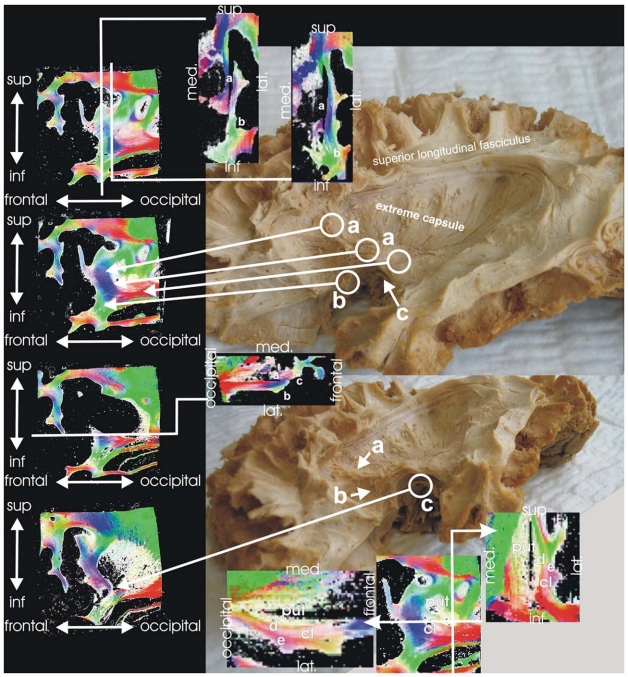


FIGURE 8 | Extreme capsule. Macroscopical fiber dissection and PLI of the white matter in the deep of the limen insulae show comparable results. Different nerve fibers can be distinguished running from the temporal to the frontal lobe: fibers from the extreme capsule to the frontal operculum (a), fibers from the extreme capsule to the orbitofrontal cortex (b), and fibers from the amygdala to the orbitofrontal cortex (c, the uncinate fasciculus). The uncinate fasciculus is localized more deeply than the other two fiber systems and can be clearly distinguished from these. The figures at the bottom show fiber orientation maps through putamen (put) and claustrum (cl), where external (d) and extreme capsule (e) can clearly be distinguished from each other.

during auditory comprehension (Saur et al., 2008). The extreme capsule seems to be part of a long association fiber pathway coursing between the inferior frontal region, the superior temporal gyrus, and the inferior parietal lobe (Makris and Pandya, 2009), which could be verified using DTI tractography of the human brain. In the monkey, these ventral and dorsal streams could also be demonstrated by means of tract tracing using radioactively labeled amino acids (Petrides and Pandya, 2009) as well as diffusion spectrum imaging (DSI; Schmahmann et al., 2007).

Anatomically, the white matter bridge between temporal and frontal lobes deep under the limen insulae is called the temporal stem (Kier et al., 2004). In this anatomical fiber channel, the extreme capsule fiber system can be clearly distinguished from the uncinate fasciculus, which is located deeper in the white matter. The uncinate fasciculus is a ventral limbic pathway (Schmahmann et al., 2007) connecting the rostral temporal region (including the amygdala) and the medial prefrontal cortex (especially the cingulate gyrus). The PLI results of the case study presented here could verify the anatomy of these fiber systems with respect to macroscopic dissection. Moreover, this topic is an example of changing concepts in the understanding of connectivity (of language circuits for instance), which was initially inspired by functional considerations.

The human cingulum bundle is another example of an association fiber pathway of the limbic system. Anatomically, the cingulum bundle is a collection of short and long association fibers which surround the corpus callosum (Nieuwenhuys et al., 1988). It is located above the corpus callosum and beneath the cingulate gyrus. The fiber bundle curves around the rostrum and genu of the corpus callosum spreading into the frontal lobe. Caudally, the bundle curves around the splenium of the corpus callosum and runs within the white matter of the parahippocampal gyrus.

In the monkey (Schmahmann et al., 2007), the fibers arise from the orbital surface (area 11 and 14) and areas 24 and 32 in the medial frontal cortex. Other fibers come from from SMA and areas 8, 9, and 46 in the dorsolateral prefrontal cortex. In addition, caudal connections also exist with the inferior parietal lobule and parahippocampal regions.

The cingulate gyrus may be subdivided into the anterior, mid, posterior, and retrosplenial cortices (ACC, MCC, PCC, and RSC; Vogt, 2005; Palomero-Gallagher et al., 2009). The midcingulate cortex (MCC) receives more input from the inferior parietal cortex and less from the amygdala compared to the perigenual cingulate cortex (Vogt and Pandya, 1987; Vogt et al., 2003). MCC contains the cingulate motor area (CMA), which has similar connections like the premotor areas with reciprocal connections with primary motor cortex and supplementary motor area (Devinsky et al., 1995). The anterior cingulate gyrus is believed to be primarily involved in executive functions especially of emotion, visceromotor and skeletomotor control, vocalization, and pain.

The anterior executive anterior cingular cortex (Devinsky et al., 1995; Allman et al., 2001) may further be subdivided into a ventral affect-dominated division, mainly connected to the amygdala and the periaqueductal gray as well as a dorsal cognition-dominated division characterized by its contribution to nociception and skeletomotor control. This region has projections to all parts of the striatum.

The PLI data regarding fiber architecture of the anterior cingulum bundle suggests a homologous parcellation. The posterior part of the supracallosal cingulum bundle receives various inputs from the adjacent white matter that apparently comes from the motor- and premotor areas and parietal cortex. In contrast, the subgenual part shows manifold connections to the orbitofrontal areas as well as with fibers directed to the amygdala. In this region, the cingulum bundle has a high variability (Bürgel et al., 2006). Our data demonstrate that these fibers are diffusely intermingled with fibers of the medial forebrain bundle and laterally with the forceps minor of the corpus callosum.

Generally, PLI is a method which supplies information relating to fiber orientation. Therefore, it provides similar information as attained from diffusion MRI, although both methods estimate fiber orientation based on different physical principles (diffusion vs. optical birefringence). The advantage of PLI is that it can provide higher resolution in comparison to diffusion MRI. For instance, a problem with diffusion-based methods is that those cannot differentiate between external and extreme capsules. PLI is able to resolve these fibers (see **Figure 8**). The resolution as used in the studies presented here was 64 μm × 64 μm × 100 μm

per voxel. However, resolution of the method can be enhanced by using a microscope for imaging (Larsen et al., 2007). Nevertheless, to maximize resolution, it is necessary to use sections that are as thin as possible, so that section thickness may the major limiting factor of resolution.

Inclination ambiguity is a second limiting aspect of PLI, as fiber inclination (out-of-plane orientation) can be estimated only between 0° and 90° (e.g., an inclination of 45° and −45° will give the same signal and cannot be distinguished from each other). However, physical improvements of the imaging system may overcome this problem in the near future. The progress in computational methods allows for elastic 3D reconstruction and finally for 3D tractography in PLI datasets (Axer et al., 2011). Although, PLI is a method which is actually still in a state of development, histological processing of the specimens is time-consuming and elaborate and as such the use of the method will be restricted to a limited number of laboratories and to a limited number of samples.

On a clinical MRI scanner, the order of a voxel size is in the mm range. Thus, one voxel can contain hundreds of thousand axons with a total length of 1 km (Alexander and Seunarine, 2011). These axons will probably have different orientations in all possible variations. Very problematic configurations involve crossing fibers or kissing fibers. Further, the classic diffusion tensor only represents one major fiber orientation. In contrast, fiber orientation distribution function (fODF) quantifies the fraction of fibers in the voxel pointing in each direction. This, therefore, represents a more differentiated estimate of the complex voxel composition. In diffusion MRI, estimations of the fODF may be achieved by the use of DSI on high-end magnets or single shell high angular resolution diffusion imaging (HARDI) without major hardware requirements or others (Hagmann et al., 2010). These methods have been specifically developed to solve the crossing fiber problem in MR. It can be used to perform fiber tracking in a much more robust and reliable manner (Tournier, 2011).

Polarized light imaging provides information regarding a single direction in each voxel. The difference to diffusion MRI, however, is the smaller voxel size, so that about 1000 PLI voxels are contained in one voxel of diffusion MRI. Therefore, the higher magnification of PLI allows for the imaging of smaller bundles of fibers and, therefore, will reduce problems of crossing and kissing fibers since inhomogeneous composition of the voxels is consequently reduced. However, PLI does not provide information about fiber number or fiber density, while the degree of myelination may be a parameter which can be estimated from the measurements. PLI as an estimate of fiber orientation is, therefore, very similar to DTI, which represents a dataset of principal diffusion directions.

MR-based methods, such as DTI, DSI, and also fMRI together with other examples, have the clear advantage that they allow for investigations of the living human brain over time in healthy or diseased conditions. However, an established gold standard for reference purposes is not yet available (Hagmann et al., 2010). To perform a multi-methodological validation, the combination of different anatomical methods has to be undertaken on the same, or at least on comparable objects. Postmortem studies are indispensable for this purpose. Several anatomical methods exist to study fiber architecture (Axer, 2011), e.g., macroscopic dissection or confocal microscopy as seen in our examples. Recently, a very promising tool to generate a highly resolved 3D data set of a whole Golgi-stained mouse brain has been published (Li et al., 2010). This system is able to generate comprehensive 3D data of a complete mouse brain, but is to date not applicable to entire human brains.

Each method only shows a selective view of the object, for example connectivity, number of fibers, myelo- and cytoarchitectural histology, or fiber orientation. The analysis of the extreme capsule system as presented here shows that the combination of different methods (e.g., DTI tractography and fMRI (Saur et al., 2010)) generates new concepts, which, in turn, may be evaluated using other neuroanatomic methods. Moreover, it represents a paradigm of the interpretation of the functional meaning of fiber tracts and connectivity, but has to be proven using different anatomical methodological approaches – since at the end, the basis of these functions is the structure of white and gray matter.

CONCLUSION

In the recent past, the discussion regarding the exploration of the human connectome as a compilation of the overall connectivity in human brain has gained increasing attention (Sporns, 2011). To date, several methods for connectivity mapping are available ranging from macroscopic to molecular scales. However, it is difficult to integrate all these multiple scaled data into one concept. It is not yet clear what the optimal scale is for efficient characterization of brain connectivity (Hagmann et al., 2010). Perhaps a courser to finer approach may be a solution achieved by using MR methods to generate "larger-scale" hypotheses which could create specified studies using "micro-scale" methods (see also DeFelipe, 2010).

However, connectivity is not only a matter of scale, but also a matter of modality and function. This is largely influenced by the scientific method which is used and the conditions under which scientific interest the study has been performed, i.e., the hypothesis that has to be proven. Challenging targets in connectivity may be variability, age-, and gender-dependent differences, asymmetry, development, degeneration, plasticity, and disease-specific changes of nerve fiber architecture.

Polarized light imaging – in this context – is a method, which can estimate fiber orientation in gross histological sections of the human brain in great detail. Although, it provides a dataset of principal fiber directions, it has the appealing advantage of a much higher resolution than MR-based methods. Therefore, PLI can generate high-resolved 3D datasets of fiber orientation of the human brain and it has, therefore, a high comparability to diffusion MR. To obtain additional information regarding axon structure and density, PLI can also be combined with classical histological stains. It brings the directional aspects of diffusion MRI into the range of histology and may represent a promising tool to close the gap between larger-scale diffusion orientation and microstructural histological analysis of connectivity.

ACKNOWLEDGMENTS

This work was partly funded by the German Research Counsil (DFG, Deutsche Forschungsgemeinschaft; grant no. Ax 20/3-1) and the German Federal Ministry of Education and Research (BMBF, grant no. 01GW0740). We thank Nasim Kroegel for her help in language editing.

REFERENCES

Aboitiz, F., Scheibel, A. B., Fisher, R. S., and Zaidel, E. (1992). Fiber composition of the human corpus callosum. *Brain Res.* 598, 143–153.

Alexander, D. C., and Seunarine, K. K. (2011). "Mathematics of crossing fibers," in *Diffusion MRI: Theory, Methods, and Applications*, ed. D. K. Jones (New York: Oxford University Press), 451–464.

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., and Hof, P. (2001). The anterior cingulate cortex. The evolution of an interface between emotion and cognition. *Ann. N. Y. Acad. Sci.* 935, 107–117.

Axer, H. (2011). "Invasive methods for tracing white matter architecture," in *Diffusion MRI: Theory, Methods, and Applications*, ed. D. K. Jones (New York: Oxford University Press), 31–42.

Axer, H., Axer, M., Krings, T., and Keyserlingk, D. G. (2001). Quantitative estimation of 3-D fiber course in gross histological sections of the human brain using polarized light. *J. Neurosci. Methods* 105, 121–131.

Axer, H., Berks, G., and Keyserlingk, D. G. (2000). Visualization of nerve fiber orientation in gross histological sections of the human brain. *Microsc. Res. Tech.* 51, 481–492.

Axer, H., Grässel, D., Steinhauer, M., Stöhr, P., John, A., Coenen, V. A., Jansen, R. H., and Keyserlingk, D. G. (2002a). Microwave dielectric measurements and tissue characteristics of the human brain: potential in localizing intracranial tissues. *Phys. Med. Biol.* 47, 1793–1803.

Axer, H., Leunert, M., Mürköster, M., Grässel, D., Larsen, L., Griffin, L. D., and Keyserlingk, D. G. (2002b). A 3D fiber model of the human brainstem. *Comput. Med. Imaging Graph* 26, 439–444.

Axer, H., and Keyserlingk, D. G. (2000). Mapping of fiber orientation in human internal capsule by means of polarized light and confocal scanning laser microscopy. *J. Neurosci. Methods* 94, 165–175.

Axer, H., Lippitz, B. E., and Keyserlingk, D. G. (1999a). Morphological asymmetry in anterior limb of human internal capsule revealed by confocal laser and polarized light microscopy. *Psychiatry Res.* 91, 141–154.

Axer, H., Stegelmeyer, J., and Keyserlingk, D. G. (1999b). Comparison of tissue impedance measurements with nerve fiber architecture in human telencephalon: value in identification of intact subcortical structures. *J. Neurosurg.* 90, 902–909.

Axer, M., Amunts, K., Grässel, D., Palm, C., Dammers, J., Axer, H., Pietrzyk, U., and Zilles, K. (2011). A novel approach to the human connectome: ultra-high resolution mapping of fiber tracts in the brain. *Neuroimage* 54, 1091–1101.

Briggman, K. L., and Denk, W. (2006). Towards neural circuit reconstruction with volume electron microscopy techniques. *Curr. Opin. Neurobiol.* 16, 562–570.

Brodmann, K. (1903). Bemerkungen zur Untersuchung des Nervensystems im polarisierten Lichte. *J. Psychol. Neurol.* 2, 211–213.

Brosseau, C. (1998). *Polarized Light – A Statistical Optics Approach*. New York: John Wiley & Sons.

Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., and Zilles, K. (2006). White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. *Neuroimage* 29, 1092–1105.

Catani, M., Howard, R. J., Pajevic, S., and Jones, D. K. (2002). Virtual in vivo interactive dissection of white matter fasciculi in the human brain. *Neuroimage* 17, 77–94.

DeFelipe, J. (2010). From the connectome to the synaptome: an epic love story. *Science* 330, 1198–1201.

Devinsky, O., Morrell, M. J., and Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain* 118, 279–306.

Finsterbusch, J. (2009). High-resolution diffusion tensor imaging with inner field-of-view EPI. *J. Magn. Reson. Imaging* 29, 987–993.

Gering, D. T., Nabavi, A., Kikinis, R., Hata, N., O'Donnell, L. J., Grimson, W. E., Jolesz, F. A., Black, P. M., and Wells, W. M. (2001). An integrated visualization system for surgical planning and guidance using image fusion and an open MR. *J. Magn. Reson. Imaging* 13, 967–975.

Geschwind, N. (1970). The organization of language and the brain. *Science* 170, 940–944.

Good, C. D., Johnsrude, I. S., Ashburner, J., Henson, R. N., Friston, K. J., and Frackowiak, R. S. (2001). A voxel-based morphometric study of ageing in 465 normal adult human brains. *Neuroimage* 14, 21–36.

Göthlin, G. (1913). *Die doppelbrechenden Eigenschaften des Nervengewebes*. Uppsala: Friedländer & Sohn.

Graf von Keyserlingk, D., and Schramm, U. (1984). Diameter of axons and thickness of myelin sheaths of the pyramidal tract fibres in the adult human medullary pyramid. *Anat. Anz.* 157, 97–111.

Grässel, D., Ringer, T. M., Fitzek, C., Fitzek, S., Kohl, M., Kaiser, W. A., Witte, O. W., and Axer, H. (2010). Wallerian degeneration of pyramidal tract after paramedian pons infarct. *Cerebrovasc. Dis.* 30, 380–388.

Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Grant, P. E., Wedeen, V., Meuli, R., Thiran, J.-P., Honey, C. J., and Sporns, O. (2010). MR connectomics: principles and challenges. *J. Neurosci. Methods* 194, 34–45.

Jones, D. K. (2011). *Diffusion MRI: Theory, Methods, and Applications*, 1st Edn. New York: Oxford University Press.

Jones, R. C. (1941). A new calculus for the treatment of optical systems. *J. Opt. Soc. Am.* 31, 488–493.

Kawaguchi, H., Obata, T., Ota, M., Akine, Y., Ito, H., Ikehira, H., Kanno, I., and Suhara, T. (2010). Regional heterogeneity and age-related change in sub-regions of internal capsule evaluated by diffusion tensor imaging. *Brain Res.* 1354, 30–39.

Kier, E. L., Staib, L. H., Davis, L. M., and Bronen, R. A. (2004). MR imaging of the temporal stem: anatomic dissection tractography of the uncinate fasciculus, inferior occipitofrontal fasciculus, and Meyer's loop of the optic radiation. *AJNR Am. J. Neuroradiol.* 25, 677–691.

Klingler, J. (1935). Erleichterung der makroskopischen Präparation des Gehirns durch den Gefrierprozeß. *Schweiz. Arch. Neurol.* 36, 247–256.

Kretschmann, H. J. (1967). On the demonstration of myelinated nerve fibers by polarized light without extinction effects. *J. Hirnforsch.* 9, 571–575.

Larsen, L., Griffin, L. D., Grässel, D., Witte, O. W., and Axer, H. (2007). Polarized light imaging of white matter architecture. *Microsc. Res. Tech.* 70, 851–863.

Li, A., Gong, H., Zhang, B., Wang, Q., Yan, C., Wu, J., Liu, Q., Zeng, S., and Luo, Q. (2010). Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain. *Science* 330, 1404–1408.

Ludwig, E., and Klingler, L. (1956). *Atlas cerebri humani*. Basel: Karger.

Makris, N., and Pandya, D. N. (2009). The extreme capsule in humans and rethinking of the language circuitry. *Brain Struct. Funct.* 213, 343–358.

Martenson, R. E. (1992). *Myelin: Biology and Chemistry*. Boca Raton: CRC Press.

Miklossy, J., and Van der Loos, H. (1991). The long-distance effects of brain lesions: visualization of myelinated pathways in the human brain using polarizing and fluorescence microscopy. *J. Neuropathol. Exp. Neurol.* 50, 1–15.

Nieuwenhuys, R., Voogd, J., and Van Huijzen, C. (1988). *The Human Central Nervous System*. Berlin: Springer.

Palomero-Gallagher, N., Vogt, B. A., Schleicher, A., Mayberg, H. S., and Zilles, K. (2009). Receptor architecture of human cingulate cortex: evaluation of the four-region neurobiological model. *Hum. Brain Mapp.* 30, 2336–2355.

Petrides, M., and Pandya, D. N. (2009). Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. *PLoS Biol.* 7, e1000170. doi: 10.1371/journal.pbio.1000170

Rudrauf, D., Mehta, S., and Grabowski, T. J. (2008). Disconnection's renaissance takes shape: formal incorporation in group-level lesion studies. *Cortex* 44, 1084–1096.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R. Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc. Natl. Acad. Sci. U.S.A.* 105, 18035–18040.

Saur, D., Schelter, B., Schnell, S., Kratochvil, D., Küpper, H., Kellmeyer, P., Kümmerer, D., Klöppel, S., Glauche, V., Lange, R., Mader, W., Feess, D., Timmer, J., and Weiller, C. (2010). Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. *Neuroimage* 49, 3187–3197.

Scheuner, G., and Hutschenreuter, J. (1972). *Polarisationsmikroskopie in der Histophysik*. Leipzig: VEB Georg Thieme.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., de Crespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.

Schmidt, W. (1924). *Die Bausteine des Tierkörpers im polarisierten Lichte*. Bonn: Cohen.

Schmitt, F. O., and Bear, R. S. (1935). The optical properties of vertebrate nerve axons are related to fiber size. *J. Cell. Comp. Physiol.* 9, 261–273.

Schüz, A., and Preißl, H. (1996). Basic connectivity of the cerebral cortex and some considerations on the corpus callosum. *Neurosci. Biobehav. Rev.* 20, 567–570.

Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042

Sullivan, E. V., and Pfefferbaum, A. (2006). Diffusion tensor imaging and aging. *Neurosci. Biobehav. Rev.* 30, 749–761.

Tournier, J.-D. (2011). "The biophysics of crossing fibers," in *Diffusion MRI: Theory, Methods, and Applications*, ed. D. K. Jones (New York: Oxford University Press), 465–481.

Vidal, B., Silveira Mello, M. L., Caseiro-Filho, A. C., and Godo, C. (1979). Anisotropic properties of the myelin sheath. *Acta Histochem.* 66, 32–39.

Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nat. Rev. Neurosci.* 6, 533–544.

Vogt, B. A., Berger, G. R., and Derbyshire, S. W. G. (2003). Structural and functional dichotomy of human midcingulate cortex. *Eur. J. Neurosci.* 18, 3134–3144.

Vogt, B. A., and Pandya, D. N. (1987). Cingulate cortex of the rhesus monkey: II. Cortical afferents. *J. Comp. Neurol.* 262, 271–289.

Wolman, M. (1970). On the use of polarized light in pathology. *Pathol. Annu.* 5, 381–416.

Wolman, M. (1975). Polarized light microscopy as a tool of diagnostic pathology. *J. Histochem. Cytochem.* 23, 21–50.

Wozniak, J. R., and Lim, K. O. (2006). Advances in white matter imaging: a review of in vivo magnetic resonance methodologies and their applicability to the study of development and aging. *Neurosci. Biobehav. Rev.* 30, 762–774.

Wright, S. J., Centonze, V. E., Stricker, S. A., DeVries, P. J., Paddock, S. W., and Schatten, G. (1993). Introduction to confocal microscopy and three-dimensional reconstruction. *Methods Cell Biol.* 38, 1–45.

Yoon, B., Shim, Y.-S., Lee, K.-S., Shon, Y.-M., and Yang, D.-W. (2008). Region-specific changes of cerebral white matter during normal aging: a diffusion-tensor analysis. *Arch. Gerontol. Geriatr.* 47, 129–138.

Zhang, J., Jones, M. V., McMahon, M. T., Mori, S., and Calabresi, P. A. (2011). In vivo and ex vivo diffusion tensor imaging of cuprizone-induced demyelination in the mouse corpus callosum. *Magn. Reson. Med.* doi: 10.1002/mrm.23032 [Epub ahead of print].

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 March 2011; accepted: 25 October 2011; published online: 11 November 2011.

Citation: Axer H, Beck S, Axer M, Schuchardt F, Heepe J, Flücken A, Axer M, Prescher A and Witte OW (2011) Microstructural analysis of human white matter architecture using polarized light imaging: views from neuroanatomy. Front. Neuroinform. ***5****:28. doi: 10.3389/fninf.2011.00028*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 30 December 2011
doi: 10.3389/fninf.2011.00034

High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging

Markus Axer[1,2]*, David Grässel[1], Melanie Kleiner[1], Jürgen Dammers[1], Timo Dickscheid[1], Julia Reckfort[1,2], Tim Hütz[1], Björn Eiben[1], Uwe Pietrzyk[1,2], Karl Zilles[1,3] and Katrin Amunts[1,4]

[1] *Institute of Neuroscience and Medicine (INM-1, INM-2, INM-4), Research Centre Jülich and Jülich Aachen Research Alliance, Jülich, Germany*
[2] *Department of Mathematics and Natural Sciences, University of Wuppertal, Wuppertal, Germany*
[3] *C. and O. Vogt Institute for Brain Research, Heinrich-Heine University Düsseldorf, Düsseldorf, Germany*
[4] *Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Aachen, Germany*

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Suzanne Haber, University of Rochester, USA
Marc Tittgemeyer, Max-Planck-Institute for Neurological Research, Germany

***Correspondence:**
Markus Axer, Institute of Neuroscience and Medicine, INM-1, Research Centre Jülich, D-52425 Jülich, Germany.
e-mail: m.axer@fz-juelich.de

Functional interactions between different brain regions require connecting fiber tracts, the structural basis of the human connectome. To assemble a comprehensive structural understanding of neural network elements from the microscopic to the macroscopic dimensions, a multimodal and multiscale approach has to be envisaged. However, the integration of results from complementary neuroimaging techniques poses a particular challenge. In this paper, we describe a steadily evolving neuroimaging technique referred to as *three-dimensional polarized light imaging (3D-PLI)*. It is based on the birefringence of the myelin sheaths surrounding axons, and enables the high-resolution analysis of myelinated axons constituting the fiber tracts. 3D-PLI provides the mapping of spatial fiber architecture in the postmortem human brain at a sub-millimeter resolution, i.e., at the mesoscale. The fundamental data structure gained by 3D-PLI is a comprehensive 3D vector field description of fibers and fiber tract orientations – the basis for subsequent tractography. To demonstrate how 3D-PLI can contribute to unravel and assemble the human connectome, a multiscale approach with the same technology was pursued. Two complementary state-of-the-art polarimeters providing different sampling grids (pixel sizes of 100 and 1.6 μm) were used. To exemplarily highlight the potential of this approach, fiber orientation maps and 3D fiber models were reconstructed in selected regions of the brain (e.g., Corpus callosum, Internal capsule, Pons). The results demonstrate that 3D-PLI is an ideal tool to serve as an interface between the microscopic and macroscopic levels of organization of the human connectome.

Keywords: connectome, human brain, method, polarized light imaging, PLI, U-fiber, systems biology, white matter

INTRODUCTION

Structure and function of the human brain are intricately linked across multiple levels of brain connectivity. It is the brain's intrinsic multiscale architecture that allows different morphological entities to be defined at different scales, essentially depending on the spatial resolution provided by the available neuroimaging techniques and the scientific objectives. A natural description of neuronal connections might therefore be organized in three levels that cover the macro-, meso-, and microscale (Sporns et al., 2005; Kötter, 2007). Various approaches for mapping the human connectome at the level of macroscopic anatomical connectivity, i.e., at millimeter scales, such as MR-based diffusion imaging (Basser et al., 1994; Pierpaoli and Basser, 1996; Conturo et al., 1999; Beaulieu, 2002; Tuch et al., 2002, 2003; Hagmann et al., 2003; Tuch, 2004; Mori and Zhang, 2006; Schmahmann et al., 2007; Naidich et al., 2008; Johansen-Berg and Rushworth, 2009) or dissection techniques (Klingler, 1935; Türe et al., 2000) were established over the past years and enriched our knowledge about fiber pathways in the human brain. To obtain and interpret experimental connectivity data at the meso- and microscale, however, is still a challenging task in terms of methodology, data handling and analysis, as well as exchange of information.

THREE-DIMENSIONAL POLARIZED LIGHT IMAGING

Axer et al. (2011) represents a novel neuroimaging technique to map nerve fibers, i.e., myelinated axons, and their pathways in human postmortem brains with are solution at the sub-millimeter scale, i.e., at the mesoscale. Polarized light imaging (Scheuner and Hutschenreiter, 1972; Brosseau, 1998; Larsen et al., 2007) utilizes an optical property of the myelin sheaths of nerve fibers known as birefringence (Göthlin, 1913; Schmidt, 1923; Schmitt and Bear, 1937). Biological samples exhibit essentially two types of birefringence referred to as intrinsic and form birefringence. The latter type reflects the textural feature of the material and the mutual birefringence of all its components (e.g., tissue and embedding), while the intrinsic birefringence is caused by individual properties of the constituting molecules (Wiener, 1912). In case of nerve fibers, birefringence is induced by the regular arrangement of lipids

and proteins in the myelin sheath (Norton and Cammer, 1984; Martenson, 1992), thus resulting in distinct optical anisotropy. The net birefringence of the neurofilaments inside the axon and the radially oriented lipid chains of the myelin sheath can be described by a single axis of optical anisotropy giving rise to uniaxial negative birefringence and, therefore, reflecting the spatial orientation of the fiber (Vidal et al., 1979).

The birefringence of brain tissue is measured by passing linearly polarized light through histological brain sections and by detecting local changes in the polarization state of light by a camera. Such principle of measurement is referred to as polarimetry and has been used in anatomical studies of the central nervous system already a century ago (Brodmann, 1903). However, in the recent past, significant advances in the 3D reconstruction of microtome sections (Dauguet et al., 2007; Singh et al., 2008; Capek et al., 2009; Palm et al., 2010), image analysis, computational techniques, and progress in understanding the interaction of polarized light with birefringent tissue (Schnabel, 1966; Fraher and MacConnaill, 1970; Oldenbourg and Mei, 1995;Oldenbourg, 1996; Oldenbourg et al., 1998; Massoumian et al., 2003; Farrell et al., 2005; Larsen et al., 2007; Axer et al., 2011) have opened up new avenues to study brain regions with complex fiber architecture at the highest level of detail. We took advantage of this progress to gain a vector field description of fiber tract orientations in histological brain sections and to reconstruct 3D fiber tract models in selected brain regions across a series of aligned sections.

This paper focuses on basic data structures gained with the latest advances in Three-dimensional polarized light imaging (3D-PLI; based on a novel tilting object stage and a polarizing microscope) and their utilization in terms of mapping the human connectome at the mesoscale. The idea is to apply basic principles of polarized light imaging in different optical setups, thus providing complementary scales of resolution to bridge the gap between structural descriptions at single fiber dimensions and at macroscopic fiber pathway levels.

MATERIALS AND METHODS

PREPARATION OF HUMAN BRAIN TISSUE

The study is based on adult postmortem human brains obtained from body donor programs at the medical departments of the Heinrich-Heine-University in Düsseldorf (Germany) and the Rheinisch-Westfälische Technische Hochschule in Aachen (Germany) in accordance with legal requirements. The clinical records were free of signs of neurological or psychiatric diseases. The brains were removed from the skull, fixed in 4% buffered formalin for at least 6 months, embedded in luxol fast blue-stained gelatine, and sectioned completely with a thickness of 100 μm thickness using a cryostat microtome (Polycut CM 3500, Leica, Germany). The maximum deviation of the section thickness from 100 μm amounts to ±5 μm. The integrity of the myelin sheaths in the brain tissue was not affected. In addition, one optic tract was prepared bilateral and sectioned parallel to the main fiber direction with thicknesses of 20, 25, 50, 70, and 100 μm.

The blockface of the frozen, gelatine-embedded brain was imaged during sectioning in order to obtain an undistorted reference image, i.e., the blockface image (Zilles et al., 2002; Amunts and Zilles, 2006; Toga et al., 2006). Hence, a CCD camera (Oscar F510, Allied Vision Technologies, Germany) with an endocentric lens (APO Rodagon N 2.8/50 mm, Linos, Germany) was mounted vertically above the blockface at a distance of 1.2 m. It provided an image dimension of 2588 × 1958 pixels. In addition, the precise distance of each blockface to the camera system was determined with an opto-electronic distance sensor (LDM 42A, Astech, Germany). The distance measurement is required to correct for perspective distortions in the images, since an endocentric lens was used and the distance between the block surface and the camera increased while the sectioning proceeded. This is due to the fact, that the microtome owns an automatized height adjustable knife. Taking the distance information into account, the set of blockface images was subjected to a single-modality 2D affine registration for the 3D (blockface) brain reconstruction (Eiben et al., 2010; Palm et al., 2010).

LARGE-AREA ROTATING POLARIMETER WITH TILTING STAGE

A fully automatized rotating polarimeter has been constructed and optimized for high-resolution PLI (Axer et al., 2011). The setup is sensitive to small local changes of the polarization state of light induced by the interaction with the anisotropic brain tissue, i.e., the myelin. Since it provides single-shot imaging of whole human brain sections (with a sensitive area of 24 cm in diameter), it is referred to as *large-area polarimeter*.

The polarimeter is equipped with a pair of crossed polarizers (XP38, ITOS, Germany), a specimen stage and a quarter-wave retarder (WP140, ITOS, Germany) positioned between the two polarizers. A customized LED light source (FZJ-SSQ300-ALK-G, iiM, Germany) illuminates the brain section with a narrow-band green wavelength spectrum (525 ± 25 nm; **Figure 1A**). By rotating all optical devices simultaneously around the stationary tissue sample, the principal axes of birefringent structures, i.e., the fiber axes, are systematically imaged by a CCD camera (AxioCam HRc, Zeiss, Germany) at discrete rotation angles ρ. The birefringence causes the measured light intensity I to vary in a sinusoidal manner with respect to the rotation angles ρ, depending on the local 3D fiber orientation (φ, α; cf. **Figure 1C**). The sinusoidal variation of light intensity as a function of the rotation angle is referred to a slight intensity profile.

$$I = \frac{I_0}{2} \cdot [1 + \sin(2\rho - 2\varphi) \cdot \sin\delta], \quad (1)$$

Where

$$\delta \approx 2\pi \cdot \frac{d \cdot \Delta n}{\lambda} \cdot \cos^2\alpha. \quad (2)$$

Each image pixel is therefore characterized by an individual light intensity profile that can be described with the Jones calculus (Jones, 1941) and the basic principles of optics (Snell's law, Huygens–Fresnel principle; cf. Saleh and Teich, 1991; **Figure 1D**). In Eq 1 the phase of the profile encodes the angle φ, which represents the individual fiber direction, i.e., the projection of the fiber axis into the section plane with respect to the null position of the polarimeter (ρ = 0°). The amplitude of the profile quantifies the phase retardation δ induced to the light wave by the myelin (**Figure 1C**). This phase retardation is a function of the

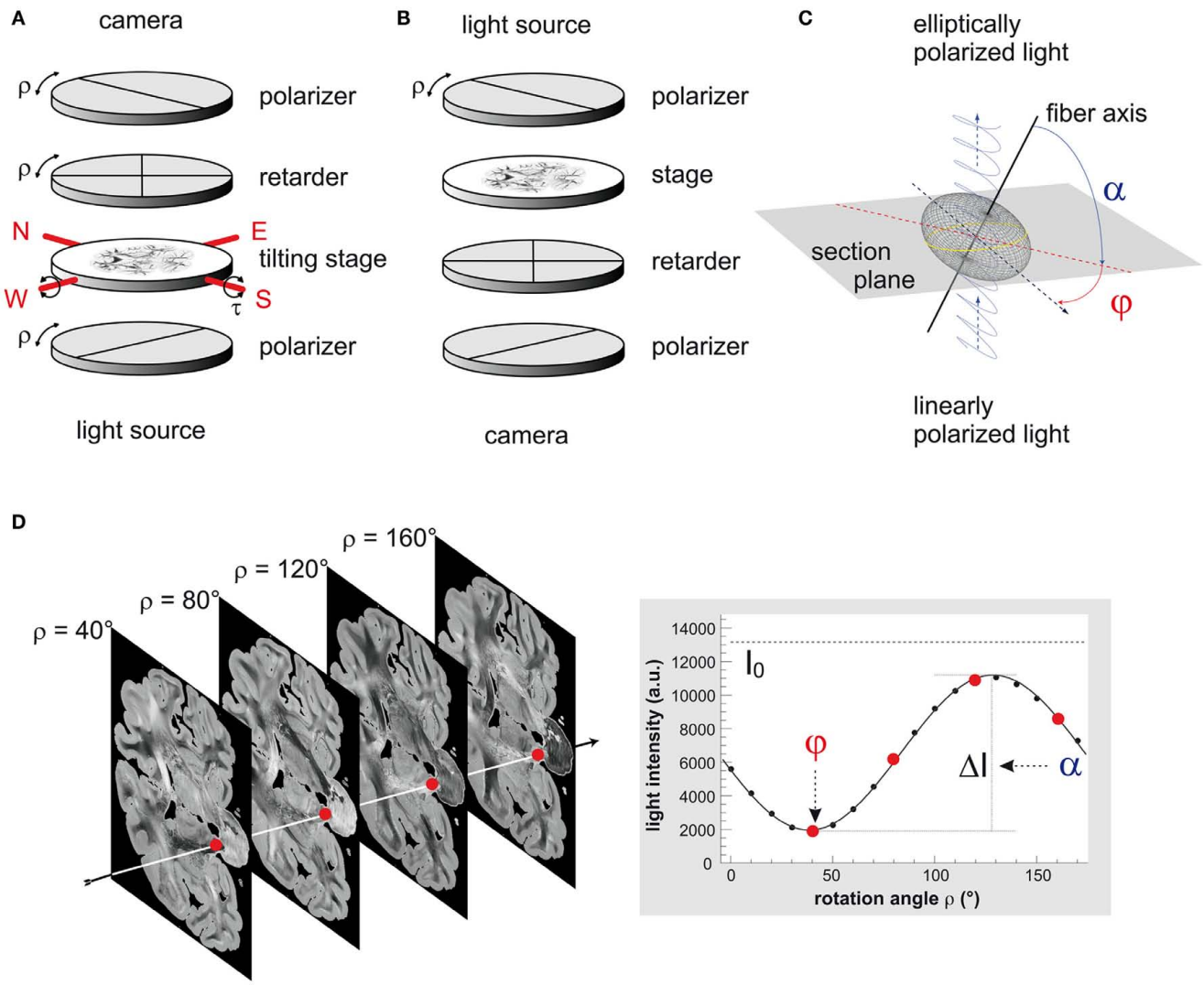


FIGURE 1 | Three-dimensional-polarimetry at a glance. (A) Scheme of the large-area rotating polarimeter with tilting stage (N-North, W-West, E-East, S-South). **(B)** Optical scheme of the polarizing microscope LMP1-1. **(C)** Scheme of the optical fiber model. The refractive index of a negative uniaxially birefringent medium, such as a myelinated axon, is described by an elliptically shaped oblate surface, the refractive index ellipsoid or indicatrix (gray mesh). A beam of linearly polarized light (blue trace) interacts locally with the myelin sheath of a single axon (black line), which induces a phase shift to the light beam. The light becomes elliptically polarized and serves as a direct measure of the orientation of the indicatrix or the prevailing local fiber orientation, respectively. In the frame coordinate system this orientation is determined by the in-plane direction angle φ and the out-of-section inclination angle α. **(D)** A typical PLI raw image data set consists of 18 images corresponding to equidistant rotation angles between 0° and 170°. Here, a selection of four images of a coronal section is shown, while the sketched arrow indicates one representative pixel. To obtain the fiber orientation, the measured light intensities are studied pixel-wise as a function of discrete rotation angles. The derived physical model provides a precise mathematical description of the measurement (continuous black line) and relates the sine phase to the direction angle φ and the amplitude to the inclination angle α. The highlighted data points correspond to the selected images.

light wavelength λ, the section thickness d, the birefringence Δn of the myelin, and the inclination angleα of the fiber's principal axis (cf. Eq 2). The birefringence depends on the wavelength of light and the temperature of the tissue. In order to keep the variations of both parameters small, the light source was water-cooled and all measurements were carried out at controlled environment temperatures (22 ± 1°C). The transmittance I_0 denotes the intensity of the incident light modified by local extinction effects.

As a fundamental technical innovation in comparison to the planar setup as specified in Axer et al. (2011) a novel specimen stage has been introduced. This stage is tiltable with respect to two perpendicular axes precisely aligned to the camera image axes. By slightly tilting the brain section around these two axes, fiber tracts can be imaged stereoscopically from four different views (within a tilt angle range of $|\tau| \leq 4°$ in east, west, north, and south direction), in addition to the standard planar view (**Figure 1A**). This innovation was driven by the known limitation of a standard planar

polarimeter where the extracted inclination angles α are naturally restricted to absolute values between 0° and 90°, and are therefore afflicted with a sign ambiguity. The tilting modifies the inclination angles of the fiber axes with respect to the planar projection plane and enables to disambiguate the inclination measurement.

Image acquisition and image processing using the large-area polarimeter

Each section was imaged at five tilts (planar with $\tau = 0°$; east, west, north, and south with $\tau = 4°$) at 18 equidistant rotation angles of the polarimeter covering an angular range between 0° and 170°. I.e., in total, 90 images were acquired for each section. The camera setup yielded RGB images with image sizes of 1388 × 1040 pixels and a pixel dimension of 100 μm × 100μm. The sampling interval can be increased up to 33 μm × 33μm using the same setup. However, for the presented studies we aimed for isotropic voxel sizes. The light intensities were sampled with a dynamic range of 14 bits per color channel.

Raw image processing, signal analysis and image registration were performed according to Axer et al. (2011). As a result, the following parameter maps were obtained:

- The *transmittance map* (I_0): it represents the pixel-wise (doubled) average map of all PLI raw images (cf. **Figure 1D**) and is a measure for the extinction of light after passing through the polarimeter and the brain tissue.
- The *retardation map* (|sin δ|): it equates to the normalized amplitudes of the light intensity profiles ($\Delta I/I_0$, cf. **Figure 1D**) and describes the magnitude of the phase shift (retardation) induced to the light wave due to the interaction with the birefringent tissue.
- The *fiber direction map* (φ): it describes the in-section direction angle, i.e., the *x*–*y* orientation of each fiber (cf. **Figure 1C**).
- The *fiber inclination map* (α): it describes the out-of-section angle, i.e., the vertical component of each fiber (cf. **Figure 1C**).

Image calibration, independent component analysis (ICA), discrete harmonic Fourier analysis, and image registration were fundamental prerequisites in this processing chain aiming for consistent data sets of local fiber orientations. To ensure comparability among different sections and brains, image calibration is applied to all raw images aiming for homogeneous and constant background intensities, i.e., measured light intensities without brain tissue in the optical path. In addition, the application of ICA to PLI raw images enables the effective separation of birefringence signals from interfering signal sources, such as electronics noise, light scatter, dust particles, or filter inhomogeneity. The reproducibility of the results gained with the combination of image calibration and ICA has been demonstrated in Dammers et al. (2010, 2011). Pixel-wise fitting of the ICA-enhanced light intensity profiles by means of discrete harmonic Fourier analysis (Glazer et al., 1996) provides fast and automatized extraction of the parameter maps listed above.

To utilize the additional information gained with the tilting stage, a precise rectification of the raw images had to be done prior to any analysis. This is due to the fact, that the tilts geometrically distort the raw images with respect to each other. To overcome this distortion, projective transformations (2D homography, cf. Fischler and Bolles, 1981; Hartley and Zisserman, 2004) between mutually tilted image pairs is recovered based on a robust matching of SIFT keypoints (scale-invariant feature transform; cf. Lowe, 2004). Since the tilting takes effect on the amplitudes of the intensity profiles, i.e., on the retardation |sin δ|, the change of the retardation with the tilting condition was evaluated pixel-wise to determine the sign of the fiber inclination. The sign describes the slope of the fiber course (uphill or downhill). By this means an *inclination sign map*(*s*) was created for each section.

Fiber tractography

To reconstruct fiber models in a volume of fiber orientation maps (FOM), the data were subjected to a standard streamline algorithm based on the Euler method (Mori, 2007) as implemented in the basic toolbox of the Matlab© framework (Mathworks Inc.). A tract is propagated from a seed point by following the local vector orientation using interpolation methods. Seed points were densely placed on the surfaces of a box confining the volume of interest (VOI). The VOIs shown here were placed primarily in areas of white matter with a heterogeneous fiber direction (e.g., the Pons). The line propagator step size was chosen to be 30% of the vector grid unit throughout the propagation process in order to minimize aberration due to coarse quantization. The resulting streamlines are defined by a sequence of vertices in 3D space.

POLARIZING MICROSCOPE

Basic principles of polarized imaging enable also to utilize microscopic devices in order to study neuroanatomical structures at the highest level of detail in 3D on small sample and tissue sizes. Therefore, a *polarizing microscope* (LMP-1, Taorad, Germany) with a pixel size of 1.6 μm × 1.6 μm in-plane has been employed (Glazer et al., 1996). In contrast to the large-area polarimeter, only one linear polarizer in the optical path is rotatable (**Figure 1B**) and a brain section has to be scanned tile-wise with overlapping fields of view using a motorized microscope stage (Märzhäuser, Germany). Pursuant to theoretical optics (Jones, 1941), Eqs 1 and 2 also apply to this type of polarimeter.

Image acquisition and image processing using the polarizing microscope

Each brain region was imaged at 18 equidistant rotation angles of the polarizer covering an angular range between 0° and 170°.A single microscope (grayscale) image comprises an area of about 3.3 mm × 3.3 mm (2048 × 2048 pixels), i.e., a whole brain section is built up of about 2500 single tiles, for example. The images were sampled with a dynamic range of 12 bits.

To produce ultra high-resolution images of large brain areas, multiple images were combined (stitched) with a Matlab© application developed in our lab. Raw image processing and signal analyses were performed as already described for the large-area polarimeter (cf. Axer et al., 2011).

RESULTS

REFERENCE DATA SET

The aligned series of blockface images represents a three-dimensional data set that serves as an undistorted high-resolution

reference. It was used to retrieve the spatial coherence within adjacent sections and to correct for tissue distortions inevitably introduced by sectioning and further steps of histological processing. In particular, the luxol fast blue-stained gelatine (cf. **Figure 2A**) turned out to be beneficial for automatized discrimination of the brain tissue from the background, i.e., the surrounding gelatine. The hue representation of the hue-saturation-value (HSV) color space provided a clear separation of the predominantly red and yellow colored fixed brain tissue from the blue color of the gelatine (**Figure 2**). The perspective error correction integrated into the rigid registration process finally led to an accurately aligned and scaled high-resolution 3D representation of the blockface brain (**Figure 2D**).

PARAMETER MAPS AND DATA STRUCTURES

The image processing chain of 3D-PLI results in a set of comprehensive parameter maps in 2D (grayscale) and 3D (colorscale), each of them highlighting distinct tissue properties and anatomical structures. One of the basic 2D parameter maps obtained by analyzing the light intensity profiles is the transmittance map (I_0; **Figure 3A**). This type of map provides a general separation of gray and white matter, but it also shows intensity gradients within the cortical and the subcortical gray matter as well as in the white matter (cf. arrow heads in **Figure 3B**). Comparison of transmittance maps of coronal sections with myelin stained sections based on the Heidenhain–Woelcke technique (Bürgel et al., 1997) revealed a striking correspondence between both modalities (**Figure 3**). However, the transmittance maps yielded more contrast across the white matter than the myelin stained sections, in particular in U-fiber regions (cf. white arrow heads in **Figures 3A,B**).

The fundamental entity of 3D-PLI to characterize the locally prevailing fiber tract orientation in a single voxel is a unit vector. The integration of all vectors for an entire brain section leads to a comprehensive 3D vector field description, i.e., a FOM. This is another, fundamental difference from the myelin stained sections, which do not enable to extract the third dimension, i.e., the inclination.

Due to the tilting stage, we were able to assign a definite preference to the out-of-section inclination angle (i.e., the sign relative to the section plane) in the FOM for the first time. Analysis of the tilted raw images approved that the tilting of 4° takes effect on the amplitudes of the intensity profiles, i.e., the retardation $|\sin\delta|$ (cf. **Figure 4A** as an example). As intended, the data revealed correlations between the tilting condition (east, west, north, and south), the retardation, and the fiber direction. The largest changes of retardation values caused by the tilting were observed in brain regions with prevailing intermediate fiber inclinations (i.e., $\alpha \approx 45°$), while fibers running within or perpendicular to the section plane showed much smaller changes. The benefit of tilting becomes obvious in **Figures 4B,C**. In general, we observed that inclination sign maps were characterized by distinct clusters dominated by the same inclination sign distributed all over the brain section (cf. **Figure 5A**). However, the tilting method also

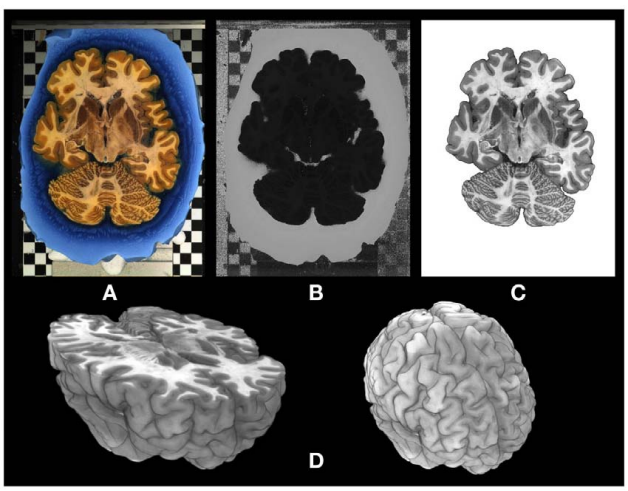


FIGURE 2 | Segmentation and 3D reconstruction of the blockface image data set. (A) Blockface image of a horizontally cut postmortem human brain represented in RGB (red-green-blue) color space. The brain is embedded in stained (luxol fast blue) gelatine. The checkerboard in the background was used for subsequent alignment of the blockface images obtained during sectioning. **(B)** Transformation of the RGB image into the HSV-color space enables an accurate segmentation of the image into tissue and background. The hue-channel of the blockface is shown here. **(C)** Segmented brain section. **(D)** 3D representations of the reconstructed blockface brain.

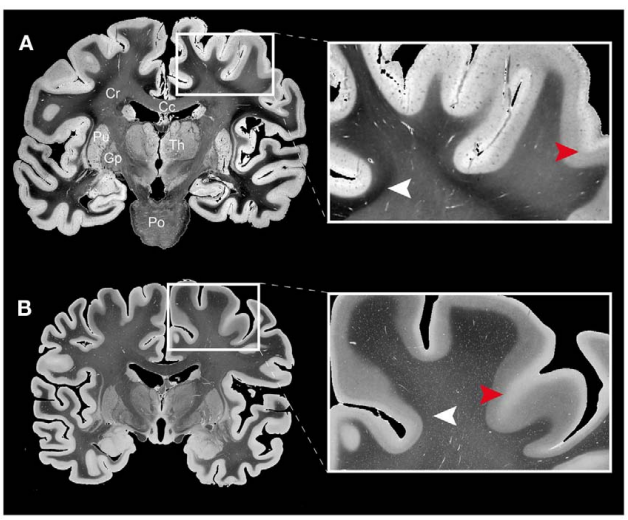


FIGURE 3 | Comparison of (A) a transmittance map of a coronal whole human brain section (100 μm thickness, gelatine embedding) and (B) a coronal section from another human brain stained with the Heidenhain–Woelcke technique for myelin (20 μm thickness, paraffin embedding). The images were scaled to the same gray value range and show global similarities in their gray level distributions. However, the enlarged regions of interest from the frontal lobe (right images) document that the transmittance map yields more contrast across the white matter regions (e.g., in U-fiber regions as indicated by the white arrow heads) than the classical histological myelin staining. For cortical regions the measured intensity gradients are similar (cf. red arrow heads). Legend: Cc, corpus callosum; Cr, corona radiata; Th, thalamus; Pu, putamen; Gp, globus pallidus; Po, pons.

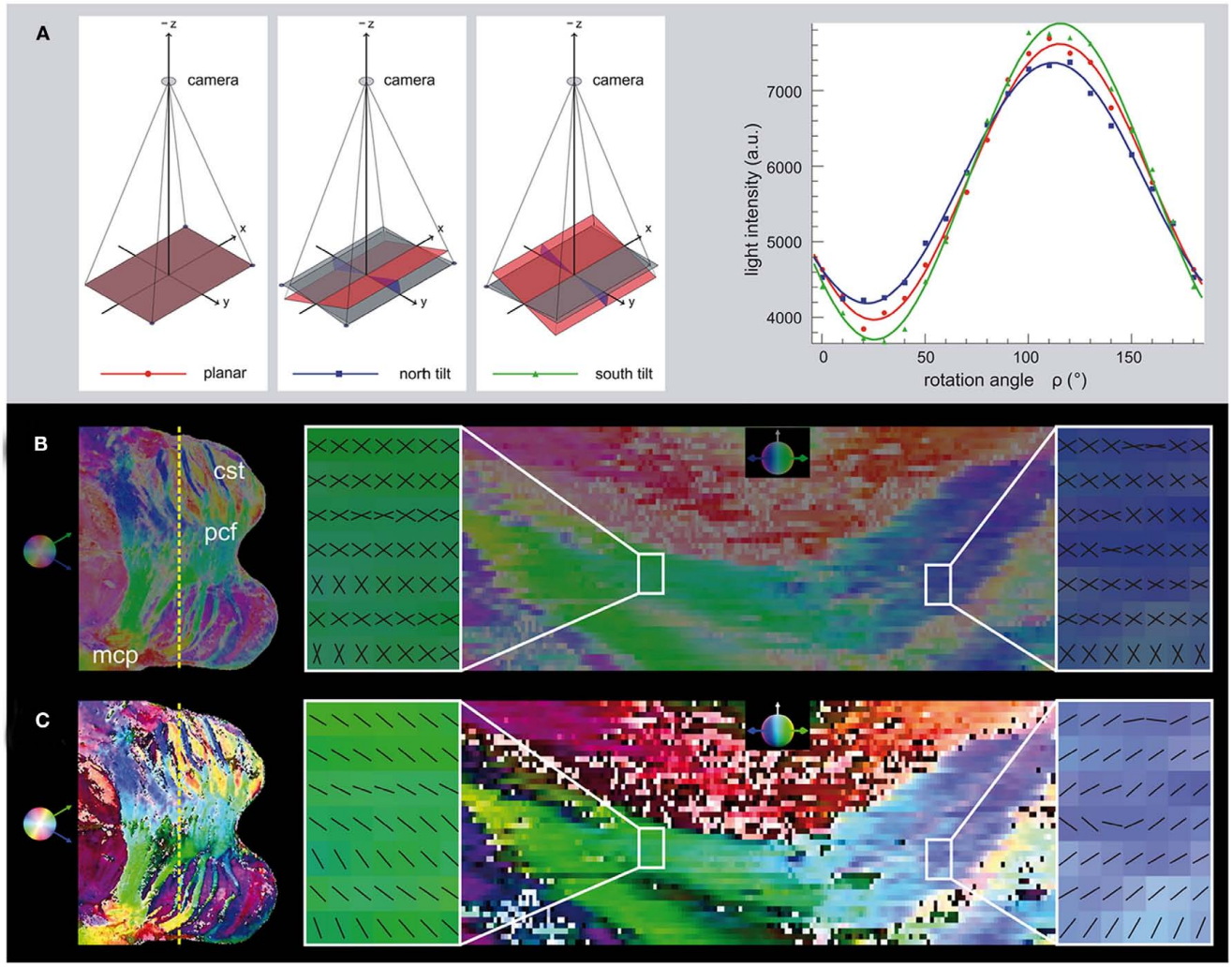


FIGURE 4 | Determination of the ambiguous inclination sign by tilting. **(A)** In addition to images acquired in the standard planar position, further information can be derived from images that are tilted in north and south direction, for example. After optical rectification, the light intensity in a single pixel is plotted for the series of polar filter rotation angles. The north, south and planar tilting positions have different amplitudes. The change in the amplitude of the signal demonstrates the change in the absolute inclination angle as a result of tilting. In this case, the south tilt yields a larger amplitude and hence a higher absolute inclination angle than the north tilt. This indicates a negative inclination sign. **(B)** The overview of the transversal section through the Pons on the left shows the cutting plane (dotted yellow line) of the coronal view on the right. The HSV-color coding shows both possible inclination signs in the same color (Hue: transversal direction, Saturation: coronal inclination, Value: constantly 0.5). The magnified regions show the two possible fiber orientations mirrored to each other, if the inclination sign is still ambiguous. **(C)** After determination of the inclination sign, a decision is made for every pixel, which is color-coded by different brightness values (>0.5: positive sign, <0.5: negative sign). The magnified regions show the resulting fiber orientations. The inclination sign was determined as negative in the left region, while a positive inclination sign was derived in the right region. The orientations agree with the course of the pontocerebellar fiber bundles running toward each other from lateral to medial. Legend: cst, cortico-spinal tract; pcf, pontocerebellar fibers; mcp, middle cerebellar peduncle.

proved to be sensitive to heterogeneous structures within regions of intense fiber intermingling such as the Corona radiata.

Two representations of a FOM turned out to be beneficial in terms of visualization and fiber tractography: (i) utilizing the HSV-color space provided an informative way to encode the determined fiber orientations based on the direction angle (φ), the inclination angle (α), and the inclination sign (s) for each image pixel. Brain regions with a rich variety of fiber orientations (e.g., the Corona radiata), but also regions with homogenous fiber courses (e.g., the Corpus callosum) become evident in an HSV-colored FOM (cf. **Figure 5A**). (ii) Transforming the angle-based description of fiber orientations into unit vectors (**Figure 5B**) resulted in a data structure that is best suited for fiber tractography using streamline algorithms. The visualization of vector data utilizes the RGB color space, where the vector components are encoded in the basic colors red, green, and blue. Though this color code is an established approach in MR diffusion imaging, it actually reduces the full spatial information of symmetrical fiber courses (compare

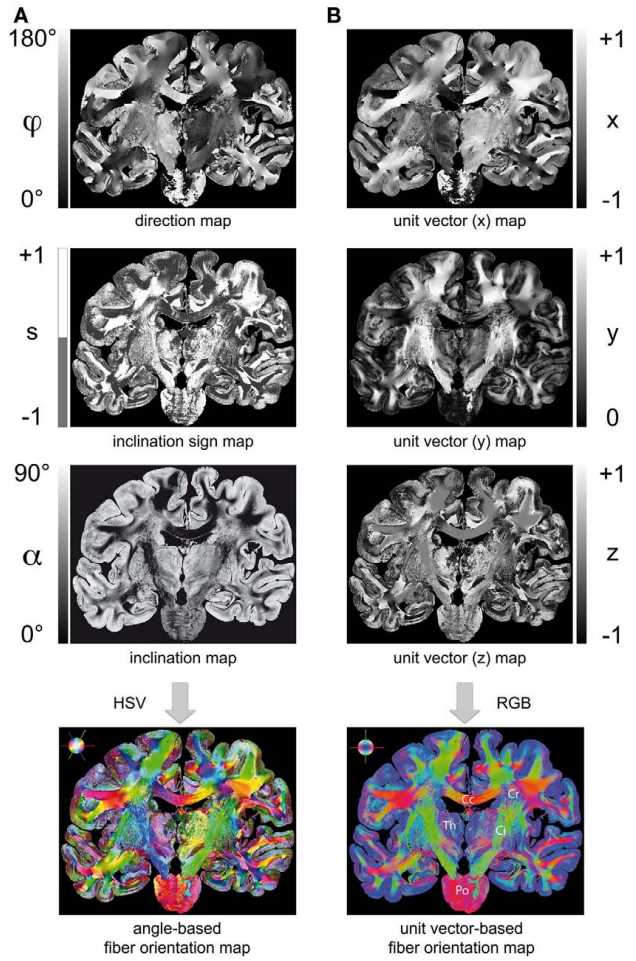


FIGURE 5 | Two ways to render a fiber orientation map (FOM): (A) Transformation of a direction map, inclination sign map, and inclination map into a FOM using the HSV-color space (encoding: $H = 2\varphi$, $S = 1 - \alpha/90^{\circ}$, $V = 1$ for $S > 0$; $H = 2\varphi$, $S = 1$, $V = 1 - \alpha/90^{\circ}$) for $S \leq 0$). (B) Transformation of the extracted fiber orientations into unit vectors ($x = \cos(s\alpha)\cdot\cos(\varphi)$, $y = \cos(s\alpha)\cdot\sin(\varphi)$, $z = \sin(s\alpha)$) and visualization in the RGB color space (encoding: $R = |x|$, $G = |y|$, $B = |z|$). Legend: Cc, corpus callosum; Cr, corona radiata; Ci, internal capsule; Th, thalamus; Po, pons.

the transversally symmetric courses of the Internal capsule in the FOMs shown in **Figure 5**). However, the RGB representation is preferred for direct comparison between 3D-PLI and results from diffusion weighted MRI (cf. Oishi et al., 2011).

3D FIBER TRACT MODELS

The 3D data set for high-resolution structural fiber mapping is composed of adjusted and stacked sectional FOMs. This 3D-FOM already provides novel insights into complex local fiber architectures, even without further computational processing such as fiber tractography.

Fiber tract models finally demonstrate the feasibility of the applied imaging technique. Therefore, fiber tractography was performed in small subsamples of a stack of 10 coronal whole brain FOMs to build continuous fiber tract models. Five regions of interest with dimensions of 2 mm × 2 mm × 1 mm distributed over the entire volume of sections were subjected to the streamline algorithm. Five selected regions are shown in **Figure 6**, where fiber models were reconstructed from 3D-FOMs in the Corpus callosum, the Capsula interna, and the Pons. Intense fiber crossings also affect 3D-PLI due to partial volume effects, however, on a much smaller scale than it is the case in diffusion weighted MRI. Some transversal fiber tracts (red) crossing the internal capsule between Putamen and Thalamus (C) are confused (blue) by the cortico-spinal mainstream (green) and do not survive the crossing.

In addition to fiber tract models in white matter regions reconstructed from 3D-FOMs obtained with the large-area polarimeter, 3D-PLI applied to the polarizing microscope images provides further, even more detailed insights into regions close to or within the cortex. To give an example, association fibers at the circular sulcus of the insula were 3D reconstructed and studied (**Figure 7**). The fiber tract models in a volume of 0.5 mm × 0.5 mm × 0.1 mm (obtained with the large-area polarimeter) follow a clear subcortical U-shape with indications of changeover into the cortex (**Figure 7B**). Two microscopic 2D-FOMs (**Figures 7C,D**) covering 2D sub-regions at the white/gray matter border of a section from the center of the stack prove the fiber crossover into the cortex. Single fibers can clearly be observed at the provided pixel size of 1.6 µm. There seems to be a gap (blue) between the white and the gray matter. This is due to the fact that the contributions of perpendicularly crossing fibers in the same voxel tend to average out thus lowering the amplitude of the measured intensity profile, i.e., partial volume effects due to the section thickness can be observed.

To demonstrate the influence of section thickness d on the retardation as well as the impacts on the polarimeter setup in terms of sensitivity, a study of an optic tract was carried out. Since the optic tract was sectioned along its main fiber direction (i.e., $\alpha = 0^{\circ}$), the determined retardation maps of the different section thicknesses reveal the largest signals to be expected in 3D-PLI. According to Axer et al. (2011), the histogram of a retardation map with resident in-plane fibers can be used to determine the correlation between the measured retardation value $|\sin\delta|$ and the inclination angle α (**Figure 8**). The mean values $\overline{|\sin\delta|}_{\alpha=0^{\circ}}$ of the histogram tails were extracted by fitting and used as reference measures for in-plane fibers of the different section thicknesses (**Figures 8A,B**). A significant increase of $\overline{|\sin\delta|}_{\alpha=0^{\circ}}$ from 0.201 at $d = 20$ µm to 0.728 at $d = 100$ µm was observed, while scattering and absorption effects become more influential at thicknesses above 100 µm thus leading to a decrease of the maximum signal (**Figure 8B**). This behavior was described by a fit function based on Eqs 1 and 2 multiplied by an exponential term. Hundred micrometer represents the maximum suitable section thickness and the highest dynamic range to separate fiber inclination angles (**Figure 8C**).

DISCUSSION

In order to create a comprehensive structural description of the neural network and its intricate fiber connections, i.e., to assemble an important part of the human connectome, a multimodal approach is certainly indispensable. Depending on data provided by complementary neuroimaging techniques, an

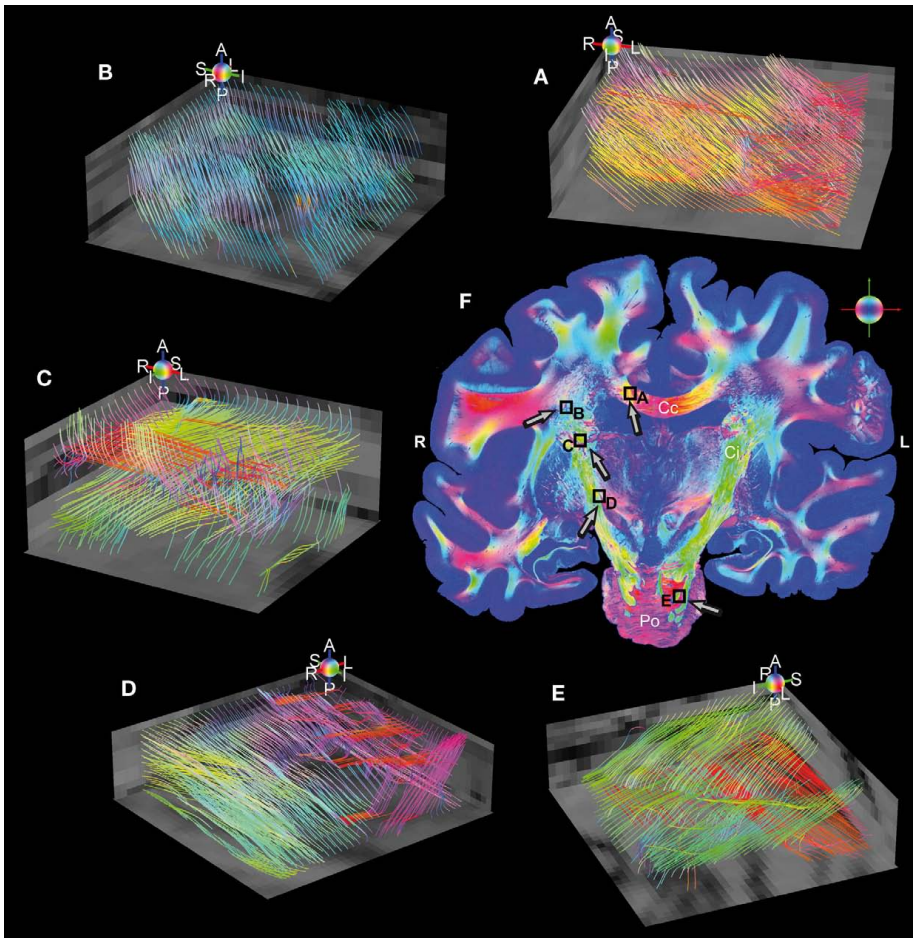


FIGURE 6 | Reconstructed fiber tract models in five regions of interest (volumes of 2 mm × 2 mm × 1 mm) sampled with an isotropic resolution of 100 μm. The individual color spheres (legend: A, anterior; P, posterior; I, inferior; S, superior; R, right; L, left) indicate the orientations of the fiber tract models. Fiber models were generated in **(A)** the Corpus callosum, **(B)** the Corona radiata, **(C,D)** the internal capsule (green), perforated by small fascicles (red and magenta) connecting the cerebral cortex with the Thalamus **(C)**, Red nucleus and Substantia nigra **(D)**, and **(E)** the Pons (green = cortico-spinal tract, red = transversal branches). Superimposed retardation maps (gray values) serve as anatomical references. **(F)** The RGB fiber orientation map is a representative of the stacked whole brain sections used for the study. The black rectangles highlight the magnified regions of interest **(A–E)** and the arrows indicate the individual observer's perspective. Legend: Cc, corpus callosum; Ci, internal capsule; Po, pons; R, right; L, left.

anatomical connectivity map could target three levels of organization (microscale, mesoscale, and macroscale) requiring reasonable cross-level integration.

MR-based diffusion imaging is the most frequently used method to visualize fiber pathways in both the living (Basser et al., 1994; Pierpaoli and Basser, 1996; Conturo et al., 1999; Beaulieu, 2002; Tuch et al., 2002, 2003; Hagmann et al., 2003; Tuch, 2004; Mori and Zhang, 2006; Schmahmann et al., 2007; Naidich et al., 2008; Johansen-Berg and Rushworth, 2009) and the postmortem (Englund et al., 2004; Larsson et al., 2004) human brain. With a spatial resolution at the millimeter scale, diffusion imaging contributes nearly exclusively to the construction of data sets at the level of macroscopic structural connectivity. However, restricted by the resolution, complex fiber networks and small fiber tracts cannot be discovered reliably at present. Furthermore, the terminal parts of fiber tracts within the cerebral cortex cannot be demonstrated.

Conversely, microscopic techniques generate data sets of impressing neuroanatomical detail, but they are limited to small sample sizes (i.e., small areas of interest in a small number of subjects). This substantially restricts their predictive power. In the recent years, anatomical connections in the human postmortem brains were studied with dissection techniques (Klingler, 1935; Türe et al., 2000), in myelin stained sections of adult human brains (Bürgel et al., 1997, 2006), or of immature brains taking advantage of heterochronic myelination of different fiber tracts during pre- and early postnatal development (Flechsig, 1901), in lesioned brains using various techniques for staining degenerating fibers (Funk and Heimer, 1967; Clarke and Miklossy, 1990), and using tract-tracing methods for discovering local connections (Burkhalter et al., 1993; Lanciego and Wouterlood, 2000). These studies have contributed to our knowledge about human brain fiber tracts, but all of these approaches suffer from severe restrictions, if fiber tracts are to be mapped in the adult human brain including their

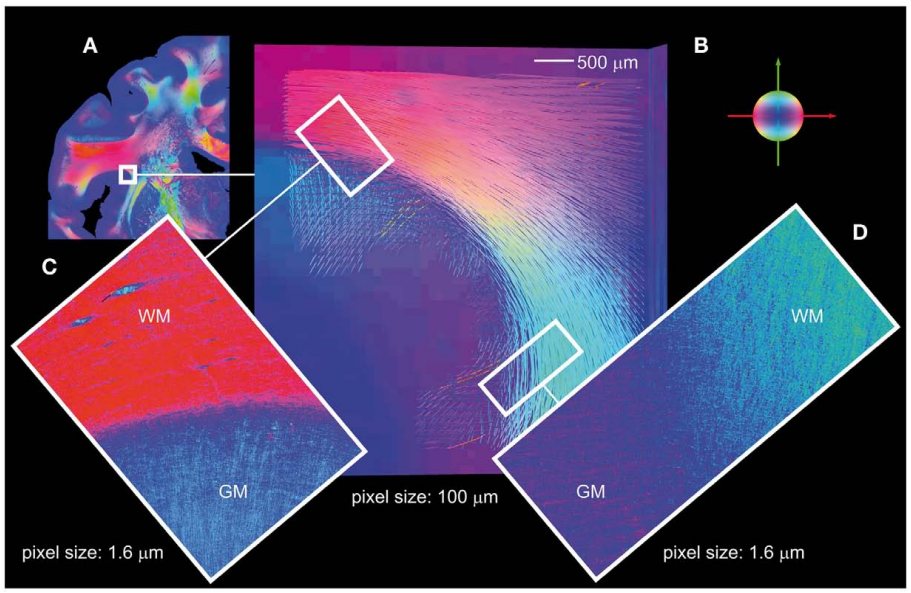


FIGURE 7 | Association fibers. (A) FOM of a coronal section with the indicated region of interest in the upper part of the circular sulcus of the insula. **(B)** Reconstructed U-fiber models (volume of 5 mm × 5 mm × 1 mm, pixel size of 100 µm) based on a stack of aligned FOMs obtained with the large-area polarimeter, covering the insular cortex and the underlying extreme capsule turning into the stem of the parietal operculum. Superficial layers of U-fibers turn into the insular cortex and into the cortex of the post-central gyrus. **(C,D)** Show the gray/white matter borders in the two regions of interest shown in **(B)**, analyzed with the polarizing microscope. Note, the two 2D-FOMs reflect the fiber orientations in a single section from the center of the stack. White matter (WM) and gray matter (GM) regions are characterized by significantly different fiber tract densities and fiber orientations. Color code of the fiber orientations: red = transversal, green = axial, blue = sagittal.

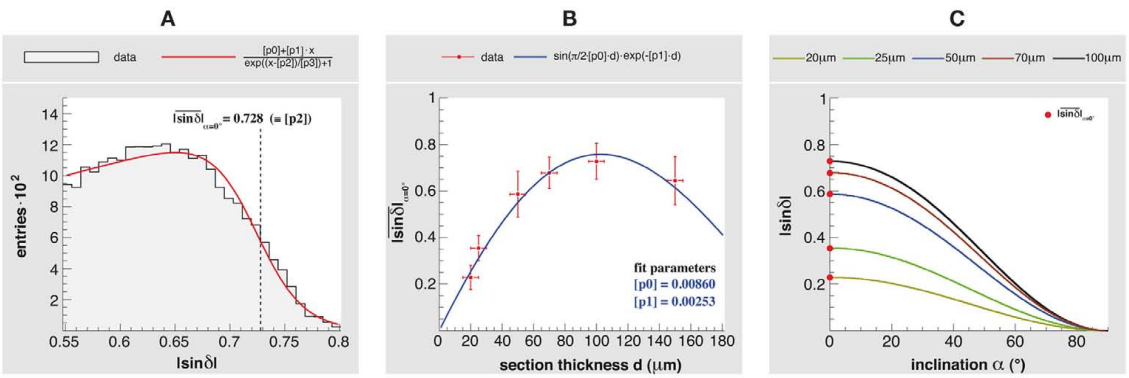


FIGURE 8 | Influence of the section thickness on the fiber inclination estimation. An optic tract was sectioned parallel to the main fiber direction with thicknesses of 20, 25, 50, 70, and 100 µm, and measured with the large-area polarimeter. **(A)** Histogram of the retardation map ($|\sin\delta| > 0.55$) of a 100 µm section and the resulting fit function (with fit parameter $[p2] = |\overline{\sin\delta}|_{\alpha=0°}$ used as a reference measure). **(B)** $|\overline{\sin\delta}|_{\alpha=0°}$ as a function of the section thickness. The error bars indicate the maximum deviation of the section thickness (±5 µm) and the maximum variance across the fitted histogram tail. **(C)** The relationship between the fiber inclination angle α and a measured retardation value is given by $|\sin\delta| = |\sin(\arcsin(|\overline{\sin\delta}|_{\alpha=0°}) \cdot \cos^2\alpha)|$.(cf. Axer et al., 2011). Since $|\overline{\sin\delta}|_{\alpha=0°}$ depends on the section thickness, the usable dynamic range to separate different fiber inclinations decreases for decreasing thicknesses.

3D courses. In contrast to studies in animals, the tight packing of different fiber tracts in the white substance, and the lack of specific tracers for *in vitro* tracking of long-distance fibers made comprehensive fiber tract mapping impossible in the adult human brain (Schmahmann and Pandya, 2009).

Obviously, there is still need for an operational link between both, the high-resolution micro-connectome and the averaged macro-connectome. In this paper, we propose to utilize 3D-PLI based on two polarimeter setups providing complementary scales of resolution, in order to close the gap between macroscopic fiber pathways revealed by diffusion weighted MRI and their real terminals located in the cortex: the polarizing microscope and the large-area polarimeter with tilting stage. The latter device has been optimized to one-shot whole brain imaging, while the polarizing

microscope is an ideal tool to disentangle complex fiber constellations in small brain areas and to follow fiber tracts to their terminals into gray matter regions. With an in-plane sampling between 33 and 100 μm, the large-area polarimeter enables the 3D modeling of long and short distance white matter tracts as shown in this paper (cf. **Figures 6** and **7**). The polarizing microscope features an in-plane pixel size down to axonal diameters (1.6 μm) together with a high sensitivity to small birefringent signals (cf. **Figure 7**). This *multiscale approach* utilizing the same technology provides data sets of importance to several branches of neuroscience, however, limited to a small number of subjects.

To successfully apply 3D-PLI to thin brain sections, a valid *bio-physical model* is required to describe the light transmittance through the polarimeter and the birefringent brain tissue. It has been shown in previous studies (Axer et al., 2000, 2011; Larsen et al., 2007) that basic principles of theoretical optics provide a sound base to estimate the light transmission, which is a measure of the locally prevailing spatial fiber orientation. Data fitting by means of discrete Fourier analysis based on the mathematical expression of this bio-physical model (cf. Eqs 1 and 2) enabled the analysis of light intensity profiles obtained for each tissue (image) pixel. This holds true for the large-area polarimeter as well as the polarizing microscope.

The used fiber model yields the local in-plane fiber direction φ with very high accuracy (±2°), while the most challenging parameter to be extracted is the inclination angle α. The interpretation of a light intensity profile is quite challenging in terms of fiber inclination, due to its complex dependencies on section thickness, birefringence, and light wavelength. In addition, the inclination angle cannot be extracted unambiguously from the theoretical model solely. The introduction of a tilting object stage into the large-area polarimeter represents a milestone for 3D-PLI, since the out-of-section fiber inclination (α) could be determined without ambiguity, for the first time. This enhances the reliability of the extracted fiber orientation significantly. Note, with a standard planar polarimeter the fiber orientation was intrinsically afflicted with an inclination sign ambiguity, which was formerly tackled with a simulated annealing approach as proposed by Larsen and Griffin (2004). However, it was demonstrated that the purely computer-aided application of the continuity heuristic could not entirely eliminate the ambiguity (Axer et al., 2011). Uncertainties evoked by this ambiguity were maximal (±90°), for fibers with intermediate inclinations (i.e., $\alpha \approx \pm 45°$), while fibers running within or perpendicular to the section plane were not affected. We found that the tilting method is most sensitive where a sign flip of the inclination has the greatest impact on the spatial fiber orientation, which are fibers with intermediate inclinations. This observation matches the expectation based on the theoretical model (cf. Eq 2) used to determine the fiber orientations.

The results presented here were mainly derived from brain sections with a thickness of 100 μm as this thickness was found to be well suited for the large-area polarimeter (Axer et al., 2001). Our study of an optic tract confirmed this result (cf. **Figure 8**). However, PLI inherently permits the reduction of section thicknesses to values considerably below 100 μm, even down to single fiber dimensions, but at the expense of the sampled dynamic range. We demonstrated this by investigating sections of the optic tract of different thicknesses (between 20 and 150 μm; cf. **Figure 8**). It is clear that the discrimination of different fiber orientations becomes more challenging and the demands on the imaging system are tightened when the section thickness decreases. In addition, the handling of whole brain cryo-sections also becomes difficult. However, investigating thin brain sections still represents an attractive approach to enhance the sampling resolution of 3D-PLI in the third dimension and to further reduce partial volume effects impairing the signal interpretation, e.g., in transition zones between fiber tracts with different orientations (**Figure 7** and Axer et al., 2011) or in regions with myelination gradients. The polarizing microscope with its high sensitivity is certainly a promising tool in terms of fiber inclination determination in thin sections (below 50 μm).

Although the tissue properties (myelination, section thickness, fiber orientation) are well integrated into the bio-physical model, to date, the 3D-PLI-based analysis does not take local variations in the axon myelination into account. As a consequence, regions characterized by significant myelination gradients (e.g., the transition domains between gray and white matter) are likely to be misinterpreted as changing fiber inclinations, if the local myelination is not considered. The misinterpretation becomes obvious in RGB FOMs (**Figure 5**), where cortical areas appear to be dominated by steep fiber orientations (blue color). Brain areas where the density of fibers diminishes significantly (e.g., subcortical nuclei) are quite challenging, since the detected amplitudes of the light intensity profiles are small, thus faking steep fibers. Therefore, fiber models were primarily reconstructed in white matter domains so far. However, in the current study we demonstrated that the extracted transmittance value I_0 is sensitive to local myelination and is, therefore, an ideal candidate to quantify the myelination in the same term as used for the determination of fiber inclinations. Cortical and subcortical gray matter could be distinguished from regions of strong myelination in the white matter, in accordance with results gained from myelin stained histological sections (cf. **Figure 3**). As a matter of fact, the transmittance basically reflects the attenuation of light when passing through the brain tissue. The observed discriminative absorption and scattering effects in the white matter are mainly due to the local shape and orientation of the refractive index ellipsoid determined by the myelination. This means that the transmittance is an indirect measure of the myelination slightly modified by the fiber orientation (which explains the intensity gradients in the white matter, Hebeda et al., 1994). Ongoing studies are currently dealing with the extension of the 3D-PLI analyses using the transmittance information.

The 3D reconstruction of gross histological brain sections poses particular challenges to the alignment techniques. Clearly, nerve fiber models of the human brain also depend on an accurate 3D volume data set of vectors derived from a registered stack of 2D-FOMs, which are individually taken from distorted histological brain sections. In order to preserve the reliability of the vector data on a sub-millimeter scale, the registration procedure needs to meet high-resolution requirements. A successful strategy to retrieve the 3D coherence among serial sections and correction for histological distortion was realized by using state-of-the-art image (affine and non-linear) registration techniques applied to blockface images (**Figure 2**) and to corresponding parameter maps (Palm et al.,

2010). By means of these instruments large global and local section deformations, unavoidably introduced during histological processing, were corrected with high quality, i.e., with a maximum displacement of 300 μm. The registration preserves the initial in-plane image resolution by using high-resolution blockface images as spatial references. In addition, the registration yields transformations related to rotation, shearing or non-uniform scaling. This information is used to re-orientate the fiber courses accordingly, finally providing an aligned stack of corrected FOMs – the 3D-FOM. High-quality image registration of large microscopic PLI data sets remains a task for future studies.

Promising approaches to reconstruct fiber tracts from 3D-FOMs are based on deterministic streamline region-to-region tractography (cf. **Figures 6** and **7**). Due to the complexity of the 3D-PLI data sets, the reconstruction is currently limited to small regions of interest (about 300 mm^3). To deal with crossing and kissing fiber constellations, partial volume effects in transitional domains between adjacent major fiber tracts, or gray/white matter borders, investigations beyond the standard streamline tractography are certainly needed. Therefore, dedicated utilization of high-performance computing and state-of-the-art tools for fiber tractography will be of importance to exploit the great potential of 3D-PLI in the near future.

CONCLUSION

Three-dimensional polarized light imaging enables the uniform high-resolution scanning of whole human brain sections, which allows for the investigation of both short-range axonal projections and long-distance links at the sub-millimeter scale. Depending on the employed optical system (large-area polarimeter or polarizing microscope), PLI is able to establish an interface between the microscopic and the macroscopic characterization of the anatomical connectivity by employing different polarimeter setups. Since the method is capable of detecting even small fiber tracts with myelination within the cerebral cortex, it is an ideal candidate to demonstrate the anatomical connections by means of *direct* measurements at complementary scales in the same object. Such multiscale data sets are valuable as they provide in 2D and 3D (i) where and to what extent fibers change orientation, (ii) to study specific tracts through regions of crossing and kissing fibers, and (iii) to tackle partial volume effects. Accurately following fibers through the complexity of white matter bundles has always been a crucial issue in MRI-based fiber tractography. 3D-PLI FOMs comprise vector fields that are comparable to the principal Eigenvectors extracted from a field of MR diffusion tensors, and are therefore best suited to be integrated in a multimodal approach to map connectional anatomy. This is extremely useful for the independent calibration and validation of next generation tracking software and the interpretation of the results derived from diffusion weighted MRI. Cross-validation of diffusion imaging and 3D-PLI in the same species is particularly desirable and will therefore be addressed in future studies.

Cross-linking of multiscale data sets from complementary neuroimaging techniques is a real challenge and requires precise definition of data structures, robust alignment tools, and interfacing modalities. In case of 3D-PLI, the blockface-based reference brain represents an operational link between the macroscopic MRI world and the high-resolution 3D-PLI modality. By means of 3D–3D registration, the reference brain can be transformed into the MR space, while the correspondence of the single blockface sections to the PLI parameter maps are still maintained. Once the interface is established, our multiscale approach for the same object is a most versatile tool to derive anatomical information (e.g., fiber orientations, myelination) at different measured scales of resolution (as exemplified in the reconstructed U-fiber region). In this way 3D-PLI provides the key technology and fundamental data structures to bridge the gap between the macro- and the micro-connectome.

ACKNOWLEDGMENTS

We would like to thank M. Cremer and F. Kocaer, Research Centre Jülich, Germany, for excellent technical assistance and preparation of the histological sections. Our work was partly supported by the Initiative and Network Fund of the Helmholtz Association within the Helmholtz Alliance on Systems Biology ("Human Brain Model").

REFERENCES

Amunts, K., and Zilles, K. (2006). "Atlases of the human brain: tools for functional neuroimaging," in *Neuroanatomical Tract Tracing 3: Molecules, Neurons, and Systems*, eds L. Zaborsky, F. G. Wouterlood, and J. L. Lanciego (New York: Springer), 566–603.

Axer, H., Axer, M., Krings, T., and Keyserlingk, D. G. V. (2001). Quantitative estimation of 3-D fiber course in gross histological sections of the human brain using polarized light. *J. Neurosci. Methods* 105, 121–131.

Axer, H., Berks, G., and Keyserlingk, D. G. V. (2000). Visualization of nerve fiber orientation in gross-histological sections of the human brain. *Microsc. Res. Tech.* 51, 481–192. '

Axer, M., Amunts, K., Gräßel, D., Palm, C., Dammers, J., Axer, H., Pietrzyk, U., and Zilles, K. (2011). A novelapproach to the humanconnectome: ultra-high resolutionmapping of fiber tracts in the brain. *Neuroimage* 54, 1091–1101.

Basser, P. J., Mattiello, J., and Le Bihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin-echo. *J. Magn. Reson. B* 103, 247–254.

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system – A technical review. *NMR Biomed.* 15, 435–455.

Brodmann, K. (1903). Bemerkungen zur Untersuchung des Nervensystems im polarisierten Lichte. *J. Psychol. Neurol.* 2, 211–213.

Brosseau, C. (1998). *Polarized Light – A Statistical Optics Approach.* NewYork: John Wiley & Sons.

Bürgel, U., Amunts, K., Hoemke, L., Mohlberg, H., Gilsbach, J. M., and Zilles, K. (2006). White matter fiber tracts of the human brain: three-dimensional mapping at microscopic resolution, topography and intersubject variability. *Neuroimage* 29, 1092–1105.

Bürgel, U., Mecklenburg, I., Blohm, U., and Zilles, K. (1997). Histological visualization of long fibre tracts in the white matter of adult human brains. *J. Brain Res.* 38, 397–404.

Burkhalter, A., Bernardo, K. L., and Charles, V. (1993). Development of local circuits in human visual cortex. *J. Neurosci.* 13, 1916–1931.

Capek, M., Bruza, P., Janacek, J., Karen, P., Kubinova, L., and Vagnerova, R. (2009). Volume reconstruction of large tissue specimensfrom serial physical sections usingconfocalmicroscopy and correction of cuttingdeformations by elastic registration. *Microsc. Res. Tech.* 72, 110–119.

Clarke, S., and Miklossy, J. (1990). Occipital cortex in man: organization of callosal connections, related myelo- and cytoarchitecture, and putative boundaries of functional visual areas. *J. Comp. Neurol.* 298, 188–214.

Conturo, T. E., Lori, N. F., Cull, T. S., Akbudak, E., Snyder, A. Z., Shimony, J. S., McKinstry, R. C., Burton, H., and Raichle, M. E. (1999). Tracking neuronal fiber pathways in the living human brain. *Proc. Natl. Acad. Sci. U.S.A.* 96, 10422–10427.

Dammers, J., Axer, M., Gräßel, D., Palm, C., Zilles, K., Amunts, K., and Pietrzyk, U. (2010). Signal enhancement in polarized light imaging by means of independent component analysis. *Neuroimage* 49, 1241–1248.

Dammers, J., Breuer, L., Axer, M., Kleiner, M., Eiben, B., Gräßel, D., Dickscheid, T., Zilles, K., Amunts, K., Shah, J., and Pietrzyk, U. (2011). Automatic identification of gray and white matter components in polarized light imaging. *Neuroimage* 59, 1338–1347.

Dauguet, J., Delzescaux, T., Conde, F., Magnin, J. F., Ayache, N., Hantraye, P., and Frouin, V. (2007). Three-dimensional reconstruction of stained histological slices and 3D non-linear registration with in-vivo MRI for whole baboon brain. *J. Neurosci. Methods* 164, 191–204.

Eiben, B., Palm, C., Pietrzyk, U., Davatzikos, C., and Amunts, K. (2010). "Perspective error correction using registration for blockface volume reconstruction of serial histological sections of the human brain," in *Bildverarbeitung für die Medizin*, eds T. M. Deserno, H. Handels, H. P. Meinzer, and T. Tolxdorff (Berlin: Springer), 301–305.

Englund, E., Sjöbeck, M., Brockstedt, S., Lätt, J., and Larsson, E. M. (2004). Diffusion tensor MRI post mortem demonstrated cerebral white matter pathology. *J. Neurol.* 251, 350–352.

Farrell, A., Rouseff, D., and McCally, L. R. (2005). Propagation of polarized light through two- and three-layer anisotropic stacks. *J. Opt. Soc. Am.* 22, 1981–1992.

Fischler, M. A., and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fittingwithapplications to image analysis and automated cartography. *Comm. ACM* 24, 381–395.

Flechsig, P. (1901). Developmental (myelongenetic) localisation of the cerebral cortex in the human subject. *Lancet* 158, 1027–1030.

Fraher, J. P., and MacConnaill, M. A. (1970). Fibre bundles in the CNS revealed by polarized light. *J. Anat.* 106, 170.

Funk, R. P., and Heimer, L. (1967). Two methods for selective silver impregnation of degenerating axons and their synaptic endings in the central nervous system. *Brain Res.* 4, 36–374.

Glazer, A. M., Lewis, J. G., and Kaminsky, W. (1996). An automatic optical imaging system for birefringent media. *Proc. R. Soc. Lond. A* 452, 2751–2765.

Göthlin, G. F. (1913). Die doppelbrechenden Eigenschaften des Nervengewebes. *Kungl. Svenska Vetenskapsakad. Handl.* 51, 1.

Hagmann, P., Thiran, J. P., Jonasson, L., Vandergheynst, P., Clarke, S., Maeder, P., and Meuli, R. (2003). DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection. *Neuroimage* 19, 545–554.

Hartley, R. I., and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge: University Press.

Hebeda, K. M., Menovsky, T., Beek, J. F., Wolbers, J. G., and van Gemehrt, M. J. C. (1994). Light propagation in the brain depends on nerve fiber orientation. *Neurosurgery* 35, 720–724.

Johansen-Berg, H., and Rushworth, M. F. (2009). Using diffusion imaging to study human connectional anatomy. *Annu. Rev. Neurosci.* 32, 75–94.

Jones, R. C. (1941). A new calculus for the treatment of optical systems. *J. Opt. Soc. Am.* 31, 488–493.

Klingler, J. (1935). Erleichterung der makroskopischen Präparation des Gehirns durch den Gefrierprozess. *Schweiz. Arch. Neurol. Psychiatr.* 36, 247–256.

Kötter, R. (2007). "Anatomical concepts of brain connectivity," in *Handbook of Brain Connectivity*, eds V. K. Jirsa and A. R. MacIntosh (Berlin: Springer), 149–167.

Lanciego, J. L., and Wouterlood, F. G. (2000). Neuroanatomical tract-tracing methods beyond 2000: what's now and next. *J. Neurosci. Methods* 103, 1–2.

Larsen, L., and Griffin, L. D. (2004). "Can a continuity heuristic be used to resolve the inclination ambiguity of polarized light imaging?" in *CVAMIA-MMBIA, LNCS 3117*, eds M. Sonka, I. A. Kakadiaris, and J. Kybic (Berlin: Springer), 365–375.

Larsen, L., Griffin, L. D., Gräßel, D., Witte, O. W., and Axer, H. (2007). Polarized light imaging of white matter architecture. *Microsc. Res. Tech.* 70, 851–863.

Larsson, E. M., Englund, E., Sjöbeck, M., Lätt, J., and Brockstedt, S. (2004). MRI with diffusion tensor imaging post-mortem at 3.0 T in a patient with frontotemporal dementia. *Dement. Geriatr. Cogn. Disord.* 17, 316–319.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* 60, 91–110.

Martenson, R. E. (1992). *Myelin: Biology and Chemistry*. Boca Raton: CRC Press.

Massoumian, F., Juskaitis, R., Neil, M. A. A., and Wilson, T. (2003). Quantitative polarized light microscopy. *J. Microsc.* 209, 13–22.

Mori, S. (2007). *Introduction to Diffusion Tensor Imaging*. Amsterdam: Elsevier B.V.

Mori, S., and Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron* 51, 527–539.

Naidich, T. P., Duvernoy, H. M., Delman, B. N., Sorensen, A. G., Kollias, S. S., and Haacke, E. M. (2008). *Duvernoy's Atlas of the Human Brain Stem and Cerebellum*. Vienna: Springer.

Norton, W. T., and Cammer, W. (1984). *Myelin*. New York: Plenum Press.

Oishi, K., Faria, A., van Zijl, P. CM., and Mori, S. (2011). *MRI Atlas of Human White Matter*. London: Academic Press.

Oldenbourg, R. (1996). A new view on polarization microscopy. *Nature* 381, 811–812.

Oldenbourg, R., and Mei, G. (1995). New polarized light microscope with precision universal compensator. *J. Microsc.* 180, 140–147.

Oldenbourg, R., Salmon, E. D., and Tran, P. T. (1998). Birefringence of single and bundled microtubules. *Biophys. J.* 74, 645–654.

Palm, C., Axer, M., Gräßel, D., Dammers, J., Lindemeyer, J., Zilles, K., Pietrzyk, U., and Amunts, K. (2010). Towards ultra-high resolution fibre tract mapping of the human brain – Registration of polarized light images and reorientation of fibre vectors. *Front. Hum. Neurosci.* 4, 1–16.

Pierpaoli, C., and Basser, P. J. (1996). Toward a quantitative assessment of diffusion anisotropy. *Magn. Reson. Med.* 33, 893–906.

Saleh, B. E. A., and Teich, M. C. (1991). *Fundamentals of Photonics*. New York: John Wiley & Sons.

Scheuner, G., and Hutschenreiter, J. (1972). *Polarisationsmikroskopie in der Histophysik*. Leipzig: VEB Georg Thieme.

Schmahmann, J. D., and Pandya, D. N. (2009). *Fiber Pathways of the Brain*. New York: Oxford University Press.

Schmahmann, J. D., Pandya, D. N., Wang, R., Dai, G., D'Arceuil, H. E., deCrespigny, A. J., and Wedeen, V. J. (2007). Association fibre pathways of the brain: parallel observations from diffusion spectrum imaging and autoradiography. *Brain* 130, 630–653.

Schmidt, W. J. (1923). Zur Doppelbrechung des Nervenmarks. *Z. Wiss. Mikrosk.* 41, 29–38.

Schmitt, O., and Bear, R. S. (1937). The optical properties of vertebrate nerve axon as related to fiber size. *J. Cell. Comp. Physiol.* 9, 261–273.

Schnabel, R. (1966). Zur Mikroskopischen Untersuchung optisch anisotroper Strukturen des Nervensystems mit circular polarisiertem Licht. *Acta Neuropathol.* 7, 180–184.

Singh, M., Rajagopalan, A., Kim, T. S., Hwang, D., Chui, H., Zhang, X. L., Lee, A. Y., and Zarow, C. (2008). Co-registration of in vivo human MRI brain images to postmortem histological microscopic images. *Int. J. Imaging Syst. Technol.* 18, 325–335.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi:10.1371/journal.pcbi.0010042

Toga, A. W., Thompson, P. M., Mori, S., Amunts, K., and Zilles, K. (2006). Towards multimodal atlases of the human brain. *Nat. Rev. Neurosci.* 7, 952–966.

Tuch, D. S. (2004). Q-ball imaging. *Magn. Reson. Med.* 52, 1358–1372.

Tuch, D. S., Reese, T. G., Wiegell, M. R., Belliveau, J. W., and Wedeen, V. J. (2003). Diffusion MRI of complex neural architecture. *Neuron* 40, 885–895.

Tuch, D. S, Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Wedeen, V. J. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magn. Reson. Med.* 48, 577–582.

Türe, U., Yasargil, M. G., Friedmann, A. H., and Al-Mefty, O. (2000). Fiber dissection technique: lateral aspect of the brain. *Neurosurgery* 47, 417–427.

Vidal, B., Silveira Mello, M. L., Caseiro-Filho, A. C., and Godo, C. (1979). Anisotropic properties of the myelin sheath. *Acta Histochem.* 66, 32–39.

Wiener, O. (1912). Die Theorie des Mischkörpers für das Feld der stationären Strömung. Abhandlung der math.-phys. Klasse k. sächs. *Ges. Wiss.* 32, 507–604.

Zilles, K., Schleicher, A., Palomero-Gallagher, N., and Amunts, K. (2002). "Quantitative analysis of cyto- and receptor architecture of the human brain," in *Brain*

Mapping: The Methods, eds J. C. Mazziotta and A. W. Toga (Amsterdam: Elsevier B.V.), 573–602.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 March 2011; accepted: 08 December 2011; published online: 30 December 2011.
*Citation: Axer M, Grässel D, Kleiner M, Dammers J, Dickscheid T, Reckfort J, Hütz T, Eiben B, Pietrzyk U, Zilles K and Amunts K (2011) High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging. Front. Neuroinform. **5**:34. doi: 10.3389/fninf.2011.00034*

frontiers in
NEUROINFORMATICS

PERSPECTIVE ARTICLE
published: 24 April 2012
doi: 10.3389/fninf.2012.00013

The importance of combining MRI and large-scale digital histology in neuroimaging studies of brain connectivity and disease

Jacopo Annese*

Department of Radiology and Biomedical Imaging, The Brain Observatory, University of California, San Diego, CA, USA

Edited by:
Olaf Sporns, Indiana University, USA

Reviewed by:
Trygve B. Leergaard, University of Oslo, Norway
Katrin Amunts, RWTH Aachen University, Germany

****Correspondence:***
Jacopo Annese, Department of Radiology and Biomedical Imaging, The Brain Observatory, 3510 Dunhill Street, San Diego, CA 92121, USA.
e-mail: jannese@ucsd.edu

One of the major issues hindering a comprehensive connectivity model for the human brain is the difficulty in linking Magnetic Resonance Imaging (MRI) measurements to anatomical evidence produced by histological methods. *In vivo* and postmortem neuroimaging methodologies are still largely incompatible in terms of sample size, scale, and resolution. To help bridge the hiatus between different approaches we have established a program that characterizes the brain of individual subjects, combining MRI with postmortem neuroanatomy. The direct correlation of MRI and histological features is possible, because registered images from different modalities represent the same regions in the same brain. Comparisons are also facilitated by large-scale, digital microscopy techniques that afford images of the whole-brain sections at cellular resolution. The goal is to create a neuroimaging catalog representative of discrete age groups and specific neurological conditions. Individually, the datasets allow for investigating the relationship between different modalities; combined, they provide sufficient predictive power to inform analyses and interpretations made in the context of non-invasive studies of brain connectivity and disease.

Keywords: histology, MRI, DTI, fibers, pathology, connectivity, human, brain

INTRODUCTION

For centuries, the human brain remained impervious to direct observation and experimentation (Penfield, 1958), but with the advent of Magnetic Resonance Imaging (MRI) this situation changed very quickly. The technique supported the fastest and most significant progress toward understanding the functional architecture of the human brain; after three decades of MRI, thousands of individuals of all ages, conditions, and denominations have been scanned, leading to the emergence of comprehensive models of brain structure and function. Functional imaging studies recently broadened their focus from measuring local phenomena associated with specific perceptual and cognitive tasks, to identifying long-range networks involved in the orchestration of neural activity (Sporns et al., 2005). These networks are revealed statistically using resting-state functional MRI (R-fMRI) based on the time course of slow fluctuations in the BOLD signal when the subject is not engaged in any particular task and when the brain is at "rest" (De Luca et al., 2006). Whether R-fMRI-derived models depend on hard-wired anatomical connectivity is an open question, so this technique is often combined with diffusion tensor imaging (DTI) to show major fiber tracts in the deep white matter (Pierpaoli et al., 1996; Bandettini, 2009). These non-invasive neuroimaging modalities are the most widely utilized methodologies for mapping brain circuitry at the system level, and thus also the primary tools of the Human Connectome Project (HCP; Akil et al., 2011).

The problem with images of the brain acquired *in vivo* with MRI and DTI is that they are limited in both resolution and contrast. The data needs be averaged across many subjects who represent broad demographic or clinical groups in order to extract significant measures. In fact, even the quantification of biomarkers that are crucial clinically for the individual patient is contingent on population-based atlases that only contain very rudimentary anatomical information. In practice, it is not possible with MRI to localize the exact borders of neighboring neuroanatomical structures and identify the underlying histological features on which radiologic images depend. The latter can only be determined postmortem.

Given the undeniable importance of validating MRI, why have neuroanatomical and histological studies have only played a small, tangential role in human brain mapping and why does the HCP, despite the large investment by the National Institute for Health (NIH), not formally include a parallel histological program? The constraints are not conceptual; everyone agrees it would be very useful to corroborate non-invasive measures with baseline information on the actual properties of brain tissue. However, DTI-based models of fiber tract orientation in the human brain are currently validated only on the basis of classic atlases (Dejerine, 1895; Flechsig, 1901; Ludwig and Klinger, 1956; Yakovlev and Lecours, 1967). These contain anachronistically zealous, yet subjectively construed illustrations and were derived from the dissection of a few undocumented brain specimens, so the level of comparison is severely limited.

The challenges in creating a comprehensive histological model of the human brain that matches the scope and usability of MRI population-based templates, including the proposed Connectome, are primarily logistical. To begin with, any study that aims at examining cellular-level features in a brain depends

on the expiration of the subject, and consequently neuroanatomical studies have been confined to those few cases where autopsies were prescribed because of clinical or legal concerns or to individuals who enrolled in tissue donation programs. Because in the first scenario the rate of autopsy has universally been decreasing (Kretzschmar, 2009), the enhancement of brain donation programs (via incentives, outreach, and education) represents the only realistic future opportunity of supporting a large-scale, neuro-histological mapping effort.

Assuming that the necessity of enrolling a large enough number of brain donors has been addressed, noting that the benefits of such donor program cannot be expected to roll out in the short-term, what pre-mortem information should characterize the specimen and how should the brain be processed in order to create images that can complement and validate MRI studies? Because the main goal is to bridge the gap between non-invasive neuroimaging and histology it would be useful to have access to radiologic images acquired *in vivo* for each case that comes to autopsy. When comparisons are made between images from different modalities in the context of the same subject all the information necessary to localize and quantify microscopic features underlying the patterns revealed in MRI images is available. Furthermore, the correspondence between images acquired *in vivo* and those from histopathology is enormously improved if the latter cover the *whole* brain, just like MRI. These methodological prerogatives are the basis for a novel neuroimaging resource that will be used to cross-reference different modalities and establish fundamental properties of brain structure at multiple scales.

The approach is exemplified by the multimodal examination of the brain of H.T., one of our program participants who passed away in May 2009. Her neuroimaging profile was created using a palette of multiple modalities that provided correlated images of the brain at different levels of resolution. The study that is described in this communication was approved by the University of California's Institutional Review Board.

CASE DESCRIPTION

H.T. was a woman of Caucasian descent born in Rochester, NY in 1920. She had 15 years of education which included two years of college. A smoker for 50 years, her only major medical concern was hypertension and she took medication to lower her blood pressure. She died of heart failure. During the last year of her life she experienced a slight decline in memory and cognitive proficiency (it should be noted that these phenomena were observed by her closest relatives; H.T. never recognized or admitted to having any impairment). The examination of the clinical scans by a trained neuroradiologist did not report major abnormalities in the brain of H.T. with the exception of several small lacunar infarcts.

MULTI-SCALE NEUROIMAGING

H.T. died of cardiac arrest at the age of 88. Because she wore a pacemaker at the time of her enrollment in our study (thus precluding her from undergoing MRI scanning while she was living), multiple scans of the brain *in situ* were acquired post-mortem. Images of the brain were acquired on an "HDX" Twinspeed EXCITE 1.5T scanner (Milwaukee, WI.) using an eight-channel phased-array head coil. Scan parameters were selected to allow robust reconstruction of the brain's cortical surface using FreeSurfer Software and automated segmentation of cortex and subcortical structures (Fischl et al., 2002, 2004; **Figure 1A**). DTI imaging consisted in the collection of 51 diffusion gradient directions (**Figure 2A**).

After the scans of the brain *in situ* were completed the brain was extracted at autopsy and suspended by the basilar artery in fixative solution (4% paraformaldehyde phosphate-buffered solution) at 4°C. After six weeks of fixation in formaldehyde, a second series of scans were performed *ex situ* (**Figure 1B**). The protocol lasted approximately 36 h, using the same sequences that were run in the previous session but with multiple acquisitions (4–32 NEX) in order to improve the signal-to-noise ratio in the fixed tissue and increase resolution.

Fixation was protracted for 10 weeks after the second MRI session and the tissue was subsequently cryo-protected by immersion in 30% sucrose. Following the removal of superficial blood vessels and pial membranes, the whole specimen was embedded in a block of 10% gelatin and frozen in a bath of chilled isopentane (−40°C). The gelatin-brain block was attached to a custom-engineered freezing stage and sectioned on a large motorized sliding microtome (Leica Microsystems Inc., Bannockburn, IL) during an uninterrupted cutting procedure. The tissue was cut at an interval of 70 microns. Digital images of the cut surface were acquired before each tissue slice was collected using a digital, single-lens reflex camera (Nikon D700; Nikon Inc., Melville, NY) that was mounted directly in line with the microtome stage. Image acquisition during the sectioning procedure produced an unabridged series of tomographic anatomical images through the brain (**Figure 1C**). These data contain excellent tissue contrast, ideal for anatomical delineations and the stack is the basis for the correct alignment and 3-D reconstruction of corresponding tissue slices.

Two adjacent, regularly spaced series of giant histological slices were stained using Thionin (Nissl staining) and a sensitive silver impregnation technique based on the original Gallyas (1979) protocol for myelinated fibers (Annese et al., 2004; **Figure 1D**). Tissue slices were mounted on large glass slides (5 × 7 inches) and the tissue was imaged on custom-engineered large-format microscope scanners. The core digitizing unit is composed of a computer-controlled microscope, a linear-encoded motorized stage, a line scanner camera, and storage server (Annese et al., 2009). The images representing the entire giant tissue slice are typically composed of 40,000 image tiles that are acquired systematically at 20× magnification (resolution: ∼0.4 μm/pixel). Image tiles are stitched into pyramidal "virtual" slices that allow for fast exploratory viewing and the topographic localization of microscopic features (Mikula et al., 2007; Weinstein et al., 2009). The images are very large (dimensions in pixels are: 334,500 × 266,200; file size on disk 250–400 GB) nevertheless, as JPG2000 compressed pyramidal files (size reduced to 15 GB), the images can be examined through increasing magnification levels while a fixed number of pixels spanning progressively smaller fields of view are shown. This technology, combined with the possibility of surveying slices at multiple locations in the 3D model of the

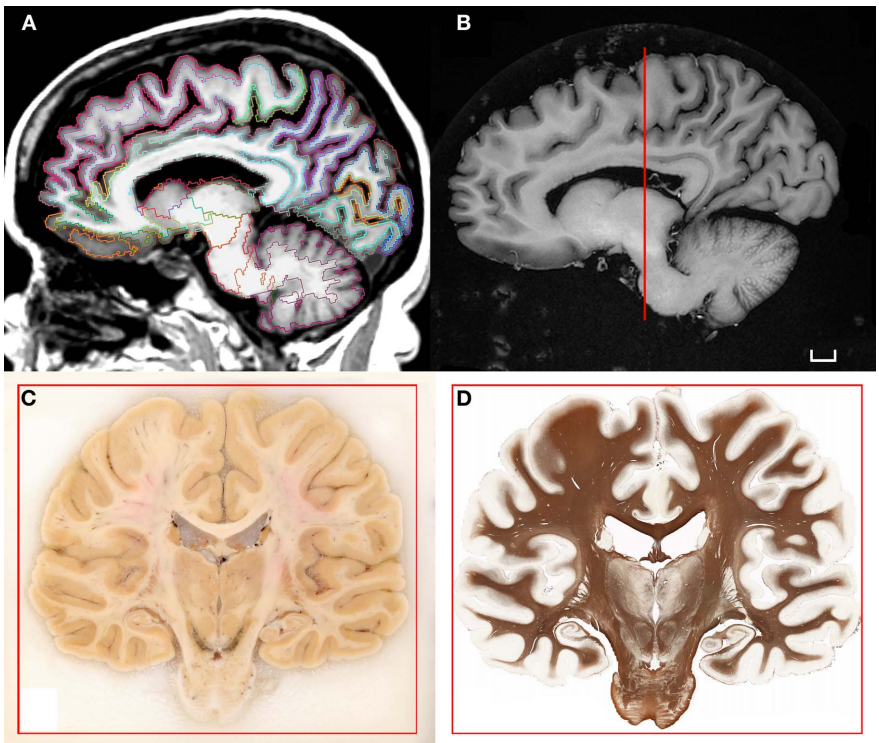


FIGURE 1 | (A) T1-weighted scan of the brain of H.T. acquired *in situ* 3D IR-SPGR, matrix size: 256 × 256, voxel size of 0.9375 × 0.9375, slice thickness = 1.2 mm. The volume underwent the automated labeling of subcortical structures and cortical gyri that is part of the morphometry pipeline included in the software Freesurfer (Fischl et al., 2002, 2004). **(B)** High-resolution T1-weighted scan of the formalin-fixed specimen imaged *ex situ*. The brain was scanned in a dedicated chamber filled with phosphate buffer (matrix size: 512 × 512, pixel size: 0.5/0.5, slice thickness: 1.2 mm). Multiple acquisitions were repeated in order to increase the signal-to-noise ratio. **(C)** Coronal blockface image from the same brain at the level of the posterior hippocampus. **(D)** Corresponding histological slice stained for myelinated fibers [protocol modified from Gallyas (1979); Annese et al. (2004)]. Scale bar: 1 cm.

brain, is the key to the proper comparison of histological and neuroimaging data sets.

DIFFERENT VIEWS OF THE DEEP WHITE MATTER

The anatomical and DTI data were aligned using tools for automated affine registration included in Amira (Developer Package; Visage Imaging Inc, San Diego, CA). A binary mask based on color segmentation (Annese et al., 2006) was created to separate brain tissue from the surrounding gelatin matrix in the tomographic images acquired during the sectioning procedure. These were reconstructed into a volume and this dataset was registered to the average B_0 volume acquired during the DTI protocol. Because the anatomical volume is composed of images that correspond to unique histological sections, it is possible to identify which sections cross any region of interest in the DTI data. The direct, landmark-based comparison of two equivalent coronal images at the level of the posterior hippocampus reveals the histological complexity underlying tensor-based maps of fiber orientation (**Figures 2A,B**).

This approach permits quantitative correlations between diffusion-based measures of fiber integrity and directionality (derived from fractional anisotropy values) with formal descriptions of axonal architecture. The latter can be produced from high-resolution histological images (**Figure 2C**) using automated (hence potentially large-scale) image analysis routines (**Figure 2D**). Quantification at the histological level produces actual values representing main fiber directionality, axonal diameter, and cross-over (Bartsch et al., 2011).

It should be noted that myelination is not the only histological feature that affects MRI signal; gliosis, and perivascular neuropathogenetic phenomena produce visible effects in MRI images of the deep white matter. In order to decipher the relationship between non-invasive imaging and the underlying histology, multiple sequences with different scanning parameters must be correlated with the co-localization of different histopathologic markers of structure and disease. Using specific antibody staining, it is possible to localize neuropathological phenomena that produce visible MRI abnormalities. Each subject-specific data set should be analyzed as a registered stack of anatomical layers that contain overlapping maps of biomarkers. In the context of our methodological pipeline, MRI and DTI are also "stains" in the sense that they show properties of the tissue that are complementary to the features detected by histochemical or immuno-cytological methods.

A better understanding of local white matter architecture at the microscopic level can help defining DTI acquisition and modeling parameters; these may need to be specific to different regions of the white-matter if the goal is to generate accurate whole-brain

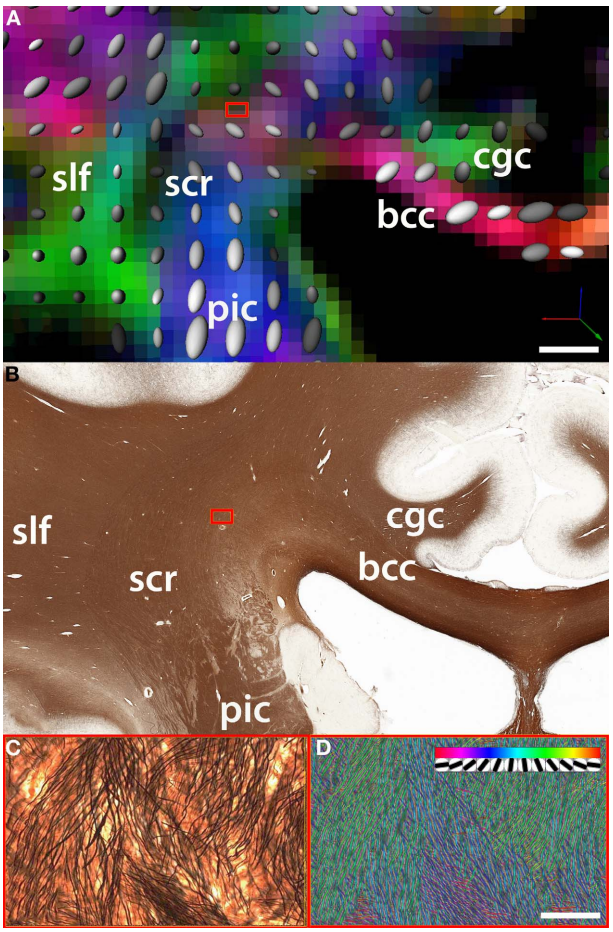


FIGURE 2 | Comparison between (A) fiber orientation maps derived from DTI and (B) the underlying myelo-architecture in the region of intersection of the fibers belonging to the corpus callosum (cc) which travel along the coronal plane (parallel to the plane of the figure) and the superior corona radiata (scr). In this example, the region of intersection between cc fibers and the scr is an area that shows significant crossing over of fibers of small caliber (blue box). **(C)** One DTI voxel-wide selection from silver-stained tissue slice (magnification = 20×). **(D)** The density and orientation of single white matter fibers is quantified using a template matching algorithm [Bartsch et al. (2011)]. The insert in panel **D** shows the discrete orientations of the templates used to classify fiber bundles. DTI acquisition parameters: image matrix size = 96 × 96, B_0 = 1000, 51 directions, FOV = 24 cm, 47 axial slices, slice thickness = 2.5 mm. Fiber track definition derived from the atlas of Oishi and colleagues (2011); bcc: body of the corpus callosum; cgc: cingulum; slf: superior longitudinal fasciculus. Scale bars: Panel B = 5 mm; Panel C = 100 microns.

models of connectivity. In general, the approach demonstrates the possibility of extending animal neuroimaging studies and resources that combine non-invasive and histological mapping (Johnson et al., 2010) to the human brain.

POTENTIAL AND FUTURE PROSPECTS

The power of our approach lies in the fact that the resolution of the images of H.T.'s brain created with different modalities spans from several millimeters (2.5 mm/voxel in our 3-D DTI data) to a few hundred nanometers (0.3 μm/pixel in the 2-D histological images). The MRI (and DTI) of the brain acquired *in situ*, containing gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) and representing the original geometry is necessary because these scans link the microscopic maps created from H.T.'s brain to the very large base of non-invasive structural and functional neuroimaging data that will be collected by the HCP. Furthermore, the neuroimaging sequences that were used to acquire images of the brain of H.T. can be matched to the standard protocols employed by the HCP and other ongoing neuroimaging projects to make direct correlations and inform the development of analysis tools. This way, the growing sample of brain specimens that will be fully characterized at the microscopic level will also be accessible as anatomical reference to large-scale neuroimaging databases. This cross-sectional data will serve as a base line reference for the interpretation of low-resolution images. One could additionally envisage "embedding" MRI data-sets, like that belonging to the brain of H.T., into the processing pipelines of the HCP and other purely non-invasive studies. The output of these pipelines could be evaluated based on corresponding histological images, providing an efficient means to algorithm optimization.

As noted in the previous section, mapping local microscopic features within a 3D model of the whole brain can lead to the formulation of "signature" models that classify neuroanatomical features in specific compartments of the brain. Histological models may help predict individual anatomy (and pathology) from the macroscopic markers that are visible and measurable in non-invasive low-resolution images. In other words, local micro-architectural patterns that can be localized consistently in a large number of subjects may eventually be used to resolve ambiguities in the interpretation of low-resolution data (Leergaard et al., 2010). The systematic characterization and classification of local micro-architectural patterns as exemplified in **Figure 2** will be extremely useful for the calibration of the next generation of fiber tracking software and interpretation of the results from diffusion imaging.

Our methodology may not be universally feasible for routine implementation. Aside from equipment and operational costs, the computational requirements for managing massive histological data-sets are considerable. However, depreciation of digital storage will occur sufficiently rapidly to justify the timely implementation of the protocols described above. Many web-based products and initiatives, like Google Maps and the Hubble Telescope, leverage on remote access and interoperability with digital images in the Gigapixel range. The digitization at 20× magnification of whole histological sections of the human brain actually produces images in the Terapixel range; these can nonetheless be shared on the web with the support of appropriate hardware and bandwidth. To demonstrate this functionality, The Brain Observatory developed the infrastructure for a web-based atlas of the human brain that can deliver Terapixel histological images representing whole brain sections at cellular resolution; these are associated with other views of the brain created in 2D and 3D using other neuroimaging modalities (Annese et al., 2012).

The development of robust quantitative methods for registration and analysis combined with the availability of adequate computational resources for handling very large data-sets will

make comparisons between neuroimaging data and the underlying microscopic features increasingly effective. Furthermore, as the number of single-subject neuroimaging studies like the one involving H.T, where we combined *in vivo* and postmortem modalities and for whom we were able to compute maps that span multiple spatial scales, increases the predictive power of the digital neuroanatomical catalog will become stronger, representing a useful resource for data mining, cross-modality validations, and algorithm development.

The example provided in this article should support the notion of multi-modal, single-subject paradigm as a complement to MRI data obtained in the context of large-scale neuroimaging initiatives such as the HCP (Marcus et al., 2011). However, a note of caution is due before the concluding remarks. Histological images only contain a final, "freeze-frame" picture of the brain's microanatomy; whereas we know that the architecture of neural circuits, from the system level to local dendritic properties, changes with maturation, behavior, and disease. Secondly, histology should not be considered unquestioningly as the "gold standard" for neuroimaging. Postmortem methods produce well-known artifacts and admittedly it is extremely difficult to produce consistent results on a large scale; nevertheless, the implementation of "industrial" and carefully controlled protocols for tissue processing can greatly increase the quality and reproducibility of microscopic data. Common standards are needed for histological processing, as well as for the acquisition, storage, and sharing of digital histological files. The importance of incorporating maps of microscopic anatomy into the HCP and other neuroimaging studies ought to justify these efforts and promote substantial investments that will provide solutions to the logistic and technical challenges that still hinder postmortem studies of the human brain (Crick and Jones, 1993). Given that modern clinical strategies for prevention and treatment of neurological disease depend largely on knowledge acquired via non-invasive MR-based tools and thus from the *indirect* observation of developmental and neuropathogenetic phenomena, it is crucial at this point in time to support a new phase of validation studies based on rigorously prepared histological material. These studies ought to be commensurate both in scale and funding to the ambitious neuroimaging initiatives like the HCP. After all, truth is in the brain of the beholder.

ACKNOWLEDGMENTS

The author would like to acknowledge the contribution of all the research staff of The Brain Observatory. In particular Mr. Paul Maechler and Dr. Hauke Bartsch had a primary role in the generation of the data and results that were discussed in this communication. Jacopo Annese would like to thank Miss Colleen Sheh for the assistance in generating the illustrations and Miss Chris Ha and Dr. Justin Feinstein for their useful comments on the manuscripts.

The work was supported by two research grants from the National Eye Institute, R01 EY018359–02 and ARRA R01 EY018359–02S1 (Jacopo Annese Principal Investigator) and a research grant from the National Institute for Mental Health R01MH084756 (Jacopo Annese Principal Investigator).

REFERENCES

Akil, H., Martone, M. E., and van Essen, D. C. (2011). Challenges and opportunities in mining neuroscience data. *Science* 331, 708–712.

Annese, J. (2010). "Deconstructing Henry. The Neuroanatomy and Neuroinformatics of the Brain of the Amnesic Patient H.M. Program No. 397.18," *Neuroscience Meeting Planner. 40th Annual Meeting of the Society for Neuroscience*. (San Diego, CA). [Online].

Annese, J., Schenker, M. N., Ha, C., Maechler, P., Sheh, C., and Bartsch, H. (2012). A public atlas of the human brain. Web-resource. http://thebrainobservatory.ucsd.edu/content/public-brain-atlas

Annese, J., Pitiot, A., and Toga, A. W. (2004). A myelo-architectonic method for the structural classification of cortical areas. *Neuroimage* 21, 15–26.

Annese, J., Schenker, M. N., and Candido, C. (2009). "Anatomical mapping and digital pathology of the human brain by Giant Whole Slide Imaging (GWSI)," *Program No. 104.5. Neuroscience Meeting Planner. 39th Annual Meeting of the Society for Neuroscience*. (Chicago, IL). [Online].

Annese, J., Sforza, D. M., Dubach, M., Bowden, D., and Toga, A. W. (2006). Postmortem high resolution 3-dimensional imaging of the primate brain. Blockface imaging of perfusion stained tissue. *Neuroimage* 30, 61–69.

Bandettini, P. A. (2009). What's new in neuroimaging methods? *Ann. N.Y. Acad. Sci.* 1156, 260–293.

Bartsch, H., Maechler, P., and Annese, J. (2011). "Classification of NeuroAIDS-related white matter abnormalities using polarized light microscopy and histological staining in relation to Magnetic Resonance Imaging (MRI)," *Program no. 782.01. Neuroscience Meeting Planner. 41st Annual Meeting of the Society for Neuroscience*. (Washington, DC). [Online].

Crick, F., and Jones, T. (1993). Backwardness of human neuroanatomy. *Nature* 361, 109–110.

Dejerine, J. (1895). *Anatomie des centres nerveux, Vol 1*. Paris: Rueff et CIE.

De Luca, M., Beckmann, C. F., de Stefano, N., Matthews, P. M., and Smith, S. M. (2006). fMRI resting state networks define distinct modes of long-distance interactions in the human brain. *Neuroimage* 29, 1359–1367.

Fischl, B., Destrieux, C., Halgren, E., Segonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. M. (2004). Automatic parcellation of the human cerebral cortex. *Cereb. Cortex* 14, 721–730.

Fischl, B., Salat, D., Albert, M., van der Kouwe, A., Killiany, R., Kennedy, D., Montillo, A., Makris, N., Rosen, B., and Dale, A. M. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.

Flechsig, P. (1901). Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. *Lancet* 2, 1027–1029.

Gallyas, F. (1979). Silver staining of myelin by means of physical development. *Neurol. Res.* 1, 203–209.

Johnson, G. A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., and Nissanov, J. (2010). Waxholm space: an image-based reference for coordinating mouse brain research. *Neuroimage* 53, 365–372.

Kretzschmar, H. (2009). Brain banking: opportunities, challenges and meaning for the future. *Nat. Rev. Neurosci.* 10, 70–77.

Leergaard, T. B., White, N. S., de Crespigny, A., Bolstad, I., D'Arceuil, H., Bjaalie, J. G., and Dale, A. M. (2010). Quantitative histological validation of diffusion MRI fiber orientation distributions in the rat brain. *PloS One* 5:8595. doi: 10.1371/journal.pone.0008595

Ludwig, E., and Klinger, J. (1956). *Atlas Cerebris Humani: The Inner Structure of the* Brain. Basel, Switzerland: Little, Brown and Company.

Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., Jenkinson, M., Laumann, T., Curtiss, S. W., and van Essen, D. C. (2011). Informatics and data mining tools and strategies for the human connectome project. *Front. Neuroinform.* 5:4. doi: 10.3389/fninf.2011.00004

Mikula, S., Trotts, I., Stone, J. M., and Jones, E. G. (2007). Internet-enabled high-resolution brain mapping and virtual microscopy. *Neuroimage* 35, 9–15.

Oishi, K., Faria, A. V., van Zijl, P. C. M., and Mori, S. (2011). MRI Atlas

of Human White Matter, 2nd Edn. London, UK: Academic Press.

Penfield, W. (1958). *The Excitable Cortex in Conscious Man*. Liverpool: University Press.

Pierpaoli, C., Jezzard, P., Basser, P. J., Barnett, A., and Di Chiro, G. (1996). Diffusion tensor MR imaging of the human brain. *Radiology* 201, 637–648.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1:e42. doi: 10.1371/journal.pcbi.0010042

Weinstein, R., Graham, A., Richter, L., Barker, G., Krupinski, E., Lopez, A. M., Erps, K. A., Bhattacharyya, A. K., Yagi, Y., and Gilbertson, J. R. (2009). Overview of telepathology, virtual microscopy, and whole slide imaging: prospects for the future. *Hum. Pathol.* 40, 1057–1069.

Yakovlev, P. I., and Lecours, A. R. (1967). "The myelogenetic cycles of regional maturation of the brain," in: *Regional development of the Brain in Early Life*, ed A. Minkowski (Oxford, Blackwell Scientific Publications), 3–65.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 April 2011; accepted: 31 March 2012; published online: 24 April 2012.

Citation: Annese J (2012) The importance of combining MRI and large-scale digital histology in neuroimaging studies of brain connectivity and disease. Front. Neuroinform. ***6****:13. doi: 10.3389/fninf.2012.00013*

frontiers in
NEUROINFORMATICS

METHODS ARTICLE
published: 22 November 2011
doi: 10.3389/fninf.2011.00029

Multiscale exploration of mouse brain microstructures using the knife-edge scanning microscope brain atlas

Ji Ryang Chung[1†], Chul Sung[1†], David Mayerich[2], Jaerock Kwon[3], Daniel E. Miller[1], Todd Huffman[4], John Keyser[1], Louise C. Abbott[5] and Yoonsuck Choe[1]*

[1] *Department of Computer Science and Engineering, Texas A&M University, College Station, TX, USA*
[2] *Beckman Institute of Advanced Science and Technology, University of Illinois, Urbana-Champaign, IL, USA*
[3] *Department of Electrical and Computer Engineering, Kettering University, Flint, MI, USA*
[4] *Research and Development, 3Scan, Bryan, TX, USA*
[5] *Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA*

Edited by:
Olaf Sporns, Indiana University, USA

Reviewed by:
Claus Hilgetag, Jacobs University Bremen, Germany
Tao Ju, Washington University in St. Louis, USA

***Correspondence:**
Yoonsuck Choe, Department of Computer Science and Engineering, Texas A&M University, 3112 TAMU, College Station, TX 77843-3112, USA.
e-mail: choe@tamu.edu

[†] Ji Ryang Chung and Chul Sung have contributed equally to this work.

Connectomics is the study of the full connection matrix of the brain. Recent advances in high-throughput, high-resolution 3D microscopy methods have enabled the imaging of whole small animal brains at a sub-micrometer resolution, potentially opening the road to full-blown connectomics research. One of the first such instruments to achieve whole-brain-scale imaging at sub-micrometer resolution is the Knife-Edge Scanning Microscope (KESM). KESM whole-brain data sets now include Golgi (neuronal circuits), Nissl (soma distribution), and India ink (vascular networks). KESM data can contribute greatly to connectomics research, since they fill the gap between lower resolution, large volume imaging methods (such as diffusion MRI) and higher resolution, small volume methods (e.g., serial sectioning electron microscopy). Furthermore, KESM data are by their nature multiscale, ranging from the subcellular to the whole organ scale. Due to this, visualization alone is a huge challenge, before we even start worrying about quantitative connectivity analysis. To solve this issue, we developed a web-based neuroinformatics framework for efficient visualization and analysis of the multiscale KESM data sets. In this paper, we will first provide an overview of KESM, then discuss in detail the KESM data sets and the web-based neuroinformatics framework, which is called the KESM brain atlas (KESMBA). Finally, we will discuss the relevance of the KESMBA to connectomics research, and identify challenges and future directions.

Keywords: mouse brain; Golgi; web-based brain atlas; multiscale; connectomics; Knife-Edge Scanning Microscopy

1. INTRODUCTION

Connectomics aims to map the full connection matrix of the brain (Sporns et al., 2005; Sporns, 2011). The fundamental assumption in connectomics is that structure defines function. To evaluate this assumption, we can consider the fact that the functional evolution of the brain has been mainly driven by that of the brain architecture and not by individual neurons (Swanson, 2003). Also, "basic circuits" of the brain have been identified as an important abstraction of brain function at the system-level (Shepherd, 2003). Furthermore, structure (connectivity) has been shown to greatly affect the dynamics of the network (Sporns and Tononi, 2002). Varying the delay distribution in a network was also found to alter its dynamics (Thiel et al., 2003), where structural analogs of delay, e.g., connection length, could also contribute to the same effect. These, taken together, indicate that obtaining the connectome can lead to a major breakthrough in understanding brain function.

Recent advances in high-throughput, high-resolution 3D microscopy methods have enabled the imaging of whole small animal brains at a sub-micrometer resolution, potentially opening the road to full-blown connectomics research. One of the first such instruments to achieve whole-brain-scale imaging at sub-micrometer resolution is the Knife-Edge Scanning Microscope (KESM; McCormick, 2003, 2004; Kwon et al., 2008; Mayerich et al., 2008b; cf. Li et al., 2010 based on the same imaging principles as that of the KESM). KESM whole-brain data sets now include Golgi (neuronal circuits; Abbott, 2008), Nissl (soma distribution; Choe et al., 2010), and India ink (vascular networks; Choe et al., 2009; Mayerich et al., 2011b). Methods related to the KESM include All-Optical Histology (Tsai et al., 2003) and Array Tomography (Micheva and Smith, 2007). There are also methods that explore much finer structural detail, such as Serial Block-Face Scanning Electron Microscopy (SBF-SEM; Denk and Horstmann, 2004), and Automatic Tape-Collecting Lathe Ultramicrotome (ATLUM; Hayworth, 2008). The resolution and size of the volume that can be imaged by the above methods vary widely (resolution of 10s of nm to 100s of nm, to volumes ranging from 10s of μm cube up to 1 cm cube; see Choe et al., 2008 for a review), but they all share the same principle of physical sectioning or physical ablation, as opposed to optical sectioning common in conventional 3D microscopy (All-Optical Histology uses a hybrid approach, physical plus optical sectioning; Tsai et al., 2003).

Data from KESM and similar approaches based on light microscopy can greatly contribute to connectomics research, by filling the critical gap between large scale, lower resolution methods like diffusion MRI (Basser and Jones, 2002; Tuch et al., 2003; Hagmann et al., 2007; Roebroeck et al., 2008) on the one hand and small-scale, higher resolution methods like SBF-SEM (Denk and Horstmann, 2004) on the other hand. It is also notable that KESM data are by nature multiscale, ranging from the subcellular ($<1\,\mu$m) to the whole organ scale ($\sim$1 cm). Due to the large volume (several tera voxels) and the multiscale nature, visualization alone is a huge challenge, before we even start worrying about quantitative connectivity analysis. Furthermore, delivering the neuronal circuit data to connectomics researchers is also a challenge, due to the same reasons as above. To solve this issue, we developed a web-based neuroinformatics platform for efficient visualization and analysis of the multiscale KESM data sets.

In this paper, we will first provide an overview of KESM, then discuss in detail the web-based neuroinformatics framework called the KESM brain atlas (KESMBA). Next, we will present the KESM data sets using KESMBA. Finally, we will discuss the relevance of the KESMBA to connectomics research, and identify challenges and future directions.

2. MATERIALS AND METHODS

2.1. SPECIMEN PREPARATION

Two mouse brains were imaged in their entirety, after being stained by Golgi-Cox for the visualization of neuronal morphology. C57BL/6J mice were anesthetized with isoflurane inhalant anesthesia. Each mouse was decapitated, the brain removed and immersed in Golgi-Cox solution that contained 1% potassium chromate, 1% potassium dichromate, and 1% mercuric chloride in distilled water. The brain was then soaked for 10–16 weeks in the dark and then washed in distilled water overnight. Additionally, it was immersed in a 5% ammonium hydroxide solution in distilled water for about one week in the dark and then again washed in distilled water for 4 h. After that, the brain was dehydrated through a graded series of ethanols starting with 50% ethanol in water and increasing to 100% ethanol over a time period of 6 weeks. Finally, it was embedded in araldite following a standard protocol (Abbott and Sotelo, 2000), with the exception that each step needed to infiltrate the brains with araldite took 24 h. KESM sectioning requires that whole-brains be completely dehydrated and infiltrated with araldite plastic. Normal plastic embedding is typically carried out on much smaller pieces of tissue, so we have modified the processing steps to allow us to embed whole mouse brains that we can cut using the KESM.

2.2. IMAGING WITH THE KNIFE-EDGE SCANNING MICROSCOPE

We used the KESM for sectioning and imaging (Mayerich et al., 2008b) two prepared mouse brains (both Golgi). **Figure 1** shows a photo of the KESM with its major components.

Each stained mouse brain, embedded in a plastic block, was mounted on the three-axis precision stage. The diamond knife-collimator assembly was used to cut sequential 1.0 μm-thick sections from the tissue blocks, while providing transmission illumination. (Note that the KESM design supports illumination through the objective (McCormick, 2004) and the original implementation already includes this design, especially for fluorescence imaging.) The light passed through the diamond knife and penetrated the tissue sections for imaging. The brain tissues stained in Golgi were imaged with a Nikon Fluor 10$\times$ objective (NA = 0.3). The actual image digitizing was performed by a DALSA CT-F3 high-sensitivity line-scan camera capturing the transmitted light, and these images were stored in the designated storage. In order to automatically control the stage movement and data acquisition process, we developed in-house control software (Kwon et al., 2008). Noise due to the knife-edge misalignment, defects in the knife blade, and knife chatter were removed through image processing algorithms including light normalization (Mayerich et al., 2007). The KESM controller also employed a stair-step cutting algorithm to minimize damage to tissue between neighboring columns (Kwon et al., 2008). After preliminary image processing for noise and distortion removal, TIFF formatted raw image files were compressed into high quality JPEG format and stored for further processing, while the original TIFF images were kept for archival purposes. We imaged horizontal sections from the brain.

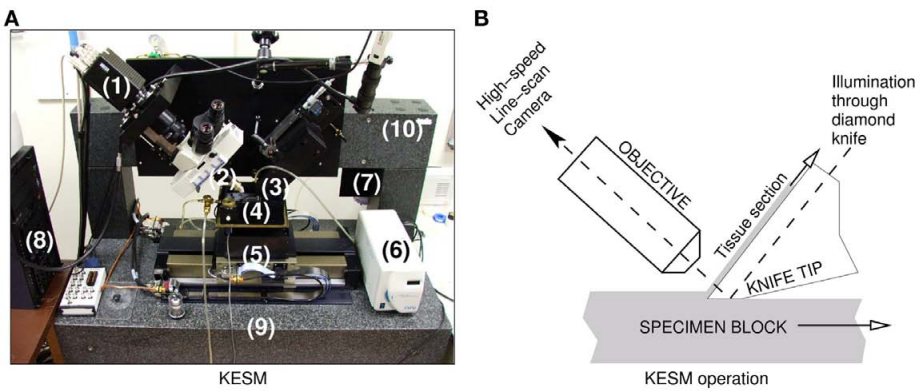


FIGURE 1 | The knife-edge scanning microscope and its operation. **(A)** The Knife-Edge Scanning Microscope and its main components are shown: (1) high-speed line-scan camera, (2) microscope objective, (3) diamond knife assembly and light collimator, (4) specimen tank (for water immersion imaging), (5) three-axis precision air-bearing stage, (6) white-light microscope illuminator, (7) water pump (in the back) for the removal of sectioned tissue, (8) PC server for stage control and image acquisition, (9) granite base, and (10) granite bridge. **(B)** The imaging principle of the KESM is shown.

2.3. THE KESM BRAIN ATLAS

The KESM brain atlas (KESMBA) framework has been designed and implemented to allow the widest dissemination of KESM mouse brain circuit data and to enable easy visual and quantitative analysis. For this, we had several design requirements, that the atlas is (1) not dependent on high-end computer hardware (e.g., expensive graphics cards), (2) not dependent on custom 3D viewing applications or plug-ins, and (3) browsable within any standard web browser.

2.3.1. Basic idea: transparent overlay with distance attenuation

The basic idea we use to meet the requirements listed above is transparent overlay of images with distance attenuation (Eng and Choe, 2008). **Figure 2** shows the concept. Overlaying an image stack containing two intertwining objects (**Figure 2A**) to get minimum intensity projection (**Figure 2B**) results in the loss of 3D information. Interleaving each image with semi-opaque blank images brings out the 3D information (**Figure 2C**). This is similar to the artistic use of haze to achieve depth effect in a 2D medium (cf. Kersten et al., 2006). In practice, raw images containing data already have semi-opaqueness in the background once made transparent, so simply overlaying them results in the same kind of effect. This simple approach, when combined with a Google Maps™-like zoomable web interface, results in a powerful browsing environment for large 3D brain data. In fact, we customized and extended the Google Maps API (version 2) to construct the KSEMBA.

2.3.2. Image processing and adding transparency

After the raw image files were acquired using the KESM, three additional image processing steps were performed to enhance the image quality suitable for the web atlas. First, to enhance visibility when overlayed, we inverted the original images with black foreground and white background to have white foreground and black background. Next, because the inverted images do not have enough luminance contrast, we performed Gamma correction with a sigmoidal non-linearity to expand the luminance contrast between foreground and background pixels within each image. Finally, we turned the background color of the image to be transparent, for layering of the images to achieve a 3D view within a web browser. The pixels were made transparent according to their gray-level value. The processed images were stored in PNG format which supports alpha channel transparency. The contrast factor and contrast center values (25 and 50, respectively) used in the gray-level transparency process were empirically selected.

2.3.3. Multiscale tile generation

Once the image processing is done, pyramidal tiles are generated. Each tile in Google Maps consists of 256×256 pixels. The pyramidal structure of the Google Maps tiles in different zoom levels is shown in **Figure 3A**. Our Golgi data sets have 8–10 columns, where each column consists of a tall stack of $2{,}400 \times 12{,}000$-pixel images. Below, we will consider the case with 8 columns. Calculation for the 10 column case can be done in a similar manner.

With the KESM image stack, we prepared tiles for 6 different zoom levels compatible with the Google Maps API's zoom level from 2 to 7. The number of tiles required at zoom level z is $2^z \times 2^z$. Therefore, the Golgi data set requires $\sum_{z=2}^{7} 2^z \times 2^z = 21{,}840$ tiles for each section, and 121,692,480 tiles for all 5,572 sections, theoretically. Fortunately, the actual number of tiles we created is 4,892 per section and 27,258,224 overall because each image section is not square-shaped and we only had to create tiles containing tissue data. **Figure 3B** shows an example of the tiles we created for the highest zoom level of 7. In the example, we had to create only 8 tiles out of 16 possible ones because the other 8 tiles were empty. Preparing a tile pyramid requires extra storage, time, and effort. Assuming that the above mentioned PNG transform did not increase the file size, in our Golgi data set, the tile pyramid ($4{,}892 \times 256 \times 256 \times 5{,}572 = 320{,}002{,}112$) causes about 39% increase in disk space usage (without tiles, the total size is $19{,}200 \times 12{,}000 \times 5{,}572 = 230{,}400{,}000$). However, once they are generated, they contribute to saving image download time for the currently viewed portion of the atlas. For example, the KESMBA has a map area of 80% browser window width $\times$ 600 pixel window

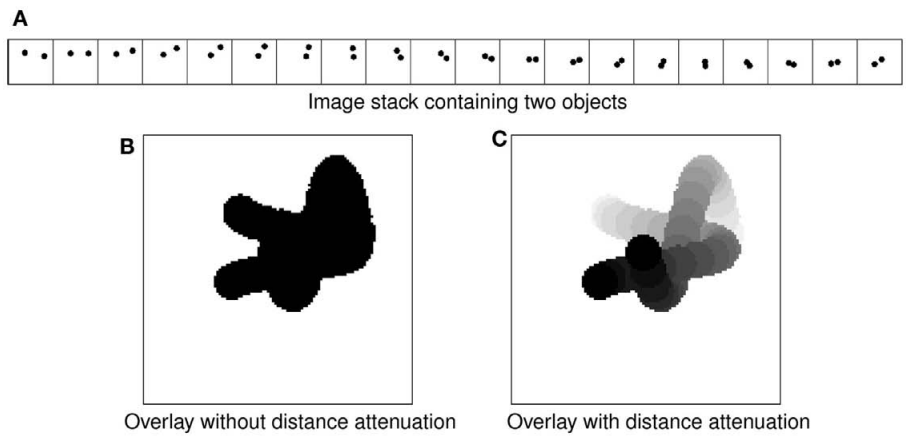


FIGURE 2 | Transparent overlay with distance attenuation. (A) An image stack containing two intertwined objects are shown. **(B)** Simple overlay of the image stack in **(A)** results in loss of 3D perspective. **(C)** Overlay with distance attenuation helps bring out the 3D cue.

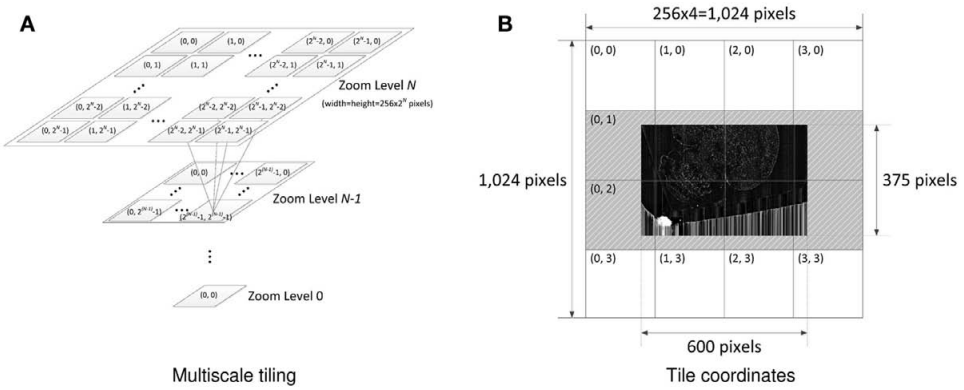


FIGURE 3 | Multiscale tiling. (A) Tile pyramid compatible with GoogleMaps. Quad-tree pyramid of tiles and the x-y coordinate indexing convention are depicted. Each tile has 256 × 256 pixels. Zoom level ranges from 0 to 7, and zoom level N has $2^N \times 2^N$ tiles. Following this tiling convention automatically enables various map functions including zoom-in/out. **(B)** Example of actual tiles. The maximum zoom level of the Golgi image section (19,200 × 12,000 pixels) is 7 [=argmin_x ($256 \times 2^x \geq \max(19200,12000)$)]. The minimum zoom level is set to 2. This particular example shows how the original image is tiled and how the tiles are named at zoom level 2. The dark image in the middle is the image section halved down 5 times (600 × 375 pixels) to fit in zoom level 2 (1,024 × 1,024 pixels). After putting the downsized image at the center, transparent image patches (gray dashed area) are added to fill the incomplete tiles so that every tile can have the same 256 × 256 pixels size. Because there is no need to generate empty tiles, only 8 tiles are created out of 16 possible ones.

height specified in a Cascading Style Sheet (CSS). Given typical screen resolution, the number of 256 × 256 sized tiles concurrently displayed on a client web browser will be ∼30 at the maximum, which is less than 1% of the original image section at the maximum zoom level (19,200 × 12,000). Once all the tiles are generated, each tile is named consistent with Google Maps tile specification. For example, a tile name 1_2_3.png denotes zoom level = 1, x-coordinate = 2, and y-coordinate = 3.

2.3.4. Web atlas based on Google maps API

To enable 3D visualization, we customized the Google Maps JavaScript API. Google Maps JavaScript API provides extensive functions required for a geographical atlas. In addition to the essential navigational functions of zooming and panning, Google Maps JavaScript API offers useful features such as zoom scale bar, double-click zoom-in, and overlaying various objects including images, text, markers, and polygons. Google Maps provides an extensive API specification and there exist a large number of private developers seeking and sharing solutions for customizing the API.

2.3.4.1. Customizations. Existing API functions were customized to fit the purpose of the KESMBA. This customization included: using custom tiles; tile overlays; user options to select the number and interval of the tile overlays; overlaying zoomable annotation; and map redraw function. The API was further extended to include: information panel; scale bar; map capture button; and *z*-axis navigation controller.

We generated a custom map type instance of the "GMapType" class to call the map tiles from the Google database to feed in the custom tiles we generated. Multiple tiles from subsequent image sections are overlaid to create a 3D effect. The summary code below provides an overview of how this is achieved using the Google Maps API.

```
////////////////////
// File: overlay.js
////////////////////

...

// 1. Create a tile layer.
customLayer=[new GTileLayer(...)];

// 2. Generate a custom tile URL.
customLayer.getTileUrl=customGetTileUrl;

// 3. Create an overlay instance of
//    the GTileLayerOverlay class.
customOverlay=new GTileLayerOverlay
(customLayer);

// 4. Create a map type instance of the
//    GMapType class.
customMap=new GMapType(...);

// 5. Create a map instance of the GMap2
//    class.
var myMap=new GMap2(document.getElementById
("map"),
     {mapTypes:[customMap]});
// 6. Add predefined custom map type into
//    the map instance.
myMap.addMapType(customMap);

// 7. Add a custom tile layer into the map
//    instance.
cMap.addOverlay(overlays);
...
```

We allow users to select the number of tiles to overlay and the interval between two overlays, so that they can freely generate the 3D view that they prefer. On top of the image tile overlays, we added another optional overlay for text annotation. This annotation can change at different zoom levels, so that it can show more global description at a distant view and detailed description at close-up. Finally, we added a "redraw" function so that when a user changes any of the above options (overlay size, overlay interval, and annotation on/off) and the map needs to be redrawn, it does not refresh the entire page but redraws only the map area. This is achieved by using the JavaScript "arguments.callee" property, with which an executing function can recursively refer to itself. When "loadFcn" is called to redraw the map, it re-invokes the original "load" function. Redrawing the map in this way does not have to resummon the entire page, and therefore is faster.

2.3.4.2. Extensions. Since the Google Maps JavaScript API is created solely for 2D geographical maps, some features necessary for the 3D brain atlas are missing and thus are not customizable. We introduced a horizontal menu bar on top of the map to include the functions that are necessary for the KESMBA. Most of the new features are achieved by using various properties of the Document Object Model (DOM). In the menu, we added a z-axis navigation function with a drop-down menu (depth navigation step size) and buttons ("+" and "−" for moving in/out). Also, users can choose whether to display the annotation layer by clicking on a checkbox. To facilitate capturing the current view on the atlas, we added an image capture button. When the button is clicked, it opens the print.html file using the windows.open method. In print.html, it gets the map area information of index.html by the "window.opener" property. Then, it copies whatever is on the map area of index.html to generate the page content of print.html using the "innerHTML" property.

Scale bar is one of the uncustomizable features of Google Maps. Therefore, we created one and attached it onto the map. We first created a ⟨div⟩ object using "createElement" method. Then, it was appended to the map div (⟨div id = "mapArea"⟩) by using "getContainer" and "appendChild" methods. To make KESMBA more informative, we created a panel to display the information of the current view. In the panel to the right, the KESMBA displays the information about the specimen, stain type, current plane of view, dimension of the image section, and the z-range of the layers in the current view. An area to display the above information dynamically is first encapsulated by ⟨span id = "xx"⟩. . .⟨/span⟩ tags, and its contents are updated using the "firstChild" and "data" properties. This way, contents of the information panel are automatically updated as the user navigates or switches between the atlases using the top menu bar. **Figure 4** shows the interface of the KESMBA containing all the above mentioned features.

3. RESULTS

In this section, we will present our two KESM Golgi data sets and results from applying the KESMBA framework to these data sets.

3.1. KESM GOLGI DATA SETS

The first Golgi brain was sectioned and imaged in 2008 (from July 7 to August 8, 2008). These results were first reported in Abbott (2008). The first Golgi data set did not include the left frontal lobe, part of the left temporal lobe, and part of the right frontal lobe due to a misconfigured frame buffer that truncated the images, although the entire brain was sectioned using the KESM. The second Golgi brain was sectioned and imaged in 2010 (from June 8 to August 4, 2010). The second data set contained the entire brain. The first Golgi data set, although partly incomplete, includes less noise than the second Golgi data set, so we decided to make available both data sets within the KESMBA framework. These results are shown in **Figures 5–7**. All data sets had a voxel resolution of $0.6\,\mu m \times 0.7\,\mu m \times 1.0\,\mu m$, so at maximum zoom, the data are quite detailed, as shown in **Figure 8**.

3.2. 3D RENDERING THROUGH IMAGE OVERLAYS

All results shown in **Figures 5–8** were from direct screenshots of the KESMBA. The 3D effect is most notable in **Figure 8**. To highlight the z-axis resolution of the KESM data sets, and to show the effectiveness of our overlaying technique, we prepared views of a fixed region in the KESM data set by varying the number of overlays (**Figures 9A–C**). As we can see from this figure, overlays are effective in rendering 3D content, all within a standard web browser without any dedicated plug-in. Another technique that we implemented that is especially helpful when viewing with a larger field of view (i.e., zoomed out) is to overlay images at a certain interval. For example, overlaying 20 images at an interval of 5 would visualize a 100-μm-thick volume (compare **Figures 9D,E**).

3.3. MULTISCALE NATURE OF THE KESM DATA

One of the main advantages of the KESMBA is that it is very easy to navigate through the data, both within a certain scale and across multiple scales. In fact, this capability assisted greatly in producing the figures in the very article. Here, we will present the multiscale nature of the KESM data and show the effectiveness of the KESMBA framework in handling such multiscale data. In **Figure 10**, we show successive snapshots of the KESMBA while zooming from the largest scale to the smallest scale. Each step of zooming in doubles the resolution, so the final panel has 32× higher resolution than the first panel.

3.4. NEURONAL CIRCUITS: LOCAL AND GLOBAL

Finally, we examine the relevance of the KESM data sets to connectomics research. Although it is true that with Golgi-Cox only ∼1% of the entire population of neurons are stained and thin myelinated axons are not stained reliably, we can still gain valuable insights from this whole-organ level data at a microscopic resolution.

KESM Golgi data sets can help advance connectomics research in two ways, (1) locally and (2) globally. At the local scale, we can investigate the basic circuits (Shepherd, 2003). Although exact connectivity cannot be established, the repeating pattern can help us refine our basic circuit model, and also use the data to validate synthetic circuits constructed based on a theoretical generative model (see, e.g., van Pelt and Uylings, 2005; Koene et al., 2009). Having access to these basic circuits from all regions in the brain is also a great benefit, as shown in **Figure 11**. This figure shows neurons from the cerebellum, inferior colliculus, thalamus, and hippocampus.

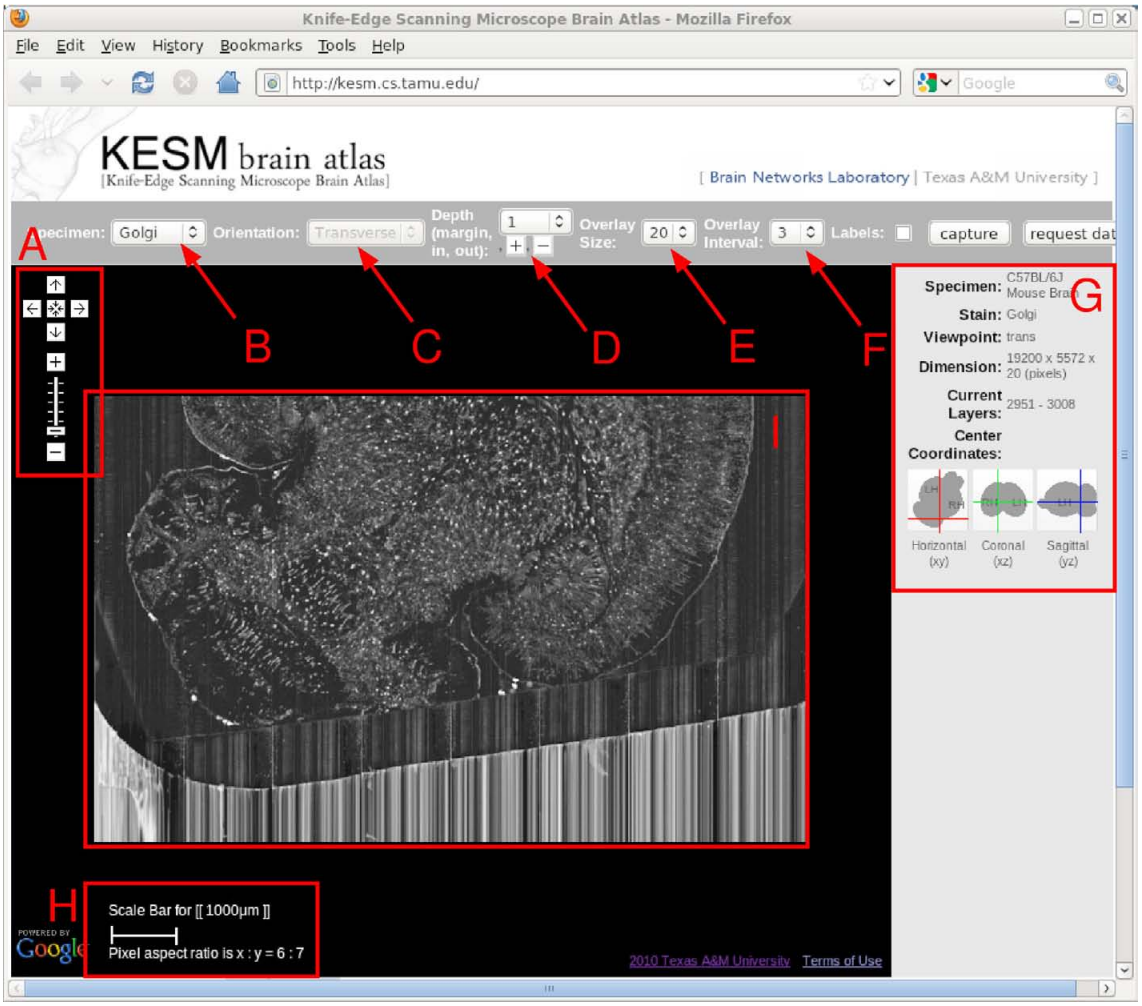


FIGURE 4 | KESM brain atlas interface. A screenshot of the KESM brain atlas running in a web browser is shown. Red markers and text were added on top for the purpose of explanation, below. **(A)** Navigation panel: panning and zoom-in/zoom-out. **(B)** Data set selection. Golgi, Golgi2, India Ink are available in the pull-down menu. **(C)** Sectioning plane orientation. Three standard planes supported (planned). **(D)** Depth navigation. Amount of movement (unit = 1 μm) in the z direction and forward (deeper, [+]) or backward (shallower, [−]) can be controlled. **(E)** Overlay count. How many images to overlay can be selected here. **(F)** Overlay interval. For high zoom-out levels, overlaying every *n* images is enough, and this helps visualize thicker sections. **(G)** Information window containing specimen meta data and current location information. **(H)** Scale bar that automatically adjust to the given zoom level. **(I)** Main display. Note that the Google logo on the bottom left is shown due to the use of the Google Maps API, and it by no means indicate any connection between the KESM data and Google.

At the global scale, certain fiber tracts show up prominently in the KESM Golgi data. For example, various commissures in the frontal lobe and dense fiber bundles in the striatum are prominently visible (**Figure 12**). Similar fiber tracts can easily be identified, such as the hippocampal commissure in the posterior part of the brain.

3.5. DOWNLOAD PERFORMANCE

The above results confirm the effectiveness of the KESMBA's pseudo 3D view method using image overlays. However, the additional image overlays mean longer download time, and it will have limited utility if the download time exceeds waiting time tolerable for the users. **Figure 13** shows the result of download time analysis of the KESMBA. Download time and download data size were measured in two modern web browsers (Internet Explorer 8.0.6 and Mozilla Firefox 3.6.8) using HttpWatch 7.0.26, a browser plug-in to monitor http traffic. Expectedly, the download time and data size were proportional to the number of overlays. Notably, Firefox took extraordinarily long with large variance, while downloading 20 overlays. The Intranet and the Internet download times for 20 overlays reached above 22 and 44 s respectively with Microsoft Internet Explorer, and 53 and 52 s respectively with Mozilla Firefox. Literature on the tolerable waiting time for a web page download presents discordant thresholds between 4 and 41 s (Selvidge et al., 2002; Galletta et al., 2004), but none of them used a web page with as much graphical content as the KESMBA. Considering the unique graphics-rich nature of the KESMBA,

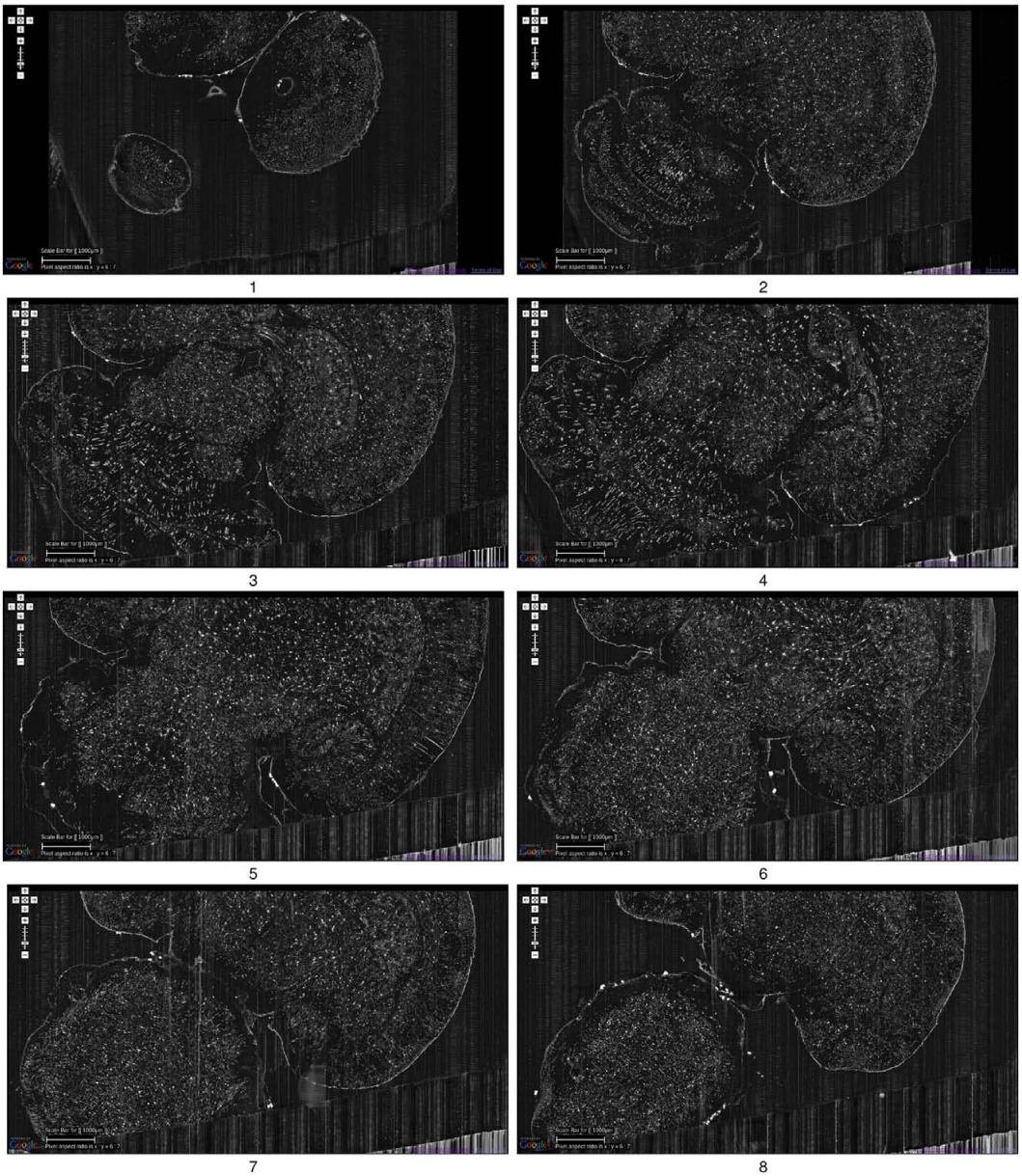


FIGURE 5 | Golgi data set 1. A fly-through of the Golgi data set 1 is shown. The data were obtained by sectioning in the horizontal plane (upper right corner: anterior, lower left corner: posterior). This is the full extent of the data that was captured. We can see that part of the left temporal lobe, left frontal lobe, and part of the right frontal lobe are cut off. Scale bar = 1 mm. Each image is an overlay of 20 images in the z direction. The z-interval between each panel is 600 μm. The numbers below the panels show the ordering. These are cropped screenshots from the KESMBA. This data set, obtained in 2008, is the first whole-brain-scale data set of the mouse at sub-micrometer resolution.

we believe the above download times are within the tolerable threshold.

4. DISCUSSION

This article presents one of the first whole-brain-scale mouse brain atlases imaged at a sub-micrometer resolution, and a novel neuroinformatics framework for rapid visualization and exploration of the massive data sets. The main value of this kind of resource is that it fills the gap between (1) the lower resolution (100s of μms), system-level (10s of cm), diffusion MRI-based tractography data and (2) the higher resolution (10s of nm), small volume (10s of μm), EM-based synaptome data. Both local and global circuit data from our KESM brain atlas are expected to contribute greatly to connectomics research. In the following, we will discuss existing brain atlas and neuronal morphology resources and draw a comparison with the KESM brain atlas, and consider potential challenges and initial solutions.

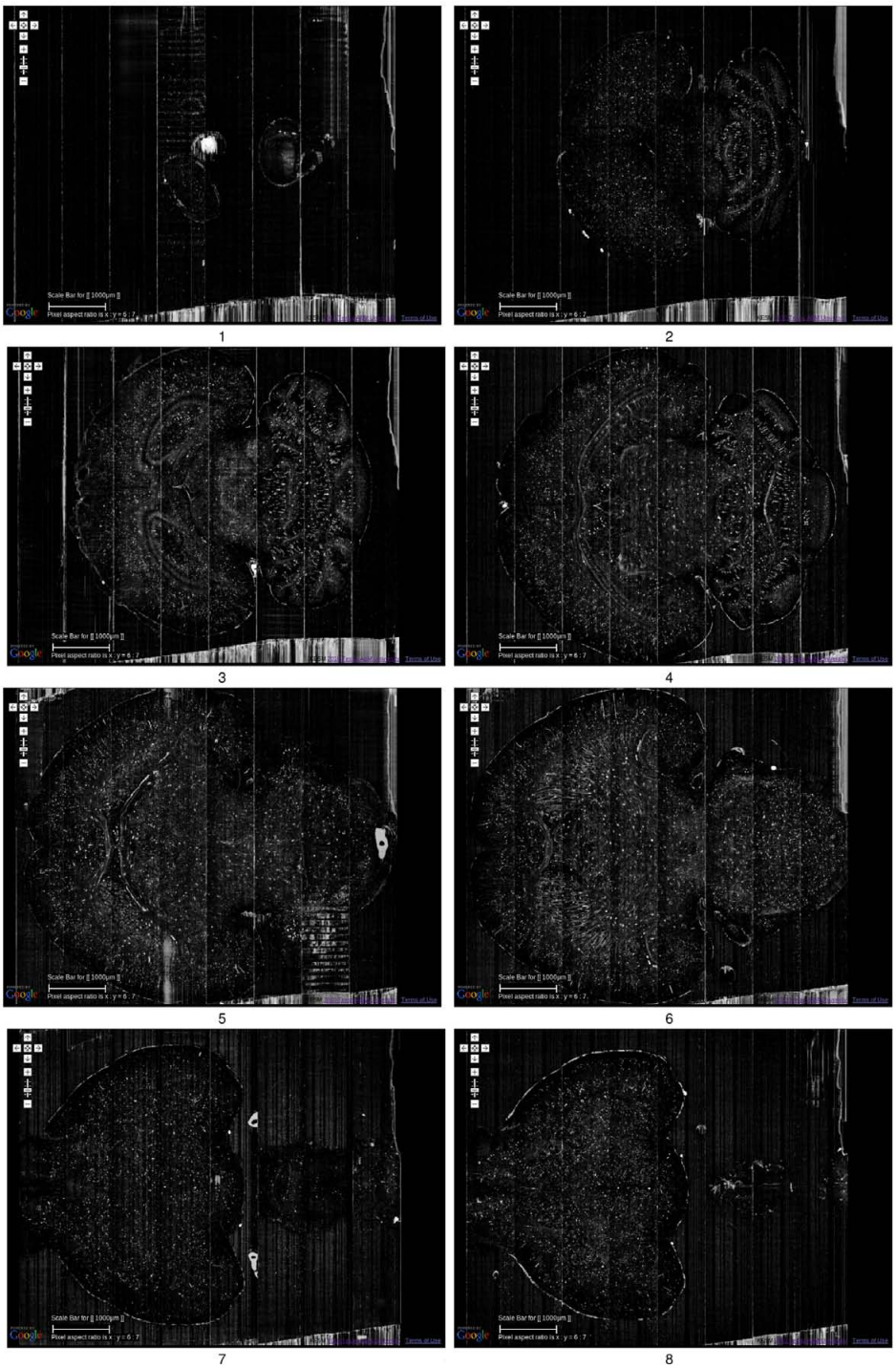


FIGURE 6 | Golgi data set 2. A fly-through of the Golgi data set 2 is shown. The data were obtained by sectioning in the horizontal plane (left: anterior, right: posterior). Scale bar = 1 mm. Each image is an overlay of 20 images in the z direction. The z-interval between each panel is 800 μm, except for the last where it was 200 μm (so that data from near the bottom of the data stack can be shown: otherwise it will overshoot into regions with no data). The numbers below the panels show the ordering. See Movie S1 in Supplementary Material for a fly-through of this data set.

4.1. BRAIN MAPS AND ATLASES

The 3D mouse brain atlas, at a typical macroscale spatial resolution of 10 μm, is an indispensable guide to navigation within the mouse brain (Paxinos and Franklin, 2001; Paxinos and Watson, 2006). Without it, the mouse brain microstructure, viewed as a database of individual neurons, is virtually unintelligible. The

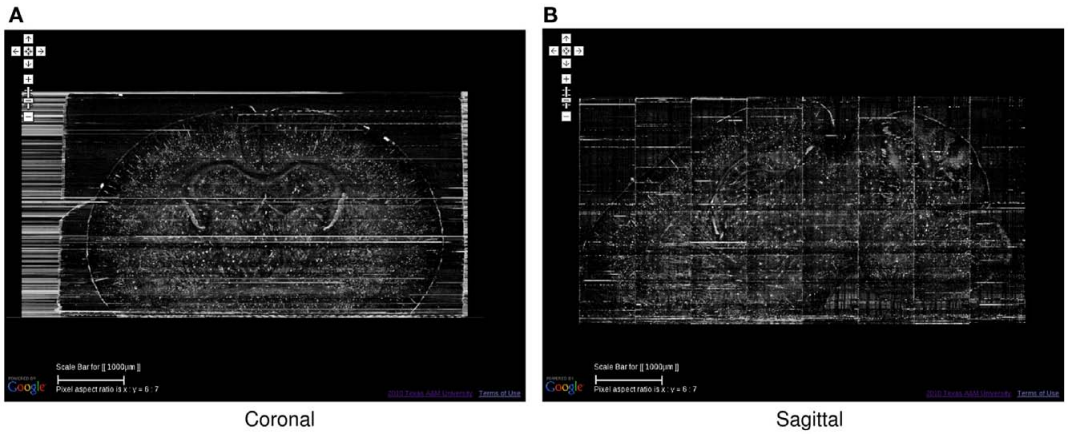


FIGURE 7 | Golgi data set 2, Coronal and Sagittal Views. The coronal **(A)** and sagittal **(B)** views of the data set in **Figure 6** are shown. Scale bar = 1 mm. These views show the superior *z*-axis resolution of the KESM data sets.

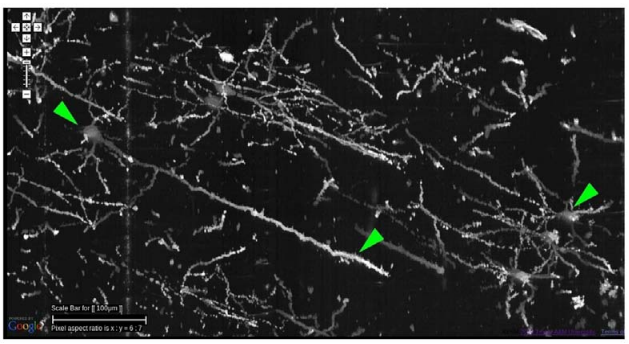


FIGURE 8 | Details from Golgi data set 1. Details from the Golgi data set 1 are shown at full resolution. This panel shows an overlay of 20 images, thus it is showing a 20-μm-thick volume. Scale bar = 100 μm. The arrow heads, from left to right, point to (1) the soma of a pyramidal cell in the cortex and (2) its apical dendrite, and (3) a couple of spiny stellate cells. Other pyramidal cells and stellate cells can be seen in the background. At this resolution, we can see dendritic spines as well.

focus of activity for standardizing anatomical structures and ontology for the mouse, like those defined in the mouse and rat atlases produced by Paxinos and Franklin (2001) and Paxinos and Watson (2006), are at this macrostructure level.

4.1.1. The Mouse Atlas Project (MAP)

MacKenzie-Graham et al. (2003) developed a probabilistic atlas of the adult and developing C57BL/6J mouse. MAP consists of not only data from Magnetic Resonance Microscopy (MRM) and histological atlases, but also a suite of tools for image processing, volume registration, volume browsing, and annotation. MAP will produce an imaging framework to house and correlate gene expression with anatomic and molecular information drawn from traditional and novel imaging technologies. This digital atlas of the C57BL/6J mouse brain is composed of volumes of data acquired from μMRI, block-face imaging, histology, and immunohistochemistry. MAP technology provides the infrastructure for the development of the Allen Brain Atlas (MacKenzie-Graham et al., 2003; see below). Also see the related Mouse BIRN (Biomedical Informatics Research Network).

4.1.2. The Allen Brain Atlas

The Allen Brain Atlas contains detailed gene expression maps for ∼20,000 genes in the C57BL/6J mouse (Lein et al., 2007). A semi-automated procedure was used to conduct *in situ* hybridization and data acquisition on 25 μm-thick sections (*z*-axis) of the mouse brain. The *x*-*y*-axis resolution of the images range from 0.95 to 8 μm. The Allen Brain Atlas is the first comprehensive gene expression map at the whole-brain level, and is currently accessed over 4 million times per month, with over 250 scientists browsing the data on a daily basis.

4.1.3. The Mouse Brain Library (MBL)

MBL is developing methods to construct atlases from celloidin-embedded tissue to guide registration of MBL data into a standard coordinate system, by segmenting each brain in its collection into 1,200 standard anatomical structures at a resolution of 36 μm (Rosen et al., 2000). Algorithms are to be designed to segment each brain in the MBL into a set of standard anatomical structures like those defined in the rat atlas produced by Computer Vision Laboratory for Vertebrate Brain Mapping at Drexel College of Medicine, whose computerized 3D atlas was built from stained sections for the mouse brain that reconstructs Nissl-stained sectional material, a 17.9-μm isotropic 3D data set, from a freshly frozen brain of an adult male C57BL/6J mouse.

4.1.4. BrainMaps.org

BrainMaps.org is an internet-enabled, high-resolution brain map (Mikula et al., 2007). The map contains over 10 million mega pixels (35 terabytes) of scanned data, at a typical resolution of ∼0.46 μm/pixel (in the *x*-*y* plane). The atlas provides an intuitive web-based interface for easy and band-width-efficient navigation, through the use of a series of subsampled (zoomed out) views of the data sets, similar to the Google Maps interface. Even though the *x*-*y* plane resolution is below 1 μm, the *z*-axis resolution is

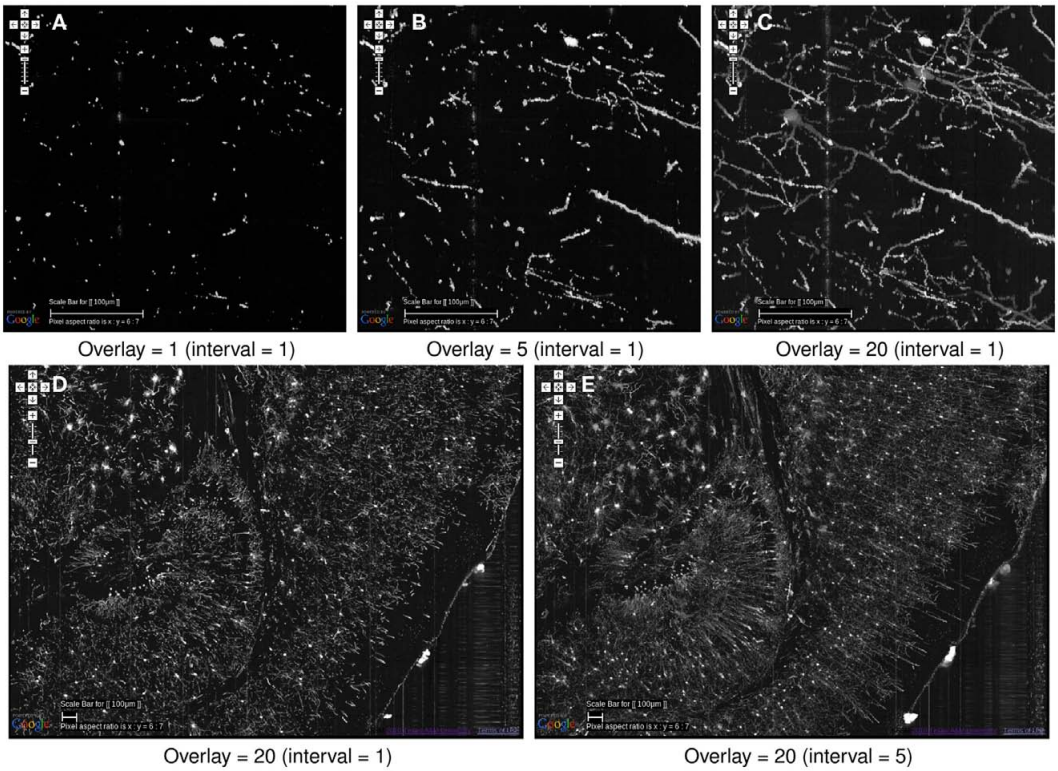


FIGURE 9 | Effectiveness of image overlays. The effect of an increasing number of overlays is shown. Scale bar = 100 μm. The data is from the same region as that from **Figure 8**. **(A)** Since each KESM image corresponds to a 1-μm-thick section, a single image conveys little information about the neuronal morphology. **(B)** Five overlayed images, corresponding to a 5-μm-thick section, begins to show some structure but it is not enough. **(C)** With twenty overlayed images, familiar structures begin to appear. **(D,E)** At a zoomed-out scale, skipping over images can be an effective strategy to view the circuits more clearly. In **(D)**, 20 overlays at an interval of 1, representing a 20-μm-thick volume is shown. In **(E)**, 20 overlays at an interval of 5 is shown, representing 100 μm. The dense dendritic arbor in the hippocampus (left), fiber tract projecting toward the hippocampal commissure (middle, top), and the massive number of pyramidal cells and their apical dendrites (right) are clearly visible only in **(E)**.

orders of magnitude lower (for example, one coronal brain set has 234 slides in it, corresponding to a sectional thickness of 25 μm). The database also serves serial sections from electron microscopy, cryo sections, and immunohistochemistry, and hosts a total of 135 data sets (as of March 2, 2011).

4.1.5. Whole-Brain Catalog (WBC)

WBC is a 3D virtual environment for exploring multiple sources of brain data (including mouse brain data), e.g., Cell Centered Database (CCDB, see below), Neuroscience Information Framework (NIF), and the Allen Brain Atlas (see above). WBC has native support for registering to the Waxholm Space, a rodent standard atlas space (Johnson et al., 2010). Multiple functionalities including visualization, slicing, animations, and simulations are supported.

In summary, there are several mouse brain atlases available, with data from different imaging modalities, but their resolution is not high enough in one or more of the *x*, *y*, or *z* axes to show morphological detail of neurons.

4.2. DATABASES OF 3D RECONSTRUCTION OF NEURONS

The low spatial resolution in existing whole-brain level brain maps and atlases have been pointed out as a major limitation. Near-micron-level reconstructions of brain areas do exist, but only for a small volume. Part of the reason is that, in many cases, the geometric reconstructions were done manually, with the aid of interactive editing tools like Neurolucida (Glaser and Glaser, 1990), Reconstruct (Fiala, 2005), or Neuron_Morpho (Brown et al., 2007).

4.2.1. The Duke/Southampton archive of neuronal morphology

This on-line archive of neuronal geometry (Cannon et al., 1998) includes full 3D representations of 124 neurons from the rat hippocampus, obtained following intracellular staining with biocytin and reconstruction using Neurolucida. The archive includes data both in the native format as supplied from the digitization software, and in a simpler, 3D standardized format (given the extension "SWC" in the archive). The data for the SWC files are obtained by fitting cell segments in three dimensions with cylinders, directly confirming the location and size of these shapes using a computer-based tracing system.

4.2.2. NeuroMorpho.org

This is a centrally curated collection of reconstructed neurons, currently containing 5793 cells (version 5, November 15, 2010) from

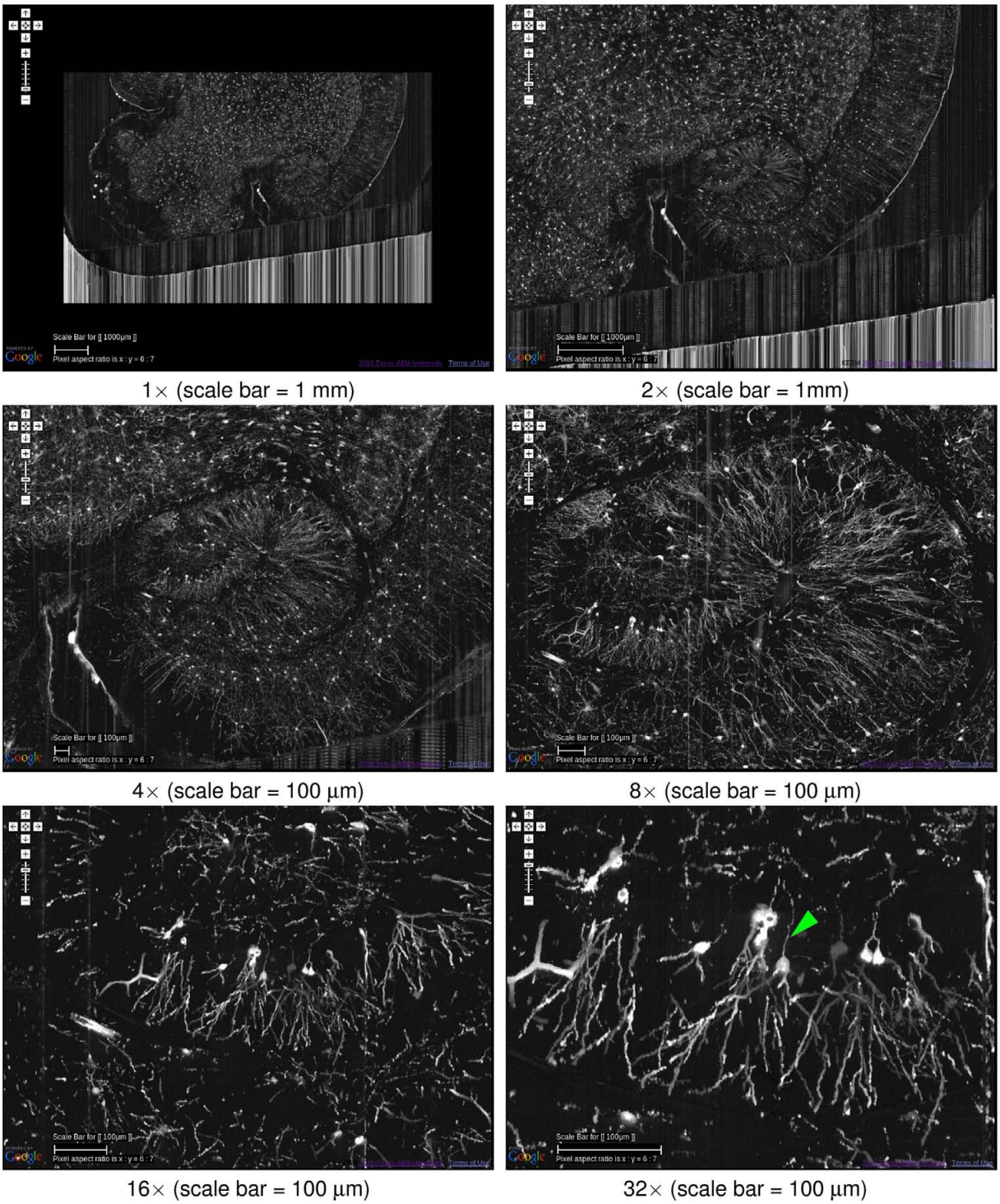


FIGURE 10 | Multiscale view of the KESMBA. A multiscale view of the KESMBA is shown (Golgi data set 1), by gradually zooming into the hippocampus (the numbers below the panels show the zoom-in sequence). All panels show an overlay of 20 sections. The first four panels are shown with an overlay interval of 5 and the last two with an interval of 1. Axons emerging from the hippocampal neurons are clearly visible (arrow head, last panel).

various species and brain regions (Ascoli et al., 2007). The data are available for download in SWC format. *L-neuron* is a modeling and analysis project that is associated with this database, where statistical features of dendritic geometry and stochastic generation of (statistically) realistic neurons are studied (Senft and Ascoli, 1999; Ascoli and Krichmar, 2000; Ascoli, 2002).

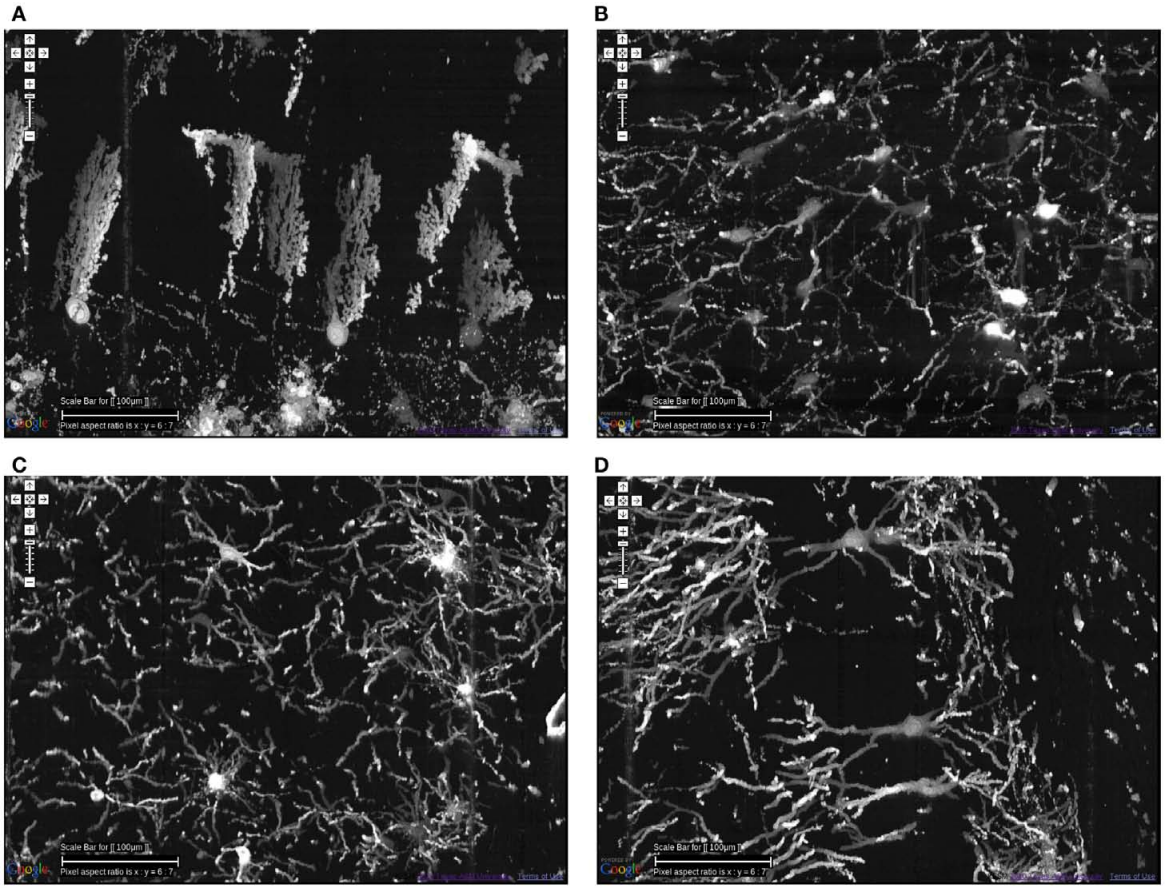


FIGURE 11 | Different types of local circuits. Different types of local circuits from the KESM Golgi data set 1 are shown. **(A)** Cerebellum. **(B)** Inferior colliculus. **(C)** Thalamus. **(D)** Hippocampus (also see **Figure 10**). See **Figure 8** for circuits in the neocortex. Scale bar = 100 µm. See Movie 2 in Supplementary Material (cerebellum, colliculi) and Movie 3 in Supplementary Material (hippocampus).

4.2.3. The cell centered database (CCDB)

CCDB houses high-resolution 3D light and electron microscopic reconstructions spanning the dimensional range from 5 nm^3 to 50 μm^3 produced at the National Center for Microscopy and Imaging Research (NCMIR; Martone et al., 2002). The current CCDB has 8391 micrograph data sets (as of March 2, 2011) in various modalities including confocal, light microscopy, electron tomography, electron microscopy, live imaging, filled cell imaging, protein imaging, and serial block-face imaging.

4.2.4. The SynapseWeb

The SynapseWeb (Fiala and Harris, 2001) is a portal into a dense network of synaptic connections and supporting structures in the gray matter of the brain that can be fully visualized only through 3D electron microscopy. It provides an interface for examining volumes of brain tissue at nanometer resolutions which have been reconstructed from serial section electron microscopy. Currently, the SynapseWeb houses three brain volumes ranging from 62 to 108 μm^3 from the CA1 regions of rat hippocampus.

In summary, there are several excellent neuronal morphology databases that serve the neuroscience community, but they are limited to a small number of neurons from limited volumes, isolated from the system-level context.

4.3. ATLASING AND NEURONAL MORPHOLOGY DESCRIPTION STANDARDS

A rapid increase in web-based resources serving neuronal morphology and atlas-scale data sets gave rise to the need for data representation standards.

The Waxholm Space (Johnson et al., 2010; Hawrylycz et al., 2011) is a new standard atlasing space for rodents. The effort to build this standard space was motivated by multiple non-standard, yet widely used coordinate spaces such as those in the Allen Brain Atlas (Lein et al., 2007) or Paxinos and Franklin's atlas (Paxinos and Franklin, 2001).

As for neuronal morphology, NeuroML has become the *de facto* standard (using XML). BrainML, on the other hand, provides an XML framework for the exchange of general neuroscience data at the whole-brain scale.

4.4. ALTERNATIVE MAPPING APIs

Geospatial interfaces have undergone massive innovation in the last decade and as we have demonstrated in this article, they can

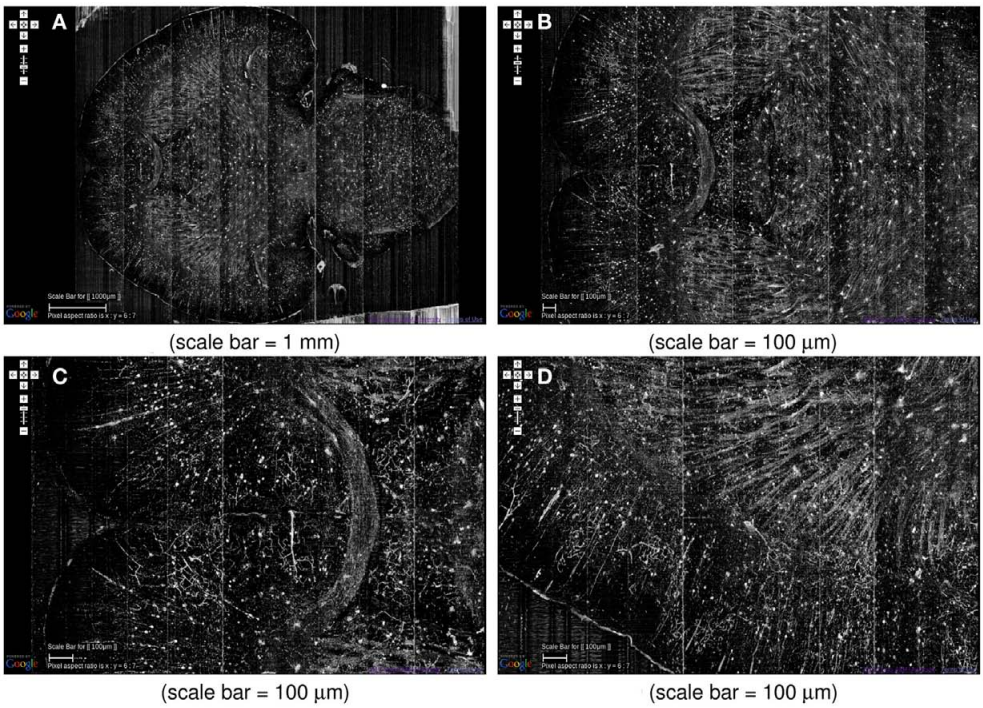


FIGURE 12 | System-level fiber tracts in the KESM Golgi data set 2. (A) Horizontal section at the level of the anterior commissure (the ")"-shaped fiber bundle) is shown (left: anterior, right: posterior). Massive fiber tracts in the striatum can also be observed. **(B,C)** Zoomed in view showing the anterior commissure near the middle. **(D)** Close-up of the fiber bundles in the striatum can be seen. A large number of apical dendrites in the adjoining cortex can also be seen.

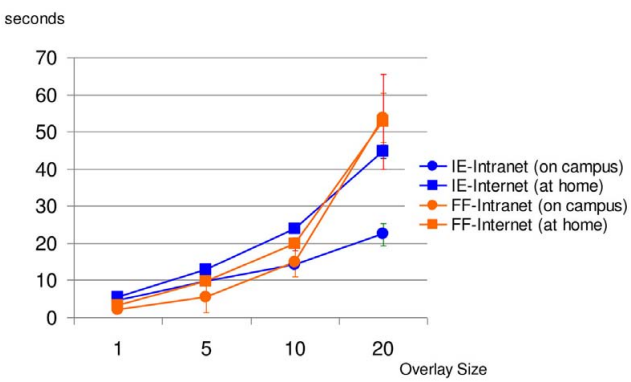


FIGURE 13 | KESMBA download performance analysis. Average of 10 download trials for each setting (browser type and overlay size) is plotted (error bars indicate standard deviation). IE, internet explorer, FF, firefox. Except for the case of Mozilla Firefox downloading 20 overlays, both the Intranet and Internet downloading times increased proportionally with the overlay size.

be very effective in presenting biological data. However, existing tools are encumbered by proprietary licensing, which limits adaptation to cosmetic levels and requires awkward workarounds to implement even basic functionality. Use of open source tools will allow for code level adaptation as opposed to API extension, and facilitate interoperability and adoption by other groups.

Tools that can help this transition include GDAL and OpenLayers. The open source GDAL[1] is a translator library for data formats maintained by the Open Source Geospatial Foundation. OpenLayers is an open source browser-based map display system using client side JavaScript. OpenLayers[2] serves up data as a service and supports the basic tile display functionality (i.e., zoom levels, layers) with custom controls used in map navigation. It also supports a number of advanced features such as layer opacity, feature opacity, vector formats and others necessary for more sophisticated user interfaces. A key feature is the ability to use disk-based caching to improve local performance, which can greatly improve performance of KESMBA-like web atlases.

Open standards and best practices are widely used in the geospatial community and contribute significantly to the interoperability of geospatial visualization across a wide range of devices. Modifications of standards such as the Web Map Service for large scale microscopy data offer potential for interoperability between neuroinformatics systems.

4.5. CHALLENGES

The KESM data sets are in a unique strategic position to help advance the field of connectomics in the short term future (5–10 years). This is due to its system-level scope combined with sub-micrometer resolution. However, there are many challenges

[1]http://gdal.org/
[2]http://openlayers.org

that need to be overcome in order to enable fully quantitative connectivity analysis such as graph theoretical analysis (Sporns, 2002; Sporns and Tononi, 2002; Kötter and Stephan, 2003) or motif analysis (Milo et al., 2002). Here, we will discuss some of these challenges and suggest strategies to overcome these challenges.

4.5.1. Establishing connectivity

One of the main issues with any approach based on light microscopy (LM) is the requirement that sparse stains like Golgi are used (which stains about 1% of the total neuronal population). Dense stains commonly used in electron microscopy will render objects in the specimen indistinguishable at the resolution permitted by LM. Furthermore, stains like Golgi do not stain thin axons, thus tracing long projections even for the sparse sample is difficult. Long-range tracers like biocytin could be a good solution, but these tracers require intracellular injection and a long transport time, so applying them at the whole-brain scale can be troublesome. An attractive possibility is to use Brainbow transgenic mice (Livet et al., 2007), combined with fluorescence imaging (recently, we have successfully imaged fluorescent proxies [10 μm beads] with the KESM using laser illumination). This way, neurons are densely labeled, but due to the variation in the emitted wavelength, even close-by neurons can be differentiated (Lichtman et al., 2008). Techniques using pseudorabies virus (PRV) can also be used to label neurons that form an actual circuit since PRV allows for trans-synaptic tracing (Smith et al., 2000; Willhite et al., 2006; Kim et al., 2011). As most recent labeling methods such as Brainbow and PRV require fluorescence imaging, further development of KESM fluorescence imaging capability will become a key requirement.

An alternative to the experimental techniques above is to estimate connectivity based on the sparse data. Methods like those proposed by Kalisman et al. (2003) can be used for this purpose. Also, a systematic simulation study can be conducted with a full synthetic circuit, by dropping a certain proportion of connections and observing the resulting change in behavior. The degree of redundancy in the connections (both for real and synthetic circuits) will play an important role here.

4.5.2. From image to structure

Another important issue is that of structural reconstruction (**Figure 14**). Together with whole-brain-scale data acquisition, structural reconstruction is a grand challenge for connectomics. The DIADEM (Digital Reconstruction of Axonal and Dendritic Morphology) Challenge and lessons learned from the first round of competition show a long road ahead of us in terms of accurate circuit tracing (Liu, 2011). The KESM data sets are basically image stacks and they do not provide quantitative morphological or connectivity data. Among different approaches we have found that vector tracing methods are fast and reliable (Can et al., 1999; Al-Kofahi et al., 2002; Mayerich et al., 2008a, 2011a; Han et al., 2009a,b). However, these approaches are not perfect and small errors can lead to topological mistakes, which can cause serious errors in establishing connectivity (Jain et al., 2010a). Jain et al. (2010b) propose the use of machine learning techniques, and this can be a promising direction. However, whatever automated methods we use the burden of validation (see, e.g., Warfield et al., 2004; Mayerich et al., 2008c) still remains and human intervention is inevitable. The question is how to make this human intervention minimal while maximizing accuracy. We are currently exploring several options: (1) multiple-choice selection from parameterized reconstruction alternatives, (2) interactive editing using graph cuts, and (3) colorized voxel-intensity-based confidence to aid in rapid editing region selection (Yang and Choe, 2009). These approaches can help combine automated reconstruction algorithms with the power of human computing (von Ahn, 2006; von Ahn et al., 2006, 2008), to enable reliable tracing of massive volumes of neuronal circuit data.

4.5.3. From structure to function

The connectome is fundamentally a static structure, an adjacency matrix. Important physiological parameters such as sign

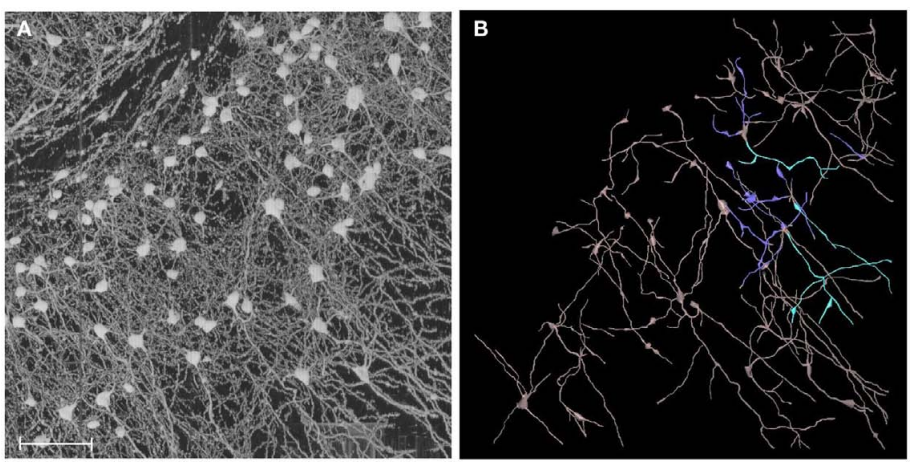


FIGURE 14 | From image to geometry. (A) A portion of KESM Golgi data set 1 is shown (cortex). Maximum intensity projection is used to show a thicker section containing a large number of neurons. Scale bar = 100 μm. **(B)** Semi-automated 3D reconstruction results are shown (partial results, using Neuromantic, Myatt and Nasuto, 2007).

(excitatory/inhibitory), weight (synaptic efficacy), and delay (axonal conduction delay) are not available. How can these and other physiological properties be inferred from just the structure? Toledo-Rodriguez et al. (2004) shows a possibly powerful solution to this: Use gene expression data. They found that gene expression and electrophysiological properties are closely correlated. The availability of very large gene expression atlases such as the Allen Brain Atlas (Lein et al., 2007; 22,000 genes), and imaging modalities such as Array Tomography that support molecular as well as EM imaging (Micheva and Smith, 2007) are great resources for this kind of approach (see, e.g., Markram, 2006). Another straight-forward yet potentially valuable approach is to start with computational simulation based on detailed neuronal morphology (cf. the Blue Brain Project; Markram, 2006). The reconstructed geometry can be used to construct multi-compartment models (see, e.g., Dayan and Abbott, 2001). Appropriate parameters such as channel conductance, capacitance, etc., need to be figured out (Vanier and Bower, 1999). Tools like NEURON, GENESIS, neuroConstruct, and NeuGEN can be used for multi-compartment simulation and parameterized synthetic circuit generation/simulation/analysis (Hines and Carnevale, 1997; Bower and Beeman, 1998; Ascoli et al., 2001; Eberhard et al., 2006; Gleeson et al., 2007; Koene, 2007; Koene et al., 2009). Data from the KESM can help narrow down on the range of various parameters for these simulations (see Druckmann et al., 2008 for parameter constraining procedures).

4.5.4. *Enabling connectomics research through neuroinformatics*

From visualization to annotation to editing and quantitative analysis, neuroinformatics tools are expected to serve as a key to the success of connectomics research. This is because the process of going from data to information and information to knowledge cannot be achieved through purely automated (fast but inaccurate) or purely manual (accurate but slow) means. Thus, an informatics platform is needed to optimally blend both automated and manual exploration and analysis methods.

The KESM brain atlas framework provides a good starting point. However, to increase its utility as a connectomics platform, it needs to be expanded to include support for all three standard sections (coronal, horizontal, and sagittal), overlay of automated reconstructions, and reconstruction editing facilities (uncertainty/confidence visualization, reconstruction alternatives, etc.). These reconstructed morphologies have to be moved one step farther to achieve the connectivity diagram needed for connectomics research. Thus, facilities to allow users to manually specify neuron-to-neuron connectivity, or allow parametric connectivity specification (e.g., connect axons and dendrites that meet a certain rule, where the rule can be specified by setting parameters such as proximity radius, etc.) are needed. With the reconstructed geometry already in place, calculating such parameterized connectivity could be done rapidly. A distribution of connection matrices can be generated based on such parametrizations, from which meaningful structural and functional properties of the connectome can be extracted.

ACKNOWLEDGMENTS

This project was funded in part by the Collaborative Research in Computational Neuroscience Program of National Science Foundation (NSF CRCNS #0905041), National Institutes of Health - National Institute of Neurological Disorders and Stroke (NIH/NINDS #1R01-NS54252), the National Science Foundation's (NSF) Major Research Instrumentation (MRI) Program (NSF MRI #0079874), the National Science Foundation's (NSF) Information Technology Research (ITR) Program (NSF ITR #CCR-0220047), Texas Higher Education Coordinating Board (ATP#000512-0146-2001), Texas A&M Research Foundation, Texas Engineering Experiment Station, and 3Scan. The appearance of the Google logo in the figures are due to the use of the Google Maps API by the KESM brain atlas. Besides that this work was done independent of Google. We would like to thank Bruce H. McCormick, inventor of the KESM, for his visionary insights, and Bernard Mesa for technical advice and instrumentation for the KESM.

SUPPLEMENTARY MATERIAL

The Movies S1, S2, and S3 for this article can be found online at http://www.frontiersin.org/Neuroinformatics/10.3389/fninf.2011.00029/abstract

Media Files: All video clips were made using MeVisLab (http://www.mevislab.de). Note that the video clips are presented to showcase the KESM data itself, and are independent of the KESMBA web interface.

Movie S1 | A video clip of a sweep-through of the entire KESM Golgi data set 2 is shown, along all three sectioning planes: Sagittal, coronal, and horizontal. Initial block width is 11.52 mm.

Movie S2 | A video clip of a small region near the cerebellum and the colliculi from the KESM Golgi data set 1 is shown. Zoom-in near the end of the clip shows a number of cerebellar Purkinje cells. The view is initially horizontal, but later on it rotates and shows sagittal sections. Initial width of the block is 2.88 mm.

Movie S3 | A video clip of a small region near the hippocampus (middle) and the cortex (bottom) from the KESM Golgi data set 1 is shown. Initial width of the block is 1.44 mm.

REFERENCES

Abbott, L. C. (2008). "High-throughput imaging of whole small animal brains with the knife-edge scanning microscope," in *Neuroscience Meeting Planner*, Program No. 504.2. Washington, DC: Society for Neuroscience.

Abbott, L. C., and Sotelo, C. (2000). Ultrastructural analysis of catecholaminergic innervation in weaver and normal mouse cerebellar cortices. *J. Comp. Neurol.* 426, 316–329.

Al-Kofahi, K. A., Lasek, S., Szarowski, D. H., Pace, C. J., Nagy, G., Turner, J. N., and Roysam, B. (2002). Rapid automated three-dimensional tracing of neurons from confocal image stacks. *IEEE Trans. Inf. Technol. Biomed.* 6, 171–187.

Ascoli, G., Krichmar, J., Scorcioni, R., Nasuto, S., and Senft, S. (2001). Computer generation and quantitative morphometric analysis of virtual neurons. *Anat. Embryol. (Berl.)* 204, 283–301.

Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). NeuroMorpho.Org: a central resource for neuronal morphologies. *J. Neurosci.* 27, 9247–9251.

Ascoli, G. A., and Krichmar, J. L. (2000). L-Neuron: a modeling tool for the efficient generation and parsimonious description of dendritic morphology. *Neurocomputing* 32–33, 1003–1011.

Ascoli, G. A. (ed.). (2002). *Computational Neuroanatomy: Principles and Methods.* Totowa, NJ: Humana Press.

Basser, P. J., and Jones, D. K. (2002). Diffusion-tensor MRI: theory, experimental design and data

analysis – a technical review. *NMR Biomed.* 15, 456–467.

Bower, J. M., and Beeman, D. (1998). *The Book of GENESIS: Exploring Realistic Neural Models with the General Neural Simulation System*. Santa Clara, CA: Telos.

Brown, K. M., Donohue, D. E., D'Alessandro, G., and Ascoli, G. A. (2007). A cross-platform freeware tool for digital reconstruction of neuronal arborizations from image stacks. *Neuroinformatics* 3, 343–359.

Can, A., Shen, H., Turner, J. N., Tanenbaum, H. L., and Roysam, B. (1999). Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. *IEEE Trans. Inf. Technol. Biomed.* 3, 125–138.

Cannon, R. C., Turner, D. A., Pyapali, G. K., and Wheal, H. V. (1998). An on-line archive of reconstructed hippocampal neurons. *J. Neurosci. Methods* 84, 49–54.

Choe, Y., Abbott, L. C., Han, D., Huang, P.-S., Keyser, J., Kwon, J., Mayerich, D., Melek, Z., and McCormick, B. H. (2008). "Knife-edge scanning microscopy: high-throughput imaging and analysis of massive volumes of biological microstructures," in *High-Throughput Image Reconstruction and Analysis: Intelligent Microscopy Applications*, eds A. R. Rao and G. Cecchi (Boston, MA: Artech House), 11–37.

Choe, Y., Abbott, L. C., Miller, D. E., Han, D., Yang, H.-F., Chung, J. R., Sung, C., Mayerich, D., Kwon, J., Micheva, K., and Smith, S. J. (2010). "Multiscale imaging, analysis, and integration of mouse brain networks," in *Neuroscience Meeting Planner*, Program No. 516.3. San Diego, CA: Society for Neuroscience.

Choe, Y., Han, D., Huang, P.-S., Keyser, J., Kwon, J., Mayerich, D., and Abbott, L. C. (2009). "Complete submicrometer scans of mouse brain microstructure: neurons and vasculatures," in *Neuroscience Meeting Planner*, Program No. 389.10. Chicago, IL: Society for Neuroscience.

Dayan, P., and Abbott, L. F. (2001). *Theoretical Neuroscience*. Cambridge, MA: MIT Press.

Denk, W., and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* 19, e329. doi:0.1371/journal.pbio.0020329

Druckmann, S., Berger, T. K., Hill, S., Schürmann, F., Markram, H., and Segev, I. (2008). Evaluating automated parameter constraining procedures of neuron models by experimental and surrogate data. *Biol. Cybern.* 99, 371–379.

Eberhard, J. P., Wanner, A., and Wittum, G. (2006). NeuGen: a tool for the generation of realistic morphology of cortical neurons and neural networks. *Neurocomputing* 70, 327–342.

Eng, D. C.-Y., and Choe, Y. (2008). "Stereo pseudo 3D rendering for web-based display of scientific volumetric data," in *Proceedings of the IEEE/EG International Symposium on Volume Graphics*. Los Angeles, CA.

Fiala, J. C. (2005). Reconstruct: a free editor for serial section microscopy. *J. Microsc.* 218, 52–61.

Fiala, J. C., and Harris, K. M. (2001). Extending unbiased stereology of brain ultrastructure to three-dimensional volumes. *J. Am. Med. Inform. Assoc.* 8, 1–16.

Galletta, D. F., Henry, R., McCoy, S., and Polak, P. (2004). Web site delays: how tolerant are users? *J. Assoc. Inf. Syst.* 5, 1–28.

Glaser, J. R., and Glaser, E. M. (1990). Neuron imaging with Neurolucida – a PC-based system for image combining microscopy. *Comput. Med. Imaging Graph.* 14, 307–317.

Gleeson, P., Steuber, V., and Silver, R. (2007). NeuroConstruct: a tool for modeling networks of neurons in 3D space. *Neuron* 54, 219–235.

Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V. J., Meuli, R., and Thiran, J.-P. (2007). Mapping human whole-brain structural networks with diffusion MRI. *PLoS ONE* 2, e597. doi:10.1371/journal.pone.0000597

Han, D., Choi, H., Park, C., and Choe, Y. (2009a). "Fast and accurate retinal vasculature tracing and kernel-isomap-based feature selection," in *Proceedings of the International Joint Conference on Neural Networks* (Piscataway, NJ: IEEE Press), 1075–1082.

Han, D., Keyser, J., and Choe, Y. (2009b). "A local maximum intensity projection tracing of vasculature in knife-edge scanning microscope volume data," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, Boston, MA, 1259–1262.

Hawrylycz, M., Baldock, R. A., Burger, A., Hashikawa, T., Johnson, G. A., Martone, M., Ng, L., Lau, C., Larsen, S. D., Nissanov, J., Puelles, L., Ruffins, S., Verbeek, F., Zaslavsky, I., and Boline, J. (2011). Digital atlasing and standardization in the mouse brain. *PLoS Comput. Biol.* 7, e1001065. doi:10.1371/journal.pcbi.1001065

Hayworth, K. (2008). "Automated creation and SEM imaging of Ultrathin Section Libraries: tools for large volume neural circuit reconstruction," in *Society for Neuroscience Abstracts*, Program No. 504.4. Washington, DC: Society for Neuroscience.

Hines, M. L., and Carnevale, N. T. (1997). The NEURON simulation environment. *Neural. Comput.* 9, 1179–1209.

Jain, V., Bollmann, B., Richardson, M., Berger, D. R., Helmstaedter, M. N., Briggman, K. L., Denk, W., Bowden, J. B., Mendenhall, J. M., Abraham, W. C., Harris, K. M., Kasthuri, N., Hayworth, K. J., Schalek, R., Tapia, J. C., Lichtman, J. W., and Seung, H. S. (2010a). "Boundary learning by optimization with topological constraints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 2488–2495.

Jain, V., Seung, H. S., and Turaga, S. C. (2010b). Machines that learn to segment images: a crucial technology for connectomics. *Curr. Opin. Neurobiol.* 20, 653–666.

Johnson, G. A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., and Nissanov, J. (2010). Waxholm space: an image-based reference for coordinating mouse brain research. *Neuroimage* 53, 365–372.

Kalisman, N., Silberberg, G., and Markram, H. (2003). Deriving physical connectivity from neuronal morphology. *Biol. Cybern.* 88, 210–218.

Kersten, M., Stewart, J., Troje, N., and Ellis, R. (2006). Enhancing depth perception in translucent volumes. *IEEE Trans. Vis. Comput. Graph* 12, 1117–1124.

Kim, D. H., Phillips, M. E., Chang, A. Y., Patel, H. K., Nguyen, K. T., and Willhite, D. C. (2011). Lateral connectivity in the olfactory bulb is sparse and segregated. *Front. Neural Circuits*, 5:5. doi:10.3389/fncir.2011.00005

Koene, R. A. (2007). "Large scale high resolution network generation: producing known validation sets for serial reconstruction methods that use histological images of neural tissue," in *International Conference on Complex Systems*. Boston, MA. [Presentation].

Koene, R. A., Tijms, B., van Hees, P., Postma, F., de Ridder, A., Ramakers, G. J. A., van Pelt, J., and van Ooyen, A. (2009). NETMORPH: a framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. *Neuroinformatics* 7, 1539–2791.

Kötter, R., and Stephan, K. E. (2003). Network participation index: characterizing component roles for information processing in neural networks. *Neural Netw.* 16, 1261–1275.

Kwon, J., Mayerich, D., Choe, Y., and McCormick, B. H. (2008). "Lateral sectioning for knife-edge scanning microscopy," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, Paris, France, 1371–1374.

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T. M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., Dee, N. R., Desaki, A. L., Desta, T., Diep, E., Dolbeare, T. A., Donelan, M. J., Dong, H. W., Dougherty, J. G., Duncan, B. J., Ebbert, A. J., Eichele, G., Estin, L. K., Faber, C., Facer, B. A., Fields, R., Fischer, S. R., Fliss, T. P., Frensley, C., Gates, S. N., Glattfelder, K. J., Halverson, K. R., Hart, M. R., Hohmann, J. G., Howell, M. P., Jeung, D. P., Johnson, R. A., Karr, P. T., Kawal, R., Kidney, J. M., Knapik, R. H., Kuan, C. L., Lake, J. H., Laramee, A. R., Larsen, K. D., Lau, C., Lemon, T. A., Liang, A. J., Liu, Y., Luong, L. T., Michaels, J., Morgan, J. J., Morgan, R. J., Mortrud, M. T., Mosqueda, N. F., Ng, L. L., Ng, R., Orta, G. J., Overly, C. C., Pak, T. H., Parry, S. E., Pathak, S. D., Pearson, O. C., Puchalski, R. B., Riley, Z. L., Rockett, H. R., Rowland, S. A., Royall, J. J., Ruiz, M. J., Sarno, N. R., Schaffnit, K., Shapovalova, N. V., Sivisay, T., Slaughterbeck, C. R., Smith, S. C., Smith, K. A., Smith, B. I., Sodt, A. J., Stewart, N. N., Stumpf, K. R., Sunkin, S. M., Sutram, M., Tam, A., Teemer, C. D., Thaller, C., Thompson, C. L., Varnam, L. R., Visel, A., Whitlock, R. M., Wohnoutka, P. E., Wolkey, C. K., Wong, V. Y., Wood, M., Yaylaoglu, M. B., Young, R. C., Youngstrom, B. L., Yuan, X. F., Zhang, B., Zwingman, T. A., and Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176.

Li, A., Gong, H., Zhang, B., Wang, Q., Yan, C., Wu, J., Liu, Q., Zeng, S., and Luo, Q. (2010). Micro-optical sectioning tomography to obtain a high-resolution atlas of the mouse brain. *Science* 330, 1404–1408.

Lichtman, J. W., Livet, J., and Sanes, J. R. (2008). A technicolour approach to the connectome. *Nat. Rev. Neurosci.* 9, 417–422.

Liu, Y. (2011). The DIADEM and beyond. *Neuroinformatics* 9, 99–102.

Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., and Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 450, 56–62.

MacKenzie-Graham, A., Jones, E. S., Shattuck, D. W., Dinov, I. D., Bota, M., and Toga, A. W. (2003). The informatics of a C57BL/6J mouse brain atlas. *Neuroinformatics* 1, 397–410.

Markram, H. (2006). The blue brain project. *Nat. Rev. Neurosci.* 7, 153–160.

Martone, M. E., Gupta, A., Wong, M., Qian, X., Sosinsky, G., Ludscher, B., and Ellisman, M. H. (2002). A cell-centered database for electron tomographic data. *J. Struct. Biol.* 138, 145–155.

Mayerich, D., Abbott, L. C., and Keyser, J. (2008a). Visualization of cellular and microvessel relationship. *IEEE Trans. Vis. Comput. Graph.* 14, 1611–1618.

Mayerich, D., Abbott, L. C., and McCormick, B. H. (2008b). Knife-edge scanning microscopy for imaging and reconstruction of three-dimensional anatomical structures of the mouse brain. *J. Microsc.* 231, 134–143.

Mayerich, D., Kwon, J., Choe, Y., Abbott, L., and Keyser, J. (2008c). "Constructing high-resolution microvascular models," in *Proceedings of the 3rd International Workshop on Microscopic Image Analysis with Applications in Biology (MIAAB 2008)*, New York, NY.

Mayerich, D., Bjornsson, C., Taylor, J., and Roysam, B. (2011a). "Metrics for comparing explicit representations of interconnected biological networks," in *IEEE Symposium on Biological Data Visualization*, Chicago, IL, 79–86.

Mayerich, D., Kwon, J., Sung, C., Abbott, L. C., Keyser, J., and Choe, Y. (2011b). Fast macro-scale transmission imaging of microvascular networks using KESM. *Biomed. Opt. Express* 2, 2888–2896.

Mayerich, D., McCormick, B. H., and Keyser, J. (2007). "Noise and artifact removal in knife-edge scanning microscopy," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, Washington, DC, 556–559.

McCormick, B. H. (2003). *The knife-edge Scanning Microscope*. Technical report, Department of Computer Science, Texas A&M University. Available at: http://research.cs.tamu.edu/bnl/

McCormick, B. H. (2004). U.S. Patent No. 6,744,572. Washington, DC: U.S. Patent and Trademark Office.

Micheva, K., and Smith, S. J. (2007). Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron* 55, 25–36.

Mikula, S., Trotts, I., Stone, J. M., and Jones, E. G. (2007). Internet-enabled high-resolution brain mapping and virtual microscopy. *Neuroimage* 35, 9–15.

Milo, R., Sen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827.

Myatt, D., and Nasuto, S. (2007). Three-dimensional reconstruction of neurons with neuromantic. *AISB Q.* 125, 1–2.

Paxinos, G., and Franklin, K. B. J. (2001). *The Mouse Brain in Stereotaxic Coordinates*, 2nd Edn. San Diego, CA: Academic Press.

Paxinos, G., and Watson, C. (2006). *The Rat Brain in Stereotaxic Coordinates*, 6th Edn. San Diego, CA: Academic Press.

Roebroeck, A., Galuske, R., Formisano, E., Chiry, O., Bratzke, H., Ronen, I., Kim, D.-S., and Goebel, R. (2008). High-resolution diffusion tensor imaging and tractography of the human optic chiasm at 9.4 T. *Neuroimage* 39, 157–168.

Rosen, G. D., Williams, A. G., Capra, J. A., Connolly, M. T., Cruz, B., Lu, L., Airey, D. C., Kulkarni, A., and Williams, R. W. (2000). "The mouse brain library@www.mbl.org," in *Proceedings of the 14th International Mouse Genome Meeting*, Narita, C6.

Selvidge, P. R., Chaparro, B., and Bender, G. T. (2002). The world wide wait: effects of delays on user performance. *Int. J. Ind. Ergon.* 29, 15–20.

Senft, S. L., and Ascoli, G. A. (1999). Reconstruction of brain networks by algorithmic amplification of morphometry data. *Lect. Notes Comput. Sci.* 1606, 25–33.

Shepherd, G. M. (ed.). (2003). *The Synaptic Organization of the Brain*, 5th Edn. New York: Oxford University Press.

Smith, B. N., Banfield, B. W., Smeraski, C. A., Wilcox, C. L., Dudek, F. E., Enquist, L. W., and Pickard, G. E. (2000). Pseudorabies virus expressing enhanced green fluorescent protein: a tool for in vitro electrophysiological analysis of transsynaptically labeled neurons in identified central nervous system circuits. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9264–9269.

Sporns, O. (2002). "Graph theory methods for the analysis of neural connectivity patterns," in *Neuroscience Databases: A Practical Guide*, ed. R. Kötter (Boston, MA: Kluwer Publishers), 171–186.

Sporns, O. (2011). *Networks of the Brain*. Cambridge, MA: MIT Press.

Sporns, O., and Tononi, G. (2002). Classes of network connectivity and dynamics. *Complexity* 7, 28–38.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi:10.1371/journal.pcbi.0010042

Swanson, L. W. (2003). *Brain Architecture: Understanding the Basic Plan*. Oxford: Oxford University Press.

Thiel, A., Schwegler, H., and Eurich, C. W. (2003). Complex dynamics is abolished in delayed recurrent systems with distributed feedback times. *Complexity* 8, 102–108.

Toledo-Rodriguez, M., Blumenfeld, B., Wu, C., Luo, J., Attali, B., Goodman, P., and Markram, H. (2004). Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. *Cereb. Cortex* 14, 1310–1327.

Tsai, P. S., Friedman, B., Ifarraguerri, A. I., Thompson, B. D., Lev-Ram, V., Schaffer, C. B., Xiong, Q., Tsien, R. Y., Squier, J. A., and Kleinfeld, D. (2003). All-optical histology using ultrashort laser pulses. *Neuron*, 39, 27–41.

Tuch, D. S., Reese, T. G., Wiegell, M. R., and Wedeen, V. J. (2003). Diffusion MRI of complex neural architecture. *Neuron* 40, 885–895.

van Pelt, J., and Uylings, H. (2005). "Natural variability in the geometry of dendritic branching patterns," in *Modeling in the neurosciences: From Ionic Channels to Neural Networks*, 2nd Edn, Chap. 4, ed. R. R. Pozanski (London: CRC Press), 79–108.

Vanier, M. C., and Bower, J. M. (1999). A comparative survey of automated parameter-search methods for compartmental models. *J. Comput. Neurosci.* 7, 149–171.

von Ahn, L. (2006). Games with a purpose. *IEEE Comput.* 39, 96–96.

von Ahn, L., Kedia, M., and Blum, M. (2006). "Peekaboom: a game for locating objects in images," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHO 2006)*, Montreal, QC, 55–64.

von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. (2008). reCAPTCHA: human-based character recognition via web security measures. *Science* 321, 1465–1468.

Warfield, S. K., Zou, K. H., and Wells, W. M. (2004). Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* 23, 903–921.

Willhite, D. C., Nguyen, K. T., Masurkar, A. V., Greer, C. A., Shepherd, G. M., and Chen, W. R. (2006). Viral tracing identifies distributed columnar organization in the olfactory bulb. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12592–12597.

Yang, H.-F., and Choe, Y. (2009). "Cell tracking and segmentation in electron microscopy images using graph cuts," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*, Boston, MA, 306–309.

Conflict of Interest Statement: Todd Huffman is employed by 3Scan. Part of the instrumentation, computing, and storage costs of this project has been funded by 3Scan.

Received: 16 March 2011; accepted: 01 November 2011; published online: 22 November 2011.

Citation: Chung JR, Sung C, Mayerich D, Kwon J, Miller DE, Huffman T, Keyser J, Abbott LC and Choe Y (2011) Multiscale exploration of mouse brain microstructures using the knife-edge scanning microscope brain atlas. Front. Neuroinform. ***5****:29. doi: 10.3389/fninf.2011.00029*

frontiers in
NEUROINFORMATICS

PERSPECTIVE ARTICLE
published: 25 October 2011
doi: 10.3389/fninf.2011.00021

Tracing activity across the whole brain neural network with optogenetic functional magnetic resonance imaging

Jin Hyung Lee [1,2,3,4,5] *

[1] Department of Electrical Engineering, University of California Los Angeles, Los Angeles, CA, USA
[2] Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, CA, USA
[3] Department of Radiology, University of California Los Angeles, Los Angeles, CA, USA
[4] Department of Neuroscience, University of California Los Angeles, Los Angeles, CA, USA
[5] Biomedical Engineering Interdepartmental Program, University of California Los Angeles, Los Angeles, CA, USA

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Kathleen S. Rockland, Massachusetts Institute of Technology, USA
Yoonsuck Choe, Texas A&M University, USA

****Correspondence:***
Jin Hyung Lee, Departments of Electrical Engineering, Psychiatry and Biobehavioral Sciences, Radiology, Neuroscience, Biomedical Engineering Interdepartmental Program, University of California Los Angeles, 420 Westwood Plaza, #68-121, Los Angeles, CA 90095, USA.
e-mail: ljinhy@gmail.com

Despite the overwhelming need, there has been a relatively large gap in our ability to trace network level activity across the brain. The complex dense wiring of the brain makes it extremely challenging to understand cell-type specific activity and their communication beyond a few synapses. Recent development of the optogenetic functional magnetic resonance imaging (ofMRI) provides a new impetus for the study of brain circuits by enabling causal tracing of activities arising from defined cell types and firing patterns across the whole brain. Brain circuit elements can be selectively triggered based on their genetic identity, cell body location, and/or their axonal projection target with temporal precision while the resulting network response is monitored non-invasively with unprecedented spatial and temporal accuracy. With further studies including technological innovations to bring ofMRI to its full potential, ofMRI is expected to play an important role in our system-level understanding of the brain circuit mechanism.

Keywords: optogenetics, BOLD, ofMRI, opto-fMRI, connectomics

Human brains form highly complex circuits where the circuit elements communicate using electrical and/or chemical signals. It consists of approximately 100 billion neurons and many more glial cells. Furthermore, the hundreds of billions of neurons and glial cells also come in various different cell types, which can be categorized based on their shape, location, genetic properties, and the chemicals used for communication. It has been extremely difficult to spatially distinguish pertinent elements, adding complexity to the understanding of a circuit's connection topology and function. Different state-of-the art methods to understand brain circuits include microscopic approaches looking at small-scale connections with electron (Briggman and Denk, 2006; Bock et al., 2011; Briggman et al., 2011) and light microscopy (Livet et al., 2007; Micheva and Smith, 2007; Choe et al., 2008), and larger scale connection topologies using light microscopy of brain sections across the whole brain (Bohland et al., 2009), and diffusion tensor MRI (Mori and Zhang, 2006). However, one of the most important approaches in understanding complex circuits, as is often used with electronic circuits, is the ability to trigger specific circuit elements with high temporal precision while monitoring the global network response in an intact circuit. The new optogenetic (Boyden et al., 2005; Zhang et al., 2006, 2007a,b) functional magnetic resonance imaging (ofMRI; Lee et al., 2010) has the potential to enable such processes for brain circuits. Optogenetic stimulation allows for cell-type specific stimulation with high temporal precision while the network response is measured non-invasively across the whole brain. This approach allows brain circuits to be systematically analyzed in normal and diseased conditions where it not only provides tracing of activity throughout the brain but has the potential to serve as an *in vivo* quantitative biomarker that can be directly correlated with normal and diseased phenotypes.

BRAIN CIRCUIT ANALYSIS AND DEBUGGING WITH ofMRI

Optogenetics (Boyden et al., 2005; Zhang et al., 2006, 2007a,b), is a revolutionary technology in which single-component microbial light-activated trans-membrane conductance regulators are introduced into specifically targeted cell types using genetic approaches allowing millisecond-scale targeted activity modulation *in vivo* (Aravanis et al., 2007). Channelrhodopsin (ChR2) is a monovalent cation channel that allows Na+ ions to enter the cell following exposure to 470 nm blue light, whereas Halorhodopsin (NpHR) is a chloride pump that activates upon illumination with 580 nm yellow light. As the optimum activation wavelength of these two proteins are over 100 nm apart, they can be controlled independently to either initiate action potential firing or suppress neural activity in intact tissue, and together may modulate neuronal synchrony. Both proteins have fast temporal kinetics, on the scale of milliseconds, making it possible to drive reliable trains of high frequency action potentials *in vivo* using ChR2 and suppress single action potentials within high frequency spike trains using NpHR. Recent developments in optogenetics also provide a wide range of additional tools including more accurate temporal kinetics (Gunaydin et al., 2010), step function control (Berndt et al., 2009), and higher sensitivity to light (Gradinaru et al., 2008; Chow et al.,

2010). Thus far, one of the greatest challenges in neuroscience has been the difficulty of selectively controlling different circuit elements due to the dense complex wiring of many different cell types. Optogenetics, by enabling *in vivo* control of genetically targeted circuit elements, represents an exciting new opportunity for addressing these complicated issues.

The ofMRI technology (Lee et al., 2010; **Figures 1** and **2**), by combining optogenetics with fMRI, allows precise control of brain circuit elements and *in vivo* visualization of the resulting causal effects on the brain. In the first study demonstrating ofMRI (Lee et al., 2010), brain circuit elements were successfully controlled and monitored based on their genetic identity, cell body location, and axonal projection target. Selective excitation of excitatory neurons with cell body in M1 cortex resulted in robust activity measurement in local cortex (**Figure 1**) and distal areas including striatum, and thalamus (**Figure 2**). It was also demonstrated that the neural activity is more accurately mapped throughout the brain using the passband bSSFP–fMRI technique (Lee et al., 2008) compared to the conventional GRE-BOLD fMRI technique (**Figure 2**). The temporal dynamics of the fMRI signal was also shown to have strong correlations with the electrophysiological measurements indicating that the fMRI hemodynamic response accurately reflects temporal neural activity pattern (Figure 2 in Lee et al., 2010). Targeting excitatory neurons in anterior and posterior thalamus also demonstrated robust local and long-range activity consistent with the existing literature regarding network connectivity of each region (Figure 4 in Lee et al., 2010). In addition, selective excitation of axonal fibers projecting from M1 cortex within the thalamus, by selective expression of ChR2 in excitatory neurons with cell body in the M1 cortex and optical stimulation in the thalamus, showed that wiring patterns in addition to genetic identity can be used to selectively target and monitor the brain circuitry (Figure 3 in Lee et al., 2010). These findings demonstrate basic feasibility on how ofMRI defines a potent tool that is suitable for functional circuit analysis as well as global phenotyping of dysfunctional circuitry.

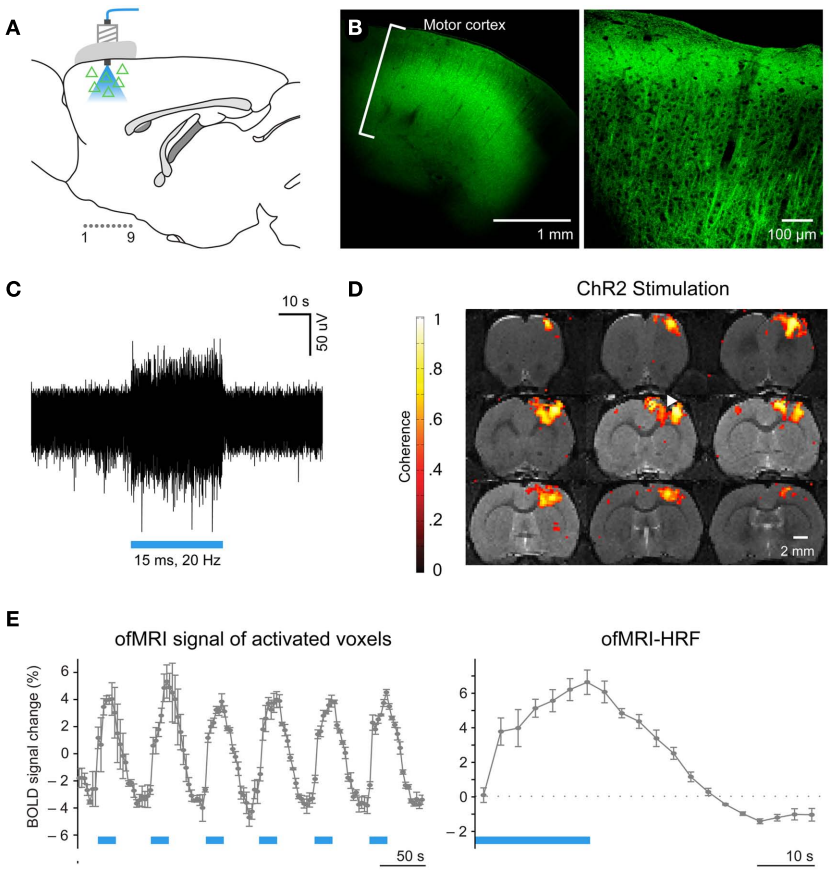


FIGURE 1 | Optogenetic functional magnetic resonance imaging enables systematic brain circuit analysis through cell-type specific stimulation and non-invasive monitoring of the activity throughout the whole brain. (A) Schematic: transduced cells (triangles) and blue light delivery shown in M1. Coronally imaged slices in **(D)** marked as "1.0.9." **(B)** Confocal images of ChR2–EYFP expression in M1. **(C)** Extracellular optrode recordings during 473 nm optical stimulation (20Hz/15 ms pulsewidth). **(D)** BOLD activation is observed at or near the site of optical stimulation in animals injected with AAV5–CaMKIIα::ChR2–EYFP (white arrowhead: injection/stimulation site). Coronal slices are consecutive and 0.5 mm thick. **(E)**, Left: ofMRI hemodynamic response during 6 epochs of optical stimulation at 20 Hz/15 ms 473 nm light stimulation repeated every 60 s (blue bars). Hemodynamic response averaged across all voxels with coherence coefficient >0.35 in motor cortex. *Right*, Mean of all stimulation epochs. (Modified Figure from Lee et al., 2010).

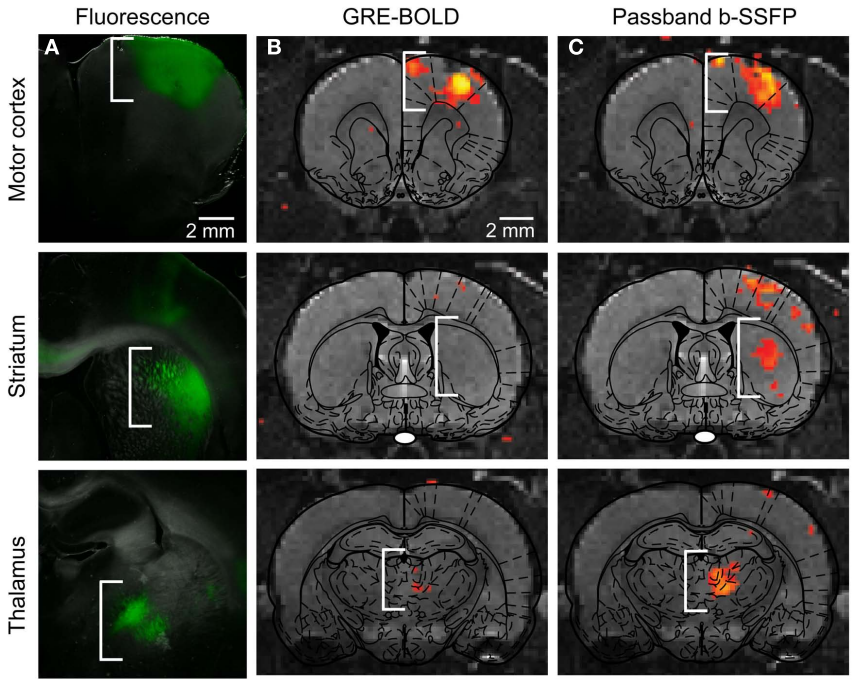


FIGURE 2 | (A) Injection of CaMKIIα::ChR2-EYFP in M1, as expected, leads to opsin expression in motor cortex, striatum, and thalamus, i.e., the primary site of injection and sites targeted by the axonal projection of expressing neurons. **(B)** Conventional BOLD fMRI activity map superimposed onto appropriate anatomical and atlas images. **(C)** Passband bSSFP-fMRI activity map superimposed onto appropriate anatomical and atlas images, which more fully captures circuit-level activity. (Modified supplemental Figure from Lee et al., 2010).

TRACING THE BRAIN CONNECTION TOPOLOGY AND ACTIVITY

To compare the nature of different information obtained using various methods to understand the connectivity of the brain, let us consider a simple model of the brain (**Figure 3A**). Assume a brain that has three distinct areas (nodes labeled A–C), with connections originating from each area with the number of connections from one node to another ranging from 1 to 4. For simplicity, let us further assume that a given node is either excitatory (red triangle) or inhibitory (blue square). We will further assume that conflicting signals received from different nodes is resolved by majority vote. For example, if four inhibitory inputs and three excitatory inputs are received, the node will be inhibited. We also assume all nodes are synchronized where τ is the time it takes for each node to trigger the next node and that external stimulation results in an immediate state change at the stimulated node, which allows us to assume a discrete time model where n is the time variable with the discrete time interval τ. While the realistic scenario is obviously much more complex than what is represented here, this simple model brain can help illustrate how different techniques address distinct pieces of the puzzle in our effort to understand the brain circuit. The model brain described above can be summarized with Eq. 1, where G is the connectivity matrix with entry values that can range from −4 to 4, X(n) and Y(n) are the stimulus input vector and the activity at a given time n with values of 0 (not active) or 1 (active). G will then be a 3 by 3 matrix, X(n) and Y(n) will be 3 by 1 vectors. u is the unit step function applied to each vector components separately.

$$Y(n+1) = u\,(GY(n) + X(n)) \quad \text{where} \quad n = 0, 1, 2 \ldots \tag{1}$$

For the model brain outlined in **Figure 3A**, the G matrix is as follows with a zero-state initial condition.

$$G = \begin{bmatrix} 0 & 0 & 0 \\ 3 & 1 & -4 \\ 2 & 1 & 0 \end{bmatrix} \tag{2}$$

$$Y(0) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{3}$$

The tracer methods (Raju and Smith, 2006) based on viral and/or chemical tracers will provide anatomical information with directionality (**Figure 3B**). For example, if an anterograde tracer that does not cross synapses is injected into node A, the injection node will be highlighted with its axonal projections showing nodes that are directly connected with node A (B, C). Different types of tracers, for example, retrograde tracers (Miyamichi et al., 2011) will give information regarding which nodes the injected region receive axonal projections from. As demonstrated with this example, the tracer-based methods have the advantage of providing information about how different areas are connected with direction and cell type information. However, the main drawback of the method is that it does not allow *in vivo* evaluation of the connection, as

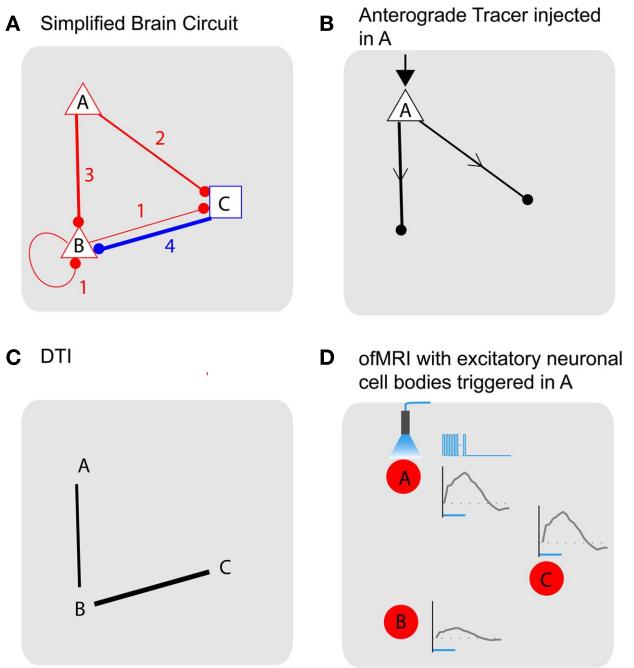


FIGURE 3 | Tracing brain pathways: simplified model of brain demonstrates nature of the different information obtained using the current state-of-the art methods. (A) Simplified model brain with three distinct areas labeled A–C. Each area has connections originating from one area connecting to another with the number of connections that range from 1 to 4. Each node is also characterized as excitatory (red triangle) or inhibitory (blue square). **(B)** Anatomical tracing with anterograde tracers that do not cross synapses. **(C)** Diffusion tensor MRI has the potential to reveal large-scale axonal connections without connection type and directional information. **(D)** ofMRI with stimulation of the excitatory neurons with cell bodies at node A.1/6τ Hz, 3τ s duration ($\tau = 4$ ms) stimulation, repeated every 1 min for 20 s duration will result in 1/6τ Hz activity in all three nodes. Node B activity will be at τ duration while A and C have 3τ duration. Node B and C activity will also be delayed by τ (**Table 1A**). The color of the node (red) represent the expected positive BOLD signal from the increased activity at all three nodes synchronized with the stimulation cycle. Blue time course on the right of node Are presents temporally encoded optogenetic stimulation and the gray time curves show the temporal firing patterns captured through ofMRI–HRF. ofMRI is expected to reflect network activity pattern determined by the cell types involved, network connection topology, and temporal stimulation pattern.

well as the fact that the number of synapses that can be crossed by existing tracers is limited. Recent developments of powerful tracers with high level of specific control (Wall et al., 2010) and manganese ion (Mn^{2+}, Pautler et al., 1998)-based MRI-sensitive *in vivo* fiber tract tracer potentially overcome some of the challenges. While individual tracer methods pose different pros and cons, the main information the tracing methods aims to obtain is the anatomical connectivity with directionality (G_{ij} vs. G_{ji}) and cell type (+, −).

Diffusion tensor imaging (DTI; Mori and Zhang, 2006) is an MR technique that relies on the restricted and/or anisotropic zero-displacement diffusion of metabolites (normally the most abundant water molecules are used) within and across different neural compartments (such as intra/extra cellular diffusion and diffusion within the myelin sheath etc.). Signal attenuation is observed whenever diffusion is present along the direction of the magnetic field gradients. The resulting diffusion ellipsoids for each voxel can then be concatenated in order to yield smooth trajectories that are assumed to be co-linear to the primary direction of the physical barriers, which cause the diffusion to be restricted/anisotropic in the first place. With DTI, large-scale connections can be measured without directionality, and without cell type information (**Figure 3C**). Due to limitations in spatial resolution and DTI technique's ability to resolve directional ambiguities when fibers cross, the information that is obtained using DTI will be coarser than those obtained using tracer methods. However, DTI allows *in vivo*, non-invasive, and whole brain scanning ability, which enables the assessment of the connections in fully intact brain with the possibility of longitudinal assessment in living humans. Ideally, DTI aims to uncover the larger values of $|G_{ij}| + |G_{ji}|$ (no direction, cell type information). In **Figure 3C**, the DTI image shows connections with $|G_{ij}| + |G_{ji}|$ values over 3 as an example.

To fully understand the brain, temporal dynamics have to be resolved in addition to the anatomical connections. Traditionally, electrophysiological stimulation and recording hold an important place in the assessment of the activity. For activity assessment in an intact circuit, electrodes are placed *in vivo* where cells at a specific location are stimulated and/or recorded. However, one of the most significant difficulties with this approach has been the lack of cell type specificity in the stimulation (Histed et al., 2009), and the limited spatial information in the readout. With the lack of cell type specificity, it is difficult to interpret the resulting signal while the limited spatial information makes it difficult to trace the activity throughout the brain. fMRI (Ogawa et al., 1992; Bandettini and Wong, 1997; Song et al., 2000), on the other hand, while being a completely non-invasive technique with whole brain spatial information, also lacks cell-type specificity in stimulation with the conventional sensory stimulations and micro-electrode based stimulations (Tolias et al., 2005). Techniques such as resting-state fMRI (Fox and Raichle, 2007), while revealing valuable, completely non-invasive information about the network level connectivity, do not provide causal information. Modeling approaches to obtain causal information include the use of granger causality (Goebel et al., 2003) and dynamic causal modeling (Stephan and Friston, 2011).

ofMRI adds valuable new information since causal communication can be directly traced throughout the brain across multiple synapses with global activity pattern information *in vivo*. For example, if node A is excited, the downstream impact of such stimulation will be visualized across the whole brain with full spatial information. Assuming a stimulation with 1/6τ Hz, 3τ s duration (where, $\tau = 4$ ms) repeated every 1 min for 20 s duration at node A [**Figure 3D**; **Table 1A**; X(n)], one can potentially expect positive BOLD signal in all three nodes, since all three nodes will have increased neural activity at 1/6τ Hz with 3τ duration at node A and C and τ duration at node B [**Table 1A**; Y(n)]. The amplitude of the ofMRI-measured response at node B could be 3 times smaller than the other nodes since it is three times less active. Alternatively, if continuous stimulation is applied at node A for 20 s every 1 min [**Table 1B**; X(n)], one can expect positive BOLD signal in only

Table 1 | Temporal dynamics of the simplified brain circuit. X(n) and Y(n) for an example stimulation of **(A)**, 1/6τ Hz, 3τ s duration (τ = 4 ms), repeated every 1 min for 20 s at node A, and **(B)**, constant stimulation repeated every 1 min for 20 s at node A. Each column represents signal value at every time interval τ.)

A

	n	0	1	2	3	4	5	6	7	8	9	10	11	12
X(n)	A	1	1	1	0	0	0	1	1	1	0	0	0	1
	B	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	0	0	0	0	0	0	0	0	0	0	0	0	0
Y(n)	A	0	1	1	1	0	0	0	1	1	1	0	0	0
	B	0	0	1	0	0	0	0	0	1	0	0	0	0
	C	0	0	1	1	1	0	0	0	1	1	1	0	0

B

	n	0	1	2	3	4	5	6	7	8	9	10	11	12
X(n)	A	1	1	1	1	1	1	1	1	1	1	1	1	1
	B	0	0	0	0	0	0	0	0	0	0	0	0	0
	C	0	0	0	0	0	0	0	0	0	0	0	0	0
Y(n)	A	0	1	1	1	1	1	1	1	1	1	1	1	1
	B	0	0	1	0	0	0	0	0	0	0	0	0	0
	C	0	0	1	1	1	1	1	1	1	1	1	1	1

node A and C since node B will not be active except for a mere τ s at the beginning of the 20-s stimulation [**Table 1B**; Y(n)].

As illustrated through this simple example, temporal encoding of the stimulation, in addition to the anatomical connections and cell type, is expected to determine how activity propagates throughout the brain. ofMRI, for the first time, offers the potential to trace such activity throughout the whole brain with temporal accuracy. While the exact relationship between the neural activity and the observed ofMRI signal remains elusive, initial studies show that neural activity patterns are strongly correlated with the ofMRI hemodynamic response function (HRF; Figure 2 in Lee et al., 2010).

DISCUSSION

While the published study (Lee et al., 2010) shows great promise in terms of the ofMRI technology's capability to accurately trace neural activity across the whole brain, further investigation on neurovascular coupling is necessary to accurately trace neural activity. While the causal roles of the excitatory neurons have been investigated, causal roles of other cell types need to be further studied. In addition, triggering of any specific cell types will result in a cascade of activity in many different cell types that are connected to the primary neuron that is being modulated. Therefore, in order to study the one to one relationship between specific cell types and the resulting HRF, combinatorial approach of excitation and inhibition using ChR2 and NpHR will be necessary (see Pearl, 2002 for the use of inhibition to infer causality). For example, in order to study the role of excitatory neurons in generating the BOLD signal in the absence of activity in any other cell types, one can stimulate the excitatory neurons with ChR2 while inhibiting all other cell types using NpHR. Furthermore, ofMRI signal most likely represents a group average of neural activity since each imaging voxel is much larger than individual neurons. Each voxel can contain both excitation and inhibition. Therefore, investigating the relationship between how different mixture of activity is reflected in the ofMRI signal will be of great importance.

In addition to uncovering the cell-type specific roles in generating distinct HRFs, one can also utilize the temporally accurate control enabled by optogenetics to investigate the relationship between temporal signaling and the corresponding HRF. Temporal encoding will most likely change the local HRF as well as the signal propagation throughout the neural network. Once the relationship between the signaling pattern and the ofMRI signal is revealed, ofMRI studies can be used to accurately infer the signaling pattern in addition to the location of the traced activity. While initial studies hint at the possibility of inferring such temporal information, further studies to understand the relationship with experiment designs specifically targeted to understand the relationship will be necessary to bring this potential capability to its full potential.

In the efforts to trace the neural activity, distinguishing activity resulting from the first, second, and higher order connectivity is of great interest. Due to the relatively slow time scale provided by fMRI studies, which is on the order of seconds, activity cannot be traced at the timescale of the neural activity. However, careful experiment designs involving anatomical knowledge, particular design of stimulation patterns, and the study of the resulting shape, delay of the HRF can potentially provide valuable information.

Additional considerations that need to be taken into account include the use of anesthesia. Initial experiments were conducted with light isoflurane anesthesia. Anesthesia is known to have an impact on the neural activity, its propagation and the resulting HRF. Precisely how it impacts the system and whether it is of importance will depend on the system under investigation. While anesthesia is likely to bring about changes in neural response, whether the specific change impacts the conclusion of a study can vary. Further studies utilizing different anesthesia and awake animals will be of crucial importance to fully understand the impact of anesthesia on ofMRI signals.

Furthermore, while the initially published study demonstrates 0.5 mm × 0.5 mm × 0.5 mm resolution scans covering most of the brain volume, future technological developments can potentially enable ofMRI at the cortical layer and sub-nucleus specific level while covering the whole brain. Combining ofMRI approaches with other methods including anatomical connectomics data, modeling approach, resting-state fMRI, electrophysiological studies, optical imaging, and behavioral readouts will also be a powerful new synergistic approach to study brain circuits.

CONCLUSION

Increasing data suggest that the brain functions as an integrated circuit with network communication across diverse brain circuit elements and that disorders result in network level dysfunction of the brain (Cardin et al., 2009; Gradinaru et al., 2009; Kravitz et al., 2010). With ofMRI, large-scale communication arising from specifically targeted neural populations can be mapped across the whole brain with spatial and temporal accuracy. This will add an

important new dimension to the anatomical connectomics project (Sporns et al., 2005) by visualizing how information propagates through the network. ofMRI based, cell-type specific, activity-dependent brain atlases can also provide a new angle to understanding the brain architecture by providing additional information in the efforts to segment different brain regions (Hilgetag et al., 2000) and in understanding the relationship between the segmented regions (Stephan et al., 2000; Honey et al., 2007).

ACKNOWLEDGMENTS

Jin Hyung Lee is supported by the NIH/NIBIB R00 Pathway to Independence Award, the NIH Director's New Innovator Award, the Okawa Foundation Research grant award, and the NSF CAREER award. Jin Hyung Lee would like to acknowledge Dr. Dae-Shik Kim for discussions and contributions on DTI and other MRI based methods and Dr. Partha Mitra for helpful discussions regarding connectomics and network dynamics.

REFERENCES

Aravanis, A. M., Wang, L. P., Zhang, F., Meltzer, L. A., Mogri, M. Z., Schneider, M. B., and Deisseroth, K. (2007). An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.* 4, S143–S156.

Bandettini, P. A., and Wong, E. C. (1997). Magnetic resonance imaging of human brain function. Principles, practicalities, and possibilities. *Neurosurg. Clin. N. Am.* 8, 345–371.

Berndt, A., Yizhar, O., Gunaydin, L. A., Hegemann, P., and Deisseroth, K. (2009). Bi-stable neural state switches. *Nat. Neurosci.* 12, 229–234.

Bock, D. D., Lee, W. C., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S., and Reid, R. C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature* 471, 177–182.

Bohland, J. W., Wu, C., Barbas, H., Bokil, H., Bota, M., Breiter, H. C., Cline, H. T., Doyle, J. C., Freed, P. J., Greenspan, R. J., Haber, S. N., Hawrylycz, M., Herrera, D. G., Hilgetag, C. C., Huang, Z. J., Jones, A., Jones, E. G., Karten, H. J., Kleinfeld, D., Kotter, R., Lester, H. A., Lin, J. M., Mensh, B. D., Mikula, S., Panksepp, J., Price, J. L., Safdieh, J., Saper, C. B., Schiff, N. D., Schmahmann, J. D., Stillman, B. W., Svoboda, K., Swanson, L. W., Toga, A. W., Van Essen, D. C., Watson, J. D., and Mitra, P. P. (2009). A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLoS Comput. Biol.* 5, e1000334. doi: 10.1371/journal.pcbi.1000334

Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* 8, 1263–1268.

Briggman, K. L., and Denk, W. (2006). Towards neural circuit reconstruction with volume electron microscopy techniques. *Curr. Opin. Neurobiol.* 16, 562–570.

Briggman, K. L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature* 471, 183–188.

Cardin, J. A., Carlen, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L. H., and Moore, C. I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 459, 663–667.

Choe, Y., Abbott, L. C., Han, D., Huang, P.-S., Keyser, J., Kwon, J., Mayerich, D., Melek, Z., and McCormick, B. H. (2008). "Knife-edge scanning microscopy: high-throughput imaging and analysis of massive volumes of biological microstructures," in *High-Throughput Image Reconstruction and Analysis: Intelligent Microscopy Applications*, eds R. Rao and G. Cecchi (Boston, MA: Artech House), 11–37.

Chow, B. Y., Han, X., Dobry, A. S., Qian, X., Chuong, A. S., Li, M., Henninger, M. A., Belfort, G. M., Lin, Y., Monahan, P. E., and Boyden, E. S. (2010). High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* 463, 98–102.

Fox, M. D., and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711.

Goebel, R., Roebroeck, A., Kim, D. S., and Formisano, E. (2003). Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn. Reson. Imaging* 21, 1251–1261.

Gradinaru, V., Mogri, M., Thompson, K. R., Henderson, J. M., and Deisseroth, K. (2009). Optical deconstruction of parkinsonian neural circuitry. *Science* 324, 354–359.

Gradinaru, V., Thompson, K. R., and Deisseroth, K. (2008). eNpHR: a *Natronomonas* halorhodopsin enhanced for optogenetic applications. *Brain Cell Biol.* 36, 129–139.

Gunaydin, L. A., Yizhar, O., Berndt, A., Sohal, V. S., Deisseroth, K., and Hegemann, P. (2010). Ultrafast optogenetic control. *Nat. Neurosci.* 13, 387–392.

Hilgetag, C. C., Burns, G. A., O'Neill, M. A., Scannell, J. W., and Young, M. P. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 91–110.

Histed, M. H., Bonin, V., and Reid, R. C. (2009). Direct activation of sparse, distributed populations of cortical neurons by electrical microstimulation. *Neuron* 63, 508–522.

Honey, C. J., Kotter, R., Breakspear, M., and Sporns, O. (2007). Network structure of cerebral cortex shapes functional connectivity on multiple time scales. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10240–10245.

Kravitz, A. V., Freeze, B. S., Parker, P. R., Kay, K., Thwin, M. T., Deisseroth, K., and Kreitzer, A. C. (2010). Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* 466, 622–626.

Lee, J. H., Dumoulin, S. O., Saritas, E. U., Glover, G. H., Wandell, B. A., Nishimura, D. G., and Pauly, J. M. (2008). Full-brain coverage and high-resolution imaging capabilities of passband b-SSFP fMRI at 3T. *Magn. Reson. Med.* 59, 1099–1110.

Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D. S., Fenno, L. E., Ramakrishnan, C., and Deisseroth, K. (2010). Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature* 465, 788–792.

Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., and Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 450, 56–62.

Micheva, K. D., and Smith, S. J. (2007). Array tomography: a new tool for imaging the molecular architecture and ultrastructure of neural circuits. *Neuron* 55, 25–36.

Miyamichi, K., Amat, F., Moussavi, F., Wang, C., Wickersham, I., Wall, N. R., Taniguchi, H., Tasic, B., Huang, Z. J., He, Z., Callaway, E. M., Horowitz, M. A., and Luo, L. (2011). Cortical representations of olfactory input by trans-synaptic tracing. *Nature* 472, 191–196.

Mori, S., and Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron* 51, 527–539.

Ogawa, S., Tank, D. W., Menon, R., Ellermann, J. M., Kim, S. -G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5951–5955.

Pautler, R. G., Silva, A. C., and Koretsky, A. P. (1998). In vivo neuronal tract tracing using manganese-enhanced magnetic resonance imaging. *Magn. Reson. Med.* 40, 740–748.

Pearl, J. (2002). Reasoning with Cause and Effect. *AI Mag.* 23, 95–111.

Raju, D. V., and Smith, Y. (2006). Anterograde axonal tract tracing. *Curr. Protoc. Neurosci.* Chap. 1, Unit 1.14.

Song, A. W., Popp, C. A., Mao, J., and Dixon, W. T. (2000). fMRI: methodology – acquisition and processing. *Adv. Neurol.* 83, 177–185.

Sporns, O., Tononi, G., and Kotter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi:10.1371/journal.pcbi.0010042

Stephan, K. E., and Friston, K. J. (2011). Analyzing effective connectivity with fMRI. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 446–459.

Stephan, K. E., Hilgetag, C. C., Burns, G. A., O'Neill, M. A., Young, M. P., and Kotter, R. (2000). Computational analysis of functional connectivity between areas of primate cerebral cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 111–126.

Tolias, A. S., Sultan, F., Augath, M., Oeltermann, A., Tehovnik, E. J., Schiller, P. H., and Logothetis, N. K. (2005). Mapping cortical activity elicited with electrical microstimulation using FMRI in the macaque. *Neuron* 48, 901–911.

Wall, N. R., Wickersham, I. R., Cetin, A., De La Parra, M., and Callaway, E. M. (2010). Monosynaptic circuit tracing in vivo through Cre-dependent targeting and complementation of modified rabies virus. *Proc. Natl. Acad. Sci. U.S.A.* 107, 21848–21853.

Zhang, F., Aravanis, A. M., Adamantidis, A., de Lecea, L., and Deisseroth, K. (2007a). Circuit-breakers: optical technologies for probing neural signals and systems. *Nat. Rev. Neurosci.* 8, 577–581.

Zhang, F., Wang, L. P., Brauner, M., Liewald, J. F., Kay, K., Watzke, N., Wood, P. G., Bamberg, E., Nagel, G., Gottschalk, A., and Deisseroth, K. (2007b). Multimodal fast optical interrogation of neural circuitry. *Nature* 446, 633–639.

Zhang, F., Wang, L. P., Boyden, E. S., and Deisseroth, K. (2006). Channelrhodopsin-2 and optical control of excitable cells. *Nat. Methods* 3, 785–792.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 March 2011; accepted: 13 September 2011; published online: 25 October 2011.

Citation: Lee JH (2011) Tracing activity across the whole brain neural network with optogenetic functional magnetic resonance imaging. Front. Neuroinform. ***5****:21. doi: 10.3389/fninf.2011.00021*

frontiers in
NEUROINFORMATICS

TECHNOLOGY REPORT
published: 27 June 2011
doi: 10.3389/fninf.2011.00004

Informatics and data mining tools and strategies for the Human Connectome Project

Daniel S. Marcus[1]*, John Harwell[2], Timothy Olsen[1], Michael Hodge[1], Matthew F. Glasser[2], Fred Prior[1], Mark Jenkinson[3], Timothy Laumann[4], Sandra W. Curtiss[2] and David C. Van Essen[2†]

[1] *Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA*
[2] *Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO, USA*
[3] *Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, University of Oxford, John Radcliffe Hospital, Oxford, UK*
[4] *Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA*

Edited by:
Trygve B. Leergaard, University of Oslo, Norway

Reviewed by:
Jan G. Bjaalie, University of Oslo, Norway
Russell A Poldrack, University of California, USA

****Correspondence:***
Daniel S. Marcus, Washington University School of Medicine, 4525 Scott Avenue, Campus Box 8225, St. Louis, MO, USA.
e-mail: dmarcus@wustl.edu

†David C. Van Essen for the WU-Minn HCP Consortium

The Human Connectome Project (HCP) is a major endeavor that will acquire and analyze connectivity data plus other neuroimaging, behavioral, and genetic data from 1,200 healthy adults. It will serve as a key resource for the neuroscience research community, enabling discoveries of how the brain is wired and how it functions in different individuals. To fulfill its potential, the HCP consortium is developing an informatics platform that will handle: (1) storage of primary and processed data, (2) systematic processing and analysis of the data, (3) open-access data-sharing, and (4) mining and exploration of the data. This informatics platform will include two primary components. ConnectomeDB will provide database services for storing and distributing the data, as well as data analysis pipelines. Connectome Workbench will provide visualization and exploration capabilities. The platform will be based on standard data formats and provide an open set of application programming interfaces (APIs) that will facilitate broad utilization of the data and integration of HCP services into a variety of external applications. Primary and processed data generated by the HCP will be openly shared with the scientific community, and the informatics platform will be available under an open source license. This paper describes the HCP informatics platform as currently envisioned and places it into the context of the overall HCP vision and agenda.

Keywords: connectomics, Human Connectome Project, XNAT, caret, resting state fMRI, diffusion imaging, network analysis, brain parcellation

INTRODUCTION

The past decade has seen great progress in the refinement of non-invasive neuroimaging methods for assessing long-distance connections in the human brain. This has given rise to the tantalizing prospect of systematically characterizing human brain connectivity, i.e., mapping the connectome (Sporns et al., 2005). The eventual elucidation of this amazingly complex wiring diagram should reveal much about what makes us uniquely human and what makes each person different from all others.

The NIH recently funded two consortia under the Human Connectome Project (HCP)[1]. One is led by Washington University and University of Minnesota and involves seven other institutions (the "WU-Minn HCP consortium")[2]. The other, led by Massachusetts General Hospital and UCLA (the MGH/UCLA HCP consortium), focuses on building and refining a next-generation 3T MR scanner for improved sensitivity and spatial resolution. Here, we discuss informatics aspects of the WU-Minn HCP consortium's plan to map human brain circuitry in 1,200 healthy young adults using cutting-edge non-invasive neuroimaging methods. Key imaging modalities will include diffusion imaging, resting-state fMRI, task-evoked fMRI, and magnetoencephalography combined with electroencephalography (MEG/EEG). A battery of behavioral and cognitive tests will also be included along with the collection of genetic material. This endeavor will yield valuable information about brain connectivity, its relationship to behavior, and the contributions of genetic and environmental factors to individual differences in brain circuitry. The data generated by the WU-Minn HCP consortium will be openly shared with the scientific community.

The HCP has a broad informatics vision that includes support for the acquisition, analysis, visualization, mining, and sharing of connectome-related data. As it implements this agenda, the consortium seeks to engage the neuroinformatics community through open source software, open programming interfaces, open-access data-sharing, and standards-based development. The HCP informatics approach includes three basic domains.

- *Data support* components include tools and services that manage data (e.g., data uploads from scanners and other data collection devices); execution and monitoring of quality assurance, image processing, and analysis pipelines and routines; secure long-term storage of acquired and processed data; search services to identify and select subsets of the data; and download mechanisms to distribute data to users around the globe.

[1]http://humanconnectome.org/consortia/
[2]http://humanconnectome.org/

- *Visualization* components include a spectrum of tools to view anatomic and functional brain data in volumetric and surface representations and also using network and graph-theoretic representations of the connectome.
- *Discovery* components are an especially important category of the HCP's informatics requirements, including user interfaces (UI) for formulating database queries, linking between related knowledge/database systems, and exploring the relationship of an individual's connectome to population norms.

The HCP is expected to generate approximately 1 PB of data, which will be made accessible via a tiered data-sharing strategy. Besides the sheer amount of data, there will be major challenges associated with handling the diversity of data types derived from the various modalities of data acquisition, the complex analysis streams associated with each modality, and the need to cope with individual variability in brain shape as well as brain connectivity, which is especially dramatic for cerebral cortex.

To support these needs, the HCP is developing a comprehensive informatics platform centered on two interoperable components: *ConnectomeDB*, a data management system, and *Connectome Workbench* (CWB), a software suite that provides visualization and discovery capabilities.

ConnectomeDB is based on the XNAT imaging informatics platform, a widely used open source system for managing and sharing imaging and related data (Marcus et al., 2007)[3]. XNAT includes an open web services application programming interface (API) that enables external client applications to query and exchange data with XNAT hosts. This API will be leveraged within the HCP informatics platform and will also help externally developed applications connect to the HCP. CWB is based on Caret software, a visualization and analysis platform that handles structural and functional data represented on surfaces and volumes and on individuals and atlases (Van Essen et al., 2001). The HCP also benefits from a variety of processing and analysis software tools, including FreeSurfer, FSL, and FieldTrip.

Here, we provide a brief overview of the HCP, then describe the HCP informatics platform in some detail. We also provide a sampling of the types of scientific exploration and discovery that it will enable.

OVERVIEW OF THE HUMAN CONNECTOME PROJECT

INFERRING LONG-DISTANCE CONNECTIVITY FROM *IN VIVO* IMAGING

The two primary modalities for acquiring information about human brain connectivity *in vivo* are diffusion imaging (dMRI), which provides information about structural connectivity, and resting-state functional MRI (R-fRMI), which provides information about functional connectivity. The two approaches are complementary, and each is very promising. However, each has significant limitations that warrant brief comment.

Diffusion imaging relies on anisotropies in water diffusion to determine the orientation of fiber bundles within white matter. Using High Angular Resolution Diffusion Imaging (HARDI), multiple fiber orientations can be identified within individual voxels. This enables tracking of connections even in regions where multiple fiber bundles cross one another. Probabilistic tractography integrates information throughout the white matter and can reveal detailed information about long-distance connectivity patterns between gray-matter regions (Johansen-Berg and Behrens, 2009; Johansen-Berg and Rushworth, 2009). However, uncertainties arising at different levels of analysis can lead to both false positives and false negatives in tracking connections. Hence, it is important to continue refining the methods for dMRI data acquisition and analysis.

R-fMRI is based on spatial correlations of the slow fluctuations in the BOLD fMRI signal that occur at rest or even under anesthesia (Fox and Raichle, 2007). Studies in the macaque monkey demonstrate that R-fMRI correlations tend to be strong for regions known to be anatomically interconnected, but that correlations can also occur between regions that are linked only indirectly (Vincent et al., 2007). Thus, while functional connectivity maps are not a pure indicator of anatomical connectivity, they represent an invaluable measure that is highly complementary to dMRI and tractography, especially when acquired in the same subjects.

The HCP will carry out a "macro-connectome" analysis of long-distance connections at a spatial resolution of 1–2 mm. At this scale, each gray-matter voxel contains hundreds of thousands of neurons and hundreds of millions of synapses. Complementary efforts to chart the "micro-connectome" at the level of cells, dendrites, axons, and synapses aspire to reconstruct domains up to a cubic millimeter (Briggman and Denk, 2006; Lichtman et al., 2008), so that the macro-connectome and micro-connectome domains will barely overlap in their spatial scales.

A TWO-PHASE HCP EFFORT

Phase I of the 5-year WU-Minn HCP consortium grant is focused on additional refinements and optimization of data acquisition and analysis stages and on implementing a robust informatics platform. Phase II, from mid-2012 through mid-2015, will involve data acquisition from the main cohort of 1,200 subjects as well as continued refinement of the informatics platform and some analysis methods. This section summarizes key HCP methods relevant to the informatics effort and describes some of the progress already made toward Phase I objectives. A more detailed description of our plans will be published elsewhere.

SUBJECT COHORT

We plan to study 1,200 subjects (300 healthy twin pairs and available siblings) between the ages of 22 and 35. This design, coupled with collection of subjects' DNA, will yield invaluable information about (i) the degree of heritability associated with specific components of the human brain connectome; and (ii) associations of specific genetic variants with these components in healthy adults. It will also enable genome-wide testing for additional associations (e.g., Visscher and Montgomery, 2009).

IMAGING

All 1,200 subjects will be scanned at Washington University on a dedicated 3 Tesla (3T) Siemens Skyra scanner. The scanner will be customized to provide a maximum gradient strength of ~100 mT/m, more than twice the standard 40 mT/m for the Skyra. A subset of 200 subjects will also be scanned at the University of Minnesota using a new 7T scanner, which is

[3]http://www.xnat.org

expected to provide improved signal-to-noise ratio and better spatial resolution, but is less well established for routine, high-throughput studies. Some subjects may also be scanned on a 10.5 T scanner currently under development at the University of Minnesota. Having higher-field scans of individuals also scanned at 3T will let us use the higher-resolution data to constrain and better interpret the 3T data.

Each subject will have multiple MR scans, including HARDI, R-fMRI (Resting-state fMRI), T-fMRI (task-evoked fMRI), and standard T1-weighted and T2-weighted anatomical scans. Advances in pulse sequences are planned in order to obtain the highest resolution and quality of data possible in a reasonable period of time. Already, new pulse sequences have been developed that accelerate image acquisition time (TR) by sevenfold while maintaining or even improving the signal-to-noise ratio (Feinberg et al., 2010). The faster temporal resolution for both R-fMRI and T-fMRI made possible by these advances will increase the amount of data acquired for each subject and increase the HCP data storage requirements, a point that exemplifies the many interdependencies among various HCP project components.

Task-fMRI scans will include a range of tasks aimed at providing broad coverage of the brain and identifying as many functionally distinct parcels as possible. The results will aid in validating and interpreting the results of the connectivity analyses obtained using resting-state fMRI and diffusion imaging. These "functional localizer" tasks will include measures of primary sensory processes (e.g., vision, motor function) and a wide range of cognitive and affective processes, including stimulus category representations, working memory, episodic memory, language processing, emotion processing, decision-making, reward processing and social cognition. The specific tasks to be included are currently being piloted; final task selection will be based on multiple criteria, including sensitivity, reliability and brain coverage.

A subset of 100 subjects will also be studied with combined MEG/EEG, which provides vastly better temporal resolution (milliseconds instead of seconds) but lower spatial resolution than MR (between 1 and 4 cm). Mapping MEG/EEG data to cortical sources will enable electrical activity patterns among neural populations to be characterized as functions of both time and frequency. As with the fMRI, MEG/EEG will include both resting-state and task-evoked acquisitions. The behavioral tasks will be a matched subset of the tasks used in fMRI. The MEG/EEG scans, to will be acquired at St. Louis University using a Magnes 3600 MEG (4DNeuroimaging, San Diego, CA, USA) with 248 magnetometers, 23 MEG reference channels (5 gradiometer, and 18 magnetometer) and 64 EEG voltage channels. This data will be analyzed in both sensor space and using state-of-the-art source localization methods (Wipf and Nagarajan, 2009; Ou et al., 2010) and using subject specific head models derived from anatomic MRI. Analyses of band-limited power (BLP) will provide measures that reflect the frequency-dependent dynamics of resting and task-evoked brain activity (de Pasquale et al., 2010; Scheeringa et al., 2011).

BEHAVIORAL, GENETIC, AND OTHER NON-IMAGING MEASURES

Measuring behavior in conjunction with mapping of structural and functional networks in HCP subjects will enable the analysis of the functional correlates of variations in "typical" brain connectivity and function. It will also provide a starting point for future studies that examine how abnormalities in structural and functional connectivity play a role in neurological and psychiatric disorders.

The HCP will use a battery of reliable and well-validated measures that assess a wide range of human functions, including cognition, emotion, motor and sensory processes, and personality. The core of this battery will be from the NIH Toolbox for Assessment of Neurological and Behavioral function[4]. This will enable federation of HCP data with other large-scale efforts to acquire neuroimaging and behavioral data and will facilitate comparison of brain-behavior relationships across studies (Gershon et al., 2010). Additional tests that are currently being piloted will be drawn from other sources.

GENETIC ANALYSES

Blood samples collected from each subject during their visit will be sent to the Rutgers University Cell and DNA Repository (RUCDR), where cell lines will be created and DNA will be extracted. Genetic analysis will be conducted in early 2015, after all Phase II subjects have completed in-person testing. Performing the genotyping in the later stages of the project will allow the HCP to take advantage of future developments in this rapidly advancing field, including the availability of new sequencing technologies and decreased costs of whole-genome sequencing. Genetic data and de-identified demographic and phenotype data will be entered into the dbGAP database in accordance with NIH data-sharing policies. Summary data look-up by genotype will be possible via ConnectomeDB.

STUDY WORKFLOW

The collection of this broad range of data types from multiple family groups will necessitate careful coordination of the various tests during in-person visits. **Figure 1** illustrates the data collection workflow planned for the high-throughput phase of the HCP. All 1,200 subjects in the main cohort will be scanned at Washington University on the dedicated 3T scanner. A subset of 200 subjects (100 same-sex twin pairs, 50% monozygotic) will also be scanned at University of Minnesota using 7T MRI (HARDI, R-fMRI, and T-fMRI) and possibly also 10.5 T. Another subset of 100 (50 same-sex twin pairs, all monozygotic) will be scanned at St. Louis University (SLU) using MEG/EEG. Many data management and quality control (QC) steps will be taken to maximize the quality and reliability of these datasets (see Data Workflow and Quality Control sections).

THE HCP INFORMATICS APPROACH

Our HCP informatics approach includes components related to data support and visualization. The Section "Data Support" discusses key data types and representations plus aspects of data processing pipelines that have major informatics implications. This leads to a discussion of ConnectomeDB and the computational resources and infrastructure needed to support it, as well as our data-sharing plans. The Section "Visualization" describes CWB and its interoperability with ConnectomeDB. These sections also include examples of potential exploratory uses of HCP data.

[4]www.nihtoolbox.org

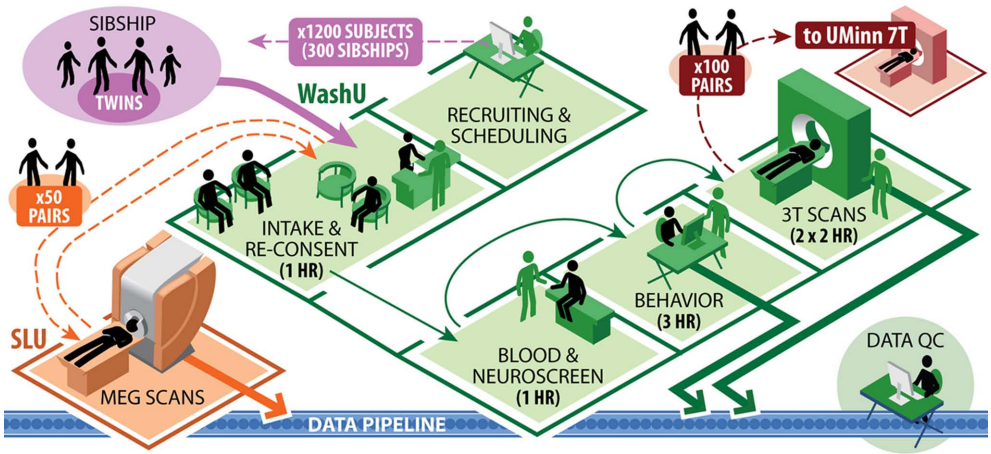


FIGURE 1 | HCP subject workflow.

DATA SUPPORT

Data types

Volumes, surfaces, and representations. MR images are acquired in a 3-D space of regularly spaced voxels, but the geometric representations useful for subsequent processing depend upon brain structure. Subcortical structures are best processed in standard volumetric (voxel) coordinates. The complex convolutions of the cortical sheet make it advantageous for many purposes to model the cortex using explicit surface representations – a set of vertices topologically linked into a 2D mesh for each hemisphere. However, for other purposes it remains useful to analyze and visualize cortical structures in volume space. Hence, the HCP will support both volumetric and surface representations for analysis and visualization.

For some connectivity data types, it is useful to represent subcortical volumetric coordinates and cortical surface vertices in a single file. This motivates introduction of a geometry-independent terminology. Specifically, a "*brainordinate*" (brain coordinate) is a spatial location within the brain that can be either a voxel (*i*, *j*, *k* integer values) or a surface vertex (*x*, *y*, *z* real-valued coordinates and a "node number"); a "*grayordinate*" is a voxel or vertex within gray matter (cortical or subcortical); a "*whiteordinate*" is a voxel within white matter or a vertex on the white matter surface. These terms (brainordinate, grayordinate, and whiteordinate) are especially useful in relation to the CIFTI data files described in the next paragraph.

When feasible, the HCP will use standard NIFTI-1 (volumetric) and GIFTI (surfaces) formats. Primary diffusion imaging data will be stored using the format MiND recently developed by Patel et al. (2010). By conforming to these existing formats, datasets generated using one software platform can be read by other platforms without the need to invoke file conversion utilities. Several types of connectivity-related data will exceed the size limits supported by NIFTI-1 and GIFTI and will instead use the recently adopted NIFTI-2 format[5]. NIFTI-2 is similar to NIFTI-1, but has dimension indices increased from 16-bit to 64-bit integers, which will be useful for multiple purposes and platforms. For the HCP, connectivity or time-series values will be stored in the binary portion of the NIFTI-2 format. Datasets whose brainordinates include both voxels and surface vertices pose special metadata requirements that are being addressed for the HCP and for other software platforms by a "CIFTI" working group (with "C" indicating connectivity). A description of CIFTI data types including example file formats has been reviewed by domain experts and is available for public comment[6]. CIFTI file formats will support metadata that map matrix rows and columns to brainordinates, parcels (see below), and/or time points, in conformance with NIFTI conventions for header extensions.

Individuals, atlases, and registration. The anatomical substrates on which HCP data are analyzed and visualized will include individual subjects as well as atlases. In general, quantitative comparisons across multiple subjects require registering data from individuals to an atlas. Maximizing the quality of inter-subject registration (alignment) is a high priority but also a major challenge. This is especially the case for cerebral cortex, owing to the complexity and variability of its convolutions. Several registration methods and atlases are under consideration for the HCP, including population-average volumes and population-average cortical surfaces based on registration of surface features. Major improvements in inter-subject alignment may be attainable by invoking constraints related to function, architecture, and connectivity, especially for cerebral cortex (e.g., Petrovic et al., 2007; Sabuncu et al., 2010). This is important for the HCP informatics effort, insofar as improved atlas representations that emerge in Phase II may warrant support by the HCP.

Parcellations. The brain can be subdivided into many subcortical nuclei and cortical areas ("*parcels*"), each sharing common characteristics based on architectonics, connectivity, topographic organization, and/or function. Expression of connectivity data as a matrix of connection weights between parcels will enable data to be stored very compactly and transmitted rapidly. Also,

[5]http://www.nitrc.org/forum/message.php?msg_id = 3738

[6]http://www.nitrc.org/projects/cifti

graph-theoretic network analyses (see below) will be more tractable and biologically meaningful on parcellated data. However, this will place a premium on the fidelity of the parcellation schemes. Data from the HCP should greatly improve the accuracy with which the brain can be subdivided, but over a time frame that will extend throughout Phase II. Hence, just as for atlases, improved parcellations that emerge in Phase II may warrant support by the HCP.

Networks and modularity. Brain parcels can often be grouped into spatially distributed networks and subnetworks that subserve distinct functions. These can be analyzed using graph-theoretic approaches that model networks as nodes connected by edges (Sporns, 2010). In the context of HCP, graph nodes can be brainordinates or parcels, and edges can be R-fMRI correlations (full correlations or various types of partial correlations), tractography-based estimates of connection probability or strength, or other measures of relationships between the nodes. The HCP will use several categories of network-related measures, including measures of segregation such as clustering and modularity (Newman, 2006); measures of integration, including path length and global efficiency; and measures of influence to identify subsets of nodes and edges central to the network architecture such as hubs or bridges (Rubinov and Sporns, 2010).

Processing pipelines and analysis streams. Generation of the various data types for each of the major imaging modalities will require extensive processing and analysis. Each analysis stream needs to be carried out in a systematic and well-documented way. For each modality, a goal is to settle on customized processing streams that yield the highest-quality and most informative types of data. During Phase I, this will include systematic evaluation of different pipelines and analysis strategies applied to the same sets of preliminary data. Minimally processed versions for each data modality will also remain available, which will enable investigators to explore alternative processing and analysis approaches.

ConnectomeDB

XNAT foundation. ConnectomeDB is being developed as a customized version of the XNAT imaging informatics platform (Marcus et al., 2007). XNAT is a highly extensible, open source system for receiving, archiving, managing, processing, and sharing both imaging and non-imaging study data. XNAT includes five services that are critical for ConnectomeDB operations. The *DICOM Service* receives and stores data from DICOM devices (scanners or gateways), imports relevant metadata from DICOM tags to the database, anonymizes sensitive information in the DICOM files, and converts the images to NIFTI formatted files. The *Pipeline Service* for defining and executing automated and semi-automated image processing procedures allows computationally intensive processing and analysis jobs to be offloaded to compute clusters while managing, monitoring and reporting on the execution status of these jobs through its application interface. The *Quality Control Service* enables both manual and automated review of images and subsequent markup of specific characteristics (e.g., motion artifacts, head positioning, signal to noise ratio) and overall usability of individual scans and full imaging sessions. The *Data Service* allows study data to be incorporated into the database. The default data model includes a standard experiment hierarchy, including projects, subjects, visits, and experiments. On top of this basic hierarchy, specific data type extensions can be added to represent specific data, including imaging modalities, derived imaging measures, behavioral tests, and genetics information. The Data Service provides mechanisms for incorporating these extensions into the XNAT infrastructure, including the database backend, middleware data access objects, and frontend reports and data entry forms. Finally, the *Search Service* allows complex queries to be executed on the database.

All of XNAT's services are accessible via an open web services API that follows the REpresentational State Transfer (REST) approach (Fielding, 2000). By utilizing the richness of the HTTP protocol, REST web services allow requests between client and server to be specified using browser-like URLs. The REST API provides specific URLs to create, access, and modify every resource under XNAT's management. The URL structures follow the organizational hierarchy of XNAT data, making it intuitive to navigate the API either manually (rarely) or programmatically. HCP will use this API for interactions between ConnectomeDB and CWB, for importing data into and out of processing pipelines, and as a conduit between external software applications and HCP datasets. External libraries and tools that can interact with the XNAT API include pyxnat – a Python library for interfacing with XNAT repositories[7]; 3D Slicer – an advanced image visualization and analysis environment[8]; and LONI Pipeline – a GUI-based pipelining environment[9].

API extensions. The HCP is developing additional services to support connectome-related queries. A primary initial focus is on a service that enables spatial queries on connectivity measures. This service will calculate and return a connectivity map or a task-evoked activation map based on specified spatial, subject, and calculation parameters. The *spatial parameter* will allow queries to specify the spatial domain to include in the calculation. Examples include a single brainordinate (see above), a cortical or subcortical parcel, or some other region of interest (collection of brainordinates). This type of search will benefit from registering each subject's data onto a standard surface mesh and subcortical atlas parcellation. The *subject parameter* will allow queries to specify the subject or subject groups to include in the calculation examples including an individual subject ID, one or more lists of subject IDs, subject characteristics (e.g., subjects with IQ > 120, subjects with a particular genotype at a particular genetic locus), and contrasts (e.g., subjects with IQ > 110 vs. subjects with IQ < 90). Finally, the *calculation parameter* will allow queries to specify the specific connectivity or task-evoked activation measure to calculate and return. Basic connectivity measures will include those based on resting-state fMRI (functional connectivity) and diffusion imaging (structural connectivity). Depending on the included subject parameter, the output connectivity measure might be the individual connectivity maps for a specific subject, the average map for a group of subjects, or the average difference map between two groups. When needed, the requested connectivity information

[7]http://packages.python.org/pyxnat/

[8]http://slicer.org

[9]http://www.loni.ucla.edu

(e.g., average difference maps) will be dynamically generated. Task-evoked activation measures will include key contrasts for each task and options to they view activation maps for a particular task in a specific subject, the average map for a group of subjects, or comparing two groups.

Importantly, connectivity results will be accessible either as dense connectivity maps, which will have fine spatial resolution but will be slower to compute and transmit, or as parcellated connectivity maps, which will be faster to process and in some situations may be pre-computed. Additional features that are planned include options to access time courses for R-fMRI data, fiber trajectories for structural connectivity data, and individual subject design files and time courses for T-fMRI data. Other approaches such as regression analysis will also be supported. For example, this may include options to determine the correlation between features of particular pathways or networks and particular behavioral measures (e.g., working memory).

When a spatial query is submitted, ConnectomeDB will parse the parameters, search the database to identify the appropriate subjects, retrieve the necessary files from its file store, and then execute the necessary calculations. By executing these queries on the database server and its associated computing cluster, only the final connectivity or activation map will need to be transferred back to the user. While this approach increases the computing demands on the HCP infrastructure, it will dramatically reduce the amount of data that needs to be transferred over the network. CWB will be a primary consumer of this service, but as with all services in the ConnectomeDB API, it will be accessible to other external clients, including other visualization environments and related databases.

User interface. The ConnectomeDB UI is being custom developed using dynamic web technologies (HTML 5, Javascript, Ajax; **Figure 2**). Building on advanced web technologies has several advantages, including streamlined access to remote data, high levels of dynamic user interaction, and portability across client systems (browsers, desktop applications, mobile devices). The interface will include two main tracks. The *Download* track emphasizes rapid identification of data of interest and subsequent download. The most straightforward downloads will be pre-packaged bundles, containing high interest content from each quarterly data release (see Data-Sharing below). Alternatively, browsing and search interfaces will allow users to select individual subjects and subjects groups by one or more demographic, genetic, or behavioral criteria. The *Visualization & Discovery* track will include an embedded version of CWB, which will allow users to explore connectivity data on a rendered 3D surface (see Visualization below). Using a faceted search interface, users will build subject groups that are dynamically rendered by CWB.

High-throughput informatics

The HCP informatics platform will support high-throughput data collection and open-access data-sharing. Data collection requirements include uploading acquired data from multiple devices and study sites, enforcing rigorous QC procedures, and executing standardized image processing. Data-sharing requirements include supporting world-wide download of very large data sets and high volumes of API service requests. The overall computing and database strategy for supporting these requirements is illustrated in **Figure 3** and detailed below.

Computing infrastructure. The HCP computing infrastructure (**Table 1**) includes two complementary systems, an elastically expandable virtual cluster and a high performance computing system (HPCS). The virtual cluster has a pool of general purpose servers managed by VMW are ESXi. Specific virtual machines (VMs) for web servers, database servers, and compute nodes are allocated from the VMW are cluster and can be dynamically provisioned to

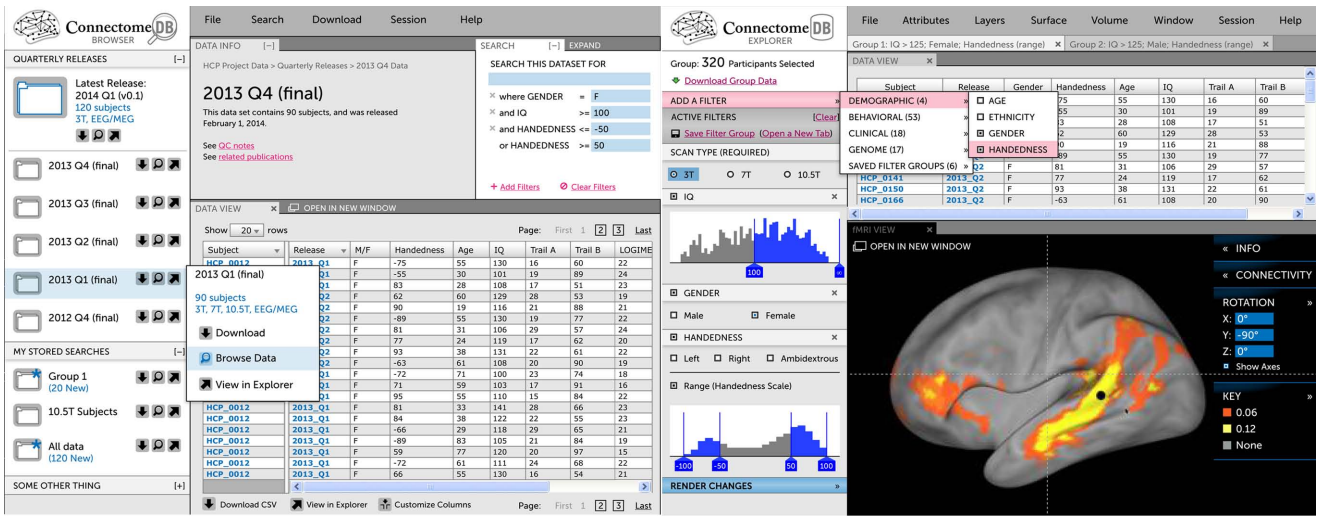


FIGURE 2 | The Connectome UI. (Left) This mockup of the Visualization & Discovery track illustrates key concepts that are being implemented, including a faceted search interface to construct subject groups and an embedded version of Connectome Workbench. Both the search interface and Workbench view are fed by ConnectomeDB's open API. (Right) This mockup of the Download track illustrates the track's emphasis on guiding users quickly to standard download packages and navigation to specific data.

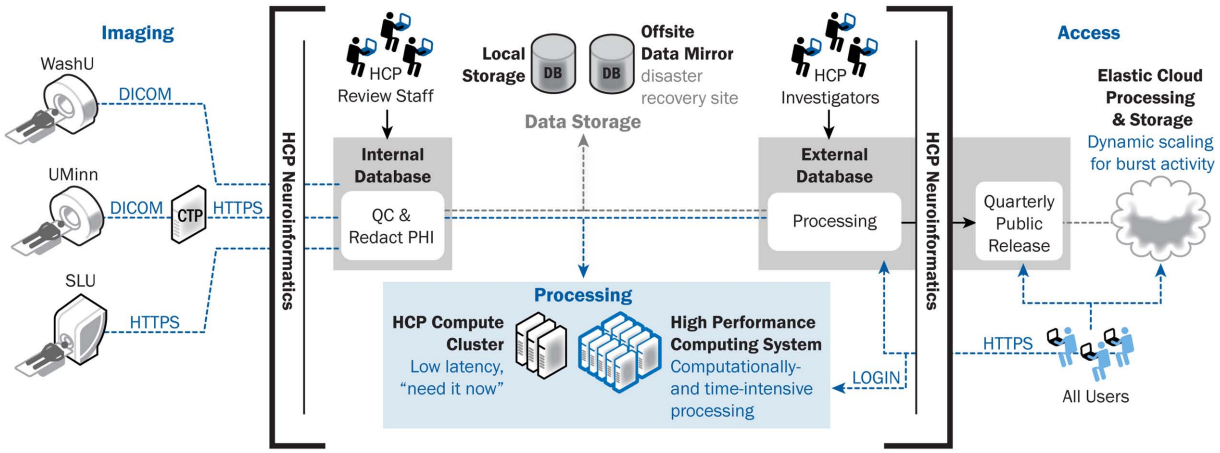


FIGURE 3 | ConnectomeDB architecture, including data transfer components. ConnectomeDB will utilize the Tomcat servlet container as the application server and use the enterprise grade, open source PostgreSQL database for storage of non-imaging data, imaging session meta-data, and system data. Actual images and other binary content are stored on a file system rather than in the database, improving performance and making the data more easily consumable by external software packages.

Table 1 | The HCP computing infrastructure.

Component	Device	Notes
Virtual cluster	2 Dell PowerEdge R610s managed byVMWare ESXi	Additional nodes will be added in years 3 and 5. Dynamically expandable using NIAC cluster.
Web servers	VMs running Tomcat 6.0.29 and XNAT 1.5	Load-balanced web servers host XNAT system and handle all API requests. Monitored by Pingdom and Google Analytics.
Database servers	VMs running Postgres 9.0.3.	Postgres 9 is run in synchronous multi-master replication mode, enabling high availability and load balancing.
Compute Cluster	VMs running Sun Grid Engine-based queuing.	Executes pipelines and on-the-fly computations that require short latencies.
Data storage	Scale-out NAS (Vendor TBD)	Planned 1 PB capacity will include tiered storage pools and 10Gb connectivity to cluster and HPCS.
Load balancing	Kemp Technologies LoadMaster 2600	Distributes web traffic across multiple servers and provides hardware-accelerated SSL encryption
HPCS	IBM system in WU's CHPC	The HPC will execute computationally intensive processing including "standard" pipelines and user-submitted jobs.
DICOM gateway	Shuttle XS35-704 Intel Atom D510	The gateway uses CTP to manage secure transmission of scans from UMinn scanner to ConnectomeDB.
Elastic computing and storage	Partner institutions, cloud computing	Mirror data sites will ease bottlenecks during peak traffic periods. Elastic computing strategies will automatically detect stress on compute cluster and recruit additional resources.

The web servers, database servers, and compute cluster are jointly managed as a single VMware ESXi cluster for efficient resource utilization and high availability. The underlying servers each include 48-GB memory and dual 6-core processors. Each node in the VMware cluster is redundantly tied back in to the storage system for VM storage. All nodes run 64-bit CentOS 5.5. The HPCS includes an iDataPlex cluster (168 nodes with dual quad core Nehalem processors and 24-GB RAM), an e1350 cluster (7 SMP servers, each with 64 cores and 256-GB RAM), a 288-port Qlogic Infiniband switch to interconnect all processors and storage nodes, and 9 TB of high-speed storage. Connectivity to the system is provided by a 4 × 10 Gb research network backbone.

match changing load conditions. Construction of the VMs is managed by Puppet (Puppet Labs), a systems management platform that enables IT staff to manage and deploy standard system configurations. The initial Phase 1 cluster includes 4 6-core physical CPUs that will be expanded in project years 3 and 5. We will partner with the WU Neuroimaging Informatics and Analysis Center (NIAC), which runs a similar virtual cluster, to dynamically expand the HCP's capacity during peak load. During extremely high load, we may also utilize commercial cloud computing services to elastically expand the cluster's computing capacity.

To support the project's most demanding processing streams, we have partnered with the WU Center for High Performance Computing (CHPC), which operates an IBM HPCS that commenced operating in 2010. Pipelines developed for the HCP greatly

benefit from the ability to run in parallel across subjects and take advantage of the vast amount of memory available in the HPCS nodes. Already, several neuroimaging packages including FreeSurfer, FSL, and Caret have been installed on the platform and are in active use by the HCP. The system utilizes a MOAB/TORQUE scheduling system that manages job priority. While the CHPC's HPCS is a shared resource openly available to the University's research community, the HCP will have assured priority on the system to ensure that the project has sufficient resources to achieve its goals.

The two HCP computing systems are complementary in that the virtual cluster provides rapid response times and can be dynamically expanded to match load. The HPCS, on the other hand, has large computing power but is a shared resource that queues jobs. The virtual cluster is therefore best for on-the-fly computing, such as is required to support web services, while the HPCS is best for computationally intensive pipelines that are less time sensitive.

The total volume of data produced by the HCP will likely be multiple petabytes (1 petabyte = 1,000,000 gigabytes). We are currently evaluating data storage solutions that handle data at this scale to determine the best price/performance ratio for the HCP. Based on preliminary analyses, we are expecting to deploy 1 PB of storage, which will require significant compromises in deciding which of the many data types generated will be preserved. Datasets to be stored permanently will include primary data plus the outputs of key pre-processing and analysis stages. These will be selected on the basis of their expected utility to the community and on the time that would be needed to recompute or regenerate intermediate processing results.

A driving consideration in selecting a storage solution is close integration with the HPCS. Four 10-Gb network connections between the two systems will enable high-speed data transmission, which will put serious strain on the storage device. Given these connections and the HPCS's architecture, at peak usage, the storage system will need to be able to sustain up to 200,000 input/output operations per second, a benchmark achievable by a number of available scale-out NAS (Network Attached Storage) systems. To meet this benchmark, we expect to design a system that includes tiered storage pools with dynamic migration between tiers.

In addition to this core storage system, we are also planning for backup, disaster recovery, and mirror sites. Given the scale of the data, it will be impossible to backup all of the data, so we will prioritize data that could not be regenerated, including the raw acquired data and processed data that requires significant computing time. We will utilize both near-line backups for highest priority data and offsite storage for catastrophic disaster recovery. As described below, our data-sharing plan includes quarterly data releases throughout Phase 2. To reduce bottlenecks during peak periods after these releases, we aim to mirror the current release on academic partner sites and commercial cloud systems. We are also exploring distribution through the BitTorrent model (Langille and Eisen, 2010).

Data workflow. All data acquired within the HCP will be uploaded or entered directly into ConnectomeDB. ConnectomeDB itself includes two separate database systems. Initially, data are entered into an internal-facing system that is accessible only to a small group of HCP operations staff who are responsible for reviewing data quality and project workflow. Once data pass quality review, they will be de-identified, including removal of sensitive fields from the DICOM headers and obscuring facial features in the high-resolution anatomic scans, transferred to a public-facing database, and shared with the public according to the data-sharing plan described below. All processing and analysis pipelines will be executed on the public-facing system so that these operations are performed on de-identified data only.

MRI data acquired at Washington University will be uploaded directly from the scanner to ConnectomeDB over the DICOM protocol on a secure private network. MRI data acquired at the University of Minnesota will be sent from the scanners to an on-site DICOM gateway configured with RSNA's Clinical Trial Processor (CTP) software. The CTP appliance will receive the data over the DICOM protocol, which is non-encrypted, and relay it to ConnectomeDB over the secure HTTPS protocol. Once the data have been uploaded, several actions will be triggered. First, XNAT's DICOM service will import metadata from the DICOM header fields into the database and places the files into its file repository. Next, a notification will be sent to HCP imaging staff to complete manual inspection of the data. Finally, a series of pipelines will be executed to generate sequence-specific automated QC metrics with flags to the HCP imaging staff regarding problematic data, and to validate metadata fields for protocol compliance. We aim to complete both manual and automated QA within 1 h of acquisition, which will enable re-scanning of individuals while they are still on-site.

MEG/EEG data will be uploaded to ConnectomeDB via a dedicated web form in native 4D format that will insure de-identification and secure transport via https. QC procedures will ensure proper linkage to other information via study specific subject IDs. EEG data will be converted to European Data format (EDF)[10] while MEG data will remain in source format.

Demographic and behavioral data will be entered into ConnectomeDB, either through import mechanisms or direct data entry. Most of the behavioral data will be acquired on the NIH Toolbox testing system, which includes its own database. Scripts are being developed to extract the test results from the Toolbox database and upload them into ConnectomeDB via XML documents. Additional connectome-specific forms will be developed for direct web-based entry into ConnectomeDB, via desktop or tablet computers.

Quality control. Initial QC of imaging data will be performed by the technician during acquisition of the data by reviewing the images at the scanner console. Obviously flawed data will be immediately reacquired within the scan session. Once imaging studies have been uploaded to the internal ConnectomeDB, several QC and pre-processing procedures will be triggered and are expected to be completed within an hour, as discussed above. First, the scans will be manually inspected in more detail by trained technicians. The manual review process will use a similar procedure as that used by the Alzheimer's Disease Neuroimaging Initiative, which includes evaluation of head positioning, susceptibility artifacts, motion, and other acquisition anomalies along a 4-point scale (Jack et al., 2008). Specific extensions will be implemented for

[10]http://www.edfplus.info/

BOLD and diffusion imaging. Second, automated programs will be run to assess image quality. Specific quality metrics are currently being developed for each of the HCP imaging modalities and behavioral paradigms. The resulting metrics will be compared with the distribution of values from previous acquisitions to determine whether each is within an expected range. During the initial months of data acquisition, the number of HCP scans contributing to these norm values will be limited, so we will seed the database with values extracted from data obtained in similar studies and during the pilot phase. As the study database expands, more sophisticated approaches will become available, including metrics specific for individual fMRI tasks (which may vary in the amount of head motion). Specific QC criteria for each metric will be developed during Phase I.

Data quality will be recorded in the database at the imaging session level and for each scan within the session. The database will include a binary pass/fail determination as well as fields for the aforementioned manual review criteria and the automated numeric QC metrics. Given the complexity and volume of image data being acquired in the HCP protocol, we anticipate that individual scans within each imaging visit will vary in quality. A single fMRI run, for example, might include an unacceptable level of motion, whereas other scans for that subject are acceptable in quality. In such cases, data re-acquisition is unlikely. The appropriate strategy for handling missing datasets will be dependent on exactly which data are absent.

Pipeline execution. The various processing streams described above are complex and computationally demanding. In order to ensure that they are run consistently and efficiently across all subjects, we will utilize XNAT's pipeline service to execute and monitor the processing. XNAT's pipeline approach uses XML documents to formally define the sequence of steps in a processing stream, including the executable, execution parameters, and input data. As a pipeline executes, the pipeline service monitors its execution and updates its status in the database. When a pipeline exits, notifications will be sent to HCP staff to review the results, following pipeline-specific QC procedures similar to those used to review the raw data. Pipelines that require short latency (such as those associated with initial QC) will be executed on the HCP cluster, while those that are more computationally demanding but less time sensitive will be executed on the HPCS.

Provenance. Given the complexity of the data analysis streams described above, it will be crucial to keep accurate track of the history of processing steps for each generated file. Provenance records will be generated at two levels. First, a record of the computational steps executed to generate an image or connectivity map will be embedded within a NIFTI header extension. This record will contain sufficient detail that the image could be regenerated from the included information. Second, higher level metadata, such as pipeline version and execution date, will be written into an XCEDE-formatted XML document (Gadde et al., 2011) and imported into ConnectomeDB. This information will be used to maintain database organization as pipelines develop over time.

Data-sharing. The majority of the data collected and stored by the HCP will be openly shared using the open-access model recommended by the Science Commons[11]. The only data that will be withheld from open access are those that could identify individual study participants, which will be made available only for group analyses submitted through ConnectomeDB. Data will be distributed in a rolling fashion through quarterly releases over the course of Phase 2. Data will be released in standard formats, including DICOM, NIFTI, GIFTI, and CIFTI.

Given the scope and scale of the datasets, our aim of open and rapid data-sharing represents a significant challenge. To address this challenge, the HCP will use a tiered distribution strategy (**Figure 4**). The first tier includes dynamic access to condensed representations of connectivity maps and related data. The second distribution tier will allow users to download bundled subsets of the data. These bundles will be configured to be of high scientific value while still being small enough to download within a reasonable time. A third tier will allow users to request a portable hard drive populated by a more extensive bundle of HCP data. Finally, users needing access to extremely large datasets that are impractical to distribute will be

[11]http://sciencecommons.org/projects/publishing/open-access-data-protocol/

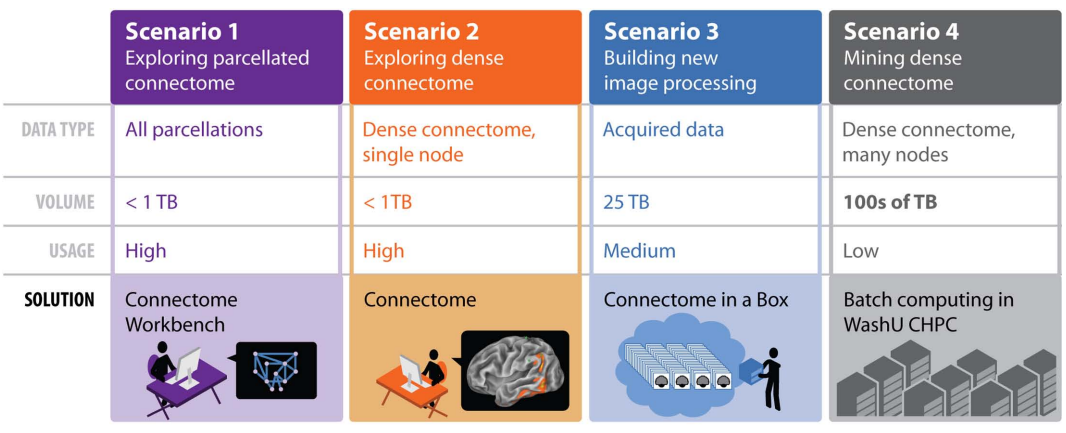


FIGURE 4 | HCP data distribution tiers.

able to obtain direct access to the HPCS to execute their computing tasks. This raises issues of prioritization, cost recovery, and user qualification that have yet to be addressed.

Some of the data acquired by the HCP could potentially be used to identify the study participants. We will take several steps to mitigate this risk. As mentioned above, sensitive DICOM header fields will be redacted and facial features in the images will be obscured. Second, the precision of sensitive data fields will be reduced in the open-access data set, in some cases binning numeric fields into categories. Finally, we will develop web services that will enable users to submit group-wise analyses that would operate on sensitive genetic data without providing users with direct access to individual subject data. For example, users could request connectivity difference maps of subjects carrying the ApoE4 allele versus ApoE2/3. The resulting group-wise data would be scientifically useful while preventing individual subject exposure. This approach requires care to ensure that requested groups are of sufficient size and the number of overall queries is constrained to prevent computationally driven approaches from extracting individual subject information.

VISUALIZATION

The complexity and diversity of connectivity-related data types described above result in extensive visualization needs for the HCP. To address these needs, CWB, developed on top of Caret software (Van Essen et al., 2001)[12] will include both browser and desktop versions. The browser-based version will allow users to quickly view data from ConnectomeDB, while the desktop version will allow users to carry out more demanding visualization and analysis steps on downloaded data.

Connectome Workbench

Connectome Workbench is based on Caret6, a prototype Java-based version of Caret, and will run on recent versions of Linux, Mac OS X, and Windows. It will use many standard Caret features for visualizing data on surfaces and volumes. This includes multiple viewing windows and many display options. Major visualization options will include (i) data overlaid on surfaces or volume slices in solid colors to display parcels and other regions of interest (ROIs), (ii) continuous scalar variables to display fMRI data, shape features, connectivity strengths, etc., each using an appropriate palette; (iii) contours projected to the surface to delineate boundaries of cortical areas and other ROIs, (iv) foci that represent centers of various ROIs projected to the surface; and (v) tractography data represented by needle-like representations of fiber orientations in each voxel.

A "connectivity selector" option will load functional and structural connectivity data from the appropriate connectivity matrix file (dense or parcellated) and display it on the user-selected surface and/or volume representations (e.g., as in **Figure 2**). Because dense connectivity files will be too large and slow to load in their entirety, connectivity data will be read in from disk by random access when the user requests a connectivity map for a particular brainordinate or patch of brainordinates. For functional connectivity data, it may be feasible to use the more compact time-series datasets and to calculate on the fly the correlation coefficients representing connectivity.

[12]http://brainvis.wustl.edu/wiki/index.php/Caret:About

Figure 5 illustrates how CWB allows concurrent visualization of multiple brain structures (left and right cerebral hemispheres plus the cerebellum) in a single window. Subcortical structures will be viewable concurrently with surfaces or as volume slices in separate windows.

Connectome Workbench will include options to display the results of various network analyses. For example, this may include concurrent visualization of network nodes in their 3D location in the brain as well as in a spring-embedded network, where node position reflects the strength and pattern of connectivity. The connection strength of graph edges will be represented using options of thresholding, color, and/or thickness. As additional methods are developed for displaying complex connectivity patterns among hundreds of nodes, the most useful of these will be incorporated either directly into CWB or via third party software.

Both the dense time-series and the parcellated time-series files provide temporal information related to brain activity. A visualization mode that plays "movies" by sequencing through and displaying each of the timepoints will be implemented. Options to view results of Task-fMRI paradigms will include both surface-based and volume-based visualization of individual and group-average data. Given that Task-fMRI time courses can vary significantly across regions (e.g., Nelson et al., 2010), options will also be available to view the average time course for any selected parcel or other ROI.

MEG and EEG data collected as part of the HCP will entail additional visualization requirements. This will include visualization in both sensor space (outside the skull) and after source localization to cortical parcels whose size respects the attainable spatial resolution. Representations of time course data will include results of power spectrum and BLP analyses.

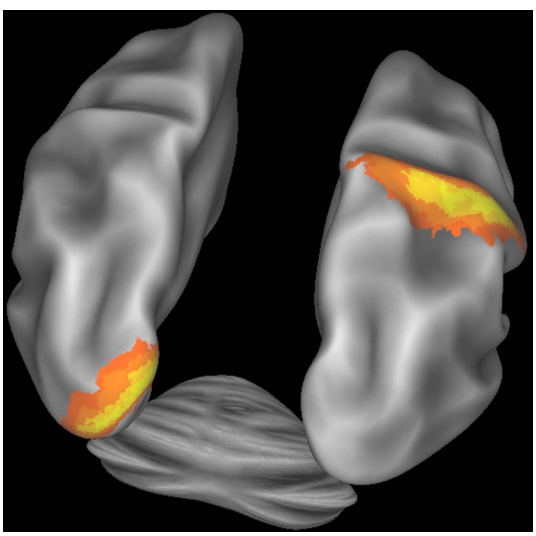


FIGURE 5 | Connectome Workbench visualization of the inflated atlas surfaces for the left and right cerebral hemispheres plus the cerebellum. Probabilistic architectonic maps are shown of area 18 on the left hemisphere and area 2 on the right hemisphere.

ConnectomeDB/Workbench Integration

Querying ConnectomeDB from Connectome Workbench. While users will often analyze date already downloaded to their own computer, CWB will also be able to access data residing in the Connectome database. Interactions between the two systems will be enabled through ConnectomeDB's web services API. CWB will include a search interface to identify subject groups in ConnectomeDB. Once a subject group has been selected, users can then visually explore average connectivity maps for this group by clicking on locations of interest on an atlas surface in CWB. With each click, a request to ConnectomeDB's spatial query service will be submitted. Similar interactive explorations will be possible for all measures of interest, e.g., behavioral testing results or task performances from Task-fMRI sessions, with the possibility of displaying both functional and structural connectivity maps.

Browser-based visualization and Querying Connectome DB. Users will also be able to view connectivity patterns and other search results via the ConnectomeDB UI so that they can quickly visualize processed data without having to download data – and even view results on tablets and smart phones. To support this web-based visualization, we will develop a distributed CWB system in which the visualization component is implemented as a web-embeddable viewer using a combination of HTML5, JavaScript, and WebGL. The computational components of CWB will be deployed as a set of additional web services within the Connectome API. These workbench services will act as an intermediary between the viewer and ConnectomeDB, examining incoming visualization requests and converting them into queries on the data services API. Data retrieved from the database will then be processed as needed and sent to the viewer.

Links to external databases

Providing close links to other databases that contain extensive information about the human brain will further enhance the utility of HCP-related datasets. For example, the Allen Human Brain Atlas (AHBA)[13] contains extensive data on gene expression patterns obtained by postmortem analyses of human brains coupled to a powerful and flexible web interface for data mining and visualization. The gene expression data (from microarray analyses and *in situ* hybridization analyses) have been mapped to the individual subject brains in stereotaxic space and also to cortical surface reconstructions. We plan to establish bi-directional spatially based links between CWB and the AHBA. This would enable a user of CWB interested in a particular ROI based on connectivity-related data to link to the AHBA and explore gene expression data related to the same ROI. Conversely, users of AHBA interested in a particular ROI based on gene expression data would be able to link to ConnectomeDB/Workbench and analyze connectivity patterns in the same ROI. A similar strategy will be useful for other resources, such as the SumsDB searchable database of stereotaxic coordinates from functional imaging studies[14]. Through the HCP's outreach efforts, links to additional databases will be developed over the course of the project.

[13]http://human.brain-map.org/

[14]http://sumsdb.wustl.edu/sums/

DISCUSSION

By the end of Phase II, the WU-Minn HCP consortium anticipates having acquired an unparalleled neuroimaging dataset, linking functional, structural, behavioral, and genetic information in a large cohort of normal human subjects. The potential neuroscientific insights to be gained from this dataset are great, but in many ways unforeseeable. An overarching goal of the HCP informatics effort is to facilitate discovery by helping investigators formulate and test hypotheses by exploring the massive search space represented by its multi-modal data structure.

The HCP informatics approach aims to provide a platform that will allow for basic visualization of the dataset's constituent parts, but will also encourage users to dynamically and efficiently make connections between the assembled data types. Users will be able to easily explore the population-average structural connectivity map, determine if the strength of a particular connection is correlated with a specific behavioral characteristic or genetic marker, or carry out a wide range of analogous queries. If the past decade's experience in the domain of genome-related bioinformatics is a guide, data discovery is likely to take new and unexpected directions soon after large HCP datasets become available, spurring a new generation of neuroinformatics tools that are not yet imagined. We will be responsive to new methodologies when possible and will allow our interface to evolve as new discoveries emerge.

The HCP effort is ambitious in many respects. Its success in the long run will be assessed in many ways – by the number and impact of scientific publications drawing upon its data, by the utilization of tools and analysis approaches developed under its auspices, and by follow-up projects that explore brain connectivity in development, aging, and a myriad of brain disorders. From the informatics perspective, key issues will be whether HCP data are accessed widely and whether the tools are found to be suitably powerful and user-friendly. During Phase I, focus groups will be established to obtain suggestions and feedback on the many facets of the informatics platform and help ensure that the end product meets the needs of the target users. The outreach effort will also include booths and other presentations at major scientific meetings (OHBM, ISMRM, and SfN), webinars and tutorials, a regularly updated HCP website[15], and publications such as the present one.

In addition to the open-access data that will be distributed by the HCP, the HCP informatics platform itself will be open source and freely available to the scientific community under a non-viral license. A variety of similar projects will likely emerge in the coming years that will benefit from its availability. We also anticipate working closely with the neuroinformatics community to make the HCP informatics system interoperable with the wide array of informatics tools that are available and under development.

While significant progress has been made since funding commenced for the HCP, many informatics challenges remain to be addressed. Many of the processing and analysis approaches to be used by the HCP are still under development and will undoubtedly evolve over the course of the project. How do we best handle the myriad of potential forks in processing streams? Can superseded pipelines be retired midway through the project or will users prefer

[15]http://www.humanconnectome.org/

for them to remain operational? What if a pipeline is found to be flawed? These and other data processing issues will require an active dialog with the user community over the course of the project. Subject privacy is another issue that requires both technical and ethical consideration. How do we minimize the risk of subject exposure while maximizing the utility of the data to the scientific community? Finally, what disruptive technologies may emerge over the 5 years of the HCP? How do we best maintain focus on our core deliverables while retaining agility to adopt important new tools that could further the scientific aims of the project? History suggests that breakthroughs can come from unlikely quarters. We anticipate that the HCP's open data and software sharing will encourage such breakthroughs and contribute to the nascent field of connectome science and discovery.

ACKNOWLEDGMENTS

Funded in part by the Human Connectome Project (1U54MH091657-01) from the 16 NIH Institutes and Centers that Support the NIH Blueprint for Neuroscience Research, by the McDonnell Center for Systems Neuroscience at Washington University, and by grant NCRR 1S10RR022984-01A1 for the CHCP; 1R01EB009352-01A1 and 1U24RR02573601 for XNAT support; and 2P30NS048056-06 for the NIAC. Members of the WU-Minn HCP Consortium are listed at http://www.humanconnectome.org/about/hcp-investigators.html and http://www.humanconnectome.org/about/hcp-colleagues.html. We thank Steve Petersen, Olaf Sporns, Jonathan Power, Andrew Heath, Deanna Barch, Jon Schindler, Donna Dierker, Avi Snyder, and Steve Smith for valuable comments and suggestions on the manuscript.

REFERENCES

Beckmann, M., Johansen-Berg, H., and Rushworth, M. F. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190.

Briggman, K. L., and Denk, W. (2006). Towards neural circuit reconstruction with volume electron microscopy techniques. *Curr. Opin. Neurobiol.* 16, 562–570.

de Pasquale, F., Della Penna, S., Snyder, A. Z., Lewis, C., Mantini, D., Marzetti, L., Belardinelli, P., Ciancetta, L., Pizzella, V., Romani, G. L., and Corbetta, M. (2010). Temporal dynamics of spontaneous MEG activity in brain networks. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6040–6045.

Feinberg, D. A., Moeller, S., Smith, S. M., Auerbach, E., Ramanna, S., Glasser, M. F., Miller, K. L., Ugurbil, K., and Yacoub, E. (2010). Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging. *PLoS ONE* 5, e15710. doi: 10.1371/journal.pone.0015710

Fielding, R. T. (2000). *Architectural Styles and The Design of Network-Based Software Architectures.* Doctoral dissertation, University of California, Irvine. Available at: http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm

Fox, M. D., and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711.

Gadde, S., Aucoin, N., Grethe, J. S., Keator, D. B., Marcus, D. S., and Pieper, S. (2011). XCEDE: an extensible schema for biomedical data. *Neuroinformatics* 1–14.

Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., and Wagster, M. V. (2010). Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurol.* 9, 138–139.

Jack, C. R. Jr., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., Whitwell, J. L., Ward, C., Dale, A. M., Felmlee, J. P., Gunter, J. L., Hill, D. L., Killiany, R., Schuff, N., Fox-Bosetti, S., Lin, C., Studholme, C., DeCarli, C. S., Krueger, G., Ward, H. A., Metzger, G. J., Scott, K. T., Mallozzi, R., Blezek, D., Levy, J., Debbins, J. P., Fleisher, A. S., Albert, M., Green, R., Bartzokis, G., Glover, G., Mugler, J., and Weiner, M. W. (2008). The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. *J. Magn. Reson. Imaging* 27, 685–691.

Johansen-Berg, H., and Behrens, T. E. (2009). *From Quantitative Measurement in-vivo Neuroanatomy.* Boston, MA: Elsevier.

Johansen-Berg, H., and Rushworth, M. F. (2009). Using diffusion imaging to study human connectional anatomy. *Annu. Rev. Neurosci.* 32, 75–94.

Langille, M. G. I., Eisen, J. A. (2010). BioTorrents: a file sharing service for scientific data. *PLoS ONE* 5(4): e10071. doi:10.1371/journal.pone.0010071

Lichtman, J. W., Livet, J., and Sanes, J. R. (2008). A technicolour approach to the connectome. *Nat. Rev. Neurosci.* 9, 417–422.

Marcus, D. S., Olsen, T. R., Ramaratnam, M., and Buckner, R. L. (2007). The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 5, 11–34.

Nelson, S. M., Cohen, A. L., Power, J. D., Wig, G. S., Miezin, F. M., Wheeler, M. E., Velanova, K., Donaldson, D. I., Phillips, J. S., Schlaggar, B. L., and Petersen, S. E. (2010). A parcellation scheme for human left lateral parietal cortex. *Neuron* 67, 156–170.

Newman, M. E. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* 103, 8577–8582.

Ou, W., Nummenmaa, A., Ahveninen, J., Belliveau, J. W., Hämäläinen, M. S., and Golland, P. (2010). Multimodal functional imaging using fMRI-informed regional EEG/MEG source estimation. *NeuroImage* 52, 97–108.

Patel, V., Dinov, I. D., Van Horn, J. D., Thompson, P. M., and Toga, A. W. (2010). LONI MiND: metadata in NIfTI for DWI. *Neuroimage* 51, 665–676.

Petrovic, V. S., Cootes, T. F. , Mills, A. M. , Twining, C. J., and Taylor, C. J. (2007). Automated analysis of deformable structure in groups of images. *Proc. British Machine Vision Conference* 2, 1060–1069.

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage.* 52, 1059–1069.

Sabuncu, M. R., Singer, B. D., Conroy, B., Bryan, R. E., Ramadge, P. J., and Haxby, J. V. (2010). Function-based inter-subject alignment of human cortical anatomy. *Cereb. Cortex* 20, 130–140.

Scheeringa, R., Fries, P., Petersson, K.-M., Oostenveld, R., Grothe, I., Norris, D. G., Hagoort, P., and Bastiaansen, M. C. M. (2011). Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. *Neuron* 69, 572–583.

Sporns, O., Tononi, G., and Kötter R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1: e42. doi:10.1371/journal.pcbi.0010042

Sporns, O. (2010). *Networks of the Brain.* Cambridge, MA: MIT Press, 375 pp.

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., and Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *J. Am. Med. Inform. Assoc.* 8, 443–459.

Vincent, J. L., Patel, G. H., Fox, M. D., Snyder, A. Z., Baker, J. T., Van Essen, D. C., Zempel, J. M., Snyder, L. H., Corbetta, M., and Raichle, M. E. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature* 447, 83-86.

Visscher, P. M., and Montgomery, G. W. (2009). Genome-wide association studies and human disease: from trickle to flood. *JAMA* 302, 2028–2029.

Wipf, D., and Nagarajan, S. A. (2009). Unified Bayesian framework for MEG/EEG source imaging. *NeuroImage* 44, 947–966.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 March 2011; accepted: 08 June 2011; published online: 27 June 2011.
Citation: Marcus DS, Harwell J, Olsen T, Hodge M, Glasser MF, Prior F, Jenkinson M, Laumann T, Curtiss SW and Van Essen DC (2011) Informatics and data mining tools and strategies for the Human Connectome Project. Front. Neuroinform. ***5****:4. doi: 10.3389/fninf.2011.00004*

frontiers in
NEUROINFORMATICS

METHODS ARTICLE
published: 06 June 2011
doi: 10.3389/fninf.2011.00003

The Connectome Viewer Toolkit: an open source framework to manage, analyze, and visualize connectomes

Stephan Gerhard[1]*, Alessandro Daducci[1], Alia Lemkaddem[1], Reto Meuli[2], Jean-Philippe Thiran[1] and Patric Hagmann[2]

[1] *Signal Processing Laboratory 5, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*
[2] *Department of Radiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland*

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Dennis Säring, University Medical Center Hamburg-Eppendorf, Germany
Daniel Marcus, Washington University in St. Louis, USA

***Correspondence:**
Stephan Gerhard, Signal Processing Laboratory (LTS5), Ecole Polytechnique Fédérale de Lausanne – STI-IEL-LTS5, CH-1015 Lausanne, Switzerland.
e-mail: stephan.gerhard@epfl.ch

Advanced neuroinformatics tools are required for methods of connectome mapping, analysis, and visualization. The inherent multi-modality of connectome datasets poses new challenges for data organization, integration, and sharing. We have designed and implemented the Connectome Viewer Toolkit – a set of free and extensible open source neuroimaging tools written in Python. The key components of the toolkit are as follows: (1) The Connectome File Format is an XML-based container format to standardize multi-modal data integration and structured metadata annotation. (2) The Connectome File Format Library enables management and sharing of connectome files. (3) The Connectome Viewer is an integrated research and development environment for visualization and analysis of multi-modal connectome data. The Connectome Viewer's plugin architecture supports extensions with network analysis packages and an interactive scripting shell, to enable easy development and community contributions. Integration with tools from the scientific Python community allows the leveraging of numerous existing libraries for powerful connectome data mining, exploration, and comparison. We demonstrate the applicability of the Connectome Viewer Toolkit using Diffusion MRI datasets processed by the Connectome Mapper. The Connectome Viewer Toolkit is available from http://www.cmtk.org/

Keywords: connectomics, connectome, neuroimaging, python, multi-modal data, data management, network analysis, visualization

1 INTRODUCTION

What nervous systems do – essentially – is to connect. Investigations into the connectivity properties of nervous systems have a long history (Douglas and Martin, 2007; Fishman, 2007). Despite many efforts, contemporary knowledge about the specificity of structural and functional connectivity is still poor. The new field of connectomics is emerging to tackle the challenge of mapping complete neural circuitry, or connectomes.

Connectomes represent the fundamental pathways on which complex spatiotemporal activity patterns evolve. In turn, these activity patterns modify underlying structural pathways. For an understanding of how activity patterns arise (physiology) and what they are able to produce and mean (behavior), it is indispensable to have connectome data (neuroanatomy) on all spatial descriptive levels.

On the cellular level of description, light, and electron microscopy are the main imaging tools for mapping neuronal circuitry. Partial and complete connectomes have been mapped in a variety of organisms and structures such as the nematode *Caenorhabditis elegans* (Ward et al., 1975; White et al., 1976), the crustacean *Daphnia*'s optic lobe (Macagno et al., 1979), cat visual cortex (Binzegger et al., 2004), macaque (Felleman and Van Essen, 1991; Markov et al., 2010), the rabbit retina (Anderson et al., 2011), the mouse interscutularis muscle (Lu et al., 2009), hippocampus (Ascoli, 2010), or *Drosophila melanogaster* (Cardona et al., 2010; Chklovskii et al., 2010; Hampel et al., 2011).

On the macroscale level of description, diffusion-weighted magnetic resonance imaging (MRI) is the main imaging technology employed for mapping the structural connectivity of the human connectome (Hagmann, 2005; Sporns et al., 2005; Sporns, 2011). Magnetic resonance connectomics (Hagmann et al., 2010) is increasingly recognized as a tool for basic and clinical neuroscience. Several methodological advances in image acquisition, reconstruction, and tractography (Wedeen et al., 2008; Johansen-Berg and Behrens, 2009) suggest that automated processing pipelines will make it possible to generate comprehensive *in vivo* whole brain statistical connectomes.

Despite the big differences in spatial scale and data size, both levels of connectome mapping consist of similar stages. Connectome mapping workflows involve image acquisition, registration and segmentation, data organization and sharing, high-throughput pipelining, analysis, and visualization. Advanced neuroinformatics tools will be required to meet challenges that each stage presents. In this article, we will focus on the development of neuroinformatics tools in the emerging field of macroscale connectomics.

The efficiency of sharing data and source code would benefit if a transdisciplinary lingua franca for programming was available. Especially in the neurosciences, where researchers with varying degrees of scientific knowledge and programming skills meet, a common programming language helps to bridge gaps between theoretical and experimental worlds of investigation. Moreover, the programming language must be high-level, cross-platform,

easy-to-learn, and have a large number of scientific libraries available. In recent years, Python[1] has become a viable alternative to Matlab, Java, or C++. More and more, Python is becoming the language of choice in scientific computing communities (Oliphant, 2007; Langtangen, 2009). Python is a free, open source, cross-platform programming language with a rapidly growing number of high-quality scientific libraries and interfaces to legacy code. A special-topic issue on "Python in neuroscience" in Frontiers in Neuroinformatics, and a number of publications (Kinser, 2008; Spacek et al., 2008; Davison et al., 2009) give an indication of the significance of the Python programming language.

Due to improved data acquisition methods, it is now possible to acquire large multi-modal datasets in projects involving thousands of subjects. For instance, the Human Connectome Project[2] is currently underway and collects Diffusion MRI, fMRI, EEG, MEG, behavioral, and genetic data in a cohort of 1200 healthy subjects. In such large-scale projects, the neuroinformatics challenges of data handling, sharing, and analysis become unnecessarily difficult without common infrastructure and data format standards. Due to their longer tradition, in the fields of volume-based and surface-based analysis in neuroimaging, standardized data formats have already been established such as NIFTI[3] for volume-based data and GIFTI[4] for surface-based data. Importantly for MR connectomics, no common format for network-based data yet exists. To approach the task of specifying such a format for connectivity-related neuroimaging data, the *Connectivity InFormatics Technology Initiative* (CIFTI) was launched. Furthermore, a dedicated program of the *International Neuroinformatics Coordinating Facility* (INCF[5]) on standards for data and metadata sharing was established.

For data management and sharing of large and multi-faceted datasets, a flexible data format is necessary. The key requirements of such a flexible data format under the macroscale connectomics perspective are severalfold:

- A standardized container format for raw and processed multi-modal datasets that is based on common neuroimaging data formats, extended by a standard format for network-based datasets.
- A minimal set of required metadata that can be extended flexibly by user-defined metadata, and that allow easy sharing of data and metadata across collaborating groups.
- The possibility of relating different data modalities to each other.
- An interface to database infrastructures.
- A mapping to an object model in common programming languages.
- To enable the storage of behavioral data.
- To enable the storage of provenance information such as processing scripts and runtime environment
- The ability to link data and concepts to semantic frameworks.
- To enable easier data visualization (Benger, 2009) and analysis.

To establish a novel data format, it must come with appropriate libraries for reading and writing. Only when it is possible to easily read, modify, and save data in the new format, does it benefit the researcher who wants to focus on analysis and visualization. There exist many standard formats for the different data modalities, but no one has yet tried to combine these multi-modal datatypes into a single format.

Complementary to common data formats, investigators in the field of macroscale connectomics will require interactive research and development environments for data analysis and visualization. An optimal solution would be an integrated neuroinformatics environment based on Python. It would need to provide a graphical user interface (GUI) with extensive libraries, an interactive scripting shell and built-in script editors with code-highlighting and debugging functionality. Graph analysis libraries are required to unravel the complex brain network organization of structural and functional systems (Bullmore and Sporns, 2009). Furthermore, such a macroscale connectomics research environment needs to support an interactive mode of analysis and visualization of multi-modal datasets. Scripting interfaces provide the required flexibility to allow the implementation of a variety of multi-modal data exploration and data mining strategies in an interactive way (Akil et al., 2011). The environment needs not only to provide a methodology to automatically perform elementary functionality, but also guidance for the performance of more complex analysis and visualization tasks. Moreover, a modular software architecture fosters contributions by the open source research community. Open interfaces facilitate the reuse of a diversity of tools and external libraries that the connectome researcher can draw from.

We used the Python programming language to develop the free and open source Connectome Viewer Toolkit. We specify the Connectome File Format (CFF) as a container data format for multi-modal neuroimaging datasets, specifying a connectome file. We present the Python-based Connectome File Format Library (cfflib) for data manipulation and data sharing of connectome files. The Connectome Viewer provides a framework for interactive visualization and analysis of connectomes and multi-modal datasets. We will illustrate the application of the Connectome Viewer Toolkit on Diffusion MRI datasets processed by the Connectome Mapper. The Connectome Mapper is a Python-based tool that currently implements structural connectome mapping. **Figure 1** summarizes the general connectome processing workflow for structural and functional data. It highlights the use of the Connectome Viewer tools presented in this article.

2 TOOLKIT DESIGN

After having outlined the key neuroinformatic challenges when developing tools for macroscale connectomics, we present the design and implementation of the Python-based Connectome Viewer Toolkit for data organization, analysis, and visualization of connectome data.

General considerations We have adhered to best-practices for open source scientific software tool development from the beginning (Baxter et al., 2006). Python provides mature software engineering tools for writing, testing, debugging, and maintaining scientific software. We have followed a very modular philosophy when designing the tools and the toolkit in general.

[1]www.python.org

[2]humanconnectome.org

[3]nifti.nimh.nih.gov

[4]www.nitrc.org/frs/? group_id = 75

[5]www.incf.org/core/programs/datasharing

We have adopted the powerful and widely used distributed version control system called *git* (Chacon, 2009). It is complemented by its biggest hosting platform *GitHub*[6]. We use GitHub as a project management platform for the support of all stages of the collaborative software development process. Our GitHub organization repositories are used for source code hosting, bug tracking, release management, code review, wiki, and visualization of the version control history and contributions.

We use the Sphinx tool[7] to create the online documentation for all tools. Sphinx uses so-called restructured text as its markup language, allowing one easily to create HTML, and PDF documentation with cross-referencing. The built-in syntax highlighting of code improves the readability of the documentation. Moreover, we provide user guides, tutorials, and example datasets for all tools.

Installation For proper software packaging and distribution, the NeuroDebian project[8] provides professional expertise and infrastructure. NeuroDebian provides a repository of neuroscience-related packages for easy installation on Debian-based Linux operating systems, such as Ubuntu. The Connectome Viewer Toolkit is distributed through the NeuroDebian repository, thereby facilitating the installation of dependencies and regular updates. For users on Windows or Mac OS X platforms, the Enthought Python Distribution[9] provides the required Python environment for the toolkit. The Enthought Python Distribution is free for academic purposes.

2.1 THE CONNECTOME FILE FORMAT

The metaphor of a container shall serve to explain how the Connectome File Format (CFF) is structured. The CFF makes a distinction between two entities: the Connectome Markup Language file and the connectome objects. The Connectome Markup Language file is a single XML-based file named *meta.cml*, containing metadata and a list of connectome objects. The schema for the Connectome Markup Language file is specified in conformance to the W3C standard XML schema 1.0 language (Thompson et al., 2004) and is available online. **Figure 2** illustrates the basic design of the CFF graphically.

We take advantage of the power and flexiblity inherent in defining objects with XML to specify connectome objects. Connectome objects wrap data files of multiple modalities. They add a layer of metadata information to the primary data file they refer to. Each connectome object holds information about its *name*, *fileformat*, *datatype*, *description*, and additionally a flexibly extensible *metadata* tag. We use the term multi-modal here to distinguish data types rather than measurement modalities. The basic categories of connectome objects are: CMetadata, CNetwork, CVolume, CSurface, CTrack, CData, CScript, CTimeseries, and CImagestack. **Table 1** describes the connectome object types in detail.

The connectome file is not confined to contain multi-modal datasets for a single datasets, but it can store multi-subject datasets. When storing multi-subject datasets, adding metadata annotation tags of the subject ids is suggested. Retrieving connectome objects grouped by the subject ids is then enabled by the grouping function *group_by_tagkey* as shown below in the listing of Section 2.2 Similarly, when individual connectome objects are tagged as belonging to a patient or control group, retrieving the corresponding datasets for group-wise comparison is possible with the same function.

Open Metadata Flexible annotation of metadata for each connectome object is enabled by the *metadata* tag. It is possible to annotate every connectome object in the container in the same way. We defined two ways to create annotations of metadata: (a) by simple tagging, (b) by structured annotation.

For (a), simple key-value pairs of the form `<tag key="number_of_nodes">83</tag>` are used to tag connectome objects. The employed keys can be used later on to group required objects easily in Python's so-called dictionaries. For instance, CNetwork objects can be

[6]github.com/LTS5
[7]sphinx.pocoo.org
[8]neuro.debian.net
[9]www.enthought.com

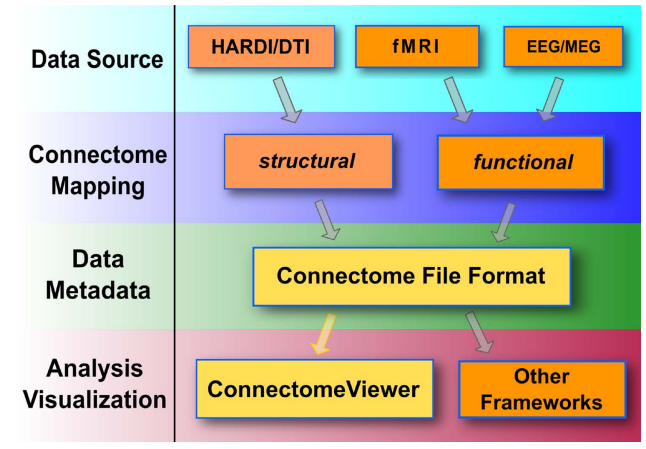


FIGURE 1 | General processing stages of a connectome workflow. The Connectome Viewer Toolkit currently supports the workflow highlighted in yellow. Mapping streams for structural data, such as the Connectome Mapper, or functional data may converge to a connectome file and can be further managed, analyzed, and visualized with the Connectome Viewer. Connectome files may be reused in other frameworks for analysis and visualization tasks.

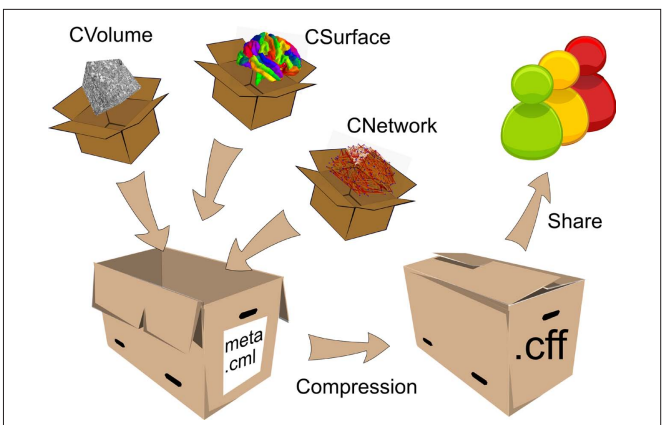


FIGURE 2 | The Connectome File Format Container. The connectome objects with reference to their primary data and metadata are depicted as small boxes. They are stored in a connectome file, represented by the open big box. After data manipulation, the connectome file can be compressed and shared with collaborators or sent to databases.

Table 1 | The variety of connectome objects the CFF supports.

Object types	Description
CMetadata	The CMetadata object describes metadata relevant to contents of the whole connectome file. We use relevant parts of the Dublin Core Metadata Terms specification (dublincore.org) to define the following tags: *title, creator, publisher, created, modified, license, references, description.* We have extended the core metadata tags with *generator, species,* and *email* tags. For each connectome object, a metadata tag can be added that expresses container-wide valid properties.
CNetwork	Networks of any sort can be stored. For MR structural connectomes, nodes represent brain regions and edges represent fiber tractography derived connections. The possibility of storing an arbitrary number of attributes per node and edge allows, for example, brain region nodes to point to ontologies that define them uniquely. Formats: GraphML, GEXF, NXGPickle, Other Types: Attribute Network, Dynamic Network, Hierarchical Network, Structural Network, Functional Network, Effective Network, …
CVolume	Volumetric, voxel-based datasets are widely used in the neuroimaging community to store many different measurement modalities. Examples: Apart from acquired raw data, brain segmentations or probability maps can be stored as 3D volumes. Formats: Nifti1, Nifti2, Other Types: Segmentation, T1-weighted, T2-weighted, PD-weighted, fMRI, Probability map, ASL, MD, FA, LD, TD, FLAIR, MRA, MRS, PET, …
CSurface	Surface-based datasets are usually stored as triangular meshes. They are often extracted from an underlying volumetric segmentation. Examples: Cortical maps for parcellations, thickness, or curvature information. Formats: Gifti, Other Types: Labeling, Surfaceset, Probability map, Surfaceset + Labeling, …
CTrack	Deterministic tractography creates sets of single polygonal lines. Examples: Reconstructed fiber bundles from Diffusion MRI Formats: TrackVis, Other Types: FACT Tractography, …
CData	Data of any type that does not fit into any other connectome object category. Examples: Phenotypic subject variables, assessment results Formats: NumPy, HDF5, XML, JSON, CSV, Pickle, TXT, Other Types: Fiber Labeling, Bval, Bvect, FPI-R, NEO-P-I-R, STAI, BIS-Test, I-S-T 2000R, …
CScript	Visualization and analysis procedures in the form of executable scripts. They may serve as provenance information for processed data. Examples: Connectome Mapper configuration script, Nipype script Formats: TXT, Python, Bash, Matlab, Other Types: Statistical Analysis, rsfMRI Connectivity Mapping, …
CTimeseries	There are plenty of time series related formats which makes it difficult to support a general one. We support generic data array containers that can store arbitrary time-series data. Formats: HDF5, NumPy, Other Types: EEG Time series, MEG Time series, fNIRS, …
CImagestack	Series of 2D images not simply representable in volume-based formats. Examples: Typical examples would be annotated slice-based atlases that represent areas as closed 2D polygons. Formats: PNG, JPG, TIFF, SVG, Other Types: Scalable Brain Atlas, …

The connectome objects are a wrapping mechanism, extending single data files by further annotations. Formats lists the file formats that are supported for reading and writing through the Connectome File Format Library. Types is a freely defined string. It usually denotes the measurement modality and is retrieved from controlled vocabularies.

grouped by their number of nodes, or connectome objects of multiple subjects can be grouped based on the subject id key. Although this scheme is very handy for everyday analysis, it is often desirable to store more structured metadata. As for (b), we adopted the Open Metadata Markup Schema odML[10] that was created to represent metadata for neurophysiology data in a bottom-up fashion. For metadata annotation, there are named *sections* that contain *properties*. The properties have a *name* and *value*, and additionally *type*, *unit*, and *uncertainty* information. This bottom-up scheme is very flexible and extensible and does not make any assumption about the terminology used. These terminologies are expected to emerge as the researchers start to annotate connectome objects using open metadata markup. Existing controlled vocabularies can already be employed with this schema. An example of the application of the metadata tag to the CMetadata object is shown in the Section 3 in **Figure 6**.

Multi-Modal Data Integration A connectome file can contain connectome objects for multiple modalities. These modalities are related to each other using unique numeric identifiers. Any given entity such as a brain structure or region of interest can potentially be represented in different modalities. Geometrically, an entity

[10]http://www.g-node.org/projects/odml

is instantiated as a set of cortical voxels in a volume and/or as a particular set of triangles in a surface mesh. The same entity may be represented by a node in a network.

We use integer values to establish this three-way relationship. Each entity is given a unique integer value. For each modality, this particular integer value is stored at defined places. For the CVolume object, the values are stored in individual voxels, thereby defining a segmentation. For the CSurface object, the values are stored as labelings on the vertices or on the faces, defining a surface-based parcellation. For the CNetwork object, we store the integer value as a node property. In **Figure 3**, we use the key *dn_correspondence_id* to store the integer value for a CNetwork object represented in the GraphML format[11]. In later analysis and visualization stages, the relevant data in the different modalities can be retrieved and combined as needed. **Figure 4** depicts this schema graphically.

Similarily, this schema is applied to connectivity information, combining the origin and target entities using their unique integer identifiers. The labeling of a subset of fibers of a whole brain tractography corresponds to the group of fibers that connect two brain regions. This connection has its correspondence in an edge of a CNetwork. Additional properties that hold for this group of

[11]graphml.graphdrawing.org

```
<graphml xmlns="http://graphml.graphdrawing.org/xmlns">
<graph id="MyConnectome" edgedefault="undirected">
<node id="1">
 <data key="dn_label">VisualArea1_RH</data>
 <data key="dn_node_position">(23.13,23.34,23.76)</data>
 <data key="dn_uri">http://www.connectome.ch/wiki/V1_(Homo sapiens)</data>
 <data key="dn_correspondence_id">1</data>
</node>
...
<edge id="e5_14" source="5" target="14">
 <data key="de_number_of_fibers">10</data>
 <data key="de_average_length_mm">22.6466</data>
</edge>
...
</graph></graphml>
```

FIGURE 3 | Networks are represented using the GraphML file format. The storage of an arbitrary number of attributes on the nodes and edges is possible. For instance, nodes denoting brain regions may link to semantic frameworks where definitions, delineation criteria, and literature references given. Standardized node positions are useful in graph layouting for comparison.

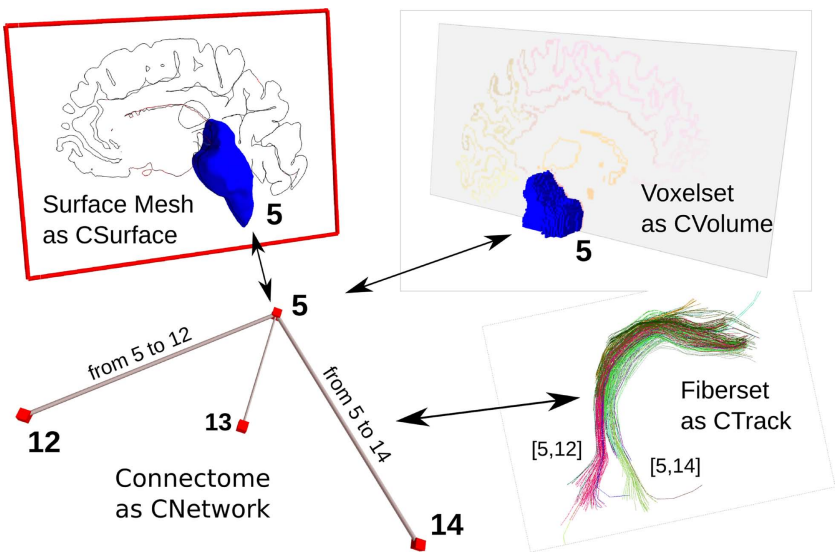


FIGURE 4 | Relationship between multi-modal connectome objects. Correspondence is established with unique (integer) identifiers between the nodes of a network, the ROI in a volumetric dataset, and the surface mesh. Analogically, a network edge has the same identifier as the fiber tracts that connect such two brain regions.

fibers, such as the number of fibers or averages of scalar values along the fibers, are stored as edge properties. For storing the label information for the fibers belonging to a network edge, we employ a CData object that represents a Nx2 NumPy array. N is the number of fibers and the first row stores integer values denoting the origin region of interest, and the second row stores the target region of interest. As a convention for the undirected fiber data derived from magnetic resonance tractography, we store the smaller integer value always in the first row. Thus, it becomes straightforward to retrieve all the fibers that connect two arbitrary regions of interest for further processing or visualization.

For time-dependent data, the same schema can be used. Time-series data from any source is stored in an NxM dimensional homogenous array. N is the number of channels and M is the number of time points. A CData object contains the labeling for the N channels to relate the series to any entity within the connectome file. Thus, time series can be defined for instance for brain regions, surface patches, network nodes, or electrodes in a very flexible manner. The CTimeseries object is used to store the array, for example in Hierarchical Data Format 5 (HDF5[12]) or NumPy (Oliphant, 2006) array format. In the CTimeseries metadata fields, additional parameters such as the sampling frequency can be stored. Additional CData objects may contain spatial position of the channels.

When psychological assessments are made in clinical trials, the datasets are often stored as spreadsheets. Bundling these data as CData within the connectome files in tabular form (CSV) or as an XML file, facilities data organization, and subsequent statistical correlation procedures with neuroimaging data.

Ultimately, all relevant multi-modal datasets for a neuroimaging study comprising multiple subjects or a single subject can be stored within a single connectome file.

2.2 THE CONNECTOME FILE FORMAT LIBRARY

The CFF specification is complemented by the Connectome File Format Library *(cfflib)* for Python. The cfflib supports (a) reading and writing of the connectome metadata markup (meta.cml) and compressed connectome files, (b) basic Input/Output of the supported file formats using supporting Python libraries, (c) a lazy loading strategy for data, (d) synchronizing files with a remote XNAT database servers, (e) setter and getter methods for updating metadata, (f) auxiliary methods for grouping objects based on metadata tag values and type.

We used the *generateDS* library (Dave Kuhlmann[13]) to create the Python object model. All classes were derived from the Connectome File Format XML schema. Subclasses provide additional methods for manipulation of the connectome files and basic loading and saving functionality. An example of an interactive Python session employing cfflib and its object model is given below:

```
# Import cfflib for usage
In [1]: import cfflib
# Load connectome markup file
In [2]: mycon = cfflib.load('meta.cml')
# Print summary of contained data
In [3]: mycon.print_ summary()
# Retrieve particular connectome object by name
In [4]: mynet = mycon.get_ by_ name('My
Connectome 83')
# Load the connectome data into memory
In [5]: mynet.load()
# Display nodes with attributes
In [6]: print mynet.data.nodes(data = True)
# On-the-fly grouping based on metadata key-values tags
In [7]: mygroup = cfflib.group_ by_ tagkey(cobj_
list = mycon.get_ all(),
tagkey = "sex", cobj_ type = ["CNetwork"],
exclude_ values = ["unknown"])
# Create list of CVolume names
In [8]: mynamelist = [vol.name for vol in mycon.
get_ connectome_ volume()]
```

Online tutorials explain extensively how to use cfflib to work with and create new multi-modal connectome files[14].

CFF data repository Via our GitHub repository[15], we provide a set of public, curated connectome datasets: single subject and group connectome files (generated with the Connectome Mapper), functional connectomes based on fMRI (Biswal et al., 2010), human atlas datasets such as Freesurfer's fsaverage, MNI152 templates, SRI24 Atlas (Rohlfing et al., 2010), and also some non-human datasets from *C. elegans*, *Macaca Mulatta*, and Mouse Brain (Johnson et al., 2010). We welcome contributions of connectome datasets under an open license for this data repository.

Database Interface We support connectome data sharing by providing an interface to remote database infrastructures. The eXtensible Neuroimaging Archive Toolkit (XNAT) is an informatics platform for managing, exploring, and sharing neuroimaging data (Marcus et al., 2007). It exposes web services using a RESTful API. Large neuroimaging initiatives, such as the Human Connectome Project, use the XNAT infrastructure for storage and sharing of large multi-subject, multi-site datasets. The Python library PyXNAT[16] interfaces to XNAT servers by their RESTful API. It supports an interactive mode of access for data selection, pulling, and pushing. Any type of data can be pushed to XNAT.

We implemented a push and pull mechanism using PyXNAT for connectome files in cfflib. After the configuration of the connection settings to the XNAT server, it is possible to push and pull connectome files to and from XNAT servers. For pushing the connectome file, the parameters project id, subject id, and experiment id must be set, in order to correctly associate the data in the project organization. All the connectome objects in the container are then submitted to the XNAT server sequentially. For the pulling operation, the same information has to be given with the storage path as an additional parameter. Submitted connectome objects can be displayed and downloaded from the XNAT web interface individually.

[12]www.hdfgroup.org

[13]www.rexx.com/~dkuhlman/generateDS.html

[14]cmtk.org/cfflib/

[15]github.com/LTS5/cffdata

[16]packages.python.org/pyxnat/

2.3 THE CONNECTOME VIEWER

We wanted to provide neuroimaging researchers with an easy entry into the world of connectome analysis with Python. We designed the Connectome Viewer as a GUI environment with a powerful scripting interface for interactive data analysis and visualization. The primary data source for the application are connectome files. The role of the Connectome Viewer is to provide a tool for the analysis and visualization of connectome files derived from different mapping streams (see **Figure 1**).

2.3.1 Dependencies

The Connectome Viewer depends on the Enthought Tool Suite (ETS[17]). ETS provides the application-building framework Envisage and the Traits and Traits UI libraries for creating GUIs. We use Mayavi (Ramachandran and Varoquaux, 2011), the main component in ETS for interactive scientific 3D data visualization, based on the popular Visualization Toolkit VTK (Schröder et al., 2006). Furthermore, we require Chaco for interactive plotting, IPython (Perez and Granger, 2007) for interactive Python shell support and cfflib for data input/output. We rely on the NetworkX (Hagberg et al., 2008) data structures for representation of networks. The dependencies of the Connectome Viewer are listed in **Table 2**. All required dependencies are installed automatically by using the NeuroDebian repository.

2.3.2 Script Generation Mechanism For Usability

The approach we took in the design of the Connectome Viewer to support synergy effects of combined analysis and visualization is based on a simple and proven paradigm. We call it the Code Oracle: for a given task, first a set of graphical user dialogs are presented for the setting of task-relevant parameters. Afterward, a Python script is automatically generated according to the these parameters. Then the script is displayed in the script editor, and can be executed without further modifications in order to perform the requested task. The generated scripts are commented, enabling the researcher to understand the commands performed by the script. It is easily possible to change the basic parameters without going back to the GUI dialogs again by modifying the script directly. Moreover, the commands necessary for reading the appropriate data, performing the visualization and analysis tasks, and generation of results can be modified in the Script Editor.

[17]code.enthought.com

2.3.3 Plugin Architecture

The Connectome Viewer as an Envisage-based application consists of a set of plugins as the primary building blocks. Plugins may contribute menus, widgets, and other functionality to the application. The core plugins are the Connectome File View, the IPython Shell, the Script Editor, and Mayavi. The Code Oracle and Python Connectome Toolbox are two additional plugins.

The Connectome File View plugin A connectome file can be loaded and saved via the menu. The plugin has a tree view widget that represents the connectome file. All connectome objects are shown in this tree view. Moreover, the loaded connectome file is accessible in the IPython shell for scripting. Double-clicking connectome objects in the tree view loads the referred data file into memory. Connectome objects can be dragged and dropped in the IPython shell for further data inspection and usage in scripts.

The IPython Shell plugin The IPython Shell plugin provides a widget with an enhanced interactive Python shell. Features such as tab-completion, automated docstring display, logging, history, and many others make it an ideal environment for interactive scientific computing.

The Script Editor plugin The Script Editor plugin enables loading, saving, and execution of Python scripts and text files. It features line numbering and syntax highlighting. Scripts can be executed directly in the IPython Shell by keybindings (Ctrl-R for executing or Ctrl-S for saving scripts).

The Mayavi plugin The Mayavi plugin is the major building block that provides advanced interactive 3D visualization and plotting (Ramachandran and Varoquaux, 2011) to the application. It exposes an easy-to-use interface to the well-known Visualization Toolkit VTK (Schröder et al., 2006). Mayavi as a stand-alone

Table 2 | Connectome Viewer library dependencies. When using NeuroDebian for the installation, all required dependencies are installed automatically.

Package	Version	Short description
Envisage	>= 3.1.2	Application-building framework similar to the Eclipse framework. Envisage is a system to define, register and use plugins to build complete applications. It is part of the Enthought Tool Suite.
Traits/ TraitsUI	>= 3.4.0	Extends the Python type declarations for improved initialization, validation, and notification. TraitsUI provides GUI-creation methods for Traits-based objects.
Mayavi	>= 3.3.2	3D Scientific Data Visualization and Plotting. For easy and interactive visualization of data and seamless integration in Envisage-based applications. Mayavi uses Traited VTK exposing a Pythonic API to VTK.
Chaco	>= 3.3.1	Interactive 2D plotting environment using Traits and TraitsUI.
IPython	>= 0.10	An enhanced interactive shell environment for scientific computing.
Fos	>= 0.1	A lightweight package for scientific 3D visualization (http://fos.me/). It supports basic visualization of dynamic networks, surfaces, and large tractography datasets and is included in the Connectome Viewer codebase.
cfflib	>= 2.0	The Connectome File Format Library. It provides functionality for manipulation of connectome files and depends on Nibabel, NumPy, and NetworkX.
Nibabel	>= 1.1.0	General library for reading and writing many neuroimaging file formats.
NumPy	>= 1.3	Homogenous, multi-dimensional array support for different data types with manipulation, and processing routines.
NetworkX	>= 1.4	Data structures and algorithms for complex network analysis.

application uses the Envisage application framework. Because all functionality is exposed as Envisage plugins, the integration as part of the Connectome Viewer was straightforward. The Mayavi plugin also provides the Mayavi Visualization Tree widget, which manages scenes, visualization objects, and filters hierarchically. Additionally, the Visualization Object Editor lets the user change all parameters of the visualization objects, filters, and scenes.

The Code Oracle plugin We encapsulated the script-generating functionality in the Code Oracle plugin. The Code Oracle plugin adds a menu to the Connectome Viewer for invoking the Code Oracle mechanism for specific task. The Code Oracle menu is structured according to analysis and visulization task on the different connectome file objects: "Surface Mesh With Labels" (CSurface); "Volume Slicer," "Volume Rendering" (CVolume); "Network Visualization," "Connection Matrix Viewer," "Network Report" (CNetwork); "Network-based statistic (NBS)" (Statistics); "Fiber Visualization" (CTrack); "XNAT Push and Pull" (Other); "Brain Extraction using BET" (Nipype).

This list presents a set of basic functionality. Several generated Code Oracle scripts may be combined and modified in the script editor for more complex tasks. The number of generated scripts is expected to grow rapidly as new use cases are discovered and algorithms developed and integrated.

The Python Connectome Toolbox plugin The Python Connectome Toolbox serves as a container for a collection of connectome-related analysis algorithms. They can also be used without the GUI. We provide within the Python Connectome Toolbox a Python wrapper to the C++ implementation bct-cpp (Williams et al., 2011) of the Brain Connectivity Toolbox (Rubinov and Sporns, 2009). We expose all of the toolbox's functions with an easy-to-use interface and parameter descriptions. The Brain Connectivity Toolbox algorithms are widely used for network analysis in the neuroimaging community. An abundance of additional well-designed network analysis libraries and Python wrappers exist that cover almost all aspects of network-based connectome analysis: NetworkX (Hagberg et al., 2008), Boost Graph Library (Siek et al., 2001), iGraph (Csardi and Nepusz, 2006), graph-tool[18], or Python-graph[19].

Furthermore, we provide within the toolbox two Network-based statistics for case–control or task-control group studies: the Network-based statistic (Zalesky et al., 2010) and Block-based statistic (Meskaldji et al., 2010). The Code Oracle plugin produces scripts that use the Python Connectome Toolbox's functionality.

2.3.4 Supporting Libraries

Through the open design of the Connectome Viewer, we encourage the use of the many libraries available in the scientific Python community. The supporting libraries presented here provide powerful tools for creative data exploration and data mining. Every library is usable from the IPython widget in the Connectome Viewer. Vice versa, contributed packages to the Connectome Viewer are usable from a pure IPython shell if they do not require the GUI.

The Neuroimaging in Python (NIPY[20]) project (Millman and Brett, 2007) is an umbrella project for various efforts to build well-written and documented open source neuroimaging libraries. Currently, NIPY consists of the following five packages. A few more helpful packages for the neuroimaging researcher have been added to the list:

Nibabel "Nibabel provides read and write access to some common medical and neuroimaging file formats, including: ANALYZE, GIFTI, NIfTI1, MINC, DICOM, MGH, TrackVis, as well as PAR/REC." Nibabel constitutes a necessary component for writing data analysis workflows, because it is crucial to retrieve and exchange data across different analysis software and computer platforms.

Dipy "Dipy is an international, free and open software project for diffusion magnetic resonance imaging analysis in Python. DiPy includes methods for reconstruction, resampling, tractography, warping, fiber clustering and visualization." (Garyfallidis et al., 2011)

Nipype "Nipype, an open source, community-developed initiative under the umbrella of NIPY, is a Python project that provides a uniform interface to existing neuroimaging software and facilitates interaction between these packages within a single workflow. Nipype provides an environment that encourages interactive exploration of algorithms from different packages, eases the design of workflows within and between packages, and reduces the learning curve necessary to use different packages." (Ghosh et al., 2010)

Nitime "Nitime is a library for time-series analysis of data from neuroscience experiments. It contains a core of numerical algorithms for time-series analysis both in the time and spectral domains, a set of container objects to represent time-series, and auxiliary objects that expose a high-level interface to the numerical machinery and make common analysis tasks easy to express with compact and semantically clear code." (Rokem et al., 2009)

NiPy "NiPy is a Python-based framework for the analysis of structural and functional neuroimaging data. It currently has a full system for general linear modeling of functional magnetic resonance imaging (fMRI)."

PyROI "PyROI is a Python package for functional neuroimaging region of interest extraction and analysis. It offers an efficient processing stream and a wide range of flexibility in the way source images are parcellated. Using PyROI, users can extract parameter and contrast effect sizes or timecourses."[21]

PyMVPA "PyMVPA is a Python package intended to ease statistical learning analyses of large datasets. It offers an extensible framework with a high-level interface to a broad range of algorithms for classification, regression, feature selection, data import and export. While it is not limited to the neuroimaging domain, it is eminently suited for such datasets." (Hanke et al., 2009)

scikit-learn "For easy-to-use and general-purpose machine learning in Python. It contains supervised learning (SVM, GLM), as well as unsupervised learning algorithms."[22]

We want to emphasize the additional tremendous potential for data exploration and data mining of connectomes using these external libraries. For a functional mapping stream, PyROI may be

[18]projects.skewed.de/graph-tool/

[19]code.google.com/p/python-graph/

[20]nipy.org

[21]web.mit.edu/mwaskom/pyroi/

[22]scikit-learn.sourceforge.net

used to extract averaged time series from regions of interest. Nitime may then be used to perform functional connectivity analysis from extracted time-series data. Nipype is ready to implement even more sophisticated data workflows using external packages and parallelize them on cluster infrastructures. For Diffusion MRI, Dipy implements reconstruction, tractography, and fiber visualization methods in a free and open source manner. Furthermore, the PyMVPA framework and the scikit-learn library provide widely used machine learning algorithms. The application of machine learning methods is facilitated for non-experts through extensive online tutorials, for instance in tutorials on fMRI decoding analysis with scikit-learn[23] or a general introduction to the PyMVPA framework[24].

2.4 CONNECTOME MAPPING WORKFLOWS

The general workflow to derive useful connectome information from primary neuroimaging data was shown in **Figure 1**. Data used here for the demonstration of the Connectome Viewer Toolkit is derived from Diffusion MRI. We reimplemented the pipeline used by Hagmann et al. (2008). We used the Traits and Traits UI libraries to build a GUI for the Connectome Mapper tool. The Connectome Mapper's processing stages are depicted **Figure 5**. All image processing stages can be parameterized and run interactively, or remotely using a configuration script. Details of the individual processing stages are available from the online documentation. The Connectome Mapper is released together with the Connectome Viewer Toolkit as the Connectome Mapping Toolkit[25], but the toolkit can be used as stand-alone application, and also with other mapping workflows, such as Nipype.

We will not detail the Connectome Mapper architecture, individual processing stages or validation issues here. We note only that the last *"Connectome Creation"* stage merges data from the two processing streams. The fibers from tractography are merged with volumetric region of interest masks denoting brain regions to form a network. This step exemplifies one instance of establishing the relationships between multi-modal data types as shown in **Figure 4**.

[23]nisl.github.com/

[24]www.pymvpa.org/tutorial.html

[25]www.cmtk.org

The multi-modal datasets produced by connectome worksflows was a major motivation to create the CFF. The product of such a mapping stream not only comprises the resulting connection matrix, but also fibers, surfaces, segmentations, labelings, other data arrays, and metadata. Capturing provenance information and storing it along with the raw and processed data is very important for later reproducibility. There are no accepted standards in the neuroimaging community for provenance tracking. In the Connectome Mapper, the log and configuration file including versions of the called executables and the operating system environment are stored for provenance everytime the pipeline is run. The CFF provides a convenient way to store this data, as well as other provenance information.

3 RESULTS

Connectome file for Multi-Modal Datasets The last processing stage of the Connectome Mapper implements an automatic conversion of original and processed data into connectome files. Using cfflib, data with relevance for subsequent analysis and visualization are packed into a connectome file. In the Connectome Mapper, the generated connectome files usually consist of the networks (CNetwork), the volumetric segmentations and the raw data (CVolume), the extracted surfaces with their brain region labeling (CSurfce), the original and filtered fibers from the deterministic tractography (CTrack), the fiber labeling and property arrays (CData), and the subject and project metadata (CMetadata). During the pipeline configuration, these subject- and project-specific metadata fields have to be entered. In a postprocessing step, additional datasets such as psychological assessment and behavioral assays can be added to the generated connectome file.

An excerpt of the connectome markup *meta.cml* for a single subject that was generated in the conversion stage of the processing pipeline is shown in **Figure 6**. It is enriched by an example section that contains information about the EEG acquisition parameters.

The connectome file may now be published to a public or private repository or database infrastructure. For instance, the connectome file can be published to an XNAT server (Marcus et al.,2007) using

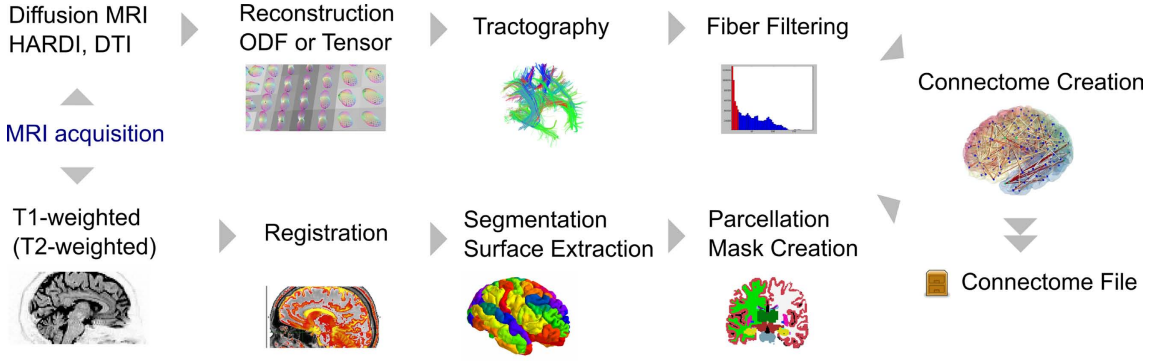


FIGURE 5 | The general processing stages of the Connectome Mapper. This pipeline is an example implementation of a structural connectome mapping stream. Here, two processing streams dealing separately with brain region (node) and white matter (edge) information eventually converge into the connectome of one particular subject.

```
<connectome xmlns="http://www.connectomics.org/cff-2"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:dcterms="http://purl.org/dc/terms/">
<connectome-meta version="2.0">
<generator>CMP1.0</generator>
<dcterms:creator>Stephan Gerhard</dcterms:creator>
<dcterms:created>2010-06-20</dcterms:created>
<dcterms:title>Single subject connectome</dcterms:title>
<dcterms:license>ODC Public Domain Dedication and Licence
</dcterms:license>
...more metadata tags...
<metadata>
 <tag key="subject_age">26</data>
 <tag key="subject_sex">M</data>
  <section title="EEG Acquisition">
   <property>
    <name>Sampling frequency</name>
    <value>256.0</value>
    <type>float</type>
    <uncertainty>0.0</uncertainty>
    <unit>Hz</unit>
   </property>
  </section>
</metadata>
</connectome-meta>
<connectome-network name="MyConnectome 83"
  src="Networks/myconnectome83.graphml"
  dtype="StructuralNetwork" fileformat="GraphML">
 <metadata>
  <tag key="number_of_nodes">83</tag>
  <tag key="hassubcortical">1</tag>
 </metadata>
 <description>Connectome with 83 brain regions.</description>
</connectome-network>
...more connectome objects...
</connectome>
```

FIGURE 6 | The content of a *meta.cml* file.

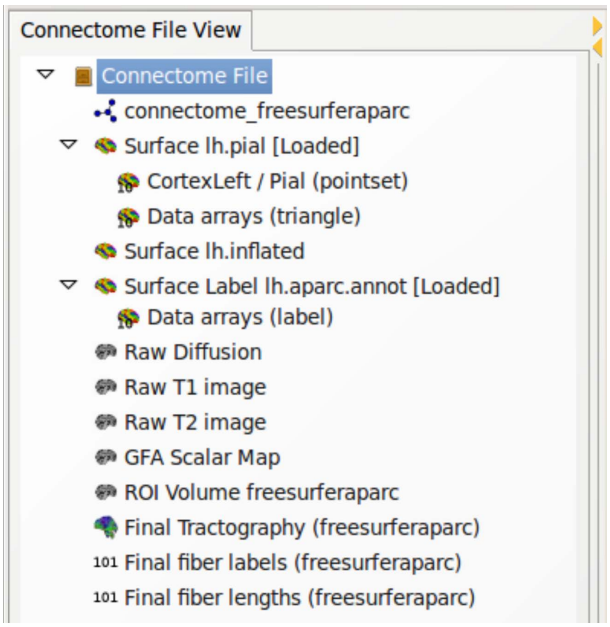


FIGURE 7 | Connectome File View as a Connectome Viewer widget. The treeview gives a convenient user interface to deal with connectome objects contained in a connectome file. Data files are loaded into memory by double-clicking. Single tree nodes can be dragged to the IPython shell for data inspection and scripting.

cfflib and PyXNAT with a few lines of Python code. The required commands are generated with the "XNAT Push and Pull" function of the Code Oracle plugin. For further analysis and visualization, the connectome file is loaded in the Connectome Viewer. The Connectome File View widget is updated and may appear as in **Figure 7**, displaying nodes for each connectome object in the connectome file.

The Connectome Viewer GUI consists of widgets contributed by its core plugins. The main window after loading a connectome file and executing a surface display script (using the Code Oracle) with an anatomical cortical parcellation (Desikan et al., 2006) is shown in **Figure 8**.

In **Figure 9**, the connectivity matrix based on particular Diffusion-derived measures for edge values is presented in an interactive user interface. The interface allows the user to zoom, drag, select the display range, switch between different edge values, and show the connecting regions when moving over an edge.

An instance of multi-modal data integration using network properties and reconstructed surfaces is depicted in **Figure 9**. A simple network metric, namely the node degree (number of adjoining edges), was computed. The nodes correspond to brain regions, delineated on a inflated brain surface. The node degree values were then used for color-coding the corresponding regions. Mayavi provides the DataSet Clipper to clip the surface mesh and reveal the view onto the subcortical nuclei. Comparatively, a much higher node degree is readily recognizable for the subcortical regions compared to the low node degree on the cortical surface.

The results of a Code Oracle analysis script to extract and cluster cortico-cortico U-fibers from tractography are shown in **Figure 10**. It uses supporting Python libraries, such as NumPy for array handling and comparison, Dipy for the local skeleton clustering algorithm (Garyfallidis et al., 2010) and Fos for visualization. Interactively, improvements in the fiber extraction and clustering parameters were made. The direct influence of parameter changes in the script are readily visible by re-executing the script. The script "U-fiber Extraction" to reproduce **Figure 10** can be invoked from the Code Oracle plugin.

Furthermore, we show an example PDF output of the network report generation mechanism, created using the "Network Report" Code Oracle (**Figure 11**). For PDF creation, ReportLab[26] is used. The figures in the report are generated using Matplotlib (Hunter, 2007) for 2D plotting. Connection matrix and node degree histograms are displayed along with basic network statistics for a selected networks contained in the currently loaded connectome file. The script that produces the reports can easily be adapted for modification of the layout or report content to include other network measures or visualizations (**Figure 12**).

4 DISCUSSION

4.1 CONNECTOME FILE FORMAT

In this article, we exemplified the application of the CFF for Diffusion MRI data and metadata. We envision more use cases of functional connectivity analysis using fMRI, EEG, MEG methods. The CFF specification is flexible enough to also accommodate such multi-modal datasets and their metadata.

[26]reportlab.com

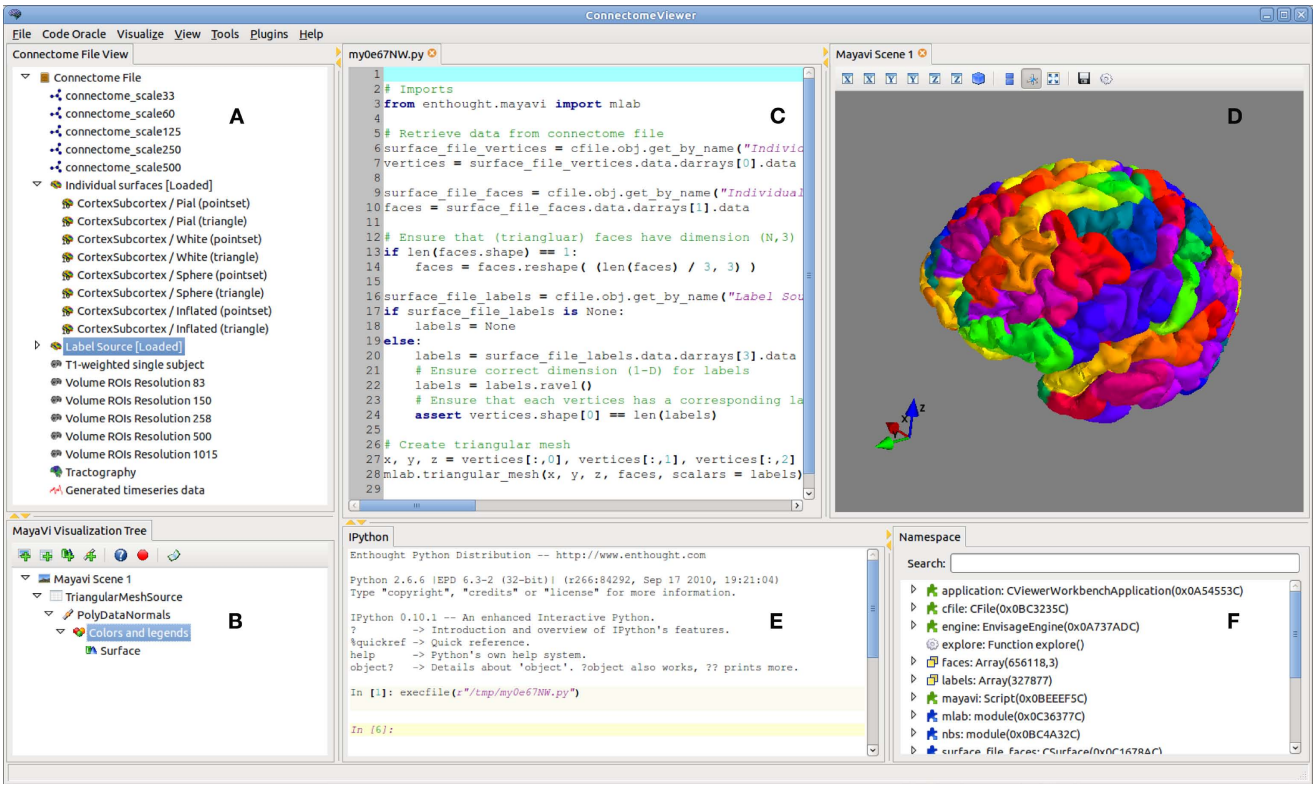


FIGURE 8 | The Connectome Viewer GUI. The main application is shown with the placeable widgets contributed by the core plugins. **(A)** The Connectome File View shows a treeview of the contents of a loaded connectome file. **(B)** The Mayavi Visualization Tree manages the visualization objects and scenes in a pipeline. **(C)** The ScriptEditor shows scripts generated with the Code Oracle with syntax highlighting. They can be manipulated and run in the IPython console. **(D)** The Mayavi Scene displays the visualized data. **(E)** The IPython shell is integrated as a widget. It exposes the loaded connectome file and other objects for interactive scripting and data inspection. **(F)** The Namespace widget displays the variables and packages currently loaded in memory and accessible in the IPython Shell.

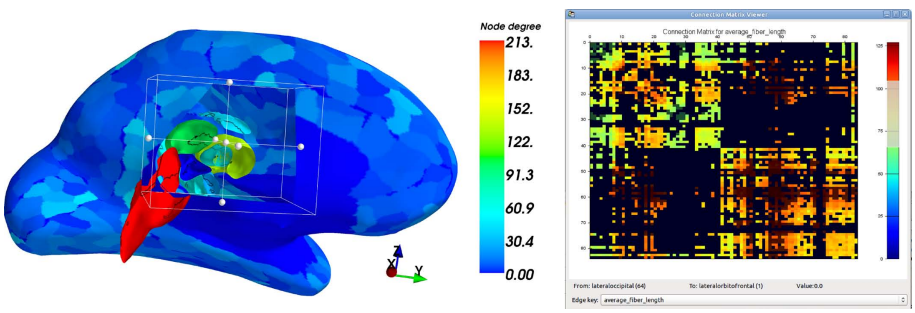


FIGURE 9 | Node degree display on the surface of an inflated left hemisphere (left). The right hemisphere and parts of the medial surface of the left hemisphere are clipped using the Mayavi DataSet Clipper. A set of left hemispheric cortical and subcortical regions is visible. The node degree of each region is coded in color. Blue corresponds to a low degree, green to a medium degree, and red to a high degree. The Connection Matrix Viewer (right). Switching between edge values and highlighting of the edge value ranges is possible. The From and To labels, and edge values are updated automatically when moving over the matrix with the mouse cursor.

One current limitation is a missing specification in the format itself of the shared spatial data space among connectome object types. For instance, it is possible that the affine transformation stored in NIFTI or GIFTI files do not map to the same common space, such as MNI152, due to different preprocessing. The annotation of the particular data spaces for individual connectome objects as metadata of the connectome file is possible.

As multi-modal or multi-subject datasets are often very large, in the order of tens or hundreds of gigabytes, having all datasets in one compressed container is disadvantageous for fast file access and manipulation. Because the *meta.cml* file stores relative references to the connectome object data files, the data files can reside on the local or remote file system in subfolders relative to the meta.cml file. The meta.cml can be loaded with

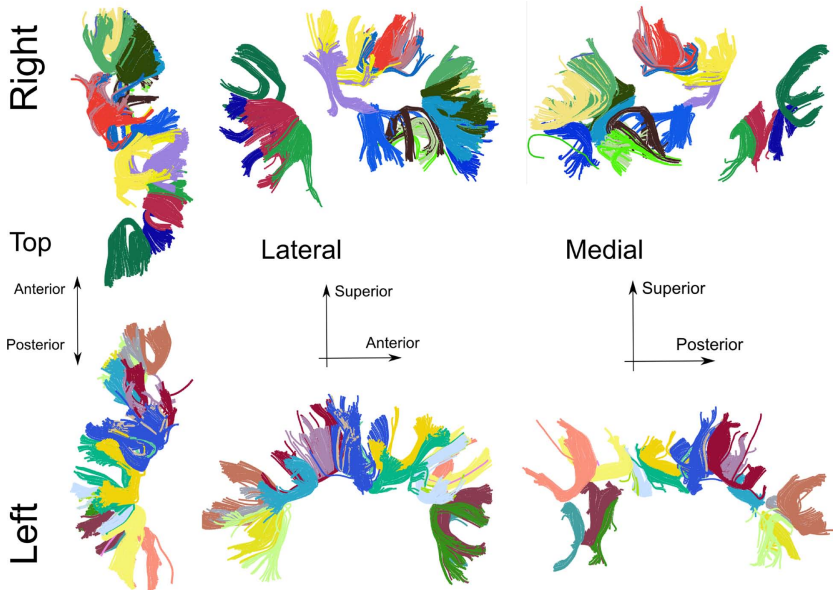


FIGURE 10 | Cortico-cortico U-shaped fibers generated with a Code Oracle script. The method to extract fibers from a subject uses criteria for fiber start and endpoint closeness, fiber curvature. The local skeleton clustering procedure (Garyfallidis et al., 2011) is used for the cluster coloring. The extracted U-fibers are cortical short distance connections hypothesized to contribute to the cortical small world network property (Bassett and Bullmore, 2006).

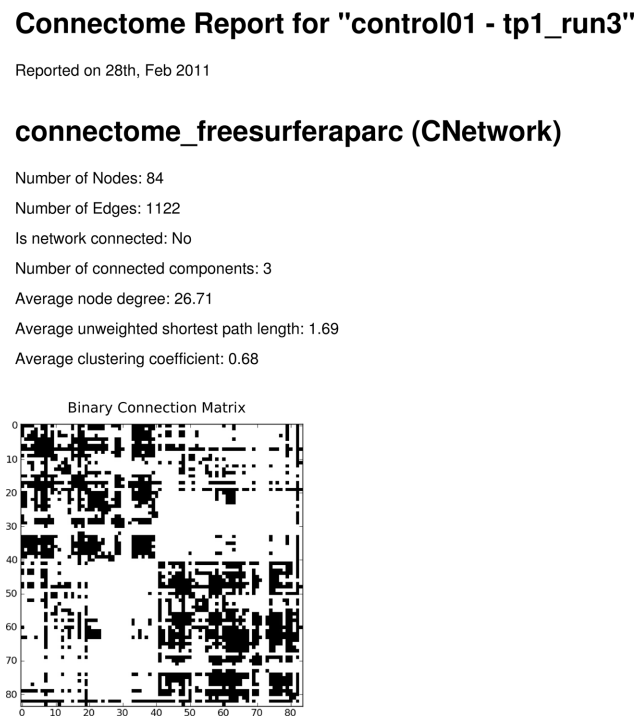


FIGURE 11 | Network PDF Report. Reports can be automatically generated using the Code Oracle "Network Report." The layout and report content can be adapted and extended with required results.

cfflib, thereby exposing the all contained connectome objects. This enables fast access and manipulation of the connectome objects without decompression and compression steps. Existing compression mechanisms for individual data files such as GZIP for NIFTI and GIFTI files reduce the data size substantially and can still be employed. Only for data exchange or distribution, the multi-modal dataset is ZIP compressed (filename ending. cff). We adopted the ZIP compression algorithm because it is widely supported and gives good performance in terms of fast compression and compression ratio on medium-sized datasets, such as single subject datasets.

The XML-Based Clinical Experiment Data Exchange Schema XCEDE2 (Gadde et al., 2011) provides an extensive metadata hierarchy for experimental context representation, provenance, and protocol information. Whereas XCEDE2 focuses on complete representation of neuroimaging studies and uses web frontends for data manipulation and annotation, the CFF metadata hierarchy is more open, flexible, and analysis-centric. All CFF data manipulation can be performed using cfflib from an interactive Python console. Experimental context representation may be stored as CData objects. For example, they may contain instances of XCEDE2 files.

The CIFTI format for connectivity-related neuroimaging data is proposed by the CIFTI and is based on the NIFTI-2 format. Currently, we employ GraphML to store connectivity graphs and associate them with surface and volume-based datasets, but adopting the CIFTI format for this purpose will be rather straightforward.

Recently, efforts to standardize terminologies and build ontologies for neuroscience have undergone a resurgence (Larson and Martone, 2009). We encourage linking to and referencing existing ontologies as much as possible. For instance, every node in a network produced by the Connectome Mapper denotes a brain structure. With attributes on these nodes, we refer to a unique Uniform Resource Identifier, as seen in **Figure 3**, pointing to

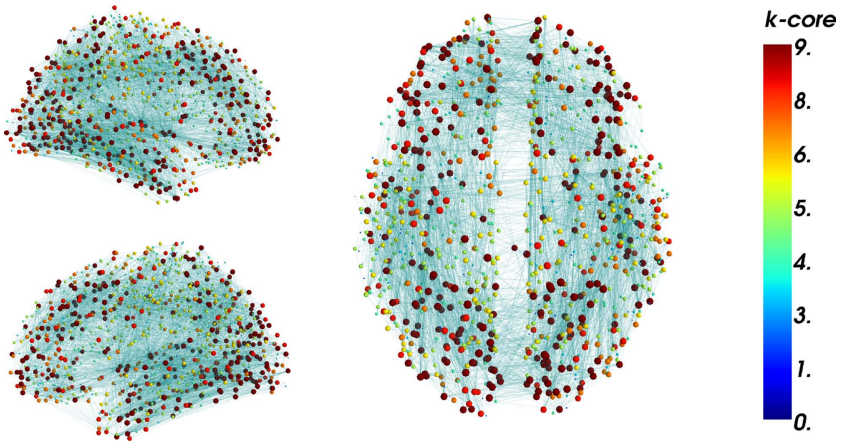


FIGURE 12 | Brain Connectivity Hairball. Display of the whole connectivity information at once does not allow for any sensible interpretation. Interactive manipulation of visualization parameters such as node position and coloring enables data exploration. Nodes are located here at the center of gravity of their underlying inflated brain region mesh. Nodes are scaled and colored according to their k-core number. The k-core number was computed using NetworkX in the IPython shell. The numbers were reused as a parameter to update the visualization interactively. Segregation of nodes with high k-core number at particular regions may be visible (Hagmann et al., 2008).

a standardized ontology or wiki. This reference allows other researchers to retrieve the concept and delineation criteria for a particular brain structure in a parcellation scheme. In addition, the annotation of metadata by the researcher, together with using standardized terminologies, will facilitate future data integration challenges of connectome datasets.

4.2 CONNECTOME FILE FORMAT LIBRARY

We used the CFF object model to create the Python library cfflib for connectome file manipulation and annotation. Within the Connectome Viewer Toolkit, cfflib naturally serves to convert multi-modal datasets and metadata from the Connectome Mapper into a connectome file. The Connectome Viewer uses cfflib to load and save connectome files. Thus, the CFF and cfflib serve well as interfacing tools.

As new file format input/ouput libraries are developed, the cfflib will be able to reuse these libraries to expose a common interface to the neuroimaging researcher dealing with multi-modal datasets of different file formats.

4.3 CONNECTOME VIEWER

We have demonstrated the usability of the Connectome Viewer as a research and development environment. Our experience has shown that integrating analysis, data manipulation, and visualization capabilities synergistically in a single application is beneficial to researchers. This is especially true for data exploration and data mining. Extending visualization applications with analysis and data manipulation functionality by including a Python shell was recently achieved in 3D Slicer (Gering et al., 2001). 3D Slicer is a comprehensive application for multi-modal visualization with over one million lines of mostly C++ code. The underlying toolkit for visualization is VTK. By using Mayavi as an interface to VTK, we can hide much of the complexity of VTK and provide an easy-to-use interface for visualization to the researcher. Mayavi provides extensive documentation and examples for many use cases, yet it allows one to use the underlying VTK objects if necessary. Similarily, the DataViewer3D (Gouws et al., 2009) uses VTK directly to provide multi-modal visualization capabilities and is Python-based. DataViewer3D does not include a Python shell for analysis and does not use the Enthought Toolsuite or Mayavi.

We have reused supporting libraries as much as possible, taking advantage from the expertise of library developers in their various domains. This has led to a comparatively small codebase for the Connectome Viewer which is approachable for contributor who may wish to create extensions. The modular plugin architecture furthers this. Through the Code Oracle script generation mechanism, the Connectome Viewer facilitates the introduction to Python scripting considerably. The scripts are easily adaptable and extendable to the needs of connectome researcher. As our experience shows, writing Python scripts feels very familiar to researchers used to other development environments such as Matlab, and can be learned in a short time period.

5 CONCLUSION

We have proposed the Connectome File Format as a convenient, easy-to-use container data format to deal with some of the heterogeneity and complexity of multi-modal neuroimaging data. The CFF *connects* multi-modal data sources and metadata in a comprehensive and flexible way. We have showed how the CFF and its accompanying Python library cfflib solve data management and integration challenges. We also foresee the usefulness of the CFF for datasets from functional neuroimaging and behavioral domains.

We have presented the Connectome Viewer, an integrated neuroinformatics research and development framework for 3D visualization and analysis. The modular plugin architecture provides means for extensibility and the Code Oracle method supports leveraging of the scripting interface by an automated script generation

mechanism for common analysis tasks. The full compatibility with the CFF facilitates cross-modal data mining, analysis and visualization in an interactive, scriptable way.

The Connectome Viewer Toolkit, its supporting libraries and the Connectome Mapper constitute the Connectome Mapping Toolkit. Altogether, this toolkit creates a unique, extensible workbench for new and ongoing macroscale connectome mapping, management, analysis, and visualization.

6 INFORMATION SHARING STATEMENT

The Connectome Viewer Toolkit is released under the terms of the open source Modified BSD license (opensource.org/licenses/bsd-license.php). Contributed packages and plugins adhere to their own open source licensing policy. All packages, documentation and example datasets can be downloaded from http://www.cmtk.org/. Furthermore, the tools are listed and associated in the Neuroimaging Informatics Tools and Resources Clearinghouse under http://www.nitrc.org/.

ACKNOWLEDGMENTS

We would like to thank the NIPY community and the open source Python community in general. Special thanks go to Leila Cammoun, Gaël Varoquaux for Mayavi development and support, Yannick Schwartz for PyXNAT, Eleftherios Garyfallidis for the Dipy package and the Fos vision, Christophe Chênes for supporting the development of cfflib, Erik Ziegler for the Nipype connectivity tutorial, and Mark Longair, Patricia Milz, and Alessandra Griffa for their helpful comments on the manuscript. This work is partially supported by the SPUM project number 33CM30-124089 of the Swiss National Science Foundation.

REFERENCES

Akil, H., Martone, M. E., and Van Essen, D. C. (2011). Challenges and opportunities in mining neuroscience data. *Science* 331, 708–712.

Anderson, J., Jones, B., Watt, C., Shaw, M., Yang, J.-H., DeMill, D., Lauritzen, J., Lin, Y., Rapp, K., Mastronarde, D., Koshevoy, P., Grimm, B., Tasdizen, T., Whitaker, R., and Marc, R. (2011). Exploring the retinal connectome. *Mol. Vis.* 17, 355–379.

Ascoli, G. A. (2010). The coming of age of the hippocampome. *Neuroinformatics* 8, 1–3.

Bassett, D. S., and Bullmore, E. (2006). Small-world brain networks. *Neuroscientist* 12, 512–523.

Baxter, S. M., Day, S. W., Fetrow, J. S., and Reisinger, S. J. (2006). Scientific software development is not an oxymoron. *PLoS Comput. Biol.* 2, e87. doi: 10.1371/journal.pcbi.0020087

Benger, W. (2009). On safari in the file format jungle – why can't you visualize my data? *Comput. Sci. Eng.* 11, 98–102.

Binzegger, T., Douglas, R. J., and Martin, K. A. C. (2004). A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.* 24, 8441–8453.

Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A. M., Ernst, M., Fair, D., Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kötter, R., Li, S. J., Lin, C. P., Lowe, M. J., Mackay, C., Madden, D. J., Madsen, K. H., Margulies, D. S., Mayberg, H. S., McMahon, K., Monk, C. S., Mostofsky, S. H., Nagel, B. J., Pekar, J. J., Peltier, S. J., Petersen, S. E., Riedl, V., Rombouts, S. A., Rypma, B., Schlaggar, B. L., Schmidt, S., Seidler, R. D., Siegle, G. J., Sorg, C., Teng, G. J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X. C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y. F., Zhang, H. Y., Castellanos, F. X., and Milham, M. P. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198.

Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., Tomancak, P., and Hartenstein, V. (2010). An integrated micro- and macroarchitectural analysis of the *Drosophila* brain by computer-assisted serial section electron microscopy. *PLoS Biol.* 8, e1000502. doi: 10.1371/journal.pbio.1000502

Chacon, S. (2009). *Pro Git.* Apress, 288. Available at: http://progit.org/book/

Chklovskii, D. B., Vitaladevuni, S., and Scheffer, L. K. (2010). Semi-automated reconstruction of neural circuits using electron microscopy. *Curr. Opin. Neurobiol.* 20, 8.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Syst.* 1695.

Davison, A. P., Hines, M. L., and Muller, E. (2009). Trends in programming languages for neuroscience simulations. *Front. Neurosci.* 3:374–80. doi: 10.3389/neuro.01.036.2009

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., and Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–80.

Douglas, R. J., and Martin, K. A. C. (2007). Mapping the matrix: the ways of neocortex. *Neuron* 56, 226–238.

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47.

Fishman, R. S. (2007). The Nobel Prize of 1906. *Arch. Ophthalmol.* 125, 690–694.

Gadde, S., Aucoin, N., Grethe, J. S., Keator, D. B., Marcus, D. S., and Pieper, S. (2011). XCEDE: an extensible schema for biomedical data. *Neuroinformatics* 1–14.

Garyfallidis, E., Brett, M., Amirbekian, B., Nguyen, C., Yeh, F.-C., Halchenko, Y., and Nimmo-Smith, I. (2011). "Dipy – a novel software library for diffusion MR and tractography," in *17th Annual Meeting of the Organization for Human Brain Mapping,* Quebec (submitted).

Garyfallidis, E., Brett, M., and Nimmo-Smith, I. (2010). "Fast dimensionality reduction for brain tractography clustering," in *16th Annual Meeting of the Organization for Human Brain Mapping,* Barcelona.

Gering, D. T., Nabavi, A., Kikinis, R., Hata, N., O'Donnell, L. J., Grimson, W. E., Jolesz, F. A., Black, P. M., and Wells, W. M. (2001). An integrated visualization system for surgical planning and guidance using image fusion and an open MR. *J. Magn. Reson. Imaging* 13, 967–975.

Ghosh, S., Burns, C., Clark, D., Gorgolewski, K., Halchenko, Y., Madison, C., Tungaraza, R., and Millman, K. (2010). "Nipype: open-source platform for unified and replicable interaction with existing neuroimaging tools," in *16th Annual Meeting of the Organization for Human Brain Mapping,* Barcelona.

Gouws, A., Woods, W., Millman, R., Morland, A., and Green, G. (2009). DataViewer3D: an open-source, cross-platform multi-modal neuroimaging data visualization tool. *Front. Neuroinformatics* 3:9. doi: 10.3389/neuro.11.009.2009

Hagberg, A. A., Schult, D. A., and Swart, P. J. (2008). "Exploring network structure, dynamics, and function using NetworkX," in *Proceedings of the 7th Python in Science Conference,* eds G. Varoquaux, T. Vaught, and J. Millman, (Pasadena, CA).

Hagmann, P. (2005). *From Diffusion MRI to Brain Connectomics.* PhD thesis, Ecole Polytechnique Fédérale de Lausanne, Lausanne, 127.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6, e159. doi: 10.1371/journal.pbio.0060159

Hagmann, P., Cammoun, L., Gigandet, X., Gerhard, S., Ellen Grant, P., Wedeen, V., Meuli, R., Thiran, J.-P., Honey, C. J., and Sporns, O. (2010). MR connectomics: principles and challenges. *J. Neurosci. Methods* 194, 34–45.

Hampel, S., Chung, P., McKellar, C. E., Hall, D., Looger, L. L., and Simpson, J. H. (2011). *Drosophila* brainbow: a recombinase-based fluorescence labeling technique to subdivide neural expression patterns. *Nat. Methods* 8, 253–259.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Fründ, I., Rieger, J. W., Herrmann, C. S., Haxby, J. V., Hanson, S. J., and Pollmann, S. (2009). PyMVPA: a unifying approach to the analysis of neuroscientific data. *Front. Neuroinformatics* 3:3. doi: 10.3389/neuro.11.003.2009.

Hunter, J. D. (2007). Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95.

Johansen-Berg, H., and Behrens, T. E. J. (2009). *Diffusion MRI: From Quantitative Measurement to In-vivo Neuroanatomy.* London: Academic Press, 490.

Johnson, G. A., Badea, A., Brandenburg, J., Cofer, G., Fubara, B., Liu, S., and Nissanov, J. (2010). Waxholm space: an image-based reference for

coordinating mouse brain research. *Neuroimage* 53, 365–372.

Kinser, J. (2008). *Python For Bioinformatics*. Sudbury, MA: Jones and Bartlett Publishers, 417.

Langtangen, H. P. (2009). *A Primer on Scientific Programming with Python (Texts in Computational Science and Engineering)*. New York: Springer, 693.

Larson, S. D., and Martone, M. E. (2009). Ontologies for neuroscience: what are they and what are they good for? *Front. Neurosci.* 3:1. doi: 10.3389/neuro.01.007.2009

Lu, J., Tapia, J. C., White, O. L., and Lichtman, J. W. (2009). The interscutularis muscle connectome. *PLoS Biol.* 7, e32. doi: 10.1371/journal.pbio.1000032.

Macagno, E. R., Levinthal, C., and Sobel, I. (1979). Three-dimensional computer reconstruction of neurons and neuronal assemblies. *Annu. Rev. Biophys. Bioeng.* 8, 323–351.

Marcus, D. S., Olsen, T. R., Ramaratnam, M., and Buckner, R. L. (2007). The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 5, 11–34.

Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M. A., Giroud, P., Ercsey-Ravasz, M., Pilaz, L. J., Huissoud, C., Barone, P., Dehay, C., Toroczkai, Z., Van Essen, D. C., Kennedy, H., and Knoblauch, K. (2010). Weight consistency specifies regularities of macaque cortical networks. *Cereb. Cortex* 21, 1254–1272.

Meskaldji, D. E., Cammoun, L., Hagmann, P., Meuli, R., Thiran, J. P., and Morgenthaler, S. (2010). Efficient statistical analysis of large correlated multivariate datasets: a case study on brain connectivity matrices. Available at: http://arxiv.org/abs/1008.1909.

Millman, K. J., and Brett, M. (2007). Analysis of functional magnetic resonance imaging in Python. *Comput. Sci. Eng.* 9, 52–55.

Oliphant, T. E. (2006). *Guide to NumPy*. Trelgol Publishing. Available at: http://www.tramy.us/

Oliphant, T. E. (2007). Python for scientific computing. *Comput. Sci. Eng.* 9, 10–20.

Perez, F., and Granger, B. E. (2007). IPython: a system for interactive scientific computing. *Comput. Sci. Eng.* 9, 21–29.

Ramachandran, P., and Varoquaux, G. (2011). Mayavi: a package for 3D visualization of scientific data. *Comput. Sci. Eng.* 13, 40.

Rohlfing, T., Zahr, N. M., Sullivan, E. V., and Pfefferbaum, A. (2010). The SRI24 multichannel atlas of normal adult human brain structure. *Hum. Brain Mapp.* 31, 798–819.

Rokem, A., Trumpis, M., and Perez, F. (2009). "Nitime: time-series analysis for neuroimaging data," in *Proceedings of the 8th Python in Science conference*, eds. G. Varoquaux, S. Van Der Walt, and J. Millman, Pasadena, CA, 68–75.

Rubinov, M., and Sporns, O. (2009). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069.

Schröder, W., Martin, K., and Lorensen, B. (2006). *Visualization Toolkit: An Object-Oriented Approach to 3D Graphics*, 4th Edn. Kitware Inc: New York 528.

Siek, J. G., Lee, L.-Q., and Lumsdaine, A. (2001). *The Boost Graph Library: User Guide and Reference Manual*. Boston, MA: Addison-Wesley Professional, 352.

Spacek, M., Blanche, T., and Swindale, N. (2008). Python for large-scale electrophysiology. *Front. Neuroinformatics* 2:9. doi: 10.3389/neuro.11.009.2008

Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042

Thompson, H. S., Beech, D., Maloney, M., and Mendelsohn, N. (2004). *XML Schema Part 1: Structures*, 2nd Edn. W3C Recommendation. Available at: http://www.w3.org/TR/xmlschema-1/

Ward, S., Thomson, N., White, J. G., and Brenner, S. (1975). Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *Caenorhabditis elegans*. *J. Comp. Neurol.* 160, 313–37. doi: 10.1002/cne.901600305

Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y. I., Dai, G., Pandya, D. N., Hagmann, P., D'Arceuil, H., and de Crespigny, A. J. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage* 41, 1267–1277.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1976). The structure of the ventral nerve cord of *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 275, 327–348.

Williams, S., Manicka, S., and Yaeger, L. (2011). *bct-cpp*. Available at: http://code.google.com/p/bct-cpp/

Zalesky, A., Fornito, A., and Bullmore, E. T. (2010). Network-based statistic: identifying differences in brain networks. *Neuroimage* 53, 1197–1207.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 February 2011; accepted: 18 May 2011; published online: 06 June 2011.
Citation: Gerhard S, Daducci A, Lemkaddem A, Meuli R, Thiran J-P and Hagmann P (2011) The Connectome Viewer Toolkit: an open source framework to manage, analyze, and visualize connectomes. Front. Neuroinform. ***5****:3. doi: 10.3389/fninf.2011.00003*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 04 July 2011
doi: 10.3389/fninf.2011.00005

Potential synaptic connectivity of different neurons onto pyramidal cells in a 3D reconstruction of the rat hippocampus

Deepak Ropireddy[1,2] and Giorgio A. Ascoli[1,2]*

[1] *Center for Neural Informatics, Structures and Plasticity, Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA*
[2] *Molecular Neuroscience Department, Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA*

Edited by:
Olaf Sporns, Indiana University, USA

Reviewed by:
Hermann Cuntz, University College London, UK
Christina Weaver, Franklin and Marshall College, USA

***Correspondence:**
Giorgio A. Ascoli, Center for Neural Informatics, Structures and Plasticity, Krasnow Institute for Advanced Study, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA.
e-mail: ascoli@gmu.edu

Most existing connectomic data and ongoing efforts focus either on individual synapses (e.g., with electron microscopy) or on regional connectivity (tract tracing). An individual pyramidal cell (PC) extends thousands of synapses over macroscopic distances (~cm). The contrasting requirements of high-resolution and large field of view make it too challenging to acquire the entire synaptic connectivity for even a single typical cortical neuron. Light microscopy can image whole neuronal arbors and resolve dendritic branches. Analyzing connectivity in terms of close spatial appositions between axons and dendrites could thus bridge the opposite scales, from synaptic level to whole systems. In the mammalian cortex, structural plasticity of spines and boutons makes these "potential synapses" functionally relevant to learning capability and memory capacity. To date, however, potential synapses have only been mapped in the surrounding of a neuron and relative to its local orientation rather than in a system-level anatomical reference. Here we overcome this limitation by estimating the potential connectivity of different neurons embedded into a detailed 3D reconstruction of the rat hippocampus. Axonal and dendritic trees were oriented with respect to hippocampal cytoarchitecture according to longitudinal and transversal curvatures. We report the potential connectivity onto PC dendrites from the axons of a dentate granule cell, three CA3 PCs, one CA2 PC, and 13 CA3b interneurons. The numbers, densities, and distributions of potential synapses were analyzed in each sub-region (e.g., CA3 vs. CA1), layer (e.g., oriens vs. radiatum), and septo-temporal location (e.g., dorsal vs. ventral). The overall ratio between the numbers of actual and potential synapses was ~0.20 for the granule and CA3 PCs. All potential connectivity patterns are strikingly dependent on the anatomical location of both pre-synaptic and post-synaptic neurons.

Keywords: rat hippocampus, potential synapses, connectivity, 3D model, CA pyramidal axons, dentate mossy fiber, CA3b interneuron axons, computational

INTRODUCTION

Mammalian brains have complex network architectures (Sporns, 2010), with each neuron connecting to thousands of others. Connectivity must be characterized at both synaptic and regional levels to advance our knowledge of cognitive and computational functions of nervous systems (Sporns et al., 2005; Buzsaki, 2007). Numerous studies recently explored structural and functional connectivity with different experimental modalities, including non-invasive imaging (Bullmore and Sporns, 2009; Honey et al., 2009; Bressler and Menon, 2010), electrophysiology (Kalisman et al., 2005), light microscopy (Ishizuka et al., 1990; Sik et al., 1993; Li et al., 1994; Wittner et al., 2007), and electron microscopy (Mishchenko et al., 2010). Non-invasive imaging such as DTI allows investigation of the whole human brain, but is only amenable to analyzing regional connectivity. Electron microscopy has complementary strengths and limitations, as it can unambiguously identify all synapses, but only in a narrow region of interest. Unfortunately, neither technique is suitable to acquire the whole synaptic connectivity of a typical cortical neuron. Light microscopy provides an optimal balance of resolution and field of view for this neuronal connectomic level bridging the micro- and macro-scale.

In particular, the connectivity maps of entire neurons can be investigated computationally based on light-level digital reconstructions of axonal and dendritic morphologies in simplified (Amirikian, 2005) or detailed (Stepanyants and Chklovskii, 2005; Escobar et al., 2008; Stepanyants et al., 2008) geometrical representations. In most such studies, putative connectivity is established on the basis of the close proximity between a pair of pre-synaptic (axonal) and post-synaptic (dendritic) segments (Kalisman et al., 2003; Stepanyants et al., 2004). These spatial overlaps have been defined as "potential synapses" because of strong evidence of spine and bouton motility and their essential role for synapse formation and memory consolidation (Lendvai et al., 2000; Chklovskii et al., 2004; Knott et al., 2006). This conceptual framework is appropriate for the mammalian hippocampus, as this region displays structural plasticity throughout adulthood and is clearly involved in learning (Squire and Zola-Morgan, 1991; Eichenbaum and Cohen 2001). To date, however, potential synapses have only been mapped in the surrounding of a neuron and relative to its local orientation rather than in a system-level anatomical reference. This limitation is particularly evident in the hippocampus, because of its peculiarly curved and laminar organization.

The methodology presented here applies an existing mathematical framework to estimate potential synaptic connectivity (Stepanyants and Chklovskii, 2005) to detailed arbor geometries embedded into a 3D reconstruction of the rat hippocampus. The hippocampal model was built by assembling the cyto-architectonic layers segmented from high-resolution histological images of thin cryostatic brain sections, and transforming the resulting space into a volumetric representation with uniformly sized voxels (Ropireddy et al., 2008). A custom-designed computational framework was developed to embed digital neuromorphological reconstructions with respect to the natural hippocampal axes (Ropireddy et al., 2011). The key advantage of this methodological extension is the ability to analyze potential synaptic maps of different neurons within the proper anatomical frame of reference. As the principal neurons of the hippocampus have long-range projections (Tamamaki et al., 1988; Li et al., 1994) forming synapses across different sub-regions and layers, this framework provides a unique opportunity to quantify potential synaptic patterns across the entire hippocampal extent.

The present analysis is based on all publicly available digital reconstructions of hippocampal pyramidal cell (PC) dendrites (Ascoli et al., 2007) and a representative sample of complete 3D axonal tracings. In particular, the pre-synaptic neuronal selection consists of a dentate granule cell (GC), 4 PCs (with somata in proximal CA3c, distal CA3c, CA3b, and CA2, respectively), and 13 CA3b radiatum and lacunosum-moleculare (LM) interneurons. For each case, we analyzed the numbers, densities, and distributions of potential synapses in every sub-region (e.g., CA3 vs. CA1), layer (e.g., oriens vs. radiatum), and septo-temporal location (e.g., dorsal vs. ventral).

MATERIALS AND METHODS

Previous work described in detail our high-resolution 3D reconstruction of the rat hippocampus from thin histological sections (Ropireddy et al., 2008) and the embedding of digitally traced neuronal morphologies (Ropireddy et al., 2011). Here, we only present a brief overview emphasizing just the procedures that specifically pertain to the present analysis.

DIGITAL 3D RECONSTRUCTION OF THE RAT HIPPOCAMPUS

A 45-day-old, male, Long–Evans hooded rat (226–237 g; Harlan, Indianapolis, IN, USA) was sacrificed by guillotine and its brain quickly removed and stored air-tight at −80°C. The brain was sectioned coronally at 16 μm in the dorso-ventral direction with a cryostat at −18°C. The 290 sections encompassing the entire hippocampus were mounted on slides and cover-slipped after Nissl staining. Slices were imaged at 3200 dpi with an EPSON scanner and contrast enhanced by the Matlab routine "*imcontrast*." The image stack was loaded into "*Reconstruct*" (Fiala, 2005) with a pixel/μm conversion factor (4.28) accounting for the measured planar shrinkage. Images were initially registered by mid-line guided manual alignment. Seven clearly identifiable cyto-architectonic layers were manually segmented (**Figure 1A**): hilus, GC, and molecular layer (ML) in the dentate gyrus (DG); and oriens (OR), PC, radiatum (RAD), and LM layers in CA3 and CA1. Serial tracing produced sets of pixels representing layer contours as closed polygons. All locations inside the boundaries of a given layer were identified by triangulation, extending a previous algorithm to yield 3D voxels from neighboring slices. Cubic voxels were thus defined with side corresponding to the inter-slice distance (16 μm). The initial registration was refined by iterative three-point average of the centroids through the longest hippocampal axis until reaching geometric convexity.

The 3D location of each voxel was mapped both onto canonical brain planes (coronal, sagittal, and horizontal), and in a natural hippocampus reference frame denoting longitudinal and transversal positions. Longitudinal coordinates were calculated as path distances from the septal to the temporal poles (**Figure 1B**). Transversal coordinates were computed by first virtually generating 60 cross-sectional planes, which displayed the classic double "C" shape corresponding to the principal cell layers. The medial axes of these granular and pyramidal contours defined the DG and CA transversal positions as the distances from the infra-pyramidal tip and from the CA3c end, respectively (**Figure 1B**). Additionally, each voxel was assigned appropriate stereotactic coordinates and tagged by its sub-region, namely DG infra- ("I") and supra-pyramidal ("S") blades, and CA3c, CA3b, CA3a, CA2, and CA1 (**Figure 1C**).

EMBEDDING OF DIGITAL NEURONAL MORPHOLOGIES

A selection of 18 fully 3D traced axons from different sub-regions of the hippocampus was obtained from NeuroMorpho.Org (Ascoli et al., 2007): 1 GC, 4 pyramidal neurons, and 13 interneurons. The somata of these pre-synaptic neurons were positioned according to the stereotactic coordinates reported in the original reports. The dentate GC and the distal CA3c pyramidal neuron were at AP = 2.5–3 mm and ML = 2.5 mm from bregma (Tamamaki and Nojyo, 1991). The proximal CA3c pyramidal neuron was at AP = 3.5 mm and ML = 2.5 mm (Wittner et al., 2007). The CA3b pyramidal neuron was at AP = 2.4 mm and ML = 3.5 mm (Li et al., 1994). The CA2 pyramidal neuron was at AP = 2–2.5 mm and ML = 2.5 mm (Tamamaki et al., 1988). All interneurons had somata in CA3b, six in the radiatum layer, and seven in LM. The range of their stereotactic coordinates were AP = 1.95–3.85 mm and ML = 2.4–2.6 mm (Ascoli et al., 2009). Since all these somatic locations for the pre-synaptic neurons are reported with a precision of 0.1 mm, they identify a range of six locations in the hippocampus reconstruction, each corresponding to voxels with 16 μm side (16 μm × 6 ≈ 0.1 mm). Cells were positioned in the location closest to the center of this range. In one case (GC) the entire range of 5 × 5 adjacent locations was tested (see Results).

The dendritic reconstructions of PCs were embedded as potential post-synaptic targets. For CA3, 54 pyramidal neurons were pooled from 5 different studies (Ishizuka et al., 1995; Turner et al., 1995; Henze et al., 1996; Carnevale et al., 1997; Jaffe and Carnevale, 1999). For CA1, 103 pyramidal neurons were pooled from 2 of the above studies and 5 additional ones (Pyapali and Turner, 1994, 1996; Ishizuka et al., 1995; Carnevale et al., 1997; Pyapali et al., 1998; Megias et al., 2001; Brown et al., 2005). In all cases, the somata were positioned based on the location ranges reported in the respective papers. As for the pre-synaptic neurons just discussed, the range of suitable voxels in the hippocampus 3D reconstruction was extended to account for the limited precision of the published positional information. Altogether, the location spans of these 10 datasets covered the entire hippocampal space invaded

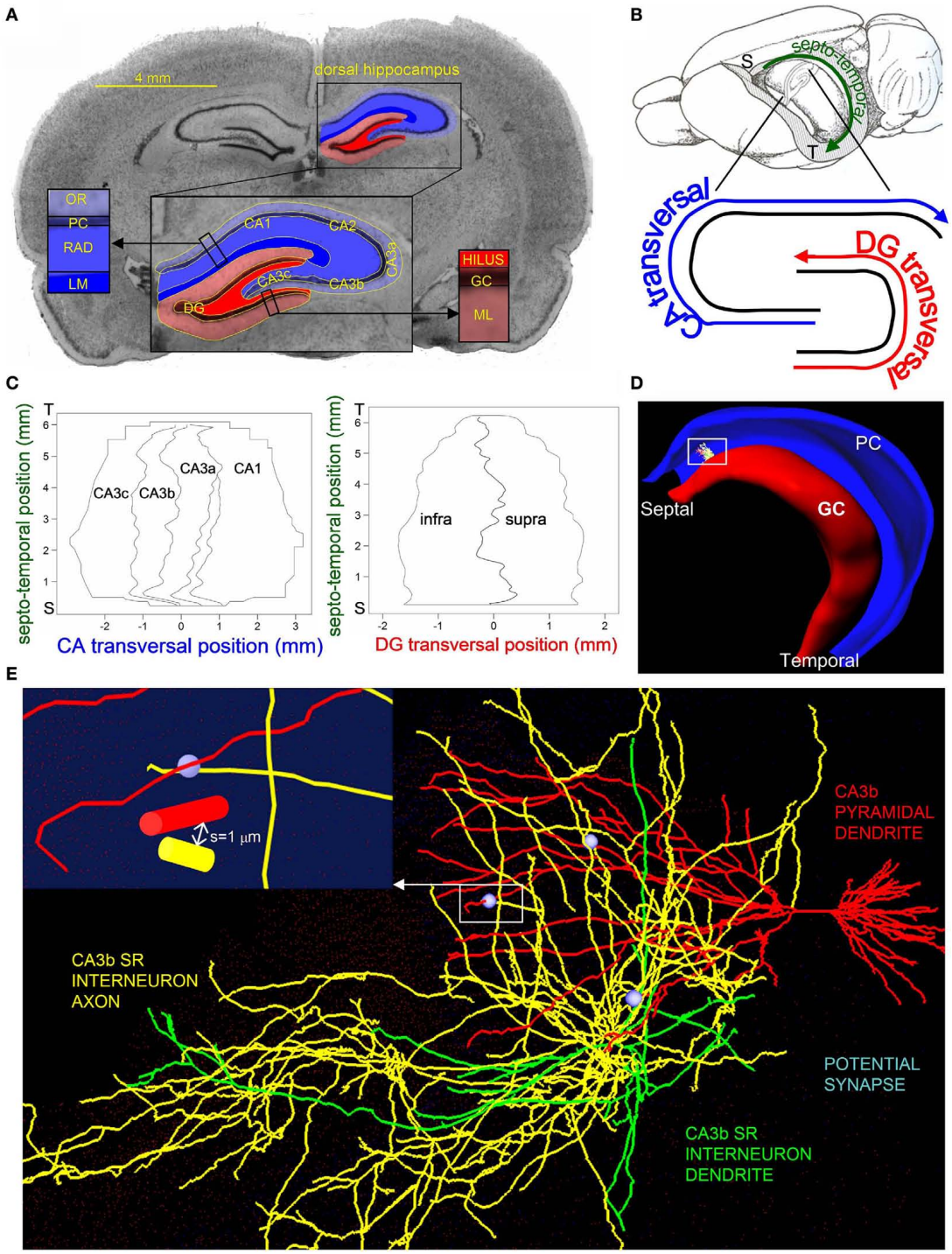


FIGURE 1 | Hippocampus 3D template and potential synapse illustration. **(A)** Nissl-stained coronal section with CA (blue) and DG (red) cyto-architectonic layers segmented in the dorsal hippocampus. The inset shows an enlarged view of the segmentation labeling the CA (CA1 and CA3a,b,c) and DG (infra and supra) sub-regions and layers. Layer abbreviations: GC, granule cell; ML, molecular; PC, pyramidal cell; RAD, radiatum; LM, lacunosum-moleculare; OR, oriens. **(B)** The septo-temporal and transversal axes of the hippocampus ("S" and "T" mark the septal and temporal poles). The rat brain depiction in lateral-medial orientation is adapted from Amaral and Witter (1989) with permission from Elsevier. **(C)** Maps marking the CA and DG sub-region borders with respect to the septo-temporal and transversal axes. **(D)** Representation of the DG (red) and CA (blue) principal layers of the hippocampus 3D reconstruction, with an embedded CA3b^{SR} interneuron and a CA3b pyramidal cell dendritic tree. **(E)** Enlarged view of the CA3b^{SR} interneuron (axon: yellow, dendrite: green) and the CA3b pyramidal cell (red), illustrating three potential synapses (light blue). Inset shows a further zoom-in on one of the potential synapses, identifying a location where the interneuron axon is within 1 μm of the CA3b pyramidal cell dendrite.

by the 18 selected axons. The pyramidal layers were densely packed by stochastically re-sampling the above 157 post-synaptic neurons, each within its appropriate voxel range. The PC layer was packed with CA3 and CA1 neurons with appropriate somatic densities so as to reflect the known total number of cells in these regions (Rapp and Gallagher, 1996).

Pyramidal and GCs were oriented in the 3D hippocampal reconstruction (**Figure 1D**) such that the principal axis of their dendritic arborization was perpendicular to the cellular layer and the secondary axis lay on the transversal plane (Scorcioni et al., 2002). To account for the natural variability observed in this general alignment (Claiborne et al., 1990; Ishizuka et al., 1995), the initial orientations were stochastically tilted between 0° and 5° around a random 0°–360° rotation. Principal and secondary axes were computed relative to the soma by single value decomposition of the digital reconstructions (Scorcioni et al., 2002) using standard numerical recipes (Press, 1988). Dendritic trees were scaled so as to reach the appropriate cyto-architectonic boundaries (ML for granule, LM for pyramidal apical, and, OR for basal trees). To ensure that the entire axonal arbors of the pre-synaptic cells fell within the 3D boundaries of the hippocampus, the orientations of these neurons were further manually fine-tuned within ±30° relative to their automatic embedding using quaternions (Hanson, 2006).

POTENTIAL SYNAPTIC CONNECTIVITY MEASUREMENT

A potential synapse occurs when a pre-synaptic segment comes in close proximity of a post-synaptic segment (**Figure 1E**). For excitatory synapses, the interaction distance is generally considered as a spine length or ~2 μm (Spacek and Hartmann, 1983; Harris and Stevens, 1989; Harris, 1999; Kalisman et al., 2003). For inhibitory synapses, the interaction distance can be approximated as the sum of the radii of the dendritic and axonal segments, typically 1 μm (Wierenga et al., 2008). The number of potential synapses can be derived from the geometry of pre- and post-synaptic neurons based on a published mathematical formula (Stepanyants and Chklovskii, 2005):

$$Np\left(\vec{R}_a,\vec{R}_d\right) = 2s\sum_{i,j} l_a^i l_d^j \left|\sin\left(n_a^i,n_d^j\right)\right| \exp\left[-\left(\vec{r}_a^{\,i}-\vec{r}_d^{\,j}\right)^2/4\sigma^2\right]/\left(4\Pi\sigma^2\right)^{3/2}$$

This equation defines the number of potential synapses between the axons of one neuron and the dendrites of a second neuron, $Np(\vec{R}_a,\vec{R}_d)$ as a function of their position in space, whereas the somata of the two cells are placed at coordinates $\vec{R}_a$ and $\vec{R}_d$, respectively. In the right hand side, s is the axo-dendritic interaction distance (in our case, 2 and 1 μm for principal cells and interneurons, respectively), l_a^i and l_d^j represent the unit-volume lengths of the axonal and dendritic segments along the unit vectors $\hat{n}_a^i$ and $\hat{n}_d^j$, and $|\sin(n_a^i,n_d^j)|$ is the absolute value of the sine of the angle between unit vectors $\hat{n}_a^i$ and $\hat{n}_d^j$. These neuronal reconstructions are taken to represent the morphology of a class of similar, but non-identical, neurons. Thus, a Gaussian filter with SD σ is applied around the centers of the axonal and dendritic segments with vectors $\vec{r}_a^{\,i}$ and $\vec{r}_d^{\,j}$. The parameter σ was fixed at 10 μm, corresponding to the average length of all segments. This equation is iteratively applied to all of the axonal and dendritic segments (indicated with i and j, respectively). Thus, the diameter measurements of the axon and dendrite segments are not used in computing the number of potential synapses.

COMPUTATIONAL DETAILS AND DATA ANALYSIS

The core computational framework was written in C/C++ and compiled with the GNU compiler under UNIX. Potential connectivity computations were executed on an SGI cluster of 80 Altix 8200 nodes, each containing two quad core Intel Xeon E5440 2.83 GHz processors with 16 GB of RAM. Instead of parallelizing the C/C++ code, we maximized the number of available nodes by submitting to the server a number of PBS (Portable Batch System) scripts that split the job into independent tasks[1]. At the same time, the code was optimized using *pthreads* (Silberschatz et al., 2009) to capitalize on the multi-processors of each node. As expected, the computing time decreased sub-linearly on the cluster machine compared to a single CPU. For instance, computing the potential connectivity for the proximal CA3c axon took ~24 h on a single computer vs. ~2 h on the cluster.

The program returns the number of potential contacts between a pre-synaptic neuron and all post-synaptic targets. Moreover, it records the anatomical locations of every potential synapse both within the hippocampus (in each of the coordinate systems described above) and in terms of distances from the somata of the pre- and post-synaptic neurons along the axonal and dendritic paths, respectively. Two-dimensional color maps (e.g., **Figure 2**) were generated using the R package "*ggplot2*"[2]. Linear dependencies were computed from these potential connectivity maps as regression fits of the septo-temporal positions weight-averaged by the corresponding numbers of potential synapses at a given transversal position. Virtual Reality Modeling Language files were generated with the freeware viewer "*view3dscene*"[3].

RESULTS

The results presented here are intended as a proof-of-concept of the 3D framework in computing the full potential connectivity of single neurons throughout system-level regional maps. As such, the research design was data-rather than hypothesis-driven.

We named the various neurons with a *region_cell-type* nomenclature convention, using subscripts AX and DE for pre- and post-synaptic cells, respectively and superscripts to specify sub-regions as needed. Thus, the dentate GC is labeled DG_GC_{AX}. The pre-synaptic PCs in various sub-regions are referred to as CA3c^{prox}_PC_{AX}, CA3c^{dist}_PC_{AX}, CA3b_PC_{AX}, and CA2_PC_{AX}. The interneurons with somata in the radiatum and LM layers are distinguished as CA3b^{R}_IN_{AX} and CA3b^{LM}_IN_{AX}, respectively. The post-synaptic PC targets are referred to as CA3_PC_{DE} and CA1_PC_{DE}.

POTENTIAL SYNAPTIC FRAMEWORK ILLUSTRATION

The dentate mossy fibers exemplify a well-known axonal arbor system within the hippocampus, with projections to all three sub-regions of the CA3 pyramidal network (**Figure 2A**). The axons of dentate GCs contact both PCs and interneurons in CA3

[1]http://www.unix-info.org

[2]http://had.co.nz/ggplot2

[3]http://vrmlengine.sourceforge.net/view3dscene.php

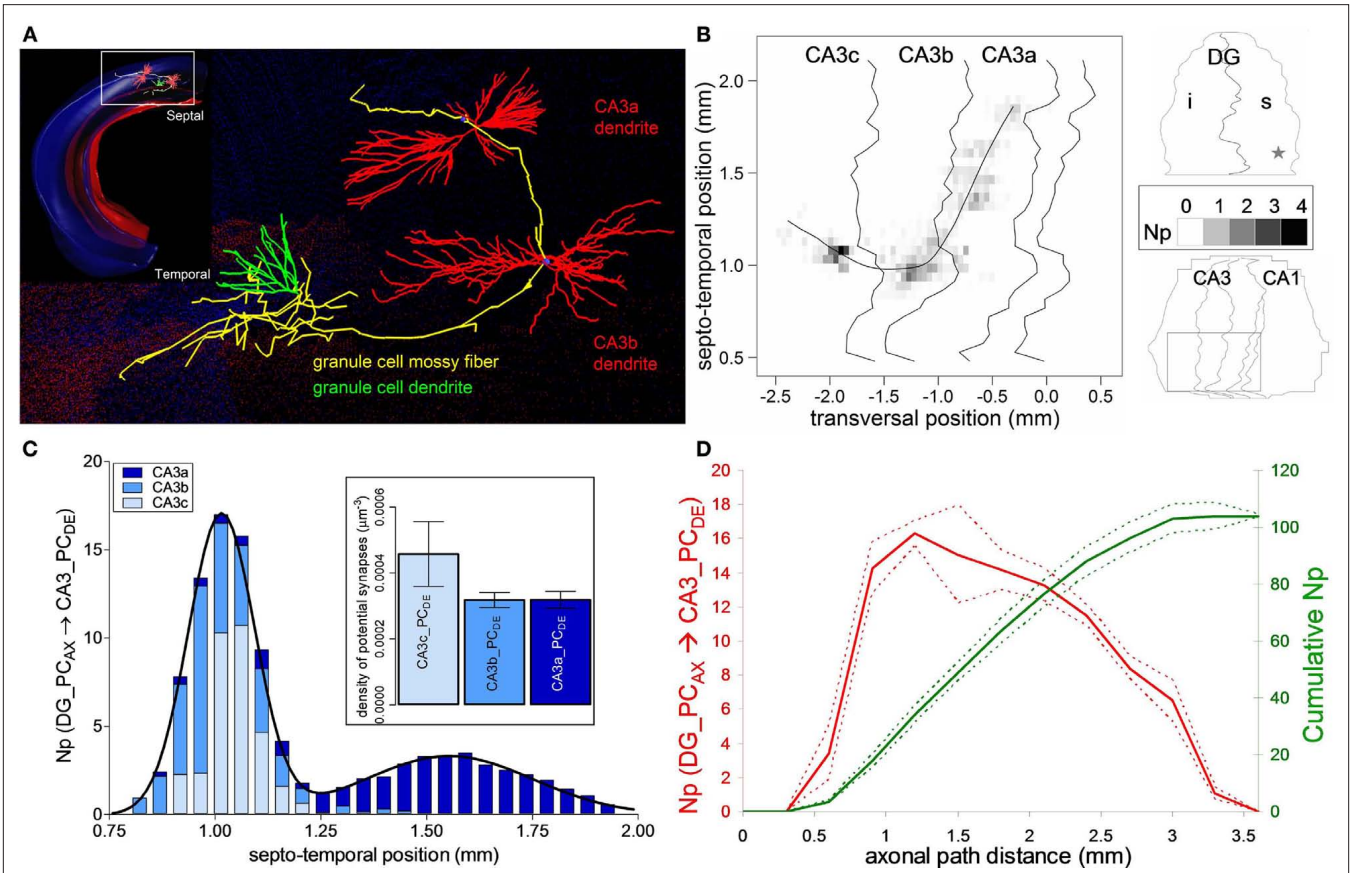


FIGURE 2 | Potential connectivity analysis for DG mossy fiber onto CA3 pyramidal cells. (A) Relative orientation of a dentate granule cell (axon: yellow, dendrite: green) and of the dendritic trees of two pyramidal cells from CA3b and CA3a. Inset shows the arbor embeddings within the 3D hippocampus reconstruction. **(B)** Potential synapse map: The average number (Np) of potential contacts (gray scale) is based on simulations of 25 pre-synaptic positions around the central granule cell location, which is marked by a star in the top (DG) right panel. The bottom (CA3/CA1) right panel demarcates the sub-regions invaded by the axon (square box), corresponding to the main panel. The fitting line is a cubic spline function ($r^2 = 0.85$) of septo-temporal vs. transversal positions weight-averaged by the potential synapse numbers. **(C)** Average number of potential contacts in each of the CA3 sub-regions, fitted with two Gaussian curves. Inset shows the density of potential synapses. **(D)** Number (left axis, red) and cumulative count (right axis, green) of potential synapses along axonal path distance. The solid and dashed colored lines correspond to averages and 95% confidence interval around the mean.

stratum lucidum (Acsady et al., 1998). Here we only considered PCs as the potential post-synaptic targets. Previous research on the organization of the dentate mossy fiber pathway revealed that the main axonal path initially traverses the CA3 region in a septal direction. Then at the CA3b/CA3a border it abruptly changes course to a temporal/caudal orientation relative to the soma position of the GC (Acsady et al., 1998). This peculiar shape of the mossy fibers is evident from the visualization of the axonal embedding within the 3D hippocampal reconstruction (**Figure 2A**) and the potential synaptic map in CA3 (**Figure 2B**). The map is built from 25 GC somatic positions around the center of the spatial range in the supra-pyramidal region corresponding to the reported stereotactic coordinates (star in the DG schematic of **Figure 2B**).

The total number of potential synapses made by the GC on pyramidal dendrites is 103.8 ± 4.1 (mean ± SD, $N = 25$), with an almost equal distribution among the three sub-regions of CA3: 32.2 ± 2.3 in CA3c, 36.6 ± 2.1 in CA3b, and 35.0 ± 1.1 in CA3a. However, both the number of potential contacts and the proportions among the sub-fields vary non-uniformly along the longitudinal axis. In particular, the distribution along the septo-temporal axis follows a double Gaussian (**Figure 2C**). The first peak is centered around the somatic septo-temporal position of the GC, and largely consists of potential synapses in areas CA3c and CA3b. The second peak (displaced ~0.6 mm temporally) is twice as wide but of smaller (~1:5) amplitude, and almost entirely corresponds to potential contacts in CA3a. In contrast to the absolute numbers of potential contacts, the potential synaptic density is significantly higher in CA3c compared to CA3b and CA3a (**Figure 2C** inset, $p < 0.0001$, unpaired two-tailed t-test). The uneven density can be explained by analyzing the number of potential synapses with respect to the axonal path distance (**Figure 2D**). The potential synaptic count rises sharply and peaks at ~1 mm within CA3c, but then decreases farther along the path in correspondence to the CA3b and CA3a sub-regions. These results parallel the empirical observations on mossy bouton and actual synapse counts within each CA3 sub-region (Acsady et al., 1998; Henze et al., 2000).

DISTINCT POTENTIAL SYNAPTIC CONNECTIVITY PATTERNS IN THE PYRAMIDAL CELL NETWORK

The axonal trees of the four examined pyramidal neurons share common general characteristics, at the same time exhibiting exquisitely distinct potential synaptic maps onto the CA3/CA1 principal neuronal network (**Figure 3**). All four arbors project to large portions of the CA3 and CA1 sub-regions, with ~3 mm longitudinal coverage. The numbers of potential synapses, however, vary widely from ~176 K for CA3c^{prox}_PC$_{AX}$ and ~160 K for CA2_PC$_{AX}$, to ~100 K for CA3c^{dist}_PC$_{AX}$ and ~80 K for CA3b_PC$_{AX}$. These differences, considerable even within sub-region, largely reflect variation in axonal length (Ropireddy et al., 2011).

Except for CA3b_PC$_{AX}$, all other three neurons have potential synapses across every CA sub-field, with relative differences in their count of potential contacts between CA3 and CA1. A prominent difference among the maps concerns their orientations with respect to the two natural axes of the hippocampus, reflected by the linear regressions of the weighted averages (**Figure 3** insets). Specifically, no particular trend is apparent for CA3c^{prox}_PC$_{AX}$ (**Figure 3A**). In contrast, CA3c^{dist}_PC$_{AX}$ (**Figure 3B**) and CA2_PC$_{AX}$ (**Figure 3D**) clearly favor septal locations in the proximal transversal regions (CA3 toward the DG), and temporal locations in the distal transversal regions (CA1 toward the subiculum), with a significantly positive correlation ($r = 0.87$ and $r = 0.98$, respectively). Conversely, CA3b_PC$_{AX}$ displays the opposite tendency (**Figure 3C**) due to a fairly uniform spread of CA3 contacts around the somatic position and a ~1 mm septal shift in CA1 ($r = -0.81$).

The distinct patterns of potential connectivity are particularly evident along the transversal axis (**Figure 4A,B**). The CA3c^{prox}_PC$_{AX}$ bimodal distribution peaks in CA3c and in mid-CA1. The multimodal shape of CA3c^{dist}_PC$_{AX}$ reveals a preference for the CA3c and CA3a sub-regions, but tails off within proximal CA1. These contrasting tendencies are summarized in the counts of potential contacts within CA3 (and its sub-fields) and CA1 (bar plots of **Figure 4A**). To quantify the opposite differences statistically, we repeated the embedding of all post-synaptic PCs 10 times with different random seeds for the exact positions and orientations. For the CA3c^{prox}_PC$_{AX}$, the potential synaptic count is 1.5 times larger in

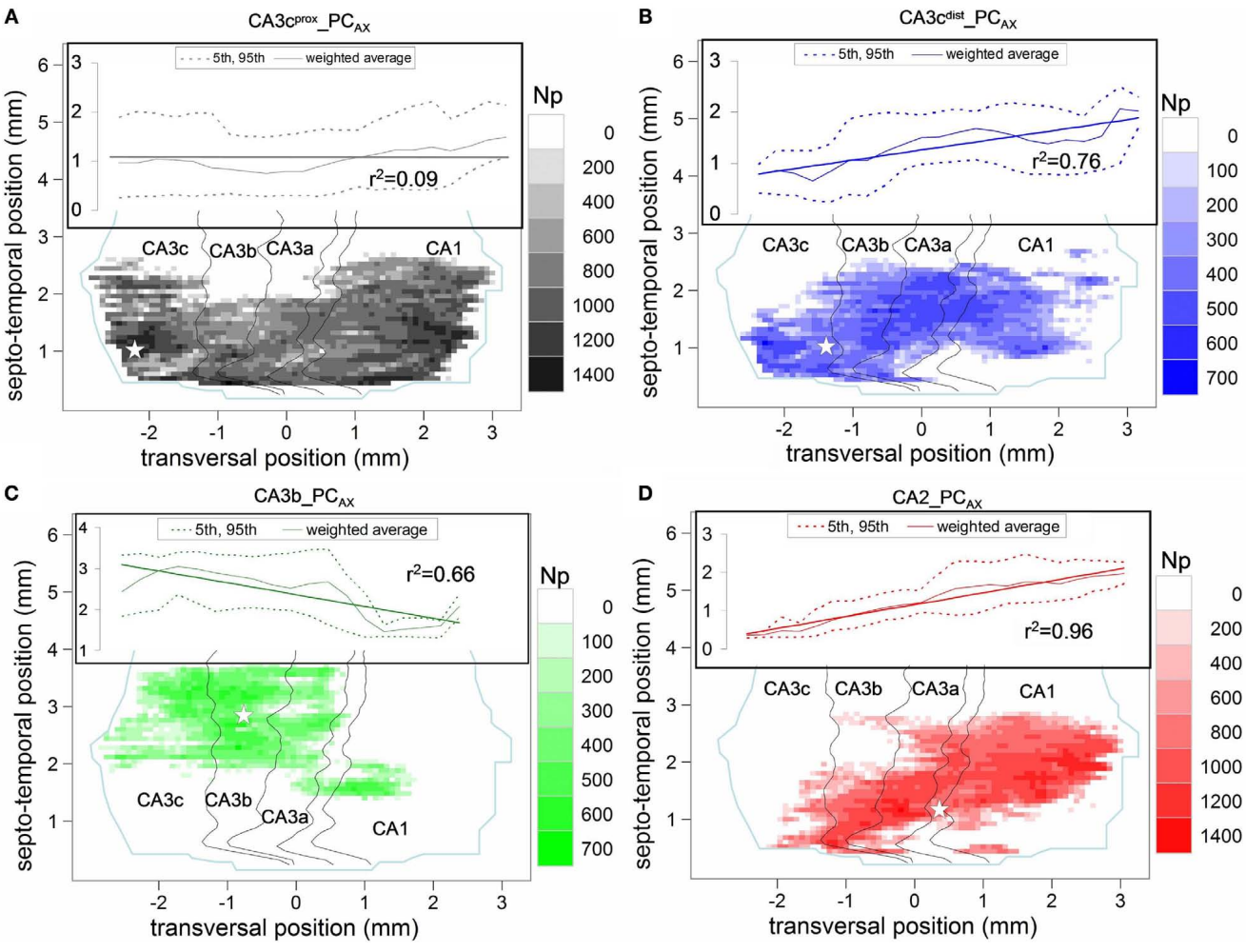


FIGURE 3 | Potential connectivity maps of four pre-synaptic CA3/2 principal neurons onto pyramidal cells. (A) CA3c^{prox}_PC$_{AX}$ (gray), **(B)** CA3c^{dist}_PC$_{AX}$ (blue), **(C)** CA3b_PC$_{AX}$ (green), and **(D)** CA2_PC$_{AX}$ (red). In all maps, the white star denotes the transversal and septo-temporal coordinates of the pre-synaptic soma. Insets show the linear regression and 95% confidence interval of the septo-temporal vs. transversal positions weight-averaged by the potential synapse numbers.

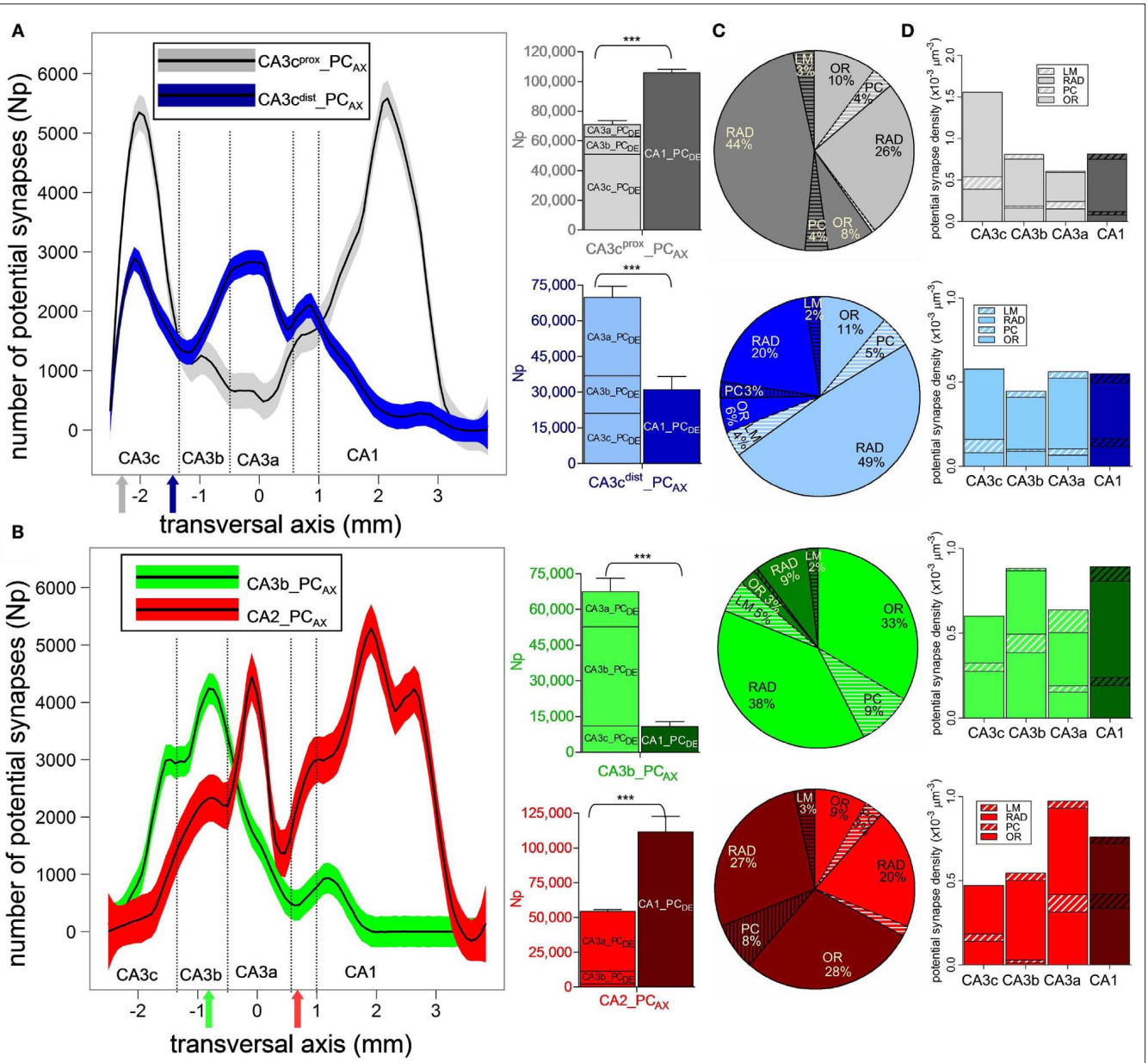


FIGURE 4 | Sub-region and layer specificity of pyramidal cell potential connectivity. (A,B) Potential synapse distributions along the transversal axis for CA3c^{prox}_PC$_{AX}$ (A, gray), CA3c^{dist}_PC$_{AX}$ (A, blue), CA3b_PC$_{AX}$ (B, green), and CA2_PC$_{AX}$ (B, red). The solid colored arrows indicate the soma transversal positions. For each of the four axonal arbors, bar plots summarize the total counts in each CA3 sub-region and in CA1. **(C)** Proportion of potential contacts in CA3 (light) and CA1 (dark) layers. The layers not primarily targeted by these axons (PC and LM) are striped. **(D)** Density of potential synapses in the layers of each CA3 sub-region (light) and CA1 (dark).

CA1 than in CA3, whereas for CA3c^{dist}_PC$_{AX}$ this ratio is 0.45 (both $p < 0.0001$, unpaired two-tailed t-test). Moreover, CA3c^{prox}_PC$_{AX}$ makes three-fourth of its potential synapses in CA3c, while half of the CA3c^{dist}_PC$_{AX}$ potential contacts is in CA3a. The CA3b_PC$_{AX}$ neuron exhibits the most localized potential connectivity, peaking around the somatic position in CA3b and extending nearly an order of magnitude fewer potential synapses in CA1 than in CA3. The CA2_PC$_{AX}$ transversal distribution is bimodal with peaks in CA3a and mid-CA1. The total number of CA1 potential contacts is twice that in CA3 ($p < 0.0001$).

The four PCs also display general similarities and distinct differences in potential connectivity across layers (**Figure 4C**). As expected, the vast majority of potential synapses are always found in radiatum (RAD) and oriens (OR), with only small fractions in the pyramidal (PC) and LM layers. The CA3 and CA1 radiatum proportions notably differentiate the four pre-synaptic neurons. For CA3c^{prox}_PC$_{AX}$, the RAD potential contacts are 1.7 times more abundant in CA1 than in CA3, while the inverse ratio (0.4) applied to CA3c^{dist}_PC$_{AX}$. In both neurons, only less than or close to one-tenth of potential synapses are contributed by OR, as opposed

to more than two-thirds by RAD. For CA3b_PC_{AX}, radiatum is still dominant in the minor CA1 component, but for the major CA3 contribution, oriens and radiatum provide an approximately even number of potential synapses. A mirror situation occurs in CA2_PC_{AX}, where radiatum has greater representation in the minor CA3 component, but oriens claims an equal split in the synaptically richer CA1 field. For both CA3b_PC_{AX} and CA2_PC_{AX}, less than 50% of potential synapses are found in RAD, and more than a third in OR.

In addition to the count of potential contacts, the potential synapse densities also vary across sub-regions and layers among the four axons (**Figure 4D**). For CA3c^{prox}_PC_{AX}, the highest density of potential synapses occurs in CA3c even though the total count is greater in CA1. In contrast, CA3c^{dist}_PC_{AX} shows similar potential synapse densities in all CA3 and CA1 sub-regions. For CA3b_PC_{AX}, the potential synapse density is highest in CA3b and CA1, again in spite of the lower CA1 total count, indicating a focal concentration of potential connectivity within a smaller area (cf. **Figure 4B**). The CA2_PC_{AX} potential synapse density is maximal in CA3a followed by CA1, where more than two-thirds of the potential contacts are made. Thus, each of the four neurons displays unique patterns of preferred density domains. Conversely, in all cases and for each sub-field, the potential synaptic density is substantially greater in radiatum than in oriens with only three exceptions in which the two values are similar: CA3c and CA3a for CA3b_PC_{AX}, and CA1 for CA2_PC_{AX}.

POTENTIAL SYNAPSE DISTRIBUTIONS ALONG THE AXONAL AND DENDRITIC PATHS

The distance from the soma of a pre-synaptic terminal along the axonal path may affect spike propagation delay and reliability. Thus, we investigated how the dependence on axonal distance of potential connectivity varies among neurons (**Figure 5A**). The four PCs exhibit linear increases with similar slopes in the cumulative

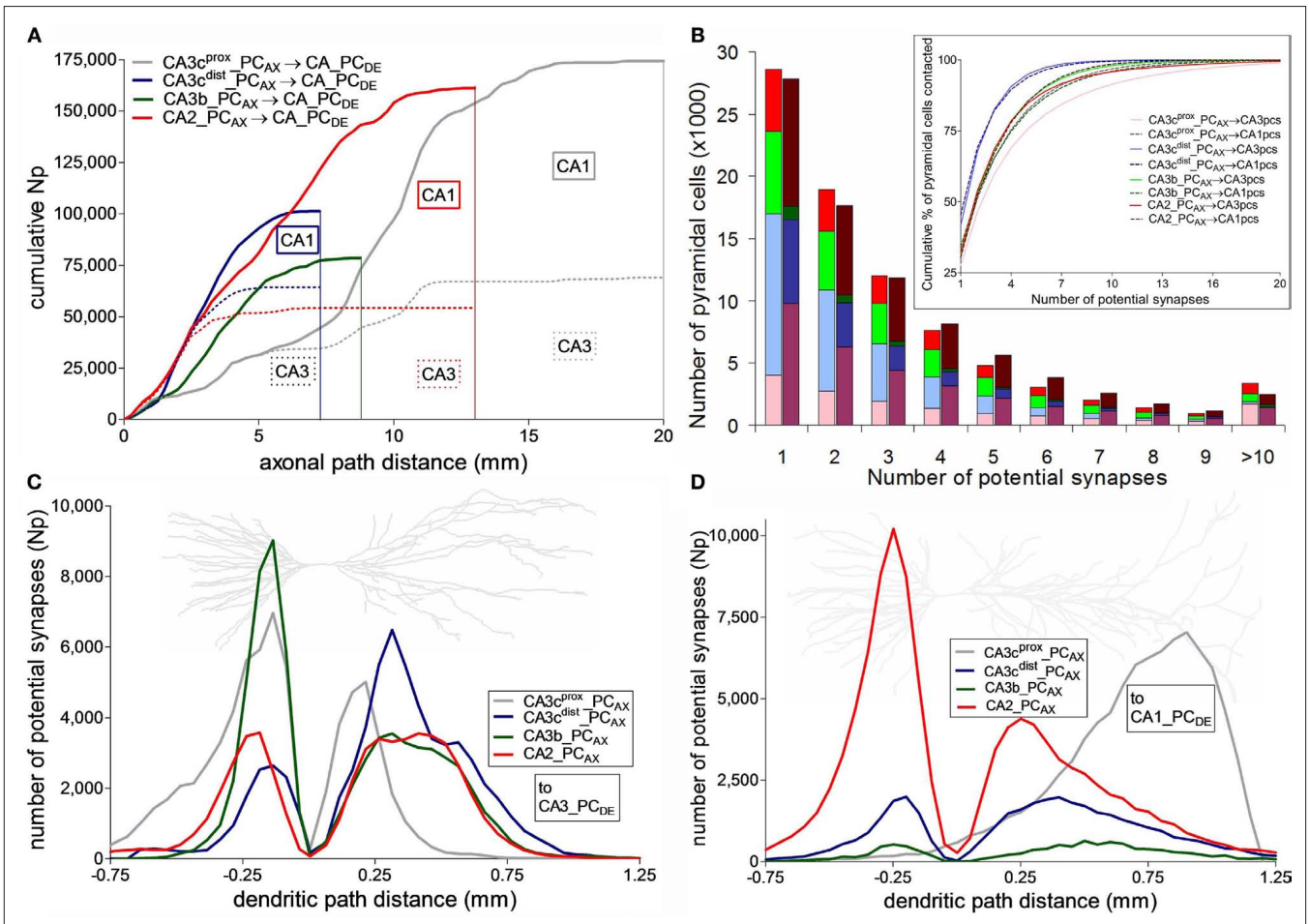


FIGURE 5 | Distribution of potential synapses along axonal and dendritic paths. (A) Cumulative count of potential synapses made by the four CA3/2 pyramidal cells as a function of their axonal path distances from the soma. The dotted colored lines correspond to the potential contacts onto CA3 pyramidal cells. The difference between the solid and dotted lines represent the cumulative potential synapse numbers onto the CA1 network. **(B)** Distribution of the number of potential synapses received by the CA3 (light) and CA1 (dark) pyramidal cells contacted by any of the four axonal arbors. Inset shows the cumulative proportion of pyramidal cells contacted with up to a given number of potential synapses. **(C,D)** Potential synapse distributions for the four CA3/2 axonal arbors along the dendrites of CA3 **(C)** and CA1 **(D)** pyramidal cells. Negative and positive path distances correspond to basal and apical dendrites, respectively. The background neuron images are representative dendritic arbors of CA3 and CA1 pyramidal cells from NeuroMorpho.Org.

count of potential synapses up to ~1.5 mm, indicating a uniform distribution of potential pre-synaptic contacts in the first part of the axonal path. Beyond this distance, the neuron with the longest axon ($CA3c^{prox}\_PC_{AX}$) has a shallow accumulation of potential contacts (mostly in CA3) within the more proximal one-third of path (the initial ~7 mm), and makes progressively more CA1 potential synapses in the next ~7 mm. The neuron with the fewest potential contacts ($CA3b\_PC_{AX}$) is also essentially confined within CA3 and in fact ends shortly after the initial 7 mm of path, though it displays a faster potential synapse increase reaching a plateau in the first ~5 mm. Within this same path, the other two neurons have the steepest (and very similar) accumulation of potential contacts. The neuron with the shortest path ($CA3c^{dist}\_PC_{AX}$) synapses mostly in CA3 for the first half of its path and only in CA1 for the second half. $CA2\_PC_{AX}$ also forms potential contacts exclusively in CA3 for the first ~2.5 mm, but then continues for ~10 mm after switching to CA1.

The four axonal arbors also differ in terms of the proportion of post-synaptic targets they contact, either out of all (225,000 CA3 and 390,000 CA1) PCs or only considering those with dendrites within the spatial reach of each axon (**Table 1**). These numbers vary widely among neurons and between CA3 and CA1. In contrast, the spatial coverage is similar among the four axons in CA3 (~40%) and CA1 (~25%). The average number of potential contacts per post-synaptic neuron ranged from ~2.2 to 5.6, with small coefficients of variation (typically below 0.1). Thus, most neurons receiving potential synapses tended to be contacted multiple times. In fact, individual connections only accounted for one-third of the cases. However, the distributions of the number of potential synapses per contacted PC varied dramatically depending on the pre-synaptic axon (**Figure 5B**). Interestingly, the grand average of potential synapses per PC over the whole post-synaptic pool (the product of overall % PCs contacted by their averaged received contacts) is similar for all axonal arbors on CA3 pyramidal dendrites (~0.3) but not on CA1 pyramidal dendrites, where it spans an order of magnitude from 0.29 for $CA2\_PC_{AX}$ to 0.03 for $CA3b\_PC_{AX}$. These divergence ratios are consistent with previous findings (Li et al., 1994).

The distance of a synapse from the soma along the dendritic path can affect signal integration both by cable filtering and active membrane properties. The distributions of potential synapses in the apical and basal dendrites of PCs reflect the patterns observed in the oriens and radiatum layers, respectively, in CA3 (**Figure 5C**) and CA1 (**Figure 5D**). All four neurons have a peak of potential synapses on the basal dendrites at a similar distance along the path (~200 μm). However, $CA3b\_PC_{AX}$ and $CA3c^{prox}\_PC_{AX}$ have the largest number of basal potential contacts in CA3 and the smallest in CA1. In contrast, $CA2\_PC_{AX}$ is dominant in CA1, and $CA3c^{dist}\_PC_{AX}$ has a similar number in both regions. The four neurons display more similar distributions of potential contacts along the CA3 apical paths, except for the more proximal peak of $CA3c^{prox}\_PC_{AX}$. The patterns along the CA1 apical paths, in contrast, are completely distinct in the four cases. $CA3c^{prox}\_PC_{AX}$ and $CA2\_PC_{AX}$ have the largest and second largest peaks, but are located at opposite extremes (distal and proximal, respectively) of the apical dendrites, almost 750 μm apart. The remaining two CA3 neurons have smaller peaks (especially $CA3b\_PC_{AX}$) at intermediate path distances.

POTENTIAL SYNAPSE DISTRIBUTIONS OF RADIATUM AND LACUNOSUM-MOLECULARE CA3b INTERNEURONS

Interneurons with somata in CA3b radiatum and LM layers have distinct electrophysiology and synaptic plasticity, yet similar axonal morphology (Ascoli et al., 2009). The collective potential synaptic map for the 13 interneurons (6 $CA3b^{R}\_IN_{AX}$ and 7 $CA3b^{LM}\_IN_{AX}$) depicts localized CA3 distributions spanning approximately the same spatial extent in the septo-temporal and longitudinal dimensions (**Figure 6A**). $CA3b^{R}\_IN_{AX}$ and $CA3b^{LM}\_IN_{AX}$ make approximately 50,000 and 40,000 potential contacts, respectively, with CA3 PC dendrites. The difference between these mean values was not statistically significantly ($p > 0.5$, unpaired two-tailed *t*-test). The spatial extent in the septo-temporal and transversal axes were 0.60 ± 0.09 mm and 2.91 ± 0.55 mm (mean ± SD, $N = 13$), respectively. The potential synaptic map area was 1.38 ± 0.23 mm^2. The coefficient of variation for these three parameters is thus under 0.2. We also computed the mean square deviation (MSD) for the same three parameters by comparing the connectivity map of each neuron with the convolved map (shown in **Figure 6A**). The corresponding measurements were 0.02 mm, 0.31 mm, and 0.09 mm^2, demonstrating a modest variation among these 13 CA3b interneurons.

Interestingly, however, the two groups of interneurons can be clearly differentiated based on their potential connectivity patterns along the transverse axis (**Figure 6B**). The number of potential

Table 1 | Percentage of pyramidal cells contacted by the four axonal CA3/CA2pc arbors and number of potential synapses they receive.

		$CA3c^{prox}\_PC_{AX}$	$CA3c^{dist}\_PC_{AX}$	$CA3b\_PC_{AX}$	$CA2\_PC_{AX}$
CA3	Overall PCs contacted (%)	5.9	12.0	8.9	4.9
	Np per PC contacted ($\mu \pm \sigma$)	5.6 ± 0.20	2.5 ± 0.18	3.4 ± 0.30	5.1 ± 0.12
	Spatial coverage (%)	41	39	45	42
	PCs contacted in spatial coverage (%)	14.4	30.8	19.8	11.7
CA1	Overall PCs contacted (%)	9.0	3.7	0.8	7.3
	Np per PC contacted ($\mu \pm \sigma$)	3.0 ± 0.07	2.2 ± 0.40	3.8 ± 0.15	4.0 ± 0.39
	Spatial coverage (%)	24	23	26	24
	PCs contacted in spatial coverage (%)	37.5	16.1	3.1	30.5

The "overall % PCs contacted" is computed based on a total of 225,000 and 390,000 CA3 and CA1 pyramidal neurons respectively. The "Np per PC contacted" only counts the target PCs receiving potential synapses. "Spatial coverage" represents the proportion of PCs with dendrites within the spatial reach of each pre-synaptic axon.

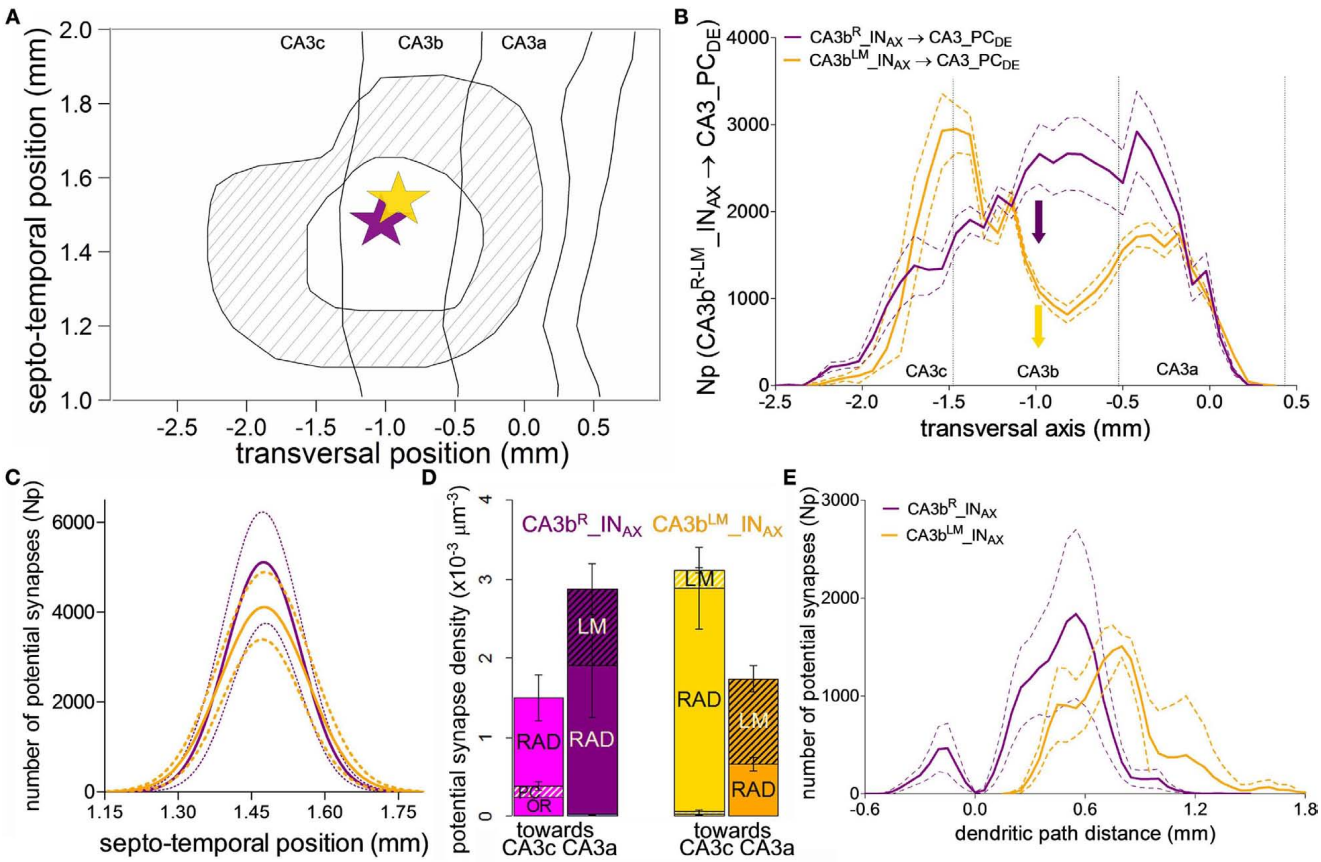


FIGURE 6 | Potential connectivity analysis for CA3b interneurons onto CA3 pyramidal cells. (A) Map of the convolved common area of potential synapses formed by the axons of 13 CA3b interneurons with somata in the radiatum (R) or lacunosum-moleculare (LM) layers onto CA3 pyramidal cell dendrites. The inner closed line represents the area in which *all* pre-synaptic cells establish potential synapses. The outer close line delineates the region in which *any* of these interneurons makes potential contacts. The star marks the average somatic location of all 13 cells. The size of the star approximately corresponds to the location range. There were no statistically significant differences in the septo-temporal and transversal positions between the R and LM neurons. **(B)** Mean potential synapse distributions and 95% confidence intervals for six $CA3b^{R}\_IN_{AX}$ (purple) and seven $CA3b^{LM}\_IN_{AX}$ (gold) along the transversal axis. The solid colored arrows indicate the average transversal positions of the somata. The vertical dashed lines demarcate the boundaries between the three CA3 sub-regions. **(C)** Gaussian fits of the means and 95% confidence intervals for the septo-temporal distributions of potential synapses. **(D)** Transversal patterns ("toward CA3c" vs. "toward CA3a") and layer specificity in potential synapse densities for the two CA3b interneuron types. **(E)** Potential synapse distribution along the dendritic path distance of CA3 pyramidal cells.

synapses from $CA3b^{R}\_IN_{AX}$ gradually increases from CA3c, peaking just past the CA3b/CA3a border, and dropping sharply within CA3a. In contrast, $CA3b^{LM}\_IN_{AX}$ displays a bimodal distribution, with a prominent potential synapse peak at the CA3c/CA3b border, a dip in the middle of CA3b, and a secondary peak in CA3a. In contrast, the potential connectivity patterns along the septo-temporal position are symmetric around the somata and indistinguishable between the two interneuron types (**Figure 6C**).

The differences along the transverse axis suggest opposite propensities of $CA3b^{R}\_IN_{AX}$ and $CA3b^{LM}\_IN_{AX}$ to make potential contacts toward CA3c and CA3a, respectively, as evidenced by potential synaptic density analysis (**Figure 6D**). Further inspection of layer specificity shows in the density bar graphs that the minor proportion of potential contacts in the oriens layer is unique of $CA3b^{R}\_IN_{AX}$ cells, and entirely confined to the CA3c region. In contrast, in both groups the concentration of potential contacts in LM grows toward CA3a. The lack of $CA3b^{LM}\_IN_{AX}$ potential synapses in oriens is reflected in the distribution along the dendritic path distance of the target PCs (**Figure 6E**), which is similar for the two interneuron types on the apical, but not basal arbors. Lastly, interneurons with higher and lower dendritic branch numbers (called HiDe and LoDe in Ascoli et al., 2009) cannot be distinguished by their potential connectivity (not shown).

DISCUSSION

The approach introduced here enables the analysis of potential connectivity patterns from individual axo-dendritic overlaps across the entire hippocampus. The spatial scale defined by the span of whole axonal arborizations bridges the conceptual levels of potential synapses and regional anatomy. The hippocampus is especially suitable for studying systems-level connectivity, because of its unique structural organization and relative wealth of cellular morphological data. At the same time, the highly diverse hippocampal principal cells and interneurons form complex

microcircuits whose computational function is only beginning to be understood. Functionally, the hippocampus is theorized to play a central role in spatial navigation and memory processing. Thus, our results might facilitate incorporation of realistic connectivity patterns into models investigating structure–function relationship.

We observe unique distributions of potential synapses across the various sub-regions and cyto-architectonic layers of the hippocampus, based on entire axonal reconstructions from seven neuron types and the dendritic arbors representing the whole CA PC network. Although axonal morphology constitutes the most prominent determinant of synaptic connectivity, the present study demonstrates the importance of integration both across data sources and data types. The 3D whole-hippocampus reconstruction enables the embedding of complete axonal arbors traced from disparate histological preparations into the same framework (e.g., Scorcioni and Ascoli, 2005; Ascoli et al., 2009; Ropireddy et al., 2011). Moreover, the dense dendritic embedding allows actual computation of potential synaptic maps, as well as quantitative analysis such as the dendritic and axonal distributions of potential contacts and the ratio between the numbers of actual and potential synapses.

In this analysis, the pre-synaptic neuron, the post-synaptic neuron, and the embedding tissue are reconstructed from different animals. Thus, this framework cannot capture the specific correlations among pairs of cells within their surroundings. At the same time, potential connectivity patterns can reveal general statistical principles of the specific interaction probabilities among particular cell classes that might apply across individuals. In particular, if validated with a larger sample, the CA3 axonal potential connectivity might indicate that proximal CA3c neurons (toward hilus) communicate primarily with CA1, while distal CA3 neurons (e.g., in CA3b) mostly form recurrent axo-dendritic overlaps within CA3. Moreover, the strikingly different potential connectivity patterns of the two CA3c neurons (proximal and distal) suggest that even principal cells from the same sub-region can have contrasting potential synaptic distributions throughout the hippocampus. While this finding may have interesting implications on the information processing in CA3 and CA1, the empirical observation must be corroborated on a representative sample of neurons before generalizing to the whole population of proximal and distal CA3 PCs.

We adopted the concept of potential synapses to characterize circuitry based on light microscopy data. Functionally, potential connectivity relates to the capability to create actual synapses leveraging the structural plasticity of dendritic spines and axonal boutons (Holtmaat and Svoboda, 2009). Evidence for this phenomenon is particularly compelling in the hippocampus, where it might bear direct functional relevance to information storage (Chklovskii et al., 2004). From this viewpoint, it is interesting to evaluate the ratio of actual to potential synapses, called connectivity fraction (Escobar et al., 2008) or filling fraction (Stepanyants et al., 2002). The only available dense electron microscopy reconstruction of a small hippocampal volume from the rat CA1 radiatum layer (Mishchenko et al., 2010) confirmed earlier estimates of connectivity fraction in the 0.20–0.25 range. Considering, as in these recent and earlier reports (Bannister and Larkman, 1995; Megias et al., 2001), bouton and/or spine counts as proxies for actual synapse numbers, we obtain values fully consistent with the existing data for that region. At the same time, our results can be further used to derive the connectivity fractions for other components of the hippocampal circuit.

Dentate mossy fibers are estimated to form ~18 mossy terminals each (Acsady et al., 1998; Henze et al., 2000), leading to connectivity fractions of 0.17–0.18 in all three sub-fields CA3a,b,c. Similarly, the bouton numbers reported for the proximal CA3c pyramidal axon (Wittner et al., 2007) combined with our potential synapse count yield an overall connectivity fraction of ~0.23. However, this value changes drastically throughout the hippocampus, from 0.06 in CA1 OR to ~0.17 in CA3 OR and RAD, to 0.32 in CA1 RAD. This differentiation by sub-region and layer emphasizes the importance of analyzing connectivity through the entire anatomical region invaded by an axonal arborization as opposed to the narrow region of interest typically covered by electron microscopy.

Summing together the potential synapses made by the four analyzed pyramidal axons on all PC dendrites, ~26% of the contacts are made in the oriens layer, and ~59% in radiatum. These fractions, virtually identical in CA3 and CA1, can be compared to the relative densities of dendritic spines observed in the same layers. The values in CA3 (Drakew et al., 1996) constitute an almost perfect match, with 28% of spines in oriens and 59% in radiatum. Such correspondence suggests equal overall connectivity fraction between basal and apical dendrites. Interestingly, the proportion of CA1 spines found in radiatum (~55%) is also very similar to the potential synapse fraction, while the value in oriens (~39%) is higher (Bannister and Larkman, 1995; Megias et al., 2001), consistent with the addition of recurrent collaterals from CA1 PCs on the basal dendrites.

A drawback of this study is that the potential connectivity patterns are based on only one axonal reconstruction for each principal neuron. This is due to the difficulty of obtaining complete digital reconstructions of projecting axons from extremely laborious *in vivo* preparations. With the more contained interneurons, axons can be reconstructed from slices, enabling analysis of larger samples. In the case of CA3b Radiatum and LM cells, we found different potential synaptic patterns across the transversal axis (i.e., among the CA3 sub-fields). Our previous intrinsic morphometric analysis could not detect these differences between the two groups (Ascoli et al., 2009). This observation stresses the importance of studying connectivity patterns within the context of a 3D system-level anatomical framework.

Based on the axonal and dendritic reconstruction data available in NeuroMorpho.Org, our computational framework can allow the estimation of potential interconnectivity of additional neuron classes in the rodent hippocampus. In particular, complete axonal reconstructions of principal neurons from posterior–ventral region of CA3, CA1a, CA1b, CA1c, subiculum, and entorhinal cortex layer 2 (e.g., Scorcioni and Ascoli, 2005) can be embedded to map their hippocampal potential connectivity. Similarly, 18 axonal reconstructions of interneurons from DG hilus, CA3 stratum lucidum, and CA1 stratum oriens can also be included in future studies. Additionally, 223 dendritic reconstructions of interneurons throughout all DG, CA3, and CA1 sub-fields and layers can be embedded to extend all potential connectivity maps

beyond the PC targets examined here. It is also expected that the number of available reconstructions will continue to grow, enabling further refinements of these results.

Comparing the potential connectivity obtained for the same pre-synaptic cells, but using the dendritic trees of different neuronal classes (e.g., interneurons vs. PCs), can also be useful as a control to examine the role of dendritic specificity in establishing potential synapses. This could be achieved, for example, by contrasting the connectivity patterns of CA3 pyramidal axons to CA1 basket cell dendrites with those to CA1 PC dendrites. An alternative possibility is to use random cable for the dendrite as a control. These extensions will be explored in future investigations.

ACKNOWLEDGMENTS

We are indebted to Dr. Diek Wheeler and Mr. David Hamilton for their valuable feedback on an earlier version of this manuscript. Grant sponsors: National Institute of Health NS39600 and NS058816, Office of Naval Research MURI N00014-10-1-0198 to Giorgio A. Ascoli.

REFERENCES

Acsady, L., Kamondi, A., Sik, A., Freund, T., and Buzsaki, G. (1998). GABAergic cells are the major post-synaptic targets of mossy fibers in the rat hippocampus. *J. Neurosci.* 18, 3386–3403.

Amaral, D. G., and Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* 31, 571–591.

Amirikian, B. (2005). A phenomenological theory of spatially structured local synaptic connectivity. *PLoS Comput. Biol.* 1, e11. doi: 10.1371/journal.pcbi.0010011

Ascoli, G. A., Brown, K. M., Calixto, E., Card, J. P., Galvan, E. J., Perez-Rosello, T., and Barrionuevo, G. (2009). Quantitative morphometry of electrophysiologically identified CA3b interneurons reveals robust local geometry and distinct cell classes. *J. Comp. Neurol.* 515, 677–695.

Ascoli, G. A., Donohue, D. E., and Halavi, M. (2007). NeuroMorpho.org: a central resource for neuronal morphologies. *J. Neurosci.* 27, 9247–9251.

Bannister, N. J., and Larkman, A. U. (1995). Dendritic morphology of CA1 pyramidal neurones from the rat hippocampus: II. Spine distributions. *J. Comp. Neurol.* 360, 161–171.

Bressler, S. L., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci. (Regul. Ed.)* 14, 277–290.

Brown, K. M., Donohue, D. E., D'Alessandro, G., and Ascoli, G. A. (2005). A cross-platform freeware tool for digital reconstruction of neuronal arborizations from image stacks. *Neuroinformatics* 3, 343–360.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198.

Buzsaki, G. (2007). The structure of consciousness. *Nature* 446, 267.

Carnevale, N. T., Tsai, K. Y., Claiborne, B. J., and Brown, T. H. (1997). Comparative electrotonic analysis of three classes of rat hippocampal neurons. *J. Neurophysiol.* 78, 703–720.

Claiborne, B. J., Amaral, D. G., and Cowan, W. M. (1990). Quantitative, three-dimensional analysis of granule cell dendrites in the rat dentate gyrus. *J. Comp. Neurol.* 302, 206–219.

Chklovskii, D. B., Mel, B. W., and Svoboda, K. (2004). Cortical rewiring and information storage. *Nature* 431, 782–788.

Drakew, A., Muller, M., Gahwiler, B. H., Thompson, S. M., and Frotscher, M. (1996). Spine loss in experimental epilepsy: quantitative light and electron microscopic analysis of intracellularly stained CA3 pyramidal cells in hippocampal slice cultures. *Neuroscience* 70, 31–45.

Eichenbaum, H., and Cohen, N. J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain.* Upper Saddle River, NJ: Oxford University.

Escobar, G., Fares, T., and Stepanyants, A. (2008). Structural plasticity of circuits in cortical neuropil. *J. Neurosci.* 28, 8477–8488.

Fiala, J. C. (2005). Reconstruct: a free editor for serial section microscopy. *J. Microsc.* 218(Pt 1), 52–61.

Hanson, A. J. (2006). *Visualizing Quaternions.* San Francisco: Morgan Kaufmann Publishers.

Harris, K. M. (1999). Structure, development, and plasticity of dendritic spines. *Curr. Opin. Neurobiol.* 9, 343–348.

Harris, K. M., and Stevens, J. K. (1989). Dendritic spines of CA1 pyramidal cells in the rat hippocampus: serial electron microscopy with reference to their biophysical characteristics. *J. Neurosci.* 9, 2982–2997.

Henze, D. A., Cameron, W. E., and Barrionuevo, G. (1996). Dendritic morphology and its effects on the amplitude and rise-time of synaptic signals in hippocampal CA3 pyramidal cells. *J. Comp. Neurol.* 369, 331–344.

Henze, D. A., Urban, N. N., and Barrionuevo, G. (2000). The multifarious hippocampal mossy fiber pathway: a review. *Neuroscience* 98, 407–427.

Holtmaat, A., and Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nat. Rev. Neurosci.* 10, 647–658.

Honey, C. J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J. P., Meuli, R., and Hagmann, P. (2009). Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2035–2040.

Ishizuka, N., Cowan, W. M., and Amaral, D. G. (1995). A quantitative analysis of the dendritic organization of pyramidal cells in the rat hippocampus. *J. Comp. Neurol.* 362, 17–45.

Ishizuka, N., Weber, J., and Amaral, D. G. (1990). Organization of intrahippocampal projections originating from CA3 pyramidal cells in the rat. *J. Comp. Neurol.* 295, 580–623.

Jaffe, D. B., and Carnevale, N. T. (1999). Passive normalization of synaptic integration influenced by dendritic architecture. *J. Neurophysiol.* 82, 3268–3285.

Kalisman, N., Silberberg, G., and Markram, H. (2003). Deriving physical connectivity from neuronal morphology. *Biol. Cybern.* 88, 210–218.

Kalisman, N., Silberberg, G., and Markram, H. (2005). The neocortical microcircuit as a tabula rasa. *Proc. Natl. Acad. Sci. U.S.A.* 102, 880–885.

Knott, G. W., Holtmaat, A., Wilbrecht, L., Welker, E., and Svoboda, K. (2006). Spine growth precedes synapse formation in the adult neocortex in vivo. *Nat. Neurosci.* 9, 1117–1124.

Lendvai, B., Stern, E. A., Chen, B., and Svoboda, K. (2000). Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex in vivo. *Nature* 404, 876–881.

Li, X. G., Somogyi, P., Ylinen, A., and Buzsaki, G. (1994). The hippocampal CA3 network: an in vivo intracellular labeling study. *J. Comp. Neurol.* 339, 181–208.

Megias, M., Emri, Z., Freund, T. F., and Gulyas, A. I. (2001). Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells. *Neuroscience* 102, 527–540.

Mishchenko, Y., Hu, T., Spacek, J., Mendenhall, J., Harris, K. M., and Chklovskii, D. B. (2010). Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. *Neuron* 67, 1009–1020.

Press, W. H. (1988). *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge: Cambridge University Press.

Pyapali, G. K., Sik, A., Penttonen, M., Buzsaki, G., and Turner, D. A. (1998). Dendritic properties of hippocampal CA1 pyramidal neurons in the rat: intracellular staining in vivo and in vitro. *J. Comp. Neurol.* 391, 335–352.

Pyapali, G. K., and Turner, D. A. (1994). Denervation-induced dendritic alterations in CA1 pyramidal cells following kainic acid hippocampal lesions in rats. *Brain Res.* 652, 279–290.

Pyapali, G. K., and Turner, D. A. (1996). Increased dendritic extent in hippocampal CA1 neurons from aged F344 rats. *Neurobiol. Aging* 17, 601–611.

Rapp, P. R., and Gallagher, M. (1996). Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. *Proc. Natl. Acad. Sci. U.S.A.* 93, 9926–9930.

Ropireddy, D., Bachus, S., Scorcioni, R., and Ascoli, G. A. (2008). "Computational neuroanatomy of the rat hippocampus: implications and application to epilepsy," in *Computational Neuroscience in Epilepsy*, eds I. Soltesz and K. Staley (San Diego: Elsevier), 71–85.

Ropireddy, D., Scorcioni, R., Lasher, B., Buzsaki, G., and Ascoli, G. A. (2011). Axonal morphometry of hippocampal pyramidal neurons semi-automatically reconstructed after in vivo labeling in different CA3 locations. *Brain Struct. Funct.* 216, 1–15.

Scorcioni, R., and Ascoli, G. A. (2005). Algorithmic reconstruction of complete axonal arborizations in rat hippocampal neurons. *Neurocomputing* 65–66, 15–22.

Scorcioni, R., Bouteiller, J., and Ascoli, G. A. (2002). A real-scale anatomical model of the dentate gyrus

based on single cell reconstructions and 3D rendering of a brain atlas. *Neurocomputing* 44–46, 629–634.

Sik, A., Tamamaki, N., and Freund, T. F. (1993). Complete axon arborization of a single CA3 pyramidal cell in the rat hippocampus, and its relationship with postsynaptic parvalbumin-containing interneurons. *Eur. J. Neurosci.* 5, 1719–1728.

Silberschatz, A., Galvin, P. B., and Gagne, G. (2009). *Operating System Concepts.* San Francisco, CA: Wiley, John & Sons.

Spacek, J., and Hartmann, M. (1983). Three-dimensional analysis of dendritic spines. I. Quantitative observations related to dendritic spine and synaptic morphology in cerebral and cerebellar cortices. *Anat. Embryol.* 167, 289–310.

Sporns, O. (2010). *Networks of the Brain.* Cambridge: MIT Press.

Sporns, O., Tononi, G., and Kotter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042

Squire, L. R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science* 253, 1380–1386.

Stepanyants, A., and Chklovskii, D. B. (2005). Neurogeometry and potential synaptic connectivity. *Trends Neurosci.* 28, 387–394.

Stepanyants, A., Hirsch, J. A., Martinez, L. M., Kisvarday, Z. F., Ferecsko, A. S., and Chklovskii, D. B. (2008). Local potential connectivity in cat primary visual cortex. *Cereb. Cortex* 18, 13–28.

Stepanyants, A., Hof, P. R., and Chklovskii, D. B. (2002). Geometry and structural plasticity of synaptic connectivity. *Neuron* 34, 275–288.

Stepanyants, A., Tamas, G., and Chklovskii, D. B. (2004). Class-specific features of neuronal wiring. *Neuron* 43, 251–259.

Tamamaki, N., Abe, K., and Nojyo, Y. (1988). Three-dimensional analysis of the whole axonal arbors originating from single CA2 pyramidal neurons in the rat hippocampus with the aid of a computer graphic technique. *Brain Res.* 452, 255–272.

Tamamaki, N., and Nojyo, Y. (1991). Crossing fiber arrays in the rat hippocampus as demonstrated by three-dimensional reconstruction. *J. Comp. Neurol.* 303, 435–442.

Turner, D. A., Li, X. G., Pyapali, G. K., Ylinen, A., and Buzsaki, G. (1995). Morphometric and electrical properties of reconstructed hippocampal CA3 neurons recorded in vivo. *J. Comp. Neurol.* 356, 580–594.

Wierenga, C. J., Becker, N., and Bonhoeffer, T. (2008). GABAergic synapses are formed without the involvement of dendritic protrusions. *Nat. Neurosci.* 11, 1044–1052.

Wittner, L., Henze, D. A., Zaborszky, L., and Buzsaki, G. (2007). Three-dimensional reconstruction of the axon arbor of a CA3 pyramidal cell recorded and filled in vivo. *Brain Struct. Funct.* 212, 75–83.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 27 April 2011; accepted: 20 June 2011; published online: 04 July 2011.

Citation: Ropireddy D and Ascoli GA (2011) Potential synaptic connectivity of different neurons onto pyramidal cells in a 3D reconstruction of the rat hippocampus. Front. Neuroinform. ***5****:5. doi: 10.3389/fninf.2011.00005*

frontiers in
NEUROINFORMATICS

REVIEW ARTICLE
published: 27 July 2011
doi: 10.3389/fninf.2011.00007

The retrosplenial cortex: intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome

Jørgen Sugar[1], Menno P. Witter[1], Niels M. van Strien[2†] and Natalie L. M. Cappaert[3*†]

[1] *Kavli Institute for Systems Neuroscience, Centre for the Biology of Memory, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway*
[2] *FMRI-group, Department of Circulation and Medical Imaging, Faculty of Medicine, Norwegian University of Science and Technology, Trondheim, Norway*
[3] *Center for NeuroScience, Swammerdam Institute for Life Science, University of Amsterdam, Netherlands*

Edited by:
Trygve B. Leergaard, University of Oslo, Norway

Reviewed by:
Rebecca D. Burwell, Brown University, USA
Mihail Bota, University of Southern California, USA
Andreas H. Burkhalter, Washington University School of Medicine, USA

****Correspondence:***
Natalie L. M. Cappaert, SILS – Center for NeuroScience, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands.
e-mail: correspondence@temporal-lobe.com

[†]Niels M. van Strien and Natalie L. M. Cappaert shared last authorship.

A connectome is an indispensable tool for brain researchers, since it quickly provides comprehensive knowledge of the brain's anatomical connections. Such knowledge lies at the basis of understanding network functions. Our first comprehensive and interactive account of brain connections comprised the rat hippocampal–parahippocampal network. We have now added all anatomical connections with the retrosplenial cortex (RSC) as well as the intrinsic connections of this region, because of the interesting functional overlap between these brain regions. The RSC is involved in a variety of cognitive tasks including memory, navigation, and prospective thinking, yet the exact role of the RSC and the functional differences between its subdivisions remain elusive. The connectome presented here may help to define this role by providing an unprecedented interactive and searchable overview of all connections within and between the rat RSC, parahippocampal region and hippocampal formation.

Keywords: connectivity, retrosplenial cortex, hippocampal formation, parahippocampal region, interactive, connectome, tract tracing

INTRODUCTION

A connectome is a comprehensive description of the network elements and connections that form the brain (Sporns et al., 2005). Such clear and comprehensive knowledge of anatomical connections lies at the basis of understanding network functions (Crick and Koch, 2003). For example, the existence of a connection between two brain regions ascertains that information transfer can occur. Likewise, when inputs from different brain regions converge onto another region, this can be interpreted as an anatomical substrate for information integration (Sporns and Tononi, 2007). Here, we review the current state of the art knowledge on connectivity of the rat retrosplenial cortex (RSC) with the hippocampal–parahippocampal region (HF–PHR). Apart from this written account, the results will be presented graphically in a comprehensive, interactive, and searchable connectome.

The RSC is the most caudal subdivision of the strip of cortex around the corpus callosum that is generally referred to as the cingulate cortex. In primates, the cingulate cortex is subdivided into an anterior and a posterior part and the most caudoventral subdivision of the posterior cingulate cortex is called RSC, whereas in rodents the RSC comprises the entire posterior cingulate cortex (Vogt and Peters, 1981). Compared to well-investigated brain regions such as the hippocampus, relatively little research has been carried out on the anatomy and functions of the RSC. Recently, discoveries that suggest an important role for the RSC in cognitive functions have sparked increased interest in its anatomy and functions. In humans, the RSC is activated in (autobiographical) memory tasks, navigation, and prospective thinking (Vann et al., 2009). These cognitive functions are also known to be affected in patients with RSC lesions, in whom topographical disorientation and learning deficits are commonly observed (Maguire, 2001). Additional evidence for the involvement of the RSC in cognitive functions stems from the observation that the RSC is a component of the so-called default mode network. The default mode network is an interconnected system of brain regions, involving the lateral and medial parietal areas, the medial frontal and medial temporal lobe region, and the RSC (Buckner et al., 2008; Greicius et al., 2009). This network of brain regions is not only active during retrieval of autobiographical memories, but also when an individual is not focused on the outside world and instead is performing internal tasks such as daydreaming, envisioning the future, retrieving memories, and probing emotions and actions of others (Buckner et al., 2008). Although the function of the default mode network and the role of the RSC within it remains elusive, it has been argued that the RSC plays an active part in memory retrieval and visualization of memories (Greicius et al., 2009).

The cognitive functions in which the RSC is engaged show a striking similarity with those that engage the medial temporal lobe system; a system that comprises the HF–PHR. The loss of cognitive capabilities as seen in patients with RSC lesions is remarkably similar to those seen in patients with HF–PHR damage (Scoville and Milner, 1957; Henderson et al., 1989; Reed and Squire, 1997; Maguire, 2001). Furthermore, in Alzheimer's disease, both HF–PHR as well as RSC show progressive atrophy (Villain et al., 2008; Raji et al., 2009; Pengas et al., 2010).

As in humans, the RSC of rats is thought to be important for a variety of cognitive tasks. RSC lesions impair performance in spatial memory tasks (Sutherland et al., 1988), allocentric working memory tasks (Vann and Aggleton, 2004), egocentric memory tasks (Cooper and Mizumori, 1999; Whishaw et al., 2001), and tasks in which animals have to detect if a spatial arrangement is novel or familiar (Vann and Aggleton, 2002). Particularly within the field of spatial learning and memory, functional attributes of the RSC show a large overlap with those of HF–PHR. The presence of so called head direction cells in the RSC, which have been implicated in navigation, provides strong support to the notion that the RSC has a role in spatial cognition (Chen et al., 1994; Muller et al., 1996; Cho and Sharp, 2001). Head direction cells fire when the animal's head points in a specific direction, and such neurons are also present in a number of subdivisions of PHR, in particular in the presubiculum, parasubiculum, and entorhinal cortex (Boccara et al., 2010). These three areas of PHR are reciprocally connected with RSC and in addition, they receive input from CA1 (Cornu Ammonis; see Nomenclature of HF and PHR) and subiculum (Insausti et al., 1997; Naber et al., 2001; Jones and Witter, 2007), which are both involved in navigation (O' Keefe and Dostrovsky, 1971; Sharp and Green, 1994). The RSC also receives direct input from CA1 and subiculum (Vogt and Miller, 1983; Van Groen and Wyss, 1990b, 1992, 2003; Insausti et al., 1997).

The functional relevance of the RSC and the striking overlap with the functional connotations attributed to HF–PHR strongly suggests a functional relationship between these areas. Knowledge about the connectome that underlies this relationship is relevant, but presently not available in an accessible format. In this review, all reported anatomical connections within the RSC and between the RSC and the HF–PHR in the rat are presented. The general patterns of connectivity will be presented in a condensed written form and specific connection patterns will be highlighted to evaluate possible functional implications. Additionally, all published connections between the RSC and HF–PHR and the intrinsic connectivity of the RSC were integrated in the already published interactive diagram of all published connections of the rat HF–PHR (Van Strien et al., 2009). The current version of the interactive and now searchable diagram, represents without doubt the fullest and most detailed account ever of the brain connections between the HF, PHR, and RSC. We hope that this contribution will help to further understanding of the functional interactions between those brain structures.

MATERIALS AND METHODS

NOMENCLATURE OF THE RSC

Multiple definitions and nomenclatures for the rat cortical mantle exist. Krieg (1946) was the first to delineate the RSC in the rat, based on the anatomical account of Brodmann, who subdivided the RSC in rabbit and named it area (A)29 (Brodmann, 1909). For this review, the nomenclature as described by Vogt et al. (2004) is followed. According to this definition, the rat RSC is subdivided into four areas referred to as A29a, A29b, A29c, and A30. Most of the connectional papers do not separate A29a from A29b and the combined region will be referred to as A29ab in this paper (**Figure 1**). Where necessary, we converted the original nomenclature used in individual papers into the nomenclature of Vogt. For this purpose, a "Rosetta table" was created, which allows easy conversion between different nomenclatures of the RSC (**Table 1**).

DELINEATION OF RSC

The RSC is a neocortical structure situated in the midline of the cerebrum. It arches around the dorsocaudal half of the corpus callosum in the rat, where it is bordered rostrally by the anterior cingulate cortex, caudoventrally by the PHR and laterally by the parietal and visual cortices. The coordinate system that defines position within the RSC is explained in **Figure 1**. The delineation of the subareas of RSC is based on cytoarchitectonic features (**Figure 2**). A29a is the most ventral subdivision and it differs from the dorsally adjacent A29b since it lacks a fully differentiated layered structure. Cytoarchitectonically, A29a has a homogenous layer II/III, while in A29b this layer is divided into a thin superficial densely packed zone and a less dense deeper zone (Vogt and Peters, 1981). A29a and A29b are distinguished from A29c most strikingly in layer III, which in A29ab has cells arranged in bands parallel to the pial surface, while in A29c layer III is thinner and the pyramidal cell bodies are randomly spaced (Van Groen and Wyss, 1990b). An additional way to compare sub-regions is by looking at chemoarchitectonic features. A29ab shows parvalbumin stained cells in layers II, V, and VI, which are not as apparent in A29c (Jones et al., 2005). In AChE stained sections A29c layer IV shows a widening and increased strength of AChE staining compared to layer IV of A29b (Vogt and Peters, 1981; Sripanidkulchai and Wyss, 1987; Van Groen and Wyss, 1990b; Jones et al., 2005). Cytoarchitectonically, A30 shows an abrupt widening and a less dense packing of layer II/III compared to A29b and A29c (Vogt and Peters, 1981; Sripanidkulchai and Wyss, 1987; Van Groen and Wyss, 1992; Jones et al., 2005). Also, A30 layer IV is wider than in A29b/A29c and A30 layer V neuronal cell bodies tend to be larger (Krieg, 1946; Van Groen and Wyss, 1992). In AChE stained sections, layer I–IV of A30 are evenly and darkly stained (Van Groen and Wyss, 1992), whereas in A29c superficial and deep parts of layer I and layer IV are most densely stained (Sripanidkulchai and Wyss, 1987).

NOMENCLATURE OF HF AND PHR

The HF is a C-shaped structure situated bilaterally in the caudal part of the brain. It is subdivided into the dentate gyrus (DG), the Cornu Ammonis (subdivided into CA3, CA2, and CA1), and the subiculum (Sub). The HF consists of three layers, a deep polymorph layer, a more superficial cell layer and on the outside a molecular layer that is almost devoid of neurons. The deep layer is called hilus in the DG and stratum oriens in CA and is not really differentiated in Sub. In DG the cell layer consists of granule cell bodies. In CA and Sub, the cell layer contains pyramidal cells. The superficially positioned molecular layer in DG and Sub is not further subdivided, whereas in CA3, it is divided into three sub-layers: stratum lucidum, stratum radiatum, and stratum lacunosum-moleculare. The lamination of CA2 and CA1 is the same, with the exception that the stratum lucidum is missing.

The PHR borders HF caudally and medially. It is subdivided into the presubiculum (PrS), the parasubiculum (PaS), the entorhinal cortex (EC), further subdivided into the medial and lateral

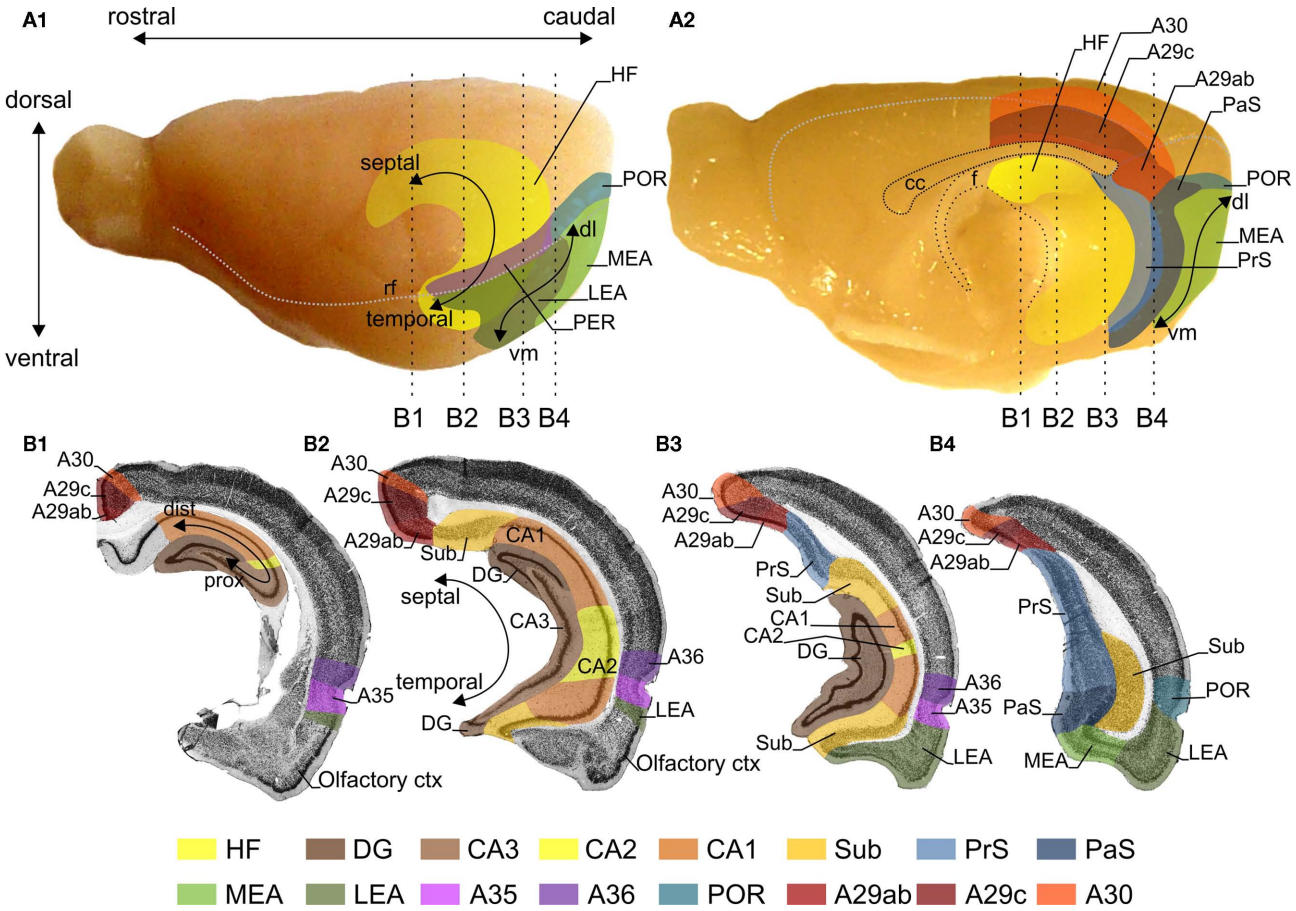


FIGURE 1 | Representations of the retrosplenial cortex (RSC), hippocampal formation (HF) and the parahippocampal region (PHR) in the rat brain. Lateral **(A1)** and midsagittal **(A2)** views of the rat brain. The RSC is subdivided into A29ab, A29c, and A30. For orientation a rostrocaudal and dorsoventral axis is indicated **(A1)**. The HF consists of the dentate gyrus (DG), CA3, CA2, CA1, and the subiculum (Sub). The PHR is subdivided into the presubiculum (PrS), parasubiculum (PaS), the entorhinal cortex, which has a lateral (LEA) and a medial (MEA) subdivision, the perirhinal cortex (PER; consisting of Brodmann areas (A)35 and A36) and the postrhinal cortex (POR). For orientation in HF the long or septotemporal axis is used [also referred to as the dorsoventral axis; **(A1,B2)**]. Another commonly used axis is referred to as the proximodistal axis which indicates a position closer to DG or closer to PHR, respectively **(B1)**. The main axis applied in case of PrS and PaS is also the septotemporal axis, in PER and POR a rostrocaudal axis and in the entorhinal cortex is a dorsolateral-to-ventromedial axis [dl and vm; **(A1,A2)**]. The dashed vertical lines in panels **(A1,A2)** indicate the levels of four coronal sections **(B1–B4)**, which are shown in **(B)**. The gray stippled line in **(A1)** represents the position of the rhinal fissure (rf) and in **(A2)** they represent the global delineation from the dorsal surface of the brain with the midsagittal and occipital surfaces. **(B)** Four coronal sections of the rat brain; the levels of these sections are indicated in **(A)**. The subfields of the HF, PHR and RSC are color-coded, see color panel below **(B)**.

entorhinal area (MEA and LEA respectively), the perirhinal cortex (PER; divided into Brodmann's areas 35 and 36) and the postrhinal cortex (POR). The PHR is generally described as having six layers. The delineation and the HF–PHR connections are extensively described in earlier publications (Witter and Amaral, 2004; Van Strien et al., 2009). The coordinate system that defines position within the HF and the PHR is explained in **Figure 1**.

DATA COLLECTION AND VISUALIZATION

A search was performed on publications reporting tract-tracing studies on intrinsic RSC and RSC – HF–PHR connections in PubMed[1] and Embase[2] (see www.temporal-lobe.com for queries). The following inclusion criteria were used: (1) tract-tracing studies or studies which report intracellular filling of single cells; (2) studies which used healthy, genetically un-altered, untreated adult rats were included; (3) publications written or translated into English or in a language using roman print. The database queries retrieved 816 papers of which 46 contained relevant information. The connectional information was retrieved from these papers, including information from tables and figures, using the following criteria: (1) it was clear where anterogradely filled terminals or retrogradely labeled cell bodies were located; (2) the location of the injection site was clearly described; (3) injection sites did not include multiple brain areas or fiber bundles; (4) lesion studies were discarded; (5) explicitly reported non-excitatory projections were excluded; (6) contralateral projections were excluded. The information about these connections was stored in a custom-made relational database

[1]www.PubMed.gov

[2]www.embase.com

Table 1 | Comparison of nomenclatures of the retrosplenial cortex (RSC).

Vogt et al. (2004)	Brodmann (1909)†	Rose (1927)‡	Krieg (1946)	Rose and Woolsey (1948)†	
A30	A29d	RSag (rostral and intermediate A30)	29c	Area cingularis	
A29c	A29c	RSgβ (caudal A30 and A29c)	29b	Area cingularis	
A29ab		RSgα		Retrosplenial area	
A29b	A29b				
A29a	A29a				
Vogt et al. (2004)	**Meibach and Siegel (1977)**	**Krettek and Price (1977)§**	**Sripanidkulchai and Wyss (1987)**	**Van Groen and Wyss (1992)**	**Shibata (1994)**
A30	RSAG	RsAg	Rag	Rdg	RSA
A29c	RSGd	RsG	Rgb	Rgb	RSG
A29ab	RSGv	RsG	Rga	Rga	RSG
A29b					
A29a					

Vogt et al. (2004)	Zilles and Wree (1995)	Burwell and Amaral (1998)	Jones et al. (2005)	Shibata et al. (2009)
A30	RSA	RSPd	Rostral and intermediate Rsd	29d
A29c	RSG	RSPv	RSv-b	29c
A29ab	RSG		RSv-a and caudal RSd	
A29b				29b
A29a				29a

In the present paper, RSC is subdivided into A29a, A29b, A29c, and A30 according to Vogt et al. (2004).
†Rabbit.
‡Mouse.
§Nomenclature described in Zilles and Wree (1995).
Rose (1927): RSag, retrosplenialis agranaluris; RSgβ, retrosplenialis granularis dorsalis; RSgα, retrosplenialis granularis ventralis; Meibach and Siegel (1977): RSAG, retrosplenialis agranularis; RSGd, retrosplenialis granularis dorsalis; RSGv, retrosplenialis granularis ventralis; Krettek and Price (1977): RsAg, agranular retrosplenial area; RsG, granular retrosplenial area; Sripanidkulchai and Wyss (1987): Rag, retrosplenial agranular cortex; Rgb, retrosplenial granular cortex b; Rga, retrosplenial granular cortex a; Van Groen and Wyss (1992):Rdg, retrosplenial dysgranular cortex; Rgb, retrosplenial granular b cortex; Rga, retrosplenial granular a cortex; Shibata (1994): RSA, retrosplenial agranular area; RSG, retrosplenial granular area; Zilles and Wree (1995):RSA, agranular retrosplenial cortex; RSG, granular retrosplenial cortex; Burwell and Amaral (1998): RSPd, dorsal retrosplenial area; RSPv, ventral retrosplenial area; Jones et al. (2005): RSd, dorsal retrosplenial cortex; RSv-b, ventral retrosplenial cortex dorsal part; RSv-a, ventral retrosplenial cortex ventral part.

(Microsoft Access; Microsoft Corporation, WA, USA). Before data was entered into this database, the accuracy was verified by at least two of the authors.

Next, results from independent retrograde and anterograde experiments were combined, such that both the layers of origin and termination could be determined. The connections were added to the existing HF–PHR connectome (Van Strien et al., 2009) which was drawn in Visio (Microsoft Corporation, WA, USA) and exported to PDF (Adobe Acrobat Pro; Adobe Systems Inc., CA, USA), see **Figure 3** for an overview of the connectome. In total, the database now includes 223 references in 710 records, describing approximately 2600 connections. Compared to the initial interactive connectome, the usability of the current version is improved by adding a search dialog, help section and a tutorial, developed using AcroDialogs and AcroButtons (Windjack Solutions, Inc., OR, USA). Updated and extended support information (e.g., manual, references, and RSC nomenclature) is available on http://www.temporal-lobe.com.

RESULTS

In the following section, the intrinsic connections between RSC subdivisions and connections between RSC subdivisions and the HF and PHR are summarized in a condensed written form (for a condensed overview see **Table 2**). These connections are also visually presented in the interactive connectome (Supplementary Material).

CONNECTIONS WITHIN RSC

A29a and A29b

Intrinsic projections are those confined to a defined cytoarchitectonic subarea. In case of A29a and A29b, reports either describe those separately or the two areas have been combined into one A29ab. We therefore deal with the two areas in this section together. The intrinsic projections of A29a originate in layer III–VI and terminate in layers I, II, III, and V. Those of A29b follow a rostral-to-rostral and a caudal-to-caudal pattern. Rostral projections originate in layers II, III, V, and VI and terminate rostrally in all layers. Caudal

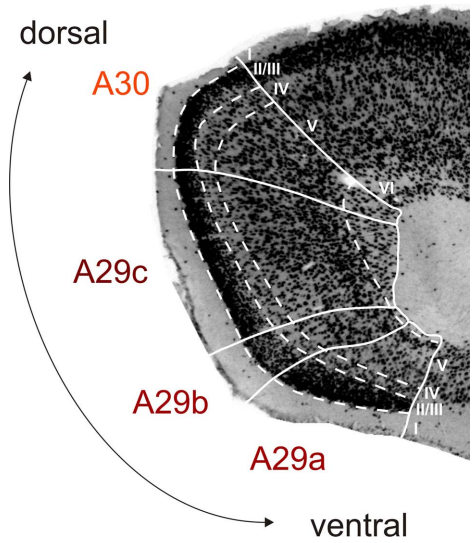


FIGURE 2 | Cytoarchitecture of rat retrosplenial cortex (RSC). Photomicrograph of a coronal section stained for NeuN (high power image taken from the section shown in **Figure 1B2**), illustrating the cytoarchitectonic characteristics of A29a, A29b, A29c, and A30. A29a has a homogenous layer II/III and lacks fully differentiated deep layers. In A29b layer II/III is divided into a thin superficial densely packed zone and a less dense deeper zone. A29c has a more differentiated layer V and a more equally dense layer II/III compared to A29b and a thinner layer IV compared to A30. A30 shows a widening and a less dense packing of layer II/III and layer V neuronal cell bodies tend to be larger. Layer VI is mostly developed in A30 and A29c and almost absent in A29a and b (for more details see Vogt et al., 2004; Jones et al., 2005).

projections terminate caudally in layers I, II, III, and V (Shibata et al., 2009). The A29a projection to rostral and caudal A29b originates in all cell layers and terminates in layers I, II, III, and V (Shibata et al., 2009). Caudal A29b projects to caudal A29a and rostral A29b projects to rostral A29a and the projections originate in all cell layers and terminate in layers I, II, III, and V (Shibata et al., 2009). The terminal patterns, when combined, are essentially in line with the reported terminal distribution in layers I, III, V, and VI for the combined area A29ab (Van Groen and Wyss, 1990b; Jones et al., 2005; Miyashita and Rockland, 2007).

Both A29a and A29b project to the entire rostrocaudal extent of A29c (Vogt and Miller, 1983; Shibata et al., 2009). Projections that arise from layer V of caudal A29a terminate in layers I, II, III, and VI of the intermediate rostrocaudal and caudal parts of A29c. Rostral A29b projects to rostral and intermediate rostrocaudal A29c, and caudal A29b projects to caudal A29c. Projections originate in layers II, III, V, and VI and terminate in all layers of A29c. When described together (Van Groen and Wyss, 1990b, 2003; Jones et al., 2005; Miyashita and Rockland, 2007), the only striking deviation from the combined separate patterns is that the terminal distribution of the projection from A29ab to rostral A29c is restricted to layers II and III (Van Groen and Wyss, 1990b).

Neurons in layers V and VI of the caudal part of A29a project to caudal levels of A30, terminating in layers I, II, III, and V (Shibata et al., 2009). In case of A29b these projections arise from the entire rostrocaudal extent, but also target caudal portions of A30, showing the same laminar distribution as those of A29a (Vogt and Miller, 1983; Shibata et al., 2009). In line with these observations, the projections from A29ab originate caudally in layer VI and more rostrally in layers III–V projecting to the midrostrocaudal portion of A30 (Van Groen and Wyss, 1990b, 1992; Jones et al., 2005).

A29c

The intrinsic connections of A29c arise from the entire rostrocaudal extent and project to the entire rostrocaudal extent (Vogt and Miller, 1983; Van Groen and Wyss, 2003; Jones et al., 2005; Miyashita and Rockland, 2007; Shibata et al., 2009). These projections originate in layers II, III, V, and VI and terminate in layers I–V. Projections from A29c to caudal A29a exist (Shibata et al., 2009) and the caudal and midrostrocaudal origin is in layer V and that of the rostrally originating projections in layers V and VI. The projections from A29c to A29ab or A29b arise from the entire rostrocaudal extent and project to the entire rostrocaudal extent (Van Groen and Wyss, 1990b, 2003; Jones et al., 2005; Miyashita and Rockland, 2007; Shibata et al., 2009). Rostrally, projections arise from layers V and VI, whereas in caudal A29c projections arise from layers II, III, V, and VI. Termination of these projections in A29b occurs in layers I, II, III, and V. The projection from A29c to A30 originates and terminates in all layers (Vogt and Miller, 1983; Sripanidkulchai and Wyss, 1987; Audinat et al., 1988; Van Groen and Wyss, 1992, 2003; Jones et al., 2005; Shibata et al., 2009). A topographical organization is present, such that rostral A29c projects to rostral and mid-rostrocaudal A30 and caudal A29c projects to the entire rostrocaudal extent of A30.

A30

The intrinsic connections of A30 arise from the entire rostrocaudal extent and project to the entire rostrocaudal extent, whereby layers II–VI project to all layers (Vogt and Miller, 1983; Audinat et al., 1988; Van Groen and Wyss, 1992; Jones et al., 2005; Shibata et al., 2009). Projections to A29a arise from the entire rostrocaudal extent of A30 and terminate caudally (Vogt and Miller, 1983; Shibata et al., 2009). At rostral and intermediate rostrocaudal levels, the projections originate in layer V, whereas caudally the projections arise from layer II–VI. A30 also projects to A29b (Vogt and Miller, 1983; Shibata et al., 2009). This projection arises rostrally from layers V and VI and caudally from layers II–VI. Termination in A29b occurs in layers I, II, III, and V. The caudal and midrostrocaudal projections of A30 to A29ab terminate in all layers (Van Groen and Wyss, 1992; Jones et al., 2005). A30 projects to A29c and the projection does not show a topographical rostrocaudal organization. Projections arise from layers II, III, V, and VI and terminate in all layers of A29c (Vogt and Miller, 1983; Van Groen and Wyss, 1992, 2003; Jones et al., 2005; Shibata et al., 2009).

RSC PROJECTIONS TO THE PHR

A29a and A29b

A29ab projects to all subdivisions of PHR. Dense projections exist from A29ab to PrS (Van Groen and Wyss, 1990a,b,d; Shibata, 1994; Jones and Witter, 2007). More specifically, layers II and V of A29ab target PrS layers I, III, V, and VI. The A29ab projection to PrS is topographically organized, such that caudal A29ab targets the entire septotemporal extent of PrS, whereas rostral A29ab targets septal PrS only. The projection to PaS originates in layers II, III,

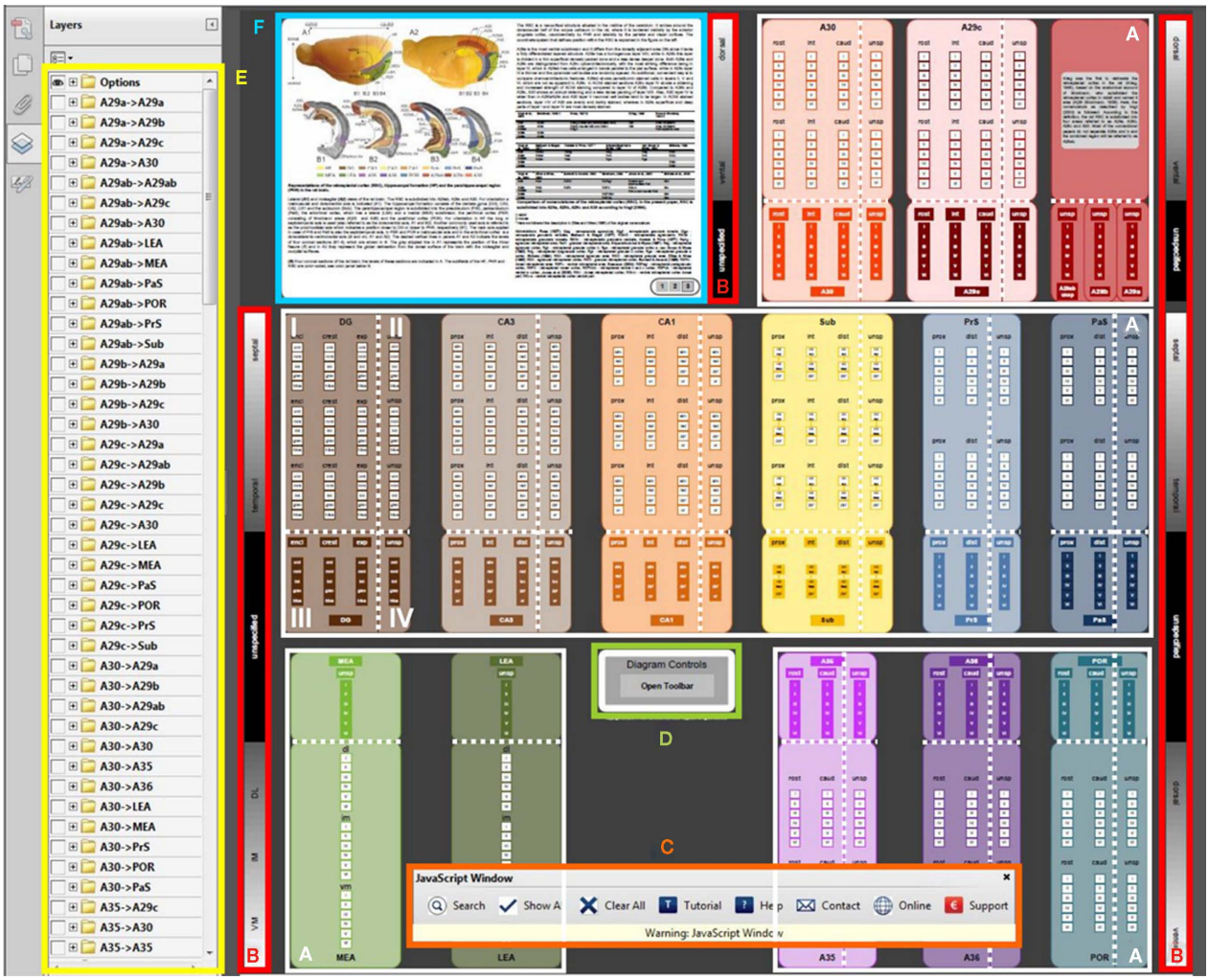
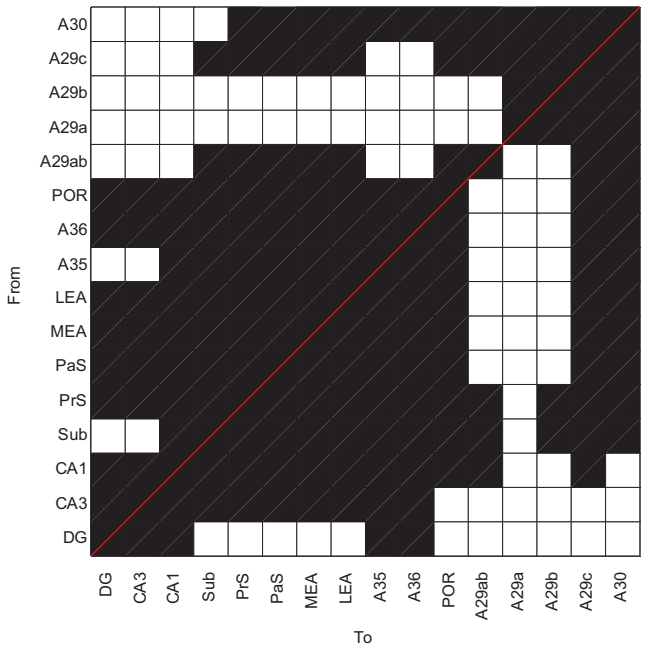


FIGURE 3 | Retrosplenial and hippocampal–parahippocampal connectome. The connectome (see ratbrain_connectome.pdf in Supplementary Material) consists of 14 large, color-coded boxes, which represent the sub-regions of the hippocampal formation, parahippocampal region, and retrosplenial cortex. In this figure, the user interface elements of the connectome are indicated with color-coded outlines and their meaning/purpose is explained. **(A)** In the white outlines, the 14 anatomical sub-regions are displayed. **(B)** The 14 anatomical sub-regions are three-dimensionally organized. However, the origin and termination of connections are not always described in full detail in the literature. Therefore, area boxes are divided into four quadrants. Quadrant I has full topological information, whereas the other quadrants have less topological detail. In *quadrant I*, the vertical axis in the connectome represents the septotemporal axis of dentate gyrus (DG), Cornu Ammonis (CA3 and CA1), subiculum (Sub), presubiculum (PrS) and parasubiculum (PaS), and the dorsoventral axis of medial and lateral entorhinal area (MEA and LEA), retrosplenial cortex (A29 and A30), perirhinal cortex (A35 and A36) and postrhinal cortex (POR). The sidebars **(B)** display the dorsoventral and septotemporal axes of the anatomical sub-regions. The horizontal axis within quadrant I and III represents the proximodistal axis in CA3, CA1, Sub, PrS, and PaS; the rostrocaudal axis in A29c, A30, A35, A36, and POR and the DG is subdivided into the inner/outer blades and crest region. Within the area boxes, the layers for each specific subarea are outlined. In *quadrant II*, the information of the vertical axis and the layers are specified, but no details of the horizontal axis are presented. In *quadrant III* the horizontal axis and the layers are represented, while in *quadrant IV*, only layer information is present. **(C)** The interactive connectome allows visualization of detailed connectivity patterns within and between sub-regions. To search for connections, use the search button on the toolbar **(C)**. The toolbar has eight buttons (from left to right): Search for connections, show all connections, clear all connections, a short tutorial on how to search for connections, a help section, contact information, a link to the project website www.temporal-lobe.com and information about how to support this project. **(D)** If the toolbar is closed, clicking the "open toolbar" button will restore it. **(E)** After a search is carried out, the retrieved connections will be drawn in the diagram between the appropriate areas and quadrants. Additionally, an eye-icon will appear in the layers panel on the left **(E)**. This is an alphabetically sorted list of "from → to" connection groups that can also be switched on or off manually. In front of each group is a "+" icon. Clicking this icon expands the list of individual connections that make up the group, allowing one to select connections originating from a specific cortical layer or according to a specific three-dimensional projection pattern. **(F)** The figure panel provides a detailed anatomical description of the retrosplenial cortex, the hippocampal formation and the parahipppcampal region, together with translation tables for nomenclature. Use the buttons in this panel to switch between the figures.

and V and the caudal A29ab projects to the whole septotemporal axis of the PaS (Van Groen and Wyss, 1990a,b; Shibata, 1994; Jones and Witter, 2007). The projections to PER originate in caudal and intermediate rostrocaudal A29ab and terminate in layers V and VI (Shibata, 1994; Jones and Witter, 2007). The POR projection originates in layer II and V and terminates in all layers (Jones

Table 2 | Anatomical connections between subareas of the hippocampal formation, the parahippocampal region and the retrosplenial cortex.



Connections that are present are shown as black squares, absent connections are shown as white squares. DG, dentate gyrus; CA, Cornu Ammonis 1 and 3; Sub, subiculum; PrS, presubiculum; PaS, parasubiculum; MEA, medial entorhinal area; LEA, lateral entorhinal area; A35: perirhinal cortex – Brodmann area 35; A36: perirhinal cortex – Brodmann area 36; POR postrhinal cortex. Brodmann area A29ab, A29a, A29b, A29c, and A30.

and Witter, 2007). The projection to LEA layers I, II, III, V, and VI and MEA layers I, V, and VI originates in caudal A29ab (Jones and Witter, 2007).

A29c

A29c also projects to all subdivisions of PHR. The projection to PrS is topographically organized along the rostrocaudal axis, such that caudal A29c projects to the entire septotemporal extent of PrS, intermediate A29c projects to intermediate and septal PrS, whereas rostral A29c only projects to septal PrS (Meibach and Siegel, 1977; Vogt and Miller, 1983; Van Groen and Wyss, 1990a,d, 2003; Shibata, 1994; Gonzalo-Ruiz and Bayona, 2001; Jones and Witter, 2007). Projections to temporal PrS originate in layers II and V, whereas projections that terminate in septal PrS are reported to originate only in layer V. Termination has been reported in all layers of PrS. The projection to PaS originates in layer V and terminates in all layers (Vogt and Miller, 1983; Finch et al., 1984; Van Groen and Wyss, 1990a; Shibata, 1994; Jones and Witter, 2007). Whether or not this projection shows a topographical organization is unclear. The projection to PER that arises from midrostrocaudal A29c terminates in layer V and VI of PER, whereas the caudal A29c projection terminates in all layers of PER (Guldin and Markowitsch, 1983; Shibata, 1994; Jones and Witter, 2007). The projection to POR originates in layers II and V of A29c and terminates in all layers (Burwell and Amaral, 1998; Jones and Witter, 2007). Both MEA and LEA are targeted by projections that originate in caudal A29c. Termination in EC has been reported to include both superficial layers and deep layers (Shibata, 1994; Jones and Witter, 2007).

A30

Similar to A29ab and A29c, A30 also projects to all subdivisions of the PHR. The projections to PrS originate in layers II and V and terminate in layers I, II, III, V, and VI (Vogt and Miller, 1983; Van Groen and Wyss, 1990a, 1992; Shibata, 1994; Jones and Witter, 2007). The topographical organization of this projection is such that rostral A30 projects to septal PrS, whereas caudal and midrostrocaudal A30 project to the entire septotemporal extent of PrS. The projection to PaS originates along the entire rostrocaudal extent of A30 (Vogt and Miller, 1983; Van Groen and Wyss, 1992; Shibata, 1994; Jones and Witter, 2007). The layers from which these projections originate have not been determined, and the projections terminate in all layers of PaS. The projections to PER originate from the rostrocaudal extent of A30. Rostral A30 projections to PER show terminal labeling in layers V and VI. Caudal projections originate in layers II, III, and V and terminate in all layers of PER (Deacon et al., 1983; Shibata, 1994; Burwell and Amaral, 1998; Jones and Witter, 2007). A30 projections to POR originate in layers II and V and terminate in all layers of POR. No topographical organization has been reported (Burwell and Amaral, 1998; Jones and Witter, 2007). A30 to EC projections show the following pattern: Layer V of A30 projects to all layers of MEA. However, caudal A30 was shown to project to LEA and MEA layers V and VI, whereas intermediate rostrocaudal A30 was shown to project to MEA layers I, II, and III (Van Groen and Wyss, 1992; Shibata, 1994; Burwell and Amaral, 1998; Jones and Witter, 2007).

PHR PROJECTIONS TO THE RSC

A29a and A29b

Some PHR sub-regions send return projections to A29ab. Septal and temporal PrS layer V neurons project to A29ab layers I–V (Finch et al., 1984; Van Groen and Wyss, 1990a,b,d). A projection from EC to A29ab has been reported, but topographical information for this projection is absent (Miyashita and Rockland, 2007). Although PaS, PER, and POR receive connections from A29ab, return projections have not been described.

A29c

Neurons in PrS layer V project to layers I and III of A29c. Septal PrS projections terminate in the whole rostrocaudal extent of A29c (Vogt and Miller, 1983; Finch et al., 1984; Van Groen and Wyss, 1990a,d, 2003). Neurons in PaS layer V project to layer II and III of A29c (Vogt and Miller, 1983; Finch et al., 1984). Similarly, projections from PER (A35, A36) and POR to A29c have been reported, but no additional information about these projections is available (Agster and Burwell, 2009). Finally, layer V of both LEA (Agster and Burwell, 2009) and MEA (Frohlich and Ott, 1980; Insausti et al., 1997; Agster and Burwell, 2009) project to A29c. MEA is known to project to layers I, II, and IV of A29c (Frohlich and Ott, 1980; Insausti et al., 1997).

A30

Neurons in layers V and VI of septal PrS project to layers I, III, IV, and V of A30 (Vogt and Miller, 1983; Finch et al., 1984; Witter et al., 1990; Van Groen and Wyss, 1992). PaS layer V

projects to A30, but the layers of termination are unknown (Vogt and Miller, 1983). A35 and A36 also project to A30 (Agster and Burwell, 2009), where caudal A36 projects to caudal A30. POR projects to A30, but layer specific information is absent (Agster and Burwell, 2009). Finally, both LEA (Agster and Burwell, 2009) and MEA (Insausti et al., 1997; Agster and Burwell, 2009) have been reported to project to A30, but only for MEA a termination in layers I and II of A30 has been described (Insausti et al., 1997).

RSC PROJECTION TO THE HF

A29a and A29b

Neurons in layer V of A29ab project to Sub in the HF (Van Groen and Wyss, 1990b; Shibata, 1994). This projection is topographically organized such that caudal A29ab projects to temporal Sub, where termination occurs in the stratum moleculare and stratum pyramidale, whereas rostral and intermediate A29ab projects to intermediate septotemporal levels of Sub, where termination occurs in the stratum pyramidale.

A29c

Also for A29c, Sub is the only HF target (Meibach and Siegel, 1977; Shibata, 1994; Gonzalo-Ruiz and Bayona, 2001; Van Groen and Wyss, 2003). This projection is topographically organized along the rostrocaudal axis, such that rostral A29c projects to septal and intermediate septotemporal levels of Sub, intermediate A29c projects to intermediate septotemporal levels and caudal A29c projects to temporal and intermediate septotemporal levels of Sub. Caudal projections to temporal Sub have been described to terminate in stratum moleculare and stratum pyramidale, whereas projections from rostral A29c to intermediate septotemporal Sub terminate in stratum pyramidale.

HF PROJECTIONS TO THE RSC

A29a and A29b

The HF projections to A29ab originate in Sub and CA1. Sub projections to A29ab terminate in layers I, II, and III and mimic the topographical organization of the retrosplenial projections to Sub. The projection from CA1 to A29ab originates from the septal portion (Van Groen and Wyss, 1990b; Naber and Witter, 1998; Miyashita and Rockland, 2007).

A29c

The HF projections to A29c also originate in Sub and CA1. The proximal part of the septal Sub projects to rostral A29c. Intermediate septotemporal and distal Sub project to caudal A29c (Vogt and Miller, 1983; Finch et al., 1984; Naber and Witter, 1998; Van Groen and Wyss, 2003; Miyashita and Rockland, 2007). These projections typically terminate in layers II and III. No information about temporal Sub to A29c projections has been published. The distal portion of septal CA1 also projects to A29c. The projection originates in neurons in stratum pyramidale and at the border between stratum moleculare and stratum radiatum and terminates in layers II, III, and IV (Meibach and Siegel, 1977; Van Groen and Wyss, 1990c, 2003; Naber and Witter, 1998; Miyashita and Rockland, 2007).

A30

The HF projections to A30 only originate in Sub. The distal portion of the septal Sub, as well as the intermediate proximodistal portion of intermediate septotemporal Sub project to layers I and II of A30 (Vogt and Miller, 1983; Finch et al., 1984; Kohler, 1985; Witter et al., 1990).

SUMMARY OF INTRINSIC RSC CONNECTIONS

There are strong intrinsic connections within the RSC subdivisions. All rostrocaudal levels within both A29c and A30 issue projections to their respective rostrocaudal extents. A29b projections have a strict topography from rostral-to-rostral and from caudal-to-caudal; A29a only has a caudal-to-caudal projection. There are also strong reciprocal connections between the RSC subdivisions. All rostrocaudal levels of one subdivision project to all rostrocaudal levels of all other subdivisions, but there are some exceptions: (1) caudal A29a projects only to caudal A29b, A29c, and A30; (2) caudal A29b does not project to rostral A29c; (3) rostral A29c does not project to caudal A30; (4) rostral and midrostrocaudal A30 only projects to caudal A29b and the return projection from A29b only terminates in caudal A30.

SUMMARY OF RSC – HF/PHR CONNECTIONS

The RSC projects to all PHR subdivisions and Sub. Only the projections of RSC to PrS and Sub show a topographical organization such that the rostrocaudal axis of origin in RSC correlates to a septotemporal terminal distribution in PrS and Sub. The projections to PrS are among the densest of RSC–PHR connections (Jones and Witter, 2007) and this is particularly true for projections from caudal RSC (Shibata, 1994). A topographical pattern for reciprocal connections has not been identified. Areas 30 and 29c receive input from the whole PHR and Sub, while PrS and Sub are the only areas which project to A29b and A29ab. Dorsal CA1 projects only to A29ab and A29c.

DISCUSSION

The connectome of the rat brain should describe all network elements and connections in a clear and comprehensible way. Compared to the comprehensiveness that a connectome implies, current knowledge is in its infancy. When considering the vast number of neurons in the rat nervous system and their connections, together with the currently available technologies to collect and handle information about them, creating a connectome is an expensive, time consuming, and complicated task. Scientists will eventually have a comprehensive map of the rat brain available, and just like the usefulness of an easily accessible map of all the roads in the world, such a connectome will be an indispensible foundation for basic and applied neurobiological research (Sporns et al., 2005). Therefore, regardless of the complexity and duration of the undertaking to create a connectome, the challenge to create a connectome of the mouse, rat, monkey, and human brain has been readily accepted by scientists [e.g., Brain Architecture Management System (BAMS), CoCoMac, The Rodent Brain Workbench, BrainNavigator, Blue Brain Project, The Allen brain atlas, Neuroscience Information Network (NIF). For a complete overview of projects see the International Coordinating Facility (INCF)].

With this publication, we present the current version of our partial rat brain connectome, which to the best of our knowledge represents all current information on the ipsilateral pathways within and between the HF, PHR, and RSC. A different approach from that of a traditional meta-analysis was taken, to create this connectome. In a traditional meta-analysis, typically only a subset of data is selected, summarized, and organized according to the author's views. The resulting reduction in detail of anatomical networks is useful for creating scientific hypotheses, but contradicts with a fundamental characteristic of a connectome: to be an exhaustive knowledge resource. Therefore, we chose the approach to present the anatomical data of the selected regions in the fullest available detail, which allows scientists to prune this information themselves to match their hypotheses, or design competing anatomical hypotheses.

We realize that the current state of knowledge is not exhaustive and hence one could argue that the connectome presented here is not a real connectome. Nevertheless, the connectome presented here provides the best approximation of a full connectome at the current point in time. With future publications we aim to continually update and expand the database. Still, users of connectomes should always keep a perspective on where the current state of knowledge stands compared to having absolute knowledge. For this reason, this discussion will first touch upon some of the challenges of anatomical connectomes that remain to be resolved, after which the potentials of connectomes will be exemplified using the information presented in this review.

CHALLENGE 1: BORDERS IN THE BRAIN

Combining data produced by many researchers, over 100 years, using many different techniques in a great number of tract-tracing experiments, leads to a number of challenges on demarcation of brain areas and designation of names in research reports. Such nomenclature issues exist not only for the RSC, but for almost all brain regions. These issues have arisen because different histological techniques produce different definitions of borders in the brain, or simply because researchers disagree on the demarcation.

Krieg (1946) was the first to delineate the RSC in the rat. Based on the nomenclature of Brodmann (1909) who subdivided the RSC in rabbit and dubbed it A29, Krieg divided it into a ventral subdivision (A29b) and a dorsal subdivision (A29c) in the rat (**Table 1**). In the 1970s of the last century, Krettek and Price (1977) divided the RSC in a granular and an agranular subdivision, the RsG and the RsAg. The granular part was further differentiated into a dorsal and a ventral part by Meibach and Siegel (1977). Vogt and Peters (1981) divided the granular subdivision into three areas, A29a, b, and c based on termination patterns of callosal fibers. In this century, Jones et al. (2005), Jones and Witter (2007) delineated the RSC borders based on parvalbumin stainings and classified the most caudal part of A29ab with A30 into the dorsal retrosplenial cortex (RSd). Although no standardization of nomenclatures has become apparent, dealing flexibly with the diverse nomenclatures may allow to efficiently generate inspiring insights into the organization of the RSC. Therefore, in the current HF–PHR–RSC connectome, we selected one nomenclature to express ourselves. For RSC we used the nomenclature of Vogt et al. (2004) and for HF–PHR the nomenclature of Witter and Amaral (2004). We provided what we call a "Rosetta table" of nomenclatures. The "Rosetta table" of nomenclatures is a necessary translation tool that lists all available nomenclatures (for a given structure in a given species) and facilitates translating between different nomenclatures. For such translations, different methods have been developed, and we applied careful comparison of all cyto- and chemoarchitectural features described in the original publications (Bota and Swanson, 2010).

CHALLENGE 2: LEVEL OF DETAIL

Aspecific reporting

Even when using technically advanced methods to clarify projection patterns, the usefulness of a scientific report on anatomical connections depends to a great extent on how detailed the authors report their results and the way they present the data in figures to support their observations. When reviewing the projections from RSC to HF and PHR, only in a few studies information was provided on the layers of origin or termination, or specific projection patterns. For example, projections from A29ab, A29c, and A30 to LEA and MEA exist (Van Groen and Wyss, 1992; Burwell and Amaral, 1998; Jones and Witter, 2007) and although LEA and MEA show an overall dorsolateral-to-ventromedial connectional and functional gradient, none of these reports provide comparable termination information. Even less specific accounts inform us that RSC projects to the EC, without indicating if the projections terminated in the lateral or medial subdivision (Audinat et al., 1988; Van Groen and Wyss, 1990b; Shibata, 1994). This is regrettable, since more detailed information about the origin and termination could be related to the function of a connection. It is known that functional differences between the LEA and MEA exist. Neurons in the MEA exhibit spatial selectivity, while LEA cells display only weak spatial modulation (Fyhn et al., 2004; Hargreaves et al., 2005). Since projections from RSC to EC mainly target MEA, it could be hypothesized that this input is relevant to spatial firing properties of MEA neurons. Unfortunately, the current level of detail is such that almost nothing is known about the topography and lamination of the HF and PHR projections to the RSC and vice versa. In our connectome, this information is displayed as connecting areas with an unspecified origin or termination. This lack of detailed information makes it difficult to predict the function of projections or small networks. Fortunately, more advanced methods to explore brain networks are continuously being developed, making it likely that future expansions of the connectome can include for example, information about the nature of postsynaptic targets or the distribution of synapses on the dendritic tree.

Displaying layer specificity

To obtain information about layer specificity of projections, ideally anterograde and retrograde tract-tracing experiments are combined. For example, an anterograde tracer injection is placed in midrostrocaudal A30, thereby discovering regions of termination including layers I, II, III, and V of rostral A29c. This experiment should be completed by placing a retrograde tracer in rostral A29c, to reveal the layers of origin of this projection in midrostrocaudal A30, i.e., layer II, III, V, and VI (Shibata et al., 2009). A strength

of our current approach is that a database of connections allows easy combination of anterograde and retrograde experiments, even across publications. But when representing this information in a connectome, should one draw projections from all origin layers to all target layers, or could it be that a more specific pattern exists, e.g., that layer II cells only project to layer V of the target region? Therefore, only when single cells are fully traced (e.g., Honda et al., 2011), one can accurately describe layer specificity of projections.

Displaying projection topographies

Apart from layer specificity, projections may be topographically organized. For example, the rostral RSC projects weaker to EC than caudal RSC. Often, such topographies show a gradient in projection strength, but one cannot rigorously claim that only caudal RSC projects to EC. In the current version of our connectome, such topographies are not visible since no information on relative density is included. We chose this approach, in view of the risk that by emphasizing some brain connections over others, the less emphasized ones may be erased from the scientific working memory. However, users should not forget that brain connections typically have different strengths and may show topographical gradients that are not apparent in the interactive connectome.

CHALLENGE 3: FUNCTIONAL CONNECTIVITY

When using our connectome, it is important to keep in mind that two connections symbolized by two similar looking lines may be different from one another for a number of reasons.

Excitatory – inhibitory, modulation

Our connectome is based on tract-tracing data and thus comes with certain limitations related to this technique. One important limitation is that tract-tracing does not reveal if a connection is excitatory or inhibitory, whereas this information is of functional relevance. By combining immunohistochemistry with either confocal or electron microscopy, it can be established if a projection is excitatory or inhibitory (Van Haeften et al., 1997). Alternatively, electrophysiological data can help determine this functional property and such information will eventually be incorporated into the connectome.

In our current connectome, known GABA-ergic connections are not specifically included, although some projections are likely GABA-ergic. There is a dense GABA-ergic projection from CA1 to the RSC, starting from all layers except the stratum lacunosum-moleculare (Miyashita and Rockland, 2007). Non-pyramidal cells, which could be GABA-ergic, of the septal Sub are projecting to layer I of the RSC as well (Van Groen and Wyss, 1992, 2003; Miyashita and Rockland, 2007). Hippocampal GABA-ergic neurons also project intra-hippocampally (Jinno, 2009). Inhibitory GABA-ergic projections are involved in regulation of neural oscillations (Somogyi and Klausberger, 2005). Theta oscillations (4–12 Hz) for instance, have been postulated to support memory formation in the HF and the cortex (Klausberger et al., 2003). Theta oscillations have also been recorded in the RSC (Borst et al., 1987; Talk et al., 2004) and are in coherence with oscillations in CA1 (Young and Mcnaughton, 2009). Although this coherence may be a product of volume conduction from the hippocampus, the GABA-ergic projections from CA1 and Sub to RSC could play a role as well.

Weak – strong connections

Most anatomical reports contain subjective descriptions of the strengths of projections. Such subjective reports are impossible to quantify and therefore such information cannot be incorporated in the connectome. Yet, in reality differences in strengths exist. For instance, the projection from A29ab to MEA is reported to terminate in both superficial and deep layers. When assessing this connection in detail, clear differences between superficial and deep layers exist. There is dense terminal labeling in layer V of MEA, whereas terminal labeling in layer III is very light (Jones and Witter, 2007). Therefore, it would be reasonable to suggest that the major effect of A29ab on MEA is mediated through layer V, not through layer III. Moreover a strong projection could arise from a small number of neurons that supply a highly collateralized terminal plexus, but a similarly strong projection may also result from many cells sending less collateralized axons to the target area. This type of information is almost never available in the literature. Yet a connectome ideally would describe this information.

CHALLENGE 4: NOT REPORTED OR NO CONNECTION?

Another challenge is how to interpret "negative information." There are two types of negative information: (1) there is data that indicate that a connection between area A and area B is very unlikely, or (2) a connection between area A and area B was not reported, but may exist. For example, the current literature assessment shows that A29c and A29ab both project to Sub, whereas data on a projection from A30 to Sub are lacking. Should one now conclude that a projection from A30 to Sub is not present in the rat brain, or could it be that this projection is present, but perhaps remained undetected or was detected but not reported? This issue has been dealt with in other databases, where connections that are explicitly reported to not exist have been collated (Stephan et al., 2000; Bota and Arbib, 2004). In a future version of our database we plan to implement a similar solution. Currently, some caution is required when drawing conclusions on the basis of the connectome alone.

CHALLENGE 5: INCLUSION OF INCORRECT DATA

Although it is assumed that researchers aim to report as accurately as possible the results of their tract-tracing experiments, it does not exclude the possibility that brain connections are reported that do not exist. Tract-tracing techniques were much refined over time, such that injections of modern tracers such as Phaseolus Vulgaris-leucoagglutinin or biotinylated dextran-amine can now be injected region or cell layer selective (Gerfen and Sawchenko, 1984; Wouterlood and Jorritsma-Byham, 1993). Compared to modern tract-tracing studies, many of the old reports are coarse and provide little information for the current connectome. Such reports are only of interest when they seem valid and provide the only available source of knowledge about a particular connection.

A bigger problem occurs when a report provides many connection details, but too little evidence is presented to confirm the claims of authors. In such cases, one can only trust the author's interpretation of the data, in the absence of proof against it. However, when sufficient proof against the existence of a connection is available, such information is registered in our connection database, but

excluded or removed from the connectome. An example of a connection that was reported but is not in the current connectome is based on an injection in A30 and A29c with terminal labeling in Sub, septal PrS and PaS (Vogt and Miller, 1983) or an injection in Sub, PrS, and CA1 (Long et al., 1995). These data were excluded from the connectome because the tracer injections covered more than one defined region or targeted fiber bundles. Another possible confounder occurs if described results do not match those in the referred figures or tables. In such cases they were either corrected to match the figure or table or if that was impossible, they were excluded entirely. Nevertheless, such "false positive" reports are taken up in the reference tables on our website, which displays in which papers connections were reported, such that scientists can evaluate such reports for themselves.

ADVANTAGE 1: EDUCATION

Apart from scientific value, a connectome is useful as an educational tool to get an overview of an ever increasing amount of literature. Instead of having to search through many papers, basic anatomical facts such as the three dimensional organization of a structure, but also complex issues such as differences in nomenclature and perhaps most complex, the numerous connections between brain regions, are neatly organized, such that users can easily get an exhaustive overview of information at the level of detail that they choose. This is useful for novice researchers who just started to learn about the organization of the brain, as well as researchers who wish to expand their research into new brain regions that are charted in the connectome.

ADVANTAGE 2: SCIENTIFIC VALUE

There are several scientific uses of connectomes. A connectome can help detect knowledge gaps, it will improve the interpretation of experimental results and may facilitate the design of new experiments. Depending on the level of detail of the available information, the scientific value varies between these uses. If little information is available, a connectome will mainly help detect knowledge gaps. When a reasonable amount of details are known, it will help better understand completed experiments. The true potential of a connectome becomes apparent only when the level of detail of information goes beyond the point where one can easily keep track of all the known facts.

Detection of knowledge gaps

Even though the current connectome may look overwhelming already, a close look immediately points to gaps in our knowledge. For example, there are no experiments evaluating the topographical organization of the connections between A29a or A29b and HF–PHR. Anterograde or retrograde injections in A29a or A29b and the topography of terminating patterns or labeling of neurons in PHR have not yet been described. Similarly, injections in HF–PHR and descriptions of the terminating patterns in A29a or A29b are missing. Another example concerns the topography of the projection from PrS to RSC. It is known that temporal PrS projects specifically to A29ab, whereas septal PrS projects to all RSC subdivisions. It is likely that a more specified topography exists, for example along the rostrocaudal axis of RSC, but such information is presently unknown.

Understanding experiments

Another useful scientific application is the graphical representation of hypotheses. Sometimes it is easier to understand a hypothesis by looking at a graphical representation of the connections. For example, it was hypothesized that information transfer between the RSC and PHR–HF is crucial for adequate navigation and spatial memory. This is supported by the observation that extensive lesions, including A29ab and A29c disrupt spatial memory tasks (Pothuizen et al., 2010). However, it is also likely that different subdivisions of the RSC have different roles in navigation and spatial memory, since lesioning A29c and A30 results in impaired performance on spatial learning tests, whereas selective lesions of A29ab do not interfere with the performance in spatial tasks (Van Groen et al., 2004; Vann and Aggleton, 2005). When looking at the connectome, it can be seen that A29ab receives information from septal and temporal PrS, septal and temporal Sub and septal CA1, whereas A30 and A29c receive inputs from LEA, MEA, A35, A36, POR, PaS, next to input from PrS, Sub, and CA1. These connectional differences may partially explain the lesion data, but likely will also lead to new concepts or new experiments.

Another example of an inferred experiment, based on the connectome follows from the topography observed in the RSC – HF–PHR projections. The traditional emphasis on the topography of RSC – HF–PHR projections relies on the observation that there is a gradient in the RSC to HF–PHR projections, such that rostral RSC (A29c) projects primarily to septal parts of HF–PHR (PrS) and caudal RSC (A29c/A30) projects both to septal and temporal HF–PHR (PrS). Upon close inspection, a more striking topography is apparent in the projections of Sub to A29c. The transverse axis of Sub relates to the rostrocaudal axis in A29c such that proximal Sub projects to the rostral part of A29c and distal Sub projects to the caudal of A29c. This topography could be of functional relevance, since LEA projects to proximal parts of Sub, whereas MEA projects distally (Tamamaki and Nojyo, 1995; Naber et al., 2001; Baks-Te Bulte et al., 2005). This suggests that MEA related input is selectively transferred to caudal portions of A29c and LEA related input to more rostral levels. Since the projections from CA1 to Sub follow a matching proximodistal organization and CA1 also projects to A29c, it would be of interest to know whether the CA1 to A29c projection is in register, such that distal CA1 projects to rostral A29c and proximal CA1 projects to caudal A29c.

The current connectome, as presented here, is far from complete and differences exist in the amount of information that is available about anatomical demarcation and connectivity of different brain areas. The HF connectivity is relatively well covered, followed by the PHR – HF connections and least is known about HF – RSC connectivity. As indicated, the amount of available information is largely decisive for the type of questions a connectome can assist with. We do not see these issues as shortcomings undermining the value of this approach. Connectomes can continuously be updated and extended and therefore will always provide an exhaustive and up to date account of the current knowledge. These two factors precisely define the value of connectomes. Moreover, anatomical connectivity characterizes the brain at an intermediate information level, allowing to easily link this to, e.g., functional properties of individual cells or effective connectivity (Friston, 2011). Such linking is a key prerequisite to fully understand brain function. A

connectome therefore does not only describe the architecture of the brain, it also is a key linking system to understand brain function at multiple scales.

ACKNOWLEDGMENTS

We like to thank Ingrid Riphagen for designing and performing the initial literature searches. Niels M. van Strien was supported by the Norwegian Research Council, Independent projects – Biology and Biomedicine grant number 197245. Menno P. Witter and Jørgen Sugar received support from the Kavli Foundation and the Norwegian Research Council, centre of excellence grant number 145993.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/Neuroinformatics/10.3389/fninf.2011.00007/abstract

REFERENCES

Agster, K. L., and Burwell, R. D. (2009). Cortical efferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. *Hippocampus* 19, 1159–1186.

Audinat, E., Conde, F., and Crepel, F. (1988). Cortico-cortical connections of the limbic cortex of the rat. *Exp. Brain Res.* 69, 439–443.

Baks-Te Bulte, L., Wouterlood, F. G., Vinkenoog, M., and Witter, M. P. (2005). Entorhinal projections terminate onto principal neurons and interneurons in the subiculum: a quantitative electron microscopical analysis in the rat. *Neuroscience* 136, 729–739.

Boccara, C. N., Sargolini, F., Thoresen, V. H., Solstad, T., Witter, M. P., Moser, E. I., and Moser, M. B. (2010). Grid cells in pre- and parasubiculum. *Nat. Neurosci.* 13, 987–994.

Borst, J. G., Leung, L. W., and Macfabe, D. F. (1987). Electrical activity of the cingulate cortex. II. Cholinergic modulation. *Brain Res.* 407, 81–93.

Bota, M., and Arbib, M. A. (2004). Integrating databases and expert systems for the analysis of brain structures: connections, similarities, and homologies. *Neuroinformatics* 2, 19–58.

Bota, M., and Swanson, L. W. (2010). Collating and curating neuroanatomical nomenclatures: principles and use of the brain architecture knowledge management system (BAMS). *Front. Neuroinform.* 4:3. doi: 10.3389/fninf.2010.00003.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbauers.* Leipzig: Verlag von Johann Ambrosius Barth.

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The brain's default network: anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* 1124, 1–38.

Burwell, R. D., and Amaral, D. G. (1998). Cortical afferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. *J. Comp. Neurol.* 398, 179–205.

Chen, L. L., Lin, L. H., Green, E. J., Barnes, C. A., and Mcnaughton, B. L. (1994). Head-direction cells in the rat posterior cortex. I. Anatomical distribution and behavioral modulation. *Exp. Brain Res.* 101, 8–23.

Cho, J., and Sharp, P. E. (2001). Head direction, place, and movement correlates for cells in the rat retrosplenial cortex. *Behav. Neurosci.* 115, 3–25.

Cooper, B. G., and Mizumori, S. J. (1999). Retrosplenial cortex inactivation selectively impairs navigation in darkness. *Neuroreport* 10, 625–630.

Crick, F., and Koch, C. (2003). A framework for consciousness. *Nat. Neurosci.* 6, 119–126.

Deacon, T. W., Eichenbaum, H., Rosenberg, P., and Eckmann, K. W. (1983). Afferent connections of the perirhinal cortex in the rat. *J. Comp. Neurol.* 220, 168–190.

Finch, D. M., Derian, E. L., and Babb, T. L. (1984). Afferent fibers to rat cingulate cortex. *Exp. Neurol.* 83, 468–485.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36.

Frohlich, J., and Ott, T. (1980). Autoradiographic analysis of projections from the medial entorhinal cortex to limbic areas. *Folia Morphol. (Praha)* 28, 240–245.

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., and Moser, M. B. (2004). Spatial representation in the entorhinal cortex. *Science* 305, 1258–1264.

Gerfen, C. R., and Sawchenko, P. E. (1984). An anterograde neuroanatomical tracing method that shows the detailed morphology of neurons, their axons and terminals: immunohistochemical localization of an axonally transported plant lectin, Phaseolus vulgaris leucoagglutinin (PHA-L). *Brain Res.* 290, 219–238.

Gonzalo-Ruiz, A., and Bayona, I. (2001). Localization of excitatory amino acid and neuropeptide markers in neurons of the subicular complex projecting to the retrosplenial granular cortex of the rat. *Eur. J. Anat.* 5, 119–132.

Greicius, M. D., Supekar, K., Menon, V., and Dougherty, R. F. (2009). Resting-state functional connectivity reflects structural connectivity in the default mode network. *Cereb. Cortex* 19, 72–78.

Guldin, W. O., and Markowitsch, H. J. (1983). Cortical and thalamic afferent connections of the insular and adjacent cortex of the rat. *J. Comp. Neurol.* 215, 135–153.

Hargreaves, E. L., Rao, G., Lee, I., and Knierim, J. J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science* 308, 1792–1794.

Henderson, V. W., Mack, W., and Williams, B. W. (1989). Spatial disorientation in Alzheimer's disease. *Arch. Neurol.* 46, 391–394.

Honda, Y., Furuta, T., Kaneko, T., Shibata, H., and Sasaki, H. (2011). Patterns of axonal collateralization of single layer V cortical projection neurons in the rat presubiculum. *J. Comp. Neurol.* 519, 1395–1412.

Insausti, R., Herrero, M. T., and Witter, M. P. (1997). Entorhinal cortex of the rat: cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. *Hippocampus* 7, 146–183.

Jinno, S. (2009). Structural organization of long-range GABAergic projection system of the hippocampus. *Front. Neuroanat.* 3:13. doi: 10.3389/neuro.05.013.2009

Jones, B. F., Groenewegen, H. J., and Witter, M. P. (2005). Intrinsic connections of the cingulate cortex in the rat suggest the existence of multiple functionally segregated networks. *Neuroscience* 133, 193–207.

Jones, B. F., and Witter, M. P. (2007). Cingulate cortex projections to the parahippocampal region and hippocampal formation in the rat. *Hippocampus* 17, 957–976.

Klausberger, T., Magill, P. J., Marton, L. F., Roberts, J. D., Cobden, P. M., Buzsaki, G., and Somogyi, P. (2003). Brain-state- and cell-type-specific firing of hippocampal interneurons in vivo. *Nature* 421, 844–848.

Kohler, C. (1985). Intrinsic projections of the retrohippocampal region in the rat brain. I. The subicular complex. *J. Comp. Neurol.* 236, 504–522.

Krettek, J. E., and Price, J. L. (1977). The cortical projections of the mediodorsal nucleus and adjacent thalamic nuclei in the rat. *J. Comp. Neurol.* 171, 157–191.

Krieg, W. J. (1946). Connections of the cerebral cortex; the albino rat; structure of the cortical areas. *J. Comp. Neurol.* 84, 277–323.

Long, Y., Hardwick, A. L., and Frederickson, C. J. (1995). Zinc-containing innervation of the subicular region in the rat. *Neurochem. Int.* 27, 95–103.

Maguire, E. A. (2001). The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. *Scand. J. Psychol.* 42, 225–238.

Meibach, R. C., and Siegel, A. (1977). Subicular projections to the posterior cingulate cortex in rats. *Exp. Neurol.* 57, 264–274.

Miyashita, T., and Rockland, K. S. (2007). GABAergic projections from the hippocampus to the retrosplenial cortex in the rat. *Eur. J. Neurosci.* 26, 1193–1204.

Muller, R. U., Ranck, J. B. Jr., and Taube, J. S. (1996). Head direction cells: properties and functional significance. *Curr. Opin. Neurobiol.* 6, 196–206.

Naber, P. A., Lopes Da Silva, F. H., and Witter, M. P. (2001). Reciprocal connections between the entorhinal cortex and hippocampal fields CA1 and the subiculum are in register with the projections from CA1 to the subiculum. *Hippocampus* 11, 99–104.

Naber, P. A., and Witter, M. P. (1998). Subicular efferents are organized mostly as parallel projections: a double-labeling, retrograde-tracing study in the rat. *J. Comp. Neurol.* 393, 284–297.

O' Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Res.* 34, 171–175.

Pengas, G., Hodges, J. R., Watson, P., and Nestor, P. J. (2010). Focal posterior cingulate atrophy in incipient Alzheimer's disease. *Neurobiol. Aging* 31, 25–33.

Pothuizen, H. H., Davies, M., Aggleton, J. P., and Vann, S. D. (2010). Effects of selective granular retrosplenial cortex lesions on spatial working memory in rats. *Behav. Brain Res.* 208, 566–575.

Raji, C. A., Lopez, O. L., Kuller, L. H., Carmichael, O. T., and Becker, J. T. (2009). Age, Alzheimer disease, and brain structure. *Neurology* 73, 1899–1905.

Reed, J. M., and Squire, L. R. (1997). Impaired recognition memory in patients with lesions limited to the

hippocampal formation. *Behav. Neurosci.* 111, 667–675.
Rose, J. E., and Woolsey, C. N. (1948). Structure and relations of limbic cortex and anterior thalamic nuclei in rabbit and cat. *J. Comp. Neurol.* 89, 279–347.
Rose, M. (1927). Gyrus limbicus anterior und Regio retrosplenialis (Cortex holoprotoptychos quinquestratificatus). Vergleichende Architektonik bei Tier und Mensch. *J. Psychol. Neurol.* 35, 65–173.
Scoville, W. B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* 20, 11–21.
Sharp, P. E., and Green, C. (1994). Spatial correlates of firing patterns of single cells in the subiculum of the freely moving rat. *J. Neurosci.* 14, 2339–2356.
Shibata, H. (1994). Terminal distribution of projections from the retrosplenial area to the retrohippocampal region in the rat, as studied by anterograde transport of biotinylated dextran amine. *Neurosci. Res.* 20, 331–336.
Shibata, H., Honda, Y., Sasaki, H., and Naito, J. (2009). Organization of intrinsic connections of the retrosplenial cortex in the rat. *Anat. Sci. Int.*
Somogyi, P., and Klausberger, T. (2005). Defined types of cortical interneurone structure space and spike timing in the hippocampus. *J. Physiol.* 562, 9–26.
Sporns, O., and Tononi, G. (2007). "Structural determinants of functional brain dynamics," in *Handbook of Brain Connectivity*, eds V. K. Jirsa and A. R. Mcintosh (Berlin: Springer-Verlag), 117–148.
Sporns, O., Tononi, G., and Kotter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042
Sripanidkulchai, K., and Wyss, J. M. (1987). The laminar organization of efferent neuronal cell bodies in the retrosplenial granular cortex. *Brain Res.* 406, 255–269.
Stephan, K. E., Zilles, K., and Kotter, R. (2000). Coordinate-independent mapping of structural and functional data by objective relational transformation (ORT). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 37–54.
Sutherland, R. J., Whishaw, I. Q., and Kolb, B. (1988). Contributions of cingulate cortex to two forms of spatial learning and memory. *J. Neurosci.* 8, 1863–1872.
Talk, A., Kang, E., and Gabriel, M. (2004). Independent generation of theta rhythm in the hippocampus and posterior cingulate cortex. *Brain Res.* 1015, 15–24.
Tamamaki, N., and Nojyo, Y. (1995). Preservation of topography in the connections between the subiculum, field CA1, and the entorhinal cortex in rats. *J. Comp. Neurol.* 353, 379–390.
Van Groen, T., Kadish, I., and Wyss, J. M. (2004). Retrosplenial cortex lesions of area Rgb (but not of area Rga) impair spatial learning and memory in the rat. *Behav. Brain Res.* 154, 483–491.
Van Groen, T., and Wyss, J. M. (1990a). The connections of presubiculum and parasubiculum in the rat. *Brain Res.* 518, 227–243.
Van Groen, T., and Wyss, J. M. (1990b). Connections of the retrosplenial granular a cortex in the rat. *J. Comp. Neurol.* 300, 593–606.
Van Groen, T., and Wyss, J. M. (1990c). Extrinsic projections from area CA1 of the rat hippocampus: olfactory, cortical, subcortical, and bilateral hippocampal formation projections. *J. Comp. Neurol.* 302, 515–528.
Van Groen, T., and Wyss, J. M. (1990d). The postsubicular cortex in the rat: characterization of the fourth region of the subicular cortex and its connections. *Brain Res.* 529, 165–177.
Van Groen, T., and Wyss, J. M. (1992). Connections of the retrosplenial dysgranular cortex in the rat. *J. Comp. Neurol.* 315, 200–216.
Van Groen, T., and Wyss, J. M. (2003). Connections of the retrosplenial granular b cortex in the rat. *J. Comp. Neurol.* 463, 249–263.
Van Haeften, T., Wouterlood, F. G., Jorritsma-Byham, B., and Witter, M. P. (1997). GABAergic presubicular projections to the medial entorhinal cortex of the rat. *J. Neurosci.* 17, 862–874.
Van Strien, N. M., Cappaert, N. L., and Witter, M. P. (2009). The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network. *Nat. Rev. Neurosci.* 10, 272–282.
Vann, S. D., and Aggleton, J. P. (2002). Extensive cytotoxic lesions of the rat retrosplenial cortex reveal consistent deficits on tasks that tax allocentric spatial memory. *Behav. Neurosci.* 116, 85–94.
Vann, S. D., and Aggleton, J. P. (2004). Testing the importance of the retrosplenial guidance system: effects of different sized retrosplenial cortex lesions on heading direction and spatial working memory. *Behav. Brain Res.* 155, 97–108.
Vann, S. D., and Aggleton, J. P. (2005). Selective dysgranular retrosplenial cortex lesions in rats disrupt allocentric performance of the radial-arm maze task. *Behav. Neurosci.* 119, 1682–1686.
Vann, S. D., Aggleton, J. P., and Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat. Rev. Neurosci.* 10, 792–802.
Villain, N., Desgranges, B., Viader, F., De La Sayette, V., Mezenge, F., Landeau, B., Baron, J. C., Eustache, F., and Chetelat, G. (2008). Relationships between hippocampal atrophy, white matter disruption, and gray matter hypometabolism in Alzheimer's disease. *J. Neurosci.* 28, 6174–6181.
Vogt, B. A., and Miller, M. W. (1983). Cortical connections between rat cingulate cortex and visual, motor, and postsubicular cortices. *J. Comp. Neurol.* 216, 192–210.
Vogt, B. A., and Peters, A. (1981). Form and distribution of neurons in rat cingulate cortex: areas 32, 24, and 29. *J. Comp. Neurol.* 195, 603–625.
Vogt, B. A., Vogt, L., and Farber, N. B. (2004). "Cingulate cortex and disease models," in *The Rat Nervous System*, 3rd Edn, ed. G. Paxinos (San Diego, CA: Elsevier Academic Press), 705–727.
Whishaw, I. Q., Maaswinkel, H., Gonzalez, C. L., and Kolb, B. (2001). Deficits in allothetic and idiothetic spatial behavior in rats with posterior cingulate cortex lesions. *Behav. Brain Res.* 118, 67–76.
Witter, M. P., and Amaral, D. G. (2004). "Hippocampal formation," in *The Rat Nervous System*, 3rd Edn. ed. G. Paxinos (Oxford: Elsevier Academic Press), 635–704.
Witter, M. P., Ostendorf, R. H., and Groenewegen, H. J. (1990). Heterogeneity in the dorsal subiculum of the rat. Distinct neuronal zones project to different cortical and subcortical targets. *Eur. J. Neurosci.* 2, 718–725.
Wouterlood, F. G., and Jorritsma-Byham, B. (1993). The anterograde neuroanatomical tracer biotinylated dextranamine: comparison with the tracer Phaseolus vulgaris-leucoagglutinin in preparations for electron microscopy. *J. Neurosci. Methods* 48, 75–87.
Young, C. K., and Mcnaughton, N. (2009). Coupling of theta oscillations between anterior and posterior midline cortex and with the hippocampus in freely behaving rats. *Cereb. Cortex* 19, 24–40.
Zilles, K., and Wree, A. (1995). "Cortex: areal and laminar structure," *in The Rat Nervous System*, 2nd Edn, ed. G. Paxinos (San Diego, CA: Elsevier Academic Press), 649–685.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 March 2011; paper pending published: 02 May 2011; accepted: 27 June 2011; published online: 27 July 2011.
Citation: Sugar J, Witter MP, van Strien NM and Cappaert NLM (2011) The retrosplenial cortex: intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome. Front. Neuroinform. ***5****:7. doi: 10.3389/fninf.2011.00007*

frontiers in
NEUROINFORMATICS

METHODS ARTICLE
published: 01 November 2011
doi: 10.3389/fninf.2011.00024

Knowledge synthesis with maps of neural connectivity

Marcelo Tallis[1], Richard Thompson[2], Thomas A. Russ[1] and Gully A. P. C. Burns[1]*

[1] *Information Sciences Institute, University of Southern California, Marina del Rey, CA, USA*
[2] *Brain Architecture Group, University of Southern California, Los Angeles, CA, USA*

Edited by:
Trygve B. Leergaard, University of Oslo, Norway

Reviewed by:
Robert C. Cannon, Textensor Limited, UK
Shawn Mikula, Max-Planck Institute for Medical Research, Germany

***Correspondence:**
Gully A. P. C. Burns, Information Sciences Institute, University of Southern California, 4676 Admiralty Way, Marina del Rey, CA 90292, USA.
e-mail: gully@usc.edu

This paper describes software for *neuroanatomical knowledge synthesis* based on neural connectivity data. This software supports a mature methodology developed since the early 1990s. Over this time, the Swanson laboratory at USC has generated an account of the neural connectivity of the sub-structures of the hypothalamus, amygdala, septum, hippocampus, and bed nucleus of the stria terminalis. This is based on neuroanatomical data maps drawn into a standard brain atlas by experts. In earlier work, we presented an application for visualizing and comparing anatomical macro connections using the Swanson third edition atlas as a framework for accurate registration. Here we describe major improvements to the NeuARt application based on the incorporation of a knowledge representation of experimental design. We also present improvements in the interface and features of the data mapping components within a unified web-application. As a step toward developing an accurate *sub-regional* account of neural connectivity, we provide navigational access between the data maps and a semantic representation of area-to-area connections that they support. We do so based on an approach called "Knowledge Engineering from Experimental Design" (KEfED) model that is based on experimental variables. We have extended the underlying KEfED representation of tract-tracing experiments by incorporating the definition of a neuronanatomical data map as a measurement variable in the study design. This paper describes the software design of a web-application that allows anatomical data sets to be described within a standard experimental context and thus indexed by non-spatial experimental design features.

Keywords: knowledge engineering, neural connectivity, tract-tracing, neuroanatomical mapping

1. INTRODUCTION

Neural connectivity may be described at three different scales: *macro* connections (representations of large-scale projections between gray matter regions); *meso*connections (descriptions of connections between defined neuron types); and *micro*-connections (accounts of connections between individual neurons; Swanson and Bota, 2010). The process of evolving the underlying knowledge representation of neural connectivity from the "macro" to the "meso" level is the high-level goal that motivates this current paper. Currently, neural connectivity in humans must be inferred from structural, diffusion, and functional magnetic resonance imaging data (Hagmann et al., 2007; Feldt et al., 2011). Tract-tracing techniques (Zaborszky et al., 2006; Lanciego and Wouterlood, 2011) permit direct observation of the structure of neuronal connections in the histologically prepared tissue of experimental animals. In general, each type of experiment generates data that needs to be interpreted in different ways. In particular, macro connectivity data tends to be described either (a) as spatial maps (often in the context of standardized atlases) or (b) as a symbolic representation of connections between named areas.

A framework for the logical representation of neural connectivity data has been succinctly proposed by Swanson and Bota (2010), in which they provide a terminology for standard neuroanatomical entities (such as "gray matter region," "fiber-tract," "amacrine cell," etc.). They also include terms pertaining to a modeling framework for neural connectivity (such as "wiring diagram," "node," and "connection"). The definitions of a "wiring diagram" and its constituents (nodes and connections) are defined to be general-purpose so that they may be re-used for the representation of macro-, meso-, or micro-connections.

The central contribution of this paper is based on tract-tracing studies that describe the topographic organization of macro connections by manually mapping the distribution of axonal fibers generated in tract-tracing experiments directly onto the plates of a neuroanatomical atlas, see (Dong et al., 2001b; Dong and Swanson, 2003, 2004a, 2006a) for examples that will be referred to later in this paper. This process of hand-drawing maps into a standard brain atlas provides an accurate registration of anatomical data into a standard atlas-based framework when performed carefully and correctly by experts (Swanson, 2004; Burns et al., 2006).This representation is only the first step of the larger goal of building systems to understand the substructure of a connection's fields of origin and termination in a knowledge engineering interface.

It is also important to note that a logical representation in a wiring diagram is an *interpretation* and that a geometrical representation in a data map is an *observation*[1]. In this paper, we

[1]The argument could be made that the Swanson-laboratory maps are also interpretations since a single case was selected to be representative of a class of experiments and then hand drawn over atlas plates by experts in an interpretative process. For the purposes of the work we describe here, we consider these maps to be direct observations.

exploit this distinction by using a knowledge engineering approach that incorporates the data maps into a knowledge representation for tract-tracing experimental observations. This permits us to use the reasoning framework of the knowledge engineering system to generate an interpretive "connection matrix" (that is topologically equivalent to the concept of a "wiring diagram" defined in Swanson and Bota, 2010) which has been implemented in the "BioScholar" system described in (Russ et al., 2011)[2].

We have implemented the interaction between spatial and semantic components directly as an executable neuroinformatics software application where an end-user may navigate between geometric data maps and connection matrix entries that pertain to them. The system is fully functional and open source (but the atlas delineations are copyright-protected and are therefore access-restricted). This software is described here as a practical methodology for understanding the spatial structure of macro connections within a framework for reasoning that allows users to describe experimental design. This is only an *initial step* toward constructing neuroinformatics infrastructure capable of representing the types of experiments to examine and model meso connections (see Discussion).

1.1. BIOMEDICAL KNOWLEDGE ENGINEERING TOOLS: NeuARt, BIOSCHOLAR, AND BAMS

We built this application by combining three preexisting independent systems which had their own particular technical focus: NeuARt and BioScholar were developed within the Biomedical Knowledge Engineering Group at ISI and the Brain Architecture Management System (BAMS) was developed by Mihail Bota and Larry Swanson at the University of Southern California (USC).

NeuArt is a system for viewing hand-drawn, atlas-based maps derived from histological labeling in brain tissue (Dashti et al., 2001). It was implemented as a downloadable Java application (Burns et al., 2006)[3] and has now been re-engineered as a Flex-based web-application. It provides a flexible windowing framework that permits the inclusion and manipulation of different types of User Interface (UI) component frames. This is similar in concept to the Eclipse integrated development environment[4].

The field of online neuroanatomical mapping systems includes several notable development efforts from large, well-supported teams such as Elsevier's Brain Navigator[5], the International Neuroinformatics Coordinating Facility's (INCF's) "Scalable Brain Atlas" system[6], the Allen Institute's Mouse, and Human Brain Atlases[7], The University of California, Los Angeles (UCLA) Laboratory of NeuroImaging's (LONI) Mouse Brain Atlasing Toolkit[8] (MBAT) and other systems from academic developers. The combined NeuARt/BioScholar application is original for two reasons. Firstly, it is based on a practical methodology that has been developed "in-house" for working neuroanatomists. Secondly, the tools that we are developing have a knowledge engineering methodology that goes beyond simply dealing with the geometrical relationships of brain structures and is concerned with the argumentation structure of scientific observations and interpretations.

"BioScholar" is a knowledge engineering and management system for biomedical experimental observations derived either directly from data or curated from the literature (Russ et al., 2011). The key distinction of this system is that the underlying knowledge representation (KR) is based on a model called "Knowledge Engineering from Experimental Design" (KEfED). In this model, we focus directly on the independent and dependent variables of a given study, so that we may make a clear distinction between statements being made that are interpretive and those that are observational. We have built and deployed web-based software for generating KEfED models and then instantiated them with data. BioScholar is hosted as a BIRN capability project (see text footnote 2); it is open source and is available as a multi-platform download for collaborators, developers, and end users from http://code.google.com/p/bioscholar.

BAMS is a mature online data repository of neuroanatomical information including brain parts, molecules, cell types, and projections[9]. It is built on a MySQL database with a PHP web-interface and contains manually curated information corresponding to all the data maps currently described in NeuARt. The system's originator (Dr. Mihail Bota) provided direct access to data so that we could then import it into our KEfED model representation. Thus, this work is concerned with the construction of a unified *spatio-semantic* neuroanatomical mapping system.

1.2. OVERVIEW OF THE KEfED FORMALISM

The KEfED model is specifically designed to be a domain-independent representation of experimental "observational assertions" where the main challenge is providing a mechanism for representing the semantics of different variables (constants, parameters, measurements, and calculations). The relationship between "tract-tracing" experiments and neural macro connectivity provides an example of this distinction (Burns, 2001; Russ et al., 2011).

An experimental design may contain many variables that are not directly needed to interpret the experimental observations given a specific interpretive model (Burns et al., 2008). The "primary observations" of a tract-tracing experiment may be summarized as follows: the experimental design involves performing an injection of a tracer chemical into a specific injection-region. The experimental animal is then euthanized and its brain processed histologically to reveal the presence of labeling which itself would include an account of the labeling's location, type, and density. Data captured in this form can then be interpreted to provide a semantic representation of a macro connection, defined here as consisting of the region where the projection has its origin, its termination, and its strength (see **Figure 1**). If the tracer chemical is an "anterograde" tracer then active axonal transport is assumed to have carried the tracer from the injection site to the regions containing transported labeling. If the tracer chemical is "retrograde" then the tracer will be carried from the terminals back to the neurons' somata. "Strength" refers to a qualitative measure of the relative magnitude of the projections (inputs and/or outputs)

[2]http://www.bioscholar.org/
[3]Available for download from http://sourceforge.net/projects/neuroscholar/
[4]http://eclipse.org/
[5]http://www.brainnav.com/
[6]http://scalablebrainatlas.incf.org/
[7]http://www.brain-map.org/
[8]http://mbat.loni.ucla.edu/
[9]http://brancusi.usc.edu/bkms/

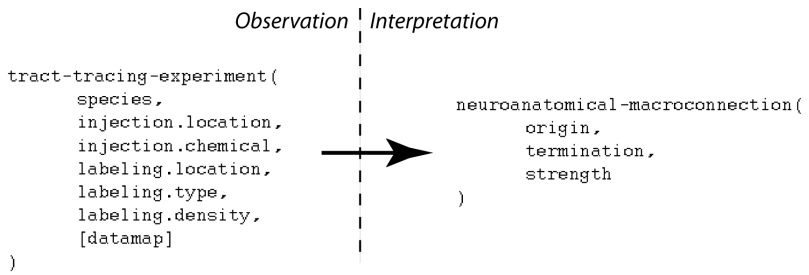


FIGURE 1 | A simplified view illustrating how a small number of variables contributing to the primary observations of a tract-tracing experiment map to the variables supporting an interpretive assertion describing a macro connection. Note the presence of an additional variable ("data map") assigned to the histological maps that may be viewed within the NeuARt system.

labeled by a tracer injection site. This is different from the notion of "connection strength" which implies a measure that addresses the number of synapses involved in a specific macro-connection and so we adopt a more conservative definition.

2. MATERIALS AND METHODS

We built this application by combining two preexisting systems: BioScholar and NeuARt. The Bioscholar system is a web-based domain-independent biocuration system based on KEfED models. The NeuARt system is a web-based neuroinformatics system for accessing and comparing neuroanatomical data registered onto brain atlases. The integration process, described below, was accomplished through our group's design and reuse of our software development framework.

We produced a dataset that included the whole set of tract-tracing experiments currently included in the NeuARt application (which effectively summarizes the entire connectivity output of the Swanson laboratory over a period of 15 years). This dataset describes 78 experiments that reported 3000 axonal projections from 58 gray matter source regions into 383 target regions. The dataset also included 1059 images depicting detailed labeling data produced by these experiments and registered against a brain atlas. This dataset was composed from preexisting databases and images collections.

Both the BioScholar and NeuARt systems are built on a Spring[10] server with a Flex 4.0 client[11], using the BlazeDS remoting architecture, SOAP web services and bare HTTP (for image files) for client-server communication. This remoting approach allows efficient server-client communication since the Java class schema operating on the server exactly mirrors the Action script class schema operating on the client. Data is stored as SVG files on disk or within a Persevere NoSQL database. The system uses the svgweb[12] library to view atlas images and maps.

2.1. BIOSCHOLAR/NeuARt INTEGRATION

The final combined application was composed of the Bioscholar and NeuARt client components since each application was web-based with their GUI behavior completely contained in the client (and implemented in Flex). Flex is a versatile, cross-platform "Rich Internet Application" (RIA) framework with a sophisticated event-handler model for user interactions. Contrary to most conventional web-based frameworks where there is a tight integration between client and server, here the presentation of data and user interaction can be contained and managed by the client programs, so that the overall integration was achieved entirely within the client.

The base client component for the combined application was based on a spatial view from the NeuARt client with added semantic views from Bioscholar. These semantic elements were a representation of a summarized connection matrix and associated experimental evidence formatted as KEfED data. We added navigational tools between the underlying knowledge and relevant maps and removed the ability for users to edit the KEfED tables. We packaged the Bioscholar client application as a library to be used in the context of the NeuARt application.

2.2. DATA MAPS

The data maps themselves were derived from a collection of Adobe Illustrator (AI) files provided directly by members of the Swanson laboratory. These files are stacks of drawings that show the observed injection sites and labeling locations hand-drawn onto atlas plates. The methodology to generate these maps was initially developed in Dr. Swanson's laboratory and is described in (Swanson, 2004). This methodology has lead to a large number of neuroanatomy publications in leading journals (Watts and Swanson, 1987; Watts et al., 1987; Canteras and Swanson, 1992a,b; Canteras et al., 1992a,b, 1994, 1995, 2010; Risold et al., 1994, 1997; Risold and Swanson, 1995, 1997a,b; Petrovich et al., 1996, 2001; Thompson et al., 1996; Petrovich and Swanson, 1997; Thompson and Swanson, 1998, 2003; Dong et al., 2001b; Dong and Swanson, 2003, 2004a,b, 2006b,c; Goto and Swanson, 2004; Goto et al., 2005; Swanson, 2005; Cenquizca and Swanson, 2007; Hahn and Swanson, 2010) system. The maps are semi-qualitative drawings of the appearance of experiments within the atlas structure. Estimates of the positional accuracy of these expert-generated maps varies by structure and number of local landmarks (or fiducial markers) but is generally on the order of 100 μm or less (R. Thompson, unpublished observation).

This data set consisted of 73 vector images (one for each coronal level in the atlas), with a large number of individual layers

[10]http://www.springsource.org/
[11]http://www.adobe.com/products/flex/
[12]http://code.google.com/p/svgweb

that each correspond to a named experiment (e.g., "ADP_PHAL" denotes a case where the tracer *Phaseolus vulgaris* Leucoagglutinin or "PHAL" was injected into the anterodorsal preoptic nucleus). The labels for these experiment layers were consistent across all images and could be used to aggregate all data corresponding to a single experiment. Given the large number of maps available, we automatically extracted all the atlas and data map images from these documents by converting each level into Scalable Vector Graphics (SVG) documents using scripts written in the AI application. We then extracted each graphic layers corresponding to an atlas level or an experimental map into separated SVG files, which could be then loaded into the NeuARt database.

2.3. GENERATING THE TRACT-TRACING EXPERIMENTS DATASET

We edited the KEfED model design for tract-tracing experiments developed previously in (Russ et al., 2011) by adding a new measurement variable to the final processing step, denoting the act of drawing these maps. We then extracted all available data from the BAMS database pertaining to the publications for which data maps were available. Some data cleaning and transformation was required to port this data into a modified KEfED model, matching terminology between BAMS and our models. This process generated the rules for relations that were used in our reasoning framework and in our base data store. The connection matrix itself (all available interpretive connection reports) was created by running a brute-force search that iterated over the brain regions of interest to find all evidence in the data for a connection. This process was time-consuming, taking about 24 h to run.

3. RESULTS

3.1. ORGANIZATION OF THE INTERFACE AND BASIC NeuARt FUNCTIONALITY

This latest version of the NeuARt application is based on a combined semantic representation of experimental design and a spatial representation of neuroanatomical maps. Under the KEfED formalism, the mechanism for this interaction is relatively straightforward: extend the existing KEfED model for tract-tracing experiments by adding a "Data Map" measurement as part of the experimental design (see **Figure 2**). The underlying logic of the experiments' relationships to their interpretations remain unchanged, but the NeuARt system now provides a specialized viewing tool for connectivity data in a neuroanatomical context. This also serves as a framework for presenting connectivity data.

The basic user interface of the NeuARt system has four panes arranged in a west-north-central-east configuration (see **Figure 3**): (A) the west pane shows coronal sections of open atlases (the system can contain several atlases); (B) the north pane shows a sagittal view of the selected atlas showing each coronal section's rostro-caudal position; (C/D) the central pane serves as the main viewing area for the various components generated within the application; the east pane has two sub sections (E) lists all maps and coronal levels, providing interactive links to the data and (F) controls for displaying particular data maps and showing their color mapping. This basic configuration mirrors that of the original NeuARt II application (Burns et al., 2006).

The system permits any of the maps in the system to be loaded directly by clicking the button marked with a [+] symbol to the bottom right. Each map denotes a single experimental case mapped over several atlas plates. Two basic components display these maps (a) views of an individual level with all data maps superimposed and (b) a view of all available levels within a single experimental case. These components appear as tabbed panes in the central panel which may itself be divided either vertically or

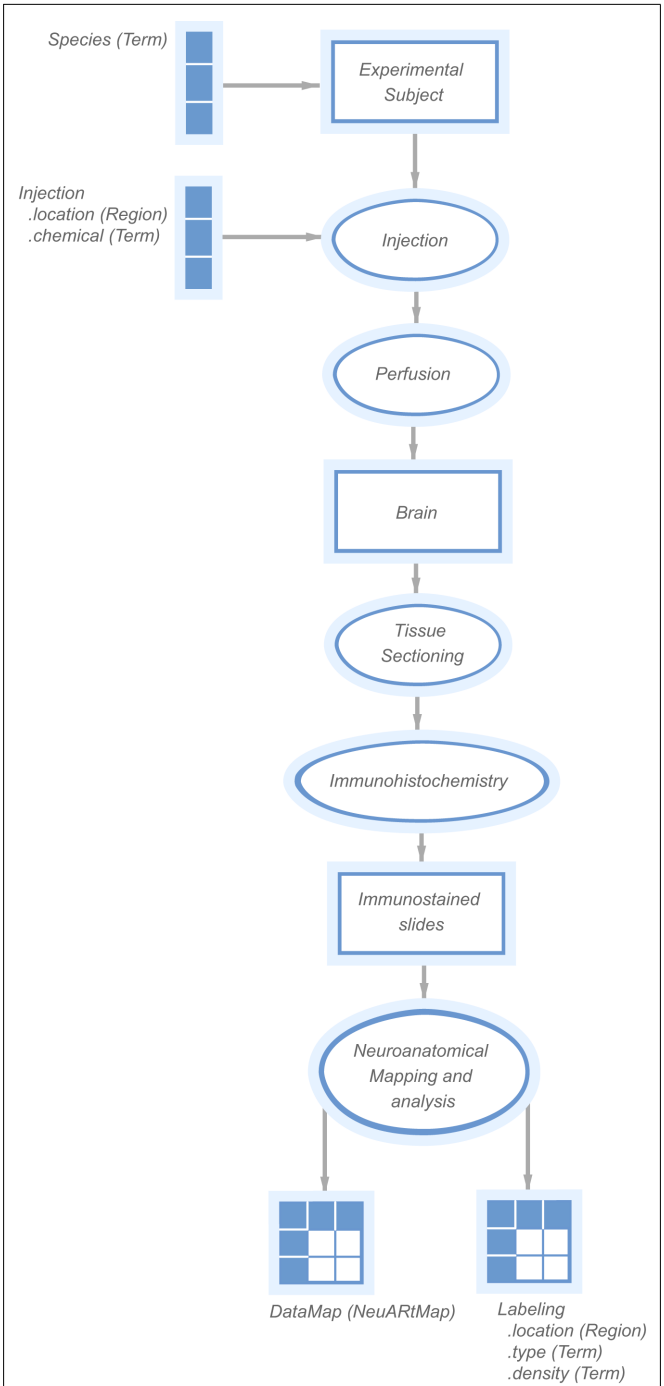


FIGURE 2 | A "KEfED" model for tract-tracing studies as it appears within the KEfED editor system. This forms the basis of the interaction between the spatial and semantic representations of these experiments.

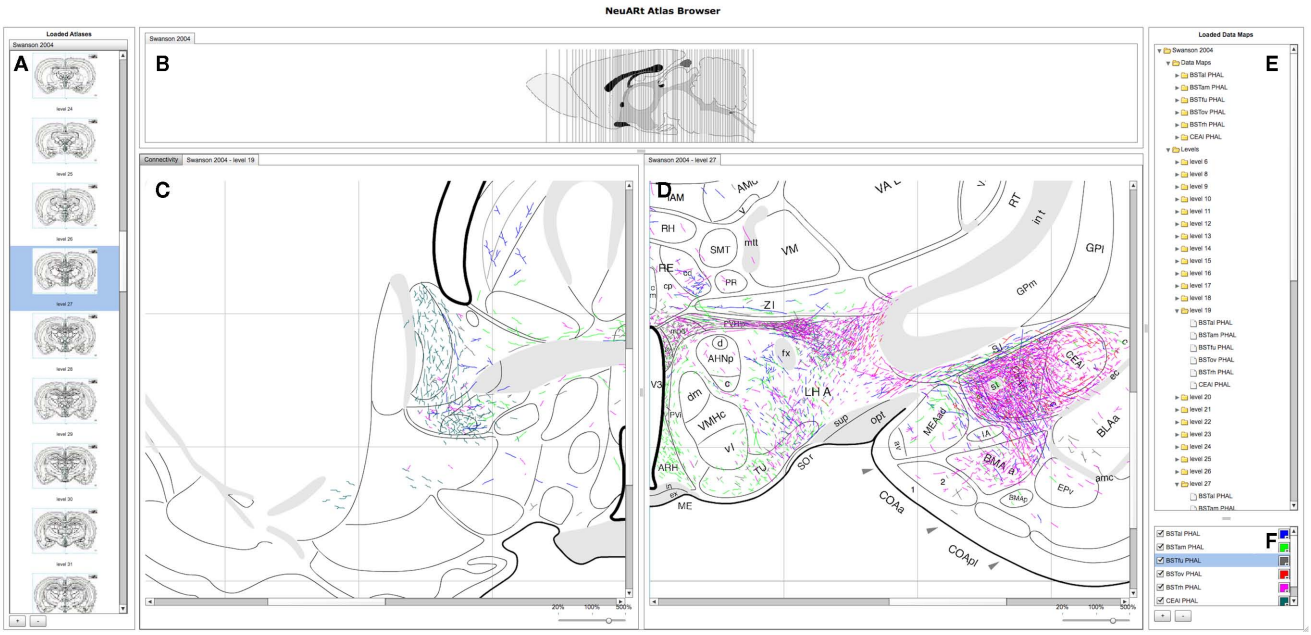


FIGURE 3 | A screenshot of the "standard" configuration of the NeuARt system specifically tailored to view neuroanatomical maps drawn in register with a brain atlas. This view shows detail of level 27 of (Swanson, 1998) with data maps pertaining to axonal labeling originating from the following subnuclei of the BST: oval (BSTov), fusiform (BSTfu; Dong et al., 2001b), rhomboid (BSTrh; Dong and Swanson, 2003), anterolateral (BSTal; Dong and Swanson, 2004a), and anteromedial (BSTam; Dong and Swanson, 2006a). We also include maps from PHAL injections into the lateral part of the Central nucleus of the Amygdala (CEAl) on the opposite side (Petrovich and Swanson, 1997). **(A)** Coronal sections for each open atlas in an "accordion" control. **(B)** A sagittal view of the open atlas (vertical lines are navigable). **(C)** Level 19 of (Swanson, 1998) in the main, vertically split central panel. **(D)** Level 27 of (Swanson, 1998) in the main, vertically split central panel. **(E)** List of the loaded data maps. **(F)** Control for selecting the maps to show and their color mapping.

horizontally to allow simultaneous viewing of multiple atlas plates. The viewer may also toggle the visibility of the maps to provide an uncluttered view. This configuration provides a functional architecture that allows other components to be added as needed to the central panel. The general-purpose nature of this central panel allowed us to implement the additional semantic functions from BioScholar.

3.2. SPATIO-SEMANTIC INTERACTIONS

We present our connectivity data summary as a "matrix" representing a directed graph of projections between the brain regions under study. As described in (Russ et al., 2011), each cell in this matrix represents an interpretive assertion about a macro connection. Within our demonstration system, we focus on representing the same connectivity data reported within the Swanson Laboratory's complete set of maps. The number of individual regions of origin and termination are quite large in this set and would generate a very large, sparse connection matrix that would be unreadable. We instead show the connections as a filtered, sortable list that the user may manipulate to show a specific subset of connections (see **Figure 4A** and **Figure 5A**; Dong et al., 2001b). This table contains three columns: Source, Target, and Strength.

Given a list of specific connections of interest, the user double-clicks on a specific interpretive assertion about a specific connection. The system then retrieves all the supporting evidence concerning that connection by executing a number of queries to an underlying knowledge base, implemented within the "PowerLoom®" architecture[13]. We are able to provide support for geometric reasoning to retrieve connection reports where either the injection site or the labeled region have relationships to the regions implicated in the connection of interest. An example of these relationships could be that the region where labeling was found is a subregion of the termination region of the connection, or that the injection site only impinged upon the connection's region of origin.

The supporting evidence is presented to the end-user as a table. Citation and PubMed ID refer to the article reporting the experiment. Chemical denotes the tracer used (e.g., *Phaseolus vulgaris* Leucoagglutinin). Label Type describes the type of cellular entity being labeled, typically *cells* for retrograde studies and *axonal fibers* for anterograde studies. Label Density is a seven-point ordinal scale to mimic the way that scientists typically describe labeling in publications (e.g., "sparse," "moderate," "dense," etc.). Injection location describes the injection site as a region defined by relations to the atlas-based regions. Label Location describes the brain regions where the labeling was observed. Data map connects to the NeuARt collection of images that document the experiment's results. Finally, Exp. Source and Exp. Label provide metadata to denote the paper and the experiment that our representation is based on. In our interface, the spatio-semantic interaction is instantiated by providing navigational links from the connection reports to their corresponding maps.

[13] http://www.isi.edu/isd/LOOM/PowerLoom/

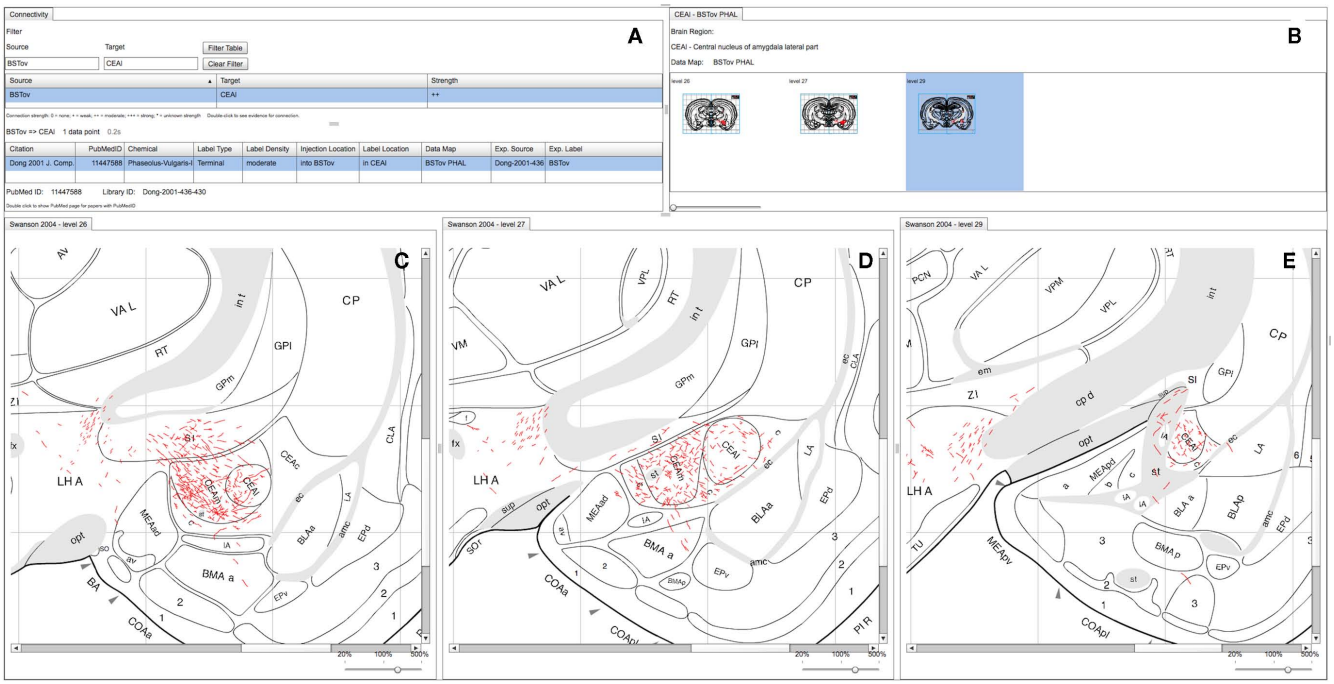


FIGURE 4 | A screenshot of the combined BioScholar/NeuARt application's central panel showing all maps pertaining to a single connection from a single piece of evidence (in this case from BSTov to the CEAl) within the central panel of the application. (A) The "connectivity matrix" panel with both interpretive summary and evidence table. **(B)** Data Map for the BSTov PHAL case limited to levels where CEAl can be found. **(C)** Level 26 of (Swanson, 1998) with BSTov PHAL data overlaid. **(D)** Level 27 of (Swanson, 1998) with BSTov PHAL data overlaid. **(E)** Level 29 of (Swanson, 1998) with BSTov PHAL data overlaid. Note, due to its time-consuming nature, data mapping generally cannot be performed on every available level.

3.3. EXAMPLE USAGE SCENARIOS

We present three scenarios to illustrate possible usage of the system as a tool for studying neural connectivity. The underlying cyto- and chemo-architecture of the bed nucleus of the stria terminalis (BST, "a rather obscure non-cortical component of the cerebral hemisphere" Dong and Swanson, 2003) was elucidated in 1989 leading to the delineation of a large number of individual nuclei within this complex cell group (Ju and Swanson, 1989; Ju et al., 1989). From 1997 to 2006, Dr. Hong Wei Dong performed an exhaustive series of tract-tracing studies describing the neural efferent connections of the constituent cell-groups of the BST generating 15 complete histological data maps, published in eight large-scale, comprehensive anatomical research articles (Dong et al., 2000, 2001a,b; Dong and Swanson, 2003, 2004a, 2006a,b,c). We use a subset of these maps to illustrate some features of the combined NeuARt/BioScholar interface.

3.3.1. Tracing circuits using maps

Figure 3 shows a high-level view of the basic configuration of the web-application with level 19 and 27 shown with data from a number of experiments. These maps are illustrative of specific features published as figures within individual studies and could not be otherwise synthesized into a single view. **Figure 3** shows the utility of the basic functionality of the NeuARt system by illustrating the connectional topography between five BST subnuclei (BSTov, BSTfu, BSTrh, BSTal, and BSTam) and the lateral part of the central nucleus of the amygdala across two atlas plates. As an illustrative example, this showcases both the process of synthesizing maps to reveal connectivity systems and also the complexity of the spatial data. Within the figure, the output connectivity of BSTrh to the CEA obscures the other labeling. The end-user may currently toggle the visibility of maps to examine patterns by interacting with the live system.

In this illustrative case, although not explicitly shown in the figure, we had used our system's query capability to discover the experiments supporting the connectivity between BST and CEA. In the following two examples we show how our toolset exploits the semantics of the KEfED data to support this type of queries.

3.3.2. Viewing all the data for an individual experiment pertaining to a specific connection

An organizing feature of the organization of the BSTov is its connections with the central nucleus of the amygdala (see Figure 10 of Dong et al., 2001a). Although a block-and-line wiring diagram could provide a summary of this projection, it cannot express any details about the spatial organization of the projection. The view shown in **Figure 4** provides a view of data pertaining to this single macro connection across all the levels for which data is available. Panel (A) of **Figure 4** shows the "connectivity matrix" summary table with its filter set to search for the required connection. The summary table shows a single report denoting the interpretative assertion concerning the presence of a connection from BSTov to CEAl. The "Evidence Table" below lists one single experimental report from (Dong et al., 2001a) describing a

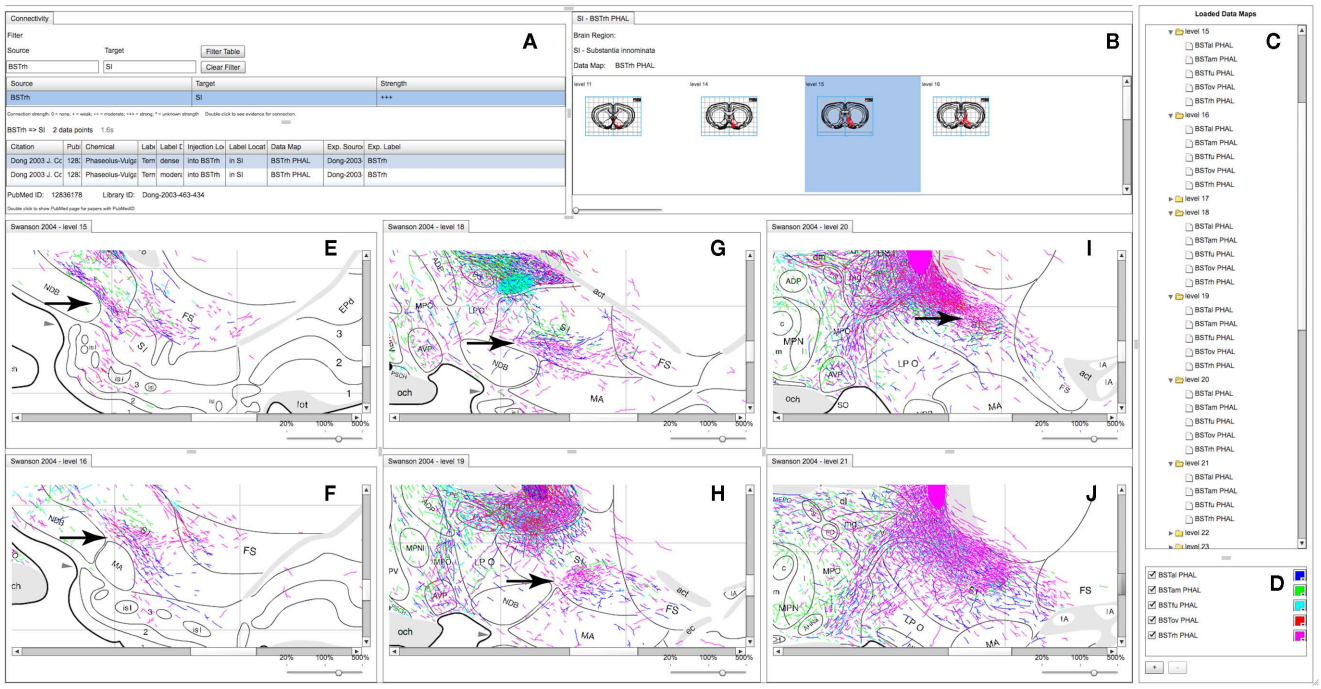


FIGURE 5 | A partial screenshot of the combined BioScholar/NeuARt application showing multiple maps possibly pertaining to the topography of the projection from multiple subnuclei of the BST to the SI. (A) The connection matrix panel showing interpretive and observational data pertaining to connections from BSTrh to SI. **(B)** One row of levels from the Data Map for the BSTrh PHAL case limited to levels where SI can be found. **(C)** List of Data Maps and Levels in the system (currently with all levels in the main panel expanded to show the maps displayed there. **(D)** Main controller for adding new Data Maps within NeuARt. **(E–J)** A selection of levels where fibers originating from BSTrh are shown in SI. These panels also include fibers from the other subnuclei shown in **(D)**.

PHAL injection into BSTov with anterograde labeling found in CEAl. The system, provides navigational tools from this summary to either the injection site or the labeling involved in the connection. It is possible to show a panel which lists each level where label is found in this case (with expandable thumbnails). Clicking on each thumbnail will generate panels with more detailed data (see **Figures 4C–E**).Note that the system does not yet lay these panels out automatically, and users must manually compose the view themselves.

This example shows that the connection to CEAm is denser than CEAl and illustrate the route that fibers take to arrive in the nucleus. It is possible to ascertain the rostro-caudal organization of the projection and provide the context of surrounding nuclei to the overall connection.

3.3.3. Examining the substructure of labeling within a larger structure

Neuroanatomical atlases allow expert users to examine the substructure of named regions. This is specifically useful when data is specifically mapped to atlas plates so that patterns between cases become evident. **Figure 5** shows maps across the substantia innominata (SI) arising from injections in multiple sub-regions of the BST where it is possible to identify a small area of consistent labeling on level 19 (see the arrow in **Figure 5H**). The description of the results of this experiment describes this labeling as part of a "rostral pathway" from BSTrh and is described as entering SI in the "rostral half and continuing to branch and generate boutons" (excerpted from page 441 of Dong and Swanson, 2003). The spatial view provides additional detail to this description. The maps in the NeuARt system can provide a view of fibers originating from other BST subnuclei within the same subregion of SI (BSTal, BSTam, and BSTov). For this to be a practical working solution, neuroanatomists would undoubtedly have to examine the experimental tissue in depth across multiple cases to be sure that any significant topography or pattern could be found. The purpose of NeuARt would be to identify possible candidates for further examination.

4. DISCUSSION AND CONCLUSION

The goal of the work underlying this paper is to describe a practical software tool to permit an end-user to access and use maps drawn onto atlas plates and then to use these maps in a semantic framework pertaining to specific experimental designs and interpretive models. We focus on tract-tracing experiments since the model is well-developed and there are available data for both maps and tract-tracing data. Our methodology is generalizable to other experimental types if they use the same hand-drawn maps and can be formulated in terms of the KEfED model. We are working to create a suite of components to serve as the basis for a general-purpose development platform for knowledge engineering tools. Flex code may also be used to generate elements that can be used by other sites and we have deployed the KEfED editor component as

an element within a Ruby-on-Rails data management application (Jacobs et al., 2009).

The neuroinformatics atlasing community has produced several systems in recent years. Notable examples include: MBAT (Lee et al., 2010), the Allen Brain Atlas (ABA)[14] (Ng et al., 2009), and the Scalable Brain Atlas[15]. Since NeuARt and KEfED were originally conceived as in-house tools for neuroanatomists directly targeting tract-tracing data, there are some differences to these other efforts. MBAT is an application toolkit that permits an investigator to query external resources (both data- and literature-centric), to register images against atlases and to compare and analyze images. It is an extensible framework where either in-house or third party tools can be included as plugins and it is a java application that must be installed locally in the user's system. Since NeuARt uses a client/server architecture to store data in a central repository to be served up to a rich internet application, it is conceivable that both the data and tooling of NeuARt could be made available to the MBAT toolkit as a plugin in the future.

There are many more similarities between the Allen Brain Atlas (ABA) and the NeuArt systems. Both ABA and NeuARt use atlases to index data and in both cases, the data was produced and registered to the atlas under a well-defined, "in-house" mapping process. The important distinction is that ABA is automated and therefore scales to process orders-of-magnitude more data. For this same reason, it lacks the detailed involvement of a human expert to place each feature from the data carefully into the atlas space. Our manual approach may be considered a more accurate method of data registration but carries a significant cost in processing speed. In the current paper, we also show the use of a representation of the logic of the underlying experiment. In neuroanatomical databases in general, this information would typically be included in the metadata associated with each mapped case. We have here taken the additional step of incorporating this knowledge into a reasoning framework that provides summarized evidence for each interpretative assertion of the system (i.e., the output of our semantic system is not an account of localized labeling density but an assertion about macro connections between brain structures).

The INCF's Scalable Brain Atlas is used to browse volume delineations of reference atlases and is notable due to it's incorporation of the Waxholm reference atlas which serves as an elegant computational vehicle for standardization based on three-dimensional coordinates taken from a standard brain. The INCF is also developing a service-oriented infrastructure to support the publication, discovery, and invocation of atlas-based resources (Hawrylycz et al., 2011). Translation of our data into this framework remains an important step to making the Swanson atlas maps useful within this broader community. An important barrier to the development of all of these tools is the closed-source nature of the most widely used atlases. There is a need for licenses that allow open development and dissemination of tools to be made available to neuroinformatics researchers.

Neuronal data is usually inherently indexed by the neuroanatomical location of phenomena or entities under study. Invariably, these location details are not easily machine-accessible, being expressed either as non-standardized terms embedded in the natural-language-based prose of a research article or as camera-lucida drawings, maps, and photographs within a paper's figures. This information, embedded in neuroscience papers, represents a massive outlay of funding, time, and expertise that is currently inaccessible to searching, comparing, and analyzing with current tools. This is an untapped resource that we hope to harness by using the system described in this paper as a possible biocuration vehicle. All of the data in this study was obtained from previously existing sources. It could be possible to use the interface itself to encourage end users to draw maps, to link photographs or other binary data types to specific brain locations.

Although our system deals only with knowledge at the level of macro connections (connections between brain regions), it provides a workable initial neuroinformatics infrastructure capable of being extended to describe meso connections (connections between cell types). In general, the logical basis for inferring knowledge about the connectivity of specific cell types is to combine tract-tracing experiments with another experiment type. This could involve, for example, concurrent immunohistochemical labeling of the tract tracers and one or more relevant neurotransmitters (or other neurochemicals) in the same tissue (see, e.g., Brinschwitz et al., 2010; Schofield, 2010) for examples of how this logic is applied to establish the projections of neuronal cell types). Another example can be found in functional neuroanatomical studies that seek to identify the behavioral or physiological relevance of given projection (see, e.g., Petrovich et al., 2005; Perez-Manso et al., 2006). These studies often use *in situ* hybridization histochemistry for the immediate-early gene *c-fos* as a marker of neuronal activation and record the number of retrogradely labeled neurons that express *c-fos* in response to the experimental manipulation. Finally, the NeuARt system provides a neuroinformatics platform that supports a third type of study for investigating networks of neural connections. This double-coinjection (DCI) tracing strategy is a recent development that utilizes combinations of tract tracers to examine meso connections systematically (Thompson and Swanson, 2010). In this approach two conventional tract tracers, one anterograde and one retrograde, are combined in solution and injected from the same pipette. When two coinjections are used in concert (four different tracers in total), all possible interactions between the two sites are labeled; that is, any given brain region sends and/or receives from one, both, or neither of the brain sites injected. The DCI approach exploits the ability to label multiple interactions directly in each experiment as a means of establishing registration across experiments. Together these features provide a framework that allows a comprehensive description of brain structure to be generated in an iterative manner from the synthesis of multiple incomplete descriptions.

Double- (triple- and quadruple-) labeling studies such as these are technically challenging and increasingly common in an effort to better understand the structural organization of brain circuits at the mesoscopic scale (Smith, 2007; Lanciego and Wouterlood, 2011). NeuARt's existing data maps provide a valuable collection to help scientists plan these types of experiments (by providing widespread access to the spatial organization of macro connections

[14]http://http://mouse.brain-map.org/
[15]http://scalablebrainatlas.incf.org/

in systems of interest). Additionally, neuroinformatics databases for knowledge from such studies would be challenging to encode with standard methods, but the flexibility of the KEfED approach could provide a vehicle to capture these data (and any maps generated from them). This, coupled with clear criteria for how specific cell types are defined based on experimental criteria (which variables need to be measured with which values), could support the definition of a system for representing meso connections.

We have presented a practical software system to show spatial and semantic views of tract-tracing experiments in a preliminary way as a demonstration of feasibility. This platform now permits us to (a) tailor the more detailed functionality of the tool to improve the end-user anatomists experience of using the system, (b) add new analysis functionality to discover patterns within the maps and interpret these patterns within the logical representation of neural connectivity and (c) broaden the scope of this methodology to new experimental types other than tract-tracing studies. Our software is open source and accessible via http://www.neuart.org/.

ACKNOWLEDGMENTS

This research is funded by the U.S. National Institutes of Health under the "NeuArt" project (MH079068-01A2), "BioScholar" project grant (GM083871). This work was also supported in part by the NIH through the following NCRR grant: the Biomedical Informatics Research Network (1 U24 RR025736-01). Special thanks to Mihail Bota for providing access to the underlying BAMS database. Thanks to Larry Swanson, Joel Hahn, and other members of the Swanson laboratory for their support.

REFERENCES

Brinschwitz, K., Dittgen, A., Madai, V., Lommel, R., Geisler, S., and Veh, R. (2010). Glutamatergic axons from the lateral habenula mainly terminate on GABAergic neurons of the ventral midbrain. *Neuroscience* 168, 463–476.

Burns, G., Cheng, W.-C., Thompson, R. F., and Swanson, L. (2006). The NeiuARt II system: a viewing tool for neuroanatomical data based on published neuroanatomical atlases. *BMC Bioinformatics* 7, 531. doi:10.1186/1471-2105-7-531

Burns, G. A. (2001). Knowledge management of the neuroscientific literature: the data model and underlying strategy of the neuroscholar system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1187–1208.

Burns, G. A., Feng, D., and Hovy, E. H. (2008). "Intelligent approaches to mining the primary research literature: techniques, systems, and examples," in *Computational Intelligence in Medical Informatics*, eds A. Kelemen, A. Abraham, Y. Chen, and Y. Liang (Berlin: Springer), 17–50.

Canteras, N. S., Ribeiro-Barbosa, E. R., Goto, M., Cipolla-Neto, J., and Swanson, L. W. (2010). The retinohypothalamic tract: comparison of axonal projection patterns from four major targets. *Brain Res. Rev.* 65, 0165–0173.

Canteras, N. S., Simerly, R. B., and Swanson, L. W. (1992a). Connections of the posterior nucleus of the amygdala. *J. Comp. Neurol.* 324, 143–179.

Canteras, N. S., Simerly, R. B., and Swanson, L. W. (1992b). Projections of the ventral premammillary nucleus. *J. Comp. Neurol.* 324, 195–212.

Canteras, N. S., Simerly, R. B., and Swanson, L. W. (1994). Organization of projections from the ventromedial nucleus of the hypothalamus: a *Phaseolus vulgaris*-leucoagglutinin study in the rat. *J. Comp. Neurol.* 348, 41–79.

Canteras, N. S., Simerly, R. B., and Swanson, L. W. (1995). Organization of projections from the medial nucleus of the amygdala: a phal study in the rat. *J. Comp. Neurol.* 360, 213–245.

Canteras, N. S., and Swanson, L. W. (1992a). Projections of the ventral subiculum to the amygdala, septum, and hypothalamus: a phal anterograde tract-tracing study in the rat. *J. Comp. Neurol.* 324, 180–194.

Canteras, N. S., and Swanson, L. W. (1992b). The dorsal premammillary nucleus: an unusual component of the mammillary body. *Proc. Natl. Acad. Sci. U.S.A.* 89, 10089–10093.

Cenquizca, L. A., and Swanson, L. W. (2007). Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. *Brain Res. Rev.* 56, 1–26.

Dashti, A. E., Burns, G. A. P. C., Simmons, D. M., Swanson, L., Ghandeharizadeh, S., Shahabi, C., Stone, J., and Jia, S. (2001). "The neuroanatomical rat brain viewer (NeuART)," in *Computing the Brain: A Guide to Neuroinformatics*, eds J. S. Grethe and M. A. Arbib (San Diego: Academic Press), 189–202.

Dong, H., Petrovich, G. D., and Swanson, L. W. (2000). Organization of projections from the juxtacapsular nucleus of the bst: a phal study in the rat. *Brain Res.* 859, 1–14.

Dong, H. W., Petrovich, G. D., and Swanson, L. W. (2001a). Topography of projections from amygdala to bed nuclei of the stria terminalis. *Brain Res. Brain Res. Rev.* 38, 192–246.

Dong, H. W., Petrovich, G. D., Watts, A. G., and Swanson, L. W. (2001b). Basic organization of projections from the oval and fusiform nuclei of the bed nuclei of the stria terminalis in adult rat brain. *J. Comp. Neurol.* 436, 430–455.

Dong, H. W., and Swanson, L. W. (2003). Projections from the rhomboid nucleus of the bed nuclei of the stria terminalis: implications for cerebral hemisphere regulation of ingestive behaviors. *J. Comp. Neurol.* 463, 434–472.

Dong, H. W., and Swanson, L. W. (2004a). Organization of axonal projections from the anterolateral area of the bed nuclei of the stria terminalis. *J. Comp. Neurol.* 468, 277–298.

Dong, H.-W., and Swanson, L. W. (2004b). Projections from bed nuclei of the stria terminalis, posterior division: implications for cerebral hemisphere regulation of defensive and reproductive behaviors. *J. Comp. Neurol.* 471, 396–433.

Dong, H. W., and Swanson, L. W. (2006a). Projections from bed nuclei of the stria terminalis, anteromedial area: cerebral hemisphere integration of neuroendocrine, autonomic, and behavioral aspects of energy balance. *J. Comp. Neurol.* 494, 142–178.

Dong, H. W., and Swanson, L. W. (2006b). Projections from bed nuclei of the stria terminalis, dorsomedial nucleus: implications for cerebral hemisphere integration of neuroendocrine, autonomic, and drinking responses. *J. Comp. Neurol.* 494, 75–107.

Dong, H. W., and Swanson, L. W. (2006c). Projections from bed nuclei of the stria terminalis, magnocellular nucleus: implications for cerebral hemisphere regulation of micturition, defecation, and penile erection. *J. Comp. Neurol.* 494, 108–141.

Feldt, S., Bonifazi, P., and Cossart, R. (2011). Dissecting functional connectivity of neuronal microcircuits: experimental and theoretical insights. *Trends Neurosci.* 34, 225–236.

Goto, M., Canteras, N., Burns, G., and Swanson, L. (2005). Projections from the subfornical region of the lateral hypothalamic area. *J. Comp. Neurol.* 493, 412–438.

Goto, M., and Swanson, L. W. (2004). Axonal projections from the parasubthalamic nucleus. *J. Comp. Neurol.* 469, 581–607.

Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V. J., Meuli, R., and Thiran, J.-P. (2007). Mapping human whole-brain structural networks with diffusion mri. *PLoS ONE* 2, e597. doi:10.1371/journal.pone.0000597

Hahn, J. D., and Swanson, L. W. (2010). Distinct patterns of neuronal inputs and outputs of the juxtaparaventricular and supraformical regions of the lateral hypothalamic area in the male rat. *Brain Res. Rev.* 64, 14–103.

Hawrylycz, M., Baldock, R. A., Burger, A., Hashikawa, T., Johnson, G. A., Martone, M., Ng, L., Lau, C., Larsen, S. D., Nissanov, J., Puelles, L., Ruffins, S., Verbeek, F., Zaslavsky, I., and Boline, J. (2011). Digital atlasing and standardization in the mouse brain. *PLoS Comput. Biol.* 7, e1001065. doi:10.1371/journal.pcbi.1001065

Jacobs, G. A., Llovet, P., Judson, I., White, L., Heimbuch, R., Jeager, C., and Burns, G. (2009). "MILO: a general purpose data repository for disease foundations and individual laboratories," in *Society for Neuroscience Annual Meeting*, Chicago, IL.

Ju, G., and Swanson, L. W. (1989). Studies on the cellular architecture of the bed nuclei of the stria terminalis in the rat: I. Cytoarchitecture. *J. Comp. Neurol.* 280, 587–602.

Ju, G., Swanson, L. W., and Simerly, R. B. (1989). Studies on the cellular architecture of the bed nuclei of the stria terminalis in the rat: II. Chemoarchitecture. *J. Comp. Neurol.* 280, 603–621.

Lanciego, J. L., and Wouterlood, F. G. (2011). A half century of experimental neuroanatomical tracing. *J. Chem. Neuroanat.* 42, 157–183.

Lee, D., Ruffins, S., Ng, Q., Sane, N., Anderson, S., and Toga, A. (2010). MBAT: a scalable informatics system for unifying digital atlasing workflows. *BMC Bioinformatics* 11, 608. doi:10.1186/1471-2105-11-608

Ng, L., Bernard, A., Lau, C., Overly, C. C., Dong, H.-W., Kuan, C., Pathak, S., Sunkin, S. M., Dang, C., Bohland, J. W., Bokil, H., Mitra, P. P., Puelles, L., Hohmann, J., Anderson, D. J., Lein, E. S., Jones, A. R., and Hawrylycz, M. (2009). An anatomic gene expression atlas of the adult mouse brain. *Nat. Neurosci.* 12, 356–362.

Perez-Manso, M., Barroso-Chinea, P., Aymerich, M., and Lanciego, J. (2006). "Functional" neuroanatomical tract tracing: analysis of changes in gene expression of brain circuits of interest. *Brain Res.* 1072, 91–98.

Petrovich, G. D., Canteras, N. S., and Swanson, L. W. (2001). Combinatorial amygdalar inputs to hippocampal domains and hypothalamic behavior systems. *Brain Res. Brain Res. Rev.* 38, 247–289.

Petrovich, G. D., Holland, P. C., and Gallagher, M. (2005). Amygdalar and prefrontal pathways to the lateral hypothalamus are activated by a learned cue that stimulates eating. *J. Neurosci.* 25, 8295–8302.

Petrovich, G. D., Risold, P. Y., and Swanson, L. W. (1996). Organization of projections from the basomedial nucleus of the amygdala: a phal study in the rat. *J. Comp. Neurol.* 374, 387–420.

Petrovich, G. D., and Swanson, L. W. (1997). Projections from the lateral part of the central amygdalar nucleus to the postulated fear conditioning circuit. *Brain Res.* 763, 247–254.

Risold, P. Y., Canteras, N. S., and Swanson, L. W. (1994). Organization of projections from the anterior hypothalamic nucleus: a *Phaseolus vulgaris*-leucoagglutinin study in the rat. *J. Comp. Neurol.* 348, 1–40.

Risold, P. Y., and Swanson, L. W. (1995). Evidence for a hypothalamothalamocortical circuit mediating pheromonal influences on eye and head movements. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3898–3902.

Risold, P. Y., and Swanson, L. W. (1997a). Chemoarchitecture of the rat lateral septal nucleus. *Brain Res. Brain Res. Rev.* 24, 91–113.

Risold, P. Y., and Swanson, L. W. (1997b). Connections of the rat lateral septal complex. *Brain Res. Brain Res. Rev.* 24, 115–195.

Risold, P. Y., Thompson, R. H., and Swanson, L. W. (1997). The structural organization of connections between hypothalamus and cerebral cortex. *Brain Res. Brain Res. Rev.* 24, 197–254.

Russ, T., Ramakrishnan, C., Hovy, E. H., Bota, M., and Burns, G. A. P. C. (2011). Knowledge engineering tools for reasoning with scientific observations and interpretations: a neural connectivity use case. *BMC Bioinformatics* 12, 351. doi:10.1186/1471-2105-12-351

Schofield, B. R. (2010). Projections from auditory cortex to midbrain cholinergic neurons that project to the inferior colliculus. *Neuroscience* 166, 231–240.

Smith, S. (2007). Circuit reconstruction tools today. *Curr. Opin. Neurobiol.* 17, 601–608.

Swanson, L. (2004). *Brain Maps: Structure of the Rat Brain*, 3 Edn. San Diego: Elsevier Academic Press.

Swanson, L. W. (1998). *Brain Maps: Structure of the Rat Brain*, 2 Edn. San Diego: Elsevier Academic Press.

Swanson, L. W. (2005). Anatomy of the soul as reflected in the cerebral hemispheres: neural circuits underlying voluntary control of basic motivated behaviors. *J. Comp. Neurol.* 493, 122–131.

Swanson, L. W., and Bota, M. (2010). Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20610–20617.

Thompson, R. H., Canteras, N. S., and Swanson, L. W. (1996). Organization of projections from the dorsomedial nucleus of the hypothalamus: a phal study in the rat. *J. Comp. Neurol.* 376, 143–173.

Thompson, R. H., and Swanson, L. W. (1998). Organization of inputs to the dorsomedial nucleus of the hypothalamus: a reexamination with fluorogold and phal in the rat. *Brain Res. Brain Res. Rev.* 27, 89–118.

Thompson, R. H., and Swanson, L. W. (2003). Structural characterization of a hypothalamic visceromotor pattern generator network. *Brain Res. Brain Res. Rev.* 41, 153–202.

Thompson, R. H., and Swanson, L. W. (2010). Hypothesis-driven structural connectivity analysis supports network over hierarchical model of brain architecture. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15235–15239.

Watts, A. G., and Swanson, L. W. (1987). Efferent projections of the suprachiasmatic nucleus: II. Studies using retrograde transport of fluorescent dyes and simultaneous peptide immunohistochemistry in the rat. *J. Comp. Neurol.* 258, 230–252.

Watts, A. G., Swanson, L. W., and Sanchez-Watts, G. (1987). Efferent projections of the suprachiasmatic nucleus: I. Studies using anterograde transport of *Phaseolus vulgaris* leucoagglutinin in the rat. *J. Comp. Neurol.* 258, 204–229.

Zaborszky, L., Wouterlood, F. G., and Lanciego, J. L. (2006). *Neuroanatomical Tract-Tracing 3: Molecules, Neurons, and Systems*. New York: Springer.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 April 2011; accepted: 13 October 2011; published online: 01 November 2011.

Citation: Tallis M, Thompson R, Russ TA and Burns GAPC (2011) Knowledge synthesis with maps of neural connectivity. Front. Neuroinform. ***5****:24. doi: 10.3389/fninf.2011.00024*

frontiers in
NEUROINFORMATICS

REVIEW ARTICLE
published: 28 February 2012
doi: 10.3389/fninf.2012.00002

Combining collation and annotation efforts toward completion of the rat and mouse connectomes in BAMS

Mihail Bota[1]*, Hong-Wei Dong[2] and Larry W. Swanson[1]

[1] *Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA*
[2] *Laboratory of Neuro Imaging, Department of Neurology, School of Medicine, University of California, Los Angeles, CA, USA*

Edited by:
Trygve B. Leergaard, University of Oslo, Norway

Reviewed by:
Claus Hilgetag, Jacobs University Bremen, Germany
Niels M. van Strien, University of Amsterdam, Netherlands
Rembrandt Bakker, Radboud University Nijmegen, Netherlands

***Correspondence:**
Mihail Bota, Department of Biological Sciences, University of Southern California, 3641 Watt Way, HEDCO Neuroscience Building, Los Angeles, CA 90089, USA.
e-mail: mbota@usc.edu

Many different independently published neuroanatomical parcellation schemes (brain maps, nomenclatures, or atlases) can exist for a particular species, although one scheme (a standard scheme) is typically chosen for mapping neuroanatomical data in a particular study. This is problematic for building connection matrices (connectomes) because the terms used to name structures in different parcellation schemes differ widely and interrelationships are seldom defined. Therefore, data sets cannot be compared across studies that have been mapped on different neuroanatomical atlases without a reliable translation method. Because resliceable 3D brain models for relating systematically and topographically different parcellation schemes are still in the first phases of development, it is necessary to rely on qualitative comparisons between regions and tracts that are either inserted directly by neuroanatomists or trained annotators, or are extracted or inferred by collators from the available literature. To address these challenges, we developed a publicly available neuroinformatics system, the Brain Architecture Knowledge Management System (BAMS; http://brancusi.usc.edu/bkms). The structure and functionality of BAMS is briefly reviewed here, as an exemplar for constructing interrelated connectomes at different levels of the mammalian central nervous system organization. Next, the latest version of BAMS rat macroconnectome is presented because it is significantly more populated with the number of inserted connectivity reports exceeding a benchmark value (50,000), and because it is based on a different classification scheme. Finally, we discuss a general methodology and strategy for producing global connection matrices, starting with rigorous mapping of data, then inserting and annotating it, and ending with online generation of large-scale connection matrices.

Keywords: connectome, neuroinformatics, data mining, data collation and annotation, neuroanatomy, mapping, BAMS

INTRODUCTION

The "connectome" concept was introduced by Sporns and his colleagues (Sporns et al., 2005). Initially it referred to the global matrix of macroconnections (i.e., axonal connections between gray matter regions seen as black box nodes; for its complete definition see BAMS Foundational Model of Connectivity Thesaurus, http://brancusi1.usc.edu/thesaurus/definition/connectome/; Swanson and Bota, 2010) for the human brain. The concept has since been refined and today has at least three specific meanings that are applied to the nervous system as a whole, not just the brain (Swanson and Bota, 2010; Akil et al., 2011). Besides the macroconnectome between gray matter regions, which includes functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) results, there is the global mesoconnectome, which is a matrix of all axonal connections between all neuron types; and there is the global microconnectome, which is a matrix of all axonal connections between all individual neurons in a particular animal or person (Swanson and Bota, 2010).

The level of abstraction of a connectome is variable, and it depends on the number of experimental variables and metadata (expressivity), associated with connectivity reports in a database. The most abstract levels are those used to construct macroconnection wiring diagrams and square matrices of gray matter regions that show qualitative connection strengths in different visual formats (Stephan et al., 2000b; Bota and Swanson, 2007a; Bohland et al., 2009; Swanson and Bota, 2010). However, they can be enhanced with other information like the spatial characteristics of injection and labeling sites, axon branching patterns, and routes taken by the axons through associated white matter tracts (Swanson and Bota, 2010).

One other factor is critical. Connection matrix construction requires the use of a single, internally consistent nomenclature for gray matter regions (macroconnectome), neuron types (mesoconnectome), and individual neurons (microconnectome). Thus, it is necessary to define relationships within and between the sets of gray matter regions and neuron types defined in different neuroanatomical and neuron nomenclatures, respectively, and individual neurons. The ultimate global connectome would thus consist of an integrated macroconnectome, mesoconnectome, and microconnectome based on an integrated nomenclature for all of them (see **Figure 1** in Bota and Swanson, 2007b).

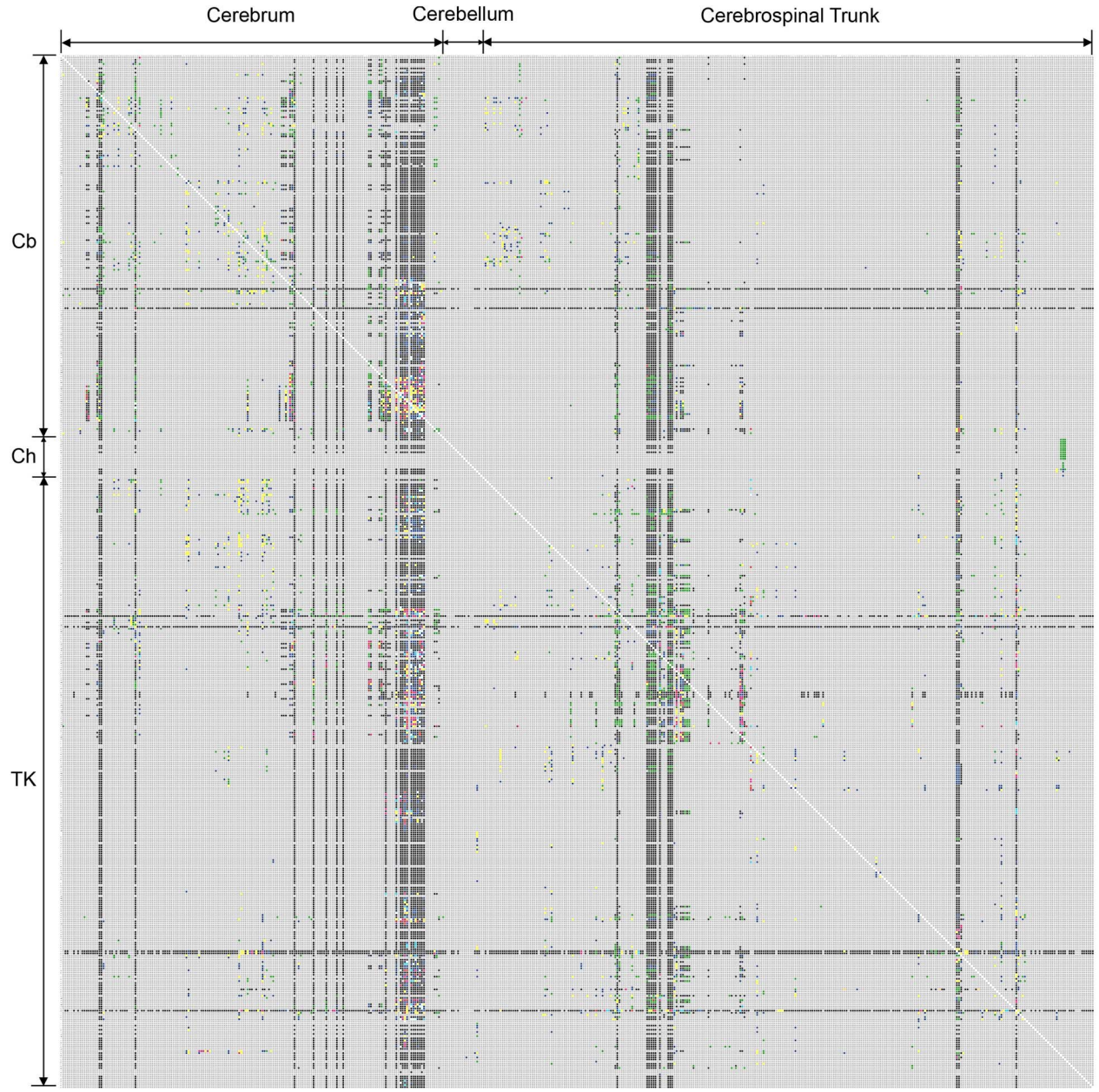


FIGURE 1 | First version of the rat CNS macroconnectome, defined in the Swanson-1998 nomenclature and constructed from data in BAMS [Bota and Swanson (2007a); Bohland et al. (2009)]. The color code used to produce both macroconnectome versions is as follows: gray—absent information; black—evidence that connection is absent; green—evidence that connection exists, without any explicit qualitative strength; red—very strong connection; pink—strong connection; yellow-to-red—medium-to-strong connection; yellow—medium or moderate connection; blue—weak or sparse connection; light blue—very weak (sparse) connection; or axons-of-passage. See Text for details.

The Brain Architecture Knowledge Management System (BAMS; http://brancusi.usc.edu/bkms; http://brancusi1.usc.edu) was designed to handle neuroanatomical information across multiple levels of vertebrate nervous system organization. It includes five interrelated *Modules* that store and process data about: (1) molecules expressed in different gray matter regions or neurons, (2) neuron types and classes, (3) gray matter regions, (4) relations between gray matter regions defined in different parcellation schemes, and (5) about connections between gray matter regions or neuron types defined in different nomenclatures and species. BAMS also includes inference engines relating gray matter regions defined in different parcellation schemes—neuron populations defined by different authors. Its backend database and inference engines currently allow construction of macroconnectomes from manually and semi-automatically inserted data, with several levels of abstraction—from simple gray

matter region wiring diagrams and 2D connectomes in different graphical formats to more detailed tabular representations of connections, augmented with spatial details that include injection and labeling sites, specific zones within regions (layers and other features), and white matter tract information (Bota et al., 2003, 2005; Bota and Swanson, 2007b).

We consider here three BAMS Modules, *Connections*, *Relations*, and *Cells* from the perspective of macroconnectome construction and relationships with mesoconnectomes, as well as the BAMS inference engine for online translation of connection reports across different nomenclatures (Bota and Swanson, 2010). The *Connections* and *Relations* Modules have been used to construct user-defined connectivity matrices and diagrams (Bota et al., 2003), as well as rat macroconnectomes since the work reported in Bota and Swanson (2007a). In this paper we present the newest version of the rat BAMS macroconnectome, which was created by using the inference engine for translation of connections reports. We also discuss the steps that have to be taken to ensure correct translation of connection reports collated from the literature. Finally, we discuss the most important challenges related to connectome construction within BAMS and argue for a collaborative and coordinated effort of the neuroinformatics community.

BAMS—MESOCONNECTOME

To reiterate: a global mesoconnectome is defined as the complete set of axonal connections between all neuron types in the nervous system of a particular species (Swanson and Bota, 2010). Depending on which methods and classification criteria are chosen, different authors may name the same or overlapping neuron sets (types or classes) differently. Therefore, one prerequisite for constructing mesoconnectomes is the construction of a single, internally consistent classification system for neuron types in a particular species. A second prerequisite is a systematic account of relationships between classification schemes used by different authors (Bota and Swanson, 2007b).

The *Cell* Module of BAMS was designed specifically to handle neuron sets (types) reports collated from the literature, as well as criteria for hierarchical classification specified by different authors. Each "is_a" relationship that relates a class to its instances is associated with a complex representation of criteria and subcriteria that are used in the literature.

BAMS also includes an inference engine that relates neuron types and classes defined in different neuron nomenclatures (Bota and Swanson, 2007b). The relations that are established across neuron populations defined or described by different authors are both qualitatively spatial (i.e., topological; Egenhofer and Franzosa, 1991), and in terms of common structural attributes. For example, two neuron sets (classes, or types) are considered identical whenever they share a common space and they have the same attributes as stated by authors, or inferred by collators.

To simplify the "knowledge maps" that can be extracted from the inserted relationships (Bota and Swanson, 2007b), we use a set of semantic relations that take into account the general spatial relations that can be defined between two neuron populations, and their compared attributes: "synonym," "includes," "partial correspondence," and "different." The details of reducing the qualitative spatial relations to a set of semantic relations, and examples of how a "knowledge map" can be constructed from the information inserted in BAMS, are discussed in Bota and Swanson (2007b). Because the sources of axonal inputs and the targets of axonal outputs for a specific neuron type or class are explicitly recorded in BAMS as classification attributes, they can be inferred using the relationships between neuron populations that include hodological criteria (i.e., the set of inputs and targets) in their definitions or descriptions. An inference engine that uses information about axonal inputs and targets for automatic classification of neurons was already implemented in BAMS. This engine automatically establishes the level of a specific neuron population within BAMS' classification scheme as well as the hierarchy of criteria, solely based on the inputs and targets, and their nature (gray matter regions or white matter tracts). Examples of such inferences are shown and discussed in detail in Bota and Swanson (2007b).

Finally, the backend structures of the *Cell* and *Connections* Modules are interrelated such that one or more connection reports about gray matter regions can be associated with the axonal connections of one or more neuron types (Bota et al., 2005). To summarize, axonal connections between neuron types or between neuron types and gray matter regions are encoded in BAMS in two different ways: first as components of macroconnections, and second as criteria for definition of neuron types and classes.

BAMS—MACROCONNECTOME

The BAMS *Connections* Module allows insertion of data and metadata at the level of macroconnections. The Entity-Relationship (ER) structure of this Module allows insertion of more than 40 qualitative, semi-quantitative, and quantitative attributes associated with a neuroanatomical connection report, as collated from the literature (Bota et al., 2005). The large number of attributes associated with any connection report (i.e., high expressivity) of the BAMS *Connections* Module allows reconstruction of connection patterns between distinct gray matter regions, including connection patterns between sets of neuron types identified in different gray matter regions.

It also allows association of macroconnection (as well as mesoconnection) reports with major white matter tracts. Thus, the *Module* can be used to reconstruct tracts in terms of contributions from one or more gray matter regions (or neuron types as described above; for details and examples see Bota et al., 2005). This BAMS feature becomes increasingly important in the context of global efforts intended to create 2D and 3D connectome maps (Hjornevik et al., 2007; Hawrylycz et al., 2011). Each connectivity report can be associated in BAMS with Atlas Levels (the serially numbered 2D maps in a brain atlas) and stereotaxic coordinates where the stain was reported. Finally, each major white matter tract of a given mammalian species can be associated with a set of neuroanatomical connections, and thus can be reconstructed in terms of the input and output regions. These three features of the BAMS *Connections* module allow it to be used as a backend data provider for reconstruction of CNS roadmaps by future visualization tools.

The high expressivity of the backend structure of BAMS's *Connections* module is necessary but not sufficient for the complete and correct insertion of connectivity data, and in the end for creating a macroconnectome using connections mapped on neuroanatomical nomenclatures (atlases) other than the standard nomenclature. The construction of any macroconnectome depends on the choice of a standard nomenclature, and it is always ideal to map new data directly onto the standard nomenclature. It also depends on the completeness and correctness of inserted data by the collators or experts.

Connectivity reports mapped on different parcellation schemes either have to be remapped by collators and curators, or the system can translate topologically the injection and labeled sites. The first option, manual remapping, does not preserve the integrity of original reports, and must be performed whenever a new nomenclature (e.g., atlas) is encountered. To preserve the original integrity of connectivity reports as collated from the associated references, remapping must become as independent as possible from the nomenclature in use. And to avoid unnecessary duplication of reports (e.g., one in the original nomenclature and the second mapped and inserted by the human expert), we employed a second option, automatic translation of connectivity data across nomenclatures.

Since standard, high-resolution, resliceable 3D computer graphics frames of reference for the rat brain and nervous system are still in the first phases of construction and testing (Hjornevik et al., 2007; Hawrylycz et al., 2011), connectome construction has to rely on a qualitative translation engine across different parcellation schemes. For this, we have constructed a special Module in BAMS, *Relations*, that allows encoding of qualitative spatial relations between nervous system parts defined in different neuroanatomical nomenclatures (e.g., atlases) in specific species. This Module also includes a large set of metadata associated with the actual process of mapping nervous system parts, performed or inserted by collators (Bota and Swanson, 2010). After the qualitative spatial relations between gray matter regions of two neuroanatomical nomenclatures defined in the same species have been inserted in BAMS, connections reports associated with either of the nomenclatures can be translated to the related one. This translation is the result of the *Connections Translations* inference engine implemented in BAMS. Full description of this engine and examples of translations are provided in Bota and Swanson.

Besides the problem of choosing a nomenclature for mapping the results of pathway tracing experiments to help establish connections (projections), the second challenge is the process of data entry itself. Any connectivity database can be populated from data collated from the literature, or it can be directly inserted by neuroanatomists, or both. The database design of the *Connections* module and its associated interfaces allow both ways of data insertion (Bota et al., 2005). BAMS is used by neuroanatomists to insert their experimental data, and manipulate in different ways the connectivity information. However, the connectivity data inserted in BAMS is mainly collated from the published literature. The collation procedure is manual and each report inserted in the system is supported by a textual annotation from the associated reference, or by collator's interpretations.

Finally, the BAMS *Connections* module is associated with a set of publicly accessible interfaces that allow construction of user-customized connections matrices. The web interface of BAMS's *Connections* module also includes inference engines that construct networks of gray matter regions, defined in specific neuroanatomical nomenclatures (Bota et al., 2003, 2005).

THE RAT MACROCONNECTOME

The first version of the rat macroconnectome was constructed from ipsilateral connectivity reports inserted in BAMS (Bota and Swanson, 2007a; Bohland et al., 2009) and used the Swanson-1998 (Swanson, 1998) nomenclature and classification hierarchy. It covered 9.4% of the entire matrix, which has 486 × 486 cells—with each cell representing a gray matter region at the bottom of the region classification hierarchy. The number of cells that are filled with any other color than gray (no data) is 22,178 (**Figure 1**).

Because the number of connection reports inserted in BAMS exceeded an internal benchmark value (50,000), we reconstructed the rat BAMS macroconnectome using the Swanson-2004 parcellation scheme (Swanson, 2004). The connection reports used in this new macroconnectome were originally mapped onto different nomenclatures recorded in BAMS; for example, Swanson (1998), Fulwiler and Saper (1984), and Moga et al. (1989). The translation of connections into the Swanson-2004 nomenclature was semi-automatic, first using the *Connections Translations* inference engine described above, and then using results validation by human agents when the relationships between Swanson-2004 nomenclature and the original nomenclature yielded equivocal results.

The increase in reports collated in BAMS's *Connections* module from the previous connectome version (Bota and Swanson, 2007a; Bohland et al., 2009) is 28.20%, from about 39,000 reports to the present value of 52,458. The present version of the rat BAMS connectome (**Figure 2**) is a matrix of 503 × 503 cells with 11.2% coverage (i.e., cells filled with any color but gray—no data). One percent coverage of the connectome matrix shown in **Figure 2** corresponds to about five completely filled columns or rows. This connectivity data increase was collated and curated from 15 newly inserted research papers collated from 2009. In addition, the results of pathway tracing experiments from more than 20 references were re-mapped, completed, or corrected.

There are two notable differences between the Swanson-1998 and Swanson-2004 connectomes shown in **Figures 1** and **2**, respectively. First, the Swanson-2004 macroconnection matrix is slightly larger than the Swanson-1998 macroconnection matrix, mostly because several gray matter regions were remapped and more finely parceled in Swanson-2004, especially the lateral hypothalamic area (LHA). However, the number of bed nuclei of the stria terminalis (BST) regions is reduced, three of them defined in Swanson-1998 nomenclature (BSTad, BSTav, BSTdl) being grouped in a single gray matter region, BSTam, in the Swanson-2004 parcellation scheme (Swanson, 2004).

The second and more important difference is the internal organization of the Swanson-1998 and Swanson-2004 nomenclatures. The main criterion for internal organization of the Swanson-2004 rat nomenclature is functional network organization (see Table B in Swanson, 2003, 2004), whereas that of Swanson-1998

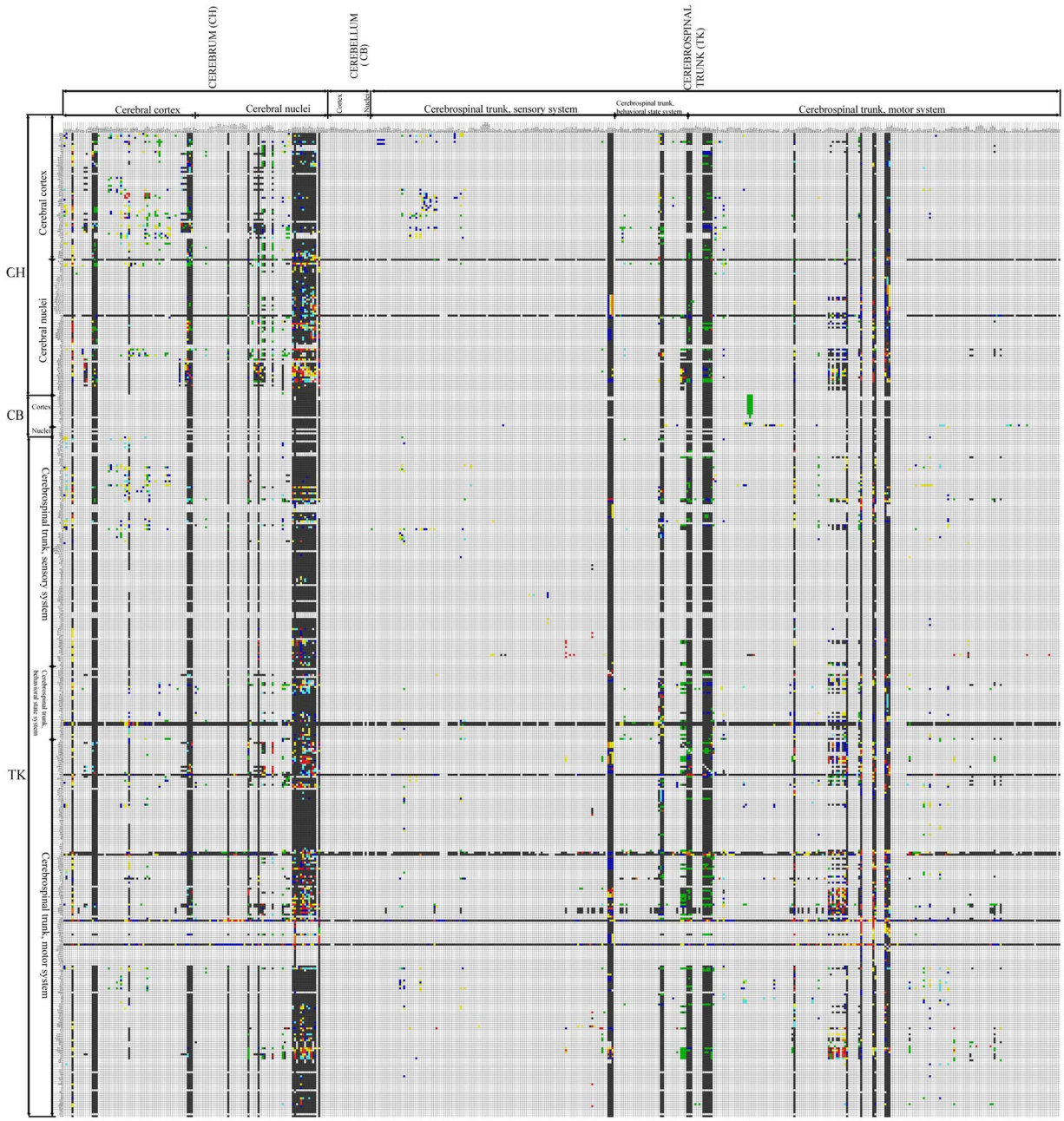


FIGURE 2 | Second version of the rat CNS macroconnectome, defined in the Swanson-2004 nomenclature and constructed from data in BAMS. The color code is identical to that used in the first version. The number of filled cells in this new rat connectome is 27,796, which represents more than 25% increase in connection data used to construct the matrix, from the 22,178 labeled cells (**Figure1**; Bota and Swanson, 2007a) in the first version of the rat connectome.

is based more on strict topographical relationships. As a result, connections of the same gray matter regions will be displayed in different columns and rows of the matrix, and under different higher-order subdivisions of the rat nervous system. The advantages of using the rat Swanson-2004 nomenclature over Swanson-1998 is thus three fold: (1) several gray matter regions are refined, (2) the nomenclature is constructed on more criteria, and (3) the nomenclature is applied consistently across the rat CNS. Hence, the internal consistency of the newest Swanson nomenclature is strengthened. Finally, the new hierarchical organization of Swanson-2004 nomenclature better integrates the structure-function relationships of rat CNS gray matter regions.

Because coverage in the newest version of the rat BAMS connectome reached a landmark value and the gray matter regions that send or receive at least one connection are not concentrated in a single CNS subdivision, we analyzed the results shown in

Figure 2. The number of connections shown to be absent (black squares) is 22,064, about 80% (79.38%) of the total. The number of regions with data about a connection to at least one other gray matter region is 317, which means that 62% of the regions in the BAMS rat connectome are associated with at least one output connection. Regions with highest numbers of inputs are in the cerebral cortex (prelimbic and infralimbic areas), cerebral nuclei (several BST nuclei), and hypothalamus (LHA regions).

The highest output connection ratio (i.e., the ratio, matrix cells in a column with data indicating an output: total number of cells in a column) shown by any gray matter region in **Figure 2** is 34%, and the highest ratio for a cerebral cortical area is 18%. In a second, more stringent and informative, step of this preliminary CNS-wide analysis we took into account only those gray matter regions with a complete set of ipsilateral outputs registered in BAMS (shown as entire filled columns in **Figure 2**). We thus determined the output connection ratio for 44 gray matter regions that include select parts of the cerebral cortex (including the subiculum), amygdalar region, lateral septal nucleus, and hypothalamus—and all parts of the BST. The output connection ratios for this subset of 44 gray matter regions range from 2% (for lateral septal nucleus subdivisions) to 34% (for LHA regions). In other words, the number of ipsilateral terminal field targets for this subset of gray matter regions ranges from 8 to 150. The average output connection ratio for this subset is 10%, which means that on average each member of the subset has about 40–50 ipsilateral gray matter region targets, out of a total of about 500 possible targets.

The average output connection ratio for the entire set of rat CNS gray matter regions may be significantly less than 10%, however, because the analyzed subset is heavily biased toward gray matter regions with very complex output patterns.

A similar situation was found for data about input connection ratios of the gray matter regions shown in **Figure 2**. Some 406 (81.2%) gray matter regions from the total of 503 have data about the reception of at least one axonal input. From data available in BAMS so far, regions that receive the most axonal inputs are mostly located in the cerebral nuclei and hypothalamus.

The connectivity data used for constructing the second BAMS rat macroconnectome is available to the neuroscience community in interactive graphical format in the newest version of BAMS: http://brancusi1.usc.edu/connections. Users can construct it online and export the data in XML or JSON formats, or as a flat image. A second XML version of the macroconnectome that includes the BAMS unique ID's of brain regions is provided in the classic version of BAMS: http://brancusi.usc.edu/bkms/brain/choose-connection.php. This additional XML version is useful for third party systems that would use the numerical ID's to bring additional gray matter region data and metadata from BAMS. Thus, the new rat macroconnectome, or parts of it, can be linked to, replicated, analyzed, or enhanced by members of the neuroscience community.

DISCUSSION

Connectome construction at all scales (micro-, meso-, or macro-) is important for analyzing and understanding global nervous system wiring diagrams, which in turn may help generate new hypotheses and design the experiments to test them. The most abstract and simplest form of a neuroinformatics-driven connectome is a 2D matrix that shows the presence or absence of connections between gray matter regions (a macroconnectome). Obviously, such connectomes can be refined by adding more information about spatial attributes like the route taken by a connection through various white matter tracts, and specific zones (differentiations) within a particular gray matter region. In the following we discuss two of the most important challenges we addressed in macroconnectome construction within BAMS: nomenclatures and data collation or annotation.

NOMENCLATURES

The construction of macroconnectomes, even in their simplest and most abstract form, needs to follow a set of rules. First, any macroconnectome must be associated minimally with an internally consistent nomenclature of gray matter regions that can be either based on published parcellation schemes, or can be constructed *de novo* by an expert or group of experts. For macroconnectomes, internal consistency of the chosen nomenclature is the necessary prerequisite ensuring that the gray matter regions used for matrix construction are distinct and do not overlap. In addition, nomenclature must cover the entire part of the nervous system under consideration and should be species specific. The nomenclatures proposed for nervous system parts in different mammalian species by different authors may or not be hierarchically organized. Any nomenclature that is also hierarchically organized according to specific sets of structural or functional criteria allows construction of connectomes that are more informative than those arranged simply alphabetically. Thus, the arrangement and size of any connectome in the graphical format of a 2D matrix depends on the nomenclature and internal classification schemes used.

Second, a macroconnectome based on data collated from the literature is the abstract form of results from many pathway tracing experiments mapped using a variety of non-identical methods and nomenclatures. The translation of connections mapped on parcellation schemes different from the connectome's standard nomenclature may be performed automatically, but the results must be verified and validated by human experts. Results produced by inference engines may be incomplete, or even contradictory, and human experts are necessary for checking them and resolving discrepancies.

DATA COLLATION

There are at least three aspects of connectivity data collation that influence the construction and usefulness of connectomes: level of detail, completeness, and correctness of inserted data. The level of detail associated with inserted data depends on the complexity of the computer-readable representation, and on how the results are presented in the literature. The simplest form of a connectivity report is "region X connects to Y," with no other details. Connectomes based on such information provide only a superficial view of CNS connectivity patterns in the species of interest, and they will be not as informative as those constructed from reports that include, for example, qualitative assessments and/or quantitative data about connections.

Thus, the richer the connection reports, the more informative the connectomes.

The high degree of abstraction in macroconnectomes that are organized in the format of a 2D matrix allows the construction of relatively simple wiring diagrams. More realistic and functionally relevant wiring diagrams need more structural and functional attributes associated with individual connections. This can be approached by relating each cell of a connectome matrix with the relevant data and metadata, as collated from the literature or inserted by experts. However, the detail level in connectivity reports collated from the literature is constrained by the mode of data presentation within them. Because a standard for presenting pathway tracing results in published references is not yet available, published connectivity data are organized in different ways by different authors. Usually, an original research article includes images of representative experimental material and more or less detailed descriptions of neuroanatomical connections. Thus, without the original results of the published pathway tracing experiments, the collation of connectivity data is best accomplished when displayed on a series of Atlas Level maps or images. Whenever possible, we collate the connectivity data from each Atlas Level presented in a published reference, and combine this information with the textual description provided by authors. This approach is necessary for qualitatively capturing the topographical details of terminal fields, and the axonal pathways and their routes. It is also useful for any neuroinformatic system that aims to reconstruct macroconnections in visual format (Tallis et al., 2011).

Regardless of the general procedure used to populate a database with pathway tracing information—either collation of the published literature or direct insertion by neuroanatomists—the process is manual and thus time consuming. Ideally, the process of connectivity data insertion should be performed in parallel with the mapping and annotation of pathway tracing experiments performed by neuroanatomists. However, this is not yet possible in an organized and large-scale way, so that collation and curation of the published literature is currently one of the most widely used methods for populating knowledge management systems. Examples of such systems include NeuroScholar (Burns, 2001), CoCoMac (Stephan et al., 2000b), BAMS (Bota et al., 2005; Bota and Swanson, 2010), and Temporal Lobe database (van Strien et al., 2009). The comprehensive collation and curation of connectivity data from published literature is also important from an historical perspective, for establishing novelty (by priority analysis) of current research results, and for future comparisons.

While the macroconnectomes of selected gray matter regions or subsytems in few mammalian species may be complete or nearly complete, it is difficult to assess the present degree of coverage for the complete macroconnectome of any particular species. We report here a coverage of about 11% for the rat CNS macroconnectome in matrix format. However, the expressiveness (in terms of attributes and associated metadata) of a neuroinformatics system is proportional to the time spent on curation and data entry. Minimally, all connectivity reports inserted in a neuroinformatics system should include information about species, standard nomenclature, methods used, and details about injection and labeled sites. Pathway tracing experiments rely on many different methods, each with unique advantages and limitations (Bota et al., 2003), so the results of pathway tracing experiments using different methods can be different, or even contradictory. As a result, information about pathway tracing methods used and about injection and labeled sites, respectively, is necessary for future evaluation of connectivity data reliability (Bota and Arbib, 2004). Further details, such as the Atlas Levels and spatial coordinates of injection and labeling sites, become very important in the context of 3D reconstructions of experimental results.

Because the prerequisites of connectome construction (pathway tracing data collation and relating gray matter regions across different parcellations) are very time consuming, the order of connectivity matrix filling in species of interest becomes important both for practical and collaborative reasons. Thus, the sequential release of updated, more complete macroconnectome versions for a species of interest allows the neuroscience community to perform statistical analyses on the released data, and to integrate it with already existent information. Moreover, each release can be seen as a benchmark toward completion of a very large-scale task. Because this task can only be done stepwise, the advantage of choosing a hierarchically organized nomenclature is obvious: it can be subdivided and reorganized as needed.

Such large-scale efforts can be accomplished only through collaboration. Completion of the rat macroconnectome both in a timely manner and with high quality data is a task that can only be done collaboratively by multiple neuroanatomy and neuroinformatics groups. Several neuroinformatics and neuroanatomy groups can work in parallel toward completion of major structural or functional divisions of the nervous system macroconnectome in a particular species. For example, each group can complete the macroconnectome for one major subdivision of a common, hierarchically organized, nomenclature. Collaboration and coordinated efforts of different groups (Bota and Swanson, 2007a; Akil et al., 2011) are already underway for neuroscience data integration at different levels of the vertebrate and invertebrate nervous systems. Associated with this, BAMS infrastructure already allows creation of collaborative mouse, rat, or macaque macroconnectomes with several systems, including the UCLA Mouse Connectome Project (http://www.mouseconnectome.org/), Rodent Brain Workbench (http://www.rbwb.org/; Zakiewicz et al., 2011), Temporal Lobe database (http://www.temporal-lobe.com; van Strien et al., 2009; Sugar et al., 2011), CoCoMac (http://cocomac.org), and the Brain Architecture Project (http://brainarchitecture.org/).

For example, the backend structure of BAMS is compatible with the recorded data and metadata associated with connection and neuron type reports, respectively, in two very important publicly available neuroinformatics applications—CoCoMac and CoCoDat—that were designed, developed, and populated by Rolf Kötter and his colleagues (Stephan et al., 2000b; Kötter, 2004; Dyhrfjeld-Johnsen et al., 2005). However, whereas the BAMS-implemented algorithm for qualitatively relating gray matter regions defined in different nomenclatures (for same species) uses the complete set of eight topological relations (Egenhofer and Franzosa, 1991; Sharma, 1986) that can be defined for a pair of convex regions, and is thus purely topological (Bota and Arbib,

2004; Bota et al., 2005), the Objective Relational Transformation (ORT) algorithm implemented in CoCoMac uses only five topological relations and a logical inference engine (Stephan et al., 2000a). Extensive discussion and comparison of both approaches is provided in Bota and Arbib (2004). Both BAMS and CoCoMac are integrated in the Neuroscience Information Framework and provide extensive information about gray matter regions and connectivity to the neuroscience community (NIF; http://www.neuinfo.org; Akil et al., 2011). A comprehensive comparison of BAMS with the major publicly accessible neuroinformatics systems developed by other groups was presented in Bota and Swanson, 2007a.

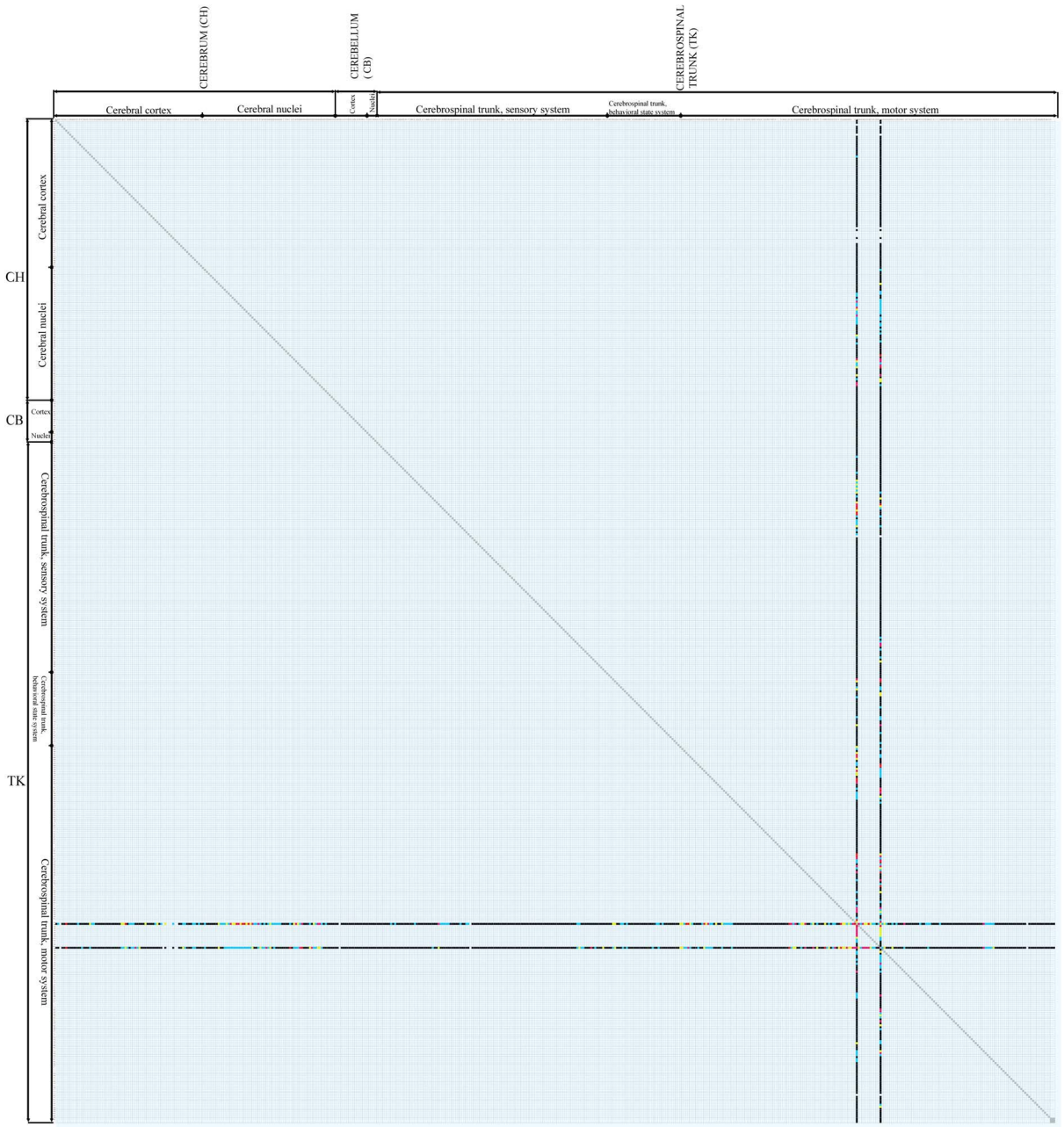


FIGURE 3 | Results of a coinjection pathway tracer analysis [Hahn and Swanson (2010)] of two nearby regions of the rat lateral hypothalamic area, the LHAjp, and LHAs, plotted on the connectome matrix for Swanson-2004 (see Figure 2). The two columns represent anterograde tracer (PHAL) data from injection sites in the LHAjp and LHAs (the leftmost and rightmost columns, respectively), whereas the two rows represent retrograde tracer (CTb) data from the same two injection sites (in different animals, in this case, though results were plotted on the same series of reference atlas level plates in the original publication). A large-scale, systematic series of coinjection sites throughout the central nervous system would gradually fill in the entire matrix. Attempts to do this in the mouse have begun in the last two years.

The backend database structure of BAMS, along with its interfaces and the data collated so far, allow us in principle and practice to construct macroconnectomes for the entire nervous system of any species (including mouse, rat, monkey, and human), using the results of pathway tracing experiments based on different methods and mapped onto different nomenclatures in the same species. The translation of connections across nomenclatures in a species is semi-automatic, and is verified by collators and experts. Using this approach, we constructed a new version of the rat macroconnectome that is the most complete connectome available to date for any vertebrate, as far as we know. This second version of the rat macroconnectome contains significantly more data than the first version, and it is based on a complete and internally consistent rat nomenclature and classification scheme for gray matter regions that facilitates network analysis. As discussed in the section above, the amount of connectivity data already inserted in BAMS allowed us to perform preliminary statistical analysis over the rat macroconnectome, and hypothesize that the average number of targets of any rat gray matter region is a maximum of 50 out of about 500.

Future work will augment the present matrix-form macroconnectome representation with spatial attributes including pathway tracer injection site and sites of connection labeling resulting from the injection. Using these attributes, we will re-implement the inference engine for evaluating connectivity data reliability (Bota and Arbib, 2004) and we will provide users with a set of tools to construct macroconnectomes that evaluate connections in different ways. In addition, we intend to complete the rat macroconnectome as best as possible from the existing literature and start constructing macroconnectomes for other species, in particular the mouse, monkey, and human. Work on the mouse connectome has already begun (mouseconnectome.org), and one novel feature here is the direct transfer of connection data from expert annotators to a reference nomenclature (Dong, 2007) in BAMS. This mouse connectome project is based on a powerful new double coinjection (DCI) method that allows to two different stereotaxically placed coinjections of an anterograde and a retrograde pathway tracer in each animal, resulting in data from four separate tracers, each labeled in a different color in each histological section (Thompson and Swanson, 2010). An example of data from two coinjections in the LHA of the rat is shown in **Figure 3**. A growing set of DCI experiment results for the mouse brain can be found at mouseconnectome.org. The networks inference engines of BAMS (Bota et al., 2003, 2005) also will be used to extract functionally relevant gray matter region networks in both rat and mouse that can be further compared.

More importantly, we will develop the backend structure and inference engines needed in BAMS to construct mesoconnectomes and microconnectomes that are fully integrated and interoperable with macroconnectomes—thus spanning and integrating the molecular, neuron, neuron type, and gray matter region levels of analysis. This extension will begin with the already implemented *Molecules* and *Cells* Modules and their relationships with the *Connections* Module.

ACKNOWLEDGMENTS

Mihail Bota and Larry W. Swanson are supported by National Institutes of Health Grant NS050792. Hong-Wei Dong is supported by Grants NIH R21MH083180 and P41RR013642. We thank Joel Hahn for providing connectivity data in electronic format and for his valuable input.

REFERENCES

Akil, H., Martone, M. E., and van Essen, D. C. (2011). Challenges and opportunities in mining neuroscience data. *Science* 6018, 708–712.

Bohland, J. W., Wu, C., Barbas, H., Bokil, H., Bota, M., Breiter, H. C., Cline, H. T., Doyle, J. C., Freed, P. J., Greenspan, R. J., Haber, S. N., Hawrylycz, M., Herrera, D. G., Hilgetag, C. C., Huang, Z. J., Jones, A., Jones, E. G., Karten, H. J., Kleinfeld, D., Kötter, R., Lester, H. A., Lin, J. M., Mensh, B. D., Mikula, S., Panksepp, J., Price, J. L., Safdieh, J., Saper, C. B., Schiff, N. D., Schmahmann, J. D., Stillman, B. W., Svoboda, K., Swanson, L. W., Toga, A. W., Van Essen, D. C., Watson, J. D., and Mitra, P. P. (2009). A proposal for coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic level scale. *PLoS Comput. Biol.* 5:e1000334. doi: 10.1371/journal.pcbi.1000334

Bota, M., and Arbib, M. A. (2004). Integrating databases and expert systems for the analysis of brain structures, connections, and homologies. *Neuroinformatics* 2, 20–59.

Bota, M., Dong, H. W., and Swanson, L. W. (2003). From gene networks to brain networks. *Nat. Neurosci.* 6, 795–799.

Bota, M., Dong, H-W., and Swanson, L. W. (2005). Brain architecture management system. *Neuroinformatics* 3, 15–48.

Bota, M., and Swanson, L. W. (2007a). Online workbenches for neural network connections. *J. Comp. Neurol.* 500, 807–814.

Bota, M., and Swanson, L. W. (2007b). The neuron classification problem. *Brain Res. Rev.* 56, 76–88.

Bota, M., and Swanson, L. W. (2010). Collating and curating neuroanatomical nomenclatures: principles of use of the Brain Architecture Management System (BAMS). *Front. Neuroinform.* 4:3. doi: 10.3389/fninf.2010.00003

Brain Architecture Project: http://brainarchitecture.org

Burns, G. A. P. C. (2001). Knowledge management of the neuroscientific literature: the data model and underlying strategy of the NeuroScholar system. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1187–1208.

Dong, H. W. (2007). *Allen Reference Atlas: A Digital Color Brain Atlas of the C57Black/67 Male Mouse.* Hoboken, NJ: Wiley.

Dyhrfjeld-Johnsen, J., Maier, J., Schubert, D., Staiger, J., Luhmann, H. J., Stephan, K. E., and Kötter, R. (2005). CoCoDat: a database system for organizing and selecting quantitative data on single neurons and neuronal microcircuitry. *J. Neurosci. Methods* 141, 291–308.

Egenhofer, M., and Franzosa, R. (1991). Point-set topological spatial relations. *Int. J. Geogr. Inf. Syst.* 5, 161–174.

Fulwiler, C. E., and Saper, C. B. (1984). Subnuclear organization of the efferent connections of the parabrachial nucleus in the rat. *Brain Res.* 319, 219–259.

Hahn, J. D., and Swanson, L. W. (2010). Distinct patterns of neuronal inputs and outputs of the juxtaparaventricular and suprafornical regions of the lateral hypothalamic area in the male rat. *Brain Res. Rev.* 64, 14–103.

Hawrylycz, M., Baldock, R. A., Burger, A., Hashikawa, T., Johnson, G. A., Martone, M., Ng, L., Lau, C., Larson, S. D., Nissanov, J., Puelles, L., Ruffins, S., Verbeek, F., Zaslavsky, I., and Boline, J. (2011). Digital atlasing and standardization in the mouse brain. *PLoS Comput. Biol.* 7:e1001065. doi: 10.1371/journal.pcbi.1001065

Hjornevik, T., Leergaard, T. B., Darine, D., Moldestad, O., Dale, A. M., Willoch, F., and Bjaalie, J. G. (2007). Three-dimensional atlas system for mouse and rat brain imaging data. *Front. Neuroinform.* 1:4. doi: 10.3389/neuro.11.004.2007

Kötter, R. (2004). Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* 2, 127–144.

Moga, M. M., Saper, C. B., and Gray, T. S. (1989). Bed nucleus of the stria terminalis: cytoarchitecture, immunohistochemistry, and projections to the parabrachial nucleus

in the rat. *J. Comp. Neurol.* 283, 283–332.

Rodent Brain Workbench: http://www.rbwb.org

Sharma, J. (1986). *Integrated spatial reasoning in geographic infomation systems: continuing topology and direction*. University of Maine, PhD Thesis.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1:e42. doi: 10.1371/journal.pcbi.0010042

Stephan, K. E., Zilles, K., and Kötter, R. (2000a). Coordinate-independent mapping of structural and functional data by objective relational transformation (ORT). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 37–54.

Stephan, K. E., Kamper, L., Bozkurt, A., Burns, G. A., Young, M. P., and Kötter, R. (2000b). Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1159–1186.

Sugar, J., Witter, M. P., van Strien, N., and Cappaert, N. L. (2011). The retrosplenial cortex: intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome. *Front. Neuroinform.* 5:7. doi: 10.3389/fninf.2011.00007

Swanson, L. W. (1998). *Brain Maps: Structure of the Rat Brain*, 2nd Edn. Amsterdam: Elsevier.

Swanson, L. W. (2003). *Brain Architecture: Understanding the Basic Plan.* New York, NY: Oxford University Press.

Swanson, L. W. (2004). *Brain Maps: Structure of the Rat Brain. A Laboratory Guide with Printed and Electronic Templates for Data, Models and Schematics*, 3rd Edn. Amsterdam: Elsevier.

Swanson, L. W., and Bota, M. (2010). Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20610–20617.

van Strien, N. M., Cappaert, N. L., and Witter, M. P. (2009). The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network. *Nat. Rev. Neurosci.* 10, 272–282.

Tallis, M., Thompson, R., Russ, T. A., and Burns, G. A. (2011). Knowledge synthesis with maps of neural connectivity. *Front. Neuroinform.* 5:24. doi: 10.3389/fninf.2011.00024

Thompson, R. H., and Swanson, L. W. (2010). Hypothesis-driven structural connectivity analysis supports network over hierarchical model of brain architecture. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15235–15239.

Zakiewicz, I. M., van Dongen, Y. C., Leergaard, T. B., and Bjaalie, J. G. (2011). Workflow and atlas system for brain-wide mapping of axonal connectivity in rat. *PLOS One* 6, e22669. doi: 10.1371/journal.pone.0022669

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 March 2011; accepted: 06 February 2012; published online: 28 February 2012.

Citation: Bota M, Dong H and Swanson LW (2012) Combining collation and annotation efforts toward completion of the rat and mouse connectomes in BAMS. Front. Neuroinform. ***6****:2. doi: 10.3389/fninf.2012.00002*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 23 September 2011
doi: 10.3389/fninf.2011.00018

Hierarchical information-based clustering for connectivity-based cortex parcellation

Nico S. Gorbach[1,2], Christoph Schütte[3], Corina Melzer[1], Mathias Goldau[4], Olivia Sujazow[1], Jenia Jitsev[1], Tania Douglas[2] and Marc Tittgemeyer[1]*

[1] *Cortical Networks Group, Max Planck Institute for Neurological Research, Cologne, Germany*
[2] *MRC/UCT Medical Imaging Research Unit, Division of Biomedical Engineering, University of Cape Town, Cape Town, South Africa*
[3] *Institute for Theoretical Physics, University of Cologne, Cologne, Germany*
[4] *Institute for Computer Science, University of Leipzig, Leipzig, Germany*

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Danielle S. Bassett, University of California Santa Barbara, USA
Saad Jbabdi, University of Oxford, UK

****Correspondence:***
Marc Tittgemeyer, Max Planck Institute for Neurological Research, Gleueler Str. 50, D-50931 Cologne, Germany.
e-mail: marc.tittgemeyer@nf.mpg.de

One of the most promising avenues for compiling connectivity data originates from the notion that individual brain regions maintain individual connectivity profiles; the functional repertoire of a cortical area ("the functional fingerprint") is closely related to its anatomical connections ("the connectional fingerprint") and, hence, a segregated cortical area may be characterized by a highly coherent connectivity pattern. Diffusion tractography can be used to identify borders between such cortical areas. Each cortical area is defined based upon a unique probabilistic tractogram and such a tractogram is representative of a group of tractograms, thereby forming the cortical area. The underlying methodology is called connectivity-based cortex parcellation and requires clustering or grouping of similar diffusion tractograms. Despite the relative success of this technique in producing anatomically sensible results, existing clustering techniques in the context of connectivity-based parcellation typically depend on several non-trivial assumptions. In this paper, we embody an unsupervised hierarchical information-based framework to clustering probabilistic tractograms that avoids many drawbacks offered by previous methods. Cortex parcellation of the inferior frontal gyrus together with the precentral gyrus demonstrates a proof of concept of the proposed method: The automatic parcellation reveals cortical subunits consistent with cytoarchitectonic maps and previous studies including connectivity-based parcellation. Further insight into the hierarchically modular architecture of cortical subunits is given by revealing coarser cortical structures that differentiate between primary as well as premotoric areas and those associated with pre-frontal areas.

Keywords: cortex parcellation, hierarchical clustering, information theory, diffusion tractography

INTRODUCTION

Subdividing the cerebral cortex into structurally and functionally distinct areas, known as cortex parcellation, arises from the notion that cortical structure reflects function. While many factors such as cytoarchitecture, myeloarchitecture, and receptor architectonics reflect the functionality of a cortical area, evidence suggests a close relationship between anatomical connectivity and functional localization within the cortex (Passingham et al., 2002). Moreover, anatomical connectivity is thought to constrain functionality and thus offers a suitable measure for differentiating between functionality of different cortical subunits. Conclusively, it has been shown at the example of the mammalian brain that structural elements of a distinct cortical region share homogeneous connectivity patterns (Hilgetag and Grant, 2000; Markov et al., 2010), which are dissimilar to those of other cortical regions and therefore determine, to some extent, the functional repertoire of that region (Stephan et al., 2000). These findings provide the basic rationale behind a methodology called connectivity-based parcellation: Structural elements with similar anatomical connectivity are grouped or clustered with the aim to segregate a cortical region of interest into functionally distinct subunits – a recent review on this topic is provided by Knösche and Tittgemeyer (2011).

In the past, information pertaining to anatomical connectivity has been mostly revealed from post-mortem and animal studies. With the advent of diffusion MRI (dMRI) and diffusion tractography, *in vivo* and *non-invasive* characterization of long-range connectivity patterns became feasible. This ultimately opened the possibility to probe the white matter structure in the human brain (Johansen-Berg and Rushworth, 2009): A convenient way to characterize anatomical connectivity of small brain areas (usually single MRI voxels) to the entire brain is the computation of probabilistic tractograms, which can be seen as an approximation (with some reservation, see Jones, 2010) to the connectivity pattern representing this brain area.

Note that, for the purpose of cortical area parcellation, probabilistic tractography does not necessarily have to accurately reflect the connectivity pattern of an individual area. The sensitivity of probabilistic tractography to differences in connectivity of cortical areas plays a much more important role. This motivates the application of tractography for connectivity-based parcellation:

When each cortical area is characterized by unique cortico-cortical connections ("connectional fingerprint"), then, any tractogram within an area should be similar.

Recently, tractography-based parcellation has been applied to a great variety of sub-cortical and cortical areas, in the macaque as well as in the human brain. These areas include the thalamus (Behrens et al., 2003; Johansen-Berg et al., 2005; Devlin et al., 2006; O'Muircheartaigh et al., 2011), basal ganglia (Lehericy et al., 2004; Sillery et al., 2005; Draganski et al., 2008), amygdala (Bach et al., 2011), midbrain (Menke et al., 2010) and cortical regions, including inferior frontal cortex (Anwander et al., 2007; Klein et al., 2007; Ford et al., 2010), premotor cortex (Tomassini et al., 2007; Schubotz et al., 2010), cingulate cortex (Beckmann et al., 2009), medial frontal (Johansen-Berg et al., 2005; Crippa et al., 2011) and insular cortex (Nanetti et al., 2009) as well as the postcentral gyrus (Roca et al., 2010).

The aforementioned attempts at clustering probabilistic tractograms, however, impose several non-trivial assumptions about the underlying structure of the data. Particularly, it is often difficult to justify the choice of a particular number of clusters *a priori*. At best, the choice of the number of cortical subunits has been subject to forming representative, meaningful cortical regions while still maintaining relative consistency across subjects. To date, two different types of clustering algorithms have been used to perform tractography-based parcellation:

(1) Similarity-based clustering methods, such as K-means clustering (Anwander et al., 2007; Klein et al., 2007; Nanetti et al., 2009) or spectral reordering (Johansen-Berg et al., 2004), employ correlation as a predefined similarity measure and thus explicitly rely on the strength of linear dependency between tractograms in order to form clusters. It is debatable, however, whether similarity between tractograms should be defined by their linear dependency to one another.
(2) Dirichlet process mixture models (Jbabdi et al., 2009) embody a Bayesian non-parametric model for clustering of probabilistic tractograms. Such stochastic processes typically assume data to be generated from a mixture of Gaussian distributions. In an application to multiple-subject parcellation of the thalamus, Jbabdi et al. (2009) represented tractograms as vectorial data and grouped them based upon a Gaussian likelihood function. Whether or not individual tractograms can be interpreted as vectors and subsequently clustered using Gaussian likelihood functions is undetermined.

A further issue concerning previous clustering attempts is that the partitioning of data into clusters that form hard borders between cortical subunits remains unjustified. The existence of a transition in cortical architecture, no matter how robust or consistent, does not necessarily signify a boundary between distinct cortical areas. An architectonic transition may instead reflect gradients or trends across the full extent of a given area. In fact it is well-known that such directional changes of cytoarchitectonic, receptor architectonic, or myeloarchitectonic properties of adjacent cortical fields can occur (Sanides, 1962; Lewis and Van Essen, 2000). A broad transition region may reflect biologically genuine gradations, such that neurons within the transition region have anatomical and/or physiological characteristics intermediate between the neighboring subdivisions. Hence, an important issue concerning parcellation is to assess the spatial extend over which such architectonic transitions occur.

Furthermore, previous clustering attempts tend to neglect the possibility of a hierarchical architecture underlying cortical subunits. Actually, brain networks are more appropriately conceived of as forming nested modules (Bassett et al., 2010; Bassett and Gazzaniga, 2011), each with a characteristic connectivity pattern – i.e., *modular hierarchies* (Kaiser and Hilgetag, 2007, 2010; Kaiser, 2011). The notion of a hierarchically modular organization of cortical subunits (Meunier et al., 2010) stems from the idea that the subunits themselves are nested into further modular structures at higher topological scales due to their similarity to one another with respect to anatomical connectivity.

The problem that an *a priori* determination of the number of clusters may not be possible clearly motivates an unsupervised clustering approach. The purpose of this study is therefore to formally adopt such an approach. Additionally, we employ an information-theoretic framework to minimize the assumptions imposed on data.

We assume for subsequent discussion that the connectivity pattern of each distinct cortical subunit retains a prototype property, referred to as exemplars in subsequent sections, such that a particular tractogram is approximately representative of the connectivity pattern of the entire cortical subunit. Further grouping of cortical subunits forms hierarchically modular structures that each contains multiple representative tractograms. Additionally, we make the prior assumption that probabilistic tractography is capable of revealing information pertaining to nested structures.

Our approach makes use of soft-constraint affinity propagation (SCAP; Leone et al., 2008) to seek exemplar tractograms that are each representative of cortical subunits. Global clusters of tractograms are formed by extracting disjoint sets of connected components each consisting of multiple exemplars. Consequently, individual global clusters are allowed to share multiple centers (i.e., exemplars) thereby allowing for the formation of irregularly shaped clusters.

The number of clusters of the global partition is determined based upon the robustness of the clustering solution against uncertainty in the data measured by clustering several bootstrap dataset samples (Fischer and Buhmann, 2003). The rationale behind this approach is to allow the uncertainty in the data to vote for the choice of exemplars and therefore the finest granularity level that gives rise to the most stable partitioning at a higher hierarchical level.

Rate distortion theory (Tishby et al., 1999) is used to stochastically map tractograms to exemplars thereby inducing a soft partition between cortical areas. A more informative nested architecture is obtained using information-theoretic agglomerative grouping (Slonim and Tishby, 1999) of cortical areas by preserving as much information as possible about the representative tractograms through the partitioning at each step of the merging sequence.

A face validation of this approach is presented using data studied in previous work, namely parcellation of the left posterior inferior pre-frontal cortex (IPC) of the human brain: Parcellation of the inferior frontal gyrus (IFG) together with the precentral gyrus (PCG) demonstrates a proof of concept of our approach. These gyri contain brain regions for which the anatomical segregation has been relatively well established (Geyer et al., 1996, 2000; Amunts et al., 2010). Both regions have been intensively studied in previous approaches to connectivity-based parcellation from our group (Anwander et al., 2007; Schubotz et al., 2010) as well as from others (Klein et al., 2007; Tomassini et al., 2007) and established reproducible results. Moreover, a modular hierarchy within the posterior IPC, conveyed through areas pointing toward primary motor, premotor, and pre-frontal brain function, is well established (Passingham, 1983; Fuster, 1997; Averbeck et al., 2009).

MATERIALS AND METHODS

dMRI DATA ACQUISITION AND PREPROCESSING

Diffusion-weighted data and high-resolution 3-dimensional (3D) T1- and T2-weighted images were acquired on a Siemens 3T Trio scanner with an eight-channel array head coil and maximum gradient strength of 40 mT/m. The diffusion-weighted data were acquired using spin-echo echo planar imaging (EPI; TR = 12 s, TE = 100 ms, 72 axial slices, resolution 1.72 mm × 1.72 mm × 1.7 mm, no cardiac gating). A GRAPPA technique (reduction factor 2.0) was chosen as parallel imaging scheme. Diffusion weighting was isotropically distributed along 60 directions (b-value = 1000 s/mm^2). Additionally, seven data sets with no diffusion weighting were acquired initially and interleaved after each block of 10 diffusion-weighted images as anatomical reference for motion correction. The high angular resolution of the diffusion weighting directions improves the robustness of the tensor estimation by increasing the signal-to-noise ratio (SNR) and reducing directional bias. To further increase SNR, scanning was repeated three times for averaging, requiring a total scan time for the dMRI protocol of approximately 45 min. dMRI data were acquired after the T2-weighted images in the same scanner reference system.

As first step in preprocessing the data, the 3D T1-weighted (MPRAGE; TR = 1300 ms, TI = 650 ms, TE = 3.97 ms, resolution 1.0 mm × 1.0 mm × 1.0 mm, flip angle 10°, 2 acquisitions) images were reoriented to the sagittal plane through the anterior and posterior commissures. Upon reorientation, the 3D T2-weighted images (RARE; TR = 2 s, TE = 355 ms, resolution 1.0 mm × 1.0 mm × 1.0 mm, flip angle 180°) were co-registered to the reoriented 3D T1-weighted images using rigid-body transformations (Jenkinson et al., 2002), implemented in FSL (http://www.fmrib.ox.ac.uk/fsl). The images without diffusion weightings were used to estimate motion correction parameters with the same registration method. The motion correction for the dMRI data was combined with the global registration to the T1 anatomy. The gradient direction for each volume was corrected using the rotation parameters. The registered images were interpolated to an isotropic voxel resolution of 1 mm and the three corresponding acquisitions were averaged. Finally, for each voxel, a diffusion tensor was fitted to the dMRI data. For presentation purposes, cortical surfaces were rendered on basis of the T1-weighted images by using Freesurfer (Dale et al., 1999).

DEFINITION OF THE REGION OF INTEREST

The region of interest was taken from the same dataset (same individual) as presented by Anwander et al., 2007, see here subject I) for the IFG and by Schubotz et al., 2010, see here subject 188) for the PCG; both regions are combined in one large region of interest for the study that is reported here. Note that the data we present reflect the left hemisphere: the left inferior frontal cortex, that is, the deep frontal operculum as well as the surface portion of the opercular and triangular part of the IFG. Since the left lateral premotor cortex cannot be determined on the basis of macroanatomical landmarks only and individual cytoarchitectonic data is not available, Schubotz et al. (2010) preselected the PCG, i.e., the anatomical region that is considered to consist of (part of) BA 4 and BA 6 (Brodmann, 1909).

PROBABILISTIC TRACTOGRAPHY

The purpose of probabilistic tractography is to characterize the connectivity pattern of cortical structural elements, denoted by seed voxels utilizing the orientation dependence of water within fiber bundles (i.e., water is more likely to diffuse along fiber bundles than across them). The 3D "random walk" method developed by (Anwander et al., 2007) attempts to quantize the connectivity pattern from a probabilistic point of view using diffusion tensor images. The random walk method describes the path taken by a particle starting from a given seed voxel and transitioning through target voxels within the white matter volume based upon local diffusivity measurements (i.e., local diffusivity measurements determine the transition probability from voxels to neighboring voxels). The probability of a particle moving to a neighboring voxel is thus greater along fiber directions. The random walk of a particle starting from the same seed voxel is repeated many times such that relative frequencies at which particles transitioned to target voxels (i.e., connectivity scores) give an appropriate measure of the probability of connectivity from particular seed voxels to target voxels. **Figure 1** illustrates the location of seed voxels at the cortical boundary and their associated probabilistic tractograms. Let each tractogram x_i be a list of connectivity scores $y_{(i)}$ for all random paths originating from a particular seed voxel to every other white matter target voxel (i.e., target voxel) such that the i-th tractogram is given by $x_i = (y^1_{(i)}, \ldots, y^a_{(i)}, \ldots, y^\eta_{(i)})$, $a \in \Omega$, where Ω denotes the set of all target voxels and η denotes the number of target voxels. Note that, for the purpose of unsupervised cortex parcellation, the set of imaging voxels compromises the whole white matter volume.

INFORMATION-BASED SIMILARITY MEASURE

Given that a pair of tractograms, x_i and x_j, are similar with respect to their connectivity, our intuition about similar tractograms arises from the notion that one tractogram x_i reveals information about connectivity associated with another tractogram x_j and vice versa. From an information-theoretic point of view, an overlap in uncertainty between tractograms x_i and x_j translates into a gain in mutual information.

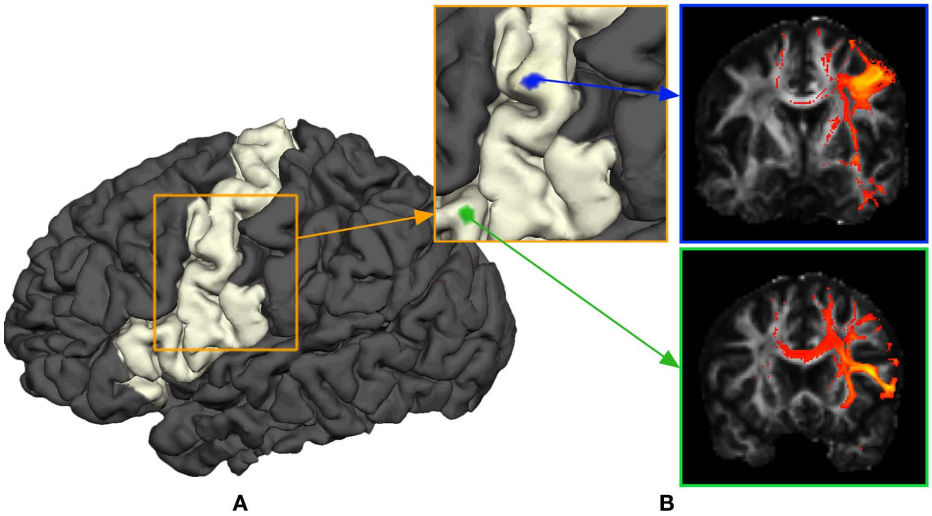


FIGURE 1 | General methodology to obtain connectivity information of a cortical area using probabilistic tractograms. (A) Seed voxels in the white matter near the white–gray matter interface form the region of interest. **(B)** A probabilistic tractogram is computed for each seed voxel.

Mutual information is one such quantity that provides a unique measure of the interdependence between tractograms:

$$I\left(x_i, x_j\right) = H\left(x_i\right) - H\left(x_i|x_j\right), \quad (1)$$

where $H(x_i)$ is the entropy and a measure of uncertainty in connectivity associated with tractogram x_i. Correspondingly, the conditional entropy $H(x_i|x_j)$ measures the remaining uncertainty in connectivity of tractogram x_i after x_j is observed. Mutual information is thus intuitively defined as the amount of uncertainty removed in x_i after observing x_j or equivalently the amount of information tractogram x_i provides about tractogram x_j. Mutual information is computed as follows:

$$I\left(x_i, x_j\right) = \sum_{y_{(i)}} \sum_{y_{(j)}} p\left(y_{(i)}, y_{(j)}\right) \log\left(\frac{p\left(y_{(i)}, y_{(j)}\right)}{p\left(y_{(i)}\right) p\left(y_{(j)}\right)}\right). \quad (2)$$

In order to compute the mutual information between two tractograms x_i and x_j we have to assume that the distribution of random variables (i.e., connectivity scores) in both tractograms are dependent upon each other. More precisely, we have to consider pairs of random variables in order to calculate the joint probability. We define a pair of random variables as $\{y^a_{(i)}, y^a_{(j)}\}$. Note that the pair of connectivity scores is defined for the *same* target voxel a, thereby preserving spatial information of tractograms. The joint occurrence of connectivity scores, $p(y_{(i)},y_{(j)})$, is then simply defined as the probability of obtaining a combination of connectivity scores in tractograms x_i and x_j for any *common* target voxel. Computing $p(y_{(i)},y_{(j)})$ is equivalent to constructing the frequency table as shown in **Figure 2**.

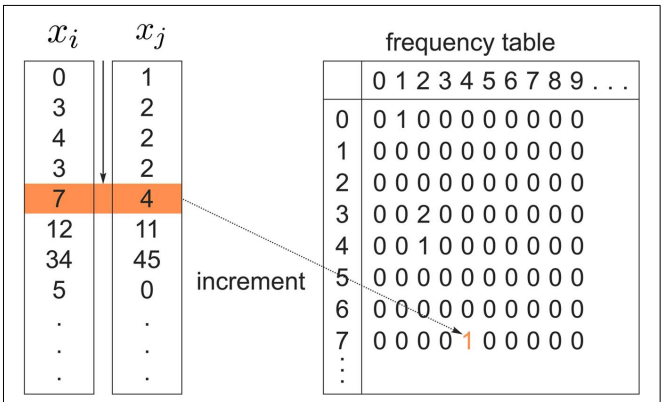


FIGURE 2 | Pairs of connectivity scores, $y^a_{(i)}$ and $y^a_{(j)}$, in common target voxels between two tractograms on the left hand side serve as coordinates of the frequency table given on the right hand side. Elements of the frequency table are incremented for each occurrence of a combination of connectivity scores $(y^a_{(i)}, y^a_{(j)})$.

IDENTIFYING REPRESENTATIVE TRACTOGRAMS

A desirable outcome of clustering probabilistic tractograms is characterizing each cortical subunit with a representative or exemplar connectivity pattern, signified as the tractographic prototype of that region. Recall that mutual information measures the dependence of tractogram x_i to tractogram x_j. Put differently, it infers the degree to which tractogram x_j is representative of tractogram x_i. The principle behind affinity propagation (Frey and Dueck, 2007) is to accumulate evidence among all pairs of tractograms to identify which tractograms are most representative of the entire cortical region of interest. The exemplar search method used in this paper is a slight variation of the original affinity propagation that softens the hard constraints inherent in the algorithm while still allowing exemplar choices that fulfill a global optimization principle. More precisely, SCAP (Leone et al., 2008) operates by iteratively updating two different messages exchanged between tractograms, denoted by "responsibility" and "availability," which

together reflect the accumulated affinity tractogram x_i has for choosing tractogram x_q as its exemplar:

$$r\left(x_i, x_q\right) \leftarrow I\left(x_i, x_q\right) - \max_{z \neq q}\left\{a\left(x_i, x_z\right) + I\left(x_i, x_z\right)\right\}, \forall x \in X,$$
$$a\left(x_i, x_q\right) \leftarrow \min\left\{0, -\rho + \sum_{z \neq i,q} \max(0, r\left(x_i, x_z\right)\right\}, \forall x \in X, \tag{3}$$

where the penalty term ρ serves as a free parameter. Tractogram x_q infers its suitability for serving as an exemplar for tractogram x_i by comparing its similarity with tractogram x_i and the maximum of similarities between tractogram x_i, corrected by availability $a(x_i,x_q)$ and all other tractograms. A positive responsibility reveals that tractogram x_i prefers tractogram x_q as its exemplar. The sum of accumulated positive incoming responsibility messages computed by availability gathers further evidence as to whether candidate exemplar x_q is a favorable exemplar for a group of tractograms. The goal of the message-passing procedure is to converge upon a set of exemplars such that the maximum net similarity of the data is attained. After convergence, the exemplar choice of x_i is extracted by selecting the candidate exemplar $\bar{x}_q$ with which tractogram x_i has maximum affinity (i.e., similarity corrected by the availability):

$$x_i \rightarrow \bar{x}_q = \arg\max_{xq}\left(s\left(x_i, x_q\right) + a\left(x_i, x_q\right)\right) \forall x \in X. \tag{4}$$

The original formulation of affinity propagation (Frey and Dueck, 2007) imposed the hard constraint that each chosen exemplar should also choose to be an exemplar for itself. SCAP relaxes the hard constraints such that a weighted availability is conveyed whenever the sum of positive responsibilities is below a penalty term ρ as shown in Eq. 3. Consequently, chosen exemplars are allowed to choose other tractograms as their exemplars (i.e., exemplars do not have to be self-exemplars) thereby forming a set of connected components. Such connected components form loops and are therefore extracted as global clusters that contain several sub-clusters of tractograms. As mentioned previously, each global cluster of tractograms contains several exemplars, which implicitly implies a nested hierarchical structure due to the association of each cortical subunit with a particular exemplar. **Figure 3** illustrates the usefulness of SCAP in revealing two levels of clustering shown for synthetic data.

Note that the penalty term ρ influences the number of global clusters K and therefore the number of exemplars. The following section discusses a means to infer the number of global clusters and therefore the optimal ρ independent of the clustering algorithm.

ESTIMATING THE NUMBER OF CLUSTERS

The method applied in this paper to assess an optimal clustering solution (i.e., to yield an automatic estimation of the number of clusters) was originally developed by Fischer and Buhmann (2003) and concerns the reliability of clustering tractograms: Uncertainty in the partitioning is quantified by clustering B bootstrap samples drawn from the original dataset. The empirical distribution of cluster assignments $\hat{p}\left(k|x_i\right)$ learned from clustering B multiset

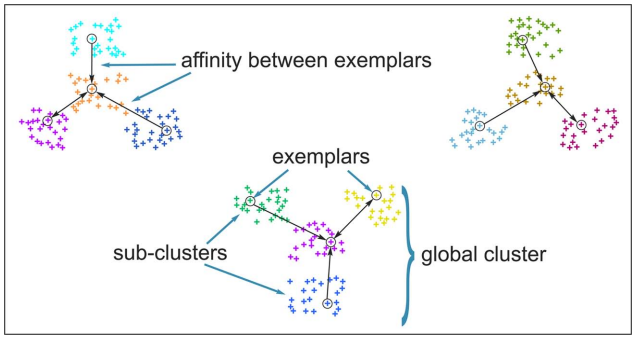


FIGURE 3 | Clustering of synthetic data to illustrate the capability of soft-constraint affinity propagation (SCAP) in capturing two levels of clustering. SCAP identifies 12 exemplars shown as circles and therefore 12 sub-clusters as well as their preferred grouping in three global clusters. Arrows indicate the affinity between exemplars at the top-level of the nested hierarchy, sub-clusters are color-coded.

replications quantifies the uncertainty in mapping tractogram x_i to cluster k for the same number of clusters across the bootstrap samples. A problem related to estimating the empirical assignment probability is to identify equivalent clusters across partitionings of different datasets. A greedy approach is to search the particular permutation π^{b+1} of cluster labels c^{b+1} of dataset X^{b+1} that maximizes the sum over all cluster assignment probabilities learned from the previous b mappings:

$$\pi^{b+1} = \arg\max_{\pi}\left\{\sum\nolimits_{i \in X^{b+1}} \hat{p}^{b}\left(\pi\left(c^{b+1}\right)\middle|x_i\right)\right\}. \tag{5}$$

The Hungarian method (Kuhn, 1955) finds the permutation π^{b+1} efficiently without having to search through $K!$ possible permutations. More precisely, the problem is formulated in terms of a weighted bipartite matching that contains two sets of nodes with each set containing a permutation of cluster labels (Fischer and Buhmann, 2003). Edges between nodes give the original assignment of label k to the assignment of a label $\pi(k)$ from a permuted set. The weight of each edge is given by:

$$w_{k\pi(k)} = \sum\nolimits_{\left(i \in X^{b+1}; c_i^{b+1} = k\right)} \hat{p}^{b}\left(\pi(k)|x_i\right). \tag{6}$$

Maximizing the sum over all possible weights using the Hungarian method with a running time of $X(K^3)$ is equivalent to solving Eq. 5:

$$\pi^{b+1} = \arg\max_{\pi}\left\{\sum\nolimits_{1 \leq k \leq K} w_{k\pi(k)}\right\}. \tag{7}$$

Finding the optimal cluster relabeling in each of the bootstrap sample allows one to quantify the reliability of clustering tractograms across different data replicates based upon their maximum likelihood given by:

$$\tilde{p} = \frac{1}{|X|}\sum\nolimits_{i \in X} \hat{p}\left(c_i^{*}|x_i\right), \tag{8}$$

where $c_i^* = \arg\max_{1\le k\le K} \hat{p}(k|x_i)$ defines the maximum likelihood mapping. Fischer and Buhmann (2003) propose a stability criterion that compares the reliability of the maximum likelihood mapping with the reliability of making random cluster assignments relative to the risk of misclassification:

$$K^* = \arg\max_K \left\{ \frac{\tilde{p}(K) - \tilde{p}_0(K)}{1 - \tilde{p}_0(K)} \right\}, \tag{9}$$

where $\tilde{p}$ and $\tilde{p}_0$ are the mean of maximum and random probability assignments, respectively. By validating the global partitioning obtained by SCAP one yields a set of exemplars that give rise to the most stable global partitioning. Such a set of exemplars proves useful in identifying the finest level of detail of the hierarchy within a bottom-up approach as introduced in a latter section.

ALLOWING FOR TRANSITIONAL BORDERS BETWEEN CORTICAL AREAS

As mentioned above, previous attempts at tractography-based parcellation have formed hard borders between cortical subunits. However, from an anatomical point of view, it is unclear whether borders between cortical structures should be distinct or have a transitional property. A soft partitioning of the data should therefore be made, in order to account for transitional regions. Such a soft partitioning is induced by means of a stochastic mapping, $p(\bar{x}_q|x_i)$ and $p(\bar{x}_q)$, in order to map tractograms to exemplars as opposed to making hard assignments. The most straightforward approach is to convert dissimilarity measures (i.e., distortion measures) into stochastic mappings using the following equations:

$$\begin{aligned} p\left(\bar{x}_q|x_i\right)^t &= \frac{p\left(\bar{x}_q\right)^t}{A} \exp\left(-\frac{1}{T} d(\bar{x}_q, x_i)\right) \quad \forall \bar{x} \in \bar{X}, \\ p\left(\bar{x}_q\right)^{t+1} &= \frac{1}{|X|} \sum_i p\left(\bar{x}_q|x_i\right)^t \qquad\qquad \forall \bar{x} \in \bar{X}, \end{aligned} \tag{10}$$

where t denotes the iteration sequence and A serves as the normalization constant. Bayes' rule is used in Eq. 10 (top set) with a likelihood function given by $\exp(-\frac{1}{T} d(\bar{x}_q, x_i))$. Note that the likelihood function contains the distortion measure $d(\bar{x}_q, x_i)$ between tractograms and exemplars together with the computational temperature T that sets the scale for converting dissimilarity measures into probabilities. The marginal probability $p(\bar{x}_q)$ in Eq. 10 (bottom) is computed by summing over all conditional probabilities.

However we require that the conditional $p(\bar{x}|x)$ as well as marginal probabilities $p(\bar{x})$ remain consistent (i.e., they do not change with respect to one another). Within an information-based framework the problem can be formulated in terms of rate distortion theory where the conditional entropy and the expected distortion determine the quality of the stochastic mapping (Tishby et al., 1999):

$$\begin{aligned} H(\bar{X}|X) &= -\frac{1}{|X|} \sum_q \sum_i p\left(\bar{x}_q|x_i\right) \log p\left(\bar{x}_q|x_i\right), \\ \langle d(\bar{x}, x)\rangle_{p(\bar{x},x)} &= \frac{1}{|X|} \sum_q \sum_i p\left(\bar{x}_q|x_i\right) d\left(\bar{x}_q, x_i\right). \end{aligned} \tag{11}$$

Variation of information serves as the distortion measure, $d(\bar{x}_q|x_i)$, between tractogram x_i and the exemplar $\bar{x}_q$. Note that conditional entropy characterizes the average information required, in bits per tractogram, to invoke a mapping of a tractogram to an exemplar without confusion (Tishby et al., 1999). Rate distortion theory characterizes the tradeoff between information rate $I(\bar{X}, X) = H(X) - H(\bar{X}|X)$ and expected distortion, where the objective is to allot membership probabilities to tractograms in order to maximize compression (i.e., equivalent to minimizing information rate) under the expected distortion constraint. Finding the rate distortion function is solved by introducing the Lagrange multiplier or inverse temperature, $\beta = 1/T$, and minimizing the corresponding functional:

$$F\left[p\left(\bar{x}|x\right)\right] = I\left(\bar{X}, X\right) + \beta \left\langle d(\bar{x}, x)\right\rangle_{p(\bar{x},x)}. \tag{12}$$

Minimization of the functional yields the set of self-consistent equations (Eq. 10) that are each iterated over convex sets of normalized distributions given by Blahut (1978). More precisely, Blahut (1978) proves that, for a given temperature, iterating over the conditional and marginal probabilities in Eq. 10 yields the global minimum of the functional F in Eq. 12. Note that both conditional and marginal probabilities remain consistent at the global minimum of F.

TESTING HIERARCHICAL ORGANIZATION OF CLUSTERS

As we introduced above, evidence suggests that clusters of anatomical connectivity patterns are organized into a hierarchical structure; whereby bottom-level clusters reveal finer structures and top-level clusters (i.e., coarser clusters reveal a collection of finer structures) within the region of interest. The SCAP approach identified cortical subunits as well as inferred their preferred grouping into a global partitioning. However, from an anatomical point of view the nested structure (i.e., preferred grouping of cortical subunits) might exist at multiple levels thereby constituting a more informative hierarchical structure.

The information that exemplars provide about all other tractograms is given by $I(\bar{X}, X)$. Forming a more informative nested structure of clusters requires merging clusters, z_k and $z_{k'}$, in the partitioning Z_m to form a coarser partitioning Z_{m-1}. However, forming Z_{m-1} from Z_m results in information loss about exemplars [i.e., $I(\bar{X}, X) > I(\bar{X}, Z_{m-1})$]. Intuitively, given the assumption that cortical subunits retain a prototype characteristic, we wish group clusters in such a way that maximal information about exemplars $I(\bar{X}, X)$, i.e. prototypes, is preserved. In order for $I(\bar{X}, Z_{m-1})$ to approximate $I(\bar{X}, X)$ as much as possible, the difference between the loss of information between merging operations should be minimal [i.e., $\min(I(\bar{X}, Z_{m-1}) - I(\bar{X}, Z_m))$ so that $I(\bar{X}, Z_{m-1}) \approx I(\bar{X}, X)$]. Slonim and Tishby (1999) demonstrate that clusters, z_k^* and $z_{k'}^*$, achieve an optimal grouping, which preserves as much information about exemplars as possible [i.e., $I(\bar{X}, Z_{m-1}) \approx I(\bar{X}, X)$], if the Jenson–Shannon distance between their conditional distributions $JS(p(\bar{X}|z_k^*), p(\bar{X}|z_{k'}^*))$, corrected for marginal probabilities, is minimal. More precisely, the clusters that we have to merge is found by minimizing $\delta I(z_k, z_{k'})$:

$$\begin{aligned} \delta I\left(z_k, z_{k'}\right) &= \left(p\left(z_k\right) + p\left(z_{k'}\right)\right) \cdot JS\left(p\left(\bar{X}|z_k\right), p\left(\bar{X}|z_{k'}\right)\right), \\ \left\{z_k^*, z_{k'}^*\right\} &= \arg\min_{k\neq k'} \left(\delta I\left(z_k, z_{k'}\right)\right), \end{aligned} \tag{13}$$

where $\delta I(z_k, z_{k'}) = I(\bar{X}, Z_{m-1}) - I(\bar{X}, Z_m)$. Essentially, the merge cost $\delta I_{\bar{X}}(z_k, z_{k'})$, is a product of the weight sum of clusters, $p(z_k) + p(z_{k'})$, and the distance between them with respect to the exemplars measured by the Jenson–Shannon divergence.

This optimization strategy was formerly introduced as agglomerative information bottleneck method (Slonim and Tishby, 1999) – **Figure 4** illustrates our application.

After merging clusters, $\hat{z} = \{z_k^*, z_{k'}^*\}$, the marginal and conditional probabilities for $\hat{z}$, $p(\hat{z})$, $p(\bar{x}|\hat{z})$, and $p(\hat{z}|x)$, are updated as follows (Slonim and Tishby, 1999):

$$
\begin{aligned}
p\left(\hat{z}\right)^{m-1} &= p\left(z_k^*\right)^m + p\left(z_{k'}^*\right)^m, \\
p\left(\bar{x}|\hat{z}\right)^{m-1} &= \frac{1}{p\left(z'\right)}\left(p\left(\bar{x}, z_k^*\right)^m + p\left(\bar{x}, z_{k'}^*\right)^m\right) \quad \forall \bar{x} \in \bar{X}, \\
p\left(\hat{z}|x\right)^{m-1} &= \left\{ \begin{array}{cc} 1 & \text{if } x \in \left\{z_k^*, z_{k'}^*\right\}^m \\ 0 & \text{otherwise} \end{array} \right\} \quad \forall x \in X,
\end{aligned}
\tag{14}
$$

where m denotes the merging sequence. Note that the sub-clusters obtained by SCAP are used as the initial hard partition between cortical subunits Z_M.

RESULTS

Assessment of clustering solutions in **Figure 5** based upon the stability criterion (Eq. 9) suggests four global clusters consisting of 15 exemplars and thus 15 cortical subunits as the most stable solution within the region of interest.

Figure 6A illustrates the preferred grouping of cortical subunits in the region of interest in four global cortical structures: The PCG is divided into two areas, a dorsal area (dPCG) and a ventral area (vPCG), then a ventral transition into the posterior IFG resides at the ventral tip of the PCG and pars opercularis of the IFG, pars triangularis of the IFG and the deep frontal operculum together form the forth group.

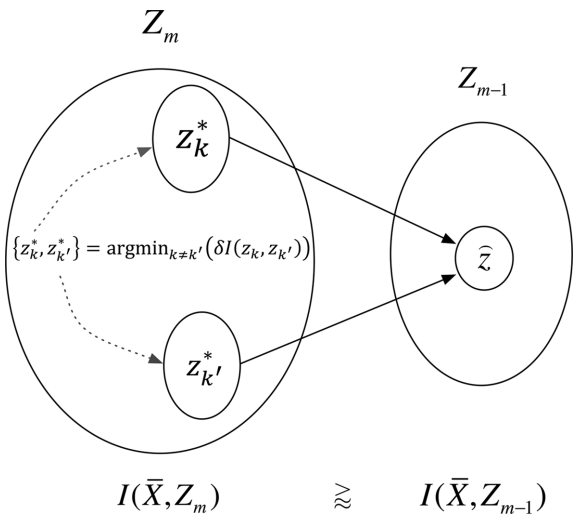


FIGURE 4 | The agglomerative information bottleneck method merges clusters z_k^* and $z_{k'}^*$ in partitioning Z_m to form a coarser partitioning $(z_k, z_{k'})$. The resulting partitioning Z_{m-1} is such that there is minimal loss of information – i.e., $I(\bar{X}, Z_{m-1}) \approx I(\bar{X}, Z_m) \approx I(\bar{X}, X)$.

Exemplars identify 15 areas in the posterior inferior frontal and precentral cortex (cf. **Figure 6B**). The dorsal PCG is subdivided into five areas: two superior–caudal and two inferior–rostral areas, and one directly bordering ventral PCG at the bank of precentral sulcus. The ventral PCG is subdivided into a superior–rostral and an inferior–caudal area. For validation purposes a part of the (inferior) postcentral gyrus was included in the region of interest; this region is appropriately clustered as a separate area (orange field in **Figure 6B**). Parcellation results suggest a transition region into the posterior IFG at the ventral tip of the PCG: The pars opercularis of the IFG, pars triangularis of the IFG, the depths of the inferior frontal sulcus and frontal operculum. The hierarchical organization of these clusters (**Figure 7**) constitutes a distinction of areas in the dorsal PCG and those belonging to the posterior ventral precentral cortex.

For the latter there is further modular organization showing distinction between areas of ventral PCG and those of posterior IFG. The parcellation results given alongside the hierarchical tree in **Figure 7** show the finest detail expressed by cortical structures. Arrows in **Figure 7** illustrate the preferred grouping of cortical subunits into four global cortical structures shown in **Figure 6A**. Notice that the same four global cortical structures emerge from the agglomerative information bottleneck method.

DISCUSSION

We propose an unsupervised information-based clustering technique for connectivity-based cortex parcellation suitable for automatic parcellation. The methodological framework used here to reveal complex properties of cortical subunits such as transitional

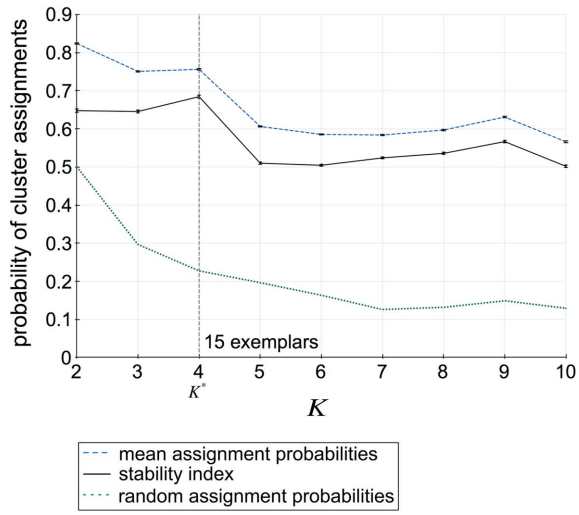


FIGURE 5 | Assignment probabilities plotted against the number of global clusters *K*. Dashed plot shows the mean assignment probability based upon the maximum likelihood mapping. Dotted plot shows the random cluster assignment probabilities. The stability index (Fischer and Buhmann, 2003) used to select the number of clusters K^* is the relative difference between mean and random cluster assignment probabilities (solid plot). The most stable partitioning K^* is given by the preferred grouping of 15 cortical subunits, characterized by 15 exemplars, into four global cortical structures.

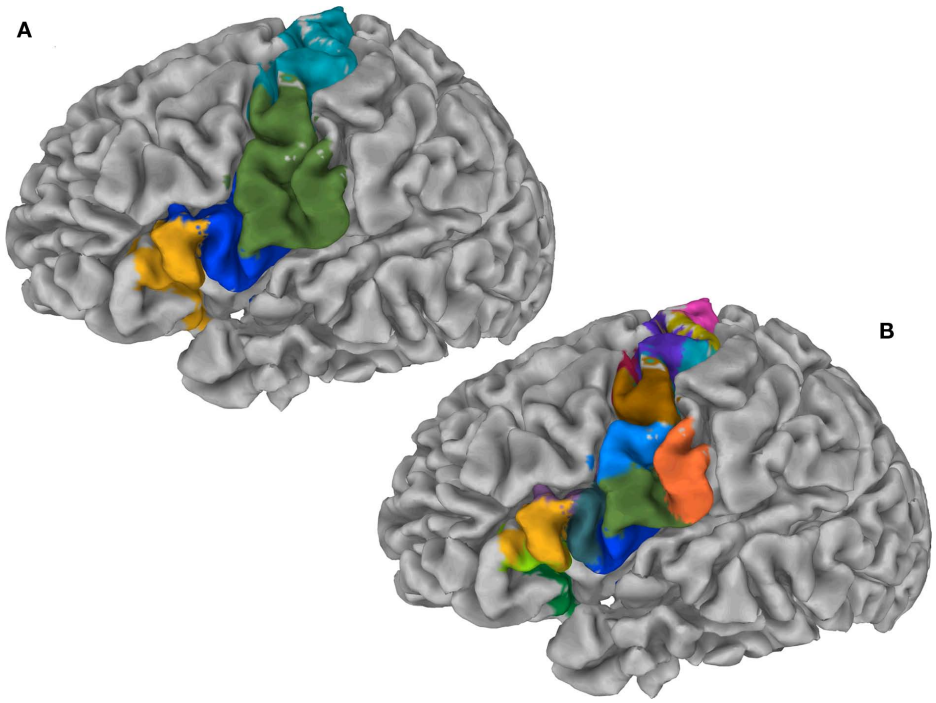


FIGURE 6 | (A) Cortex parcellation of the IFG together with the PCG showing four global cortical structures on the gray matter surface. **(B)** Cortex parcellation of the same ROI at finest level of detail expressed by a hierarchy indicates 15 cortical subunits.

borders as well as a modular hierarchical architecture is summarized in **Figure 8**.

A proof of principle of the approach yields anatomically sensible results.

ANATOMICAL INTERPRETATION

The parcellation results advocate that dorsal PCG fields can be separated in agreement with the suggestion that this area consists of two premotor areas (Schubotz et al., 2010), as well as primary motor cortex (Geyer et al., 1996) and the frontal eye field at the rostral bank of precentral sulcus and the ventral branch of posterior superior frontal sulcus (Amiez and Petrides, 2009).

Concerning the convexity of PCG, the average Talairach z coordinate of the border between ventral and dorsal areas was 49, consistent with other reports from functional imaging studies (Rizzolatti et al., 2002) and connectivity-based parcellation (Tomassini et al., 2007; Schubotz et al., 2010). The delineation of the sub-fields in the posterior ventral precentral cortex accurately resembles results from cytoarchitectonic and multireceptor studies as those were recently reported by Amunts et al. (2010). This includes previously unknown areas, such as the ventral precentral transitional cortex 6r1, anterior and posterior areas 45a and 45p, and areas in the frontal operculum op8 and op9.

Anatomically disjoint areas were distinguished (**Figure 9**), consistent with Amunts et al. (2010), one being located in the depths of the inferior frontal sulcus, the other immediately rostrally to the ventral premotor area. Both areas were found at the junction of the inferior frontal and the precentral sulcus and therefore may correspond to the previously described inferior frontal junction region (IFJ, Brass et al., 2005; Amunts and Von Cramon, 2006). Strikingly, our results accurately reflect the delineation of areas concerning the IFJ obtained by Derrfuss et al. (2009). Note that the fMRI data used by Derrfuss et al. (2009) were taken from the same subject (subject 2 in Derrfuss et al., 2009). Our results therefore suggest a specific connectivity underlying IFJ, rendering this region as a distinct anatomical area.

The merging of the postcentral region (orange field in **Figure 6B**) with the ventral PCG at a rather high hierarchical level seems to be supported by findings in non-human primates, implying dense bidirectional connections between the rostral portion of the inferior parietal lobule and the adjacent opercular area, i.e., ventral premotor area 6 (cf., e.g., Schmahmann and Pandya, 2007). However, whether this suggestion is indeed evident in tractography-based connectivity scores remains to be studied in detail, and specifically with respect to limitations potentially imposed by the choice of a particular tractography method and the underlying diffusion model.

METHODOLOGICAL ISSUES

A well-known difficulty of most clustering algorithms is the choice of an appropriate similarity measure, since this ultimately determines the cluster structure that can be inferred from the data – i.e., elements within the same cluster share a common similarity quantified by the respective measure. Intuitively, clustering of tractograms should be based upon capturing the shape of probabilistic tractograms. In other words, probabilistic tractograms should be grouped together if they have similar shape. Such tractograms are represented as volumes containing connectivity scores for each

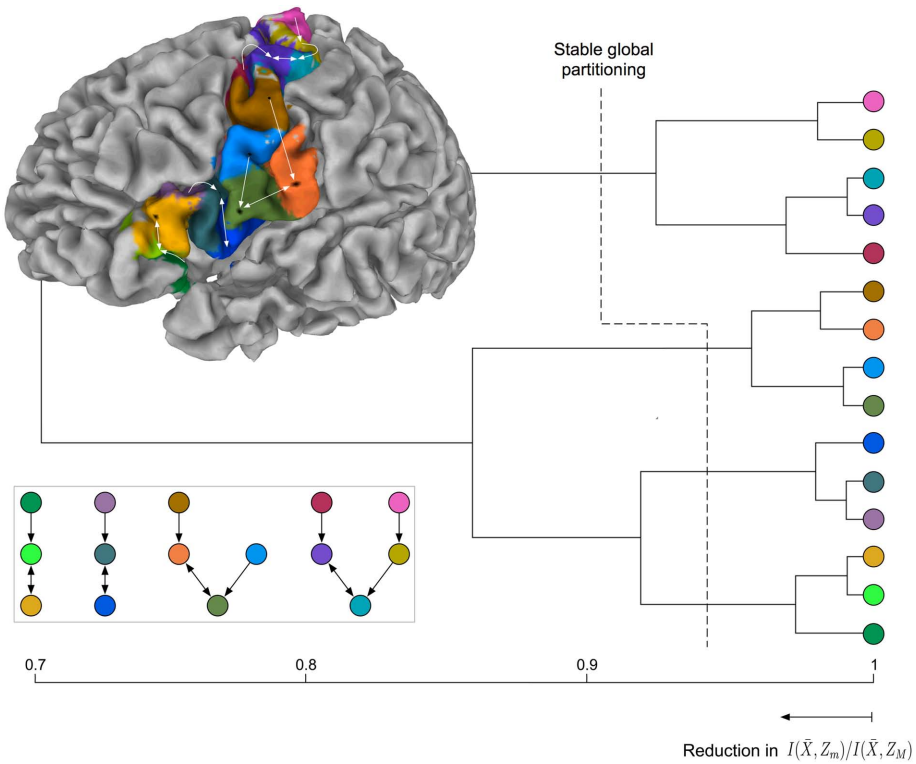


FIGURE 7 | Hierarchical tree of the nested hierarchical structure constructed by the agglomerative information bottleneck method (Slonim and Tishby, 1999). The cardinality of the initial partition equals the number of exemplars, $M = 15$. The length of each branch gives the normalized information loss $I(\bar{X}, Z_m)/I(\bar{X}, Z_M)$ due to merging operations. Bottom left: Internal organization of representative tractograms within four global cortical structures obtained by SCAP. Top left: Cortex parcellation showing the finest level of detail. Black dots indicate the location of exemplars while arrows show the preferred grouping of cortical subunits obtained by SCAP.

target voxel. Defining their shape is therefore not straightforward. We define two tractograms as having similar shape if their connectivity scores in *common* (i.e., *corresponding*) target voxels (not any target voxel) are similar. The similarity measure should therefore involve a *pairwise* comparison of connectivity scores with pairs of connectivity scores given by $\{y^a_{(i)}, y^a_{(j)}\}$, where a denotes the particular target voxel *common* to both tractograms i and j.

In order to compute how probable it is that two tractograms have similar shape we consider the joint occurrence of connectivity scores $p(y_{(i)},y_{(j)})$ for any *common* target voxel; $p(y_{(i)},y_{(j)})$ is computed by constructing the frequency table using the frequency of occurrence of pairs of connectivity scores $\{y^a_{(i)}, y^a_{(j)}\}$ within all *common* target voxels. Constructing the frequency table as shown in **Figure 2** already involves pairwise comparisons of connectivity scores in *common* target voxels. If $p(y_{(i)},y_{(j)})$ is high for all connectivity scores among common target voxels in tractograms i and j it follows that tractograms i and j are likely to have similar shape. Mutual information measures the dependency of one tractogram on another tractogram. Since we are interested in capturing the shape of tractograms, mutual information, as computed in Eq. 2, measures how dependent the shape of one tractogram is on the shape of another tractogram. Moreover, mutual information will capture any type of dependency including linear and non-linear dependencies between the shapes of tractograms.

An issue that draws less attention is the dependency of similarity measures on the representation of the data – i.e., different transformations of the data will produce different similarity quantities. Given that, in our application, we have limited knowledge about the structure of clusters or about which type of relation should be considered, the similarity measure should be invariant to data representation. Mutual information has the useful property of being independent of representation of the data – i.e., different invertible transformations on individual tractograms will yield the same mutual information quantity.

Another difficulty of clustering algorithms is their associated degree of freedom that influences the partitioning, mostly with regard to the number of clusters. Typically used cluster validity criteria are determined heuristically and favor compactness and separability of clusters. The method proposed in this study to infer the number of global clusters suggests an intuitive notion of a sensible partitioning. That is, a stable clustering solution should be resistant to noise in the data (Buhmann, 2010). More precisely, uncertainty in the data gives rise to uncertainty in the clustering solution. A sensible global partitioning is one for which the uncertainty in the data has minimum influence on the clustering solution. Note that the stability criterion (cf. Eq. 12) is dependent upon the ordering of the bootstrap samples, particularly if the first sample leads to a poor clustering solution. To circumvent this

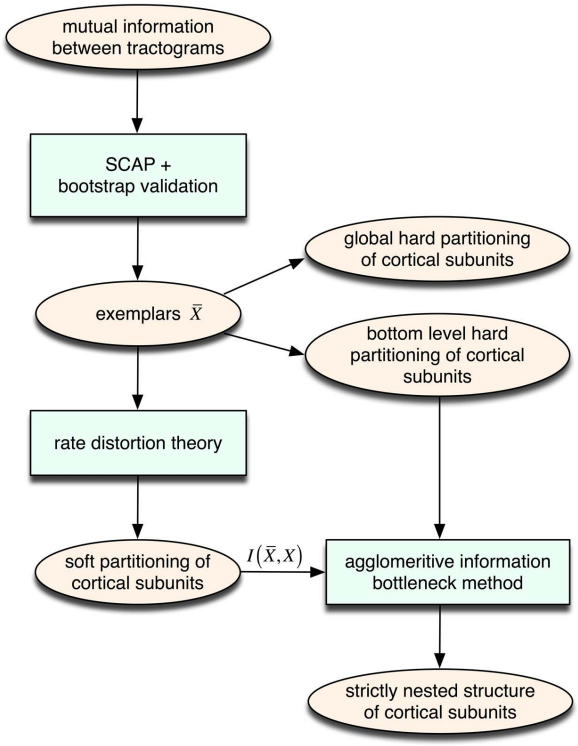


FIGURE 8 | Overview of the unsupervised framework used in this study. SCAP together with clustering assessment using bootstrap sampling is used to obtain exemplars that each characterizes individual cortical subunits within a global and bottom-level partitioning. Thereafter, rate distortion theory is used to map tractograms to exemplars. The agglomerative information bottleneck method used information gained from rate distortion theory [i.e., $I(\bar{X}, X)$] to construct a more informative nested structure of cortical subunits. The sub-clusters obtained by SCAP define the bottom-level partitioning of the hierarchy.

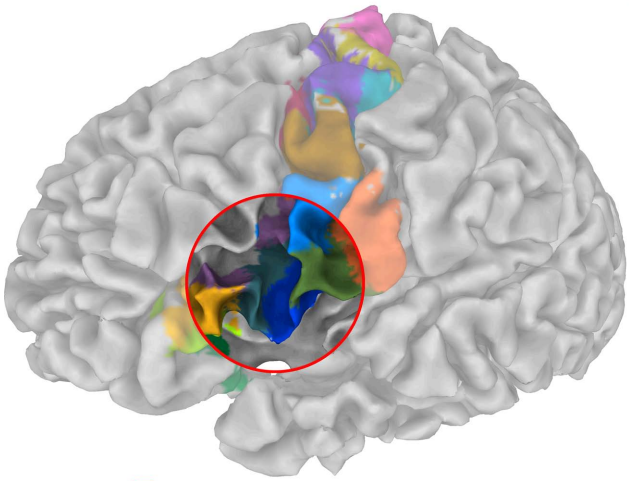


FIGURE 9 | Parcellation of the sub-fields in the posterior ventral precentral cortex indicate a striking resemblance with cytoarchitectonic and multireceptor studies as those were recently reported by Amunts et al. (2010) – cf. their Figure 9.

potential difficulty, we suggest computing the stability criterion for different permutations of bootstrap sample orderings in order to avoid the influence of initial poor solutions on the stability criterion.

Essentially, our method rests upon the assumption that each cortical subunit possesses a "prototype" characteristic: an exemplar tractogram sufficiently describes the connectivity pattern of the entire cortical subunit and is thus representative of that subunit. Among the previously used clustering algorithms for connectivity-based parcellation, *k*-means clustering and Dirichlet process mixture models (used together with a Gaussian likelihood), have the advantage of searching for central tendencies (i.e., means) in data which are useful for identifying exemplar tractograms. Such techniques, however, rely on random sampling that allow for unlucky pruning decisions that cannot be recovered from and consequently lead to poor solutions. Affinity propagation (Frey and Dueck, 2007) used in this study is a recently developed algorithm that avoids random sampling by simultaneously considering all tractograms as potential exemplars.

Our clustering approach is inherently different to previous techniques since global clusters of tractograms define groups of cortical subunits whereas the sub-clusters form the cortical subunits themselves. Given the prior assumption that a strictly nested structure may exist, our approach associates the partitioning of cortical subunits reflecting the finest level of detail with the partitioning that brings about the most stable grouping of cortical subunits at a higher level of the hierarchy. The agglomerative information bottleneck method (Slonim and Tishby, 1999) provides more levels of clustering on the basis of preserving as much information as possible about the partitioning with respect to representative tractograms. In contrast to other traditional hierarchical clustering methods we insist upon defining the finest level of detail of the partitioning other than simply associating it with the maximum number of clusters (i.e., every tractogram is its own cluster). While, anatomically, very fine detailed hierarchical organization of cortical subunits may exist, the level of detail that probabilistic tractography is capable of revealing is, among other things, limited by resolution offered by diffusion-weighted images.

The prototype-based characteristic of cortical subunits allows for modeling the transition between cortical subunits as the uncertainty in mapping tractograms to cortical subunits using rate distortion theory. A simple parameter, namely the "temperature" T, controls the sensitivity of the uncertainty in the aforementioned mapping to the similarity between tractograms and representative tractograms. Tuning T therefore determines the level of fuzziness in the partitioning (see **Figure 10**).

Note that T influences the hierarchical structure obtained by the agglomerative information bottleneck method since it determines the dependency of the data on the exemplars. Tuning the temperature to infinity maximizes the uncertainty of mapping tractograms to exemplars and causes conditional probabilities to be insensitive to similarity between tractograms. The resulting hierarchy is thus constructed at random. Conversely, decreasing the temperature results in a hierarchy that is more governed by the dependency of tractograms to exemplars. We propose to select the

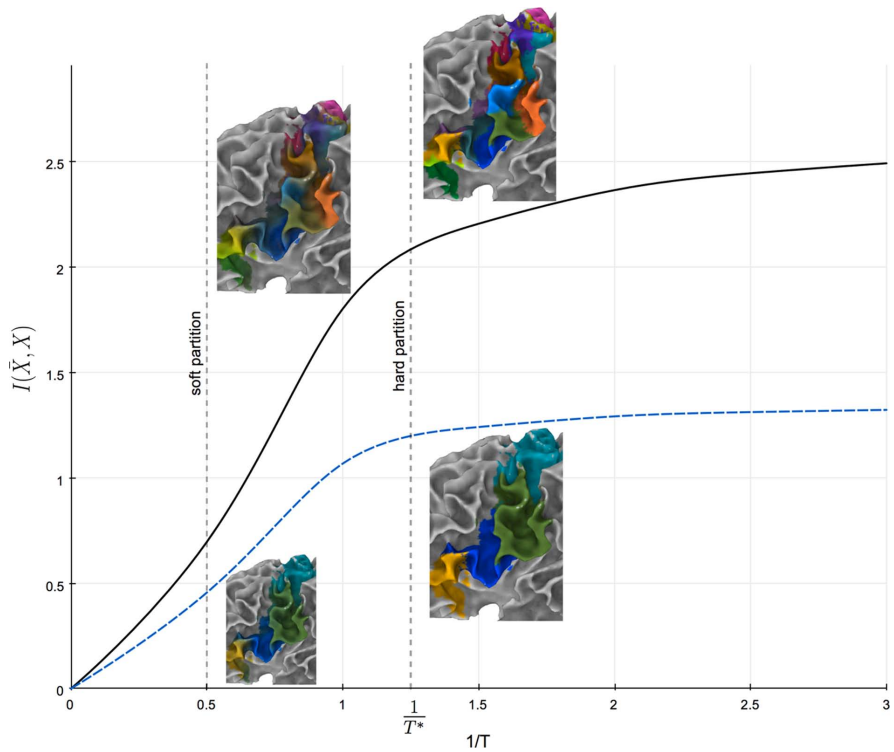


FIGURE 10 | Information rate $I(\bar{X},X)$ plotted against the inverse temperature $1/T$. Solid and dashed plots illustrate the relationship between information rate and inverse temperature for 15 cortical subunits and 4 global cortical structures, respectively. $1/T^*$ gives the temperature for which the information rate has only slight changes for lower temperatures.

highest possible "temperature" for which the information rate has minimal change for lower temperatures as shown in **Figure 10**. Maximal amount of information is thus used to construct the hierarchy while taking into account some uncertainty in mapping tractograms to exemplars.

FUTURE WORK

A further step towards understanding the organization of cortical subunits is to study the consistency or heterogeneity of hierarchically modular cortical subunits across subjects. Individual variability, however, is an important issue in anatomical studies, because any given area (even a primary sensory area) can vary in size by twofold or more (Filiminoff, 1932; Maunsell and Van Essen, 1987; Uylings et al., 2005) and because the consistency with which each area is located with respect to topographic boundaries has important implications for physiological and neuroimaging studies. In this respect, a meaningful parcellation should be assumed to exist in all subjects with similar location, shape, and connectivity pattern.

Note that it is unclear whether or not cortical subunits possess the aforementioned "prototype" (i.e., exemplar) characteristic. A clustering method that avoids defining cluster "prototypes" might therefore prove more suitable for parcellating regions without *a priori* knowledge (Slonim et al., 2005).

Furthermore, an obvious limitation of any model-based approach to reveal structure in the data is that it already assumes that there is structure in the data. In the context of hierarchically modular cortical subunits, hierarchical clustering models are already forced to find nested structures in the data. The question of whether connectivity patterns, quantified by probabilistic tractograms, prefer to be grouped in a nested structure or not should be addressed and is formulated in terms of model validation.

Buhmann (2010) performs model validation based upon an indispensible requirement that the solution should be generalizable under the influence of noise. More precisely, a nested structure of cortical subunits should be generalizable given noise in the diffusion measurements, such as noise from the MR scan, physiological noise, motion affects, etc. Assessing the generalizability performance is given by a tradeoff between the informativeness as well as the robustness against noise of the nested clustering solution. Intuitively, a nested structure showing more branching is more informative. The question is to be answered is therefore: How informative can the nested structure of cortical subunits be (i.e., how many branches if any should the dendrogram have) without fitting the sampling noise?

ACKNOWLEDGMENTS

The authors wish to thank D. Y. von Cramon for his continuing support and his endurance to teach us tedious anatomy. Further appreciation for insightful discussions is given to J. M. Buhmann, Swiss Federal Institute of Technology Zurich, and M. Lässig, University of Cologne. The work was supported by the German Ministry of Education and Research (Grants No. 01GW0772 and 01GW0612).

REFERENCES

Amiez, C., and Petrides, M. (2009). Anatomical organization of the eye fields in the human and non-human primate frontal cortex. *Prog. Neurobiol.* 89, 220–230.

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., and Zilles, K. (2010). Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol.* 8, e1000489. doi: 10.1371/journal.pbio.1000489

Amunts, K., and Von Cramon, D. Y. (2006). The anatomical segregation of the frontal cortex: what does it mean for function? *Cortex* 42, 525–528.

Anwander, A., Tittgemeyer, M., Yves Von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of Broca's area. *Cereb. Cortex* 17, 816–825.

Averbeck, B. B., Battaglia-Mayer, A., Guglielmo, C., and Caminiti, R. (2009). Statistical analysis of parieto-frontal cognitive-motor networks. *J. Neurophysiol.* 102, 1911–1920.

Bach, D., Behren, S. T., Garrido, L., Weiskopf, N., and Dolan, R. (2011). Deep and superficial amygdala nuclei projections revealed in vivo by probabilistic tractography. *J. Neurosci.* 31, 618–623.

Bassett, D. S., and Gazzaniga, M. S. (2011). Understanding complexity in the human brain. *Trends Cogn. Sci. (Regul. Ed.)* 15, 200–209.

Bassett, D. S., Greenfield, D. L., Meyer-Lindenberg, A., Weinberger, D. R., Moore, S. W., and Bullmore, E. T. (2010). Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. *PLoS Comput. Biol.* 6, e1000748. doi: 10.1371/journal.pcbi.1000748

Beckmann, M., Johansen-Berg, H., and Rushworth, M. F. S. (2009). Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. *J. Neurosci.* 29, 1175–1190.

Behrens, T. E. J., Johansen-Berg, H. J., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. a. M., Boulby, P. A., Barker, G. J., Sillery, E. L., Sheehan, K., Ciccarelli, O., Thompson, A. J., Brady, M., and Matthews, P. M. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nat. Neurosci.* 6, 750–757.

Blahut, R. E. (1978). Computation of channel capacity and rate distortion theory. *IEEE Trans. Inf. Theory* 18, 460–473.

Brass, M., Derrfuss, J., Forstmann, B. U., and Yves Von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends Cogn. Sci. (Regul. Ed.)* 9, 314–316.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellaufbaues.* Leipzig: Barth.

Buhmann, J. M. (2010). "Information theoretic model validation for clustering," in *International Symposium on Information Theory* (Austin, TX: IEEE).

Crippa, A., Cerliani, L., Nanetti, L., and Roerdink, J. B. T. M. (2011). Heuristics for connectivity-based brain parcellation of SMA/pre-SMA through force-directed graph layout. *Neuroimage* 54, 2176–2184.

Dale, A. M., Fischl, B., and Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9, 179–194.

Derrfuss, J., Brass, M., Yves Von Cramon, D. Y., Lohmann, G., and Amunts, K. (2009). Neural activations at the junction of the inferior frontal sulcus and the inferior precentral sulcus: interindividual variability, reliability, and association with sulcal morphology. *Hum. Brain Mapp.* 30, 299–311.

Devlin, J. T., Sillery, E. L., Hall, D. A., Hobden, P., Behrens, T. E. J., Nunes, R. G., Clare, S., Matthews, P. M., Moore, D. R., and Johansen-Berg, H. (2006). Reliable identification of the auditory thalamus using multimodal structural analyses. *Neuroimage* 30, 1112–1120.

Draganski, B., Kherif, F., Kloppel, S., Cook, P. A., Alexander, D. C., Parker, G. J. M., Deichmann, R., Ashburner, J., and Frackowiak, R. S. J. (2008). Evidence for segregated and integrative connectivity patterns in the human basal ganglia. *J. Neurosci.* 28, 7143–7152.

Filiminoff, I. (1932). Über die variabilitat der grosshirnrindenstruktur. Mitteilung II. Regio occipitalis beim erwachsenen Menshchen. *J. Psychol. Neurol.* 44, 1–96.

Fischer, B., and Buhmann, J. M. (2003). Bagging for path-based clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 1411–1415.

Ford, A., Mcgregor, K. M., Case, K., Crosson, B., and White, K. D. (2010). Structural connectivity of Broca's area and medial frontal cortex. *Neuroimage* 52, 1230–1237.

Frey, B. J., and Dueck, D. (2007). Clustering by passing messages between data points. *Science* 315, 972–976.

Fuster, J. M. (1997). *The Prefrontal Cortex: Anatomy, Physiology and Neuropsychology of the Frontal Lobe.* Baltimore: Lippincott/Williams & Wilkins.

Geyer, S., Ledberg, A., Schleicher, A., Kinomura, S., Schormann, T., Bürgel, U., Klingberg, T., Larsson, J., Zilles, K., and Roland, P. E. (1996). Two different areas within the primary motor cortex of man. *Nature* 382, 805–807.

Geyer, S., Matelli, M., Luppino, G., and Zilles, K. (2000). Functional neuroanatomy of the primate isocortical motor system. *Anat. Embryol.* 202, 443–474.

Hilgetag, C. C., and Grant, S. (2000). Uniformity, specificity and variability of corticocortical connectivity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 7–20.

Jbabdi, S., Woolrich, M. W., and Behrens, T. E. J. (2009). Multiple-subjects connectivity-based parcellation using hierarchical Dirichlet process mixture models. *Neuroimage* 44, 373–384.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. M. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841.

Johansen-Berg, H., Behrens, T., Sillery, E., Ciccarelli, O., Thompson, A., Smith, S., and Matthews, P. (2005). Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. *Cereb. Cortex* 15, 31–39.

Johansen-Berg, H. J., Behrens, T. E. J., Robson, M. D., Drobnjak, I., Rushworth, M. F. S., Brady, M., Smith, S. M., Higham, D. J., and Matthews, P. M. (2004). Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 101, 13335–13340.

Johansen-Berg, H. J., and Rushworth, M. (2009). Using diffusion imaging to study human connectional anatomy. *Annu. Rev. Neurosci.* 32, 75–94.

Jones, D. (2010). Challenges and limitations of quantifying brain connectivity in vivo with diffusion MRI. *Imaging Med.* 2, 341–355.

Kaiser, M. (2011). A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage* 57, 892—907.

Kaiser, M., and Hilgetag, C. C. (2007). Development of multi-cluster cortical networks by time windows for spatial growth. *Neurocomputing* 70, 1829–1832.

Kaiser, M., and Hilgetag, C. C. (2010). Optimal hierarchical modular topologies for producing limited sustained activation of neural networks. *Front. Neuroinform.* 4:8. doi: 10.3389/fninf.2010.00008

Klein, J. C., Behrens, T. E. J., Robson, M. D., Mackay, C. E., Higham, D. J., and Johansen-Berg, H. J. (2007). Connectivity-based parcellation of human cortex using diffusion MRI: establishing reproducibility, validity and observer independence in BA 44/45 and SMA/pre-SMA. *Neuroimage* 34, 204–211.

Knösche, T. R., and Tittgemeyer, M. (2011). The role of long-range connectivity for the characterization of the functional-anatomical organization of the cortex. *Front. Syst. Neurosci.* 5:58. doi: 10.3389 fnsys.2011.00058. [Epub ahead of print].

Kuhn, H. (1955). The Hungarian method for the assignment problem. *Naval Res Logist. Q.* 2, 83–97.

Lehericy, S., Ducros, M., Van De Moortele, P.-F., Francois, C., Thivard, L., Poupon, C., Swindale, N. V., Ugurbil, K., and Kim, D. (2004). Diffusion tensor fiber tracking shows distinct Corticostriatal circuits in humans. *Ann. Neurol.* 55, 522–529.

Leone, M., Sumedha, and Weigt, M. (2008). Unsupervised and semi-supervised clustering by message passing: soft-constraint affinity propagation. *Eur. Phys. J. B* 66, 125–135.

Lewis, J. W., and Van Essen, D. C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 428, 112–137.

Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M. A., Giroud, P., Ercsey-Ravasz, M., Pilaz, L. J., Huissoud, C., Barone, P., Dehay, C., Toroczkai, Z., Van Essen, D. C., Kennedy, H., and Knoblauch, K. (2010). Weight consistency specifies regularities of macaque cortical networks. *Cereb. Cortex* 21, 1254–1272.

Maunsell, J. H., and Van Essen, D. C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. *J. Comp. Neurol.* 266, 535–555.

Menke, R. A., Jbabdi, S., Miller, K. L., Matthews, P. M., and Zarei, M. (2010). Connectivity-based segmentation of the substantia nigra in human and its implications in Parkinson's disease. *Neuroimage* 52, 1175–1180.

Meunier, D., Lambiotte, R., and Bullmore, E. T. (2010). Modular and hierarchically modular organization of brain networks. *Front. Neurosci.* 4:200. doi: 10.3389/fnins.2010.00200

Nanetti, L., Cerliani, L., Gazzola, V., Renken, R., and Keysers, C. (2009). Group analyses of connectivity-based cortical parcellation using repeated k-means clustering. *Neuroimage* 47, 1666–1677.

O'Muircheartaigh, J., Vollmar, C., Traynor, C., Barker, G., Kumari, V., Symms, M., Thompson, P., Duncan, J., Koepp, M., and Richardson, M. (2011). Clustering probabilistic tractograms using independent component analysis applied to the thalamus. *Neuroimage* 54, 2020–2032.

Passingham, R. E. (1983). *The Frontal Lobes and Voluntary Action.* Oxford: Oxford University Press.

Passingham, R. E., Stephan, K. E., and Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci.* 3, 606–616.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Curr. Opin. Neurobiol.* 12, 149–154.

Roca, P., Tucholka, A., Rivière, D., Guevara, P., Poupon, C., and Mangin, J.-F. (2010). "Inter-subject connectivity-based parcellation of a patch of cerebral cortex," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (New York, NY: Springer-Verlag).

Sanides, F. (1962). Architectonics of the human frontal lobe of the brain. With a demonstration of the principles of its formation as a reflection of phylogenetic differentiation of the cerebral cortex. *Monogr. Gesamtgeb. Neurol. Psychiatr.* 98, 1–201.

Schmahmann, J. D., and Pandya, D. N. (2007). *Fiber Pathways of the Brain.* New York: Oxford University Press.

Schubotz, R. I., Anwander, A., Knösche, T. R., Von Cramon, D. Y., and Tittgemeyer, M. (2010). Anatomical and functional parcellation of the human lateral premotor cortex. *Neuroimage* 50, 396–408.

Sillery, E., Bittar, R. G., Robson, M. D., Behrens, T. E. J., Stein, J., Aziz, T. Z., and Johansen-Berg, H. (2005). Connectivity of the human periventricular-periaqueductal gray region. *J. Neurosurg.* 103, 1030–1034.

Slonim, N., Atwal, G. S., Tkacik, G., and Bialek, W. (2005). Information-based clustering. *Proc. Natl. Acad. Sci. U.S.A.* 102, 18297–18302.

Slonim, N., and Tishby, N. (1999). "Agglomerative information bottleneck," in *Advances in Neural Information Processing Systems* (MIT Press) 12, 617–623.

Stephan, K. E., Hilgetag, C. C., Burns, G. A., O'Neill, M. A., Young, M. P., and Kötter, R. (2000). Computational analysis of functional connectivity between areas of primate cerebral cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 111–126.

Tishby, N., Pereira, F., and Bialek, W. (1999). "The information bottleneck method," in *Proceedings of the 37th Annual Allerton Conference on Communication, Control, and Computing*, eds B. Hajek and R. C. Sreenivas, Urbana, IL, 368–377.

Tomassini, V., Jbabdi, S., Klein, J. C., Behrens, T. E. J., Pozzilli, C., Matthews, P. M., Rushworth, M. F. S., and Johansen-Berg, H. J. (2007). Diffusion-weighted imaging tractography-based parcellation of the human lateral premotor cortex identifies dorsal and ventral subregions with anatomical and functional specializations. *J. Neurosci.* 27, 10259–10269.

Uylings, H., Rajkowska, G., Sanz-Arigita, E., Amunts, K., and Zilles, K. (2005). Consequences of large interindividual variability for human brain atlases: converging macroscopical imaging and microscopical neuroanatomy. *Anat. Embryol.* 210, 423–431.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 March 2011; accepted: 26 August 2011; published online: 23 September 2011.

Citation: Gorbach NS, Schütte C, Melzer C, Goldau M, Sujazow O, Jitsev J, Douglas T and Tittgemeyer M (2011) Hierarchical information-based clustering for connectivity-based cortex parcellation. Front. Neuroinform. ***5****:18. doi: 10.3389/fninf.2011.00018*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 22 July 2011
doi: 10.3389/fninf.2011.00010

Integrating temporal and spatial scales: human structural network motifs across age and region of interest size

Christoph Echtermeyer[1,2], Cheol E. Han[3], Anna Rotarska-Jagiela[4], Harald Mohr[5], Peter J. Uhlhaas[4] and Marcus Kaiser[2,3,6]*

[1] *Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky University, Oldenburg, Germany*
[2] *School of Computing Science, Newcastle University, Newcastle-upon-Tyne, UK*
[3] *Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Republic of Korea*
[4] *Department of Neurophysiology, Max-Planck Institute of Brain Research, Frankfurt a. M., Germany*
[5] *Goethe University, Institute of Psychology, Frankfurt a. M., Germany*
[6] *Institute of Neuroscience, The Medical School, Newcastle University, Newcastle-upon-Tyne, UK*

Edited by:
Olaf Sporns, Indiana University, USA

Reviewed by:
Alex Fornito, University of Melbourne, Australia
Lucas Antiqueira, University of Sao Paulo, Brazil

****Correspondence:***
Marcus Kaiser, School of Computing Science, Newcastle University, Claremont Tower, Newcastle-upon-Tyne NE1 7RU, UK.
e-mail: m.kaiser@newcastle.ac.uk

Human brain networks can be characterized at different temporal or spatial scales given by the age of the subject or the spatial resolution of the neuroimaging method. Integration of data across scales can only be successful if the combined networks show a similar architecture. One way to compare networks is to look at spatial features, based on fiber length, and topological features of individual nodes where outlier nodes form single node motifs whose frequency yields a fingerprint of the network. Here, we observe how characteristic single node motifs change over age (12–23 years) and network size (414, 813, and 1615 nodes) for diffusion tensor imaging structural connectivity in healthy human subjects. First, we find the number and diversity of motifs in a network to be strongly correlated. Second, comparing different scales, the number and diversity of motifs varied across the temporal (subject age) and spatial (network resolution) scale: certain motifs might only occur at one spatial scale or for a certain age range. Third, regions of interest which show one motif at a lower resolution may show a range of motifs at a higher resolution which may or may not include the original motif at the lower resolution. Therefore, both the type and localization of motifs differ for different spatial resolutions. Our results also indicate that spatial resolution has a higher effect on topological measures whereas spatial measures, based on fiber lengths, remain more comparable between resolutions. Therefore, spatial resolution is crucial when comparing characteristic node fingerprints given by topological and spatial network features. As node motifs are based on topological and spatial properties of brain connectivity networks, these conclusions are also relevant to other studies using connectome analysis.

Keywords: network analysis, network motifs, structural connectivity, human

1 INTRODUCTION

The set of connections in the brain can be described as the connectome (Sporns et al., 2005). Connectome data is currently becoming available at different levels of the structural organization: from neuronal networks of connections between neurons to fiber-tract networks between brain regions. These data promise to give valuable new insights, but analyses integrating different data sets are challenging. One problem of data integration across labs is that the raw data might show different voxel resolutions due to different magnetic field strengths of the magnetic resonance imaging (MRI) scanner. In addition, different labs might use different parcelation routines or brain atlases. This leads to different spatial resolutions and consequently to different structural connectivity networks (eventually with different numbers of nodes) (Bassett et al., 2010; Fornito et al., 2010; Hayasaka and Laurienti, 2010; Zalesky et al., 2010). Another problem of data integration is caused by age differences between subjects. There is a certain evidence that characteristics of the connectome vary over age (Fair et al., 2009; Uhlhaas et al., 2009; Fan et al., 2010; Hagmann et al., 2010). Given a network analysis result of two studies, say concerning small-world features (Watts and Strogatz, 1998), are differences due to the study population's differences in connectivity or are they due to differences in age or network resolution? In this article, we observe how topological and spatial network features change across age and network resolution for structural connectivity, based on diffusion tensor imaging (DTI), in healthy human subjects. Using single node motifs, we also apply a network mapping technique, which yields a compact and easily comparable representation of complex network structures.

1.1 SCALES OF A HIERARCHICAL ORGANIZATION

Hierarchy is a central feature in the organization of complex biological systems and particularly the structure and function of neural networks (Kaiser and Hilgetag, 2010; Kaiser et al., 2010). In neural networks, the term hierarchy may be understood in several different ways and can apply to topological, spatial, temporal as well as functional properties. These features can be observed at different scales or resolutions.

1.1.1 Temporal scales

Network analysis is often presented as the analysis of a network at a given time, but many networks change both over time and with temporal resolution. Structural connectivity, given by anatomical

connections between neurons or brain regions, may change due to activity-dependent plasticity (Butz et al., 2009). Functional connectivity, given by correlations in the activity patterns of network nodes, can change for different tasks for a subject or due to different system states (e.g., awake vs. asleep). In our case, connectivity might also change over longer time-scales, as part of brain development.

Recovered networks can appear differently depending on temporal resolution. At a fine-grained temporal resolution of a few hundred milliseconds, cortical tissue can show distinct oscillations, e.g., in the α, γ, or δ frequency bands (Roopun et al., 2008). For more coarse-grained time-resolutions with sample lengths of several seconds or minutes such distinct frequency peaks disappear and the frequency distribution displays a $1/f$ power-law behavior instead (Buzsaki, 2006).

Here, we observe temporal scale in terms of snapshots of structural connectivity at different points in time. The human brain undergoes large structural developments until the age of around 20–25 years. These developments coincide with large functional changes during child development and during the teenager years. They further coincide with changes in functional connectivity (Uhlhaas et al., 2009). Such changes during brain network development are not only of interest for the study of healthy subjects, but also for many psychiatric disorders like schizophrenia (Uhlhaas and Singer, 2010), which have an onset around the time when the brain network matures (age 18–25 years). We therefore test the potential of single node motifs to integrate information across temporal network "snapshots".

1.1.2 Spatial scales

The network structure not only depends on the time of a "snapshot" but also on the spatial resolution. One can distinguish the micro- and macro-connectome as the connectivity between individual neurons and connections between brain regions, respectively (DeFelipe, 2010). In this study of structural connectivity we focus on the macro-connectome level based on MRI data [see Seung (2009) for more information about neuronal connectivity instead].

The macro-connectome can be analyzed at different levels of spatial resolution and thus different levels of brain network organization. With a low spatial resolution, it corresponds to connectivity between brain areas where network nodes correspond to brain regions. The resulting networks show features of modular and small-world networks (Hilgetag et al., 2000; Sporns et al., 2000) and typically consist of up to 100 nodes per cortical hemisphere for the primate brain. At the mesoscale, connectivity between regions of interest (ROIs) of the same size, e.g., 1 cm^2 cortical surface area, can be studied. Such networks consist of around 1000 nodes for the human brain (Hagmann et al., 2008). With even higher resolution, the microscale of the macro-connectome could be studied, i.e., connections between cortical columns (Mountcastle, 1997). However, such networks with nodes representing one cortical column each would consist of 1,250,000 nodes for the human brain (both hemispheres, not including subcortical structures) and identifying columns with diameters ranging from 200 μm to 1 mm (Hubel et al., 1977; Horton and Adams, 2005) in humans is (currently) beyond the reach of standard MRI. Such extremely high resolutions might only be achieved with higher magnetic fields or extremely high magnetic fields in post mortem studies.

We therefore study the global and regional scale of the macro-connectome. Importantly, this level allows us to assess whether motifs which are characteristic for a certain cortical region are also characteristic for all parts of that cortical area or whether a region displays a variety of motifs at higher spatial resolutions. Also, changes in spatial resolution affect spatial measurements, which form part of the motif detection. It is often interesting to observe how many connections go to nearby targets and how many extend over a long distance, potentially linking different components of the neural network. This can be readily observed using a histogram of the connection lengths of a network. It is known that the probability that two neurons are connected decays almost exponentially with distance (Hellwig, 2000; Schüz et al., 2005; Kaiser et al., 2009): connections over a long distance are less likely than short-distance connections. In different organisms, ranging from neuronal connectivity in *C. elegans* and layers in the rat visual cortex to fiber-tract connectivity in the macaque, the actual distribution was approximated best by a Gamma probability density function (Kaiser et al., 2009). Looking at different spatial resolutions within humans, we would expect that the proportion of short-distance connections relative to long-distance connections increases when the spatial resolution is increased. We will therefore observe how the length distributions change over different spatial scales.

1.2 SINGLE NODE MOTIFS

Networks can be analyzed following different concepts: to investigate a single network, for example, analysis can focus on its nodes, comparing them through *connectional fingerprints* based on the absolute values of node properties (e.g., node degree or local clustering coefficient; Passingham et al., 2002). On a more abstract level, nodes of brain networks can be regarded with respect to their different organization levels. These approaches look at one network of a particular subject, but networks can also be compared across different subjects – or different networks might exist for the same subject across time. However, mapping between networks becomes difficult if the number of nodes and edges varies, such that direct comparisons (i.e., link-to-link or node-to-node) between structures are ruled out. In contrast to such methods of cross-comparison, the approach used in this paper is less sensitive to these changes: although a network with a slightly higher number of nodes could show additional motifs, most nodes are non-outliers making such changes less likely. Similarly, a network with slightly more edges might show absolute changes in network features without affecting the outliers of the feature distribution leading to similar single node motifs. Again, these comparisons of the number and classes of motifs become possible for non-identical numbers of nodes or edges but only if the numbers do not differ too much. These motif-classes are significantly different from each other and they allow to classify network nodes at the same level even if they appear in different networks.

The concept of *node motifs* – a combination of local network features – complements that of *network motifs*, which are specific connectivity patterns that have been used to characterize networks before (Milo et al., 2002). An example for a node motif are highly connected nodes or hubs that affect spreading phenomena and can be important components of a network. More complex, multi-dimensional node motifs, which are characterized by multiple features in combination, specify nodes more comprehensively. With

these more precise descriptions, new kinds of motifs can be formulated and Costa et al. (2009) presented a routine for their detection and specification. We proposed first improvements and procedures for automating parameter choices for this method (Echtermeyer et al., 2011), such that large numbers of networks can be processed and compared to each other. This enhanced technique is applied to brain networks in this paper.

1.3 COMPARING MOTIF COUNTS ACROSS NETWORKS

The usefulness of mapping brain networks (which itself are abstract representations) to motif-classes is demonstrated for two scenarios: first, regional connectomes of 53 children (aged 12–23 years) are compared to show how the fingerprint of a DTI network can change over time. Second, we test the effect of different connectome-resolutions on the resulting network-fingerprint. Therefore, DTI networks at different levels of organization were generated by changing the level of granularity from 414 ROIs to up to 1615 ROIs.

Comparing networks from the same subject but at different granularities has so far been complicated by the critical dependence of standard approaches on the resolution of the network (Antiqueira et al., 2010). We show how our novel node fingerprints can be used to compare different levels of connectome organization.

Note that comparing motifs has several advantages over comparing network features. Features, such as the small-world measures of clustering coefficient and characteristic path length, critically depend on the edge density of a network. Therefore, changes in these measures could be either due to changes in network organization (e.g., modularity) or due to a difference in the number of connections. To identify the changes related to network organization, the edge density needs to be constant when comparing networks. However, this leads to a severe problem: when transforming raw data of connection strengths between nodes to binary data, each network needs a different threshold for establishing a connection (Rubinov and Sporns, 2010). To avoid the problem of either having different edge densities or different thresholds, researchers increasingly use weighted networks without thresholding (Bullmore and Bassett, 2011). For our analysis, we expect that differences in edge density are not only due to inter-subject variability, as in other studies, but are more heavily influenced by changes across different temporal or spatial scales. Therefore, we observe how network measures, and as a result node motifs, change over scales. These changes can be due to changes in organization or edge density.

Although a detailed analysis is outside the scope of this paper, we show the relation between both scales in **Figure 1**. Whereas edge density decreases with network size, there are no significant changes between age groups when looking at networks with the same spatial resolution. However, we would like to point out that (a) motifs change over time and can therefore be used to find characteristic changes in the developing brain and (b) motifs change over spatial scales, therefore comparisons between studies are only meaningful when the same (or a comparable) spatial resolution has been used.

2 MATERIALS AND METHODS

2.1 RECRUITING AND DATA RECORDING

We recruited 53 participants (age between 12 and 23 years) from local community and high schools. The study received ethics committee approval and written informed consents were obtained from the participants' parents for under-age participants, or from themselves if older than 18 years after study procedures were described. Participants were screened for a history of psychiatric and neurological disorders and current drug abuse using ophthalmological assessment including monocular and binocular visual acuity. Participants were grouped into four age-categories (**Table 1**).

A 3-Tesla scanner (Siemens Trio, Erlangen, Germany) was used at the Brain Imaging Centre (BIC), Frankfurt am Main, Germany. We recorded T1 weighted MRI (voxel-size 1 mm × 1 mm × 1 mm) and Diffusion Tensor MRI (voxel-size 2 mm × 2 mm × 2 mm). To minimize head motion, we used tightly padded clamps on the eight-channel head coil. The T1 weighted MRI was recorded with the following parameters: 176 slices, field of view (FoV) 256 mm, repetition time (TR) 2250 ms, and echo time (TE) 2.6 ms. Three Diffusion Tensor Images per participant were recorded with the following parameters: 60 slices, FoV 192 mm, TR 7600 ms, TE 91 ms, 60 directions b-vectors with b-factor of 1000 and 10 b0-images.

2.2 DATA PROCESSING AND NETWORK CONSTRUCTION

We used FreeSurfer to obtain surface meshes of the boundary between gray matter and white matter from T1 anatomical brain images[1], and Diffusion toolkit along with TrackVis (Wang et al., 2007) to obtain stream-line tractography from diffusion tensor

[1]http://surfer.nmr.mgh.harvard.edu

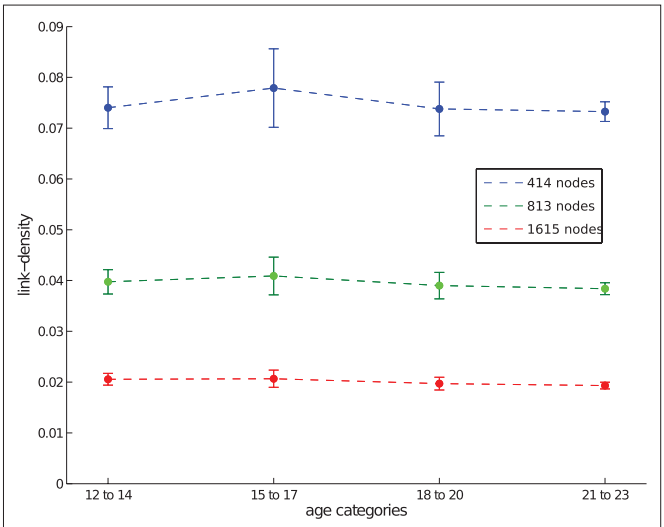


FIGURE 1 | Edge density for different spatial resolutions, given by the number of network nodes, and different age groups. Data points show the average edge density of all networks belonging to a certain age group and spatial resolution. Corresponding SDs are shown by error-bars.

Table 1 | Ranges of age-categories.

Category	Age range (years)	Number of subjects
1	12–14	9
2	15–17	20
3	18–20	16
4	21–23	8

images. We registered surface meshes into brains of the diffusion tensor images to extract networks. FreeSurfer provides parcelation of anatomical regions of cortices (33 for each hemisphere) based on the Deskian atlas (Fischl et al., 2004b; Desikan et al., 2006) and subcortices (Fischl et al., 2002, 2004a). We subdivided those anatomical regions of cortices into many of similar surface area using our own algorithm. It is known that the size of each ROI affects network connectivity and various network measures (Hagmann et al., 2007; Zalesky et al., 2010). The core of our algorithm is an expectation–maximization (EM) algorithm. First, we choose the expected number of subdivisions (400, 800, and 1600) for cortical regions (we do not subdivide subcortical areas) and decided the number of subdivisions based on the ratio of each surface area over whole surface area in the average subject. Then we subdivide each anatomical cortical region to the assigned number of subdivisions using the EM algorithm, minimizing variance of subdivisions' surface area sizes. The ratio of variance to mean does not exceed 10% in the average subject. In each subject, this ratio was higher (400: 414.78 ± 140.06 mm^2, 33.77%; 799: 207.65 ± 77.51 mm^2, 37.33%; 1601: 103.63 ± 44.33 mm^2, 42.78%, statistics of cortical ROIs only) because the size of each subject's anatomical region varied (2468.06 ± 1813.10 mm^2, 73.46%, which represents individual subjects' regions were 12 ± 30% bigger than the average subject's; ratio of variance to mean is 26.49%). Another useful statistic is the ratio of interquartile range to median: larger values represent higher dispersion. The values (**Table 2**) are smaller than those determined by Fornito et al. (2010).

We projected this new parcelation template into each subject's surface. Using this projection procedure we kept topological consistency across all subjects. We also included a few selected subcortical areas in the ROI list: nucleus accumbens, Amygdala, Caudate, Hippocampus, Pallidum, Putamen, and Thalamus. Thus the actual number of ROIs (414, 813, and 1615, respectively) is slightly larger than specified. These three numbers of ROIs on the brain and resulting networks are shown in **Figure 2**. Before further processing the diffusion tensor images, we serialized three diffusion tensor images and b-vectors for each participant; thus, a collated image of each subject has 30 b0-images and 180 diffusion images. The eddy current was corrected through FSL[2]. We used Diffusion toolkit along with TrackVis (Wang et al., 2007) with fiber assignment by continuous tracking (FACT) algorithm (Mori et al., 1999) and 35° of angle threshold. After registering the surface meshes (FreeSurfer[3]) of each subject to his/her DTI, we generated cortical volume ROIs, which are voxels in the gray matter. Also the selected subcortical areas (see above) were registered to DTIs and used as ROIs. Then, we used the UMCP toolbox[4] to obtain connectivity matrices from the defined and registered ROIs and tractography. This counts the number of fibers from a tractography from the Diffusion toolkit, between all pairs of defined ROIs' gray matter; the fiber counts between all ROI-pairs yield the weight-matrix. We also computed the average fiber lengths between ROIs (if there is no connection between a pair, the length is set to zero). The fiber length is based on the actual three-dimensional trajectory of the fiber tract and may be larger than the Euclidean distance between connected nodes. Note that long fiber length usually also corresponds to a large Euclidean distance between ROIs: earlier studies in fiber tracts in primates have shown that 85% of the fibers go in a straight line or are only slightly curved (Hilgetag and Barbas, 2006). However, only a fully bended (U-turn) fiber could result in a long fiber-tract connection to spatially nearby nodes.

Diffusion tensor imaging fiber-tracking yielded weighted matrices where weights indicate the probability of connections between corresponding ROIs. Non-zero weights have been interpreted as an indication for a link when constructing the brain networks. The resulting adjacency matrices were similarly sparse for each network resolution (**Figure 1**, average link-densities 7.5, 4.0, and 2.0% for networks of 414, 813, and 1615 nodes, respectively).

2.3 NETWORK ANALYSIS

Single node motifs were identified from the constructed networks by applying a technique presented by Costa et al. (2009) with additions in order to choose parameters automatically (Echtermeyer

[2]http://www.fmrib.ox.ac.uk/fsl/

[3]http://surfer.nmr.mgh.harvard.edu/

[4]http://ccn.ucla.edu/wiki/index.php/UCLA_Multimodal_Connectivity_Package

Table 2 | Region of interest (ROI) surface area (mm^2) statistics with ratio of interquartile range to median.

Parcelation	Median	Interquartile range	Ratio
aparc	1935	2811	1.45
414	395	183	0.46
813	195	101	0.52
1615	95	56	0.59

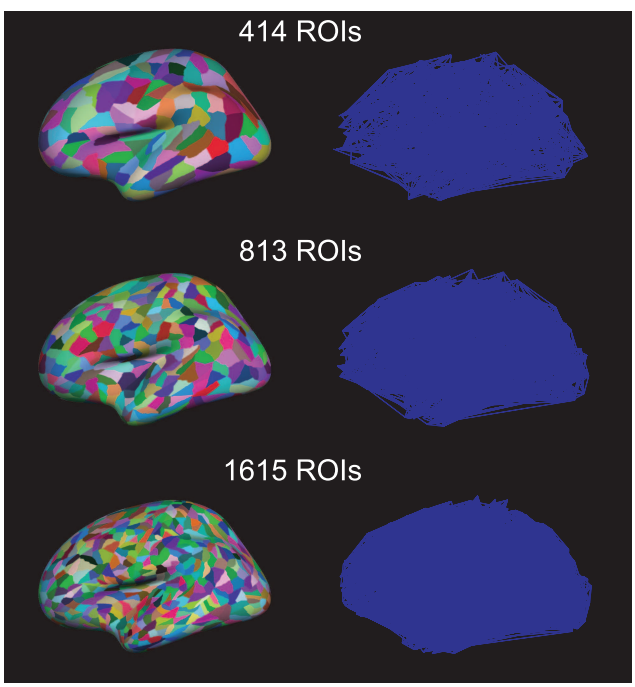


FIGURE 2 | Region of interests (ROIs) on the brain (left) and the resulting networks (right) with different numbers of ROIs.

et al., 2011). The method yields a global characterization of the network based on its local properties. Here, each node is quantified by nine different local measures, which are motivated in the next paragraph. We chose such a large number of local measures, some of which share similarities, to ensure that as many network features as possible are considered. Redundancies, where two features are highly correlated, are removed during the node-motif detection, which is explained further below.

A nodes' *degree* is its number of connections to other nodes; dividing this by the number of all links in the network yields the *normalized node degree K*. The average over all neighbors' degrees is called the *normalized average degree r*. (Nodes that are directly linked to each other are called *neighbors*.) Degrees of a node's immediate neighbors can further be taken into account by their respective *coefficient of variation cv*. Connectivity among neighbors of a node is quantified by the *clustering coefficient cc*, which reflects how many of all possible connections between neighbors actually exist (Watts and Strogatz, 1998; Kaiser et al., 2007a). The *hierarchical clustering coefficient* of level two cc_2 extends this concept to connections between neighbors' neighbors (Costa and Silva, 2006). To what degree a node's neighbors connect to the same target is quantified by the *locality index loc*, which is based on the matching index (e.g., Kaiser and Hilgetag, 2004a). The six measures mentioned so far quantify topological aspects of the network, while the next three measures also take into account spatial features. The *average connection length acl* for a node is the average length of fiber tracts to its neighbors and the furthest away neighbor determines the *maximum connection length mcl*. The average over connection lengths to nodes two steps away, i.e., neighbors' neighbors, is called the *average indirect reach air*. Further details and measures can be found in the literature (Albert and Barabási, 2002; Newman, 2003; Newman et al., 2006; Costa et al., 2007b).

The motif-detection workflow starts by applying the above listed network measures (**Figure 3**): local network measures are calculated for all nodes of the network, which yields a nine-dimensional characterization (*feature vector*) for each node. This high-dimensional

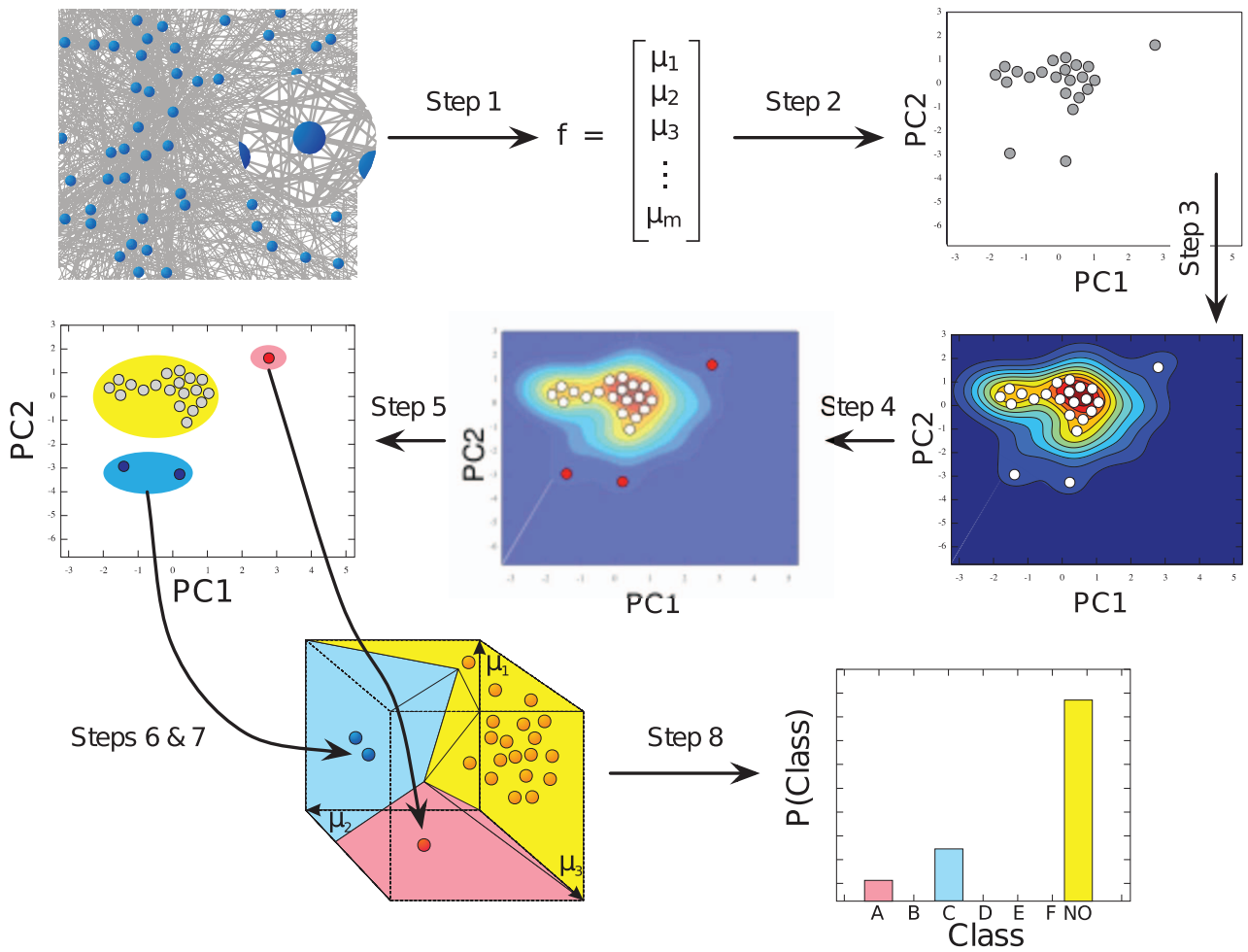


FIGURE 3 | Figure from Echtermeyer et al. (2011). Illustration of network analysis workflow: high-dimensional characterization of all network nodes through local network measures μ_i (Step 1) is compacted to two dimensions (Step 2) in order to estimate a probability distribution (Step 3), which is used to identify nodes with uncommon features (Step 4). All nodes are grouped (Step 5) to form high-dimensional motif-regions (Step 6), which are eventually joined, if too close to each other (Step 7). The number of nodes in each motif-region yields a fingerprint of the network (Step 8).

representation is simplified by exploiting similarities between different measures (Costa et al., 2007b): using principal component analysis (PCA; Johnson and Wichern, 2007, Chapter 8), correlation between measures is removed and the feature vectors are reduced to two dimensions (*PCA-plane*). The position on the PCA-plane characterizes the nodes and allows for comparisons: nodes close to each other share similar features whereas well separated ones differ. The next step is to estimate how likely specific combinations of features are. This is done by smoothing over points in the PCA-plane using the Parzen window approach (Parzen, 1962; Duda et al., 2001, Chapter 4.3) with width of the smoothing kernel scaled according to the SD along the corresponding principal component axis (Echtermeyer et al., 2011). Thereby, each node is assigned a probability, which is used to distinguish those with common and rare features. The w many nodes with uncommon properties, termed *outlier nodes*, were determined as follows. Given all nodes' probabilities $p = (p_k)_{k=1,\ldots,n}$ (sorted increasingly), the respective mean $\bar{p}$, and SD $\sigma(p)$, the number of singular nodes w is chosen as

$$w = \arg \max_{k:p_k<\bar{p}\ -\sigma(p)} p_{k+1} - p_k, \tag{1}$$

or $w = 0$, if all probabilities are greater than $\bar{p} - \sigma(p)$. Nodes corresponding to the smallest w probabilities are the identified outliers. Next, the outlier nodes were assigned to clusters each of which represents a different *motif-group*, i.e., a set of features. This was done by first centering equally sized ellipses on each point and determining their overlap. Maximal sets of nodes whose ellipses form a connected area yield both the number k of motif-groups as well as the groups themselves. (The above mentioned mechanisms have been evaluated and discussed in detail in an earlier publication (Echtermeyer et al., 2011).) Motif-groups correspond to nine-dimensional *motif-regions* in feature space, such that all nodes within a motif-group are similar with respect to the local network measures used in the beginning. In the final step, motif-regions from all networks were reduced to the final nine motif-regions by subsequently joining the closest two of them. The pairwise distances have been determined with a modified Mahalanobis measure (Mahalanobis, 1936). The number of nodes that express a particular motif yields a *fingerprint* of the network, which characterizes the network and allows for comparisons (Costa et al., 2009), e.g., by considering the *motif-diversity*, which is the number of unique motifs k that were identified. Another possibility is to compare the number of outlier nodes w.

3 RESULTS

3.1 MOTIF CHANGES WITH AGE

Networks were first analyzed with respect to age-dependent changes. Therefore, outlier motifs were determined for all networks to see whether certain motifs might only be expressed in subjects of similar age. This was not the case, but the total number of outlier nodes w shows a characteristic pattern with age (**Figure 4**, solid blue lines). Subjects of young ages (13–14 years) show 3–4 outlier nodes on average – a number that is reduced throughout the teenager years and thereafter (15–20 years) until a sudden and predominant peak at the age of 21. Subjects aged 22–23 years have again fewer, but still several outlier nodes. This pattern is qualitatively preserved for different network resolutions (**Figure 4** top to bottom). Quantitatively, the number of outlier nodes w rises with larger networks.

To determine which changes are statistically significant a random permutation test has been applied: subjects were randomly assigned an age (12–23 years) while ensuring that each age group had the same size as for the original data. Subsequently, each group's mean w and k were compared to those of the original data to compute the p-values (using a total of 500,000 permutations). Results, shown directly in **Figure 4**, represent how far each group is to the average over age. Note that except for the lowest network resolution, the large spike at age 21 does not lead to significant increase in w (and k). This is because our data only include one subject aged 21 and further data would be needed to confirm that the deviations at this age are indeed significant across all scales of resolution.

A pattern that is very similar to that of the outlier number w is the diversity k of expressed motifs. The number k is high and low for nearly the same ages as was the number of outliers w before (**Figure 4**, dashed orange lines), which is also reflected by the strong correlation between w and k (Pearson correlation 0.94, 0.88, and 0.87 for 414-, 813-, and 1615-node networks, respectively). Peaks at low (13–14 years) and high (21 years) ages can be seen with a dip in between. This pattern fits networks with 414 nodes best, whereas networks with the higher resolution (813 and 1615 nodes) show an increasingly flat profile except for the later peak (21 years). Note that absolute motif-diversity k is relatively similar for all network resolutions, which was not the case for the actual number of outliers w.

Both, number of outliers w and motif-diversity k are interesting on their own, but their strong correlation provides additional information. Although the absolute values differ, similarity of the curves' profiles is independent of network resolution. We thus find the very robust effect that motif-diversity is coupled to the number of outlier nodes. In other words, whenever more outliers occur, they also become more diverse.

3.2 MOTIF CHANGES WITH NETWORK RESOLUTION

The previous section has already shown that network resolution affects outlier quantity, w, but not the relative distribution across the age range (**Figure 4**, solid blue lines). Different from that, motif-diversity, i.e., how many different motifs are expressed, k, is nearly invariant on an absolute scale, in contrast to its relative distribution (**Figure 4**, dashed orange lines). We tested whether the number w and diversity k of outliers, at different spatial scales, are correlated with cognitive skills of our subjects. Unfortunately, measures of intelligence yielded by four sub-scores of the Wechsler intelligence test were only available for 44 out of our 53 subjects. For all comparisons, the Pearson correlation was $r \leq 0.5$. Therefore, we cannot confirm a link between intelligence and w or k. However, intelligence might still be related to a specific combination of motifs; a possibility which could only be tested given a larger cohort of subjects. Next, we analyze the actual frequencies with which motifs are expressed by different age groups (**Table 1**).

Using the same raw DTI-data, networks have been constructed with different resolutions, i.e., differently sized ROIs, for each subject. More fine-grained ROIs parcelate the brain into more elements, each of which is represented by a network node. We were interested in the effect that different network resolutions have on the occurring node motifs (**Table 3**; **Figure 5**) and therefore, networks with 414, 813, and 1615 nodes were analyzed and compared to each other.

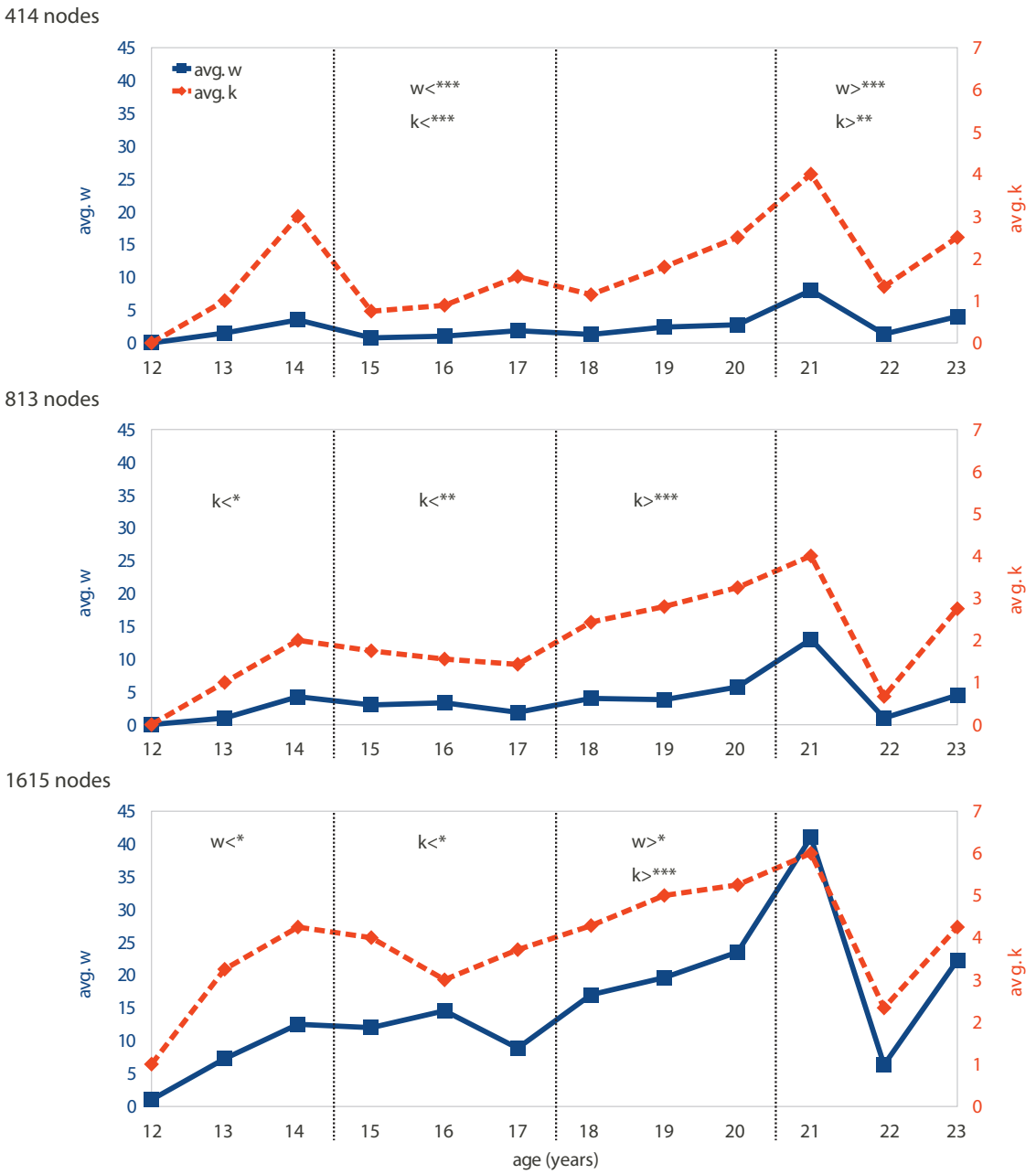


FIGURE 4 | Motif-expression changing with age: both the number of outlier nodes in a network *w* (solid blue line) and the diversity of motifs expressed *k* (dashed orange line) vary with subject age. The time-dependent patterns of *w* and *k* are shown for different network resolutions (rows). Age groups indicated by dashed vertical lines. Significantly de- or increased values for *w* and *k* are indicated by symbols < and >, respectively (*90, **95, and ***99% significance). Note that our data only include one subject aged 21 and further data would be needed to confirm significance of deviations at this age.

The results show that motif-expression strongly depends on network resolution (**Figure 6**). Whereas motif 4 is most frequent for the 414-node resolution, it is nearly absent in the highest resolution of 1615 nodes. Other motifs also show large fluctuations between spatial resolutions with either increasing or decreasing frequency. Motifs can also be characteristic for a single scale. For example, motifs 8, 9, and 10 are most prevalent across age groups at an intermediate resolution of 813 nodes. This shows that the spatial resolution is critical for assessing the frequency of motifs and, presumably, also for its underlying spatial and topological features.

Motif counts also depend on age. For 414-node networks, many motifs only occur for one or two out of the four age groups. Whereas this might be influenced by the low number of nodes in the network, we also see age-based motif frequency changes for higher spatial resolutions. Looking at the resolution of 1615 nodes, the

Table 3 | Properties of most frequent motifs.

Motif	*K*	*r*	cv	loc	cc	cc_2	acl	mcl	air
3	↓	↓	↓	↑	↑	↓	…	↓	↓
4	…	↑	…	↑	↓	…	…	↓	…
5	↓	…	↓	↑	↓	↑	↑	↑	↑
8	↓	↓	↑	↑	↓	↑	↑	…	↑
9	↓	…	↓	↑	↑	↑	↑	↑	↑
10	↓	…	↓	↑	↓	↑	…	↓	…

Symbols …, ↑, and ↓ indicate normal, elevated, and decreased values of K, normalized node degree; r, normalized average degree; cv, coefficient of variation of neighbors' degrees; loc, locality index; cc, clustering coefficient; cc_2, hierarchical clustering coefficient of level two; acl, average connection length; mcl, maximum connection length; air, average indirect reach, respectively.

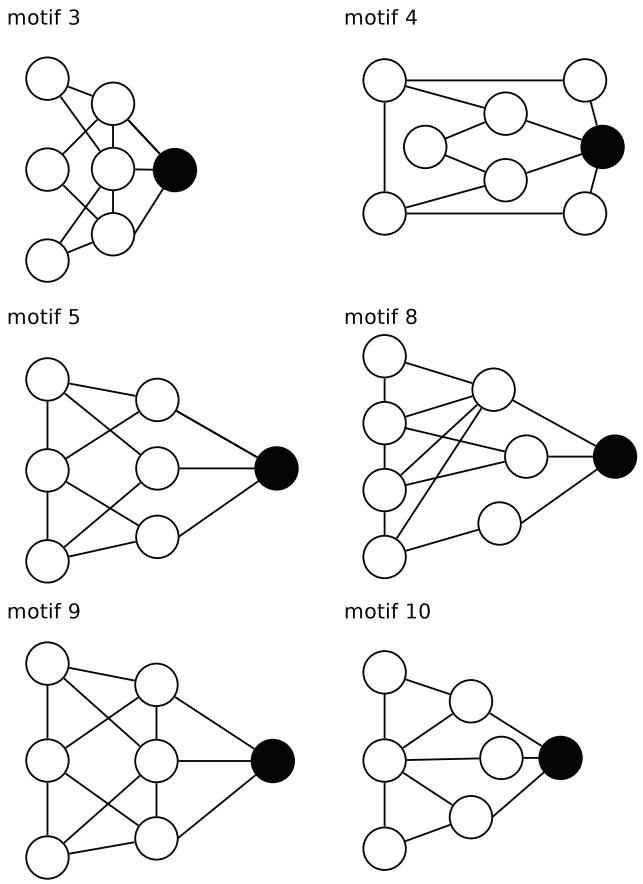


FIGURE 5 | Stylized illustration of most frequent motifs (Table 3): motif 3 Provincial node with connections in the direct (high cc) but not the larger neighborhood (low cc_2, air, and mcl). **motif 4** *Provincial hub* with more connections than its neighbors (high *r*) that are less connected between themselves (low *cc*). **motif 5** *Global bottleneck* with few (low *K*) but long-range connections (high *acl, mcl, air*) that reach beyond the local neighborhood (low *cc*). **motif 8** *Global uniform bottleneck* sharing properties of motif 5 but connected to nodes with similar degrees (low *cv*). **motif 9** *Global local bottleneck* sharing properties of motif 5 but also having well-connected neighbors (high *cc*) thus better informing local circuits. **motif 10** *Provincial bottleneck* with only short-range connections (low *mcl*) and few connections between its neighbors (low *cc*) leading to a large influence on the local circuit.

number of nodes showing motifs 3 and 5 increases with age whereas motifs 8 and 9 become less frequent. These age-dependent trends are less visible at 813-node resolution and they disappear for the 414-node networks.

3.3 FIBER LENGTH DISTRIBUTION FOR DIFFERENT NETWORK SIZES

A possible explanation for the observed dependence on network resolution could be fiber length based network measures (namely: average connection length, maximum connection length, and average indirect reach). These measures take into account the determined fiber trajectory between ROIs, i.e., their fiber-tract length. It is thus to be expected that the number of ROIs impacts on the corresponding fiber lengths. We tested this by observing the fiber length distribution for different network resolutions (**Figure 7**). Across all resolutions, short fibers are more frequent than long fibers. This exponential tail, which can be fitted by a Gamma distribution (Kaiser et al., 2009), has been reported before not only at the level of fiber tracts (Kaiser and Hilgetag, 2004b), but also at the level of connections between neurons within cortical areas (Hellwig, 2000). One notable difference between spatial resolutions is the increase in the number of short fibers (length < 10 mm) increasing by more than 50% from the lowest to the highest spatial resolution. This can be due to the smaller size of network nodes where surface regions that before belonged to the same node are now separate nodes on the surface. Therefore, a short fiber tract between these nodes becomes feasible. Looking at the maximum range of fibers, spatial resolution does not seem to influence maximum fiber lengths. However, differences in average fiber length, in particular due to an increased number of short fiber connections, might occur for particular nodes. The pattern for average indirect reach is more complex and can be influenced by changes in both short and long fibers. Overall, given a three-fold increase in spatial resolution, the fiber length distributions remain remarkably similar.

3.4 MOTIF CONSISTENCY ACROSS SPATIAL SCALES

How consistent are motifs across different resolutions? If a region splits into several regions for higher resolutions, do the daughter regions show the same motif as the parent region? Interestingly, this does not seem to be the case. **Figure 8** shows a typical example of motif locations for one subject. For 414 nodes, one location shows motif 3. One part of this location then shows motif 8 for the resolution of 813 nodes. For 1615 nodes, part of the original region shows motif 3 again, whereas another part shows motif 9. This is an interesting point when talking about the connectivity between brain regions: a pattern shown at the low-resolution level (e.g., cortical regions or Brodmann areas) may consist of several diverse patterns when observing connectivity at a higher resolution.

4 DISCUSSION

We observed how the distribution (fingerprint) of characteristic motifs of single network nodes changed for human structural connectivity (DTI) during brain development. We studied the role of both temporal scales (age groups from 12 to 23 years) and spatial scales (414, 813, and 1615 nodes) and found that, first, the number and diversity of motifs in a network are strongly correlated. Second, comparing different scales, the number and diversity of motifs varied across the temporal (subject age) and spatial (network

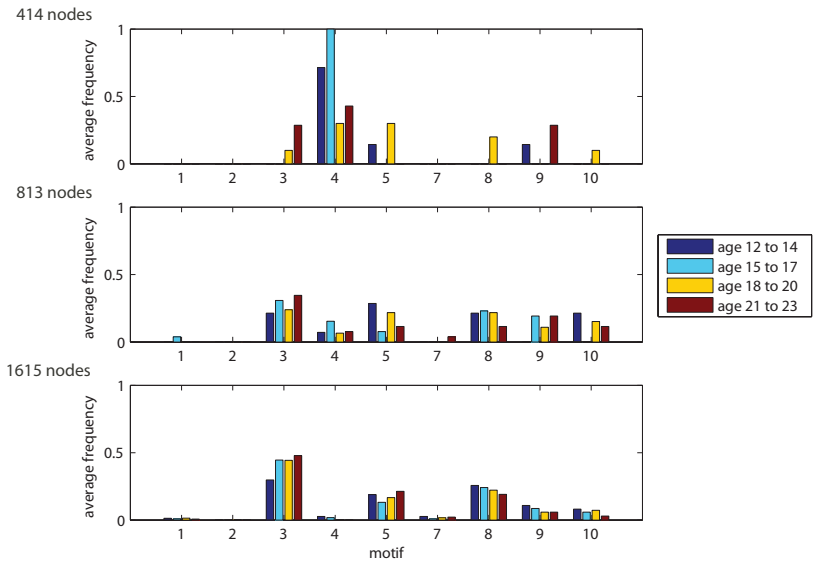


FIGURE 6 | Motif-expression changing with network resolution. Plots show distribution of outlier nodes among motifs 1–5 and 7–10. Motif 6 (not shown) corresponds to the remaining 98% network nodes with common features (regular nodes).

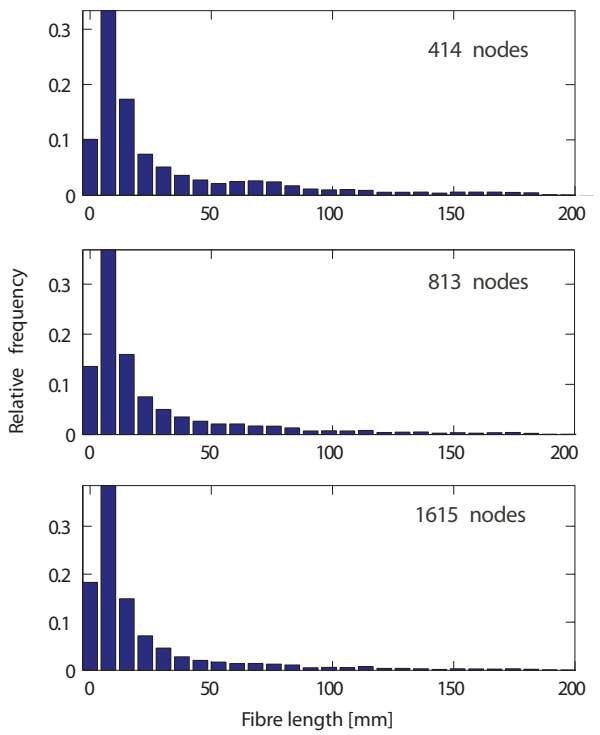


FIGURE 7 | Fiber length distribution, as the length of the trajectory in millimeter, for different network resolutions. Relative frequencies for low (414 nodes, top), medium (813 nodes, middle), and high (1615 nodes, bottom) spatial resolution. Note that longer fibers (>200 mm) occurred so infrequently that corresponding bars (not shown) would be invisible.

resolution) scale (with certain motifs only occurring at one spatial scale or for a certain age range). Third, the sub-regions of a node, using a higher spatial resolution, may or may not include the original motif at the lower resolution. Therefore, both the type and localization of motifs differ for different spatial resolutions. This indicates that spatial resolution is crucial when comparing characteristic node fingerprints given by topological and spatial network features. This result is in line with previous studies that observed the role of sampling on brain network properties (Antiqueira et al., 2010) and it additionally shows the influence of the temporal scale.

Networks can be characterized at different levels. Aggregate measures that characterize a network as a whole can be used to distinguish different network types, e.g., small-world (Watts and Strogatz, 1998; Hilgetag et al., 2000; Sporns et al., 2000), scale-free (Eguiluz et al., 2005; Kaiser et al., 2007b), modular (Hilgetag et al., 2000), or hierarchical (Kaiser et al., 2010). Networks can also be characterized at the level of individual components. Looking at the topological features of individual nodes, a first study by Passingham et al. (2002) has linked the network features of macaque cortical areas with the function of each area. Whereas this approach considered all nodes of the network, there are also ways to search for "special" nodes of a system. Following the ideas of scale-free networks, brain areas with a large number of connections – so-called network hubs – might be crucial for integrating or distributing information (Kaiser et al., 2007b). Sporns et al. (2007) were able to detect and classify different kinds of hubs of cortical structural connectivity. However, hubs are just one type of outliers where one or more node features (here: the number of connections of a node) differ from the average value of nodes in the network. In two previous manuscripts (Costa et al., 2009; Echtermeyer et al., 2011), we have developed a tool that systematically searches for and classifies nodes which differ from the majority of nodes in a network. This approach includes the joint analysis of multiple features, which include topological as well as spatial properties (Costa et al., 2007a).

Our measures are based on outliers, which do not show the overall trend as shown in previous studies (spatial resolution dependent changes: Zalesky et al. (2010), Bassett et al. (2010), Fornito et al.

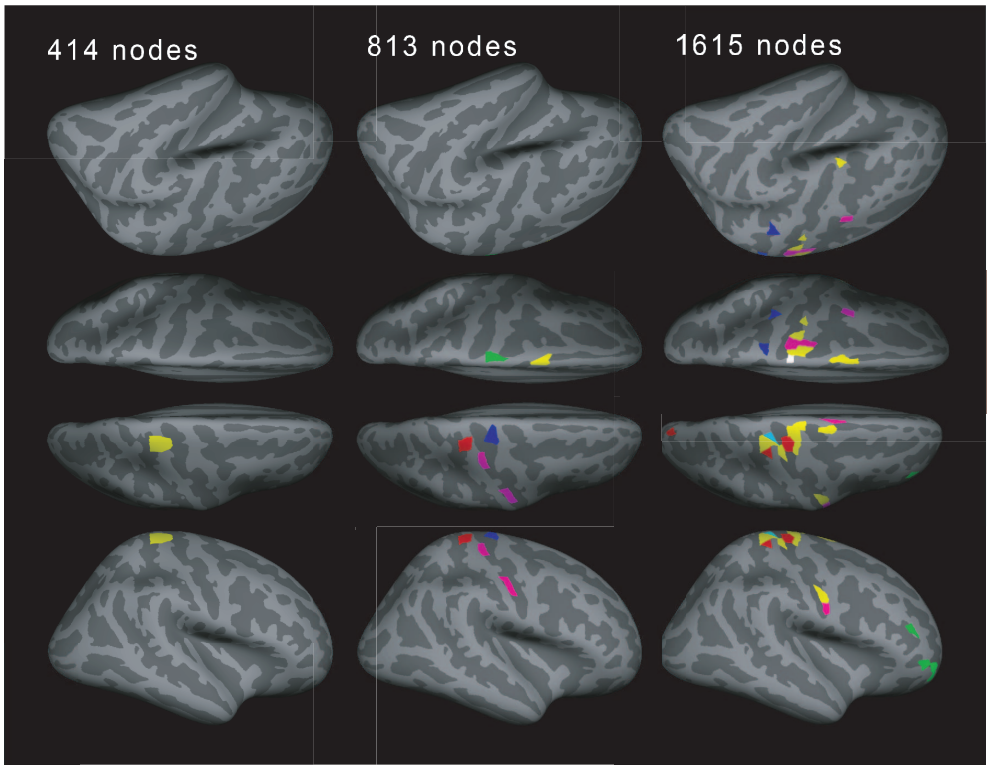


FIGURE 8 | Example for the motif distribution in one subject (19 years old) for different spatial resolutions. Different number of regions of interests (ROIs, left: 414, middle: 813, and right: 1615) with different views: left lateral view, left superior view, right superior view, and right lateral view in order from the top. Yellow: motif 3, magenta: motif 4, cyan: motif 7, red: motif 8, green: motif 9, blue: motif 10 (motifs 1, 2, and 5 were not present).

(2010), Hayasaka and Laurienti (2010), age-dependent changes: Fan et al. (2010), Hagmann et al. (2010), Fair et al. (2009); Uhlhaas et al. (2009), **Table A1** in Appendix). Thus, it is difficult to directly relate our results to their findings. However, our report that edge density decreased with increased spatial resolution is consistent with previous studies of structural connectivity (Bassett et al., 2010; Zalesky et al., 2010). For the temporal resolution, most previous studies showed no overlap with the age range of our study. Hagmann et al. (2010) performed a study with overlapping temporal scales, but only observed ages up to 18 years and thus only about half of the temporal scales considered in our study. Studies with larger overlap focused on functional rather than structural connectivity (Uhlhaas et al., 2009; Fan et al., 2010).

Here we noted that surface areas of used ROIs varied around 33–43% compared to their mean (shown only for the cortical ROIs because subcortical ROIs were not subdivided, **Table 2**). This variation stems from the variation of anatomical regions' surface areas (ratio of its SD to its mean is 73.5%). The ratio of interquartile range to median showed comparatively small ROI variation (Fornito et al., 2010). To keep the topological consistency, this might be the best we can do. At least, this is better than using just anatomical regions whose variation is much higher (73.5% variation).

This node fingerprint study is a proof of principle for the current technique showing the influence of spatial and temporal scales on network comparison. However, the current results could also be influenced by the following points. First, the number of subjects per age group was not identical (**Table 1**). Therefore, findings in age groups with fewer members might have been influenced by relatively few outliers. Whereas an identical number of subjects per group is desirable, for our study there was also a trade-off with the age range: using the same subject numbers would have led to groups where the age ranges differ, e.g., covering 3 years for one group and 5 years for another. We decided for a different number of subjects per group in order to perform the comparison across the temporal scale. Second, single node motifs become more robust for larger network sizes. As node motifs are outliers, only a small percentage of all nodes will be characterized as node motifs. Therefore, for low-resolution networks, node motifs might only occur for few subjects. In addition, their number might be a poor estimate of the underlying frequency in human connectivity networks. For this reason, we have only reported networks with at least 414 nodes, leaving out networks with a parcelation into 110 cortical and subcortical regions. Third, node features but not necessarily motifs depend on the edge density of the network. In our developmental networks, we observed significant edge density changes for different spatial resolutions ranging from more than 7% for 414 nodes to 2% for 1615 nodes. However, for each network resolution, edge densities remained comparable for different temporal scales (age groups; **Figure 1**). Fourth, deterministic tracking cannot capture crossing fibers which can be done in probabilistic tracking (Behrens et al., 2007), in Diffusion spectrum imaging (Wedeen et al., 2008; Bassett et al., 2010), or in high-angular resolution diffusion imaging

(HARDI; Tuch, 2004; Zalesky et al., 2010). However, fiber length distributions, influencing our three spatial features, were similar across resolutions (**Figure 7**).

Network Science has led to a wide range of tools for analyzing neural systems (Sporns et al., 2004; Costa et al., 2007b; Bullmore and Sporns, 2009; Rubinov and Sporns, 2010; Kaiser, 2011). Whereas the characterization of individual networks, e.g., as small-world or scale-free, is now possible, comparing different networks is still a challenge. Our study shows how node properties, given by characteristic single node motif fingerprints, can be compared *between* different networks. Studying single node properties is only one way to characterize networks with alternatives being (a) the study of global properties including local and global efficiency (Latora and Marchiori, 2001; Achard and Bullmore, 2007) as well as modularity (Newman, 2006), (b) the comparison of the cluster organization of different networks, and (c) direct comparison of network matrices (Crofts and Higham, 2009). However, it is problematic for these measures that network features are influenced by the number of nodes and edges of a network; feature changes between networks might simply be due to different edge densities in the compared systems (van Wijk et al., 2010). For our networks, edge densities did not significantly differ across age but they differed across spatial resolutions. Whereas the underlying topological and, to some extent, spatial features might have changed in our networks, our motif fingerprints do not depend on the absolute values for network measures but on their distribution: an overall increase or decrease in the average value of a measure will have less of an influence on the number of node motifs as these depend on statistical outliers and not on absolute values of a measure.

Nonetheless, nodes that present characteristic node motifs, as well as hubs, are rare within networks. These fingerprints can therefore only be applied for high-resolution networks (≥414 nodes) and large groups of subjects. Fortunately, large subject cohorts are currently being recruited in several initiatives including the Human Connectome Project for structural connectivity and the 1000 Functional Connectome Project (Biswal et al., 2010). In addition, fingerprints will also be useful for high-resolution networks at the scale of the micro-connectome observing connections between individual neurons (DeFelipe, 2010). Another potential application lies in the analysis of multi-electrode array recordings, which nowadays can record from more than 4000 channels (Sernagor et al., 2010). We therefore made this tool available within the CARMEN initiative[5] for developing electrophysiology analysis tools (Smith et al., 2007). The tool is also available on our website[6].

5 CONCLUSION

In this study we found that (a) node motifs change over time and can therefore be used to find characteristic changes in the developing brain and (b) motifs change over spatial scales, therefore comparisons between studies are only meaningful when the same (or a comparable) spatial resolution has been used. Our results also indicate that spatial resolution has a higher effect on topological measures whereas spatial measures, based on fiber lengths, remain more comparable between resolutions. As node motifs are based on topological and spatial properties of brain connectivity networks, these conclusions are also relevant to other studies using network analysis. Another important aspect is the analysis of differences between healthy controls and subjects with brain disorders that can arise during brain network development, such as schizophrenia and epilepsy; we hope that our node fingerprint approach will be useful for detection and specification in these cases.

ACKNOWLEDGMENTS

Marcus Kaiser and Christoph Echtermeyer were funded by EPSRC (EP/G03950X/1) and the CARMEN e-science Neuroinformatics project (http://www.carmen.org.uk) funded by EPSRC (EP/E002331/1). Marcus Kaiser and Cheol Han also acknowledge support by the WCU program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R32-10142).

[5]http://www.carmen.org.uk

[6]http://www.biological-networks.org/

REFERENCES

Achard, S., and Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS Comput. Biol.* 3, e17. doi: 10.1371/journal.pcbi.0030017

Albert, R., and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.* 74, 47–97.

Antiqueira, L., Rodrigues, F. A., van Wijk, B. C. M., Costa, L. D. F., and Daffertshofer, A. (2010). Estimating complex cortical networks via surface recordings – a critical note. *Neuroimage* 53, 439–449.

Bassett, D. S., Brown, J. A., Deshpande, V., Carlson, J. M., and Grafton, S. T. (2010). Conserved and variable architecture of human white matter connectivity. *Neuroimage* 54, 1262–1279.

Behrens, T. E., Berg, H. J., Jbabdi, S., Rushworth, M. F., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: what can we gain? *Neuroimage* 34, 144–155.

Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A. M., Ernst, M., Fair, D., Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kotter, R., Li, S. J., Lin, C. P., Lowe, M. J., Mackay, C., Madden, D. J., Madsen, K. H., Margulies, D. S., Mayberg, H. S., McMahon, K., Monk, C. S., Mostofsky, S. H., Nagel, B. J., Pekar, J. J., Peltier, S. J., Petersen, S. E., Riedl, V., Rombouts, S. A., Rypma, B., Schlaggar, B. L., Schmidt, S., Seidler, R. D., Siegle, G. J., Sorg, C., Teng, G. J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X. C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y. F., Zhang, H. Y., Castellanos, F. X., and Milham, M. P. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198.

Bullmore, E. T., and Bassett, D. S. (2011). Brain graphs: graphical models of the human brain connectome. *Annu. Rev. Clin. Psychol.* 7, 113–140.

Butz, M., Wörgötter, F., and Van Ooyen, A. (2009). Activity-dependent structural plasticity. *Brain Res. Rev.* 60, 287–305.

Buzsaki, G. (2006). *Rhythms of the Brain*. Oxford: Oxford University Press.

Costa, L. D. F., Kaiser, M., and Hilgetag, C. C. (2007a). Predicting the connectivity of primate cortical networks from topological and spatial node properties. *BMC Syst. Biol.* 1, 16. doi: 10.1186/1752-0509-1-16

Costa, L. D. F., Rodrigues, F. A., Travieso, G., and Boas, P. R. V. (2007b). Characterization of complex networks: a survey of measurements. *Adv. Phys.* 56, 167–242.

Costa, L. D. F., Rodrigues, F. A., Hilgetag, C. C., and Kaiser, M. (2009). Beyond the average: detecting global singular nodes from local features in complex networks. *Europhys. Lett.* 87, 18008.

Costa, L. D. F., and Silva, F. N. (2006). Hierarchical characterization of complex networks. *J. Stat. Phys.* 125, 841–876.

Crofts, J. J., and Higham, D. J. (2009). A weighted communicability measure applied to complex brain networks. *J. R. Soc. Interface* 6, 411–414.

DeFelipe, J. (2010). From the connectome to the synaptome: an epic love story. *Science* 330, 1198–1201.

Desikan, R., Se'gonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Albert, M., and Killiany, R. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage* 31, 968–980.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern Classification,* 2nd Edn. New York, NY: Wiley Interscience.

Echtermeyer, C., da Fontoura Costa, L., Rodrigues, F. A., and Kaiser, M. (2011). Automatic network fingerprinting through single-node motifs. *PLoS ONE* 6, e15765. doi: 10.1371/journal.pone.0015765

Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., and Apkarian, A. V. (2005). Scale-free brain functional networks. *Phys. Rev. Lett.* 94, 018102.

Fair, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., Schlaggar, B. L., and Petersen, S. E. (2009). Functional brain networks develop from a "local to distributed" organization. *PLoS Comput. Biol.* 5, e1000381. doi: 10.1371/journal.pcbi.1000381

Fan, Y., Shi, F., Smith, J. K., Lin, W., Gilmore, J. H., and Shen, D. (2010). Brain anatomical networks in early human brain development. *Neuroimage* 54, 1862–1871.

Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., and Dale, A. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.

Fischl, B., Salat, D., van der Kouwe, A., Makris, N., Ségonne, F., Quinn, B., and Dalea, A. (2004a). Sequence-independent segmentation of magnetic resonance images. *Neuroimage* 23(Suppl. 1), S69–S84.

Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D., Busa, E., Seidman, L., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., and Dale, A. (2004b). Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14, 11–22.

Fornito, A., Zalesky, A., and Bullmore, E. T. (2010). Network scaling effects in graph analytic studies of human resting-state fMRI data. *Front. Syst. Neurosci.* 4:22. doi: 10.3389/fnsys.2010.00022

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biol.* 6, e159. doi: 10.1371/journal.pbio.0060159

Hagmann, P., Kurant, M., Gigandet, X., Thiran, P., Wedeen, V., Meuli, R., and Thiran, J. (2007). Mapping human whole-brain structural networks with diffusion mri. *PLoS ONE* 2, e591. doi: 10.1371/journal.pone.0000597

Hagmann, P., Sporns, O., N. Madan, Cammoun, L., R. Pienaar, Wedeen, V., Meuli, R., Thiran, J., and Grant, P. (2010). White matter maturation reshapes structural connectivity in the alte developing human brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19067–19072.

Hayasaka, S., and Laurienti, P. J. (2010). Comparison of characteristics between region- and voxel-based network analysies in resting-state fmri data. *Neuroimage* 50, 499–508.

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biol. Cybern.* 82, 111–121.

Hilgetag, C. C., and Barbas, H. (2006). Role of mechanical factors in the morphology of the primate cerebral cortex. *PLoS Comput. Biol.* 2, e22. doi: 10.1371/journal.pcbi.0020022

Hilgetag, C. C., Burns, G. A. P. C., O'Neill, M. A., Scannell, J. W., and Young, M. P. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 91–110.

Horton, J. C., and Adams, D. L. (2005). The cortical column: a structure without a function. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 837–862.

Hubel, D., Wiesel, T., and LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 278, 377–409.

Johnson, R. A., and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*, 6th Edn. Englewood Cliffs, NJ: Prentice Hall.

Kaiser, M. (2011). A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage* 57, 892–907.

Kaiser, M., Goerner, M., and Hilgetag, C. C. (2007a). Criticality of spreading dynamics in hierarchical cluster networks without inhibition. *New J. Phys.* 9, 110.

Kaiser, M., Martin, R., Andras, P., and Young, M. P. (2007b). Simulation of robustness against lesions of cortical networks. *Eur. J. Neurosci.* 25, 3185–3192.

Kaiser, M., and Hilgetag, C. C. (2004a). Edge vulnerability in neural and metabolic networks. *Biol. Cybern.* 90, 311–317.

Kaiser, M., and Hilgetag, C. C. (2004b). Modelling the development of cortical systems networks. *Neurocomputing* 58–60, 297–302.

Kaiser, M., and Hilgetag, C. C. (2010). Optimal hierarchical modular topologies for producing limited sustained activation of neural networks. *Front. Neuroinform.* 4:8. doi: 10.3389/fninf.2010.00008

Kaiser, M., Hilgetag, C. C., and Kötter, R. (2010). Hierarchy and dynamics of neural networks. *Front. Neuroinform.* 4:112. doi: 10.3389/fninf.2010.00112

Kaiser, M., Hilgetag, C. C., and van Ooyen, A. (2009). A simple rule for axon outgrowth and synaptic competition generates realistic connection lengths and filling fractions. *Cereb. Cortex* 19, 3001–3010.

Latora, V., and Marchiori, M. (2001). Efficient behavior of small-world networks. *Phys. Rev. Lett.* 87, 198701.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proc. Indian Natl. Sci. Acad. B Biol. Sci.* 2, 49–55.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827.

Mori, S., Crain, B., Chacko, V., and van Zijl, P. (1999). Three dimensional tracking of axonal projections in the brain by magnetic resonance imaging. *Ann. Neurol.* 45, 265–269.

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain* 120(Pt 4), 701–722.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Rev.* 45, 167–256.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proc. Natl. Acad. Sci. U.S.A.* 130, 8577–8582.

Newman, M. E. J., Barabási, A.-L., and Watts, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press.

Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Stat.* 33, 1065–1076.

Passingham, R. E., Stephan, K. E., and Kötter, R. (2002). The anatomical basis of functional localization in the cortex. *Nat. Rev. Neurosci.* 3, 606–616.

Roopun, A. K., Kramer, M. A., Carracedo, L. M., Kaiser, M., Davies, C. H., Traub, R. D., Kopell, N. J., and Whittington, M. A. (2008). Temporal interactions between cortical rhythms. *Front. Neurosci.* 2:145–154. doi: 10.3389/neuro.01.034.2008

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* 52, 1059–1069.

Schüz, A., Chaimow, D., Liewald, D., and Dortenman, M. (2005). Quantitative aspects of corticocortical connections: a tracer study in the mouse. *Cereb. Cortex* 16, 1474–1486.

Sernagor, E., Maccione, A., Hennig, M. H., Gandolfo, M., Eglen, S. J., and Berdondini, L. (2010). "Changing dynamics of spontaneous waves during retinal development: a novel panretinal perspective achieved with the active pixel sensor (aps) 4,096 electrodes array," in *7th International Meeting on Substrate-Integrated Microelectrode Arrays*, Reutlingen.

Seung, H. S. (2009). Reading the book of memory: sparse sampling versus dense mapping of connectomes. *Neuron* 62, 17–29.

Smith, L., Austin, J., Baker, S., Borisyuk, R., Eglen, S., Feng, J., Gurney, K., Jackson, T., Kaiser, M., Overton, P., Panzeri, S., Quiroga, R. Q., Schultz, S., Sernagor, E., Smith, V., Smulders, T., Stuart, L., Whittington, M., and Ingram, C. (2007). "The CARMEN e-science pilot project: neuroinformatics work packages," in *Proceedings of the UK e-Science All Hands Meeting*, Nottingham, 591–598.

Sporns, O., Chialvo, D. R., Kaiser, M., and Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends Cogn. Sci. (Regul. Ed.)* 8, 418–425.

Sporns, O., Honey, C. J., and Kötter, R. (2007). Identification and classification of hubs in brain networks. *PLoS ONE* 2, e1049. doi: 10.1371/journal.pone.0001049

Sporns, O., Tononi, G., and Edelman, G. M. (2000). Theoretical neuroanatomy: relating anatomical and functional connectivity in graphs and cortical connection matrices. *Cereb. Cortex* 10, 127–141.

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042

Tuch, D. S. (2004). Q-ball imaging. *Magn. Reson. Med.* 52, 273–289.

Uhlhaas, P. J., Roux, F., Singer, W., Haenschel, C., Sireteanu, R., and Rodriguez, E. (2009). The development of neural synchrony reflects late maturation and restructuring of functional networks in humans. *Proc. Natl. Acad. Sci. U.S.A.* 106, 9866–9871.

Uhlhaas, P. J., and Singer, W. (2010). Abnormal neural oscillations and synchrony in schizophrenia. *Nat. Rev. Neurosci.* 11, 100–113.

van Wijk, B. C. M., Stam, C. J., and Daffertshofer, A. (2010). Comparing brain networks of different size and connectivity density using graph

theory. *PLoS ONE* 5, e13701. doi: 10.1371/journal.pone.0013701

Wang, R., Benner, T., Sorensen, A., and Wedeen, V. (2007). Diffusion toolkit: a software package for diffusion imaging data processing and tractography. *Int. Soc. Magn. Reson. Med.* 15, 3720.

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442.

Wedeen, V. J., Wang, R. P., Schmahmann, J. D., Benner, T., Tseng, W. Y. I., Dai, G., Pandya, D. N., Hagmann, P., D'Arceuil, H., and de Crespigny, A. J. (2008). Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. *Neuroimage* 41, 1267–1277.

Zalesky, A., Fornito, A., Harding, I. H., Cocchi, L., Yücel, M., Pantelis, C., and Bullmore, E. T. (2010). Whole-brain anatomical networks: does the choice of nodes matter? *Neuroimage* 50, 970–983.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 April 2011; accepted: 02 July 2011; published online: 22 July 2011.

Citation: Echtermeyer C, Han CE, Rotarska-Jagiela A, Mohr H, Uhlhaas PJ and Kaiser M (2011) Integrating temporal and spatial scales: human structural network motifs across age and region of interest size. Front. Neuroinform. ***5****:10. doi: 10.3389/fninf.2011.00010*

APPENDIX

Table A1 | Comparison of our findings with previous studies.

Study	Findings
Spatial scale	As the number of nodes increases
Ours	In SC (414, 813, and 1615 nodes), ↓ ED, ↑*w* (see **Figure 4**)
Zalesky et al. (2010)	In SC (6 different scales between 82 and 4000 nodes), ↓ ED, ↑ small worldness, ↑ γ, ↓ global efficiency, changes in nodal rank degree, and betweenness centrality
Bassett et al. (2010)	In SC (12 different scales between 54 and 880), ↓ ED, ↑ γ, ↑ λ, conserved hierarchy, ↑ Rentian scaling
Hagmann et al. (2010)	In SC (66 and 241 nodes), ↓ ED, ↑ CC, ↓ efficiency, ↓ node strength
Fornito et al. (2010)	In rsFC (7 different scales between 84 and 4320 nodes), ↓ average correlation, ↑ size of largest component, ↓ path length, ↑ CC (for low ED), ↓ CC (for high ED) ↑ small worldness, changes in nodal rank degree
Hayasaka and Laurienti (2010)	In rsFC (voxel-based nodes with three different voxel sizes and region-based nodes), changes in node degree distribution ↑ CC, ↑ path length (while increasing voxel sizes, but it decreased with region-based node)
Temporal scale	As the age increases
Ours	In SC (12–23 y/o), a characteristic pattern over spatial scales (see **Figure 4**)
Hagmann et al. (2010)	In SC (18 m/o to 18 y/o), ↓ mean ADC, ↑ mean FA, ↓ ED, ↓ CC, ↑ Efficiency, ↑ Node strength, ↑ SC–FC correlation, no changes in modularity and no major changes in module composition after 2 y/o, no significant changes in betweenness centrality
Fan et al. (2010)	In SC (1 m/o, 1 y/o, 2 y/o, and adult), ↑ global efficiency, ↑ cost efficiency, a peak at 2 y/o in local efficiency and modularity, ↑ size of largest component in modules, changes in module assignment, and participation coefficient
Fair et al. (2009)	In rsFC (7–31 y/o), no significant changes in optimized modularity Q, CC and CPL. Changed module assignment and the number of long-distance correlations ↑.
Uhlhaas et al. (2009)	In EEG (6–21 y/o), strong correlation between neural synchrony and cognitive performance. As ages increase, the neural synchrony was increasing in early adolescence, then decreasing in late adolescence, and finally increasing again in adult

SC, structural connectivity; rsFC, resting state functional connectivity; m/o, months-old; y/o, years-old; ED, edge density; CC, clustering coefficient; CPL, characteristic path length; ADC, apparent diffusion coefficient; FA, fractional anisotropy; γ, CC/CC_{rand}; λ, CPL/CPL_{rand}.

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 23 September 2011
doi: 10.3389/fninf.2011.00020

Modeling the connectome of a simple spinal cord

Roman Borisyuk[1]*, Abul Kalam al Azad[1], Deborah Conte[2], Alan Roberts[2] and Stephen R. Soffe[2]

[1] *School of Computing and Mathematics, University of Plymouth, Plymouth, UK*
[2] *School of Biological Sciences, University of Bristol, Bristol, UK*

Edited by:
Olaf Sporns, Indiana University, USA

Reviewed by:
Brian Mulloney, University of California Davis, USA
Ronald L. Calabrese, Emory University, USA

****Correspondence:***
Roman Borisyuk, School of Computing and Mathematics, University of Plymouth, Plymouth PL4 8AA, UK.
e-mail: rborisyuk@plymouth.ac.uk

In this paper we develop a computational model of the anatomy of a spinal cord. We address a long-standing ambition of neuroscience to understand the structure–function problem by modeling the complete spinal cord connectome map in the 2-day old hatchling *Xenopus* tadpole. Our approach to modeling neuronal connectivity is based on developmental processes of axon growth. A simple mathematical model of axon growth allows us to reconstruct a biologically realistic connectome of the tadpole spinal cord based on neurobiological data. In our model we distribute neuron cell bodies and dendrites on both sides of the body based on experimental measurements. If growing axons cross the dendrite of another neuron, they make a synaptic contact with a defined probability. The total neuronal network contains ∼1,500 neurons of six cell-types with a total of ∼120,000 connections. The anatomical model contains random components so each repetition of the connectome reconstruction procedure generates a different neuronal network, though all share consistent features such as distributions of cell bodies, dendrites, and axon lengths. Our study reveals a complex structure for the connectome with many interesting specific features including contrasting distributions of connection length distributions. The connectome also shows some similarities to connectivity graphs for other animals such as the global neuronal network of *C. elegans*. In addition to the interesting intrinsic properties of the connectome, we expect the ability to grow and analyze a biologically realistic spinal cord connectome will provide valuable insights into the properties of the real neuronal networks underlying simple behavior.

Keywords: tadpole, spinal cord, developmental approach, connections, axon growth

INTRODUCTION

A key to understanding the operation of any central nervous neuronal network is knowledge of the architecture of that network: where the neurons, dendrites, axons, and synapses are located in a three-dimensional structure. This detailed architecture of interneuronal connections provides a framework into which accumulated experimental information can be mapped. It may then be possible to model the activity of the network using these connections. In most cases, the morphology of nervous systems, or even regions of nervous systems, is highly complex, making it extremely challenging to define the real connection architecture of the component networks. Within the vertebrates, there is now extensive information on the brainstem and spinal cord neurons and networks that control locomotion (e.g., Ziskind-Conhaim et al., 2010; Kiehn, 2011). However, in all cases, the detailed architecture of these networks remains ill defined. The best understood locomotor networks have been described in lower vertebrates like the adult lamprey (Grillner, 2003), zebrafish larva (McLean and Fetcho, 2007), and frog tadpole (Roberts et al., 2010). Studying developing animals reminds us that all these networks have to self-assemble and grow appropriate connectivity. In an attempt to get insights into both the development and the connection architecture that it produces, we have used detailed experimental knowledge of the identity and synaptic connections of young tadpole spinal and brainstem neurons (Li et al., 2007) to model the "growth" of a biologically realistic connectome for this simple, vertebrate brainstem, and spinal cord network.

In the developing frog tadpole spinal cord, we have detailed information on the brainstem and spinal neurons active during swimming, their physiology, synaptic connections, and morphology (Roberts et al., 2010). On the basis of a large dataset of paired recordings, we proposed that the location or geography of axons and dendrites plays a fundamental role in establishing synaptic connectivity during early development (Li et al., 2007). Simple factors such as morphogen gradients controlling dorso-ventral soma, dendrite, and axon positions may constrain the synaptic connections made between different types of neuron sufficiently as the spinal cord first develops and in this way allow functional networks to form rapidly. For example, if "geographically" the dendrites of some neurons are located mainly dorsally while the axons of other neurons are located mostly ventrally, then it is unlikely that they will form synapses. This analysis implies that detailed cellular recognition between spinal neuron types may not be necessary for the reliable formation of functional networks which generate early behavior like swimming. If we model such mechanisms, the process of network formation is based on an interplay between deterministic and stochastic components. Repeated simulations of such a model result in different coupling architectures which nevertheless have many similarities and common features. The most important requirement is that different detailed connection

patterns result in the same functionality, in the case of the tadpole spinal cord: a pattern of spiking activity corresponding to swimming.

We have used a "developmental" approach to modeling the connectome of the young *Xenopus* tadpole spinal cord, which means that connections are not prescribed but rather appear as a result of the developmental process of axon growth. A first, simplified mathematical model (Li et al., 2007; Borisyuk et al., 2008) was based on a non-linear system of difference equations with a random component. This model included only four parameters and was fitted to a wide variety of experimental measurements to reproduce successfully the patterns of axon growth of a number of different neuron types. Our model is based on experimental evidence from a database of anatomical and electrophysiological information which has been collected over 30 years (University of Bristol, Alan Roberts and Steve Soffe Lab). This information has now allowed us to construct a connectome using anatomical data to allocate positions of cell bodies, dendrites, and places where axons originate, and then to use an algorithm to grow the axons and to specify synaptic connections where axons cross dendrites. An important feature of this approach to axon growth is that these models are biologically realistic; statistical characteristics of the axons generated are not distinguishable from experimental measurements. Repetition of the axon growth results in generation of the complete connectome. The total modeled neuronal network contains nearly 2,000 neurons of six types with around 120,000 synapses. The anatomical model contains random components; therefore, each repetition of the reconstruction procedure generates a connectome which differs from others, though all networks show common features.

Studying and modeling the connectome of the tadpole spinal cord is interdisciplinary and involves diverse aspects from informatics, mathematics, and biology. In this paper we focus on methodological aspects of modeling the connectome which are relevant to neuroinformatics. Starting from our accumulated experimental data, we present a theoretical study in which we consider axon growth modeling, our reconstruction algorithm and analysis of the resulting connectome before providing a preliminary assessment of the possible significance of the connectome for understanding the functioning of the tadpole spinal cord. We are not, at this stage, describing a functional model of the spinal cord.

MATERIALS AND METHODS

EXPERIMENTAL MEASUREMENTS FOR BUILDING A CONNECTOME MODEL

The anatomical details used to inform the connectome modeling were obtained using a variety of techniques. These details have been published previously but, since they are fundamental to construction of the connectome, they are briefly reviewed here. Most measurements were made in isolated nervous systems of chemically fixed tadpoles where individual neurons had been filled with neurobiotin. Filling was done using whole-cell recording electrodes. After the recording, tadpoles were fixed and processed to show the neurobiotin. The CNS was then removed and mounted between glass coverslips so that neurons could be viewed, traced, and photographed from either side at ×200 on a bright-field microscope (Li et al., 2002). This allowed identification of different types of neuron using their anatomical features and measurement of the dorso-ventral and longitudinal positions of their cell bodies, dendrites, and axonal processes when viewed from the side (Li et al., 2001, 2007). By joining together the left and right lateral views of the spinal cord along their ventral edge, we obtain views and data on neurons in a two-dimensional plan view of the CNS as if it had been opened along its dorsal midline (**Figure 1A**). Additional information was obtained by backfilling groups of neurons following application of the marker HRP to muscles or CNS and allowing it to travel up the axons to label cell bodies and dendrites. Some information was also obtained following immunocytochemical labeling for transmitters (glycine or GABA) or transcription factors (*engrailed*; Roberts et al., 1988; Roberts, 2000; Li et al., 2004).

Longitudinal distributions of somata for the populations of the main spinal cord neuron types are given in Yoshida et al. (1998),

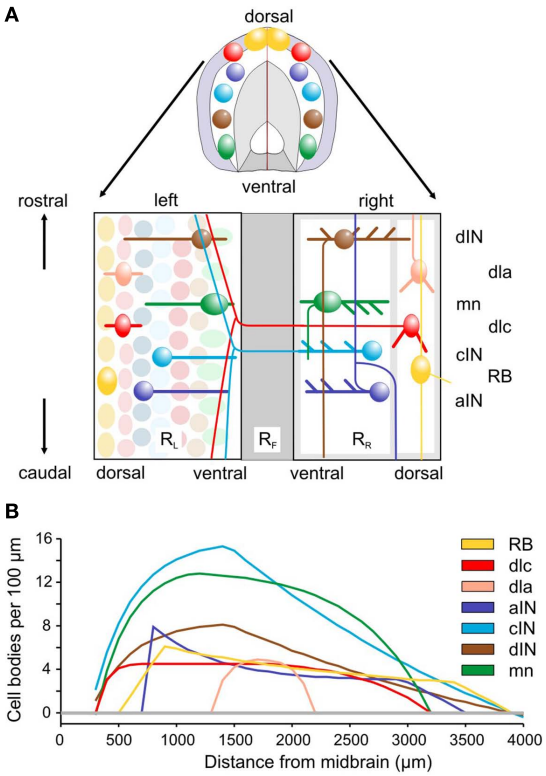


FIGURE 1 | Spinal cord anatomy and experimental measurements. (A) A short length of spinal cord shown in section and after cutting down the dorsal midline and opening flat. Rectangles containing neurons represent the two sides (R_L and R_R), separated by the ventral floor plate (dark gray rectangle, R_F). Examples of the cell body positions (ellipses), dendrites (thick lines), and axon projections (thin lines) are illustrated. Neuron types are listed on the right: RB, Rohon Beard sensory neuron; dla, dlc, dorsolateral ascending and dorsolateral commissural sensory interneurons; dIN, cIN, aIN, descending, commissural, and ascending premotor interneurons; mn, motoneurons. **(B)** Longitudinal distributions of neuron cell body numbers (per 100 µm). The curves show smoothed, theoretical distributions based on current anatomical estimates and updated from Li et al. (2001). The color coding indicated for each cell type is used throughout this paper.

Roberts et al. (1999), and Li et al. (2002) and summarized in **Figure 1B**. The best estimates are based on immunocytochemistry (aIN and cIN). Other estimates are based on counts of neurons labeled anatomically following HRP backfilling (dIN, dlc, and MN), or following intracellular labeling with neurobiotin (dla). In each case the distribution is approximated by simple equations describing the numbers of cell bodies per 100 μm lengths of each side of the spinal cord.

As an initial measure of dendrite distributions, the dorsal and ventral-most extents of dendritic branching have been measured for a sample of neurons of each type. For each neuron, the dendritic field was assumed to extend evenly between these dorso-ventral extremes. Measurements of individual fields were then combined to provide an overall dorso-ventral distribution for each neuron type (Li et al., 2007).

SIMPLE MODEL OF AXON GROWTH

In this Section we consider our simple mathematical model of axon growth (Li et al., 2007; Borisyuk et al., 2008). This model has been studied in detail and has been used here for generation of the connectome model of the whole spinal cord. For the convenience of the reader we include here a brief review of this simple model. For modeling axon growth, the tadpole spinal cord is considered as a horizontal cylinder, opened along the top (i.e., the most dorsal position) with, on each side, a very thin 10 μm layer (the marginal zone) separated ventrally by a further strip, the floor plate. We neglect the third dimension corresponding to the thickness of the marginal zone and, for modeling, consider the two sides of the spinal cord as two rectangles connected by the ventral floor plate (**Figure 1A**). Thus, positions within each rectangle are labeled by coordinates (x, y), where x corresponds the longitudinal axis giving rostro-caudal (RC) position along the body and y corresponds to the vertical, dorso-ventral (DV) axis. The size of each rectangle is standardized: $600 < x < 4{,}000, 0 < y < 100$ (these units are given in micrometers). It is important to note that, for the young developing tadpole, this length corresponds to the majority of the CNS controlling motor behavior; it only omits more rostral regions concerned with some additional sensory functions.

A simple mathematical model is described by a system of three difference equations (i.e., time is discrete) corresponding to the growth angle θ and two positions of the growth cone, the structure which forms the tip of the growing axon and which characterizes the direction of growth. The rostro-caudal (RC) horizontal position (coordinate) is denoted by the x and the dorso-ventral (DV) vertical coordinate is denoted by y. These equations describe in mathematical notations a simple rule of axon growth: if the number of iterations is large then the growth angle tends to zero (flattening the axon trajectory) therefore the axon tends to grow horizontally at a prescribed DV position which is denoted by $\bar{y}$. The model also includes a random variable which perturbs the growth angle at each step of iterations (Li et al., 2007; Borisyuk et al., 2008).

The model equations are:

$$\begin{aligned} x_{n+1} &= x_n + \Delta\cos(\theta_n) \\ y_{n+1} &= y_n + \Delta\sin(\theta_n) \\ \theta_{n+1} &= (1-\gamma)\theta_n + \mu(y_n - \bar{y}) + \xi_n \end{aligned} \quad (1)$$

where, x_n is the RC coordinate of the current axon position at the nth iteration, y_n is the DV coordinate of the current axon position; θ_n is the current growth angle. Parameters are: γ $(0 < \gamma < 1)$, μ $(\mu > 0)$, $\bar{y}$ $(0 < \bar{y} < 100)$. The random variable ξ_n is independent and uniformly distributed in the interval $(-\alpha, \alpha)$, where α is a parameter. Parameter Δ is the elongation of the axon in one iteration ($\Delta = 1$ μm). The parameter γ characterizes the tendency of an axon to grow straight in a horizontal direction. The parameter $\bar{y}$ represents the dorso-ventral position of an axon attractor and the parameter μ characterizes a rate of attraction.

This simple model is used below to reconstruct a connectome of the spinal cord. Parameter values were calculated using an optimization procedure to fit the model to available experimental measurements of both ascending and descending axons of different cell-types. The axon growth procedure does not start from the cell body so the initial portion of the axon is ignored (for details see Li et al., 2007 and Borisyuk et al., 2008).

RECONSTRUCTION ALGORITHM FOR A CONNECTOME MODEL

The connectome we describe includes almost all the spinal cord cell-types. Since these spinal neurons form populations that extend uninterrupted into the caudal part of the hindbrain, and since this caudal part of the hindbrain is needed for generating sustained locomotion (Li et al., 2006), we have included it in what we refer to as the spinal cord connectome. To model the connectome of the young tadpole spinal cord, we have used the simple axon growth model (see above) together with a new network reconstruction algorithm. Unlike the one reported previously (Borisyuk et al., 2008), networks on both sides of the spinal cord are reconstructed and biological data from experimental measurements have been used to assign distributions of cell bodies, dendrites, and axon lengths. Modeling the connectome began by assigning cell body distributions along the spinal cord; dendrites were added to these cell bodies and lastly axons were grown using our simple growth model.

The procedure for assigning cell body distribution is applied independently to left and right side of the CNS. The region of CNS being considered here is defined by RC coordinates from 600 to 4000 μm (measured from the tadpole midbrain; the region from 600 to ∼850 μm is caudal hindbrain, see above). The length of this region we divide to small spatial steps $D = L/m$, where L is the total length of the spinal cord ($L = 3{,}400$ μm) and m is the number of spatial steps (in our model the spatial step or a gap between cell bodies is $D = 1.5$ μm). The spinal cord is divided lengthwise into 34 subintervals (100 μm each) and the cell body distribution procedure is executed independently for each subinterval. Each subinterval has $100/D$ positions where neurons can be allocated; some of these positions will be empty. The goal is to allocate the neurons of different types in a random order within each subinterval keeping the prescribed number of neurons of each type (in fact, taken from experimental measurements). To reach the goal the following procedure is executed for each subinterval.

Suppose that cell-types are labeled by numbers: 1 – means RB cell-type, 2 – dlc cell-type, 3 – aIN cell-type, 4 – cIN cell-type, 5 – dIN cell-type, 6 – mn cell-type. First of all the number of cells of each type (1–6) inside the subinterval is calculated according

to the longitudinal distribution of neuron cell body numbers (per 100 μm) which are denoted by n_1,. . .,n_6. Thus, neurons of cell type 1 (i.e., RB) are allocated at the first n_1 positions with a gap D between neurons. Neurons of cell type 2 (i.e., dlc) are allocated at the next n_2 positions with a gap D between neurons, etc. There will be some empty spaces inside the subinterval because the number of possible positions is larger than the total number of neurons inside the subinterval. After that the positions are shuffled. It means all pairs of consecutive positions (from low to higher coordinate) are swapped. More precisely, consider a pair of consecutive positions, e.g. $(i, i+1)$ where the position i is either occupied by a neuron of cell-type Q_i or is empty and the $(i+1)$ position is either occupied by the neuron of cell-type Q_{i+1} or is empty. Swap means that position i is occupied by either the neuron of cell-type Q_{i+1} or is empty and the position $(i+1)$ is occupied either by the neuron of cell-type Q_i or is empty. This shuffling procedure is repeated 10,000 times to provide a random uniform distribution of neurons and empty positions of different cell-types inside the subinterval.

After filling all subintervals with appropriate numbers of neurons, the total number of neurons of each cell type can be calculated. These numbers, shown for one side of the body in **Table 1**, are then used to independently and randomly populate the other side of the spinal cord. Thus, the total number of neurons on one side is 944 and the total number of neurons in the spinal cord is 1,888. **Figure 2** shows an example of the longitudinal distribution of cIN cells along the left side of the spinal cord.

The dendrite of each neuron is allocated the same RC position as its cell body. The dendrite is represented by a vertical bar, the coordinates of whose ventral and dorsal extremes are randomly selected within the overall DV interval of 100 μm. Values are based on experimental measurements that provide distributions of pairs of low (l) high (h) DV coordinates for the dendrites of each cell-type: (l_k, h_k), $k = 1,2,. . .K$, where K is the total number of dendrite measurements for a given cell-type. These data are used to generate a random distribution of dendrite DV extents for each cell type.

Cell body and dendrite distributions are summarized in a text file. In this file, numbers of neurons on the left side range from 1 to 944 and the numbers of neurons on the right side range from 945 to 1,888. For each neuron the cell type is given as well as the rostro-caudal coordinate of the cell body (in a range from 600 to 4,000 μm). It is assumed that the RC coordinate of the cell body coincides with the RC coordinate of the dendrite. The DV position of the dendrite is represented by the most dorsal coordinate of the dendrite in μm and the most ventral coordinate of the dendrite in μm (in a range from 0 to 100). It should be noted that neurons of RB cell-type have no dendrites and, therefore, dorsal and ventral positions are set to zero.

Table 1 | Total neuron numbers per side.

Cell type	#
1 (RB)	107
2 (dlc)	86
3 (aIN)	97
4 (cIN)	271
5 (dIN)	135
6 (mn)	248

GENERATION OF AXONS AND SYNAPTIC CONTACTS

Once the neuron cell bodies and dendrites are positioned, the process of axon generation starts and axons are grown for each neuron. All axons are generated sequentially and independently from each other using our simple model of axon growth (see above; Li et al., 2007; Borisyuk et al., 2008).

To start growth of an axon the dorso-ventral coordinates of the initial point (x_0, y_0) and the initial angle of the axon growth (θ_0) are specified. After that, by applying the iterative formulas (1), the first point of axon (x_1, y_1) will be generated and continuing these iterations the whole axon will be generated: (x_i, y_i), $i = 0,1,2. . .,k$, where k is the total number of points of the generated axon. In fact, the length of the generated axon is a random number where the shape of the random distribution is derived from the experimental measurements. Thus the number k corresponds to the selected axon length.

In the current version of the connectome reconstruction model, axon growth starts from a point whose RC coordinate is the same as the RC coordinate of the cell body. For commissural neurons (dlc and cIN) the initial stage of axon growth from the cell body to the ventral floor plate, crossing the floor plate and growing up the opposite side of the spinal cord to a branching point is not considered. Instead, the initial point of axon growth has the same RC coordinate as the cell body but on the opposite side. The DV coordinate of the start point, the initial angle of axon growth, and the length of generated axon are randomly selected based on distributions from experimental data (for details see Li et al., 2007). Equation 1 describe an iterative process of axon growth from the start point continuously to the prescribed the axon length. Each iteration increases the axon length by 1 μm. Because the iterative process includes a random component, different axons

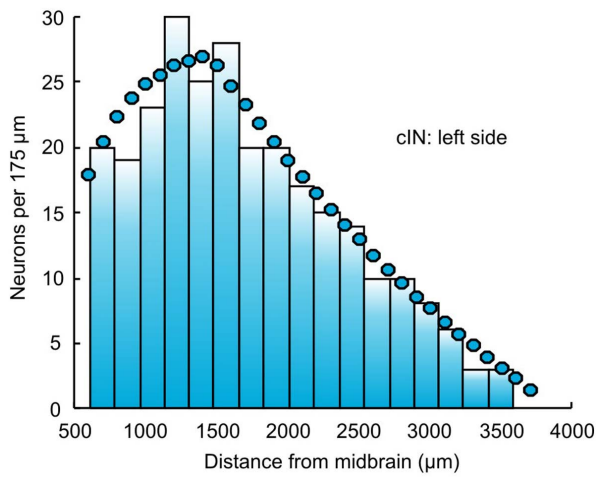


FIGURE 2 | The distribution of cIN neurons along the left side of the body. Bars indicate the connectome distribution (bin width is 175 μm); circles indicate the experimental estimate.

starting from the same initial conditions have different trajectories. Thus, each run of the connectome model generates a slightly different connectome. However, all generated connectomes have similar properties. The axon data are stored in a further text file. **Figure 3** compares the DV distributions of experimentally measured axons (from Li et al., 2007) with those for generated axons in the connectome.

Once the axons are grown for all the neurons, the connectome is completed by assigning the locations of synapses. Every time that a growing axon crosses the dendrite of another neuron, a synapse can form with a probability 0.46 (this value is based on experimental observations: Li et al., 2007). Information about all synapses is recorded to a text file. Thus, the connectome data are distributed between three text files which contain information on cell bodies and dendrites, axons, and synapses.

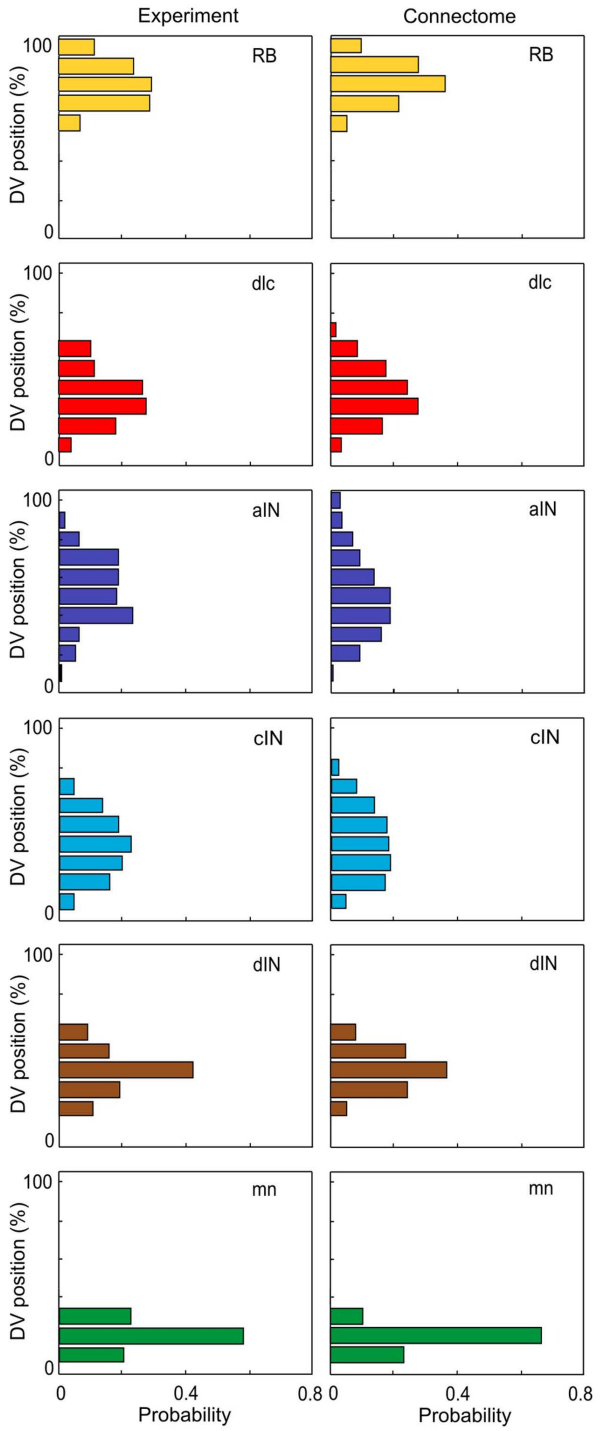


FIGURE 3 | Distributions of axon projections in the DV axis. There is a good match between distributions from experimental measurements (left, based on Li et al., 2007) and the connectome (right). In each case DV is divided into 10 bins.

RESULTS

VISUALIZING THE TADPOLE SPINAL CORD CONNECTOME

The connectome model shows a complex structure of connections. In this section we first present some approaches to visualization of the connectome. We then present some analysis to allow us to illustrate the kinds of information that the connectome model yields. In considering the spinal cord, some particular features of the connectome should be noted. For example, the fact that neurons are distributed on two sides of the spinal cord is very important; the cell bodies of dlc and cIN cell-types are located on one side but their axons project to the opposite side. The neurons of these cell-types therefore make their connections onto neurons on the opposite side. **Figure 4A** shows a fragment of the connectome for the left side and shows cell bodies, dendrites, axons, and synapses. It is clear that the dorsal part of the spinal cord (top in **Figure 4A**) visually appears predominantly "yellow" due a high number of RB axons. Similarly, the ventral part (bottom part of **Figure 4A**) visually appears predominantly "green" due to the fact that axons of motor neurons are allocated there. **Figure 4B** shows a close-up of the same fragment. It is clear that visualization of the connectome is a difficult problem and we plan to address this problem in more detail in a separate publication. All figures and graphs presented here are derived from one particular generated connectome because, although each reconstruction results in a new connectome, the statistical characteristics of each connectome are the same.

Schematically the connectome can be described in the following way. There are two networks, one on each side of the CNS: left side and right side networks. Each network contains 944 neurons of six cell-types. The left side network contains non-commissural neurons with mutual connections and commissural neurons which also receive connections from non-commissural neurons but send their connections to the right side network. The right side network has the same organization. Thus, the connectome model contains left side and right side networks of mutually coupled excitatory (RB, dIN, mn) and inhibitory (aIN) neurons and these two networks are connected by inhibitory commissural neurons (dlc and cIN cell-types).

Each side of the connectome contains about 30,000 internal connections (both excitatory and inhibitory synapses) and there are about 30,000 inhibitory connections from left to right and the same from right to left. One way to visualize the pattern of outgoing synapses from the axons of neurons in the connectome is shown in **Figure 5**. This method of visualization summarizes the

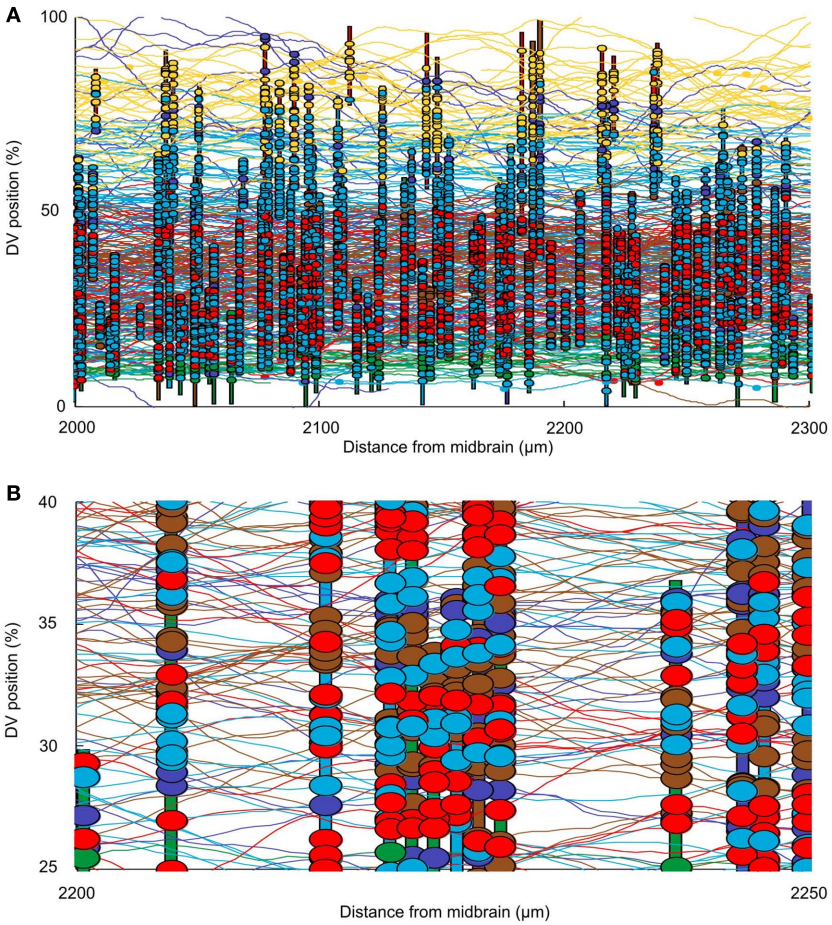


FIGURE 4 | Visualizing the connectome. (A) A 300-μm long region of the connectome showing the network on one side of the spinal cord. Note: this is ∼10% of the longitudinal extent of the model. Vertical colored bars indicate the dendrites of individual neurons; horizontal curved lines indicate the trajectories of grown axons. Dendrites and axons are color coded for cell type as in **Figure 1**. Small colored ellipses indicate individual synaptic contacts. In each case, the contact is color coded according to the presynaptic cell type. **(B)** A small fragment of one side of the connectome, color coded as in **(A)**.

spatial distribution of synapses from the axons of individual neurons of each type along the length of the spinal cord. For example: it highlights a relatively local range of ascending and descending connections from dlc neurons compared with aINs (**Figure 5A,B**). A similar contrast can be made between the synapses from the descending axons of dINs and mns (**Figures 5C,D**). However, it is clearly difficult to resolve patterns of individual connections and again illustrates difficulties in visualizing the connectome.

A second way to visualize the synapses in the connectome is from the perspective of the post-synaptic neurons, by mapping the incoming synapses onto their dendrites (**Figure 6**). This approach allows us to give a clear impression of the spatial distribution of incoming synapses. For example, it shows how synapses from cINs (on the opposite side) and dINs onto mns are generally more dense on neurons located rostral to ∼2,500 of the RC coordinate (**Figures 6A,B**). Also, while synapses between some neuron pairs span much of the DV extent of the dendrites, others (like those from cINs or dlcs onto other dlcs) are restricted to less dorsal dendrites (**Figures 6C,D**).

ANALYZING THE STRUCTURE OF THE CONNECTOME

The total number of connections in the connectome is 122,221 of which 51,530 are ascending and 70,691 are descending. The distribution of these connections between neuron types is illustrated in **Figure 7** and listed in **Table 2** The largest number of outgoing connections (summation by rows in **Table 2**) is 44,221 made from cINs to other neurons (i.e., the number of synapses which are located on axons of cIN neurons). Among these connections from cIN neurons, the largest number (15,798) are onto other cIN neurons (on the opposite side). The next largest number (12,056) of connections from cIN are onto dendrites of motoneurons. Overall, the smallest number of outgoing connections is 5,613 from RB to neurons of other neuron types.

To analyze the pattern of incoming synaptic connections, we examined the distances from which particular neuron types receive incoming synapses from other particular types. These distances were calculated by the following procedure. A neuron of cell type $p(p = 1,2,...6)$ was fixed. The longitudinal (RC) distances of all presynaptic neurons of cell type $q(q = 1,2,...6)$ with sufficiently long axons to contact the fixed neuron were then calculated. We

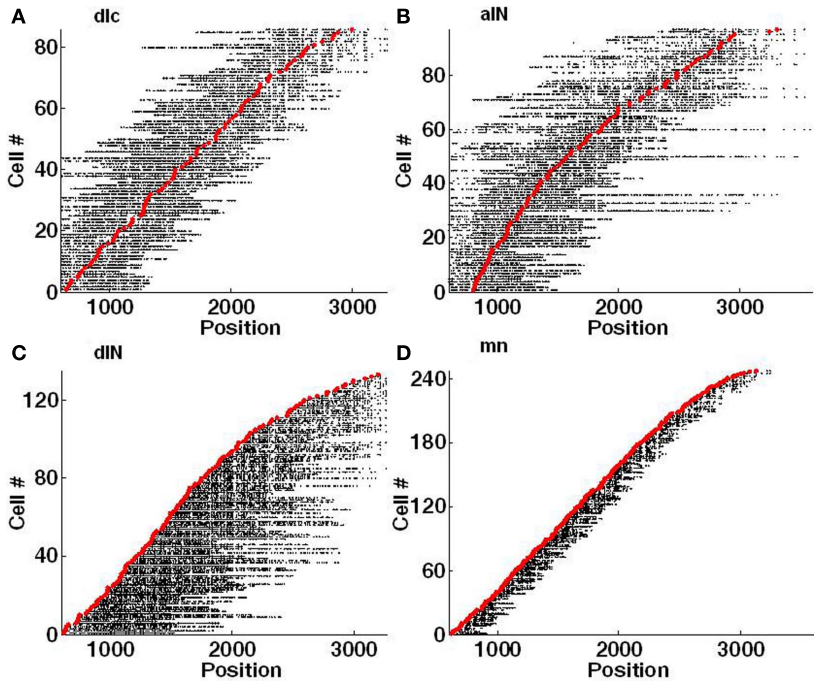


FIGURE 5 | Visualization of individual axons with synapses. Each red marker shows the RC position of a soma and each horizontal line of black dots indicates the RC positions of individual synapses on each single axon. Synapses on the left of the soma are located on an ascending axon (toward the head) and synapses on the right side are located on a descending axon (toward the tail). **(A)** dIc. **(B)** aIN. **(C)** dIN; note that the ascending axons present on a few spinal cord dINs were not included in this connectome. **(D)** mn.

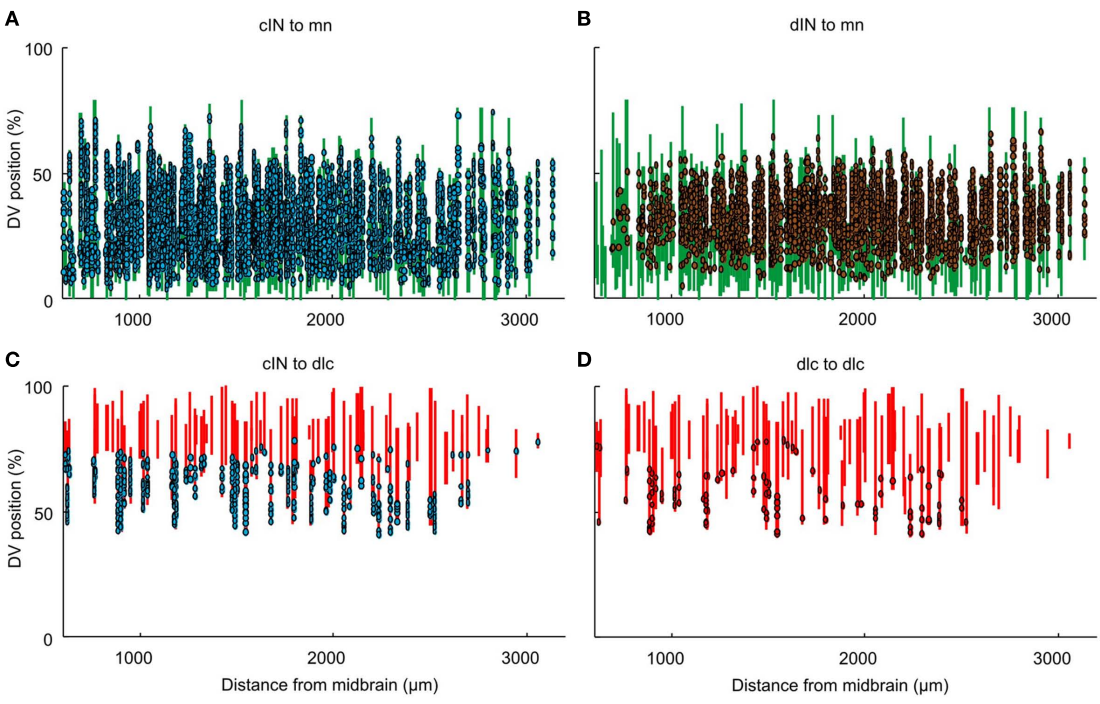


FIGURE 6 | Visualization of individual dendrites with synapses. Each vertical line of the figure corresponds to an individual dendrite. Dots indicate the DV positions of individual synapses. Dendrites and presynaptic contacts (dots) are color coded as in **Figure 1**. See text for details. **(A)** from cIN to mn. **(B)** from dIN to mn. **(C)** from cIN to dIc. **(D)** from dIc to dIc.

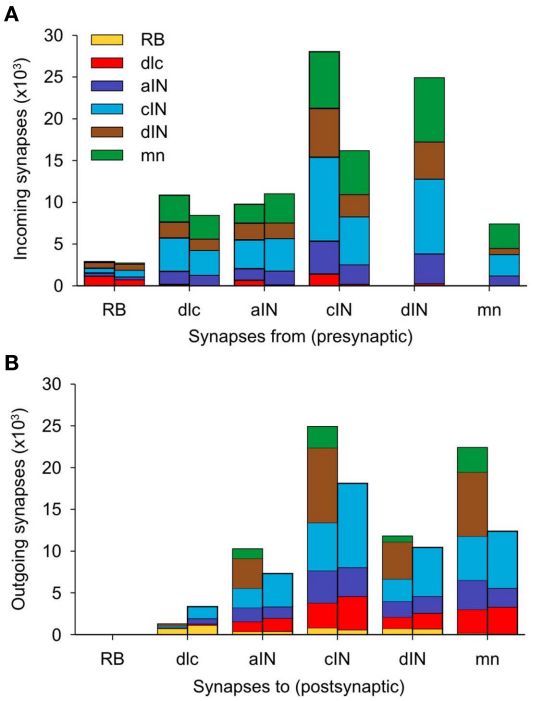


FIGURE 7 | Distribution of synapses in a single connectome model. (A) For each neuron type, the two stacked columns show outgoing synapses onto all other neuron types; post-synaptic neuron types (color coded as in **Figure 1A**) receive incoming synapses from descending axons (right columns) and ascending axons (left columns). **(B)** Like **(A)**, but each stacked column shows the distribution of incoming synapses from all other cell-types, made via descending axons (left columns) and ascending axons (right columns). See **Table 2** for data.

Table 2 | Distribution of connections between all types of presynaptic neurons and all post-synaptic neurons in the connectome.

Presynaptic neurons	Post-synaptic neurons						Total from
	RB	dlc	aIN	cIN	dIN	mn	
RB	0	1,842	744	1,373	1,408	246	5,613
dlc	0	187	2,780	6,982	3,266	6,057	19,272
aIN	0	757	3,044	7,336	3,879	5,769	20,785
cIN	0	1,585	6,275	15,798	8,507	12,056	44,221
dIN	0	247	3,576	8,955	4,447	7,706	24,931
mn	0	0	1,169	2,573	718	2,939	7,399
Total to	0	4,618	17,588	43,017	22,225	34,773	122,221

Table 3 | Probabilities of connections between all types of presynaptic neurons and all post-synaptic neurons in the connectome.

Presynaptic neurons	Post-synaptic neurons					
	RB	dlc	aIN	cIN	dIN	mn
RB	0.00	0.33	0.13	0.24	0.25	0.04
dlc	0.00	0.01	0.14	0.36	0.17	0.31
aIN	0.00	0.04	0.15	0.35	0.19	0.28
cIN	0.00	0.04	0.14	0.36	0.19	0.27
dIN	0.00	0.01	0.14	0.36	0.18	0.31
mn	0.00	0.00	0.16	0.35	0.10	0.40
total	0.00	0.04	0.14	0.35	0.18	0.28

did not take into account distances between sides for neurons with commissural axons. Each histogram shows the distribution of distances from pre- to post-synaptic cells for neurons of cell-types p, q($p = 1,2,\ldots 6$, $q = 1,2,\ldots 6$). These histograms show a variety of different distributions (**Figure 8**). The distribution of synapses from aIN to mn shows a simple distribution in which most connections are from neighboring neurons with numbers from more rostral and more caudal aINs decreasing with distance (**Figure 8A**). In contrast, connections onto dlcs from other dlcs at the same longitudinal position (though with cell bodies on the opposite side) are absent. Instead, connections, while relatively low overall, are broadly bimodal, being highest from some distance rostrally and caudally (**Figure 8B**). In a third example, connections from cINs to dlcs (again on opposite sides) are relatively few and are also roughly bimodal, but with an even stronger proportion of connections at some distance and particularly from more caudal cINs (**Figure 8C**). These types of differences will arise when axons change their DV position as they grow away from the cell body or have different DV positions when they grow in different directions. Synapses from dINs to mns are exclusively descending and decrease with distance (**Figure 8D**).

As well as examining the patterns of incoming and outgoing synapses, we have used the connectome to explore the probabilities of connection between neurons of different types (**Table 3**). As an example, the first row shows the probabilities of connections from RB to all other cell-types (dlc, aIN, cIN, dIN, mn). These range from 0.33 (connections to dlc) down to 0.04 (connections to mn). Remarkably, contact probabilities derived from the connectome model correlate well (Pearson correlation coefficient 0.62; $p < 0.001$) with previous values obtained from experimental pairwise recordings of connections between different cell-types (Li et al., 2007). This correlation supports the view that the connectome model is biologically realistic.

GENERAL PROPERTIES OF THE TADPOLE SPINAL CONNECTOME

In addition to details of the distributions of neurons, their axons and particularly their synaptic connections, the connectome model can also be analyzed to allow comparison with connections found in other systems.

Each part of the connectome on either side of the CNS can be characterized as a scale-free network. This property was shown by calculating the degree distribution of the connectome in which the nodes are individual neurons. Each node's degree is defined as the total number of connections (incoming plus outgoing) for each node. The degree distribution of a scale-free network follows a power law. **Figure 9A** shows the distribution of node degree for the whole connectome. It is clear that this distribution is very different from the power law. However, the distribution for one side

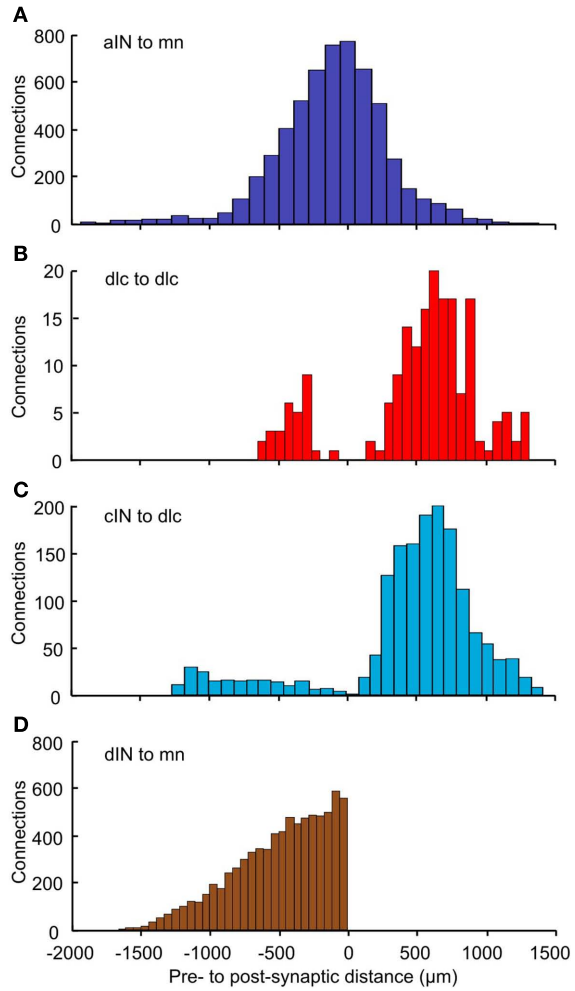


FIGURE 8 | Distributions of the number of connections as a function of distance. For each pair of neuron types, distance zero is the soma position of the post-synaptic neurons; the bars show the relative positions of the presynaptic neurons. Connections are shown: **(A)** from aIN to mn; **(B)** from dlc to dlc; **(C)** from cIN to dlc; **(D)** from dIN to mn. Colors (as in **Figure 1**) indicate the presynaptic neuron type. Bin widths: **(A)** = 111 μm; **(B)** = 67 μm; **(C)** = 90 μm; **(D)** = 55 μm.

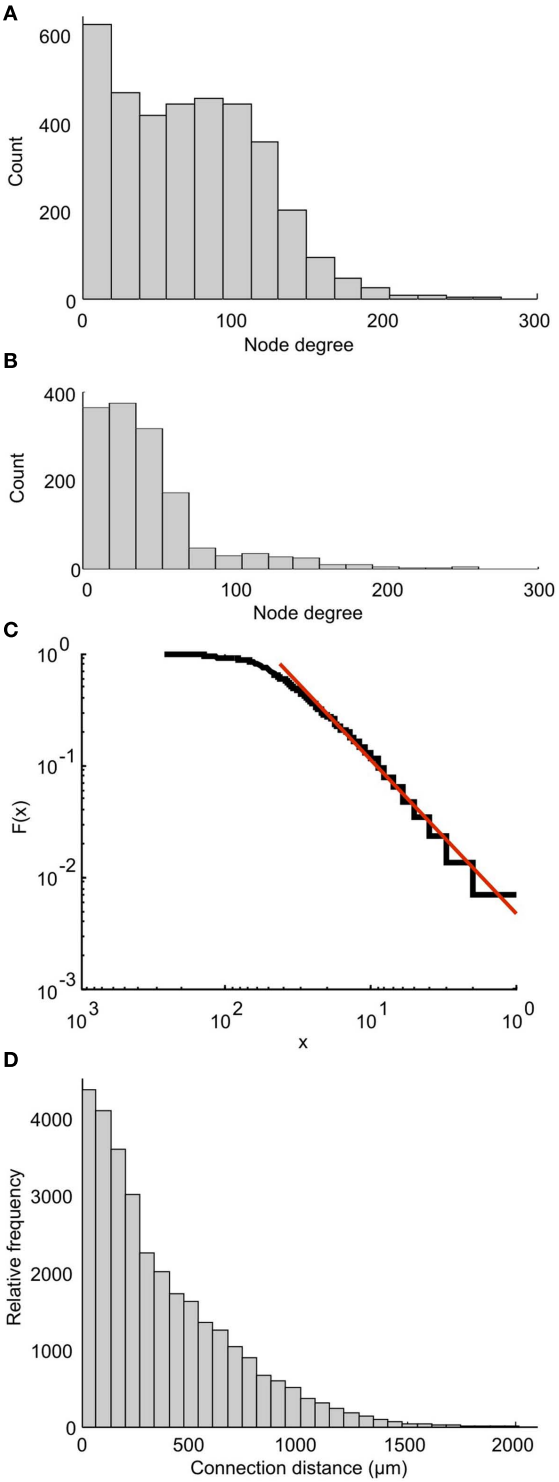


FIGURE 9 | Properties of the connectome. Node degree distributions for the whole connectome **(A)** and one side of the connectome **(B)** show that only the network on one side shows a distribution similar to the power law, and is therefore scale-free. **(C)** The cumulative function from **(B)** (blue) and a linear approximation (red) plotted on a log scale. **(D)** The overall distribution of connection distances for synapses between neurons on a single side of the tadpole connectome.

of the connectome (**Figure 9B**) is very similar to the power law. A plot of cumulative function (on a log scale) can be closely approximated by a straight line (**Figure 9C**), showing that the networks on each side of the tadpole spinal connectome are of the scale-free type. Including both sides of the spinal cord destroys the free-scale nature.

We have also analyzed the distribution of overall connection lengths, combining lengths from all neuron types and ignoring the direction of connection. It has recently been shown that this distribution is similar for different neural networks of different animals (Kaiser et al., 2009). In the tadpole spinal cord connectome, connection distances ranged from 1.5 to 1941 μm (**Figure 9D**). The distribution is similar to the connection length distribution for the large network of *C. elegans* (see Figure 2D of Kaiser et al., 2009).

Fitting a Gamma Distribution gives the shape parameter $a = 0.53$ and scaling parameter $b = 0.06$ for the tadpole (compared with $a = 0.541$ and $b = 0.419$ for *C. elegans*).

Lastly, the edge density (or connectivity) characterizes how many potential connections between network nodes are really present. For a directed network, each of N nodes can be connected with some other nodes where the number of other nodes ranges from 0 to $(N - 1)$. The formula for the edge density is $d = E/[N(N - 1)]$, where E is the number of directed connections present in the network; N is the number of nodes (neurons; Kaiser, 2011). The edge density of the tadpole connectome is 0.03 or 3%. This figure is comparable with the edge density of *C. elegans* (Kaiser, 2011), which is 3.85% (for corticocortical connections of the mammalian brain, the edge density ranges from 10 to 30%).

DISCUSSION

A key feature of the connectome modeling that we describe here is that it is based on detailed (though of course still not complete) knowledge of the neuronal components of the comparatively simple tadpole nervous system (Roberts et al., 2010). Furthermore, this nervous system belongs to a very simple animal at an early point in its development, but one with a well-defined set of simple behavioral responses. These features are important for several reasons. Firstly, we have already been able to base construction of the connectome on detailed biological data. Secondly, we can compare the connectivity with real patterns of synaptic contact between the component neuron types. Thirdly, we have the possibility of interpreting the connectome in terms of its known biological function(s) in generating behavior.

With the exception of *C. elegans* (White et al., 1986) and perhaps the crustacean stomatogastric nervous system (see: Harris-Warrick et al., 1992), it has not yet been possible to reveal a whole connectome purely through experimental analysis. The modeled connectome on the other hand allows us to consider features of the whole network as well as to peer in to look at individual neurons. Obvious values of the connectome model are as a source of predictions for future experiments and in perhaps revealing unexpected properties. For example, a striking difference was revealed in the patterns of longitudinal distance from which neurons can receive incoming synapses. Unexpectedly, and in contrast to other connections, the largest numbers of incoming synapses to dlc neurons from other dlc neurons or cIN neurons on the opposite side are not from neurons at the same longitudinal position but from some distance more caudal. The biological significance of this is not immediately clear, but it must now be considered. Testing predictions will in turn allow us to refine the model.

One simple prediction from the connectome is the number of synapses an individual neuron is likely to receive. In the case of motoneurons, this can be compared to an independent experimental estimate. Assuming an equal distribution of synapses, the connectome would predict an average number of synapses per mn as: total synapses onto mn (34,773) per total mn (496) = 70 synapses. An experimental estimate based on combined light and electron microscopy data estimated an overall average of 118–236 per mn depending on synapse spacing on each dendrite (Roberts et al., 1999). Two factors which could explain the lower prediction from the connectome are axons contacting the motoneurons from outside the region modeled (descending from the brain) or underestimating the extent of the mn dendritic tree (see below). Given the uncertainties, this reasonable match between prediction and observation reinforces our confidence that the connectome model is biologically realistic.

The tadpole connectome model describes neuronal networks connecting between the two sides of the CNS. Biologically, this interconnection underpins control of motor responses like rhythmic swimming which alternate on left and right sides (Roberts et al., 2010). An extension of the connectome model is therefore to use the connectivity revealed to produce a functional model of swimming. Although the details are beyond the scope of the present paper, it has now been possible to combine the network structure of the connectome with Hodgkin–Huxley-type model neurons to produce a functional network for swimming. Although at an early stage, this preliminary functional model is able to "swim," generating rhythmic firing of the relevant neuron populations in a sequence that alternates between sides and progresses down the body on each cycle, mimicking the properties of real swimming behavior.

Modeling the connectome of the tadpole is clearly an ongoing process and very much linked to accumulating the biological data that is needed to guide its construction. We are in the process of introducing what is the next stage in axon growth modeling, a gradient model of axon guidance. The importance of this new model will be its greater basis in biological reality, replacing rather artificial parameters of growth with more realistic responses to morphogen gradients. A further addition to be introduced when suitable biological data are available will be a more realistic description of the dendritic branches of the various spinal cord neuron types. It is likely that this will alter the DV pattern of connections, but it is too soon to predict whether the change will be rather subtle or more significant.

The construction and analysis of models of nervous system networks continues to be an area of huge interest (e.g., see: Bullmore and Sporns, 2009; Cahalane et al., 2011; Prettejohn et al., 2011). Many studies aim to understand the properties of highly complex areas of the nervous system, typically regions of the adult brain, where the detailed experimental information is still very limited and, of necessity, the analysis remains primarily theoretical. It is important to stress that the strength of our approach lies in the way we can use our considerable biological knowledge of a relatively simple system to grow our networks with realistic connectivity using simple rules of axon growth and synapse formation. This situation has allowed us to aim for a close match between the modeled network and real biology.

ACKNOWLEDGMENTS

This work is supported by BBSRC grants (BB/G006369/1 and BB/G006652/1).

REFERENCES

Borisyuk, R., Cooke, T., and Roberts, A. (2008). Stochasticity and functionality of neural systems: mathematical modelling of axon growth in the spinal cord of tadpole. *Biosystems* 93, 101–114.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–198.

Cahalane, D. J., Clancy, B., Kingsbury, M. A., Graf, E., Sporns, O., and Finlay, B. L. (2011). Network structure implied by initial axon outgrowth in rodent cortex: empirical measurement and models. *PLoS ONE* 6, e16113. doi: 10.1371/journal.pone.0016113

Grillner, S. (2003). The motor infrastructure: from ion channels to neuronal networks. *Nat. Neurosci.* 4, 573–586.

Harris-Warrick, R. M., Marder, E., Selverston, A. I., Moulins, M. (eds). (1992). *Dynamic Biological Networks: The Stomatogastric Nervous System.* Cambridge: MIT Press.

Kaiser, M. (2011). A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage* 57, 892–907.

Kaiser, M., Hilgetag, C., and van Ooyen, A. (2009). A simple rule for axon outgrowth and synaptic competition generates realistic connection lengths and filling fractions. *Cereb. Cortex* 19, 3001–3010.

Kiehn, O. (2011). Developmental and functional organization of spinal locomotor circuits. *Curr. Opin. Neurobiol.* 21, 100–109.

Li, W.-C., Cooke, T., Sautois, B., Soffe, S. R., Borisyuk, R., and Roberts, A. (2007). Axon and dendrite geography predict the specificity of synaptic connections in a functioning spinal cord network. *Neural Dev.* 2, 17.

Li, W.-C., Higashijima, S., Parry, D. M., Roberts, A., and Soffe, S. R. (2004). Primitive roles for inhibitory interneurons in developing frog spinal cord. *J. Neurosci.* 24, 5840–5848.

Li, W.-C., Perrins, R., Soffe, S. R., Yoshida, M., Walford, A., and Roberts, A. (2001). Defining classes of spinal interneuron and their axonal projections in hatchling *Xenopus* tadpoles. *J. Comp. Neurol.* 441, 248–265.

Li, W.-C., Soffe, S. R., and Roberts, A. (2002). Spinal inhibitory neurons that modulate cutaneous sensory pathways during locomotion in a simple vertebrate. *J. Neurosci.* 22, 10924–10934.

Li, W.-C., Soffe, S. R., Wolf, E., and Roberts, A. (2006). Persistent responses to brief stimuli: feedback excitation among brainstem neurons. *J. Neurosci.* 26, 4026–4035.

McLean, D. L., and Fetcho, J. R. (2007). Using imaging and genetics in zebrafish to study developing spinal circuits in vivo. *Dev. Neurobiol.* 68, 817–834.

Prettejohn, B. J., Berryman, M. J., and McDonnell, M. D. (2011). Methods for generating complex networks with selected structural properties for simulations: a review and tutorial for neuroscientists. *Front. Comput. Neurosci.* 5:11. doi: 10.3389/fncom.2011.00011

Roberts, A. (2000). Early functional organisation of spinal neurons in developing lower vertebrates. *Brain Res. Bull.* 53, 585–593.

Roberts, A., Dale, N., Ottersen, O. P., and Storm-Mathisen, J. (1988). Development and characterization of commissural interneurons in the spinal cord of *Xenopus laevis* embryos revealed by antibodies to glycine. *Development* 103, 447–461.

Roberts, A., Li, W.-C., and Soffe, S. R. (2010). How neurons generate behaviour in a hatchling amphibian tadpole: an outline. *Front. Behav. Neurosci.* 4:16. doi: 10.3389/fnbeh.2010.00016

Roberts, A., Walford, A., Soffe, S. R., and Yoshida, M. (1999). The motoneurons of the axial swimming muscles in hatchling *Xenopus* tadpoles: their features, distribution and central synapses. *J. Comp. Neurol.* 411, 472–486.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. B Lond. Biol. Sci.* 314, 1–340.

Yoshida, Y., Roberts, A., and Soffe, S. R. (1998). Axon projections of reciprocal inhibitory interneurons in the spinal cord of young *Xenopus* tadpoles and implications for the pattern of inhibition during swimming and struggling. *J. Comp. Neurol.* 400, 504–518.

Ziskind-Conhaim, L., Fetcho, J. R., Hochman, S., MacDermott, A. B., and Stein, P. S. G. (eds). (2010). *Neurons and Networks in the Spinal Cord.* Oxford: Blackwells.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 March 2011; accepted: 29 August 2011; published online: 23 September 2011.

Citation: Borisyuk R, al Azad AK, Conte D, Roberts A and Soffe SR (2011) Modeling the connectome of a simple spinal cord. Front. Neuroinform. ***5****:20. doi: 10.3389/fninf.2011.00020*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 24 January 2012
doi: 10.3389/fninf.2011.00035

Meeting the memory challenges of brain-scale network simulation

***Susanne Kunkel*[1,2]*, *Tobias C. Potjans*[3,4], *Jochen M. Eppler*[3], *Hans Ekkehard Plesser*[5,6], *Abigail Morrison*[1,2,6] *and Markus Diesmann*[3,4,6,7]**

[1] *Functional Neural Circuits Group, Albert-Ludwig University of Freiburg, Freiburg im Breisgau, Germany*
[2] *Bernstein Center Freiburg, Albert-Ludwig University of Freiburg, Freiburg im Breisgau, Germany*
[3] *Institute of Neuroscience and Medicine (INM-6), Computational and Systems Neuroscience, Research Center Jülich, Jülich, Germany*
[4] *RIKEN Computational Science Research Program, Wako, Japan*
[5] *Department of Mathematical Sciences and Technology, Norwegian University of Life Sciences, Ås, Norway*
[6] *RIKEN Brain Science Institute, Wako, Japan*
[7] *Medical Faculty, RWTH Aachen University, Aachen, Germany*

Edited by:
Claus Hilgetag, Jacobs University Bremen, Germany

Reviewed by:
Romain Brette, Ecole Normale Supérieure, France
Fernando Perez, University of California at Berkeley, USA

***Correspondence:**
Susanne Kunkel, Bernstein Center Freiburg, Hansastraße 9a, 79104 Freiburg im Breisgau, Germany.
e-mail: kunkel@bcf.uni-freiburg.de

The development of high-performance simulation software is crucial for studying the brain connectome. Using connectome data to generate neurocomputational models requires software capable of coping with models on a variety of scales: from the microscale, investigating plasticity, and dynamics of circuits in local networks, to the macroscale, investigating the interactions between distinct brain regions. Prior to any serious dynamical investigation, the first task of network simulations is to check the consistency of data integrated in the connectome and constrain ranges for yet unknown parameters. Thanks to distributed computing techniques, it is possible today to routinely simulate local cortical networks of around 10^5 neurons with up to 10^9 synapses on clusters and multi-processor shared-memory machines. However, brain-scale networks are orders of magnitude larger than such local networks, in terms of numbers of neurons and synapses as well as in terms of computational load. Such networks have been investigated in individual studies, but the underlying simulation technologies have neither been described in sufficient detail to be reproducible nor made publicly available. Here, we discover that as the network model sizes approach the regime of meso- and macroscale simulations, memory consumption on individual compute nodes becomes a critical bottleneck. This is especially relevant on modern supercomputers such as the Blue Gene/P architecture where the available working memory per CPU core is rather limited. We develop a simple linear model to analyze the memory consumption of the constituent components of neuronal simulators as a function of network size and the number of cores used. This approach has multiple benefits. The model enables identification of key contributing components to memory saturation and prediction of the effects of potential improvements to code before any implementation takes place. As a consequence, development cycles can be shorter and less expensive. Applying the model to our freely available Neural Simulation Tool (NEST), we identify the software components dominant at different scales, and develop general strategies for reducing the memory consumption, in particular by using data structures that exploit the sparseness of the local representation of the network. We show that these adaptations enable our simulation software to scale up to the order of 10,000 processors and beyond. As memory consumption issues are likely to be relevant for any software dealing with complex connectome data on such architectures, our approach and our findings should be useful for researchers developing novel neuroinformatics solutions to the challenges posed by the connectome project.

Keywords: brain-scale simulation, memory consumption, supercomputer

1. INTRODUCTION

There has been much development of the performance and functionality of neuronal network simulators in the last decade. These improvements have so far largely been aimed at investigating network models that are at or below the size of a cubic millimeter of cortex (approximately 10^5 neurons) and within the context of the current dominant laboratory high-performance computing paradigm, i.e., moderately sized clusters up to hundreds of nodes and SMP machines (e.g., Lytton and Hines, 2005; Morrison et al., 2005; Migliore et al., 2006; Plesser et al., 2007; Pecevski et al., 2009). However, there is a growing interest in performing simulations at the scale of multiple brain areas or indeed the entire brain. Firstly, the

predictive power of models of the local cortical circuit is severely limited. Such models can account for all the local synapses, but each neuron receives 50% of its synapses from external sources. Thus half of the inputs to the neurons remains unaccounted for, and must be replaced by external currents or random spike input. Secondly, such simulations are a necessary accompaniment to the development of the brain connectome (Sporns et al., 2005), as a structural description of the brain is by definition static. By performing dynamic simulations of the meso- and macroscale networks, we can check the consistency of the anatomical data, integrate it into models, identify crucial missing parameters, and constrain ranges for yet unknown parameters (Potjans and Diesmann, 2011). Thirdly, no part of the brain works in isolation; functional circuits are only closed at the brain-scale. Simulation studies of the interaction of multiple brain areas are necessary to understand how they coordinate their activity to generate brain function. Finally, experimentalists interested in brain function use imaging techniques such as fMRI and MEG and mass recordings such as the LFP to identify the brain's functional circuits and uncover the dynamics of the interaction between bottom-up and top-down processing. Models of corresponding size are required to create predictions for such measurements.

Brain-scale simulations will necessarily be orders of magnitude larger than the local network models discussed above, and will exceed the capacity of the small-to-medium sized clusters available to researchers within their own research facilities. Fortunately, massively parallel computing architectures installed at dedicated high-performance computing centers are becoming ever more available. In particular, the Blue Gene architectures, incorporating very large numbers of processors with moderate clock speed and relatively small amounts of RAM, provide computational power efficiently, and are consequently increasing in popularity. To exploit these new possibilities for brain-scale simulations, we need simulation software that will scale up to tens or hundreds of thousands of processes. Although very large scale networks have been previously investigated (e.g., Ananthanarayanan and Modha, 2007; Izhikevich and Edelman, 2008; Ananthanarayanan et al., 2009), the underlying simulation technologies have not been described in sufficient detail to be reproducible by other research groups, and so the value of these studies to the neuroscientific community remains somewhat limited.

A major challenge to the scalability of neuronal simulators on such architectures is the limited RAM available to each core. Data structure designs that are reasonable and efficient in the context of clusters of hundreds of processes may be insufficiently parallelized with respect to architectures that are two or three orders of magnitude larger. Insufficient parallelization results in substantial serial memory overhead, thus restricting the maximum network size that is representable on the architecture. We therefore conclude that a systematic approach to understanding memory consumption and designing data structures will be of benefit to any research team attempting to extend the scalability of a neuronal simulator, or indeed any other application, from moderate to very large cluster sizes.

In this manuscript we develop a technique for analyzing the memory consumption of a neuronal simulator with respect to its constituent components on the basis of a linear memory model, extending the approach presented in Plesser et al. (2007). We apply it here to the specific example of the freely available Neural Simulation Tool NEST (Gewaltig and Diesmann, 2007), however the principles are sufficiently general to be applied to other simulators, and with some adaptation to other distributed applications that need to store large numbers of objects. In particular, although we apply the model to a simulator designed for efficient calculation of point neuron models, the technique is equally applicable to simulators optimized for anatomically detailed multi-compartment neuron models. We demonstrate that the model allows the software components that are most critical for the memory consumption to be identified. Moreover, the consequences of alternative modifications to the design can be predicted. As a result of these features, development effort can be concentrated where it will have the greatest effect, and a design approach can be selected from a set of competing ideas in a principled fashion. Finally, we show that this technique enables us to remove major limitations to the scalability of NEST, and thereby increase the maximum network size that can be represented on the JUGENE supercomputer by an order of magnitude.

The conceptual and algorithmic work described here is a module in our long-term collaborative project to provide the technology for neural systems simulations (Gewaltig and Diesmann, 2007). Preliminary results have been already presented in abstract form (Kunkel et al., 2009).

2. MATERIALS AND METHODS

Typically, memory consumption in a distributed application is dependent not only on the size of the problem but also on the number of processes deployed. Moreover, the relative contributions of the various components of an application to its total memory usage can be expected to vary with the number of processes. To investigate this issue in a systematic way, in the following we develop a model which captures the contributions of the main software components of a neuronal network simulator to the total memory usage (Sec. 2.1). We applied the model to the Neural Simulation Tool NEST. Specifically, the parameterization of the model with theoretically determined parameters (Sec. 2.3) or empirically determined values (Sec. 3.1) is based on version NEST2.0-rc4 using Open MPI 1.4.3 on one core of a 12-core AMD Opteron 6174 running at 2.2 GHz under the operating system Scientific Linux release 6.0.

2.1. MODEL OF THE MEMORY USAGE OF A NEURONAL NETWORK SIMULATOR

Our primary assumption is that the memory consumption on each process is a function of M, the total number of processes:

$$\mathcal{M}(M,N,K) = \mathcal{M}_0(M) + \mathcal{M}_\mathrm{n}(M,N) + \mathcal{M}_\mathrm{c}(M,N,K) \quad (1)$$

where $\mathcal{M}_0(M)$ is the base memory consumption for an empty network, $\mathcal{M}_\mathrm{n}(M, N)$ is the memory consumed by N neurons distributed over M processes, and $\mathcal{M}_\mathrm{c}(M, N, K)$ denotes the additional memory usage that accrues for K incoming synaptic connections for each neuron.

The first term $\mathcal{M}_0(M)$ is a serial overhead that consists of the fundamental infrastructure required to run a neuronal network

simulator, i.e., the data structures for each process that support the essential tasks sustaining the simulation flow such as event scheduling and communication. It also includes the libraries and data buffers of the message passing interface (MPI) which are necessary for parallelization.

Assuming neurons are distributed evenly across processes and each neuron receives K incoming synapses, $N_M = N/M$ neurons and $K_M = N_M K$ synapses are represented on each process. We previously developed an approximation of the total memory consumption of a neuronal simulator based on the consumption of these objects (Plesser et al., 2007). However, additional data structures may be needed, for example to enable sanity checks during network construction and efficient access to neurons and synapses during simulation. In this manuscript, we will refer to the contribution of these data structures to the memory consumption as neuronal and connection infrastructure, respectively. These data structures are maintained on each process and are a source of serial overhead which increases proportionally with the network size N. Taking into consideration not only neurons and synapses but also the required infrastructures, the memory usage that accrues when creating and connecting neurons is given by:

$$\mathcal{M}_\mathrm{n}(M,N) = N m_\mathrm{n}^0 + (N - N_M)\, m_\mathrm{n}^\emptyset + N_M \left(m_\mathrm{n}^+ + m_\mathrm{n}\right) \quad (2)$$

$$\mathcal{M}_\mathrm{c}(M,N,K) = N m_\mathrm{c}^0 + N_\mathrm{c}^\emptyset m_\mathrm{c}^\emptyset + \left(N - N_\mathrm{c}^\emptyset\right) m_\mathrm{c}^+ + K_M m_\mathrm{c} \quad (3)$$

where m_n is the memory consumed by one neuron and m_c is the memory consumed by one synapse. The neuronal infrastructure causes m_n^0 serial overhead per neuron and, additionally, m_n^+ per local neuron and $m_\mathrm{n}^\emptyset$ per non-local neuron. The connection infrastructure causes m_c^0 serial overhead per neuron and, additionally, m_c^+ per neuron with local targets and $m_\mathrm{c}^\emptyset$ per neuron without local targets. Our model is based on the assumption of random connectivity. Specifically, each neuron draws K inputs from N possible sources, with repetition allowed. The probability that a neuron i does not select a given neuron j as one of its inputs is $(1 - 1/N)^K$. Therefore, the probability that a given neuron has no local targets on a specific process is given by $p_\emptyset = (1 - 1/N)^{K_M}$ which results in an expected number of $N_\mathrm{c}^\emptyset = p_\emptyset N$ neurons without local targets on each process. For a specific implementation of a neuronal network simulator, these expressions could be somewhat simplified; for example, the smaller of $m_\mathrm{n}^\emptyset$ and m_n^+ could be absorbed into m_n^0. However, in this manuscript we will be evaluating alternative designs with varying values for these components. We will thus use the full forms of (2) and (3) at all times to ensure consistency.

Our model does not account for any network topology or hierarchy, and it does not consider other types of nodes such as stimulating and recording devices, as the contribution of such devices to the total memory consumption is typically negligible.

2.2. IDENTIFICATION OF SIMULATOR COMPONENTS CONTRIBUTING TO MEMORY CONSUMPTION TERMS

Although the aim of this manuscript is to present a general method for analyzing the memory consumption of massively distributed simulation software and planning design alterations, to demonstrate the usefulness of the technique we apply it to the specific example of NEST (Gewaltig and Diesmann, 2007). **Figure 1** illustrates the key data structures in NEST that contribute to the parameters of the memory consumption model (2) and (3). Many other simulators are organized along similar lines (e.g., Migliore et al., 2006; Pecevski et al., 2009), however the precise data structures used are not critical for the method. A different design will simply result in different contributions to the memory model parameters.

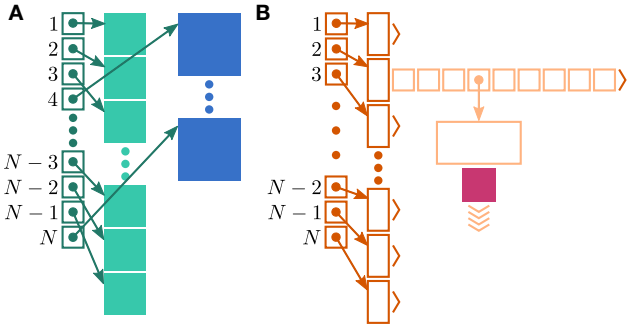


FIGURE 1 | Neuronal and connection infrastructure. The color of elements denotes which memory term they contribute to: m_n^0 (dark green), $m_\mathrm{n}^\emptyset$ (mid green), m_n (blue), m_c^0 (dark orange), m_c^+ (light orange), m_c (pink). **(A)** Neuronal infrastructure for $M = 4$. A `vector` of length N (dark green) is maintained of pointers to either local nodes (blue squares) or `Proxy` nodes (filled green squares). **(B)** Connection infrastructure. A `vector` of length N (dark orange) is maintained of pointers to initially empty `vectors` (dark orange boxes with single chevrons). If a node makes a synaptic connection to a target on the local machine, the length of the `vector` is extended to the number of synapse types available in the software (light orange boxes). A `GenericConnector` object is initialized for the type of the synaptic connection (large light orange box), and a pointer to it laid in the corresponding entry of the `vector`. The new synaptic connection (pink box) is stored in a `deque` (four chevrons) within the `GenericConnector`.

In NEST 2.0-rc4, from hereon referred to as the original implementation, the neuronal infrastructure is organized as a `vector` of length N on each process of pointers to local nodes and `Proxy` nodes which represent non-local neurons (see **Figure 1A**). This structure allows access to neurons on the basis of their unique global identifier, or GID. Access is required for functionality such as setting or querying their variables and connecting neurons. Getting and setting variables does not require the existence of `Proxy` nodes; assuming the simulation is described by a single serial instruction set, rather than a specific set for each process, all that is required is that each process does nothing when it reads the instruction to manipulate the variables of a non-local neuron. These nodes become important when creating connections, as a `Proxy` node contains information about the model of the non-local node it is representing. As not all models support all possible types of connection, by querying a `Proxy` node a process can determine whether a connection specified in its instruction set is valid. In addition to the persistent pointer to each node shown, a second persistent pointer exists that allows a hierarchy of sub-networks to be maintained. It is omitted from all diagrams in this manuscript for clarity. Simulation of the network relies on an additional `vector` of persistent pointers to local nodes. Thus m_n^0 comprises the two persistent pointers used for access, m_n and $m_\mathrm{n}^\emptyset$ consist of the

size of a local node and the size of a `Proxy` node, respectively, and m_n^+ is the persistent pointer to local nodes used during simulation.

In NEST, synapses are stored on the same process as their post-synaptic target. When a neuron spikes, its GID is communicated at the next synchronization point to all processes, which activate the corresponding synapses. Certainly for clusters on which the number of machines is less than the average number of post-synaptic targets, this structural organization reduces the bulk of the data to be communicated (Morrison et al., 2005). In the original implementation, the connection infrastructure which mediates the local dispatch of spikes is provided by a `vector` of length N on each process, as shown in **Figure 1B**. The contents of the `vector` are pointers to further structures which contains the synapses sorted by type. The inner structure is a `vector` which is initially empty; the first time a neuron is connected to a local target, the `vector` is extended to the length of the total number of distinct built-in and user-defined synapse types. The contents of the `vector` are pointers to `GenericConnector` objects which contain `deques` of synapses of their respective types. This `vector` of `vectors` allows the synapses of a particular source neuron, and synapses of a particular type, to be located very quickly. A `GenericConnector` object for a given synapse type is only instantiated if a neuron actually has synapses of that type. Consequently, the pointer to a `vector` and the empty inner `vector` contribute to m_c^0, m_c is the size of a single synapse and m_c^+ comprises the pointers to `GenericConnector` objects, the objects themselves and their corresponding empty `deques`. As there is no representation of non-local synapses, $m_c^\emptyset$ is zero.

Naturally, other simulation software may not have the constraints described above, or may have additional ones. For example, if all node types support all event types, there is no need to perform validity checking of connections, and the data structures chosen will reflect that. Similarly, if a simulation software has only one type of synapse, no structures are needed to separate different types. Independent of the similarity of the data structures of a specific application to those described above, a careful classification of objects as contributory terms to the different model parameters must be carried out.

2.3. THEORETICAL DETERMINATION OF THE MODEL PARAMETERS

We can determine the memory consumptions m_n and m_c directly by counting the number and type of data they hold. This can either be done by hand, for example by counting 7 variables of type `double` with 8 B per `double`, or by using an inbuilt `sizeof()` function. All neurons derive either directly from the base class `Node` that takes up 56 B of memory or from the intermediate class `ArchivingNode` that uses at least 184 B; the greater memory consumption in the latter case is mainly due to a `deque` that stores a certain amount of the neuron's spiking history in order to efficiently implement spike-timing dependent plasticity (Morrison et al., 2007).

Many types of neurons and synapses are available within NEST. They vary in the number of state variables and parameters that need to be stored. Hence, m_n varies across models. In this manuscript, we will use a leaky integrate-and-fire neuron model with alpha-shaped post-synaptic currents (NEST model name: `iaf_neuron`), which consumes $m_n = 424$ B. On the 64-bit architecture chosen for our investigations, a pointer consumes 8 B, therefore $m_n^0 = 16$ B (two persistent pointers) and $m_n^+ = 8$ B (one persistent pointer). The `Proxy` class derives directly from `Node` without specifying any further data, such that each `Proxy` consumes $m_n^\emptyset = 56$ B.

Similarly, the amount of memory consumed by an individual synapse m_c depends on the amount of data stored by the model. We will assume a synapse type that implements spike-timing dependent plasticity that stores variables common to all synapses in a shared data structure (NEST model name: `stdp_synapse_hom`), such that $m_c = 48$ B. The memory overhead for each potential source neuron is the pointer to a `vector` (8 B) and the size of an empty `vector` (24 B), thus $m_c^0 = 32$ B. For the sake of simplicity, in this manuscript we will only consider the case that a neuron has exactly one type of outgoing synapse. The additional memory consumption for a neuron with local targets, once the `vector` has been expanded to be able to store the nine built-in synapse types available, therefore consists of 9×8 B for pointers to `GenericConnector` objects and one `GenericConnector` object of size 104 B, including the empty `deque` (80 B) that contains the synapses, i.e., $m_c^+ = 176$ B. No additional overhead occurs for neurons that do not have local targets, therefore $m_c^\emptyset = 0$ B.

3. RESULTS

3.1. EMPIRICAL PARAMETER ESTIMATION

Counting the number of bytes per element or using an inbuilt `sizeof()` function as in Sec. 2.3 neglects any overhead that a container or the operating system may introduce for the allocation of this memory. Therefore, in this section we measure the actual memory consumption of NEST for a variety of networks to determine the parameters of (2) and (3) empirically; the results are shown in **Figure 2**.

To estimate the contribution of MPI to the total memory consumption of NEST at start-up, we compare the measured memory usage when running NEST compiled with MPI support to that used when NEST is compiled without MPI support. On the JUGENE system, we determine that the memory usage of MPI is at most 64 MB and that of NEST 30 MB (for an empty network). For simplicity, we assume a base memory usage of $\mathcal{M}_0(M) = 94$ MB throughout this paper.

To determine the parameters of (2), we create unconnected networks of neurons (type `iaf_neuron`) on a single process. We increase the number of neurons from 100,000 to 500,000 in steps of 100,000 and measure the memory consumption at each network size. The slope of the linear fit to the data gives the memory consumed by an `iaf_neuron` object plus the memory overhead per node (local or remote), i.e., $m_n + m_n^+ + m_n^0 = 1090$ B. As the overhead per node consists of two persistent pointers and the overhead per local node of one persistent pointer, in the following we assume $m_n^0 = 16$ B, $m_n^+ = 8$ B and $m_n = 1072$ B. Here, we slightly increase the estimate of m_n such that it is divisible by eight. To determine the memory usage of `Proxy` representations of non-local nodes, we repeat the measurements for $M = 2$, such that half of the created nodes are local and half are non-local. Note that memory measurements are still taken on a single process. As

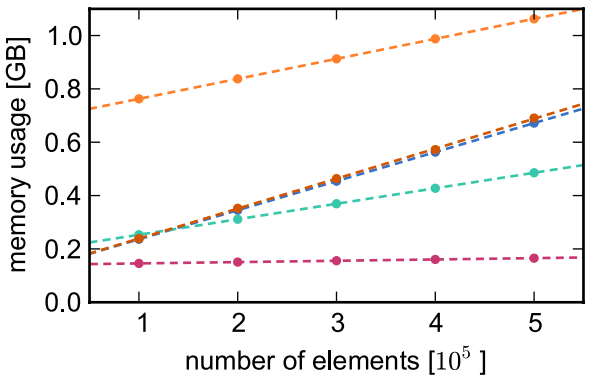


FIGURE 2 | Measured memory usage of simulator components as functions of the number of elements. Memory usage on one process for an unconnected network as a function of the number of nodes (blue circles), an unconnected network distributed over two processes as a function of the number of nodes (green circles), a network with exactly one synaptic connection as a function of the number of nodes (dark orange circles), a single neuron as a function of the number of autapses (pink circles), and a network of 5×10^5 neurons as a function of the number of autapses (light orange circles). The dashed lines indicate linear fits to the respective data sets. All networks created on a single process unless otherwise stated, with neurons of type `iaf_neuron` and synapses of type `stdp_synapse_hom`.

m_n, m_n^+, and m_n^0 have already been determined as described above, we can easily extract the amount of memory that is taken up by one `Proxy` node $m_n^\emptyset$ from the slope of the linear fit, resulting in $m_n^\emptyset = 56$ B to the nearest multiple of eight.

To parameterize (3), we first repeat the measurements on a single process for increasing network size, with exactly one synaptic connection included at each network size. The existence of a single synapse ensures that the connection infrastructure is initialized but almost entirely empty, similar to **Figure 1B**. The slope of the linear fit is therefore $m_n + m_n^+ + m_n^0 + m_c^0 = 1125$ B, resulting in $m_c^0 = 32$ B to the nearest multiple of eight. We then measure the amount of memory consumed by a single synapse of type `stdp_synapse_hom` by creating one `iaf_neuron` on a single process with connections to itself. We increase the number of autapses from 100,000 to 500,000 in steps of 100,000 and measure the memory consumption at each network size. The slope of the linear fit to the data reveals $m_c = 48$ B to the nearest multiple of eight. Finally, we determine the additional memory overhead for a neuron with local targets m_c^+. We create a network of 5×10^5 neurons and successively connect each neuron to itself exactly once. We increase the number of autapses from 100,000 to 500,000 in steps of 100,000 and measure the memory consumption at each network size. The slope of the linear fit to the measurements comprises $m_c + m_c^+ = 748$ B, resulting in $m_c^+ = 704$ B to the nearest multiple of eight. In the original implementation there is no additional overhead for nodes without local targets, therefore $m_c^\emptyset = 0$ B.

Our results show that the theoretical considerations discussed in Sec. 2.3 do not always accurately predict the true memory consumption. Having made the assumption that the theoretical values for m_n^0 and m_n^+ are correct, we determine a value for $m_n^\emptyset$ which is close to its theoretical value (i.e., 57 B compared to 56 B). However, the empirically measured value for m_n is substantially larger than its theoretical value (1074 B compared to 424 B). Similarly, assuming that the theoretical value for $m_c^\emptyset$ is correct, we measure values for m_c^0 and m_c that are near to their theoretical values (35 B compared to 32 B and 50 B compared to 48 B, respectively). The empirically measured value for m_c^+ is markedly larger than its theoretical value (700 B compared to 176 B). The fact that the empirically determined values for $m_n^\emptyset$, m_c^0, and m_c are so close to their theoretical values suggests that it was reasonable to make the assumption that m_n^0, m_n^+, and $m_c^\emptyset$ are correctly predicted by their theoretical values.

The empirical values that are well predicted by their theoretical values measure the memory consumption of objects containing only static data structures, whereas the two empirical values that are underestimated by their theoretical values, m_n and m_c^+, measure the memory consumption of objects that also contain dynamic data structures, specifically `deque` containers. The underestimation is due to the fact that dynamic data structures can be allocated more memory when they are initialized than the results of the `sizeof()` function reveal. In the case of the `deque`, memory is pre-allocated in anticipation of the first few entries. Increasing the number of autapses for each neuron from one to two does not cause the memory consumption to increase by 48 B per neuron, because the synapses are stored in this pre-allocated memory; a new section of memory is only allocated when the originally allocated memory is full (data not shown). However, this is exactly the extreme case we need to be concerned with for large scale networks/clusters: each neuron having either zero or a very small number of local targets on any given machine. Therefore in the following we will use the empirically determined values, at the risk of a certain inaccuracy for small networks/clusters.

3.2. ANALYSIS OF MEMORY CONSUMPTION

On the basis of our model for the memory consumption of a neuronal simulator (Sec. 2.1) with empirically determined parameters (Sec. 3.1), we calculated the usage of memory by NEST for the individual components and in total. Here and in the rest of the manuscript, we assume $K = 10{,}000$ synapses per neuron, representing a worst-case connectivity at least for cortical network models. The results for the cases of strong scaling (i.e., fixed network size on an increasing number of processes) and of weak scaling (i.e., network size increases proportionally to the number of processes) are displayed in **Figure 3**. The strong scaling results demonstrate that although the memory that is consumed by one neuron ($m_n = 1072$ B) is two orders of magnitude greater than the memory that is consumed by one synapse ($m_c = 24$ B), neurons contribute far less to the overall memory consumption. As the number of synapses is four orders of magnitude greater than the number of neurons, the contribution of neurons to the total memory consumption is negligible, whereas the contribution of synapses dominates the total memory consumption for less than around 300–700 processes. At these small-to-moderate cluster sizes, the size of an individual synapse plays a substantial role in determining the maximum neuronal network size that will fit on a given cluster. As the number of processes increases, the

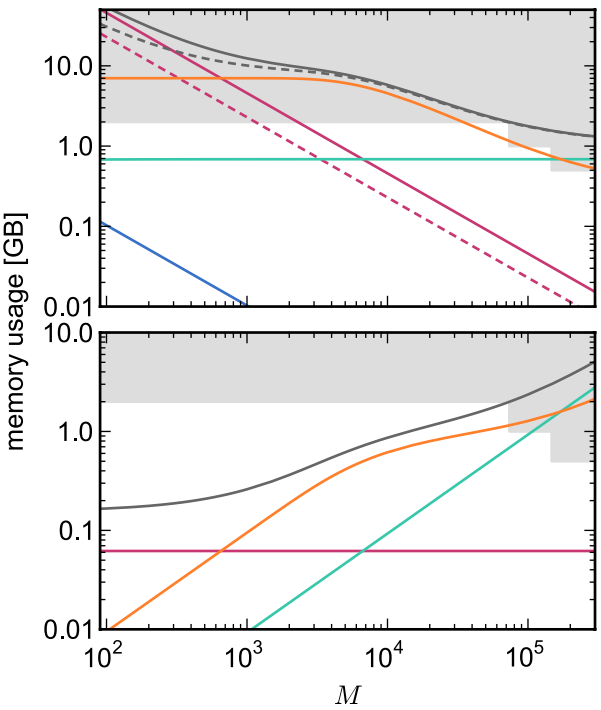


FIGURE 3 | Prediction of component and total memory usage per process as a function of the number of processes. Upper panel: Memory usage for a network containing $N = 9.5 \times 10^6$ neurons and $K = 10{,}000$ synapses per neuron: memory consumed by neurons (blue, $m_n = 1072$ B), neuronal infrastructure (turquoise, $m_n^0 = 16$ B, $m_n^+ = 8$ B, $m_n^{\emptyset} = 56$ B), synapses (pink), connection infrastructure (orange, $m_c^0 = 32$ B, $m_c^+ = 704$ B, $m_c^{\emptyset} = 0$ B), and total memory (gray). Dashed lines show memory consumption using a light-weight static synapse with $m_c = 24$ B, solid lines using a more complex plastic synapse model with $m_c = 48$ B. The gray area indicates a regime that lies beyond the memory capacity of the JUGENE Blue Gene/P machine. Lower panel: Memory usage for a network with $N_M = 129$ neurons on each machine and $K = 10{,}000$, such that $N = 9.5 \times 10^6$ just fits on $M = 73{,}728$ processes with 2 GB RAM each. Line colors and shaded area as in upper panel.

relative contribution of the size of synapse objects decreases and the curves for total memory consumption converge.

For cluster sizes greater than 700, the dominant component of memory consumption is the connection infrastructure; for greater than 3000–7000 processes, the neuronal infrastructure also consumes more memory than the synapses. The neuronal infrastructure is proportional to N for all but very small values of M (2); the connection infrastructure memory consumption is proportional to N up to around $M = 10{,}000$ (3). This results in constant terms in the case of strong scaling (**Figure 3**, upper panel) and linearly increasing terms in the case of weak scaling (**Figure 3**, lower panel). For $M > 10{,}000$ the increase in memory of the connection structure is sub-linear with respect to N, because the number of neurons with no local targets $N_c^{\emptyset}$ increases. However, the absolute value of the memory consumption of the connection infrastructure remains very high, and is only overtaken by the consumption of the neuronal infrastructure at around $M = 200{,}000$.

We can therefore conclude that the serial terms in the memory consumption of neuronal and connection infrastructures represent major limiting factors on the simulation of brain-scale models in NEST.

3.3. EVALUATION OF DESIGN STRATEGIES

In the previous section we determined that infrastructures, rather than the actual objects of interest in a simulation, are the most memory consuming components for large numbers of processes. In this section we demonstrate that, having identified a problem, the memory modeling approach can be used to predict the effects of design alterations, and thus provide a foundation for deciding on which optimization to focus resources. The commonality between the issues of neuronal infrastructure and connection infrastructure is that as network and machine sizes increase, the infrastructures become increasingly sparse. For a network containing $N = 9.5 \times 10^6$ neurons on $M = 73{,}728$ processes with 2 GB RAM each, there are only 129 neurons on each machine (see **Figure 3**) and thus almost 9.5×10^6 entries in the neuronal infrastructure to account for non-local neurons. Similarly, under the assumption of random connectivity, around 87% of neurons will not have any local targets on a given machine. Thus, taking this sparseness into consideration will help to avoid serial overheads that grow with the network size. Note that for smaller clusters and network sizes, sparseness is not a major consideration. For less than around 700 processes, the serial overheads are dominated by the memory consumption of synapses. In this regime, it is reasonable to use data structures that do not exploit sparseness.

There are two requirements on data structures to replace the neuronal and connection infrastructure. The first requirement is that memory overheads for non-present items are extremely low, so that sparseness is effectively exploited. The second requirement is that the data structures enable efficient search routines. This requirement is based on the assumptions that a distributed simulation described by a single serial instruction set, or script, and that a pre-synaptic spike is delivered to all processes and then dispatched to the local targets. If the former assumption is false, then it is not necessary to search the neuronal infrastructure for neurons that may be located on other machines. If the latter is false, it is not necessary to search the connection infrastructure for local targets of a neuron that may not have any on that specific process. If the assumptions are true, the requirement is that a search routine exists for the candidate data structure that terminates rapidly, both when a searched-for item is present and when it is not present. The original implementation of the neuronal infrastructure, consisting of a `vector` of pointers to local neurons and proxies of non-local neurons, fulfills the second requirement but not the first. A well-loaded hash map fulfills the first requirement but not the second, as the worst-case look-up time (which can be expected to occur frequently for high degrees of sparseness) is $\mathcal{O}(n)$ for typical implementations, where n is the number of items stored. The worst-case look-up time has been improved to $\mathcal{O}(\log n)$ and even $\mathcal{O}(1)$ but at the cost of higher memory consumption (Dietzfelbinger et al., 1994). Two candidates which fulfill both requirements are sparse tables (Silverstein, 2005) and cuckoo hash maps (Pagh and Rodler, 2004).

Sparse tables store entries in groups of equally sized successive index ranges; by default $s_{gr} = 48$ entries per group with an overhead of $m_{gr} = 10\,\mathrm{B}$ per group; each group contains a contiguous block of entries and a bitmap of size s_{gr}, which indicates whether there is an entry for the given index or not; to store entries with maximum index n the total overhead of the sparse table (i.e., not including the memory consumption of the stored items) is $n(m_{test} + m_{gr}/s_{gr})$ where $m_{test} = 1$ bit.

Cuckoo hashing is an interesting alternative to standard hashing as it provides a look-up time that is constant in the worst-case (Pagh and Rodler, 2004). In its original formulation it consists of two hash functions and two hash tables. To store an item x, the first hash function is used to generate a location. If the location is already occupied by an item y, the new item x replaces the stored item y, and the second hash function is used to generate a location for item y in the second hash table. Should this location also be occupied by an item z, the items are swapped, and then the first hash function is used to generate a location for the displaced item z in the first hash table. The process continues, alternating hash functions and tables, until an empty location is found or a predefined maximum number of iterations is reached. In the latter case, a rehash is performed. This concept was later extended to d-ary cuckoo hashing, i.e., d hash functions that code for alternative positions in one hash table, where the capacity threshold c_h depends on the number of hash functions: $c_h = 0.49, 0.91, 0.97, 0.99$ for $d = 2, 3, 4, 5$ (Fotakis et al., 2003). In common with other imperfect hashing schemes, i.e., where there is no guarantee that the hash function will generate a unique location for each key, the keys need to be stored with the items to allow identification at look-up time. This results in an overhead of m_{key} per entry. To store n entries the hash size should be at least n/c_h, which entails a total memory usage (i.e., including the memory consumption of the stored items) of $n(m_{key} + m_{val})/c_h$, where m_{val} is the memory consumption of payload per entry. In the rest of the manuscript, we will consider tertiary cuckoo hashing schemes ($c_h = 0.91$).

3.3.1. Neuron infrastructure

Let us first consider an alternative approach to implementing the neuronal infrastructure. The infrastructure has two critical functions, related to access and connection, respectively. Firstly, it has to enable a rapid determination of whether a given global index (GID) refers to a local or remote node, and return a pointer to the node if it is local. In the original implementation of the neuronal infrastructure, this functionality is ensured by a `vector` of pointers to local neurons and proxies for the non-local neurons (**Figure 4A**). The locality of a node with a given GID can be determined very quickly, as the GID of a node determines which index of the `vector` to examine. If following the pointer results in a non-proxy node, then by definition the node with that GID is local. Conversely, if following the pointer results in a `Proxy` node, then the node is non-local. Similarly, returning a valid pointer once a node has been determined to be local is trivial.

Secondly, as not all neuron models support all possible connection types, the neuronal infrastructure must allow the model of a node with a given GID to be determined, independent of whether it is local or remote. In NEST a model is a separate object from the nodes it creates. Local nodes and `Proxy` nodes explicitly store a model index (MID), which can be used to access a local model object. Therefore, `Proxy` nodes can uniquely identify the model of their remote counterparts. With access to this information, a process can check whether a connection described in the instruction set is valid.

One strategy to exploit the sparseness in the neuronal infrastructure is to use look-up tables. A `vector` of size N_M that stores pointers to local neurons is maintained, see **Figure 4B**. This is supplemented by a look-up table of size N that maps a node's GID to its local index in the node `vector` if it is local, and to a special value if it is remote. This enables a swifter determination of node locality, as fewer indirections are involved. Returning a valid pointer to a local neuron is also trivial in this design. To access the type information of a non-local neuron, an additional

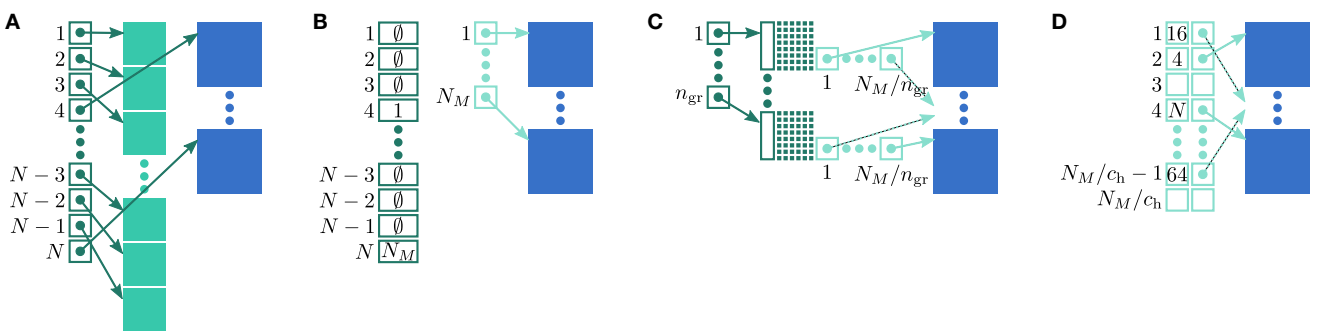


FIGURE 4 | Alternative designs for the neuronal infrastructure. In all panels $M = 4$, and the color of elements denotes which memory term they contribute to: m_n^0 (dark green), $m_n^\emptyset$ (mid green), m_n^+ (light green), m_n (blue). **(A)** Original implementation. A `vector` of length N (dark green) is maintained of pointers to either local nodes (blue squares) or `Proxy` nodes (filled green squares). **(B)** Look-up tables implementation. A `vector` of length N (dark green) is maintained that contains the local indices in a `vector` of length N_M (light green) for local nodes and a special value (here ∅) for remote nodes. The shorter `vector` contains pointers to the local nodes. **(C)** Sparse table implementation. A `vector` of length n_{gr} (dark green) is maintained of pointers to sparse groups, that is $n_{gr} = N/s_{gr}$ denotes the number of sparse groups. A sparse group has an overhead of m_{gr} (vertical dark green boxes) and s_{gr} bits (tiny dark green squares) that are set to 1 if an item is present at the given index and 0 if it is absent. The N_M/n_{gr} local items in each group (light green squares) are pointers to local nodes. **(D)** Cuckoo hashing implementation. A hash table of size N_M/c_h (light green squares) is maintained that contains pointers to local neurons as payload and also the keys of the items (i.e., the global indices of the stored nodes).

look-up table can be used that maps ranges of GIDs to their MID. Usually ranges for a given MID can be expected to be quite large, as it is substantially more convenient to create blocks of neurons of the same model than to create them individually whilst varying models. Thus the MID for a given GID can be obtained by performing a divide-and-conquer search of a table that is expected to be several orders of magnitude smaller than N. As the memory consumed by the model look-up table is negligible, the memory usage of the neuron infrastructure when using look-up tables is given by (2) where $m_n^0 = 4$ B now accounts for the size of the data type that stores the local index of nodes in the node look-up table, and $m_n^+ = 24$ B accounts for the two persistent pointers to local neurons used for access and the one persistent pointer used during simulation. As there are no proxies, $m_n^\emptyset = 0$ B.

An alternative strategy is to replace the `vector` of pointers to nodes and proxies with a sparse table that stores pointers to local neurons, and represents with one extra bit whether an item is present (local) or absent (remote). This is depicted in **Figure 4C**. As with the previous strategy, this enables a swift determination of locality and return of valid pointers to local neurons. Similarly, this data structure must be supplemented with a model look-up table to provide access to the models of remote nodes. The memory usage for this design is as in (2) with $m_n^0 = m_{test} + m_{gr}/s_{gr}$ where $m_{test} = 1$ bit, and $s_{gr} = 48$ and $m_{gr} = 10$ B are the sparse table group size and memory overhead per group as discussed above, $m_n^+ = 24$ B and $m_n^\emptyset = 0$ B.

A third strategy is to use cuckoo hashing to store the pointers for local nodes, as shown in **Figure 4D**, which reduces the overhead per neuron to $m_n^0 = 0$ B. The memory usage for this design is therefore given by (2) with $m_n^+ = (m_{key} + m_{val})/c_h + 16$ B. Here, $m_{key} = 4$ B is the data type of the key, i.e., the GID of a node (for ranges up to 10^9), $m_{val} = 8$ B is the payload, i.e., the size of a persistent pointer to a node, and $c_h = 0.91$ is the capacity threshold of the hash map as discussed above. The final term of 16 B accounts for the two additional persistent pointers to each local neuron. Also in this design, the total removal of proxies results in $m_n^\emptyset = 0$.

All three strategies remove the need for the concept of a `Proxy` node which stands in for its remote counterpart to fulfill various roles in the simulator functionality. The ideal memory usage for neurons would be (2) with $m_n^0 = m_n^+ = m_n^\emptyset = 0$, i.e., no neuronal infrastructure to account for non-local neurons and no persistent pointers required for local neurons. By calculating the memory usage as a function of the number of processes assuming the three strategies above, we can see to what extent these strategies deviate from the ideal case. This is displayed in **Figure 5**. All three strategies result in reduced memory consumption with respect to the original design described in Sec. 2.3, for all values of M. For $M < 1000$ both the look-up table and the sparse table design are close to ideal. For $M > 1000$, the sparse table design results in a greater reduction of memory consumption with respect to the original implementation, and is in fact very close to optimal. The cuckoo hashing design gives very similar results to the sparse tables design for $M < 100{,}000$, but marginally improved results for greater numbers of processes (data not shown). As sparse tables and cuckoo hashing result in a more substantial improvement in memory consumption than look-up tables for $M > 1000$, we can discard look-up tables as a design candidate. We can further conclude that the improvements to memory consumption obtained by cuckoo hashing compared to sparse tables do not compensate for their greater complexity. Thus, we conclude that a sparse table combined with a light-weight model look-up table is the best of the three strategies to reduce the memory footprint of the neuronal infrastructure.

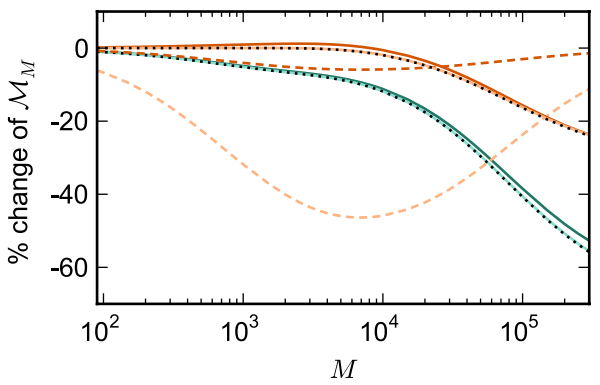


FIGURE 5 | Prediction of change in memory usage per process for infrastructure components for different design choices with respect to the memory usage of the original design. Percentage change of the total memory usage for a network of $N_M = 1000$ neurons on each process predicted as a function of the number of processes when implementing neuronal infrastructure as a look-up table (dark green) or a sparse table (light green) and when implementing connection infrastructure as a hash map with capacity threshold $c_h = 0.91$ (solid dark orange) or a sparse table (solid light orange). The dotted black curves show the ideal cases that there is no memory overhead for neuronal infrastructure, $m_n^0 = m_n^+ = m_n^\emptyset = 0$, and no memory overhead for connection infrastructure, $m_c^0 + m_c^\emptyset = 0$. Assuming no optimizations to exploit sparseness are carried out, the dashed light orange curve shows predicted change to memory usage for synapses stored in `vectors` instead of `deques`, and the dashed dark orange curve indicates the predicted change for synapses stored in `deques`, for a reduced size of the `vector` containing pointers to `GenericConnector` objects.

3.3.2. Connection infrastructure

The functionality that any implementation of the connection infrastructure must deliver is as follows. Firstly, it must be able to mediate spikes that are communicated to the process. Assuming that each arriving spike event knows the GID of the neuron that produced it (Morrison et al., 2005), the connection infrastructure must be able to determine rapidly whether the node corresponding to that GID has local targets. If so, it must provide access to the synapses that communicate the spiking information from the node to its targets. If communication between processes is implemented such that spikes are guaranteed to be sent only to processes on which they have targets, rather than to all processes, this requirement can be relaxed to swift access to the local synapses of a given node GID, without the initial determination of their existence.

As synapses can be of heterogeneous types, in order to allow the user to query the state of the network it must also be possible to locate all synapses of a given type for a given pre-synaptic node GID. The original implementation of the connection infrastructure, as shown in **Figure 6A**, consists of a `vector` of length N of pointers to `vectors` of pointers

to GenericConnectors, one per type of synapse that the node makes to local targets. The GenericConnectors contain deques of synapses of their respective types. Therefore the existence of local targets for a node with a given GID can easily be determined by testing whether the pointer in the vector of length N at index GID points to a non-empty vector. Allowing access to the relevant synapses is then trivial. As the synapses are sorted into vectors according to their types, searching for particular synapses of a particular type for a given node is straightforward.

Figure 5 shows that improvements in memory consumption can be obtained by reducing the additional memory overhead for nodes with local targets, m_c^+. Storing synapses in vectors instead of deques reduces memory consumption substantially up to a maximum of 46% at around $M = 7000$, after which the amount of improvement reduces for larger M. For $M < 50{,}000$, this adaptation results in greater improvements to the memory consumption than exploitation of sparseness in the neuronal or connection infrastructure. To make this prediction, we estimated m_c^+ for the case that a GenericConnector object contains its synapses in a vector. An empty vector consumes 24 B in comparison to 80 B for an empty deque, therefore the theoretical value for m_c^+ reduces from 176 B to 120 B. However, as discussed in Sec. 3.1, the theoretical values can substantially underestimate the actual values; we measured $m_c^+ = 704$ B which implies that the deque is approximately seven times larger than its theoretical value. This is likely to be less problematic for vectors than for deques, as the memory allocation of vectors is more straightforward. We therefore make a pessimistic estimate of $m_c^+ = 264$ B, assuming that a vector also consumes seven times more memory. However, any reasonable estimate would produce qualitatively similar results. We therefore conclude that implementing synapse containers as vectors rather than deques is a useful step to improve the memory consumption. A lesser degree of improvement can be achieved by reducing the size of the vector of pointers to GenericConnector objects. Instead of one entry for each available synapse type, the length of the vector can be reduced to the number of synapse types the source neuron actually has on the local machine. This necessitates increasing the size of the GenericConnector object to include a synapse type index. This adaptation results in a maximum of 6% improvement for 7000 processes. **Figure 6B** illustrates the strategies for reducing m_c^+ which are described above.

Disregarding potential improvements to the m_c^+ component, we can also consider how to exploit the sparseness of the distribution of synapses over the machine network. One possibility to account for the increasing number of neurons without local targets is to replace the vector of length N with a sparse table, as shown in **Figure 6C**. This allows very rapid determination of whether a node has local targets; if local targets exist, access to the corresponding synaptic connections is also swift. For nodes with local targets, the internal structure of a vector of pointers to GenericConnectors is maintained, to allow per-type storage of synapses. The memory usage for this design is as in (3) with $m_c^0 = m_{\text{test}} + m_{\text{gr}}/s_{\text{gr}} = 2.67$ bits, $m_c^+ = 736$ B, and $m_c^\emptyset = 0$ B. Note that m_c^+ has absorbed 32 B from m_c^0 in the original implementation. This accounts for an empty vector (and pointer to it) which is now only initialized for nodes which have local targets, rather than all nodes.

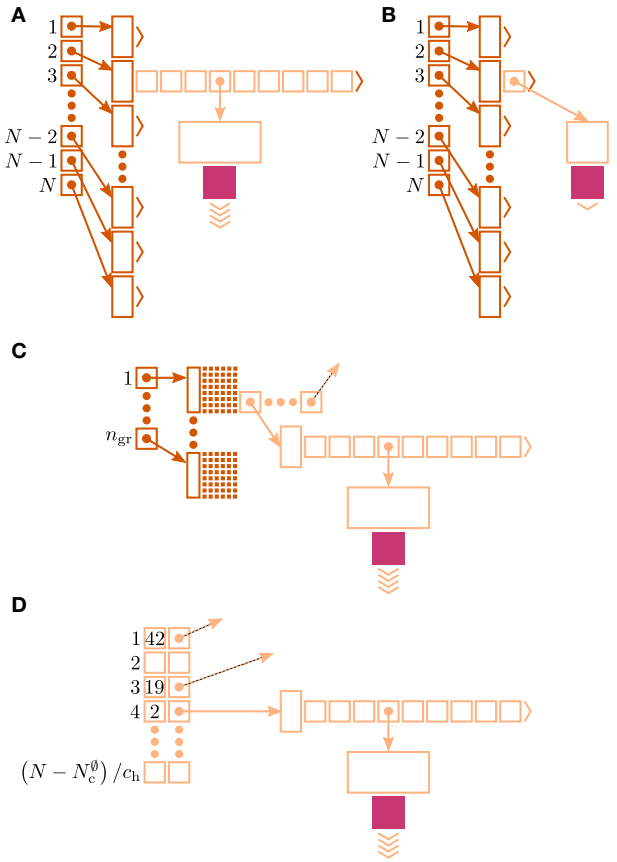


FIGURE 6 | Alternative designs for the connection infrastructure. The color of elements denotes which memory term they contribute to: m_c^0 (dark orange), m_c^+ (light orange), m_c (pink). **(A)** Original implementation. A vector of length N (dark orange) is maintained of pointers to initially empty vectors (dark orange boxes with single chevrons). If a node makes a synaptic connection to a target on the local machine, the length of the vector is extended to the number of synapse types available in the software (light orange boxes). A GenericConnector object is initialized for the type of the synaptic connection (large light orange box), and a pointer to it laid at the corresponding index of the vector. The new synaptic connection (pink box) is stored in a deque (four chevrons) within the GenericConnector. **(B)** Light-weight internal structure implementation. As in (A), but the vector of synapse types is only as long as the number of synapse types a node actually has on the local machine, and synapses are stored in a vector (single chevron) rather than a deque, which decreases the size of the GenericConnector. **(C)** Sparse table implementation. A vector of length n_{gr} (dark orange) is maintained of pointers to sparse groups, that is $n_{gr} = N/s_{gr}$ denotes the number of sparse groups. A sparse group has an overhead of m_{gr} (vertical dark orange boxes) and s_{gr} bits (tiny dark orange squares) that are set to 1 if an item is present at the given index and 0 if it is absent. The local items in each group (light orange squares) are pointers to vectors of GenericConnector objects and their respective deques of synapses (light orange/pink), as in (A). **(D)** Cuckoo hashing implementation. A hash table of size $(N - N_c^\emptyset)/c_h$ (light orange squares) is maintained that contains the keys of the items (i.e., the global indices of the pre-synaptic nodes) and as payload, pointers to vectors of GenericConnector objects and their respective deques of synapses (light orange/pink), as in (A).

As with the neuronal infrastructure, an alternative design strategy is cuckoo hashing. The hash table contains $N - N_c^{\emptyset}$ items, where $N_c^{\emptyset}$ is the number of nodes with local targets, therefore a table of size $(N - N_c^{\emptyset})/c_h$ is required. Each item consists of a pointer to a `vector` of pointers to `GenericConnector` objects, i.e., the same internal structure as in the original implementation. As for the neuronal infrastructure in Sec. 3.3.1, the keys of the items, in this case the GIDs of the pre-synaptic neurons, must be stored with the items. This is illustrated in **Figure 6D**. In this case, total memory consumed by synapses and connection infrastructure is given by (2) with $m_c^0 = m_c^{\emptyset} = 0$ B, $m_c^+ = (m_{key} + m_{val})/c_h + 736$ B where $m_{key} = 4$ B, $m_{val} = 8$ B, and $c_h = 0.91$.

The ideal memory usage would be for the case that there was no infrastructure overhead for neurons without local targets, i.e., as in (3) with $m_c^0 = 0$, $m_c^+ = 736$ B, and $m_c^{\emptyset} = 0$ B. **Figure 5** shows the predicted deviations of the sparse tables and cuckoo hashing strategies from the ideal case as a function of the number of processes M. For $M < 4000$ neither strategy improves memory consumption with respect to the original implementation; sparse tables require approximately the same amount of memory, and cuckoo hashing requires slightly more. This is because, unlike the situation with the neuronal infrastructure, there is no sparseness to be exploited on small numbers of processes. Assuming each node selects K pre-synaptic source nodes from a total of N nodes distributed evenly on M machines, the probability that a given node has no targets on a given machine is $p_{\emptyset} = (1 - 1/N)^{NK/M}$. With $K = 10{,}000$ and N ranging between 10^5 and 3×10^8 as in **Figure 5**, 5% of the connection infrastructure is empty for $M = 3338$, 50% of the connection infrastructure is empty for $M = 14{,}427$, and 95% of the connection infrastructure is empty for $M = 194{,}957$. For $M < 350{,}000$ the sparse table design results in a greater improvement to the overall memory consumption with respect to the original implementation than the cuckoo hashing design, and is in fact close to ideal. For larger values of M, the cuckoo hashing design results in marginally better memory consumption. Due to the better performance of the sparse table implementation for numbers of processes in the relevant range and the greater complexity of cuckoo hashing, we conclude that a sparse table is the better of the two designs for exploiting sparseness in the connection infrastructure for $M > 3000$.

3.4. ANALYSIS OF LIMITS ON NETWORK SIZE

In this section, we investigate to what extent the improvements to the memory consumption identified in Sec. 3.3.1 and Sec. 3.3.2 increase the maximum problem size that can be represented on a given machine. **Table 1** summarizes the parameters of the memory consumption model for the original implementation (see Sec. 2.2) and the parameters after adaptation of the data structures. The adapted values are for the case that sparse tables are used to implement neuronal and connection infrastructure. Moreover, synapses are contained in `vectors` rather than `deques`, and the length of the `vector` of pointers to `GenericConnector` objects has been reduced as suggested in Sec. 3.3.2. The empirical memory measurements for the adapted data structures are carried out as described in Sec. 3.1. The memory consumption for a neuron and synapse remain constant, $m_n = 1072$ B and $m_c = 48$ B. Once again, we see that the empirical value for m_c^+ (136 B) exceeds the theoretical value (96 B) due to the presence of a dynamic data structure, but the discrepancy for a `vector` is not as large as that for a `deque` in the original implementation. The empirical value for m_n^0 is less than the theoretical value whereas the empirical value for m_c^0 is greater, although they are measuring the overhead of the same type of data structure. This is likely to be due to the limits of measurement accuracy, as the expected value is so small.

Figure 7A shows the predicted memory required for the infrastructure components and the total memory consumption as a function of the number of processes M for two networks. With 1056 neurons on each process, the maximum network size is 7.8×10^7, which just fits on 73,728 processes with 2 GB RAM each, i.e., using one core per machine node of JUGENE. Using the original implementation, the largest network that could be represented with 1056 neurons per process would be 1.7×10^6 neurons on 1642 processes with 2 GB RAM each. With a load of 129 neurons on each machine, a maximum network size of $N = 9.5 \times 10^6$ is predicted (see **Figure 3**). Thus, the maximum network size that can be represented on the available machine has increased by an order of magnitude. In addition, the improved infrastructures allow a much more effective use of the entire machine. A network of 5.9×10^7 neurons can be distributed over all 294,912 cores of JUGENE (0.5 GB RAM each) with a load of 201 neurons per core (**Figure 7**). For the original implementation, the maximum network size that could be represented on the entire machine is 3.1×10^6 neurons with a load of 10 neurons per core, thus the adaptations have increased the size of the network representable on the entire machine by a factor of nearly twenty.

The memory consumption model accurately predicts the actual memory usage of NEST. **Figure 7B** shows the memory required per process for a simulation in which each neuron receives 10,000 randomly chosen inputs. As predicted, the largest network that could be represented on JUGENE using the original implementation contains 9.5×10^6 neurons distributed over 73,728 processes with 2 GB RAM each. The measured memory consumption for the corresponding weak scaling experiment with 129 neurons per process is well fitted by the theoretical prediction. As for the previous measurements, these results are obtained on a single core by falsely instructing the application that it is one of M processes. Thus, it only creates the desired N/M local neurons, but all infrastructures are initialized on the assumption that there are an additional $N - N/M$ neurons on remote processes. When using the adapted implementation, the maximum network size on 73,728 processes increases to 7.8×10^7 neurons. The memory consumption for the corresponding weak scaling experiment with 1056 neurons per process is well fitted by the theoretical prediction for $M > 2000$. For smaller numbers of processes, the model underestimates the true memory consumption. This is due to the `vectors` containing the synapses doubling their size when they are filled past capacity, thus allocating more memory than is strictly needed. As the number of local synapses per source neuron decreases, the `vectors` are more optimally filled, leading to a convergence of the measured memory consumption to the theoretical prediction. A better fit can be obtained by accounting for the sub-optimal filling of `vectors` in the model, but as we are more concerned with correct prediction of the model for large numbers of processes, we

Table 1 | Parameters of the memory model. The theoretically and empirically determined parameters of (2) and (3) for the original implementation and the adapted implementation. All empirical values in bytes given to the nearest multiple of eight; empty entries indicate that the corresponding theoretical value was assumed to be accurate to enable determination of other parameters in that row.

	m_n^0	$m_n^\emptyset$	m_n^+	m_c^0	$m_c^\emptyset$	m_c^+
ORIGINAL						
Theory	16 B	56 B	8 B	32 B	0 B	176 B
Actual	–	56 B	–	32 B	–	704 B
ADAPTED						
Theory	2.67 bits	0 B	24 B	2.67 bits	0 B	96 B
Actual	1.38 bits	–	–	1.38 B	–	136 B

omit this extension of the model for the sake of simplicity. Note that this effect does not occur for the original implementation, as the synapses are stored in `deques`, which allocate memory in constant-sized blocks.

The shallow slope of the memory consumption curve for the adapted implementation means that doubling the number of available processes will approximately double the size of the network that can be represented to 1.5×10^8. In sharp contrast, the steep slope of the memory consumption curve for the original implementation means that doubling the machine size would only increase the size of the maximum network to 1.2×10^7, a factor of 1.3.

The adaptations to the data structures alter not only the absolute memory usage of the various components of NEST, but also the relative usage. **Figure 8** shows the proportion of the total memory consumed by each of the components in the original and the adapted implementation. The cross-over point at which the neuronal infrastructure requires more memory than the synapses is eliminated in the adapted implementation; the memory consumption for the neuronal infrastructure is essentially negligible for a large range of processes, having been reduced to under 3 bits per node. The point at which the memory usage of the connection infrastructure is greater than that of the synapses is shifted from $M = 700$ for the original implementation to $M = 4000$ for the adapted implementation. However, even this shifted value is comparatively low with respect to the number of processes available on a large Blue Gene machine. At $M = 200{,}000$, the connection infrastructure accounts for nearly 70% of the total memory usage, synapses for just over 20%, and neurons approximately 5%. It is therefore clear that any further increases to the maximum network size that can be simulated on a machine with in the order of 100,000 processes can only be achieved by further reducing the overhead for the connection infrastructure. Given that the connection infrastructure also exploits the sparseness of the distribution of synapses over the processes, the question arises of why it is still the dominant memory component for large M. The answer is that although the first term of (3) has also been reduced to less than 3 bits per node and so is negligible, the third term $(N - N_c^\emptyset)m_c^+$ is not. This term expresses the memory consumption of the data structure containing the synapses with local targets, which is comparatively large, $m_c^+ = 136$ B.

The potential benefits of reducing m_c^+ are illustrated in **Figure 9**, which shows the largest network size that could be represented as a function of the number of processes. The original implementation shows a saturating behavior, whilst the adapted implementation shows an approximately linear relationship. The ideal case is that there is no memory overhead for neuronal or connection infrastructure. By reducing the overhead for neurons with local targets from $m_c^+ = 136$ B to 16 B, a close to optimal performance could be achieved. Therefore we can conclude that the next priority for reducing memory consumption should be optimizing this aspect of the connection infrastructure, rather than reducing the sizes of the neuron and synapse objects themselves, or attempting to further exploit sparseness in the data structures.

4. DISCUSSION

Saturation of speed-up at distressingly low numbers of processes is a very common experience when parallelizing an initially serial application, occurring when the contribution of the serial component of an algorithm to the total computation time outweighs that of the parallel component. Here, we consider the analogous problem of saturation of the available memory, which occurs when serial components of data structures consume all the memory available for each process, such that increasing the number of processes available does not increase the size of the problem that can be represented. The shift from saturation in computation time to saturation in problem size may be a characteristic of technological progression: an algorithm that performs perfectly well in a serial program may be poorly designed with respect to later parallelization, i.e., has large serial computational overhead. Re-designing algorithms with the previously un-dreamed-of parallelization in mind results in better parallel algorithms and extends the realm of scalability to much higher numbers of processes. When trying to extend the realm of scalability to ever larger problems and cluster sizes, a data structure design that performs perfectly well on the moderate scale may turn out to be poorly designed with respect to massive parallelization, i.e., it contributes a substantial serial term as well as a parallel term to the total memory usage. The serial overhead is not noticeable or problematic on moderate scales, but becomes dominant on large scales, resulting in a constant term in strong scaling investigations and a linear term in weak scaling investigations.

To analyze this problem in a systematic fashion, we developed a model for the memory consumption of a neuronal simulator which allows us to separate serial and parallel contributions. In this manuscript we apply the model to understanding and improving the memory consumption of NEST (Gewaltig and Diesmann, 2007). This entails taking the constraints of NEST's fundamental

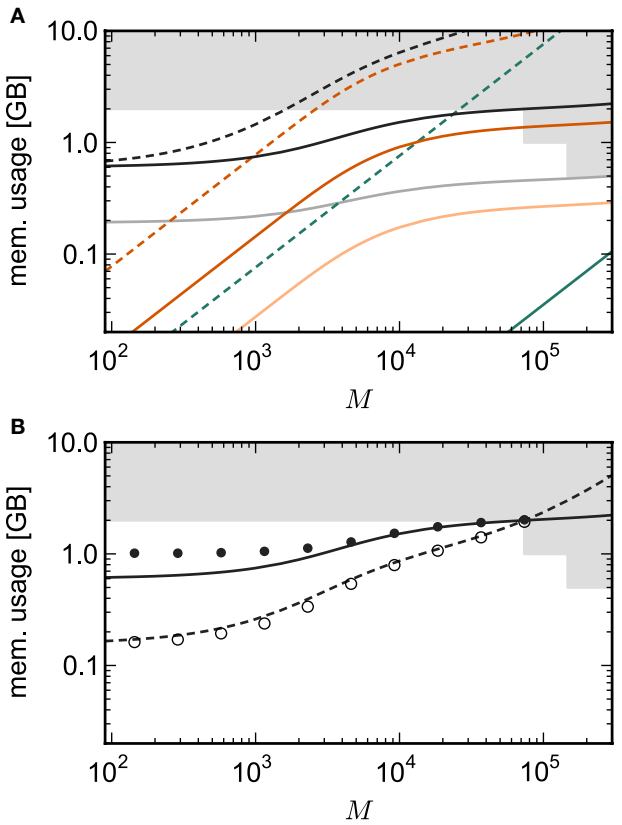


FIGURE 7 | Prediction and measurement of total memory usage per process. (A) Predicted memory usage as a function of the number of processes for a network of $N_M = 1056$ neurons on each machine such that $N = 7.8 \times 10^7$ just fits on $M = 73{,}728$ processes with 2 GB RAM each. Solid curves indicate the memory usage for the adapted implementation, and dashed curves the original implementation. Usage shown for neuronal infrastructure (dark green), connection infrastructure (dark orange), and total memory (black). Dashed curves indicate memory consumption predicted using the original data structures to implement infrastructure. The light orange curve indicates the memory consumption of the connection infrastructure and the light gray curve indicates the total memory consumption for a network with $N_M = 201$ neurons such that $N = 5.9 \times 10^7$ just fits on $M = 294{,}912$ processes with 0.5 GB RAM each. **(B)** Measured memory usage as a function of the number of processes. Original implementation: memory consumption for a network of $N_M = 129$ neurons on each machine such that $N = 9.5 \times 10^6$ just fits on 73,728 processes (open circles). Adapted implementation: memory consumption for a network of $N_M = 1056$ neurons on each machine such that $N = 7.8 \times 10^7$ just fits on 73,728 processes each (filled circles). The theoretical predictions of memory consumption for these networks are given by the dashed and solid curves, respectively.

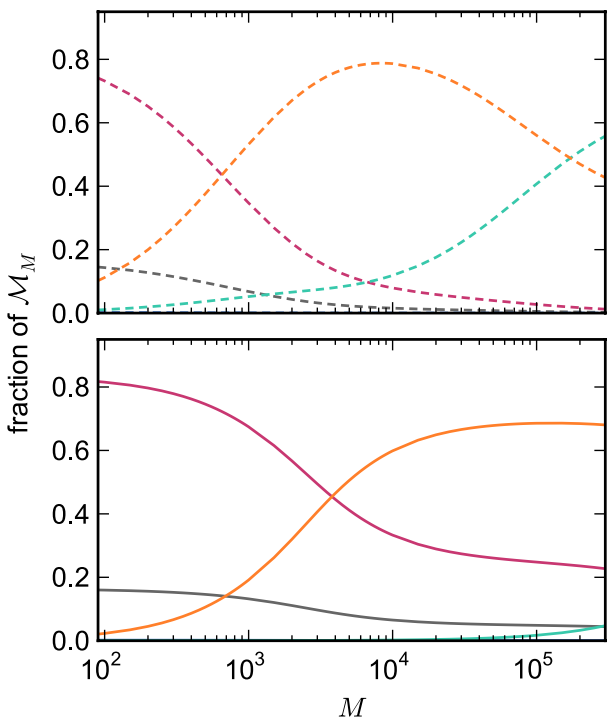


FIGURE 8 | Relative memory consumption per process as a function of the number of processes. Fraction of total memory consumption predicted for the following components: NEST and MPI (gray), synapses (pink), connection infrastructure (orange), neuron infrastructure (green) for the case that the number of neurons per process is set to $N_M = 1000$. Upper panel: original implementation. Lower panel: adapted implementation.

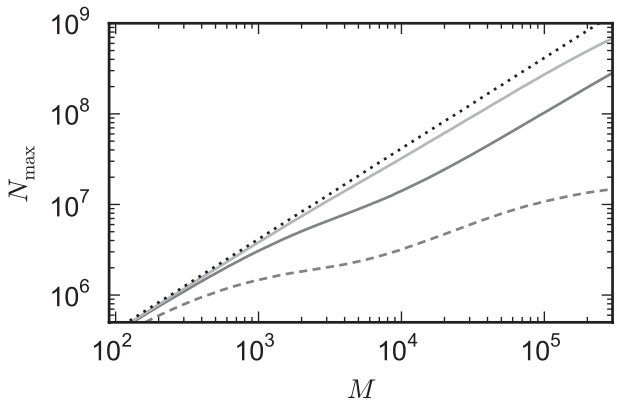


FIGURE 9 | Prediction of maximum network size as a function of the number of processes. Original implementation (dashed curve), adapted implementation with $m_c^+ = 136$ B (solid dark gray curve), adapted implementation assuming $m_c^+ = 16$ B (solid light gray curve). The dotted curve indicates the predicted maximum network size for the ideal case, i.e., no overhead for neuronal or connection infrastructure ($m_n^0 = m_n^\emptyset = m_n^+ = m_c^0 = m_c^\emptyset = m_c^+ = 0$). Curves are based on the assumptions of 2 GB RAM per process and 10,000 incoming connections per neuron.

design into consideration, for example that it must be possible to determine the model of every node, regardless of locality, and that spikes are distributed to all processes and then dispatched by the local synapses. However, the expressions are general enough to be used for any neuronal simulator, in particular those developed to simulate networks of complex compartmental neuron models. For a given simulator, the data structures of the simulator must be analyzed to determine which objects contribute to which parameters of the memory model. If a simulator has different

underlying constraints from NEST, this will reduce or expand the range of possible data structures that can be considered as alternative designs. The role of the model is to evaluate the consequences of potential design choices in terms of memory consumption, not to identify the range of possible solutions. Moreover, the principles are applicable to any distributed application that needs to maintain a large number of objects, although in such cases the terms of the memory model would have to be adapted.

By empirically determining the parameters of the model, we found that the theoretical values for the memory consumption of some data structures underestimate their actual memory consumption. In particular, dynamic data structures typically use more memory than would be predicted by the sum of the size of the empty structure and the size of their contents, due to the dynamic allocation of memory. The additional memory overhead of dynamic data structures depends on their type and size. A `deque` from the C++ standard template library typically allocates a section of memory that is substantially bigger than that needed to store a single element. As further elements are stored, the total memory consumed does not increase until all that memory is filled. It then allocates another section of the same size. Therefore, `deques` are not a good choice for storing small numbers of elements if conserving memory is important. The typical strategy for a `vector` from the standard template library is to initially allocate little or no memory for storage. When the capacity of the `vector` is exceeded, the memory allocation is doubled. As a result, `vectors` are memory efficient for storing small numbers of items, but for large numbers the difference between the allocated memory and the desired memory can be quite substantial. This discrepancy can be eliminated if the number of items to be stored is known in advance. At present, network models tend to be expressed in terms of connection probabilities or distributions of the number of incoming or outgoing synapses. Consequently, whereas it is generally possible to predict the number of neurons in a network model, the same is not true for the number of synaptic connections a neuron makes to targets on a specific machine.

Using the empirically parameterized model, we found that the ranking of components of the simulator with respect to their memory consumption varied with the numbers of processes. For tens to hundreds of processes, synaptic connections account for most of the memory consumption in a fixed size problem. At around 700 processes, the connection infrastructure that allows the synapses to be stored and accessed becomes the most dominant component, followed by the neuronal infrastructure at around 7000 processes. This analysis revealed that serial terms of the infrastructure components cause a saturation in the scalability of the simulator with respect to memory consumption.

As well as analyzing the memory consumption for the implemented data structures, we showed that the technique can also be applied to predicting memory consumption for alternative designs. These two features enable us to determine, in a principled fashion, which components are the highest priority to be improved, and which of multiple competing design strategies will be most effective. By comparing the memory consumption of potential designs to ideal cases, i.e., where there is no overhead for infrastructure, it is also possible to determine if there is anything more to be gained by improving a given data structure.

Whereas this approach would be useful for any range of processes, it is particularly helpful for large values of M. Without a principled approach to analyzing memory consumption as described here, developers are reduced to assumptions and intuition. These can both be misleading, especially for extremely large numbers of objects with a high degree of sparseness, resulting in long development cycles and wasting precious time allocations on supercomputers.

In the case of NEST, we discovered that exploiting sparseness in the neuronal and connection infrastructures was the most promising way to reduce memory usage, rather than, for example, reducing the size of neurons or synapses. It turned out that a design based on sparse tables results in greater improvements to the memory consumption than one based on look-up tables. A design based on cuckoo hashing consumes a similar amount of memory to a sparse table design, but is more complex. Therefore, we selected sparse tables as the most memory efficient strategy to exploit sparseness. Reducing the memory overhead for neurons with local synapses m_c^+ was also useful, indeed for architectures with a number of processes less than 50,000, this is the single most effective way to reduce memory consumption. Taken together, these adaptations increased the maximum network size that can be represented on the JUGENE supercomputer by an order of magnitude. This is assuming a high number of synapses per neuron – our calculations are based on $K = 10{,}000$ throughout. Clearly, a lower value would allow larger networks to be represented. Similarly, the assumption of random connectivity represents a worst-case, as it results in neurons typically having zero or one synapses on a given process. Topological considerations that result in fewer and longer lists of outgoing synapses for a given source neuron will also decrease the memory consumption, and so allow larger problems to be represented.

Meso- and macroscale simulations will for the next 10 years be bound to very large scale hardware, where scientists depend on using the architectures they have access to at computer centers. To be a useful tool, freely available simulators such as NEST need to work efficiently across a range of such architectures, as opposed to being optimized for a given architecture (e.g., Ananthanarayanan and Modha, 2007). This requires them to have a design that allows infrastructure optimizations without affecting user- and model-developer interfaces, models of the software, e.g., its memory consumption as developed here, to provide an understanding of bottlenecks and help in prioritizing optimizations, and a modular design that separates concerns as much as possible. The benefit of such a modular design is adequately demonstrated by the fact that we could perform all the described adaptations without requiring any changes at the user or model-developer level. Ultimately, the increasing range of hardware sizes may demand greater flexibility in the design of data structures, such that the size of the hardware determines which data structures are used at either compile or run-time. In the absence of the constraint of flexibility across architectures, a further use of the memory model is to help prioritize development resources with respect to a particular intended problem or machine. For example, for maximum cluster sizes below 10,000, exploiting sparseness could be considered a priority for the neuronal infrastructure but not for the connection infrastructure (see **Figure 5**).

Irrespective of the range of M for which the application should exhibit good performance, in general, trade-offs must be negotiated between memory consumption and computation time. Current trends indicate that hardware will move to increasingly multicore designs with massive processing power and limited and slow memory. Our adaptations here reflect this trend, as we replace algorithmically simple but memory-demanding representations with compact, but algorithmically more complex representations. For example, storing data in a hash table rather than a `vector` requires applying the hash function to the item key to determine its location in the table, which is computationally more expensive. Similarly, one can imagine a design where synapses are not addressable by the index of the pre-synaptic neuron, as in all potential designs considered here, but in a contiguous block of all synapses of the same type. This would reduce the memory overhead for neurons with local targets substantially, but would necessitate additional indirections or a searching strategy to locate the outgoing synapses for a given neuron and thus may increase the run-time. On the other hand, reducing the memory footprint can increase cache efficiency and thus decrease the run-time, despite additional indirections. The model we have proposed here does not account for computation time, and thus cannot predict these effects.

In Morrison et al. (2005) we presented the data structures and algorithms that allow networks in the order of 10^5 neurons with 10^9 synapses to be routinely simulated on small clusters. As a result, local cortical networks with realistic connectivity and sparseness could be studied for the first time. Here, we develop adaptations to those data structures that allow networks of the order of 10^7 neurons and 10^{11} synapses to be simulated on a popular modern supercomputer architecture, the Blue Gene, whose characteristics are a very large number of processors and a modest amount of RAM per core. These advances allow networks with macroscopic connectivity to be studied, in which some functional circuits are closed and the majority of synapses are accounted for. Networks of 10^8 neurons and 10^{12} synapses are also realizable, by either increasing the number of cores available, or reducing the overhead for neurons with local targets m_c^+, see **Figure 9**.

However, we are then still about three orders of magnitude away from the human brain. What are the next challenges, assuming the trends of increasing number of cores but low RAM per core continue? If 1 GB per core is a reasonable prediction of the amount of memory available for each core, the next decisive barrier is at 10^9. Although we have reduced the serial overhead in the neuronal and connection infrastructure to a matter of a few bits by using sparse tables, when the number of neurons approaches the number of bytes available in a core's RAM, these serial terms will dominate the total memory consumption and cause the maximum network size to saturate. Therefore, data structures must be implemented that eradicate any serial overhead, at the cost of a greater number of computational operations required to test whether neurons are local or have local targets, and also to access the objects that are stored locally. By this size the degree of sparseness is so great that all-to-all communication would be extremely inefficient. A more sophisticated approach to route the exchange of spikes between neurons on different processes will be necessary, for example by mapping the topology of the model network to that of the underlying machine architecture (see Kozloski and Wagner, 2011).

Between 10^9 and 10^{11} neurons, the full size of the human brain, no decisive barrier is expected according to our model; greater network sizes should be realizable by increasing the machine size. In this regime the communication mechanisms are likely to become the cause of saturation, analogous to the serial terms in the memory consumption of data structures between 10^5 and 10^7 neurons and the parallelizability of simulation algorithms between 10^3 and 10^5 neurons.

ACKNOWLEDGMENTS

This study was conceived with our late colleague and friend Rolf Kötter. We dedicate this work to him. Partially funded by JUGENE Grant JINB33, BMBF Grant 01GQ0420 to BCCN Freiburg, EU Grant 15879 (FACETS), EU Grant 269921 (BrainScaleS), DIP F1.2, the Helmholtz Alliance on Systems Biology (Germany), the Next-Generation Supercomputer Project of MEXT (Japan), the Research Council of Norway under grant 178892/V30 (eNeuro), Neurex, and the Junior Professor Program of Baden-Württemberg.

REFERENCES

Ananthanarayanan, R., Esser, S. K., Simon, H. D., and Modha, D. S. (2009). "The cat is out of the bag: cortical simulations with 10^9 neurons and 10^{13} synapses," in *Supercomputing 09: Proceedings of the ACM/IEEE SC2009 Conference on High Performance Networking and Computing*, Portland, OR.

Ananthanarayanan, R., and Modha, D. S. (2007). "Anatomy of a cortical simulator," in *Supercomputing 2007: Proceedings of the ACM/IEEE SC2007 Conference on High Performance Networking and Computing* (New York, NY: Association for Computing Machinery).

Dietzfelbinger, M., Karlin, A., Mehlhorn, K., Meyer auf der Heide, F., Rohnert, H., and Tarjan, R. E. (1994). Dynamic perfect hashing: upper and lower bounds. *SIAM J. Comput.* 23, 738–761.

Fotakis, D., Pagh, R., Sanders, P., and Spirakis, P. (2003). "Space efficient hash tables with worst case constant access time," in *STACS 2003, Vol. 2607 of Lecture Notes in Computer Science*, eds H. Alt and M. Habib (Berlin: Springer), 271–282.

Gewaltig, M.-O., and Diesmann, M. (2007). NEST (Neural Simulation Tool). *Scholarpedia* 2, 1430.

Izhikevich, E. M., and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. U.S.A.* 105, 3593–3598.

Kozloski, J., and Wagner, J. (2011). An ultrascalable solution to large-scale neural tissue simulation. *Front. Neuroinform.* 5:15. doi: 10.3389/fninf.2011.00015

Kunkel, S., Potjans, T. C., Abigail, M., and Markus, D. (2009). "Simulating macroscale brain circuits with microscale resolution," in *Frontiers in Neuroscience Conference Abstract: Neuroinformatics 2009*. doi: 10.3389/conf.neuro.11.2009.08.044.

Lytton, W. W., and Hines, M. L. (2005). Independent variable time-step integration of individual neurons for network simulations. *Neural Comput.* 17, 903–921.

Migliore, M., Cannia, C., Lytton, W. W., Markram, H., and Hines, M. (2006). Parallel network simulations with NEURON. *J. Comput. Neurosci.* 21, 119–223.

Morrison, A., Aertsen, A., and Diesmann, M. (2007). Spike-timing dependent plasticity in balanced random networks. *Neural Comput.* 19, 1437–1467.

Morrison, A., Mehring, C., Geisel, T., Aertsen, A., and Diesmann, M. (2005). Advancing the boundaries of high connectivity network simulation with distributed computing. *Neural Comput.* 17, 1776–1801.

Pagh, R., and Rodler, F. F. (2004). Cuckoo hashing. *J. Algorithms* 51, 122–144.

Pecevski, D., Natschläger, T., and Schuch, K. (2009). PCSIM: a

parallel simulation environment for neural circuits fully integrated with python. *Front. Neuroinform.* 3:11. doi:10.3389/neuro.11.011.2009

Plesser, H. E., Eppler, J. M., Morrison, A., Diesmann, M., and Gewaltig, M.-O. (2007). "Efficient parallel simulation of large-scale neuronal networks on clusters of multiprocessor computers," in *Euro-Par 2007: Parallel Processing, Volume 4641 of Lecture Notes in Computer Science*, eds A.-M. Kermarrec, L. Bougé, and T. Priol (Berlin: Springer-Verlag), 672–681.

Potjans, T. C., and Diesmann, M. (2011). The cell-type specific connectivity of the local cortical network explains prominent features of neuronal activity. *Neurons Cogn.* arXiv:1106.5678.

Silverstein, C. (2005). Implementation of sparse_hash_map, dense_hash_map, and sparsetable. Available at: http://google-sparsehash.googlecode.com/svn/trunk/doc/implementation.html

Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi:10.1371/journal.pcbi.0010042

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 July 2011; accepted: 09 December 2011; published online: 24 January 2012.

Citation: Kunkel S, Potjans TC, Eppler JM, Plesser HE, Morrison A and Diesmann M (2012) Meeting the memory challenges of brain-scale network simulation. Front. Neuroinform. ***5****:35. doi: 10.3389/fninf.2011.00035*

frontiers in
NEUROINFORMATICS

ORIGINAL RESEARCH ARTICLE
published: 29 July 2011
doi: 10.3389/fninf.2011.00012

Large-scale analysis of gene expression and connectivity in the rodent brain: insights through data integration

Leon French*[1,2], *Powell Patrick Cheng Tan*[1] *and Paul Pavlidis*[2,3]

[1] Bioinformatics Graduate Program, University of British Columbia, Vancouver, BC, Canada
[2] Centre for High-Throughput Biology, University of British Columbia, Vancouver, BC, Canada
[3] Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

Edited by:
Olaf Sporns, Indiana University, USA

Reviewed by:
Jason Bohland, Boston University, USA
Lior Wolf, Tel-Aviv University, Israel

****Correspondence:***
Paul Pavlidis, Centre for High-Throughput Biology, University of British Columbia, 2185 East Mall, Vancouver, BC, Canada V6T 1Z4.
e-mail: paul@chibi.ubc.ca

Recent research in *C. elegans* and the rodent has identified correlations between gene expression and connectivity. Here we extend this type of approach to examine complex patterns of gene expression in the rodent brain in the context of regional brain connectivity and differences in cellular populations. Using multiple large-scale data sets obtained from public sources, we identified two novel patterns of mouse brain gene expression showing a strong degree of anti-correlation, and relate this to multiple data modalities including macroscale connectivity. We found that these signatures are associated with differences in expression of neuronal and oligodendrocyte markers, suggesting they reflect regional differences in cellular populations. We also find that the expression level of these genes is correlated with connectivity degree, with regions expressing the neuron-enriched pattern having more incoming and outgoing connections with other regions. Our results exemplify what is possible when increasingly detailed large-scale cell- and gene-level data sets are integrated with connectivity data.

Keywords: connectome, genome, transcriptome, neuron, oligodendrocyte

INTRODUCTION

Understanding gene function requires the analysis of interactions among them, and ultimately unraveling the function of the genome will require comprehending how all of the parts interoperate in complex networks. An analogous situation exists for the brain and its regional connectome (Bota et al., 2003; Sporns et al., 2005; Lichtman and Sanes, 2008; Biswal et al., 2010; Sporns, 2011). Given the relationships between these two systems (genome and connectome), as well as the fact they are both complex networks, it is natural to ask how analysis of one can inform understanding of the other. Indeed, the integrated analyses of the connectome with other modalities will be critical to understanding brain function. In this paper our modality of interest is gene expression, for which extensive information exists.

It is obvious that the connectome is related to the genome. Axon pathfinding, target recognition, synapse formation, and plasticity are tightly controlled by gene expression (Ressler et al., 2002; Polleux et al., 2007). The function of synapses requires the coordinated expression of genes directing the synthesis of neurotransmitters in the presynaptic cell and of receptors in the post-synaptic cell. Because high throughput experimental technologies for studying the genome are well developed, in many ways our understanding of gene expression and gene networks is better than for the connectome (though this situation is changing rapidly). This allows the collection of large data sets describing gene expression patterns at high levels of resolution. It is increasingly feasible to use this molecular level information to elucidate neuroanatomy.

Analysis of connectomes with transcription data began with the nematode *C. elegans* because neuron-level connectivity and gene expression levels are known. (White et al., 1986; Harris et al., 2010). Neuron-level gene expression data in *C. elegans* is not available for all genes, but there is enough to perform reasonably large-scale analyses. The earliest study integrated the connection and expression profiles of 280 neurons and 292 genes (Varadan et al., 2006). Varadan and colleagues employed a systems-based approach to discover logical gene expression-based rules that predict connectivity. Within the resulting gene modules they found high levels of "multivariate synergy," suggesting statistically interacting genes were more important than single genes. The authors extracted several gene sets that correlate expression in pre- and post-synaptic neurons to presence of gap and chemical synapses. Interestingly, gene sets which contained the most information about the formation of synapses included cell adhesion molecules, transcription factors and axon guidance cues.

Kaufman et al. (2006) performed a similar analysis. They found a more general statistical relationship between gene expression and connectivity. Their analysis employed a co-variation correlation assay, also known as a Mantel test. The Mantel test correlates similarity or distance measures across common objects (in this case, neurons). The Mantel correlations found by Kaufman et al. (2006) were up to 0.18. This signal, while statistically significant, is not strong enough to allow prediction of connectivity from gene expression. Using an optimization method, Kaufman et al. (2006) identified a set of 15 genes whose expression patterns carried the most information about connectivity. Similar to the results of Varadan et al. (2006), they found that a statistically significant number of these were previously linked to synaptogenesis, neuron type, axon guidance, and development.

A third *C. elegans* study, by Baruch et al. (2008) focused on finding relationships between gene expression and certain aspects of synapse formation (Baruch et al., 2008). They used expression profiles to model the type of synapse (e.g., electrical or chemical) between

connected neurons. Like Varadan et al. (2006) they employed a machine learning method to find gene expression-based logical rules, and the genes found to be most predictive of connection type often had known functional roles in neural development.

Similar analyses are starting to appear for the mammalian brain, though in terms of data the situation is the opposite of that for the worm: gene expression is more fully described than connectivity. Dong et al. (2009) provided a fascinating glimpse into the relationships between brain wiring and gene expression in the mammalian brain (Dong et al., 2009). They studied the Allen Mouse Brain Atlas (ABA) for spatial gene expression profiles that segmented the hippocampal field CA1 along its longitudinal axis. Nine of the genes that segmented the CA1 field had concordant expression patterns in the lateral septal nucleus, apparently reflecting the patterns of projections between the respective dorsal and ventral aspects of the two regions. Dong et al. (2009) were able to interpret the CA1 segmentation from the perspective of brain function and connectivity. They noted that the ventral half is linked to goal-oriented and autonomic response while the dorsal half plays roles in navigation.

Recently we contributed a larger-scale analysis of gene expression and connectivity, considering 142 rodent brain regions (French and Pavlidis, 2011). While the connectivity data we used is not a complete connectome, extensive expression data is available for 17,530 genes. Using methods similar to those of Kaufman et al. (2006), we found a Mantel correlation of 0.25 between connectivity and gene expression. An important aspect of our analysis was to account for the fact that connectivity and gene expression are spatially correlated (nearby brain regions are more likely to share connections and expression patterns). Explicit control for this spatial autocorrelation allowed us to identify a signal attributable to connectivity (Mantel correlation of 0.13). A reduced set of genes with a higher Mantel test score was obtained by using the same greedy optimization procedure as Kaufman and colleagues. This reduced set of genes was enriched for axon guidance genes and contained a few of the genes noted in the CA1 study (Dong et al., 2009). This set was also enriched for genes previously linked to autism. While the forebrain and interbrain divisions showed the strongest signals, the interpretation of the results was focused on properties of the genes instead of the connections or brain regions. Recently, a second study replicating the finding of correlations between gene expression and connectivity appeared (Wolf et al., 2011). Wolf and colleagues showed that machine learning methods could be used to predict connectivity from gene expression patterns in a statistically significant manner, for approximately one half of tested brain regions. They showed that genes known to be associated with schizophrenia, autism, and attention deficit disorder are enriched in their gene sets that predict connectivity. Although the authors did not perform correction for the effect of spatial autocorrelation, they tested the robustness of the connectivity data and the quality of the expression images from the Allen Brain Atlas.

A limitation of previous studies integrating gene expression and connectivity is the challenge of interpreting the patterns observed in terms of other parameters such as cellular composition of different brain regions. In the current paper, we extend our earlier work, starting with a directed search for expression patterns of interest. We hypothesized that expression patterns that strongly distinguish brain regions from each other might be functionally relevant and potentially related to connectivity. We were specifically interested in gene pairs with expression patterns showing strong negative correlations across multiple brain regions. We then use connectivity data as well as information on cell-type-specific gene expression to further dissect and ascribe biological meaning to the patterns we identified. In addition to identifying a novel pattern of gene expression in the mouse brain, our analysis serves as a demonstration of how a complex gene expression pattern can be dissected using multiple data types including connectivity.

MATERIALS AND METHODS

Data and methods were based on that used in our previous study (French and Pavlidis, 2011). Briefly, mouse brain gene expression data is from high-resolution colorimetric *in situ* hybridization observations in the Allen Mouse Brain Atlas (ABA; Lein et al., 2007). To obtain brain region connection degree we counted the number of incoming or outgoing connections from the up-propagated rat brain connectivity data we previously extracted from the Brain Architecture Management System (BAMS; Bota et al., 2005). For neuroanatomical matching and selecting we again used non-overlapping regions and included virtual regions that better mapped between the ABA and BAMS brain region names. The region mappings are accessible as Supplementary Table S1 in French and Pavlidis (2011). Supplementary data and software are available at http://www.chibi.ubc.ca/NEOE.

GENE EXPRESSION DATA

We employed the expression energy quantifications of the ABA images. For each image set the expression energy of every voxel is defined as the product of expression area and expression intensity (Ng et al., 2009). Pixels are averaged within voxels and brain regions to provide a single expression energy value for each brain region. To reduce computation time and filter genes of low and constant expression values we restricted our analysis to genes for which ABA has expression patterns in coronal sections. This set of 4261 image series (3976 genes) were assayed by ABA in the coronal plane because they showed marked regional expression patterns in the sagittal plane (Ng et al., 2009). Most "housekeeping" genes which tend to have widespread expression are not present in the set. Some genes were represented by more than one imageseries (that is, there are replicate data sets in the Allen Atlas), which were kept separate in our analysis. To create a single expression profile for a set of genes we averaged the expression values per region.

For analysis of expression data alone, we used 150 non-overlapping ABA regions. When connectivity data was used the regions were limited to those for which we had connectivity data: 112 regions for outgoing, 141 for incoming connectivity, and 142 resulting from joining the two.

STATISTICAL ANALYSIS

To compare expression energy to spatial location and connectivity degree we compute Spearman rank correlation coefficients (ρ). Statistical significance was established by resampling 1000 gene sets of the same size to generate empirical null distributions. This provides the probability that an equally sized gene set randomly chosen from the set of all genes scores a higher correlation. We used linear regression for computing partial correlation coefficients. Principal component analysis (PCA) was performed after rescaling

the gene profiles to a common mean and variance. We employed the complete-linkage agglomeration method for hierarchical clustering with the Euclidean distance function.

CELL-TYPE ENRICHED GENE LISTS

Cell-type enriched gene sets were extracted from the "The Transcriptome Database for Astrocytes, Neurons, and Oligodendrocytes" (Cahoy et al., 2008). The database contains gene expression profiles of cell-type purified mature mouse forebrain samples. Mouse gene symbols were extracted from Supplementary Tables S4–S6 of Cahoy et al. (2008). These tables provide lists of astrocyte, neuron, and oligodendrocyte enriched genes. After removing genes that are not in the ABA coronal gene set, 716 astrocyte, 831 neuron, and 571 oligodendrocyte enriched genes remain.

GENE ONTOLOGY ENRICHMENT

We used the ErmineJ software to extract overrepresented Gene Ontology (GO) groups (Ashburner et al., 2000; Gillis et al., 2010). The set of 3976 coronal genes formed the background gene list for the over-representation analysis. GO groups were limited to the biological process division and required 5–300 annotated genes.

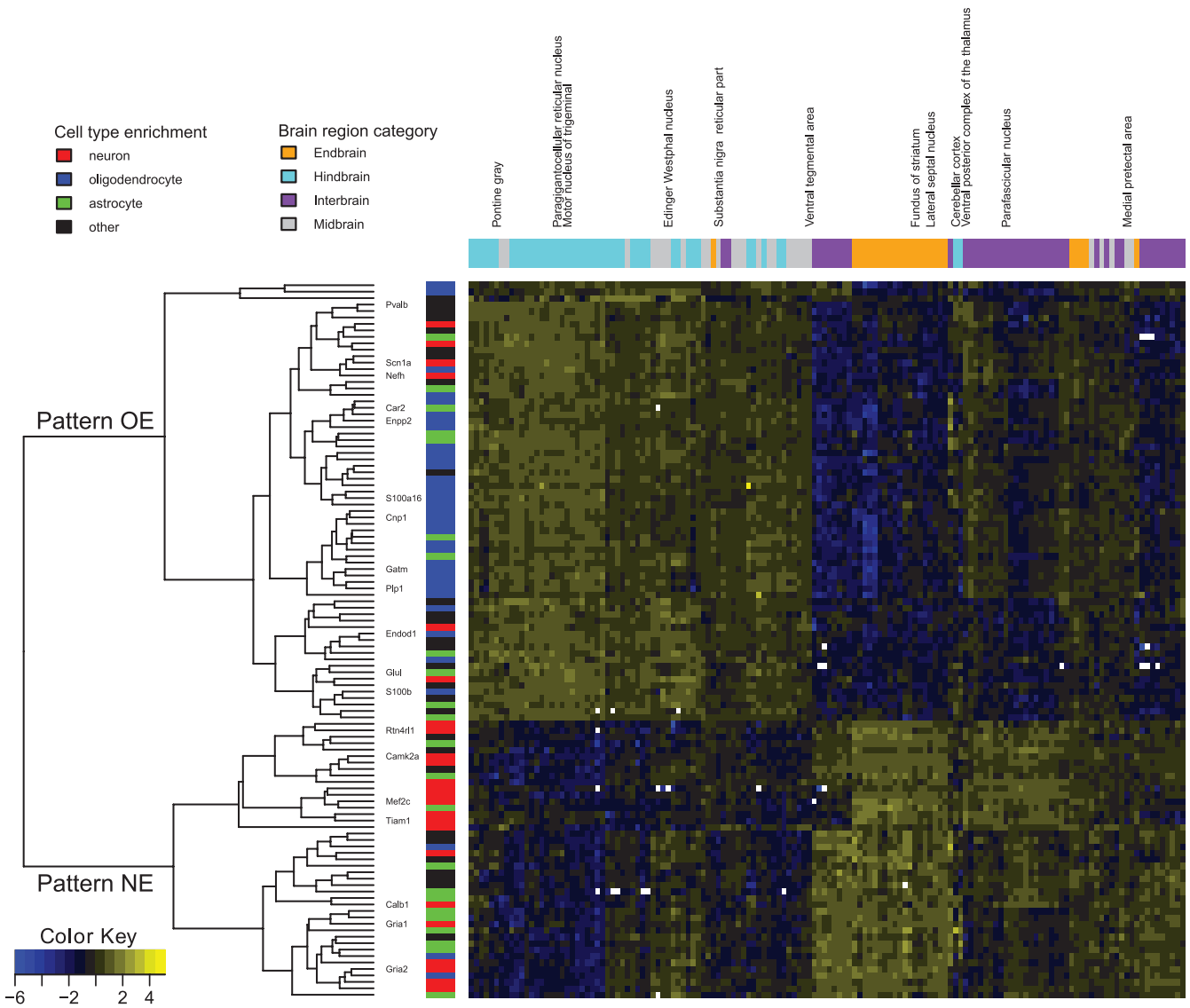


FIGURE 1 | Expression patterns of genes involved in the top 912 negative expression correlations. Normalized expression is color coded, ranging from blue (low) to yellow (high) and in white for missing values. Genes mentioned in the article are labeled. Gene membership in the transcriptome database for astrocytes (green), neurons (red), and oligodendrocytes (blue) is marked (Cahoy et al., 2008). The dendrogram shows the split between pattern NE and pattern OE. Brain regions are colored as orange for endbrain, cyan for hindbrain, purple for interbrain, and gray for midbrain. Expression data for each gene was normalized to mean zero and variance one for contrast.

ORTHOLOG ASSIGNMENT

For each gene we extracted its homologous sequences from the HomoloGene database (build version 64; Wheeler et al., 2007). HomoloGene groups were used to convert the mouse gene identifiers to genes from *S. cerevisiae* (yeast), *C. elegans* (worm), and *D. melanogaster* (fly).

RESULTS

To identify genes showing strong negatively correlated expression patterns with other genes, we ranked all pairs of genes in the data set by their Spearman correlations across 150 ABA brain regions, and considered pairs with the strongest negative correlations. By filtering gene–gene correlations at a maximum Spearman's rank correlation coefficient (ρ) of −0.72 we selected the 912 most anti-correlated gene pairs. We choose this stringent but arbitrary threshold because we wanted a small list that could be manually examined for interesting relationships, though our findings proved to hold for other reasonable selection thresholds.

Our first observation was that this list of 912 pairs includes only 102 different genes, indicating there would be strong positive correlations present within this set, rather than numerous distinct patterns. Hierarchical clustering and visualization of the expression patterns of these genes (**Figure 1**) shows that the original 912 inversely correlated patterns are essentially one inverse relationship corresponding to two gene expression profiles. Visualization of all gene–gene correlations within the set demonstrates this relationship with a clear bimodal distribution with peaks at −0.6 and 0.7 (**Figure A1** in Appendix). To further examine the inverse relationship we use clustering to divide the data into two sets: pattern NE (43 image series, 40 genes, **Table 1**) and pattern OE (68 image series, 62 genes, **Table 2**). This choice of names will be clarified later in our results. Supplemental Data sheets 1 and 2 list NCBI and Allen image series identifiers. **Figure 2** shows expression energy images in the sagittal plane for a pattern NE (CamK2a) and OE gene (S100b). The average profiles of these patterns are strongly negatively correlated (Spearman's rank correlation (ρ) = −0.88). Given the strength of this pattern, although it only includes a small fraction of the genes studied, we asked if it might correspond to patterns uncovered by PCA. We found the pattern NE and OE genes are strongly separable in PC2 (**Figure 3**) and the mean loadings in PC1 differ significantly (p-value < 0.001). Thus these patterns correspond to major trends in the data.

Inspection of the gene names and symbols suggested that pattern NE was enriched for neuron-associated genes such as calcium/calmodulin-dependent protein kinase II alpha (CamK2a; Ouimet et al., 1984) and calbindin-28K (Calb1; Pfeiffer et al., 1989). In contrast, several glial cell markers appear in the pattern OE list: carbonic anhydrase II (Car2; Ghandour et al., 1979, 1980), S100b (Ghandour et al., 1981; Rosengren et al., 1986), and glutamine synthetase (Glul; Wu et al., 2005). Also, one neuron marker, neurofilament high molecular weight (Nefh) appears in the pattern OE list (Letournel et al., 2006). We note that none of the ABA regions are white matter tracts (most are small nuclei), so the pattern does not reflect a simple contrast between gray and white matter.

Gene Ontology enrichment analysis allowed us to objectively quantify these trends. The GO provides extensive annotations of genes that allow testing for enrichment of specific functions, subcellular localizations, or processes. By looking for annotations overrepresented in patterns NE or OE we find several interesting groups, though none reach significance after multiple test correction. For pattern NE the top ranked groups include "regulation of transport" (GO:0051049, p-value = 8.3×10^{-5}) and "regulation of neurotransmitter secretion" (GO:0046928, p-value = 0.0035). Pattern OE is enriched for groups such as "potassium ion transport" (GO:0006813, p-value = 0.0047), "cellular ion homeostasis"

Table 1 | Pattern NE gene symbols and names.

Gene symbol	Name
6720401G13Rik	RIKEN cDNA 6720401G13 gene
Calb1	Calbindin-28K
CamK2a	Calcium/calmodulin-dependent protein kinase II alpha
Camkv	CaM kinase-like vesicle-associated
Cenpf	Centromere protein F
Cox6a2	Cytochrome c oxidase, subunit VI a, polypeptide 2
Cpne2	Copine II
Cpne7	Copine VII
Cyln2	Cytoplasmic linker 2
Dusp6	Dual specificity phosphatase 6
E2f1	E2F transcription factor 1
Egr3	Early growth response 3
Fos	FBJ osteosarcoma oncogene
Gria1	Glutamate receptor, ionotropic, AMPA1 (alpha 1)
Gria2	Glutamate receptor, ionotropic, AMPA2 (alpha 2)
Grik5	Glutamate receptor, ionotropic, kainate 5 (gamma 2)
Heatr5b	HEAT repeat containing 5B
Hpcal4	Hippocalcin-like 4
Itm2c	Integral membrane protein 2C
Kalrn	Kalirin, RhoGEF kinase
Ly6h	Lymphocyte antigen 6 complex, locus H
Mef2c	Myocyte enhancer factor 2C
Mef2d	Myocyte enhancer factor 2D
Nnat	Neuronatin
Ntrk2	Neurotrophic tyrosine kinase, receptor, type 2
Ogt	O-linked *N*-acetylglucosamine (GlcNAc) transferase (UDP-*N*-acetylglucosamine: polypeptide-*N*-acetylglucosaminyl transferase)
Pdgfra	Platelet derived growth factor receptor, alpha polypeptide
Pea15	Phosphoprotein enriched in astrocytes 15
Pkia	Protein kinase inhibitor, alpha
Ppap2b	Phosphatidic acid phosphatase type 2B
Prkcc	Protein kinase C, gamma
Psg16	Pregnancy specific glycoprotein 16
Ptprz1	Protein tyrosine phosphatase, receptor 1 type Z, polypeptide
Rtn4rl1	Reticulon 4 receptor-like 1
Shisa9	Shisa homolog 9 (*Xenopus laevis*)
Sirpa	Signal-regulatory protein alpha
Slc27a1	Solute carrier family 27 (fatty acid transporter), member 1
Tiam1	T-cell lymphoma invasion and metastasis 1
Tnrc4	Trinucleotide repeat containing 4
Unc84a	Unc-84 homolog A (*C. elegans*)

Table 2 | Pattern OE gene symbols and names.

Gene symbol	Name
3632451O06Rik	RIKEN cDNA 3632451O06 gene
Acyp2	Acylphosphatase 2, muscle type
Adssl1	Adenylosuccinate synthetase like 1
Ankrd34b	Ankyrin repeat domain 34B
Arhgef10	Rho guanine nucleotide exchange factor (GEF) 10
Armc2	Armadillo repeat containing 2
Aspa	Aspartoacylase (aminoacylase) 2
B630019K06Rik	RIKEN cDNA B630019K06 gene
Bcat1	Branched chain aminotransferase 1, cytosolic
Cables2	Cdk5 and Abl enzyme substrate 2
Car2	Carbonic anhydrase 2
Cldn11	Claudin 11
Cnp1	Cyclic nucleotide phosphodiesterase 1
Cnp1	Cyclic nucleotide phosphodiesterase 1
Cryab	Crystallin, alpha B
Cyp27a1	Cytochrome P450, family 27, subfamily a, polypeptide 1
Daam2	Disheveled associated activator of morphogenesis 2
Ddt	d-Dopachrome tautomerase
Dip2a	DIP2 disco-interacting protein 2 homolog A (Drosophila)
Elovl5	ELOVL family member 5, elongation of long chain fatty acids (yeast)
Endod1	Endonuclease domain containing 1
Enpp2	Ectonucleotide pyrophosphatase/phosphodiesterase 2
Fa2h	Fatty acid 2-hydroxylase
Fts	Fused toes
Galnt6	UDP-*N*-acetyl-alpha-d-galactosamine: polypeptide *N*-acetylgalactosaminyltransferase 6
Gatm	Glycine amidinotransferase (l-arginine: glycine amidinotransferase)
Glra1	Glycine receptor, alpha 1 subunit
Glul	Glutamate–ammonia ligase (glutamine synthetase)
Gprc5b	G protein-coupled receptor, family C, group 5, member B
Hcn2	Hyperpolarization-activated, cyclic nucleotide-gated K+ 2
Kcng4	Potassium voltage-gated channel, subfamily G, member 4
Kctd9	Potassium channel tetramerization domain containing 9
Klk6	Kallikrein 6
Lgi3	Leucine-rich repeat LGI family, member 3
Limk1	LIM-domain containing, protein kinase
Map2k6	Mitogen activated protein kinase kinase 6
Mmel1	Membrane metallo-endopeptidase-like 1
Nefh	Neurofilament, heavy polypeptide
Nifun	NifU-like N-terminal domain containing
Nrg1	Neuregulin 1
Pacs2	Phosphofurin acidic cluster sorting protein 2
Plekhb1	Pleckstrin homology domain containing, family B (evectins) member 1
Plp1	Proteolipid protein (myelin) 1
Pnkd	Paroxysmal non-kinesigenic dyskinesia
Prune2	Prune homolog 2 (Drosophila)
Pvalb	Parvalbumin
Qdpr	Quininoid dihydropteridine reductase
Rnd2	Rho family GTPase 2
Rnf13	Ring finger protein 13
S100a16	S100 calcium binding protein A16
S100b	S100 protein, beta polypeptide, neural
Scn1a	Sodium channel, voltage-gated, type I, alpha
Sema7a	Sema domain, immunoglobulin domain (Ig), and GPI membrane anchor, (semaphorin) 7A
Serpinb1c	Serine (or cysteine) peptidase inhibitor, clade B, member 1c
Sgpp2	Sphingosine-1-phosphate phosphatase 2
Slc12a2	Solute carrier family 12, member 2
Slc39a14	Solute carrier family 39 (zinc transporter), member 14
Slc44a1	Solute carrier family 44, member 1
Slc4a2	Solute carrier family 4 (anion exchanger), member 2
Slc6a5	Solute carrier family 6 (neurotransmitter transporter, glycine), member 5
Syt2	Synaptotagmin II
Vamp1	Vesicle-associated membrane protein 1
Zfyve9	Zinc finger, FYVE domain containing 9

(GO:0006873, p-value = 0.013), and "regulation of membrane potential" (GO:0042391, p-value = 0.0015). The full GO analysis is available as Data sheet 3 in Supplementary Material.

By linking homologous sequences we quantified how evolutionary recent the pattern NE and OE genes are. Surprisingly, only three of the pattern NE genes had a homolog in yeast, worm, or fly genomes (7.5%, p-value = 0.00023, hypergeometric test). The pattern OE group had 23 (37%, $p = 0.067$) of earlier origin, slightly more than the faction seen in the entire coronal gene set (32%). Both sets had about the expected number of detected orthologs in the human genome.

We used a third bioinformatics approach to test whether these two patterns might reflect differences in cellular populations, using the Transcriptome Database for Astrocytes, Neurons, and Oligodendrocytes (Cahoy et al., 2008). **Figures 1 and 4** show that pattern NE is enriched for genes identified by Cahoy et al. (2008) as being neuron enriched (p-value = 0.0016, hypergeometric test). In contrast, pattern OE has half the number of expected number of neuron-enriched genes (p-value = 0.015). For the Cahoy oligodendrocyte genes the opposite pattern appears, with 29 genes in pattern OE (p-value < 0.0001). Genes from the Cahoy "astrocyte" gene set were represented approximately equally in both sets at the expected proportions. Similar results were obtained by using the lists of oligodendrocyte and neuron-enriched gene sets from the ABA (Lein et al., 2007). These strong cell-type signals led us to label the two gene sets as neuron enriched (NE) and oligodendrocyte enriched (OE).

The results presented thus far are limited to information obtained at the gene level. While the two profiles seem to have a relationship to cell type, we wanted to test if they provide infor-

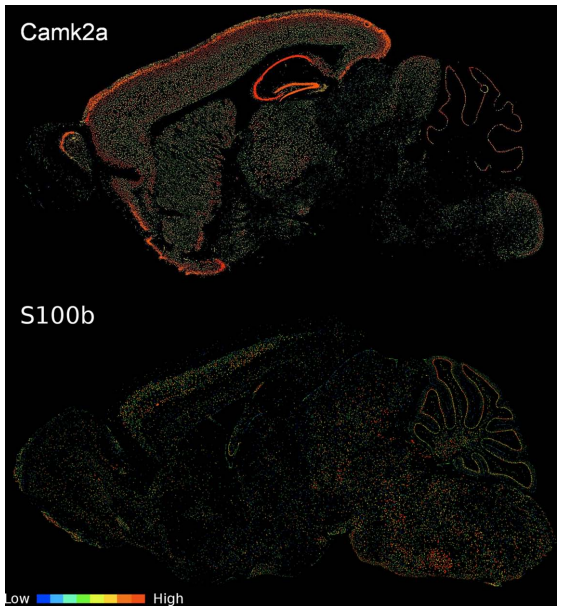


FIGURE 2 | Sagittal expression energy images of a pattern NE and OE gene. CamK2a displays pattern NE (image series 79360274) and S100b shows pattern OE (image series 924). Images were downloaded from the ABA web site (http://www.brain-map.org). While all expression information for the analysis is from coronal assays, we selected a sagittal view to better show interregional variability in a single section.

mation about higher-level brain structure. Our next analysis stage incorporated information on spatial locations within the brain and connectivity.

We first summarized patterns NE and OE as the average of the expression patterns of the gene sets. While pattern OE has slightly lower expression levels on average, the two patterns have very similar variance. This expression pattern across regions was found to be significantly correlated with the anterior–posterior axis: regions that have high pattern OE expression tend to be at the posterior end of the brain (Spearman's $\rho = 0.81$), with the opposite true of pattern NE ($\rho = -0.76$). We previously noted that regions in the posterior end of the brain had fewer connections ($\rho = 0.55$; French and Pavlidis, 2011). Accordingly we found that the expression patterns correlated with the number of connections the regions have. For incoming connectivity degree the Spearman correlations are 0.49 and −0.54 for pattern NE and OE respectively (141 brain regions). For the 112 regions that have at least one report of an outgoing connection the correlations are 0.32 and −0.44 for pattern NE and OE respectively. Joining the incoming and outgoing connections provides 142 brain regions with correlations of 0.48 (pattern NE) and −0.59 (pattern OE). This means that higher expression of pattern NE is found in "hub-like" regions with many connections, and high expression of pattern OE is observed in "relay-like" regions with few connections. The relationship is shown in **Figure 5** with regions of high connectivity degree with low pattern OE expression and high pattern NE expression. All of the above correlations are significant at $p < 0.001$. It is important to note that the entire

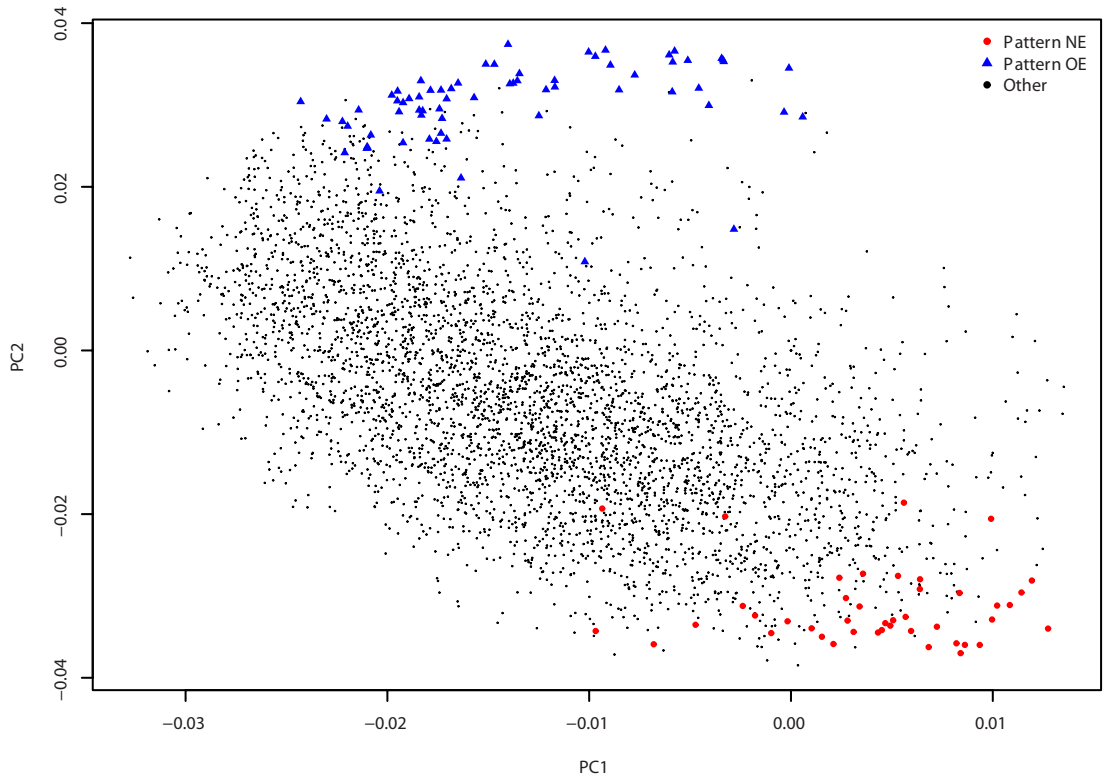


FIGURE 3 | Principal components analysis. Gene loadings for pattern NE (red circles), pattern OE (blue triangles), and all other genes (small black circles) are plotted. The first two principal components, PC1 (16.4% of the variance) and PC2 (11.8% of the variance), separate the two patterns.

coronal gene set has substantial correlations of expression levels to anterior–posterior axis ($\rho = 0.29$), incoming ($\rho = -0.19$), outgoing connection degree ($\rho = -0.25$). This spatial correlation reflects a bias in the coronal set gene selection, which favored genes expressed in the cortex and hippocampus (Ng et al., 2009). Against this baseline, the anterior–posterior expression gradient of the pattern NE and pattern OE genes is still very high.

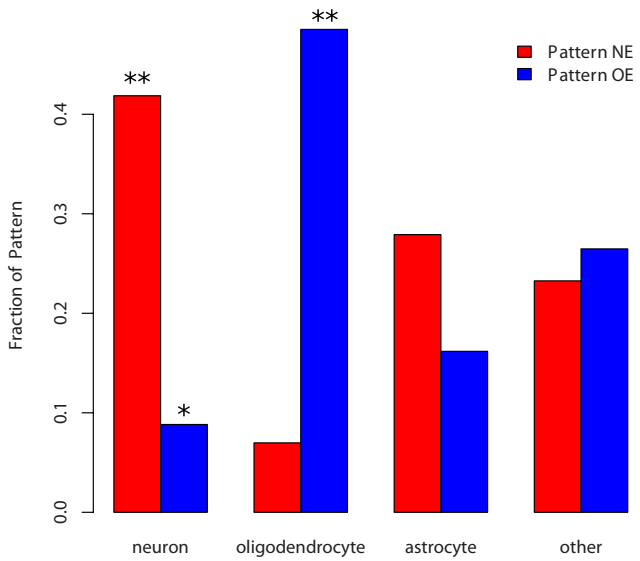


FIGURE 4 | Fraction of cell-type enriched genes appearing in the two patterns. *p*-values below 0.05 are marked by * and below 0.005 with **. Neuron-enriched genes are overrepresented in the NE list and underrepresented in the OE list. Oligodendrocyte genes are overrepresented in the OE list but not significantly underrepresented in the NE list.

Because of the known relationship between spatial location in the brain and patterns of connectivity, we sought to correct for this in our analysis of the NE and OE patterns, using partial correlations. We found that the correlations with incoming connectivity degree are still significant after correction for anterior–posterior location, with correlations of 0.20 (pattern NE) and −0.30 (pattern OE). Similarly, the outgoing degree correlations were still significant, though reduced in magnitude: 0.07 (pattern NE, p-value = 0.001) and −0.30 (pattern OE). Correlations to the combined degree across 142 regions are 0.16 (pattern NE) and −0.35 (pattern OE; all of the above correlations are significant at $p < 0.001$ unless otherwise noted). A similar analysis carried out using the full Cahoy "neuron" and "oligodendrocyte" lists show similar trends, albeit much weaker than patterns NE and OE. Expression of the Cahoy astrocyte-enriched genes is not significantly correlated with connectivity degree or anterior–posterior axis ($p > 0.1$).

Given the relationship of the NE and OE sets to connectivity, it was of interest to test how these gene sets relate to our previous results (French and Pavlidis, 2011). In that work we presented gene sets that had optimal expression correlations with connectivity (see Introduction for more detail). Unlike the present study, those genes were selected on the basis of correlation with connectivity patterns, not the number of connections. Thirteen of the pattern NE genes and five pattern OE genes overlap with the connectivity-

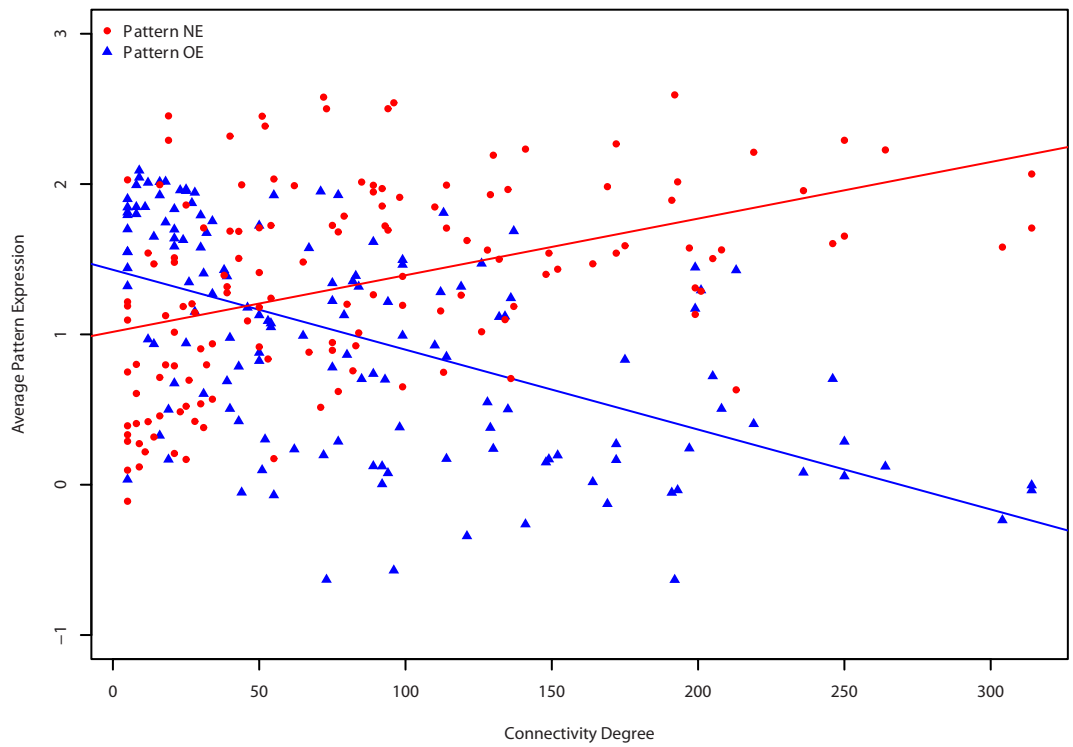


FIGURE 5 | Relationships between degree and expression patterns. Connectivity degree is plotted against average pattern NE (red circles) and OE (blue triangles) expression levels for each brain region. Degree for the 142 regions is the sum of both incoming and outgoing connections.

optimized gene set from French and Pavlidis (2011). Accordingly, using the same analysis method of French and Pavlidis (2011) to evaluate relationships between expression and connectivity patterns, the pattern NE and OE genes (pooled) were not significant. In an analysis more closely related to the one undertaken here, French and Pavlidis (2011) also ranked genes by correlation of expression level with connectivity degree, but this was not studied in any detail. Unsurprisingly, the pattern NE and OE sets have strong overlaps with these lists: 89% of the pattern OE genes have significant negative degree correlations and 65% of pattern NE genes have significant positive degree correlations.

DISCUSSION

In this paper we have shown how a complex expression pattern in the rodent brain can be dissected in terms of genes, cell types, spatial location, and connectivity. To our knowledge, the expression patterns we identified have not been previously described. However, previous work has uncovered possible links between neuroanatomy, gene expression, and cell type. Using a voxel-based PCA on a subset of the ABA data, Bohland et al. (2009b) noted that the two most separable structures, the striatum and cerebellum, contain a relatively large number of GABAergic inhibitory neurons. There are a number of differences between the analysis of Bohland et al. (2009b) and ours, including the use of voxels vs. brain regions and the choice of genes analyzed, so it is not easy to compare them (indeed it appears the components in the two PCAs are not equivalent), but it is likely that at least some of the highly weighted genes in the pattern identified by Bohland et al. (2009b) are genes in the pattern we found. A second study has examined a link between expression and connectivity for two specific brain regions (Ng et al., 2009). Using the anatomic gene expression atlas (AGEA) Ng et al. (2009) visualized correlated expression profiles of the parafascicular nucleus and the ventral posterior complex. The ventral posterior complex is a "relay nucleus" and has fewer connections than the hub-like parafascicular nucleus. The AGEA visualization demonstrated that the regions have diverse expression correlation maps that might reflect their diverse function (Ng et al., 2009). In agreement with this result, in our analysis the highly connected parafascicular nucleus has high expression of the neuron-enriched pattern NE compared to the ventral posterior complex. For the oligodendrocyte enriched pattern OE the opposite is true. Our results are consistent with the idea that degrees of connectivity might be reflected in expression pattern.

By comparing to our previous results on connectivity correlations with expression (French and Pavlidis, 2011), we found the patterns seems to contain weak information about connection partners. However, when compared to the background gene set neither of pattern NE and OE gene lists carry significant information about connectivity patterns beyond number of connections. As expected the pattern NE and OE lists do overlap significantly with the previously extracted lists of 887 and 1127 genes that had expression levels positively and negatively correlated with degree (p-value < 0.0001, hypergeometric test). In contrast our lists provide a much smaller set of genes that provide an interesting link to cell-type distribution. Thus patterns NE and OE are distinct from the connectivity-related patterns of French and Pavlidis (2011).

Patterns NE and OE are suggestive of differences in the relative proportion of neuronal and glial cell populations in the brain regions in which they are expressed. We further hypothesize that the correlations these patterns have with connectivity might be explained in terms of highly connected regions having more neurons, and concomitantly fewer oligodendrocytes. However, we could not rigorously test these ideas here because measurements of glia-to-neuron ratios across many brain structures do not appear to be readily available. More detail about the nature of connectivity supported by the pattern NE and OE regions could also provide insight; in particular the connectivity data we used does not detail if the connections are highly myelinated, inhibitory, or excitatory. We also found that the pattern NE genes have a more recent evolutionary origin, while the pattern OE genes tend to be more ancient. This agrees with past work that found evolutionary expansion and regional variation of synaptic genes that are expressed primarily in neurons (Pocklington et al., 2006; Emes et al., 2008).

We note that the connectivity data we employ does not form a complete connectome. The connectivity data we use lacks information about connections that have been shown not to exist. In addition, many brain region pairs have not been studied in a curated tract tracing experiment and may or may not be connected. Of these three cases only one (connected but not known) would increase connectivity degree of a region. Large increases in connectivity degree will affect our results but small changes in connectivity degree are unlikely to change the correlations because we measure Spearman's rank correlation coefficient. However, we expect additional connectivity data for regions with few reported connections will allow deeper analysis. Further, use of the BAMS connectivity data requires pooling of the underlying voxel-based gene expression data into brain regions. This limits our results to less than half of the brain by volume but prevents large regions from dominating the analysis. A larger analysis at the voxel level may result in more robust inverse correlations. However, associations to connection degree could not be performed because voxel level connectivity data is limited for mouse (Moldrich et al., 2010).

Our analysis required the integration of several complex data sets, illustrating several methodological problems that hinder such efforts. Mapping between anatomical atlases presents a significant challenge in linking transcriptomics to connectomics. While genomics has mostly sorted out how to reference specific genes (Gerstein et al., 2007), it is much harder to identify and delineate a specific brain region (Bohland et al., 2009a; Hawrylycz et al., 2011). In *C. elegans* the stable number of neurons allows each one to be given a unique identifier, but in more complex organisms even within a specific atlas it can be hard to map brain regions across atlases. For example, in the BAMS database we found differences between the 1998 atlas and 2004 rat brain atlases (Swanson, 1999, 2004). Although mappings between the two atlases are formalized and accessible, only 60% of the regions have mappings (Swanson and Bota, 2010). CoCoMac, a tract tracing database of Macaque connectivity has spent significant effort to reconcile the atlases and like BAMS it provides information on equal, overlapping, and enclosing brain regions (Stephan et al., 2000; Kotter, 2004; Kotter and Wanke, 2005). Using CoCoMac, Modha and Singh (2010)

were able to merge the 379 parcellation schemes and over 16,000 mapping relations to create the largest wiring diagram for the Macaque brain. These formalized brain maps will play an important role in future multimodal analyses of the nervous system. Overall, limitations in our ability to interpret these results stress the need for highly detailed neuroinformatics databases of many modalities (Akil et al., 2011).

In conclusion, we identified a novel expression pattern in the rodent brain that correlates with patterns of connectivity and measures of cellular composition. Future work will be aimed at further dissecting these and other patterns, including the potential relationships they may have with behavioral mutations in mice or neuropsychiatric disorders in humans.

ACKNOWLEDGMENTS

We are greatly thankful to the providers of the data sets relied upon in this study: the Allen Institute for Brain Research, Cahoy et al. and The Brain Architecture Centre. We thank Kevin She and Jesse Gillis for valuable discussion. We thank the reviewers for their helpful comments and suggestions. This work was supported in part by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (371348). Further support was provided by a National Institutes of Health grant GM076990 to Paul Pavlidis, the Canadian Foundation for Innovation (Leaders Opportunities Fund), the Michael Smith Foundation for Health Research (Career Investigator award to Paul Pavlidis), and the Canadian Institutes of Health Research (New Investigator Salary Award to Paul Pavlidis). Powell Patrick Cheng Tan is supported by the CIHR Strategic Training Program in Bioinformatics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at http://www.frontiersin.org/neuroinformatics/10.3389/fninf.2011.00012/abstract/

Data sheet 1 | Pattern NE identifiers.

Data sheet 2 | Pattern OE identifiers.

Data sheet 3 | Complete pattern OE and NE Gene Ontology enrichment results.

REFERENCES

Akil, H., Martone, M. E., and Van Essen, D. C. (2011). Challenges and opportunities in mining neuroscience data. *Science* 331, 708–712.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–29.

Baruch, L., Itzkovitz, S., Golan-Mashiach, M., Shapiro, E., and Segal, E. (2008). Using expression profiles of *Caenorhabditis elegans* neurons to identify genes that mediate synaptic connectivity. *PLoS Comput. Biol.* 4, e1000120. doi: 10.1371/journal.pcbi.1000120

Biswal, B. B., Mennes, M., Zuo, X. N., Gohel, S., Kelly, C., Smith, S. M., Beckmann, C. F., Adelstein, J. S., Buckner, R. L., Colcombe, S., Dogonowski, A. M., Ernst, M., Fair, D., Hampson, M., Hoptman, M. J., Hyde, J. S., Kiviniemi, V. J., Kotter, R., Li, S. J., Lin, C. P., Lowe, M. J., Mackay, C., Madden, D. J., Madsen, K. H., Margulies, D. S., Mayberg, H. S., Mcmahon, K., Monk, C. S., Mostofsky, S. H., Nagel, B. J., Pekar, J. J., Peltier, S. J., Petersen, S. E., Riedl, V., Rombouts, S. A., Rypma, B., Schlaggar, B. L., Schmidt, S., Seidler, R. D., Siegle, G. J., Sorg, C., Teng, G. J., Veijola, J., Villringer, A., Walter, M., Wang, L., Weng, X. C., Whitfield-Gabrieli, S., Williamson, P., Windischberger, C., Zang, Y. F., Zhang, H. Y., Castellanos, F. X., and Milham, M. P. (2010). Toward discovery science of human brain function. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4734–4739.

Bohland, J. W., Bokil, H., Allen, C. B., and Mitra, P. P. (2009a). The brain atlas concordance problem: quantitative comparison of anatomical parcellations. *PLoS ONE* 4, e7200. doi: 10.1371/journal.pone.0007200

Bohland, J. W., Bokil, H., Pathak, S. D., Lee, C. K., Ng, L., Lau, C., Kuan, C., Hawrylycz, M., and Mitra, P. P. (2009b). Clustering of spatial gene expression patterns in the mouse brain and comparison with classical neuroanatomy. *Methods* 50, 105–112.

Bota, M., Dong, H. W., and Swanson, L. W. (2003). From gene networks to brain networks. *Nat. Neurosci.* 6, 795–799.

Bota, M., Dong, H. W., and Swanson, L. W. (2005). Brain architecture management system. *Neuroinformatics* 3, 15–48.

Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., Xing, Y., Lubischer, J. L., Krieg, P. A., Krupenko, S. A., Thompson, W. J., and Barres, B. A. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* 28, 264–278.

Dong, H. W., Swanson, L. W., Chen, L., Fanselow, M. S., and Toga, A. W. (2009). Genomic-anatomic evidence for distinct functional domains in hippocampal field CA1. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11794–11799.

Emes, R. D., Pocklington, A. J., Anderson, C. N., Bayes, A., Collins, M. O., Vickers, C. A., Croning, M. D., Malik, B. R., Choudhary, J. S., Armstrong, J. D., and Grant, S. G. (2008). Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat. Neurosci.* 11, 799–806.

French, L., and Pavlidis, P. (2011). Relationships between gene expression and brain wiring in the adult rodent brain. *PLoS Comput. Biol.* 7, e1001049. doi: 10.1371/journal.pcbi.1001049

Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., Emanuelsson, O., Zhang, Z. D., Weissman, S., and Snyder, M. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 17, 669–681.

Ghandour, M. S., Langley, O. K., Labourdette, G., Vincendon, G., and Gombos, G. (1981). Specific and artefactual cellular localizations of S 100 protein: an astrocyte marker in rat cerebellum. *Dev. Neurosci.* 4, 66–78.

Ghandour, M. S., Langley, O. K., Vincendon, G., and Gombos, G. (1979). Double labeling immunohistochemical technique provides evidence of the specificity of glial cell markers. *J. Histochem. Cytochem.* 27, 1634–1637.

Ghandour, M. S., Langley, O. K., Vincendon, G., Gombos, G., Filippi, D., Limozin, N., Dalmasso, D., and Laurent, G. (1980). Immunochemical and immunohistochemical study of carbonic anhydrase II in adult rat cerebellum: a marker for oligodendrocytes. *Neuroscience* 5, 559–571.

Gillis, J., Mistry, M., and Pavlidis, P. (2010). Gene function analysis in complex data sets using ErmineJ. *Nat. Protoc.* 5, 1148–1159.

Harris, T. W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W. J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., Fernandes, J., Han, M., Kishore, R., Lee, R., Muller, H. M., Nakamura, C., Ozersky, P., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E. M., Tuli, M. A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L. D., Spieth, J., and Sternberg, P. W. (2010). WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.* 38, D463–D467.

Hawrylycz, M., Baldock, R. A., Burger, A., Hashikawa, T., Johnson, G. A., Martone, M., Ng, L., Lau, C., Larson, S. D., Nissanov, J., Puelles, L., Ruffins, S., Verbeek, F., Zaslavsky, I., and Boline, J. (2011). Digital atlasing and standardization in the mouse brain. *PLoS Comput. Biol.* 7, e1001065. doi: 10.1371/journal.pcbi.1001065

Kaufman, A., Dror, G., Meilijson, I., and Ruppin, E. (2006). Gene expression of *Caenorhabditis elegans* neurons carries information on their synaptic connectivity. *PLoS Comput. Biol.* 2, e167. doi: 10.1371/journal.pcbi.0020167

Kotter, R. (2004). Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* 2, 127–144.

Kotter, R., and Wanke, E. (2005). Mapping brains without coordinates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 751–766.

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T. M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., Dee, N. R., Desaki, A. L., Desta, T., Diep, E., Dolbeare, T. A., Donelan, M. J.,

Dong, H. W., Dougherty, J. G., Duncan, B. J., Ebbert, A. J., Eichele, G., Estin, L. K., Faber, C., Facer, B. A., Fields, R., Fischer, S. R., Fliss, T. P., Frensley, C., Gates, S. N., Glattfelder, K. J., Halverson, K. R., Hart, M. R., Hohmann, J. G., Howell, M. P., Jeung, D. P., Johnson, R. A., Karr, P. T., Kawal, R., Kidney, J. M., Knapik, R. H., Kuan, C. L., Lake, J. H., Laramee, A. R., Larsen, K. D., Lau, C., Lemon, T. A., Liang, A. J., Liu, Y., Luong, L. T., Michaels, J., Morgan, J. J., Morgan, R. J., Mortrud, M. T., Mosqueda, N. F., Ng, L. L., Ng, R., Orta, G. J., Overly, C. C., Pak, T. H., Parry, S. E., Pathak, S. D., Pearson, O. C., Puchalski, R. B., Riley, Z. L., Rockett, H. R., Rowland, S. A., Royall, J. J., Ruiz, M. J., Sarno, N. R., Schaffnit, K., Shapovalova, N. V., Sivisay, T., Slaughterbeck, C. R., Smith, S. C., Smith, K. A., Smith, B. I., Sodt, A. J., Stewart, N. N., Stumpf, K. R., Sunkin, S. M., Sutram, M., Tam, A., Teemer, C. D., Thaller, C., Thompson, C. L., Varnam, L. R., Visel, A., Whitlock, R. M., Wohnoutka, P. E., Wolkey, C. K., Wong, V. Y., Wood, M., Yaylaoglu, M. B., Young, R. C., Youngstrom, B. L., Yuan, X. F., Zhang, B., Zwingman, T. A., and Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176.

Letournel, F., Bocquet, A., Perrot, R., Dechaume, A., Guinut, F., Eyer, J., and Barthelaix, A. (2006). Neurofilament high molecular weight-green fluorescent protein fusion is normally expressed in neurons and transported in axons: a neuronal marker to investigate the biology of neurofilaments. *Neuroscience* 137, 103–111.

Lichtman, J. W., and Sanes, J. R. (2008). Ome sweet ome: what can the genome tell us about the connectome? *Curr. Opin. Neurobiol.* 18, 346–353.

Modha, D. S., and Singh, R. (2010). Network architecture of the long-distance pathways in the macaque brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13485–13490.

Moldrich, R. X., Pannek, K., Hoch, R., Rubenstein, J. L., Kurniawan, N. D., and Richards, L. J. (2010). Comparative mouse brain tractography of diffusion magnetic resonance imaging. *Neuroimage* 51, 1027–1036.

Ng, L., Bernard, A., Lau, C., Overly, C. C., Dong, H. W., Kuan, C., Pathak, S., Sunkin, S. M., Dang, C., Bohland, J. W., Bokil, H., Mitra, P. P., Puelles, L., Hohmann, J., Anderson, D. J., Lein, E. S., Jones, A. R., and Hawrylycz, M. (2009). An anatomic gene expression atlas of the adult mouse brain. *Nat. Neurosci.* 12, 356–362.

Ouimet, C. C., Mcguinness, T. L., and Greengard, P. (1984). Immunocytochemical localization of calcium/calmodulin-dependent protein kinase II in rat brain. *Proc. Natl. Acad. Sci. U.S.A.* 81, 5604–5608.

Pfeiffer, B., Norman, A. W., and Hamprecht, B. (1989). Immunocytochemical characterization of neuron-rich rat brain primary cultures: calbindin D28K as marker of a neuronal subpopulation. *Brain Res.* 476, 120–128.

Pocklington, A. J., Cumiskey, M., Armstrong, J. D., and Grant, S. G. (2006). The proteomes of neurotransmitter receptor complexes form modular networks with distributed functionality underlying plasticity and behaviour. *Mol. Syst. Biol.* 2, 2006 0023.

Polleux, F., Ince-Dunn, G., and Ghosh, A. (2007). Transcriptional regulation of vertebrate axon guidance and synapse formation. *Nat. Rev. Neurosci.* 8, 331–340.

Ressler, K. J., Paschall, G., Zhou, X. L., and Davis, M. (2002). Regulation of synaptic plasticity genes during consolidation of fear conditioning. *J. Neurosci.* 22, 7892–7902.

Rosengren, L. E., Kjellstrand, P., Aurell, A., and Haglid, K. G. (1986). Irreversible effects of dichloromethane on the brain after long term exposure: a quantitative study of DNA and the glial cell marker proteins S-100 and GFA. *Br. J. Ind. Med.* 43, 291–299.

Sporns, O. (2011). The human connectome: a complex network. *Ann. N. Y. Acad. Sci.* 1224, 109–125.

Sporns, O., Tononi, G., and Kotter, R. (2005). The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* 1, e42. doi: 10.1371/journal.pcbi.0010042

Stephan, K. E., Zilles, K., and Kotter, R. (2000). Coordinate-independent mapping of structural and functional data by objective relational transformation (ORT). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 37–54.

Swanson, L. W. (1999). *Brain Maps: Structure of the Rat Brain*. Amsterdam: Elsevier.

Swanson, L. W. (2004). *Brain Maps: Structure of the Rat Brain*, 3rd Edn. Oxford: Academic Press.

Swanson, L. W., and Bota, M. (2010). Foundational model of structural connectivity in the nervous system with a schema for wiring diagrams, connectome, and basic plan architecture. *Proc. Natl. Acad. Sci. U.S.A.* 107, 20610–20617.

Varadan, V., Miller, D. M. III, and Anastassiou, D. (2006). Computational inference of the molecular logic for synaptic connectivity in *C. elegans*. *Bioinformatics* 22, e497–e506.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Chetvernin, V., Church, D. M., Dicuccio, M., Edgar, R., Federhen, S., Geer, L. Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D. J., Madden, T. L., Maglott, D. R., Ostell, J., Miller, V., Pruitt, K. D., Schuler, G. D., Sequeira, E., Sherry, S. T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R. L., Tatusova, T. A., Wagner, L., and Yaschenko, E. (2007). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 35, D5–D12.

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 314, 1–340.

Wolf, L., Goldberg, C., Manor, N., Sharan, R., and Ruppin, E. (2011). Gene expression in the rodent brain is associated with its regional connectivity. *PLoS Comput. Biol.* 7, e1002040. doi: 10.1371/journal.pcbi.1002040

Wu, Y., Zhang, A. Q., and Yew, D. T. (2005). Age related changes of various markers of astrocytes in senescence-accelerated mice hippocampus. *Neurochem. Int.* 46, 565–574.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 25 March 2011; accepted: 18 July 2011; published online: 29 July 2011.
*Citation: French L, Tan PPC and Pavlidis P (2011) Large-scale analysis of gene expression and connectivity in the rodent brain: insights through data integration. Front. Neuroinform. **5**:12. doi: 10.3389/fninf.2011.00012*

APPENDIX

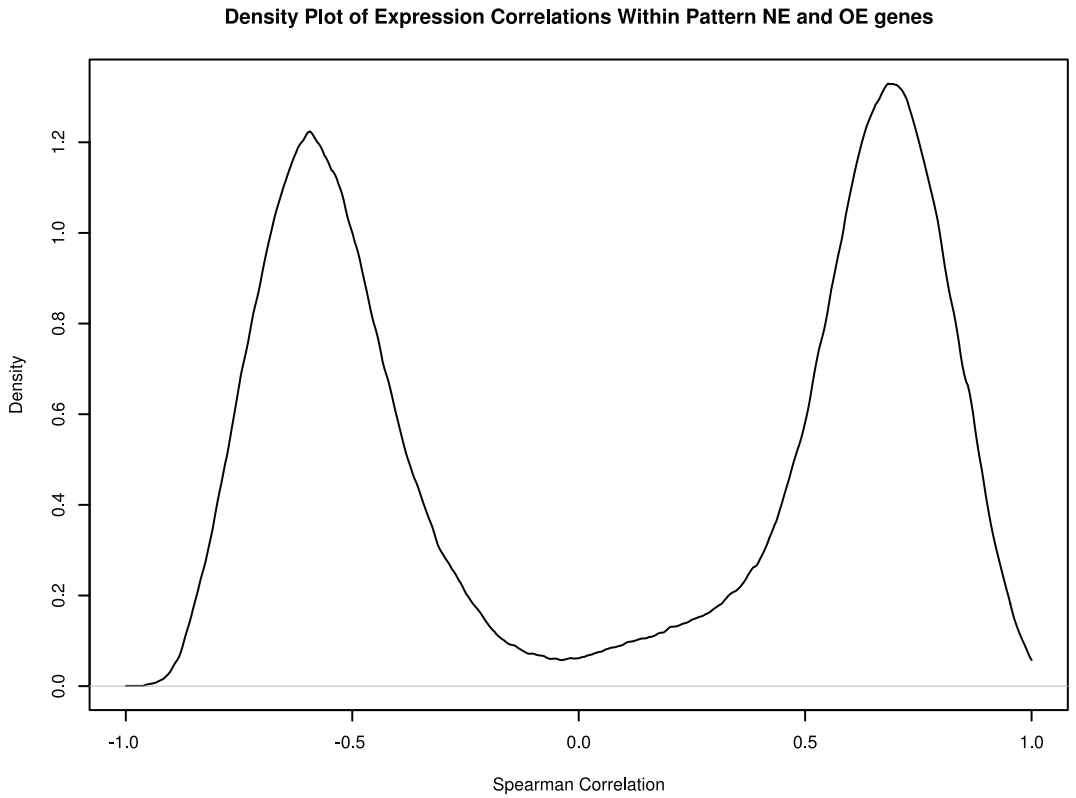


FIGURE A1 | Density plot of expression correlations within pattern NE and OE genes.

frontiers in
NEUROINFORMATICS

PERSPECTIVE ARTICLE
published: 27 December 2012
doi: 10.3389/fninf.2012.00030

CoCoMac 2.0 and the future of tract-tracing databases

Rembrandt Bakker[1,2,3]*, Thomas Wachtler[3] and Markus Diesmann[2,4,5]

[1] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands
[2] Institute of Neuroscience and Medicine 6, Research Center Jülich, Jülich, Germany
[3] Department Biology II, Ludwig-Maximilians-Universität München, Munich, Germany
[4] Faculty of Medicine, RWTH Aachen University, Aachen, Germany
[5] RIKEN Brain Science Institute, Wako, Saitama, Japan

Edited by:
Trygve B. Leergaard, University of Oslo, Norway

Reviewed by:
Trygve B. Leergaard, University of Oslo, Norway
Claus Hilgetag, Jacobs University Bremen, Germany
Olaf Sporns, Indiana University, USA

***Correspondence:**
Rembrandt Bakker, Institute of Neuroscience and Medicine 6, Research Center Jülich, 52425 Jülich, Germany.
e-mail: r.bakker@donders.ru.nl

The CoCoMac database contains the results of several hundred published axonal tract-tracing studies in the macaque monkey brain. The combined results are used for constructing the macaque macro-connectome. Here we discuss the redevelopment of CoCoMac and compare it to six connectome-related projects: two online resources that provide full access to raw tracing data in rodents, a connectome viewer for advanced 3D graphics, a partial but highly detailed rat connectome, a brain data management system that generates custom connectivity matrices, and a software package that covers the complete pipeline from connectivity data to large-scale brain simulations. The second edition of CoCoMac features many enhancements over the original. For example, a search wizard is provided for full access to all tables and their nested dependencies. Connectivity matrices can be computed on demand in a user-selected nomenclature. A new data entry system is available as a preview, and is to become a generic solution for community-driven data entry in manually collated databases. We conclude with the question whether neuronal tracing will remain the gold standard to uncover the wiring of brains, thereby highlighting developments in human connectome construction, tracer substances, polarized light imaging, and serial block-face scanning electron microscopy.

Keywords: CoCoMac, macaque, connectivity, database, axonal tracing

INTRODUCTION

The Frontiers in Neuroinformatics Special Topic "Mapping the Connectome" is dedicated to the memory of Rolf Kötter, founding father of the Macaque connectivity database CoCoMac (Stephan et al., 2001; Kötter, 2004). This database contains the results of about 300 published axonal tract-tracing studies, and another 150 studies on brain atlases and nomenclature. In the last 3 years of his life, while continuously stepping up the fight against the disease that threatened his life, Rolf Kötter energetically led an effort to improve the informatics and databasing aspects of the CoCoMac database. After his tragic death in 2010, we have stepped forward to keep this ongoing project alive. We decided to keep the CoCoMac website cocomac.org in a "frozen" but functional state and refer to it as CoCoMac 1.0. The newly developed CoCoMac 2.0 is hosted at cocomac.g-node.org.

In this article we present the status of CoCoMac 2.0, and compare its neuroinformatics aspects to six connectome-related projects that have recently been introduced or updated; the names in ***bold italics*** are used to reference the project throughout this paper:

1. The rat ***temporal-lobe project*** of Sugar et al. (2011). This project covers a limited part of the brain, but maximizes the level of detail that can be obtained from published tracing studies.
2. The Brain Architecture Management System (***BAMS***) of Bota et al. (2012), a well-established system for storage, retrieval, and searching of connectivity data at all levels of detail, largely populated with rat tract-tracing data, but also connected to a new ***mouse connectome***[1] initiative.
3. The ***Connectome Viewer*** Toolkit (Gerhard et al., 2011), which provides advanced surface, volume, and graph rendering tools in a Python package.
4. The rat connectivity component of the ***rodent brain workbench*** (Zakiewicz et al., 2011), providing public access to raw tract-tracing images, registered to a common atlas space.
5. The ***Allen mouse connectivity atlas***[2], with open access to raw data of viral, cell type-specific tracers.
6. The ***neuroVIISAS*** rat connectome system of Schmitt and Eipert (2012), a software package that covers the entire pipeline from connectivity database and atlasing to brain simulations

The organization of this paper is as follows: we briefly present the history of CoCoMac and discuss its unique features; we then present components of the new CoCoMac and a longer term perspective. Along the way, we compare the various informatics aspects to the six connectome-related projects.

CoCoMac 1.0

CoCoMac 1.0 started in the late 1990s as an MS Access (Microsoft Corporation, WA, USA) database, to collect data from tract-tracing studies in the macaque brain. Although the research questions in these studies vary widely, they have in common

[1] mouseconnectome.org
[2] http://connectivity.brain-map.org/

that injections are done in the brain, and the location and density of labeled axon terminals (anterograde tracer) or cell bodies (retrograde tracer) are observed. The biggest challenge is to accurately capture *where* in the brain the injections and labeled sites are. Macaque tracing studies almost invariably use a *brain map* that parcellates the brain into named regions, based on cytoarchitectonic criteria (Kötter and Wanke, 2005). The tracing result is summarized in textual statements such as "the injection was in region A, with spillover to region B" and "dense labeling was observed in region C." These textual statements degrade the tracing result, with its cellular level detail, to a macroscopic level description.

While populating the database, it was quickly discovered that in many cases, a tracing study adapts an existing brain map into a modified version. To combine tracing statements that use competing and evolving brain maps, the spatial relations between newly defined brain regions and older definitions must be known. In CoCoMac these relations are called *inter-map-relations*. They are usually provided by the paper that describes the modified brain map.

From the injections, labeled-sites, and inter-map-relations we can *compute* a macro-connectome: the matrix that describes the connection strength from each region in a given brain map to all regions in that same map. The computations involve the propagation of tracing data from its original brain map to the user-selected brain map. The objective relational transformation (ORT) algorithm by Stephan et al. (2000) pioneered this procedure. ORT uses one additional piece of knowledge: to what *extent* is the brain region covered by labeled sites (complete, partial, or none). This allows ORT to make inferences such as: "if region A has complete coverage of labeled sites, and B is a subregion of A, then also B has complete coverage of labeled sites."

CoCoMac 1.0 contains several visionary aspects that have helped it to become the largest resource of its kind. First it does not store any type of data that requires expert interpretation. Second, it does contain an abundance of *precision of description code* (PDC) statements that answer questions such as "How well does a paper describe the location of injection X?" with possible answers "only in text," "only in a figure," and so on. Whenever a case of conflicting data occurs, PDC statements can be used to give some data more importance than others.

DATA ENTRY, SEARCH, AND DISPLAY

Data entry in CoCoMac 1.0 is done in MS Access. This approach is not suitable for online data entry. The entire ORT algorithm is implemented in Visual Basic (VB) script, with pieces in Java. Performance was not an issue in the early days of CoCoMac, but now that it contains over 8000 brain sites, the ORT-based computation of a connectivity matrix takes months on a PC, due to highly inefficient code.

Visual Basic scripts also form the basis of the CoCoMac.org website. One can search for literature, inter-map-relations, and connectivity using a variety of preset criteria, and retrieve results in the form of expandable HTML tables, or XML that adheres to a published XML schema: http://cocomac.org/www/cocomac.xsd. The XML service has been used to embed connectivity data into other neuroinformatics resources including the Neuroscience Information Framework (Gardner et al., 2008), SumsDB (Dickson et al., 2001), and BrainInfo.org (Bowden et al., 2012).

The latest addition to CoCoMac 1.0 is the CoCoMac-Paxinos-3D viewer (Bezgin et al., 2009). This Java-based tool shows CoCoMac connectivity in a 3D rendering of a Macaque brain, parcellated according to the Paxinos et al. (2000) atlas.

LESSONS LEARNED

CoCoMac 1.0 has grown to become a well-established resource, the largest of its kind, cited in 140 research articles (Google Scholar citations of Stephan et al., 2001). However, its further expansion and usage are severely limited: search functionality is restricted to three use cases; ORT calculations by the current implementation are prohibitively slow; data-entry can only be done internally.

These considerations led to the decision in 2007 that CoCoMac should be rebuilt upon an open source database engine (MySQL[3]) and scripting language (PHP[4], along with an efficient ORT implementation, a web-based data entry system, and an online graphical display of the results.

CoCoMac 2.0 vs. OTHER CONNECTOME PROJECTS

Here we discuss the new features of the CoCoMac 2.0 system and how they compare to the six connectome-related projects mentioned in the introduction.

DATA ENTRY

Ideally, the data entry system not only changes the database contents, but also tracks who made the change, when, and why. We have developed a data entry system that presents each publication in CoCoMac as a hierarchical tree where nodes can be added and edited (Bakker et al., 2008). Every submitted form is stored in a table which contains all information to recreate the CoCoMac database from time zero to any desired instance of time. A view-only version of this system is available at http://cocomac.g-node.org/dataentry2010. A user management strategy with clearly defined user roles and permissions is under development.

None of the other connectome projects provides online, community-driven data entry as of today. Contributions to *BAMS* and *neuroVIISAS* must be submitted by spreadsheet. The temporal-lobe project uses MS Access internally and does not expose the database to the web. The Allen *mouse connectivity atlas* is populated by automated pipelines. The *Connectome Viewer* provides a standard file format to disseminate connectivity data. This format is a good candidate for data sharing among connectome projects. It stores the connectome as a graph, where each node typically represents a brain region, and can be assigned user-defined attributes.

It is surprising that a generic process as "web-based data entry with full provenance" is not readily available as an open source project. Besides the CoCoMac initiative, several approaches are under way in the context of lab automation, e.g., Grewe et al. (2011).

[3]http://www.mysql.com

[4]http://www.php.net

DATA MINING: THE CoCoMac SEARCH WIZARD

With data mining we here refer to the extraction of relevant data from a single connectome database. A different branch of data mining is text mining, which could potentially replace the human data collator. However, given the key role that figures play in published tracing studies, we only see this work if authors of tracing studies provide results in a structured markup language.

CoCoMac 2.0 exposes its underlying database to the public, and aims to be fully open access. At the most basic level there is an SQL query interface: http://cocomac.g-node.org/sql_query. More user oriented is the *interactive query wizard* that generates complex SQL queries based on a series of intuitive search criteria. **Figure 1** shows a screenshot of the wizard[5]. The wizard is not specific to CoCoMac, but can be applied to *any* relational database. It automatically discovers its structure and foreign key relations (Data Sheet 1 in Supplementary Material).

In contrast to CoCoMac 1.0, the new search wizard does not define any use cases, but rather makes *every* table searchable. It provides filters for all properties of the table and its nested relatives. On the output side, the wizard returns data in a JSON format, which is converted to an interactive web page by client-side JavaScript. Among the other connectome projects, *BAMS* is the one with the most extensive search system. The various modules of the system contain custom search components and tabular output. As in CoCoMac, connectome matrices can be generated dynamically and provide links to the underlying literature sources. The *temporal-lobe* connectome only allows searching for connections.

CONNECTOME GENERATION

Connectome generation lies at the heart of a connectome project. Methods that retrieve a full connectome from a single brain clearly have an advantage: they do not need to merge data from different brains, parcellations, and nomenclatures. Three such methods are discussed in the concluding section.

The *rodent work bench*, *Allen mouse connectivity atlas*, and *mouseconnectome.org* all provide images of the labeled sites. They allow browsing of injections by brain site. The *Allen mouse connectivity atlas* also exposes estimated densities of labeled sites. Textual statements derived from the *mouseconnectome.org* data will be available through *BAMS*.

The *temporal-lobe project* has chosen a standard brain map, and mapped all relevant tracing data onto it while entering the data into the system, thus relying on expert knowledge of the data collator.

Both *BAMS* and *CoCoMac* rely on some variant of ORT to bring injections and labeled sites into a common space. *BAMS*, in addition, has all regions mapped onto the Swanson rat atlas (Swanson, 1992) during the data entry phase. With the reimplementation and simplification of ORT in CoCoMac 2.0, computing a connectivity matrix has become a matter of seconds, and is available online at http://cocomac.g-node.org/axonal_projections. The implementation details are beyond the scope of this article, but we point out one important consideration. Translating information from one brain map to another relies on inter-map-relations, such as "A is identical to B" and "B is a subregion of C." Such relations can also be nested: a subregion of a subregion is still a subregion, etc. The new ORT procedure starts with computing a matrix of all possible relations (identical, subregion, superregion, overlap) between all brain sites. In theory, this should result in *at most one* relation between any region pair. But for about 5% of the relations, conflicting versions exist. Wrong relations can have a detrimental effect on the connectivity matrix. CoCoMac 1.0 rather arbitrarily gave preference to identity relations. Modha and Singh (2010) produced a consistent spatial relations matrix by letting "is

[5]http://cocomac.g-node.org/search_wizard

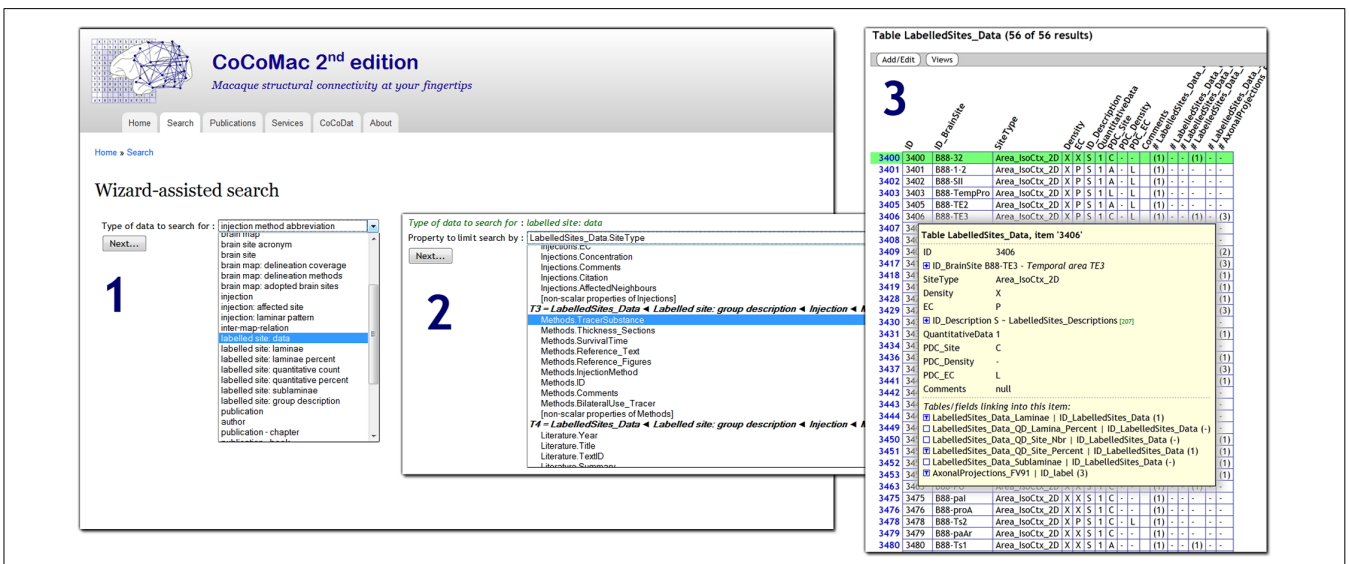


FIGURE 1 | Screenshot of the CoCoMac 2.0 search wizard (http://cocomac.g-node.org/search_wizard), with three steps highlighted: (1) selection of the data type (any database table); (2) selection of search criterion (field from table that is recursively related to data type); (3) interactive result table, expanding the tree shows all related data.

subregion of" prevail over "is identical to" statements. Both solutions neglect the fact that conflicts point to errors in the database or its literature sources. For CoCoMac 2.0 we have generated a list of core conflicts (Data Sheet 2 in Supplementary Material), which we are currently using to improve consistency across brain maps.

COMPLETENESS AND LEVEL OF DETAIL

The *temporal-lobe* project defines a complete connectome as one that "contains all available knowledge," and gets close to completeness for a selected set of brain regions. The 300 tracing studies contained in CoCoMac cover a substantial part of the available literature.

Another useful definition of completeness is: a connectome that covers the entire brain. A lower bound on completeness is that every brain region gets injected once, but given the variety of tracer characteristics, multiple injections are needed. The rodent brain workbench provides only six tracer injections, while the Allen mouse connectivity atlas systematically covers the entire brain. We have recently shown (Bezgin et al., 2012) that CoCoMac has full coverage, with between 2 and 80 injections (average 20) for the cortical regions of the Paxinos et al. (2000) atlas. If we require additional detail, such as laminar origin or termination patterns, the available data is reduced by a factor of four and covers most but not all regions.

The *temporal-lobe* project is a showcase of what level of detail can be obtained in a tract-tracing-based connectome. Connection strengths are allowed to vary within a brain region, by dividing it in nine subparcels. Within the subparcels, layer-specific termination patterns are indicated. Complete layer specificity cannot be obtained: anterograde tracers reveal layers of termination; retrograde tracers reveal layers of origin. But undetermined is whether there is full connectivity between these layers of origin and termination.

CoCoMac, *BAMS*, and most other projects distinguish between weak, moderate, and strong levels of labeled site density, but rarely do labeled sites get counted. Recent studies of Markov et al. (2011); Mikula et al. (2012) in which labeled cell bodies were counted showed variations by at least five orders of magnitude. This is essential data for computer simulations of brain tissue.

A limitation of all databases that rely on published tracing studies is that the textual descriptions of injection and labeled sites degrade the micro level accuracy of the tracing data to the macro level. A strategy to overcome this is to discard the legacy data and start a new, large-scale tracing effort that uses a standard atlas, records the stereotaxic location of injections and labeled sites, and registers all the individual brains to a common space. The *rodent brain workbench*, *mouseconnectome.org* and the *Allen mouse connectivity atlas* are such efforts. They expose the raw data in an online repository, to which automated approaches for counting labeled sites can be applied. Equivalent resources for macaque data are not available for two reasons: (1) ethical concerns restrict new experiments that duplicate legacy data,.and (2) the spatial location of brain regions has a much larger inter-subject variability, so that knowing the stereotaxic coordinate of an injection is not always more accurate than knowing its brain region.

GRAPHICAL DISPLAY OF CONNECTIVITY

Display of connectivity is essential for disseminating results to non-experts. We are feeding CoCoMac connectivity results to the web-based atlas display engine at http://scalablebrainatlas.incf.org (SBA, Bakker et al., 2011, where users can query CoCoMac in a single mouse click. In turn, SBA redirects the user back to the CoCoMac website for a full traceback of the displayed connectivity. **Figure 2** demonstrates this procedure.

The *Connectome Toolkit* is another generic system for graphical display; this standalone, python-based package allows the creation

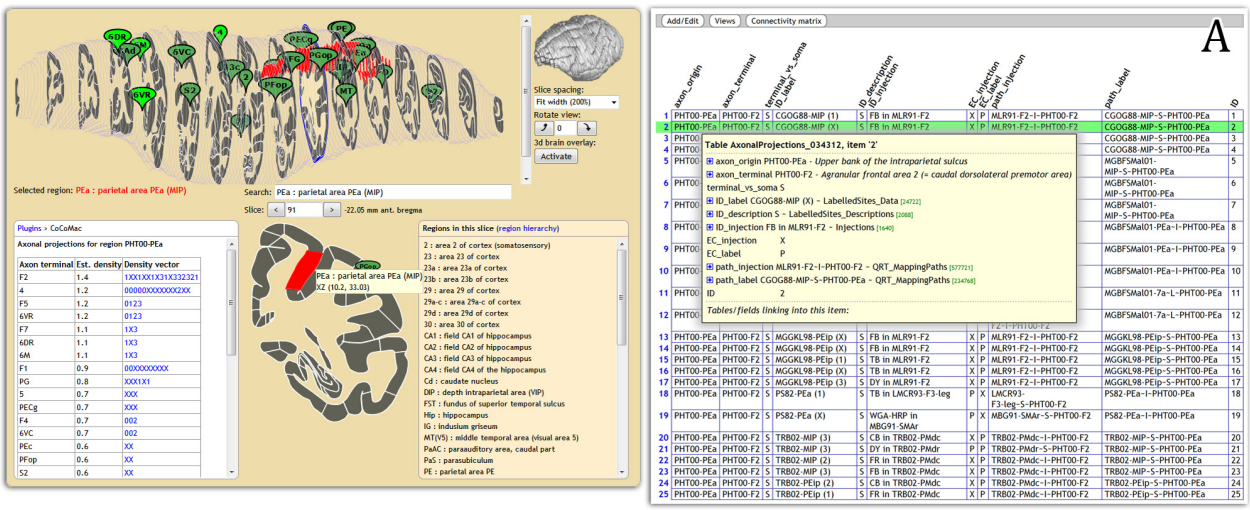


FIGURE 2 | Connectivity display using the Scalable Brain Atlas display engine (scalablebrainatlas.incf.org/cocomac), here displaying axonal projections that originate in parietal area PEa.

The insert **(A)** shows the result when clicking on a link in the connectivity table, each row can be expanded all the way down to the literature sources.

of advanced scenes in which connectivity graphs and volumetric or surface based brain data can be blended, colored, and sliced.

The *temporal-lobe* project provides a poster-size interactive PDF document in which connections can be displayed by clicking on a region of interest. The document is well structured; but what is missing is a traceback to the original publications.

NeuroVIISAS finally has built-in rendering of connections on a 3D brain atlas. In addition, spiking activity generated by the built-in brain simulator makes the brain regions flash.

FUTURE PERSPECTIVE

The future of tract-tracing connectomes hinges on three key questions:

1. Will tracing studies be obliterated by emerging new technologies?
2. How do we best deal with the existing, published data?
3. How do we best set up new experiments?

Among emerging technologies that compete with tract-tracing are diffusion based MRI, polarized light imaging (Axer et al., 2011), and serial block-face scanning electron microscopy (SBFSEM). They share as a big advantage that a complete connectome can be derived from a single brain, thus eliminating the loss of resolution that occurs when data from different brains is combined. Diffusion MRI plays a key role in the Human Connectome Project (Marcus et al., 2011), where invasive techniques are not an option. This project is unique in that its primary aim is not to publish a connectome, but rather make the best possible data and analysis tools publicly available that will allow others to do so. Also applicable to humans is the PLI method, which uses polarized light to measure the orientation of myelinated fibers in thin postmortem brain slices. This is done at microscopic resolution in the in-plane dimensions. The challenge lies in the across-slice dimension, where slices have to be realigned after cutting and non-uniform shrinking. This same problem initially hampered electron microscopy approaches, but with the introduction of serial block-face scanning it became possible to obtain resolutions below 50 nm in *all* dimensions (Denk and Horstmann, 2004). Such resolutions may become available for an entire mouse brain (Mikula et al., 2012) in about 10 years, and promise to reveal full axonal and dendritic trees – provided that automated techniques become available that digest petabytes of data without human intervention.

For the coming decades, tract-tracing will remain the gold standard (Sporns, 2010), because it provides (1) unequivocal proof for the existence of long-range connections; (2) directionality; (3) layer specificity, and (4) sufficient detail to estimate quantitative connection strengths. In addition, recently developed tracer substances can reveal connections that are (5) cell type-specific and (6) part of local, multisynaptic circuits (Vercelli et al., 2000).

The question on how to best deal with existing, published tracing data has been largely addressed in this manuscript. One further aspect is to provide large-scale brain simulators with realistic parameter estimates for layer- and cell type-specific connectivity. This requires an integrative approach that combines tract-tracing studies with cell reconstructions and multi cell patch clamp studies (Potjans and Diesmann, 2011).

For the final question on how to design new tract-tracing studies, the *rodent brain workbench*, *mouse connectome*, and *Allen mouse connectivity atlas* pave the way: register sliced individual brains to a standard brain, and provide public access to the raw tracing images. The approach has four remaining challenges: (1) providing access to large data sets is not financially rewarding; (2) brain parcellations needed to convert imaging data to textual statements are often copyrighted; (3) automated extraction of connectivity data is a daunting task, highly dependent on the *meta data*; and (4) especially for primates, ethical guidelines require that legacy data are *not discarded*. The International Neuroinformatics Coordinating Facility (INCF[6]) aims to relief these challenges by promoting standards, open access, and developing a global federated data space. The present connectome projects each have unique properties, which are best combined by supporting a common interface and data format. That would make it possible to cross-check connectivity data from multiple databases, in the same way as CoCoMac already uses redundant inter-map-relations to improve consistency across brain parcellations.

ACKNOWLEDGMENTS

The authors thank Dick Heeren for his contributions to the CoCoMac 2.0 code base. Supported by the German INCF Node (BMBF grant 01GQ0801), the Helmholtz Alliance on Systems Biology, JUGENE Grant JINB33, the Next-Generation Supercomputer Project of MEXT, Japan, and EU Grant 269921 (BrainScaleS). The Scalable Brain Atlas is conducted within the INCF Program on Ontologies of Neural Structures.

SUPPLEMENTARY MATERIAL

Supplementary material for this artical can be found online at http://www.frontiersin.org/Neuroinformatics/10.3389/fninf.2012.00030/abstract

[6]www.incf.org

REFERENCES

Axer, M., Grässel, D., Kleiner, M., Dammers, J., Dickscheid, T., Reckfort, J., et al. (2011). High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging. *Front. Neuroinform.* 5:34. doi: 10.3389/fninf.2011.00034

Bakker, R., Cannon, R. C., Howell, F. W., and Kötter, R. (2008). The new CoCoMac database: a community project. *Front. Neuroinform.* Conference Abstract: *Neuroinformatics 2008.* doi: 10.3389/conf.neuro.11.2008.01.139

Bakker, R., Diesmann, M., and Wachtler, T. (2011). Plugin architecture for the INCF Scalable Brain Atlas: CoCoMac and beyond. *Front. Neuroinform.* Conference Abstract: 4th INCF Congress of Neuroinformatics. doi: 10.3389/conf.fninf.2011.08.00066.

Bezgin, G., Reid, A. T., Schubert, D., and Kötter, R. (2009). Matching spatial with ontological brain regions using Java tools for visualization, database access, and integrated data analysis. *Neuroinformatics* 7, 7–22.

Bezgin, G., Vakorin, V. A., van Opstal, A. J., McIntosh, A. R., and Bakker, R. (2012). Hundreds of brain maps in one atlas: registering coordinate-independent primate neuro-anatomical data to a standard brain. *Neuroimage* 62, 67–76.

Bota, M., Dong, H.-W., and Swanson, L. W. (2012). Combining collation and annotation efforts toward completion of the rat and mouse connectomes in BAMS. *Front. Neuroinform.* 6:2. doi: 10.3389/fninf.2012.00002

Bowden, D. M., Song, E., Kosheleva, J., and Dubach, M. F. (2012). NeuroNames: an ontology for the BrainInfo

portal to neuroscience on the web. *Neuroinformatics* 10, 97–114.

Denk, W., and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* 2:e329. doi: 10.1371/journal.pbio.0020329

Dickson, J., Drury, H., and Van Essen, D. C. (2001). The surface management system (SuMS) database: a surface-based database to aid cortical surface reconstruction, visualization and analysis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1277–1292.

Gardner, D., Akil, H., Ascoli, G. A., Bowden, D. M., Bug, W., Donohue, D. E., et al. (2008). The neuroscience information framework: a data and knowledge environment for neuroscience. *Neuroinformatics* 6, 149–160.

Gerhard, S., Daducci, A., Lemkaddem, A., Meuli, R., Thiran, J.-P., and Hagmann, P. (2011). The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Front. Neuroinform.* 5:3. doi: 10.3389/fninf.2011.00003

Grewe, J., Wachtler, T., and Benda, J. (2011). A bottom-up approach to data annotation in neurophysiology. *Front. Neuroinform.* 5:16. doi: 10.3389/fninf.2011.00016

Kötter, R. (2004). Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics* 2, 127–144.

Kötter, R., and Wanke, E. (2005). Mapping brains without coordinates. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 360, 751–766.

Marcus, D. S., Harwell, J., Olsen, T., Hodge, M., Glasser, M. F., Prior, F., et al. (2011). Informatics and data mining tools and strategies for the human connectome project. *Front. Neuroinform.* 5:4. doi: 10.3389/fninf.2011.00004

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., et al. (2012). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex.* doi: 10.1093/cercor/bhs270 [Epub ahead of print].

Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., et al. (2011). Weight consistency specifies regularities of macaque cortical networks. *Cereb. Cortex* 21, 1254–1272.

Mikula, S., Binding, J., and Denk, W. (2012). Staining and embedding the whole mouse brain for electron microscopy. *Nat. Methods* 9, 1198–1201.

Modha, D. S., and Singh, R. (2010). Network architecture of the long-distance pathways in the macaque brain. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13485–13490.

Paxinos, G., Huang, X.-F., and Petrides, M. (2000). *The Rhesus Monkey Brain in Stereotaxic Coordinates.* New York: Academic Press.

Potjans, T. C., and Diesmann, M. (2011). *The Cell-type Specific Connectivity of the Local Cortical Network Explains Prominent Features of Neuronal Activity.* Available at: http://arxiv.org/abs/1106.5678

Schmitt, O., and Eipert, P. (2012). neuroVIISAS: approaching multiscale simulation of the rat connectome. *Neuroinformatics.* doi: 10.1007/s12021-012-9141-6

Sporns, O. (2010). *Networks of the Brain.* Cambridge, MA: MIT Press.

Stephan, K. E., Kamper, L., Bozkurt, A., Burns, G. A., Young, M. P., and Kötter, R. (2001). Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1159–1186.

Stephan, K. E., Zilles, K., and Kötter, R. (2000). Coordinate-independent mapping of structural and functional data by objective relational transformation (ORT). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 37–54.

Sugar, J., Witter, M. P., van Strien, N. M., and Cappaert, N. L. M. (2011). The retrosplenial cortex: intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome. *Front. Neuroinform.* 5:7. doi: 10.3389/fninf.2011.00007

Swanson, L. W. (1992). *Brain Maps: Structure of the Rat Brain.* Amsterdam: Elsevier.

Vercelli, A., Repici, M., Garbossa, D., and Grimaldi, A. (2000). Recent techniques for tracing pathways in the central nervous system of developing and adult mammals. *Brain Res. Bull.* 51, 11–28.

Zakiewicz, I. M., van Dongen, Y. C., Leergaard, T. B., and Bjaalie, J. G. (2011). Workflow and atlas system for brain-wide mapping of axonal connectivity in rat. *PLoS ONE* 6:e22669. doi: 10.1371/journal.pone.0022669

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 April 2012; accepted: 07 December 2012; published online: 27 December 2012.

Citation: Bakker R, Wachtler T and Diesmann M (2012) CoCoMac 2.0 and the future of tract-tracing databases. Front. Neuroinform. ***6****:30. doi: 10.3389/fninf.2012.00030*