COLLABORATIVE EFFORTS FOR UNDERSTANDING THE HUMAN BRAIN
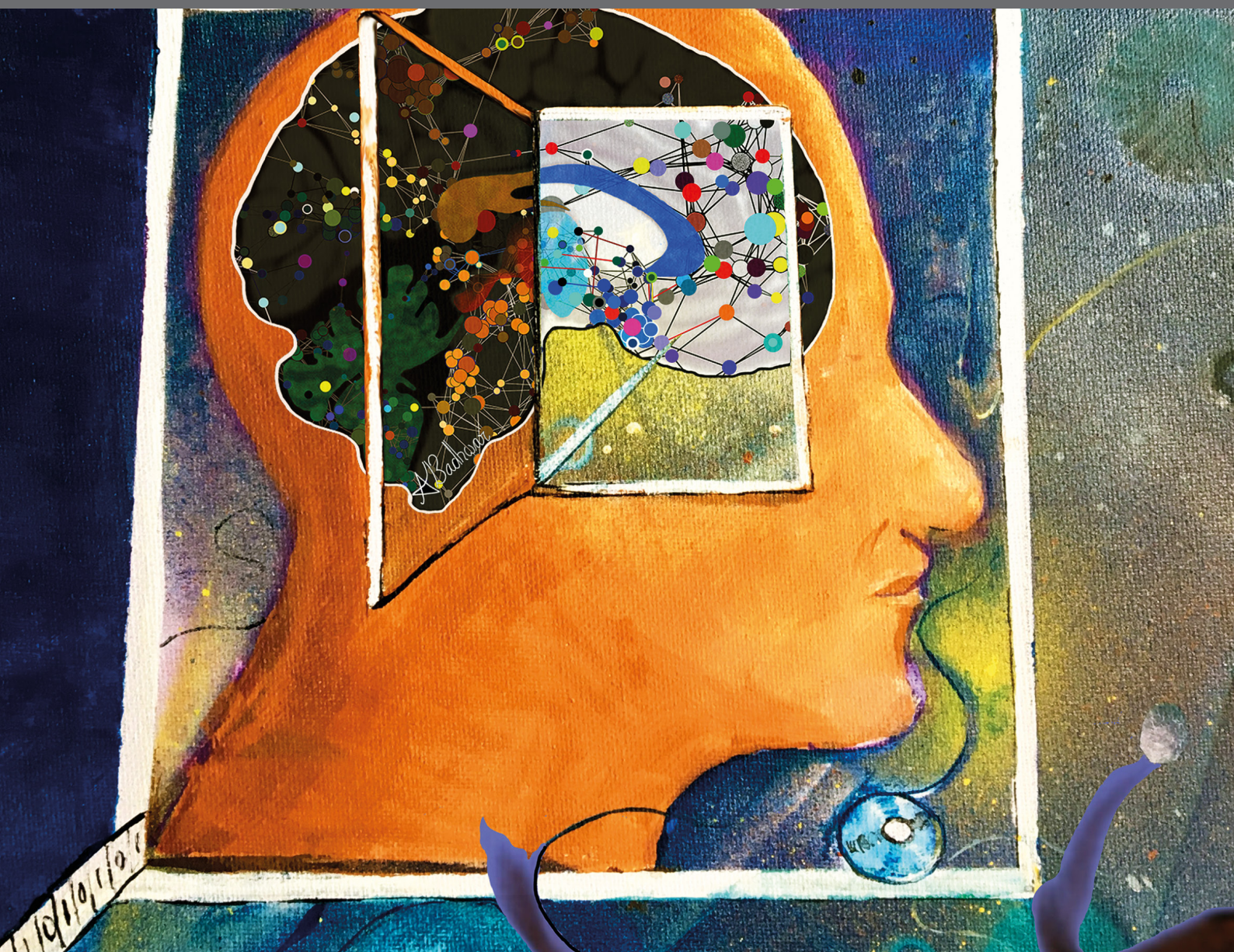
EDITED BY: Sook-Lei Liew, Lianne Schmaal and Neda Jahanshad

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

COLLABORATIVE EFFORTS FOR UNDERSTANDING THE HUMAN BRAIN

Topic Editors:

Sook-Lei Liew, University of Southern California, United States

Lianne Schmaal, Orygen, The National Centre of Excellence in Youth Mental Health, Australia; Centre for Youth Mental Health, The University of Melbourne, Australia

Neda Jahanshad, Mark and Mary Stevens Neuroimaging and Informatics Institute, University of Southern California, United States

Cover art by Urmila Das, AmanPreet Badhwar, and Peter Kochunov. The background was painted to look deep and dark, like the universe, by artist Urmila Das, who wanted to create a feeling of the mystic unknown. The lines and dots are there to give an idea of "connecting", not only with other people, but also to link the past and the future. The focus is on the head. Urmila chose the color orange to represent compassion and kindness, a symbol for humankind, and the open door is for new ways and opportunities. Through the open door we see the brain, an organ of infinite scale and mystery. Artist and scientist AmanPreet Badhwar attempts to capture the complexity of the brain by using brain science data and emulating the feel of inter-connected galaxies in the universe (or, as she calls it, the "brainverse"). Embedded in the background are artworks by scientist Peter Kochunov, generated using probabilistic tractography of fiber tracts going through the corpus callosum and the fornix, further reinforcing the concept of multiple and complex connections, be it in the brain or the universe.

The human brain is incredibly complex, and the more we learn about it, the more we realize how much we need a truly interdisciplinary team to make sense of its intricacies. This eBook presents the latest efforts in collaborative team science from around the world, all aimed at understanding the human brain.

Table of Contents

frontiers
in Neuroinformatics

Editorial: Collaborative Efforts for Understanding the Human Brain

Sook-Lei Liew [1,2]*, Lianne Schmaal [3,4] and Neda Jahanshad [2]

[1] *Chan Division of Occupational Science and Occupational Therapy, University of Southern California, Los Angeles, CA, United States,* [2] *Keck School of Medicine, Mark and Mary Stevens Neuroimaging and Informatics Institute, University of Southern California, Marina del Rey, CA, United States,* [3] *Orygen, The National Centre of Excellence in Youth Mental Health, Parkville, VIC, Australia,* [4] *Centre for Youth Mental Health, The University of Melbourne, Melbourne, VIC, Australia*

Editorial on the Research Topic

Collaborative Efforts for Understanding the Human Brain

Advancements in the technologies and methods used to study the brain have improved our capability to collect, share, and analyze large, detailed, datasets including brain imaging, genetic, andv behavioral data. This data-rich environment has allowed researchers to study the complex relationships between the structure and function of the brain throughout the lifespan with different behaviors or even clinical disease states. Today, research aimed at understanding the human brain necessitates new collaborative efforts that bring together domain experts across neuroscience, computer science, biology, engineering, statistics, medicine, and clinical practice in order to maximize the impact of these large and diverse datasets.

In this Research Topic, we highlight many novel and exciting aspects of this collaborative effort to study the human brain. This issue begins by providing perspectives on the evolving field of reproducible science, which is a main goal for collaborations in the field, and also highlights the emerging use of web-based applications for bringing together researchers from around the world. Next, several papers describe new large-scale neuroinformatics platforms that centralize data collection, simplify data management, implement rigorous quality control, and integrate complex multi-modal neuroimaging, genetic, and behavioral datasets. These are followed by several papers describing new software analysis pipelines for neuroimaging data (e.g., PET imaging and stroke MRI imaging), which provide flexible yet standardized and reproducible tools for data analysis. This issue also contains several reports comparing different ways of collecting and analyzing large, multi-site data, and provides new insights into best practices for multi-site diffusion MRI acquisition, neuroimaging data analysis, and imaging genetics analyses. Finally, several papers introduce new methods for analyzing multi-site data, including decentralized voxel-based analyses, hybrid mesio-temporal lobe segmentation, and analyses for post-traumatic epilepsy data. Overall, these papers demonstrate the breadth of work focused on bringing researchers together to decode the mysteries of the human brain.

PERSPECTIVES ON REPRODUCIBLE SCIENCE

As research becomes more collaborative, excitement around issues of reproducible and open science are growing. In this special issue, several authors provided insight on how to perform reproducible and collaborative research using cutting-edge open-source tools. For example, Kennedy et al. describe work emerging from ReproNim: A Center for Reproducible NeuroImaging Computation and detail principles of reproducible science, focusing on publications that can be re-executed. In doing so, they highlight a number of ReproNim

tools for data management, analysis, and reporting. Building on this, Keshavan and Poline share their perspective on a new wave of collaborative science happening over web-based platforms. They outline how the internet allows for improved sharing of data and research projects, and they share numerous web-based tools for everything from crowd-sourced data analysis to collaborative writing.

FRAMEWORKS AND PLATFORMS TO FACILITATE DATA SHARING

The implementation of these principles for open science is also detailed in several papers describing new neuroinformatic platforms for managing and harmonizing large multi-modal datasets. First, Mohaddes et al. described a neuroinformatics platform developed for the Canadian Consortium on Neurodegeneration in Aging (CCNA). This framework uses the LORIS data management system and supports the acquisition, storage, curation and dissemination of imaging, genetic, clinical, and biospecimen data related to aging. Next, Das et al. describe an integrated "-omics" framework, also using the LORIS platform, for analyzing multiple types of "-omics" datasets, including genomics, imaging, and behavior. Similarly, Vaccarino et al. describe the Ontario Brain Institute's Brain-CODE platform, which facilitates clinical, neuroimaging, and molecular data management, analysis and sharing in one consolidated, open-source platform. Finally, Rotenberg et al. describe a use case of the Brain-CODE platform from the Center for Addiction and Mental Health (CAMH). This platform provides an environment for centralized data capture, visualization, and analysis for psychiatric data. Importantly, all of these platforms and examples emphasize key issues in data management, such as privacy, data permissions, and quality control. They also all focus on integrating complex multimodal data sources in a manner that is easy to curate and share.

SOFTWARE PIPELINES FOR HARMONIZED ANALYSES

While the neuroinformatics platforms aim to simplify data management, there are also attempts to develop software pipelines that will standardize data analysis. Funck et al. describe the APPIAN (Automated Pipeline for PET Image Analysis) toolbox, which allows for the robust, reproducible analysis of PET imaging data with many options for flexible processing. For stroke MRI data, Ito et al. describe the PALS (Pipeline for Analyzing Lesions After Stroke) toolbox, which supports large-scale lesion analysis and quality control with many user-defined options for analysis. Both are good examples of software pipelines that can harmonize data analysis across research sites and improve reproducibility of results.

HARMONIZATION OF PROTOCOLS AND METHODS

In addition to new platforms and software, another important issue in collaborative research is to evaluate and compare different methods of data acquisition and analysis. To this end, a number of papers examined different methods for acquiring and analyzing data in order to harmonize data collection and analysis across research sites. Zavaliangos-Petropulu et al. compared six different diffusion tensor imaging (DTI) scanner acquisition protocols acquired across 47 different research sites from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Although they found differences in diffusion metrics based on the imaging protocol, they were able to successfully pool the data across the sites and protocols into one cohesive dataset. Boedhoe et al. from the ENIGMA Obsessive-Compulsive Disorder Working Group used multi-site data from 38 research sites to compare statistical approaches to pooling MRI-derived measures, and found that for this type of data, mega-analytic approaches are favorable to meta-analytic analyses. Kochunov et al. compared different methods for estimating heritability from imaging genetics data using a host of tools across multiple datasets. They found that although the different methods yielded different results depending on the dataset and the approach, incorporating several homogenization steps prior to estimating heritability was effective in producing converging results across methods. Each of these important contributions not only shows how multi-site data can be affected by different methods, but also provides recommendations for improving harmonization of the data.

METHODS FOR MULTI-SITE DATA ANALYSES

Finally, several papers described new methods for analyzing multi-site data. Gazula et al. proposed new decentralized methods for structural (voxel-based morphometry) and functional (dynamic functional connectivity) analyses, and compared this with standard centralized methods. They found that the decentralized methods worked equally well as centralized methods but are more flexible for use with multi-site data, opening the doors for large-scale collaborative analyses without bulky data-transfers. Kim et al. discussed a new hybrid template approach for automated segmentation of mesiotemporal structures, including the hippocampus, amygdala, and parahippocampal gyrus, which reliably performs better than existing segmentation methods across multiple datasets. Finally, Duncan et al. describe preliminary methods and results for analyzing multi-site data in individuals with epilepsy following traumatic brain injury from the multi-site Epilepsy Bioinformatics Study for Antiepileptogenic Therapy (EpiBioS4Rx) study, which collects MRI, EEG, and intracranial EEG from humans and animals. Overall, these papers reveal new methods specific for multi-site data analysis.

CONCLUSIONS AND FUTURE DIRECTIONS

Work presented in this Research Topic collectively highlights the growing trend for collaborative efforts for the neurosciences. These collaborations come in the form of developing tools for external researchers to access or contribute data, developing

methods that are confirmed to be robust across MRI datasets and acquisitions, and empirically testing harmonization methods for diverse datasets. Through the tools, methods, and results presented in this issue and beyond, researchers around the world are teaming up to ensure this new era of science provides robust, reliable and internationally meaningful findings to drive the understanding of the human brain forward.

AUTHOR CONTRIBUTIONS

All three authors served as editors of this Research Topic and contributed to the conceptualization of this Research Topic. The editorial was drafted by SL-L and revised and reviewed by LS and NJ.

FUNDING

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.

frontiers
in Neuroinformatics

Everything Matters: The ReproNim Perspective on Reproducible Neuroimaging

David N. Kennedy[1]*, Sanu A. Abraham[2], Julianna F. Bates[1], Albert Crowley[3],
Satrajit Ghosh[2], Tom Gillespie[4], Mathias Goncalves[2], Jeffrey S. Grethe[4],
Yaroslav O. Halchenko[5], Michael Hanke[6], Christian Haselgrove[1], Steven M. Hodge[1],
Dorota Jarecka[2], Jakub Kaczmarzyk[2], David B. Keator[7], Kyle Meyer[5],
Maryann E. Martone[4], Smruti Padhy[2], Jean-Baptiste Poline[8], Nina Preuss[3],
Troy Sincomb[4] and Matt Travers[3]

[1] Eunice Kennedy Shriver Center, Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA, United States, [2] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, United States, [3] TCG, Inc., Washington, DC, United States, [4] Department of Neuroscience, University of California, San Diego, San Diego, CA, United States, [5] Department of Psychological and Brain Sciences, Dartmouth College, Dartmouth, NH, United States, [6] Institute of Psychology, University of Magdeburg, Magdeburg, Germany, [7] Department of Psychiatry and Human Behavior, University of California, Irvine, Irvine, CA, United States, [8] Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

There has been a recent major upsurge in the concerns about reproducibility in many areas of science. Within the neuroimaging domain, one approach is to promote reproducibility is to target the re-executability of the publication. The information supporting such re-executability can enable the detailed examination of how an initial finding generalizes across changes in the processing approach, and sampled population, in a controlled scientific fashion. ReproNim: A Center for Reproducible Neuroimaging Computation is a recently funded initiative that seeks to facilitate the "last mile" implementations of core re-executability tools in order to reduce the accessibility barrier and increase adoption of standards and best practices at the neuroimaging research laboratory level. In this report, we summarize the overall approach and tools we have developed in this domain.

Keywords: reproducibility, neuroimaging, data model, publication, re-executability

INTRODUCTION

There has been a recent major upsurge in the concerns about reproducibility in many areas of science (Ioannidis, 2005, 2011; Button et al., 2013). The reasons for the concern are numerous, and there are numerous practices in the scientific field that have been found to exacerbate the problem. At a high level, a premium is put on novel, high-profile publications (in contrast to replications and negative findings) and a specific p-value (typically 0.05) as a proxy for truth has been adopted (Simonsohn et al., 2014; Wasserstein and Lazar, 2016). These aspects, in the context of a scientific reporting system that is out of touch with the digital age, have combined to create a perfect storm of practices that do not readily support the transparency needed to embrace reproducibility more substantively (Martone, 2015; Starr et al., 2015).

In acknowledgment of this situation, each scientific field is forced to re-examine the best-practices that are expected of practitioners in that field. Each field grapples with what reproducibility looks like within the context of that field. Neuroimaging provides a lens on various biological processes, and how these biological processes change over the course of development, and

in the face of pathological insult. As the biological process is the ultimate target of the neuroimaging inquiry, the question of reproducibility relates principally to the conclusions reached about such processes. A true biological inference about a population or process should generalize to other valid ways of observing that process and other samples of that population. In the quest to advance the overall reproducibility of neuroimaging science, one approach is to target the re-executability of the publication; the basic, current building block of the dissemination of scientific knowledge. The information supporting such re-executability can enable the detailed examination of how an initial finding generalizes across changes in the processing approach, and sampled population, in a controlled scientific fashion (see **Figure 1A**). It is only in the context of a systematic ability to probe a finding that the true generalizability of a claim can emerge.

It can be argued that "everything matters" in the generalizability of the traditional neuroimaging publication. The issues already identified span all levels of the experimental ecosystem:

- Computational environments matter (Glatard et al., 2015);
- Tool selection matters (Tustison et al., 2014; Dickie et al., 2017);
- Tool version matters (Dickie et al., 2017);
- Statistical model matters (Tan et al., 2016);
- Study population characteristics matter.

In the context of all these things that matter, what is an appropriate approach that investigators in this field should take? Our position is that the key to a comprehensive understanding of the published neuroimaging literature is to comprehensively, and in a machine-accessible manner, describe each of the elements of the experiment: input data, processing steps, computational environment, statistical assessment, and complete results (Ghosh et al., 2017a). The human understandable interpretations and claims, typical of a publication, can then exist around these machine-readable [and hence Findable, Accessible, Interoperable and Reusable (i.e., FAIR; Wilkinson et al., 2016)] elements. The existence of this machine readable and actionable *provenance* (the description of the origins of all elements of the publication) is what is needed to trace back and validate the underpinnings of a claim, and the starting point for the systematic examination of that claims' generalizability.

Within the neuroimaging community, the prognosis for the ability to establish a complete description of accessible elements for all parts of the publication is quite good. The field has good data standards [DICOM[1], NIfTI[2], BIDS[3] (Gorgolewski et al., 2016), MINC[4], etc.], excellent platforms for the sharing of code and data management and sharing (Git, GitHub, DataLad, OSF, etc.), there are ample raw data

repositories (XNAT[5] (Herrick et al., 2016), NITRC-IR[6] (Kennedy et al., 2016), NIMH Data Archive (NDA)[7], International Neuroimaging Data-sharing Initiative (INDI)[8] (Mennes et al., 2013), Human Connectome Project (HCP)[9] (Marcus et al., 2013), OpenNeuro[10], etc.), numerous workflow systems (Nipype[11] (Gorgolewski et al., 2011), LONI Pipeline[12] (Rex et al., 2003), etc.), package and execution management systems (NeuroDebian[13], Docker[14], NeuroDocker[15], Singularity[16], NITRC-CE[17], etc.), and several outlets to disseminate results (NeuroVault[18], BrainSpell[19], NeuroSynth[20], etc.). Importantly, a standard data model for the description of all these research elements, the Neuroimaging Data Model (NIDM)[21] (Keator et al., 2013), is also in place to facilitate and distribute semantically annotated and unambiguous representations of the complete experimental cycle. As such, the main barrier to the generation of re-executable publications which foster reproducibility and generalizability is not the core resources, but rather the ease of use alongside the acceptance of best practices (Eglen et al., 2017; Nichols et al., 2017), in the typical neuroimaging laboratory. In addition to knowing that the resources for reproducibility exist, the community needs to embrace an approach of "Reproducible by Design" (as opposed to reproducibility as an afterthought). ReproNim: A Center for Reproducible Neuroimaging Computation is a recently funded initiative that seeks to facilitate the "last mile" implementations of these core tools in order to reduce the accessibility barrier and increase adoption of standards and best practices at the research laboratory level.

REPRONIM APPROACH

In the remainder of this report, we provide an annotated perspective on the ReproNim vision for the re-executable publication. For this purpose, we concentrate on a laboratory data acquisition centric version of the research workflow. Other workflows (i.e., data query from accessible data resources) can be envisioned, but will be outside the purview of this report. **Figure 1B** depicts a stylized version of the data workflow in a typical neuroimaging experiment. Current publication practice focuses on human readable descriptions of the detailed data

---

[1]https://www.dicomstandard.org/current/

[2]https://nifti.nimh.nih.gov/nifti-1/

[3]http://bids.neuroimaging.io/

[4]https://en.wikibooks.org/wiki/MINC/SoftwareDevelopment/MINC2.0_File_Format_Reference

[5]https://www.xnat.org/

[6]https://www.nitrc.org/ir/

[7]https://data-archive.nimh.nih.gov/

[8]http://fcon_1000.projects.nitrc.org/

[9]https://www.humanconnectome.org/

[10]https://openneuro.org/

[11]https://nipype.readthedocs.io/en/latest/

[12]http://pipeline.loni.usc.edu/

[13]http://neuro.debian.net/

[14]https://www.docker.com/

[15]https://github.com/kaczmarj/neurodocker

[16]https://www.sylabs.io/docs/

[17]https://www.nitrc.org/ce/

[18]https://neurovault.org/

[19]http://brainspell.org/

[20]http://neurosynth.org/

[21]http://nidm.nidash.org/
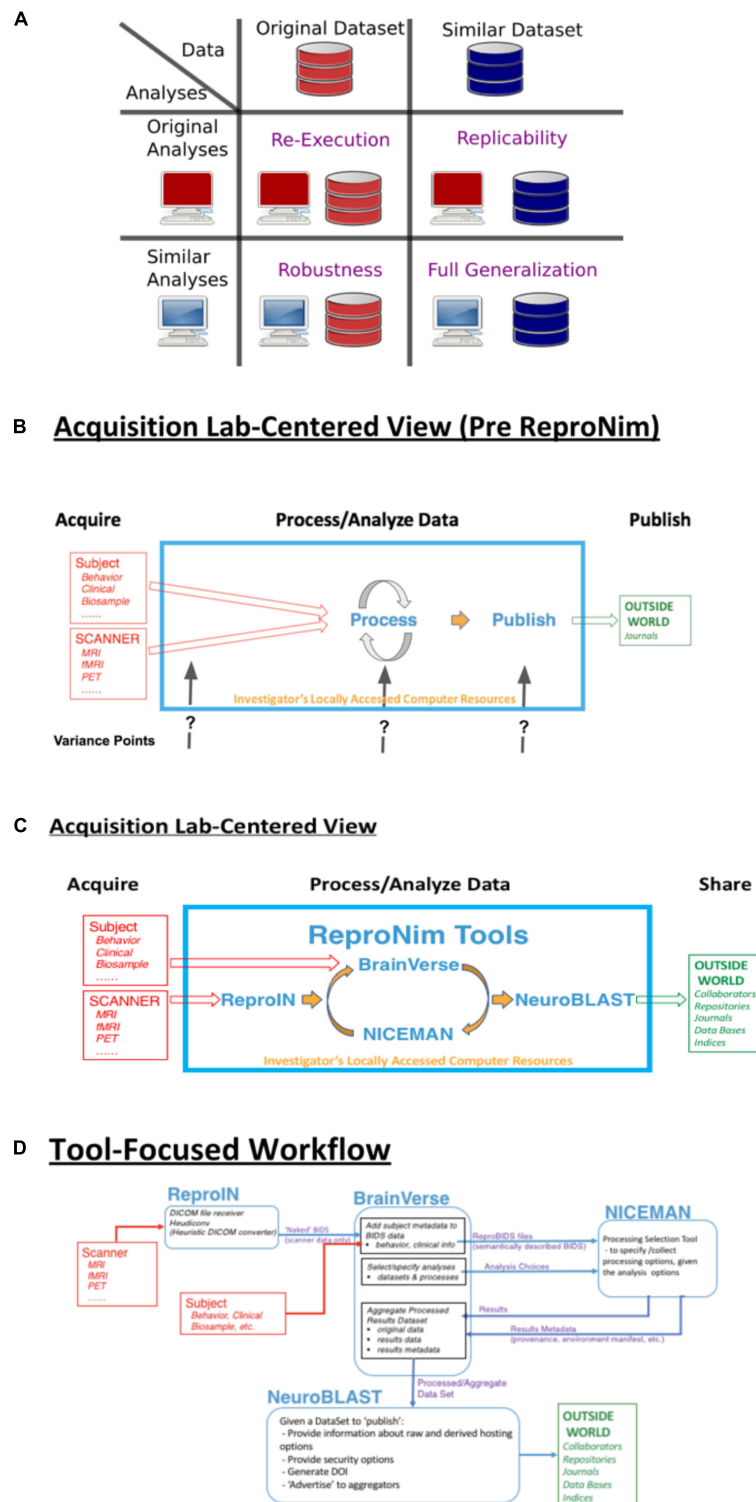
FIGURE 1 | ReproNim conceptual workflows. **(A)** Pictorial depiction of the concepts of re-executability (same data, same analysis), replication (same analysis, similar data), robustness (same data, similar analysis), and generalization (similar data, similar analysis). Adapted from multiple sources, including Dr Drummond (2009), Peng (2011); Hong (2015), Goodman et al. (2016); Whitaker (2016), and Allard (2018). **(B)** General neuroimaging data workflow: Imaging data and behavioral/clinical measures enter into a local analysis, generate results that then get published. Substantial variability in the published literature exists in how the data, analysis and results are described. **(C)** The ReproNim vision of the general neuroimaging data workflow where control of the data model and machine-readable markup is invoked to completely represent the data workflow, processing and results using the tools of ReproIn, BrainVerse, NICEMAN, and NeuroBlast. **(D)** Detailed data transformations and markup as the data work their way through the planned analysis and tools.

collection, the processing workflow and environment, and the statistical procedures and results. Therefore, across the field, there is vast variability in the detail, precision and completeness of these published descriptions. This variance in description may contribute to the limited ability of the field to replicate findings. Because we do not know exactly what a given paper did or observed, when a subsequent paper examines a similar topic it is impossible to parse similarities and differences in results appropriately. **Figure 1C** overviews the ReproNim vision for taking control of these variance points, through instrumentation that generates machine-readable provenance in each of the following areas: experimental data description and versioning (NIDM-E), processing workflow (NIDM-WF), and results (NIDM-R). While the analytic processing steps for a neuroimaging workflow using any processing tool (SPM[22], FSL[23], FreeSurfer[24], AFNI[25], etc.) will remain identical and completely under the researcher's control, we will insert simple "wrapper" functionality that manage the conversion and markup of incoming imaging data (ReproIn), markup of subject-specific observations and experiment-specific analysis plans (BrainVerse), interrogation and management of execution environments (NICEMAN), and the distribution of the results to user-identified, appropriate and FAIR data repositories (NeuroBLAST). The data transformations and annotations that these tools impart upon the data flow are illustrated in **Figure 1D**.

MATERIALS AND METHODS

In this section we will review the current status of the key tools that are in place to support the re-executable publication. Each resource will be summarized in terms of its purpose, how to access it, and its functionality as of this writing.

ReproIn

ReproIn is a specification and a software platform to fully automate acquisition, preparation and layout of collected MRI data in the BIDS data structure with DataLad version management, so they will be ready for local distribution and processing in a scalable and flexible manner, while retaining all provenance information from the moment of their creation, in order to ease later sharing or publication.

ReproIn is accessed from the ReproIn Github repository[26].

To not reinvent the wheel, the software development of ReproIn is largely done through contribution to existing software projects: *HeuDiConv*[27] – a flexible DICOM converter for organizing brain imaging data into structured directory layouts; and *DataLad*[28] – a modular version control platform and distribution for both code and data including entire

containerized computation environments via the DataLad-containers extension and automated execution provenance recording within version control systems (VCS) using DataLad's "run" functionality to provide a fully re-executable VCS-tracked analysis record. The ReproNim project actively contributes to those existing solutions to provide all necessary components for computationally reproducible research.

General features of ReproIn include:

- A flexible naming convention for study description and acquisition details to be used at an MR scanner console that extends the information typically available in DICOM metadata, to allow for an automated translation of MR scans in DICOM format into BIDS datasets.
- A HeuDiConv ReproIn heuristic implementation[29] to process and validate the above BIDS specification[30].
- Support for automated metadata generation by HeuDiConv (e.g., to tag potentially sensitive information) using DataLad's metadata capabilities.
- Datasets can be incrementally expanded with new acquisitions, as well as merged with any changes (new data, adjusted templates) from the data acquisition server.
- Optional automatic obfuscation of time stamps in the VCS records to protect privacy of study participants.
- Standalone Docker and Singularity containers for turnkey processing and analysis deployment.

BrainVerse

BrainVerse is a cross-platform software framework and collaborative desktop application to help researchers annotate the research workflow from experimental planning to execution of analysis. Annotation includes semantic coding of all data elements, as well as the merging of the imaging data and behavioral/clinical data streams, resulting in semantically marked up BIDS data structures (the so called "ReproBIDS" datasets) also under DataLad version management. Key application areas include:

- Harmonization with the NIMH Data Archive (NDA): Allows importing and curating the NDA schemas to generate collection instruments that support harmonization of variables within and across project.
- Project planner and executor: Allows creating a plan for an experimental protocol in a project and collecting data using harmonized and reusable forms.
- NIDM term editor: Allows the community to search for and build a common descriptive vocabulary around neuroimaging.

BrainVerse is accessed from the BrainVerse website/Github repository[31].

General features include:

---

[22]https://www.fil.ion.ucl.ac.uk/spm/

[23]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki

[24]https://surfer.nmr.mgh.harvard.edu/

[25]https://afni.nimh.nih.gov/

[26]https://github.com/ReproNim/reproin

[27]https://github.com/nipy/heudiconv

[28]https://www.datalad.org/

[29]https://github.com/nipy/heudiconv/blob/master/heudiconv/heuristics/reproin.py

[30]https://github.com/nipy/heudiconv/blob/master/heudiconv/heuristics/reproin_validator.cfg

[31]https://github.com/ReproNim/brainverse

- GitHub based login and authorization;
- Cross-platform support (based on the Electron[32] framework) on desktops with an option for server based installation;
- NDA harmonizer and editor:
  - Import, select, and preview forms,
  - Edit forms,
  - Push to common repository on ReproNim,
  - Pull curated forms from ReproNim repository;
- Project planner and form-based data collector:
  - Create project execution plan with multiple session support,
  - Reuse/Create session instruments,
  - Add participants to project and collect data using project plan,
  - Export collected data to CSV files for visualization and analysis;
- NIDM term editor:
  - Display of terms from NIDM owl files,
  - Edit terms and send review requests.

NICEMAN

NICEMAN is a specification and software system that supports the management of computation environments and computations, targeting the neuroimaging domain. It provides:

- A specification to describe environments consistently across available data and software distributions (e.g., VCS such as git, Debian-based systems, Conda Python distribution, Singularity and Docker images),
- A software platform to allow convenient discovery, description, and management of the computation environment(s) so that they could be easily traced (to automatically collect the specification from existing environment or a dedicated process), validated (to satisfy necessary requirements), compared (to determine how requirements differ), satisfied (to create a new or adjust an existing environment), execute necessary computation(s) and interface the output(s).

NICEMAN developed openly and accessible on Github[33]. General features include:

- *retrace* command allows users to establish a detailed description of the environment given an initial specification (e.g., from reprozip[34], Nipype's.trig PROV) or from a list of files provided on the command line. It generates tracing information that is sufficient for re-establishing the environment (origins, versions, etc.) for Debian-based systems, VCS (svn and git), and Conda,

- Support of Docker, shell (via ssh or localhost) environments for scripted or interactive sessions, with a centralized resources manager (ls command to list available environments/resources and query their status), and with basic support for Docker and Amazon Web Services (AWS) backends life-time (bootup/shutdown),
- Create/install commands to fulfill the specification and provide the requested environment (via Docker, AWS, etc.).
- Support for Singularity and Docker environments.

NeuroBlast

NeuroBlast is a share, search and discovery service. The NeuroBlast service facilitates data sharing (raw and results) of known existing repositories and assists users in the data discovery process to find matching/similar studies based on a combination of task, analysis, and activation patterns. This novel environment utilizes all information about a study, enabling researchers to select appropriate sharing sites, and find similar studies utilizing a number of different similarity metrics. This service employs deep semantics, building from terminologies managed by InterLex and its associated ontology, to enhance the search for similar data sets utilizing multiple features for comparison.

InterLex can be accessed at InterLex.org and the ontology can be accessed from its GitHub repository[35].

RESULTS

In this section, we briefly summarize a couple of example use-cases that demonstrate the ReproNim vision in action.

Tools Matter

Shared neuroimaging data is an important means of promoting an open and reproducible neuroimaging analysis culture. The Autism Brain Imaging Data Exchange (ABIDE1) dataset (Di Martino et al., 2014) is a premier example of shared neuroimaging data that promotes exploration of the factors related to the autism diagnosis relative to features accessible in structural and resting state functional MRI in over 1000 subjects. There are many factors related to the reproducibility of neuroimaging findings, including selection of software tools. In this report, we take advantage of the ABIDE Preprocessed Connectomes project[36] which has performed a comparative analysis of ABIDE1 data using three widely used structural analysis software tools: FreeSurfer (Fischl et al., 2002), versions 5.1 and 5.3, and ANTS (Avants et al., 2011). In an ideal world, regional thickness data would be independent of the specific software tool used to generate the result, when applied to common data. We utilize this dataset to evaluate the extent to which the selection of a software tool matters, and provide a common open source platform to support further exploration of these results. We identified the subset of (976 cases (from the 1112 ABIDE1 original cases)) that had completed all three analyses and are available at the ABIDE Preprocessed Connectomes site.

---

[32]https://electronjs.org/
[33]https://github.com/repronim/niceman
[34]https://www.reprozip.org/#

[35]https://github.com/SciCrunch/NIF-Ontology
[36]http://preprocessed-connectomes-project.org/abide/
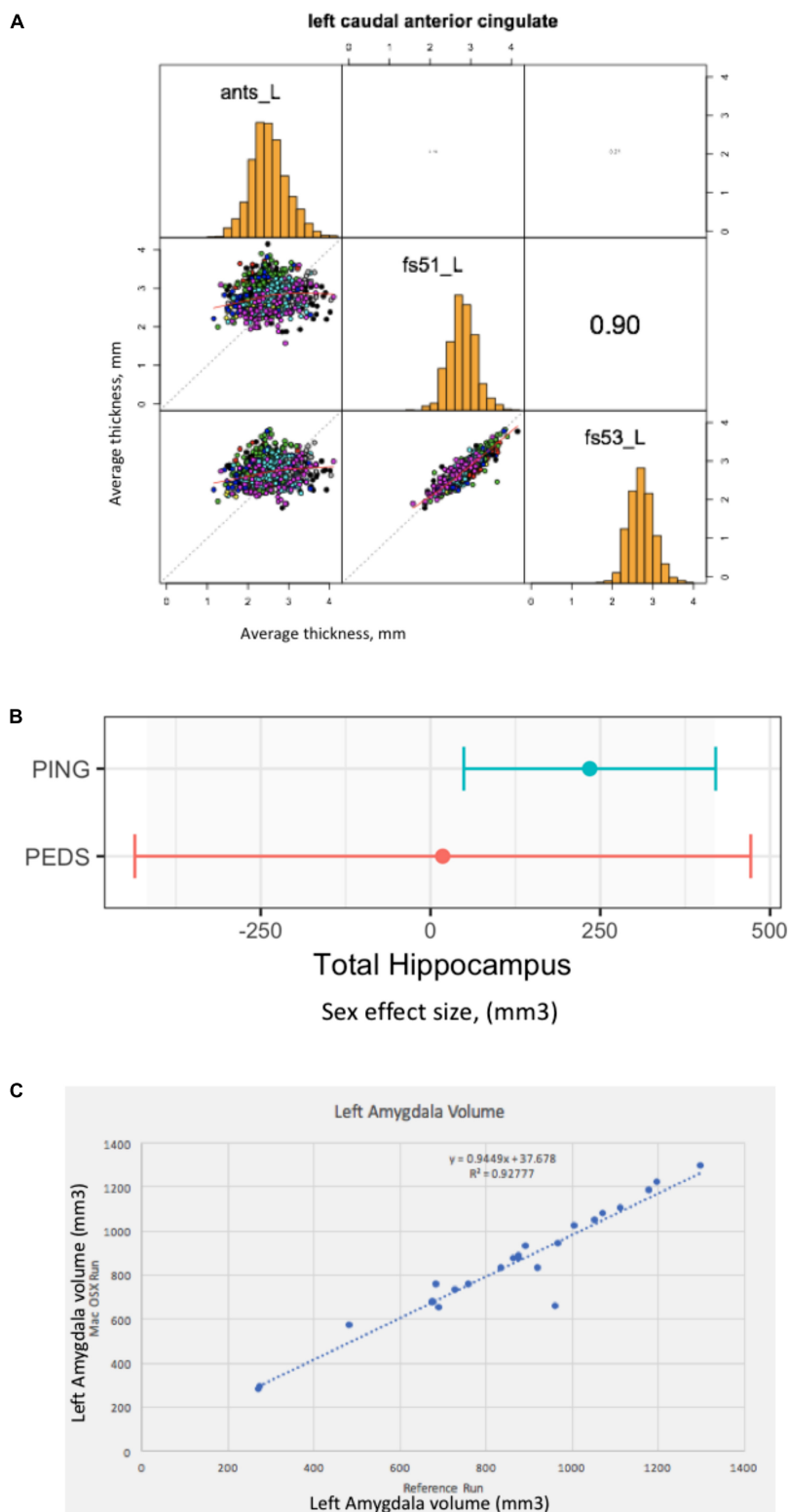
FIGURE 2 | Everything matters. **(A)** Tools matter: Same data (976 ABIDE1 cases), different tools (FreeSurfer 5.1 and 5.3 and ANTS). For a specific anatomic region (left caudal anterior cingulate cortex), we show a matrix of the between tool comparisons. On the diagonal (from upper left to lower right) we see the distribution histogram of average left caudal anterior cingulate cortex thicknesses for ANTS, FreeSurfer 5.1 and FreeSurfer 5.3, respectively. The three scatter plots (left column,

(Continued)

FIGURE 2 | Continued

middle, left column bottom and middle column bottom) show the between tool scatter plots and regression line for these data for: ANTS vs. FreeSurfer 5.1 (Pearson's correlation coefficient $r = 0.16$); ANTS vs. FreeSurfer 5.3 (Pearson's correlation coefficient $r = 0.21$); and FreeSurfer 5.1 vs. FreeSurfer 5.3 (Pearson's correlation coefficient $r = 0.90$), respectively. **(B)** Sample size matters: Same analysis (FreeSurfer 5.3 and a statistical model looking at gender effects in hippocampus volume) as a function of the large-scale publically available structural imaging data in typically developing children in ~2005 (NIH PEDS, $N = 325$) and ~2011 (PING, $N = 1239$). The plot shows the observed effect size and 95% confidence interval for the total hippocampal volume for these two cohorts. **(C)** Computational Environment Matters: Same data, same workflow, different workflow operating system environments results in different results, as shown for the volume of the left amygdala in subset of 24 cases. See text for further details.

The result of this effort is a publically available GitHub repository[37], which identifies the specific cases that are included, contains summary data tables of the volume and surface area results of the three analysis tools, software to load these data tables into the R statistical software analysis package (R reader), and an R script to correlate the corresponding analytical results between the different structural analysis runs. The surface-based results are represented as average cortical thickness for each of the 62 (31 bi-laterally represented) anatomic regions in the Desikan-Killiany-Tourville (DKT) atlas (Klein and Tourville, 2012). For each anatomic region, we calculate the three inter-tool result correlations (FreeSurfer 5.1 vs. FreeSurfer 5.3; FreeSurfer 5.1 vs. ANTS, and Freesurfer 5.3 vs. ANTS). Findings can be summarized as follows. The mean and range of region-wise correlation were observed as follows between the various tool-pair combinations: ANTS vs. FreeSurfer 5.1 mean regional correlation = 0.43, [minimum = 0.19 (rostralanteriorcingulate L), maximum = 0.59 (superiortemporal R)]; ANTS v FreeSurfer 5.3 mean regional correlation = 0.47, [minimum = 0.19 (caudalanteriorcingulate R), maximum = 0.67 (superiortemporal R)]; FreeSurfer 5.1 vs. FreeSurfer 5.3 mean regional correlation = 0.87, [minimum = 0.76 (insula R), maximum = 0.93 (paracentral L)]. The FreeSurfer analysis in this data presents excellent inter-version (5.1–5.3) commonality. There are, however, substantial differences between the regional thickness results between the FreeSurfer and ANTS analysis. As an example, the scatter plots and distributions for the left caudal anterior cingulate is shown in **Figure 2A**.

Sample Size/Quality Matters

In this example, we look at the potential gender effect of total hippocampal volume in typically developing children, and how an observation of this effect can evolve over time as a function of the imaging technology and the amount of available data. We model total hippocampus volume as a function of gender, covarying for age, sex by age interaction, site, and total cerebral volume. We used state-of-the-art at the time data available from two national typically developing cohorts. We first look at the gender effect as observed in ~2005 from the NIH Pediatric Database ($N = 325$ (159 males/166 females); aged 4.2–18.4 years) (Evans and Brain Development Cooperative Group, 2006). We also look at data from the PING cohort (Pediatric Imaging, Neurocognition and Genetics) (Jernigan et al., 2016), as released in ~2011 (total $N = 1239$ (644 males/595 females), aged 3–20 years). We applied a common analysis (FreeSurfer 5.3) using default parameters to each of these datasets in house. These

results are shown in **Figure 2B**. In this case, we note a lack of significant gender dimorphism of the total hippocampus seen in children from the PEDS cohort ($p = 0.9379$). However, the PING dataset documents a significant gender effect for the total hippocampus volume ($p = 0.013269$). While sample size is one of the differences between these studies, it is also the case the image quality and acquisition technology had evolved in the years between these two studies. Nevertheless, we feel that this type of observation is reflects the types of conclusions that are often gleaned from the literature: observations that are not significant based upon older, smaller N studies may not generalize to newer, larger N studies. The tightening of the error bars around a specific observation can be attributed to many sources, not the least of which, in this case is the sample size. Indeed, the observed effect size in the PING sample falls within the observed range of the older, smaller PEDS distribution of observations.

Simple Re-executable Publication

In this last example, we document a set of procedures, which include supplemental additions to a manuscript, that unambiguously define the data, workflow, execution environment and results of a neuroimaging analysis, in order to generate a verifiably re-executable publication. Re-executability provides a starting point for examination of the generalizability and reproducibility of a given finding. We have provided an example "publication" with four supplementary files (Ghosh et al., 2017a), the: (1) data file, (2) workflow file, (3) execution environment specification, and (4) results. In this example, the data is from 24 publically accessible typically developing subjects between the ages of 10–15 that have a structural scan at 3 Tesla available from the 1000 Functional Connectomes Project at NITRC (doi 10.18116/C6C592; Kennedy, 2017). The workflow is a FSL-based (version 5.0.9) assessment of total brain, gray and white matter and subcortical structural volumes and is accessible at doi: 10.5281/zenodo.800758, (Ghosh et al., 2017b). The execution environment is controlled through the use of Docker; the docker image is available at https://github.com/ReproNim/simple_workflow. Finally, the complete results of the reference run are stored in the expected_output folder of the GitHub repository[38]. By sharing the results of this reference run, as well as the data workflow, and a program to compare results from different runs, we can enable others to verify that they can arrive at the exact same result (if they use the exact same workflow and execution environment), or how close they come to the reference results if they utilize a different computational system (that may differ

---

[37]https://github.com/companat/compare-surf-tools

[38]https://github.com/ReproNim/simple_workflow/tree/1.1.0/expected_output

in terms of operating system, software versions, etc.). **Figure 2C** demonstrates the imprecision of "the same data and workflow" run (in this case left amygdala volumes for each of the 24 subjects) on different hardware platforms (Docker Debian 8.7 (Reference Run) vs. Mac OS X 10.12.4), documenting the importance of taking control over the complete description of all elements of the reported research publication. Ideally, while the amygdala volume will differ by subject, the same workflow when rerun should yield the line of identity. It is the case that when the same Docker image is run, the identical results are generated. However, as illustrated in **Figure 2C**, running the same workflow on a Debian 8.7 vs. Mac OS X 10.12.4 system the results deviate substantially from the expected relationship.

SUMMARY

In this perspective we have reviewed the ReproNim vision and rationale for enhancing the reproducibility of the neuroimaging literature through an emphasis on individual publication re-executability. A given publication, if published in a completely re-executable fashion, forms the basis for future systematic explorations of the generalization of the observations through independent manipulation of the data and processing details separately. Reproducible claims and conclusions are supported by findings that are generalizable to data beyond that originally reported and should be demonstrated to be robust with respect to details of the analytic approach. The key to controlling the re-executability of the publication is the generation and reporting, at all stages of the process, machine readable provenance documentation that details the input data sources, the analysis workflow, the statistical model, the execution environment and the complete results. Since we know that all these factors matter,

a good scientific report should be able to describe each of these factors unambiguously.

Time will tell if the tools and procedures promoted by the ReproNim effort (or other efforts) to enhance publication level re-executability will be successful. We can assert that the majority of neuroimaging publications to date do not expose this complete set of publication details explicitly. We envision a future re-executability check list that can be retrospectively applied by the community to the corpus of publications (or, better yet, used by reviewers of publications prospectively) that generates a catalog of compliant elements on a publication by publication basis. One can then observe, over time, the extent to which the exposure of publication elements (input data, workflow, execution environment, complete results) increases. Efforts are underway to generate more compelling scientific examples of the re-executable publication in response to exploring the generalizability of specific findings in the autism and schizophrenia literature.

AUTHOR CONTRIBUTIONS

All authors participated in the conception of the project, the development of various software aspects and writing of this manuscript.

FUNDING

REFERENCES

Allard, T. (2018). *Down the Rabbit Hole. A 101 on Reproducible Workflows with Python.* Available at: https://bitsandchips.me/Talks/PyCon#/title.

Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., and Gee, J. C. (2011). An open source multivariate framework for N-Tissue segmentation with evaluation on public data. *Neuroinformatics* 9, 381–400. doi: 10.1007/s12021-011-9109-y

Button, K. S., John, P. A. I., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Di Martino, A., Yan, C. G., Li, Q., Denio, E., Castellanos, F. X., Alaerts, K., et al. (2014). The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Mol. Psychiatry* 19, 659–667. doi: 10.1038/mp.2013.78

Dickie, E., Hodge, S., Craddock, R., Poline, J. B., and Kennedy, D. (2017). Tools matter: comparison of two surface analysis tools applied to the ABIDE dataset. *Res. Ideas Outcomes* 3:e13726. doi: 10.3897/rio.3.e13726

Dr Drummond, C. (2009). *Replicability is not Reproducibility: nor is it Good Science.* Available at: http://cogprints.org/7691/

Eglen, S. J., Marwick, B., Halchenko, Y. O., Hanke, M., Sufi, S., Gleeson, P., et al. (2017). Toward standard practices for sharing computer code and programs in neuroscience. *Nat. Neurosci.* 20, 770–773. doi: 10.1038/nn.4550

Evans, A. C., and Brain Development Cooperative Group (2006). The NIH MRI study of normal brain development. *NeuroImage* 30, 184–202. doi: 10.1016/j.neuroimage.2005.09.068

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* 33, 341–355.

Ghosh, S. S., Poline, J. B., Keator, D. B., Halchenko, Y. O., Thomas, A. G., Kessler, D. A., et al. (2017a). A very simple, re-executable neuroimaging publication. *F1000Res.* 6:124. doi: 10.12688/f1000research.10783.1

Ghosh, S. S., Halchenko, Y., Poline, J. B., and Kennedy, D. (2017b). *ReproNim/Simple_Workflow: Release 1.1.0 (Version 1.1.0).* Zenodo. doi: 10.5281/zenodo.800758

Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., et al. (2015). Reproducibility of neuroimaging analyses across operating systems. *Front. Neuroinform.* 9:12. doi: 10.3389/fninf.2015.00012

Goodman, S. N., Fanelli, D., and Ioannidis, J. P. A. (2016). What does research reproducibility mean? *Sci. Transl. Med.* 8:341s12. doi: 10.1126/scitranslmed.aaf5027

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* 5:13. doi: 10.3389/fninf.2011.00013

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing

and describing outputs of neuroimaging experiments. *Sci. Data* 3:160044. doi: 10.1038/sdata.2016.44

Herrick, R., Horton, W., Olsen, T., McKay, M., Archie, K. A., and Marcus, D. S. (2016). XNAT central: open sourcing imaging research data. *NeuroImage* 124(Pt B), 1093–1096. doi: 10.1016/j.neuroimage.2015.06.076

Hong, N. C. (2015). *Open Software for Open Science*. Available at: https://www.fosteropenscience.eu/foster-taxonomy/open-source-open-science

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124

Ioannidis, J. P. A. (2011). Excess significance bias in the literature on brain volume abnormalities. *Arch. Gen. Psychiatry* 68, 773–780. doi: 10.1001/archgenpsychiatry.2011.28

Jernigan, T. L., Brown, T. T., Hagler, D. J., Akshoomoff, N., Bartsch, H., Newman, E., et al. (2016). The pediatric imaging, neurocognition, and genetics (PING) data repository. *NeuroImage* 124(Pt B), 1149–1154. doi: 10.1016/j.neuroimage.2015.04.057

Keator, D. B., Helmer, K., Steffener, J., Turner, J. A., Van Erp, T. G. M., Gadde, S., et al. (2013). Towards structured sharing of raw and derived neuroimaging data across existing resources. *NeuroImage* 82, 647–661. doi: 10.1016/j.neuroimage.2013.05.094

Kennedy, D. (2017). *ReproNim Simple Workflow Test Dataset. ReproNim*. Available at: http://iaf.virtualbrain.org/slp/10.18116/C6C592

Kennedy, D. N., Haselgrove, C., Riehl, J., Preuss, N., and Buccigrossi, R. (2016). The NITRC image repository. *NeuroImage* 124, 1069–1073. doi: 10.1016/j.neuroimage.2015.05.074

Klein, A., and Tourville, J. (2012). 101 labeled brain images and a consistent human cortical labeling protocol. *Front. Neurosci.* 6:171. doi: 10.3389/fnins.2012.00171

Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., et al. (2013). Human connectome project informatics: quality control, database services, and data visualization. *NeuroImage* 80, 202–219. doi: 10.1016/j.neuroimage.2013.05.077

Martone, M. E. (2015). FORCE11: building the future for research communications and e-scholarship. *BioScience* 65, 635–635. doi: 10.1093/biosci/biv095

Mennes, M., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2013). Making data sharing work: the FCP/INDI experience. *NeuroImage* 82, 683–691. doi: 10.1016/j.neuroimage.2012.10.064

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., et al. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20, 299–303. doi: 10.1038/nn.4500

Peng, R. D. (2011). Reproducible research in computational science. *Science* 334, 1226–1227. doi: 10.1126/science.1213847

Rex, D. E., Ma, J. Q., and Toga, A. W. (2003). The LONI pipeline processing environment. *NeuroImage* 19, 1033–1048. doi: 10.1016/S1053-8119(03)00185-X

Simonsohn, U., Nelson, L. D., and Simmons, J. P. (2014). P-Curve: a key to the file-drawer. *J. Exp. Psychol. Gen.* 143, 534–547. doi: 10.1037/a0033242

Starr, J., Castro, E., Crosas, M., Dumontier, M., Downs, R. R., Duerr, R., et al. (2015). Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Comput. Sci.* 1:e1. doi: 10.7717/peerj-cs.1

Tan, A., Ma, W., Vira, A., Marwha, D., and Eliot, L. (2016). The human hippocampus is not sexually-dimorphic: meta-analysis of structural MRI volumes. *NeuroImage* 124, 350–366. doi: 10.1016/j.neuroimage.2015.08.050

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., et al. (2014). Large-scale evaluation of ANTs and freesurfer cortical thickness measurements. *NeuroImage* 99, 166–179. doi: 10.1016/j.neuroimage.2014.05.044

Wasserstein, R. L., and Lazar, N. A. (2016). The ASA's statement on p-values: context, process, and purpose. *Am. Stat.* 70, 129–133. doi: 10.1080/00031305.2016.1154108

Whitaker, K. (2016). *Making Your Research Reproducible*. Cambridge: University of Cambridge.

Wilkinson, M. D., Dumontier, M., Jsbrand Jan Aalbersberg, I., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Conflict of Interest Statement: AC, NP, and MT are employed by TCG, Inc.

The remaining 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.

From the Wet Lab to the Web Lab: A Paradigm Shift in Brain Imaging Research

Anisha Keshavan[1]* and Jean-Baptiste Poline[2,3]

[1] Department of Speech and Hearing, Institute for Neuroengineering, eScience Institute, University of Washington, Seattle, WA, United States, [2] Faculty of Medicine, McConnell Brain Imaging Centre, Ludmer Centre for Neuroinformatics and Mental Health, Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, Canada, [3] Henry H. Wheeler Jr. Brain Imaging Center, Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, United States

Web technology has transformed our lives, and has led to a paradigm shift in the computational sciences. As the neuroimaging informatics research community amasses large datasets to answer complex neuroscience questions, we find that the web is the best medium to facilitate novel insights by way of improved collaboration and communication. Here, we review the landscape of web technologies used in neuroimaging research, and discuss future applications, areas for improvement, and the limitations of using web technology in research. Fully incorporating web technology in our research lifecycle requires not only technical skill, but a widespread culture change; a shift from the small, focused "wet lab" to a multidisciplinary and largely collaborative "web lab."

1. INTRODUCTION

The internet is ubiquitous and infiltrating every aspect of our lives by way of the web browser. Desktops, tablets, and cell phones have web browsers, but also televisions, game consoles, wristwatches, cars, glasses, and even refrigerators can effortlessly display all the information that resides on the internet. Information that, in theory, includes nearly all scientific knowledge.

The web browser has transformed our scientific practices, by giving us access to an almost infinite information resource. It provides a flexible and immediate platform for publishing research products. It gives us access to powerful computing platforms and databases. It enables us to collect large amounts of data from many people (e.g., citizen science). It is absolutely essential for communication and scientific collaboration. And above all, its main strength is its transportability; science, particularly in computational fields such as neuroimaging, can be performed anywhere (given a speedy internet connection).

Scientific collaboration is becoming increasingly important as computing technology enables us to rapidly collect and analyze data. The result of this data deluge is that we have an increased need for interdisciplinary, collaborative research. A combination of scientists with domain specific knowledge and those with a intimate grasp of computer science, data wrangling, and statistics/machine learning are needed to fully capitalize on the potential of large datasets.

We have witnessed enormous leaps of scientific knowledge that were a direct result of large scale collaborations, like the Human Genome Project, the Large Hadron Collider, ITER (research in nuclear fusion), and LIGO (to measure gravitational waves) to name a few. And it was primarily because of a large scientific collaboration at CERN where one of the most transformative technologies of the late 21st century was born: the World Wide Web.

Around the same time as the invention of the web came the invention of functional magnetic resonance imaging (Ogawa et al., 1990) in 1990, which revolutionized neuroscience research on brain-behavior relationships. Enabled with the ability to image brain function, neuroimaging researchers have been collecting vast amounts of data to answer more complex questions about the relationship between brain structure and function. And as a result, neuroimaging researchers are collecting large amounts of data, and encountering the same roadblocks and bottlenecks that come with any "big data" science. Here, we propose that by more deeply incorporating web technology into the lifecycle of neuroimaging research, we can not only accelerate neuroscience discoveries but also develop and test novel neuroscience questions. In the following sections, we discuss the paradigm shift that web technology brings to the scientific research lifecycle in terms of two main principles: collaboration and communication.

In addition, the use of web technology should have an impact on today's reproducibility crisis (Collins and Tabak, 2014). It has become clear in several fields of the life sciences that our current research practices are not best adapted to the production of robust and replicable results. Web technologies with their capacity to scale are key for the emergence of solutions to this crisis.

2. COLLABORATION

2.1. Data Sharing and the Web

One may remember the first attempts at data sharing in functional neuroimaging, the fMRI data center (Van Horn and Gazzaniga, 2013), and the difficulty of getting and reusing data sent over on compact discs or DVDs. Creating a culture of data sharing has many advantages: it can lead to more rapid scientific discovery for basic science and clinical research, can improve data quality, reduce costs, and improve reproducibility, and is in some cases a requirement made by funding agencies (Poline et al., 2012; Poldrack and Gorgolewski, 2014; Madan, 2017a). Some researchers argue that it is an ethical imperative (Brakewood and Poldrack, 2013) to maximize a subject's contributions, especially in clinical trials (Bauchner et al., 2016). But just because the data is shared, it doesn't mean the data can be found.

First and -possibly- foremost, browsers are the doors to the four principles of FAIR (Findable, Accessible, Interoperable, and Reusable), a set of guidelines developed by stakeholders in academia, industry, and funding agencies to promote data reuse (Wilkinson et al., 2016). We review them briefly here in the context of the web technology:

1. Findable: In order for scientists to discover data that may be of use for their research questions, datasets need to be indexed within a central database server, with appropriate metadata such that search engine algorithms can efficiently perform queries, and most importantly, with a browser-based user interface for researchers to submit queries and display results.
2. Accessible: the standard HTTP protocol used by browsers and web servers is open, free, and can provide authentication if needed.

3. Interoperable: all browsers speak the same language, regardless of their base operating system. Data description should adopt standards and convention to enable reuse across datasets, for instance through linked data technologies (Berners-Lee, 2009).
4. Reusability requires critically a community effort, to define relevant metadata and to standardize metadata reporting. This can be streamlined with web interfaces.

A key feature of the FAIR principles is that when possible they should be applicable not only to humans, but also to machines. For instance, datasets should be findable by "bots" by being tagged with the appropriate machine-readable metadata.

Neuroimaging groups have developed web portals that make it easy for other researchers to query, explore, contribute, and share both raw data and derived data. The COINS web platform (Scott et al., 2011) provides data management tools, an intuitive user interface, and was built with an emphasis for PHI security and multisite collaborations. The LORIS platform (Das et al., 2012) includes a web portal for data management and data quality control with neuroimaging viewers. The LONI Image Data Archive (LONI-IDA) is a long-term, centralized, HIPAA-compliant relational database archive for researchers to upload and share their data (Van Horn and Toga, 2009); as of this writing, the LONI-IDA has provided over 50 million downloads and over 1 million uploads to the archive. Web application such as these reduce the technical overhead to find, share, and aggregate data, and should ideally become standard practice for all large data collection efforts in neuroimaging.

The accessibility (FAIR-ness) of derived data is key to meta- and mega- analyses. A prominent example is the ENIGMA project (Thompson et al., 2014), which disseminated standardized analysis scripts to be able to co-analyze (e.g., a mega-analysis) a set of individual center's results, by sharing derived data rather than raw data. A mega-analysis strategy is especially optimal in cases where raw data sharing is not feasible. For task and resting state fMRI, the NeuroVault (Gorgolewski et al., 2015) web application enables scientists to upload fMRI statistical maps (e.g., derived data) in the standardized MNI space, and link to their publications; this platform includes both volume and surface-based visualization, and can enable more accurate meta- and mega-analyses. For diffusion imaging, the Automated Fiber Quantification (AFQ) package (Yeatman et al., 2012) has an associated web-viewer (Yeatman et al., 2018) and vault[1] to easily share derived AFQ data in a standardized format. Building software that returns derived data in standardized formats and lowers barriers to sharing these derivatives with the neuroimaging community will facilitate meta- and mega-analyses in future years.

In the past, sharing data was a technical challenge (Van Horn and Gazzaniga, 2013); now, it is easier to share data even if the data are not part of a large consortium. The OpenNeuro web application enables researchers to upload and share their neuroimaging data as long as the data follow a community-developed standard to organize and describe neuroimaging

---

[1]http://afqvault.org

datasets called the Brain Imaging Data Structure (BIDS) (Gorgolewski et al., 2016). Adopting standards for how data are stored enables sharing by reducing the overhead needed to curate heterogeneous datasets, and therefore promotes interoperability and reusability of data (Tenopir et al., 2011). Examples of standardized data formats outside of the neuroimaging field include the Open Geospatial Consortium (Castronova et al., 2013) and the Ecological Metadata Language (Fegraus et al., 2005).

In general, the FAIR principles do not stipulate how data sharing should be incentivized. The adoption of FAIR principles requires financial support as well as community adoption. While the OpenNeuro project has been funded by the NIH[2], the BIDS standard that it relies upon is, importantly, starting to be adopted by a wide community. The standard has recently been endorsed by the International Neuroinformatics Coordinating Facility (INCF)[3], and is recommended by several journals. Funding agencies (e.g., the Wellcome Trust[4]) are increasingly asking that a wider set of research products are shared with the community to increase reuse and maximize the funding impact on research. The set of tools that facilitate the conversion of small datasets to BIDS format is also growing (see the BIDS starter kit[5]), which may mitigate the need for long-term funding. Concurrently, training material to educate researchers to adopt the BIDS format is being actively developed by ReproNim (e.g the "FAIR data" module[6]).

In the genomic community, the Bermuda principles (Contreras, 2011) led to the establishment of few large public databases, but the brain imaging community has been less unified. This led to a variety of large or small initiatives, such as ADNI (Mueller et al., 2005), BIRN (Keator et al., 2008), BrainMap (Laird et al., 2011), INDI (Mennes et al., 2013), OpenfMRI (Poldrack et al., 2013), OMEGA (Niso et al., 2016), OpenNeuro (Gorgolewski et al., 2017a), Schizconnect portal(Wang et al., 2016), Healthy Brain Network (Alexander et al., 2017) to name a few [for more, see (Eickhoff et al., 2016)], and more recently the funder-based National Data Archive. Specialized tools to discover these resources and their content are improving fast [see for instance Scicrunch (Grethe et al., 2014)].

Efforts have begun in the neuroimaging community to create centralized resources to find openly released neuroimaging datasets. A very simple yet valuable collection was collaboratively compiled on the social coding platform Github[7]. OpenMorph,[8] (Madan et al., 2018), is a curated list of open access datasets that can be used to study brain morphology. It includes sample sizes, types of MRI modalities, the associated publications and a link to each project's web portal to download the data. Anyone

can contribute to this collection by creating a GitHub[9] account and editing the document. The DataLad (Halchenko et al., 2018) project has developed a crawler to index the data from various scientific data portals for a unified interface from which to download these datasets from the command line interface on their computers. DataLad also hosts a web application to interactively explore the various datasets that have been indexed. We hope to see more aggregation of open neuroimaging datasets in the future, with accessible web interfaces to query and explore all our resources.

More generally, platforms like Zenodo (https://zenodo.org), Dryad https://datadryad.org/, and the Open Science Framework https://osf.io give researchers generous storage for their datasets and assign digital object identifiers (DOIs) to datasets. This means that researchers who primarily collect data can get credit via citations, potentially alleviating concerns about "research parasites" (Longo and Drazen, 2016) that prevent some from openly sharing data. Our scientific culture is in part a roadblock to data sharing (Tenopir et al., 2011). Ideally, moving away from placing importance on only the first and last authors during grant and career reviews may incentivize data sharing and large collaborations. It is clear that technical challenges are not the only barrier to data sharing; we discuss the social and ethical challenges with data sharing in the "pitfalls" section. For an overview of the resources on data sharing, data analysis, and data collection, see **Figure 1**.

2.2. Collaborative Work and the Web
2.2.1. Collaborative Data Analysis Through the Web
Data, albeit the foundation of most work, is only the first element of a research project. The reusability of other research products such as software, libraries, scripts, and pipelines or workflows, has traditionally been poor, with the exception of a few neuroimaging software packages [e.g., SPM (Friston et al., 1994), FSL (Smith et al., 2004), and Freesurfer (Fischl, 2012)]. With a greater ease of dissemination and search of these objects, research is entering a phase of accelerated efficiency, providing building blocks for fast construction of a new analysis. Todays researcher in neuroimaging is able to search for and download an entire software environment in a Docker[10] container and launch complex pre-processings and analyses. Neurodocker (Kaczmarzyk et al., 2018) makes it possible in a single command line to create an environment with all necessary software specific version for an analysis. Reprozip (Chirigati et al., 2016) makes it possible to trace all the dependencies of a single command and create reusable packages that rerun the exact command, even on a different system. fMRIprep (Esteban et al., 2019) and MRIQC (Esteban et al., 2017) provide environments for fMRI preprocessing or MRI quality control. Work that may have taken a post-doc or a graduate student a few months can take now a few days if not a few hours. This order of magnitude acceleration factor has been made possible because (1) these projects are often highly collaborative and often will have inputs from tens

---

[2]https://www.braininitiative.nih.gov/funded-awards/openneuro-open-archive-analysis-and-sharing-brain-initiative-data

[3]https://www.incf.org/node/295

[4]https://wellcome.ac.uk/funding/guidance/guidelines-good-research-practice

[5]https://github.com/bids-standard/bids-starter-kit

[6]http://www.reproducibleimaging.org/module-FAIR-data/00-Introduction-to-Module/

[7]https://www.github.com

[8]https://github.com/cmadan/openmorph

[9]https://www.github.com

[10]https://docker.com

FIGURE 1 | Overview of discussed collaborative scientific web tools. General resources for data sharing include. (1) Zenodo https://zenodo.org; (2) Dryad: https://datadryad.org/; (3) OSF: https://osf.io; Neuroimaging specific data sharing resources include (4) COINS (Scott et al., 2011) (5) LORIS (Das et al., 2012) (6) LONI-IDA (Van Horn and Toga, 2009) (7) NeuroVault (Gorgolewski et al., 2015) For general data analysis: (8) Project Jupyter (Ragan-Kelley et al., 2014; Kluyver et al., 2016) To access cloud resources with Jupyter notebooks, try: (9) Binder https://mybinder.org/; (10) Colaboratory https://colab.research.google.com/notebook; (11) Azure Notebooks https://notebooks.azure.com; For neuroimaging specific cloud computing, see (12) OpenNeuro (Gorgolewski et al., 2017a), https://openneuro.org; (13) CBRAIN (Sherif et al., 2014) (14) BrainLife (Hayashi and Pestilli, 2017), https://brainlife.io; (15) BrainCode (Vaccarino et al., 2018), https://www.braincode.ca/; For data analysis with citizen science, see (16) Zooniverse (Simpson et al., 2014), https://zooniverse.org; and for neuroimaging-specific projects, see: (17) Brainspell (Badhwar et al., 2016), https://brainspell.org; (18) BrainBox (Heuer et al., 2016), http://brainbox.pasteur.fr; (19) Mindcontrol (Keshavan et al., 2017a), https://mindcontrol-hbn.herokuapp.com; (20) braindr (Keshavan et al., 2018), https://braindr.us; For behavioral experiments, web services such as (21) psiTurk (Gureckis et al., 2016), https://psiturk.org; (22) expfactory integrate with Amazon mTurk. (Sochat et al., 2016), https://expfactory.org.

of individuals leveraging social coding platforms (e.g., Github), and (2) the communication of the technologies and repositories through web based platforms.

Cloud computing provides unlimited, scalable, computing resources (provided enough financial resources), but can be difficult to interface with because it requires specialized expertise. Through web interfaces, cloud computing can be made accessible such that domain specific researchers can reap its full benefits. OpenNeuro (Gorgolewski et al., 2017a), currently hosted on Amazon Web Services, enables researchers to upload BIDS-compatible datasets and then run analyses via BIDS-Apps (Gorgolewski et al., 2017b) on the AWS cloud for free, given that the data is publicly shared after a certain grace period. The Canadian Brain Imaging Research Platform (CBRAIN) web platform (Sherif et al., 2014) can bring together heterogeneous data sources and compute grids into one, secure web interface. The BrainLife[11] (Hayashi and Pestilli, 2017) web application is in development to provide

researchers with an intuitive interface to cloud computing resources, enable data sharing, and the publishing of results with clear provenance. The Brain-Code (Vaccarino et al., 2018)[12] web portal and data management/analysis platform aims to foster collaboration and data discovery across various clinical brain disorders.

The Jupyter project (Ragan-Kelley et al., 2014; Kluyver et al., 2016) has been actively developing a web-based scientific notebook interface for various programming languages (Julia, Python, R, and more). Researchers can interact with various programming kernels on a web interface that can be deployed locally, or on the cloud. The resulting notebook can be shared as a website, with not only code displayed but also the resulting figures, and associated documentation that is formatted in Markdown, which can also render equations. The Jupyter notebook comes with the ability to write interactive widgets, such as javascript-based sliders that let users explore various parameter spaces of the functions they write. Interactive plotting libraries,

---

[11]https://brainlife.io

[12]https://www.braincode.ca/

like Plotly[13] can be integrated within the Jupyter notebook, enabling researchers to create rich, interactive data visualizations. The Binder[14] project, as of this writing, provides a free service to host instances of Jupyter notebooks on the cloud. Azure notebooks[15] and Google Colaboratory[16] also provide similar notebook hosting services. Currently, Colaboratory provides access to GPUs instances, which are incredibly useful for deep learning projects. Services that enable easy deployment of notebooks and their associated computing environments will vastly improve the transportability of research objects; we therefore encourage neuroinformatics researchers to take advantage of these web services.

2.2.2. Collaborative Writing on the Web

In the past, collaboratively preparing manuscripts might only have been possible with those in a scientist's immediate vicinity. With the web browser, email drastically improved the collaborative writing process, but it is still a slow, serial process of emailing documents back and forth. Google Docs[17] was a breakthrough web application that parallelized the manuscript preparation process by enabling multiple authors to simultaneously write, edit, comment, and even chat with each other. Version control, tracking changes, and generous free cloud storage means researchers are much less likely to lose their work. Microsoft Word, the most widely used software for preparing manuscripts, offers an "edit in the browser" feature for realtime collaborative editing[18]. For reference management, Paperpile[19] interfaces nicely with Google Docs. For those who prefer to prepare manuscripts with LaTeX, services such as Overleaf[20] and Authorea[21] compile latex on the cloud, removing the technical overhead of setting up latex locally and compiling the document. Collaborators who are less familiar with LaTeX can now easily contribute to these manuscripts. See **Table 1** for a summary of collaborative writing web applications.

GitHub[22], "the social coding platform", has simplified and improved the collaborative writing of software. Github provides a visual representation of the somewhat complicated git version control system. GitHub repositories contain the full codebase for a project, all the changes that have been made, and who made them (via git). Users can "Fork" GitHub repositories, which makes a copy of the code to their account. They can then make changes to the code and send the changes back to the original repository via "Pull Requests," which begins a discussion thread for others to comment on the code (called a code review). GitHub also provides an "Issues" page for each repository, where users can discuss any issues and ask the community for help. Continuous integration software testing can be automatically run

TABLE 1 | Summary of collaborative tools for writing manuscripts and code on the web.

Name	URL	Comment
Google Drive	https://drive.google.com	Write manuscripts, spreadsheets, etc
Office 365	https://office.com	Write manuscripts MS Word online.
Paperpile	https://paperpile.com	Reference manager for google docs
Overleaf	https://overleaf.com	Write manuscripts (LaTEX)
Authorea	https://authorea.com	Write manuscripts (LaTEX, HTML)
GitHub	https://github.com	Write code
GitLab	https://gitlab.com	Write code
Travis-CI	https://travis-ci.com	Test code (links to GitHub/Lab)
Circle-CI	https://circleci.com	Test code (links to GitHub/Lab)

on the cloud once changes to the code are pushed to GitHub, by web-hooks to services like Travis CI[23] and Circle CI[24], which provide a generous free tier for open source projects. GitHub repositories can also host static websites; this is extremely useful for hosting code documentation. GitLab[25] is an open source alternative to GitHub, which can be deployed by researchers in cases where they need a private git web application. Many open source neuroimaging tools are built collaboratively on GitHub, such as Nipype[26] (Gorgolewski et al., 2011) , Dipy[27] (Garyfallidis et al., 2014), and Nilearn[28] (Abraham et al., 2014), to name a few. By developing open source neuroimaging software packages on social coding web interfaces, researchers are able to engage a much larger community of contributors than would have been possible in the earlier days of the web.

2.2.3. The Web for Mass Collaboration: Citizen Science and Crowdsourcing

The web browser is particularly well suited for citizen science and crowdsourcing; this is becoming necessary as neuroimaging datasets grow, and data analysis bottlenecks arise when massive amounts of data need visual inspection. In the astronomy community, the Galaxy Zoo (Lintott et al., 2008) web application was successful at engaging citizen scientists in visually classifying galaxies. This project evolved into a more general citizen science platform called the Zooniverse (Simpson et al., 2014), which enables researchers from any domain to engage citizen scientists in annotating their data. In the neuroscience field, EyeWire (Kim et al., 2014) and Mozak (Roskams and Popović, 2016) have

---

[13]www.plot.ly
[14]https://mybinder.org/
[15]https://notebooks.azure.com
[16]https://colab.research.google.com/notebook
[17]https://drive.google.com
[18]https://support.office.com/en-us/article/collaborate-on-word-documents-with-real-time-co-authoring-7dd3040c-3f30-4fdd-bab0-8586492a1f1d
[19]https://paperpile.com/
[20]https://www.overleaf.com
[21]https://www.authorea.com
[22]https://www.github.com

[23]https://travis-ci.org
[24]https://www.circleci.com
[25]https://gitlab.com
[26]https://www.github.com/nipy/nipype
[27]https://www.github.com/nipy/dipy
[28]https://www.github.com/nilearn/nilearn

gamified the tracing of neurons. The EyeWire project was able to engaged over 100,000 citizen scientists from all over the world to collaboratively trace the neurons of the human retina. Such a massive engagement of collaborators would not have been possible without the web browser.

The neuroimaging community is just beginning to engage citizen scientists as a resource in our data analyses. The Brainspell (Badhwar et al., 2016) web application was developed to manually annotate fMRI coordinate tables that were automatically extracted by Neurosynth https://neurosynth.org (Yarkoni et al., 2011), which itself is a web application to perform coordinate-based fMRI meta-analyses. BrainBox (Heuer et al., 2016) and Mindcontrol (Keshavan et al., 2017a) are web applications to annotate MRI volumes (e.g., to edit segmentations). Recently, a mobile-optimized and gamified web application called braindr (Keshavan et al., 2018) was developed to perform quality control on images from the Healthy Brain Network initiative. At the time of this writing, braindr has engaged over 400 citizen scientists and over 100,000 annotations. Image labels were aggregated by weighting citizen scientists based on how well their ratings matched an expertly labeled "gold standard" subset of images. A deep learning network was then trained from these aggregated labels to automatically rate image quality to near perfect accuracy. Hybrid human-computer approaches for quality control seem the most promising (Esteban et al., 2018), such as "triaging" image reviews based on machine-learning output probability scores for Freesurfer image segmentation as in Klapwijk et al. (2018). Whether citizen science applications can go beyond quality control and toward more complex tasks like image segmentation and registration remains to be explored.

The cognitive science and psychology communities often utilize paid crowdsourcing web platforms like Amazon Mechanical Turk (mTurk) to run behavioral experiments with large, diverse populations. The psiTurk (Gureckis et al., 2016) and ExpFactory (Sochat et al., 2016) frameworks enable scientists to interface with mTurk and create reusable web-based psychology experiments. For image processing, the quanti.us (Hughes et al., 2018) platform can be used to interface with mTurk to crowdsource the segmentation of biological images. In neuroimaging, Ganz et al. (2017) showed it was feasible to crowdsource the detection of Freesurfer (Fischl, 2012) cortical surface delineation errors on mTurk. We expect to see more utilization of citizen science, gamification, and paid crowdsourcing platforms in neuroimaging research, and there are still many open questions about which strategies (citizen science vs. paid crowdsourcing) and task designs are better suited for various analyses, as well as how to properly acknowledge the contributions of citizen scientists [see (Hunter and Hsu, 2015) for a proposed method].

2.3. Pitfalls

Even though the benefits of the web browser for scientific collaborations are evident, using the web for our research comes with some drawbacks or difficulties. Collaboration requires the sharing of data, and while some argue that data sharing is an ethical imperative (Brakewood and Poldrack, 2013; Bauchner et al., 2016), one must consider the risks of reidentification of our

subjects, particularly for clinical research. True deidentification is difficult because of linked metadata (Narayanan and Shmatikov, 2008; de Montjoye et al., 2015). For example, in Narayanan and Shmatikov (2008), researchers identified pseudo-anonymized Netflix users by linking data with metadata from another website (IMDB). In de Montjoye et al. (2015), researchers proved that pseudo-anonymized credit card data could be reidentified provided just four spatiotemporal points. Research in differential privacy (Sarwate et al., 2014) might alleviate some of these risks; regardless, it is important that subjects are made aware of the risk in the consent process. The Open Brain Consent website[29] is a collaborative effort to provide resources that aid researchers in the IRB process for sharing data, writing the consent form, and tools for the anonymization of neuroimaging data.

Legal obligations concerning personal data handling are evolving and the recent European Union General Data Protection Regulation (GDPR) will likely change the requirements for participants control over their personal data. This will need to be considered at all stages of the research data lifecycle. While a full discussion on the legal and ethical aspects of data dissemination and reuse is out of the scope of this article [see for example (Marelli and Testa, 2018) on the GDPR] it is clear that legal and regulatory constraints are going to shape the implementation and use of web based data dissemination and retrieval tools, and this will require increased attention and human resources in the future. The challenge will be to constantly adapt our infrastructures and practices to the new regulations, which will require continuous software development.

Another drawback of using web technology for collaboration, in terms of sharing data, accessing cloud resources for analysis, and distributing work, is bit rot (Baker et al., 2006; Cerf, 2011). Bit rot refers to the eventual degradation of information stored on electronic media; for example, information stored on floppy disks is likely not accessible for most of us. Web technology is advancing rapidly: the browsers we use now look nothing like they did a decade ago. Some websites that were built in the past do not work with modern browser technology, and most websites from the past are not available to us anymore. A decade from now, many of the links presented in this article may no longer exist. Servers cost money, and domain names are charged annually. Software needs to be consistently maintained to be compatible with current technology. Efforts such as the Internet Archive[30] and Digital Object Identifier (DOI) system are working to preserve the information on the web, and in the case of DOI, provide persistent links to our research articles. But we need to work with funding agencies to ensure we have the resources to maintain scientific output, outside of our research articles, that depend on web technology. We also need to work with publishers to ensure our full scientific output, including the web technology that is used to produce it, can be fully preserved.

Finally, web-based research depends on a stable and fast internet connection. Such infrastructure may not be available to scientists in developing countries, which further drives

---

[29]https://open-brain-consent.readthedocs.io/en/stable/
[30]https://archive.org/

inequalities and will decrease the diversity of our scientific community. It is important to keep this in mind when designing web applications, by optimizing websites for slow internet connections, and building offline support.

3. COMMUNICATION

Scientific work in the public imagination is still often thought to be a rather solitary activity of independent individuals, sometimes attracting introverted personalities. But actual scientific work is largely communication, where a large proportion of time is spent thinking of the best way to communicate research to collaborators, to scientific communities, to the public, and to funders. Different scientific fields have different levels of interdependencies. A researcher in a specialized mathematical subfield like non Riemannian geometry could be mostly working on their own, but fields like neuroscience or the biomedical sciences are highly multidisciplinary. The ability to absorb and reuse research from other laboratories is most often critical for progress, as the systems studied are both too complex and too interdependent to be understood by individuals or single labs. While conferences and in-person meetings are traditional methods for communicating research, the web now expands scientific communication to a completely new level, by removing time delays and scalability constraints. Now, even social network communication tools are used for the benefit of scientific communication.

3.1. Local Networks Communication

The small or medium size laboratory structure [5–15 people (Conti and Liu, 2015; Cook et al., 2015)] is still the predominant basic research structure in universities and research institutes, and these are mostly set up such that in person meetings are practical. Nevertheless, it is common that one or several members of the laboratory are temporarily located in another institution or building and the meeting will occur through web video conferences. The number of companies proposing free or paid services that may include capacity to share documents has multiplied during the past few years (the authors count at least 7 web video conference systems as of today, for instance Zoom[31], Webex[32], BlueJeans[33], Skype[34], Google Hangouts[35], appear.in[36], GoToMeeting[37], etc, as well as project management systems such as Trello[38] or Asana[39]), allowing for unprecedented efficiency even in local communication. A key aspect of some of these communication tools is their capacity to record the meeting (audio-video) permitting delayed communication and traceability of discussion points, ideas or decisions, as well

as scaling for larger groups. Another key aspect is that the use of these tools allow a group to immediately scale to non local members.

3.2. Scholarly Communication

A neuroimaging or neuroscience researcher's work is heavily influenced, if not directed by, the search for funding and progression in academia career. As these mostly still depend on the quality and number of publications, it is clear that publishing activity is central to a researcher's academic life.

The current publishing industry is still very much influenced by how this activity used to be at the turn of the twentieth century, at a time when manuscripts had to be manually typed and printed, and distribution of journals was achieved through mail. Today, the article remains a standard for scholarly communication, even though an increasing number of researchers realize that the actual scholarship may actually reside in the code and data rather than the article[40]. Jon Claerbout, a professor from Stanford University, argues that an article about a computational result is advertising, rather than scholarship. The actual scholarship is the full software environment, code and data, that produced the result (Donoho, 2010). The web has transformed the industry and is *de facto* the new media for scholarly communication, but somehow less rapidly and less radically than it could have. Most traditional journals are still shipping some printed copies of their editions, while a very large number of "on-line only" journals with an open access policy have emerged with a business model based on article processing charges (ACP), occasionally generating low quality content, but a highly profitable business (for a long list of questionable publishers, see the Beall's List[41]). We note that Beall's list does not necessarily have the level of granularity required as it can address general publishers rather than specific journals.

Even when the web is adopted as the communication media, the very large majority of the articles are based on HTML and PDF, with almost none of the modern visualization and interactive figure components that can be delivered by modern JavaScript libraries (e.g., D3.js[42]). In neuroimaging, a number of open source, browser-based visualization tools have been developed. Javascript brain viewers like BrainBrowser (Sherif et al., 2015), papaya.js[43], XTK.js[44] (Haehn et al., 2014), and AMI library (Bernal-Rusiel et al., 2017) enable researchers visualize neuroimaging data in the browser. Interactive, linked data dashboards have been built as outputs of neuroimaging software, like ROYGBIV[45] (Keshavan et al., 2017b; Klein et al., 2017), AFQ-Browser[46] (Yeatman et al., 2018), and MRIQC has a web-based viewer to visually inspect outputs (Esteban et al.,

---

[31]https://zoom.us/
[32]https://www.webex.com/
[33]https://www.bluejeans.com/
[34]https://www.skype.com/en/
[35]https://hangouts.google.com/
[36]https://appear.in
[37]https://www.gotomeeting.com/
[38]https://trello.com
[39]https://asana.com/

[40]https://www.researchtrends.com/issue-31-november-2012/force11-gains-momentum-creating-the-future-of-research-communications-and-e-scholarship/
[41]https://beallslist.weebly.com/standalone-journals.html
[42]https://d3js.org/
[43]https://github.com/rii-mango/papaya
[44]https://github.com/xtk/X
[45]http://roygbiv.mindboggle.info
[46]https://yeatmanlab.github.io/AFQBrowser-demo

2017). The Open Anatomy Browser[47] (Halle et al., 2017) hosts a variety of atlases with collaborative viewing. These tools have greatly simplified the process of building and sharing complex, interactive visualizations. For example, researchers may deploy an AFQ-Browser visualization of their data with two simple commands (afqbrowser-assemble, afqbrowser-publish). These interactive figures may go much beyond the convenience of a better view of the result; they allow to test for the potential robustness or sensitivity with data input or methods in a way that cannot be provided by static figures. In such a case, some parts of the scholarship need to be communicated by interactive figures, but few publishers are able to provide the infrastructure for hosting such "interactive articles".

Recently, the rise of documents able to mix code and narrative such as R-markdown or Jupyter notebooks also provide researchers with new opportunities for communicating full fledged research objects. Some publishers already have embraced these new possibilities. For instance, eLife is working with Stencila[48], designed to be documents that "... are self-contained, interactive and reusable, containing all the text, media, code and data needed to fully support the narrative of research discovery" to foster more reproducible and reusable research, see eLife. In the near future, systems such as Binder (Jupyter et al., 2018) will allow not only to publish and review the computational documents but also provide with a container and the environment for a fully re-executable publication. The new web tools are not only key to provide us with ways of publishing a more complete set of research objects, they also allow for new review workflows to be implemented. For instance, Frontiers developed a platform that intended to make the interaction between reviewers and authors more efficient. Tools such as https://web.hypothes.is/[49] permit readers to annotate only specific parts of an article and may in the future be re-used by a review system. Such a review system could associate expert reviews and open community based reviews.

The web is also transforming how research communities meet for discussions by creating virtual conferences. A number of virtual conferences have been successfully organized in the past, removing the constraints of space and travels, while still allowing for questions and answer sessions monitored online (see for instance neuroscience-2018[50]). A recent twitter conference was recently organized (the Brain Twitter Conference) which could scale easily to tens of thousands of participants. These events are much easier to organize in a short time and less costly if not free for attendees. They also are possible to attend by all researchers independently of possible travel and funding restrictions and are only limited by time zone constraints. For example, Chris Madan advocates using Twitter for science in (Madan, 2017b), and see his associated blog post[51] on this topic.

In the same spirit, global Brainhack events gather locally groups of neuroinformaticians who collaborate on software development projects, and are also sharing courses and seminars across locations. The latest Brainhack[52] event took place in 16 countries and gathered more than 1000 participants in 5 different time zones. The University of Washington hosts various week-long summer schools or "hack weeks" (e.g., Astrohackweek, Geohackweek, and Neurohackweek) to promote education and training, tool development, community building, and interdisciplinary research by combining pedagogy with project-based learning (the "hacking") around a specific domain (Huppenkothen et al., 2018). They found that this combination is particularly effective at fostering collaborations and promoting best practices. Through collaborative web applications like GitHub, the projects started at these hackweeks have continued even after the events ended, and have resulted in measurable scientific output [for details, see Huppenkothen et al. (2018)]. Data analysis challenges hosted by conferences or symposia like MICCAI[53] bring researchers together to solve problems in the field, even if they cannot be present at the conference, and these groups collaboratively publish their results [for example see Commowick et al. (2018) for the results of a multiple sclerosis lesion segmentation challenge]. A curated list of biomedical image challenges can be found at https://grand-challenge.org/challenges/. We expect these types of events to be more frequent in the future, limiting the ecological, time, and cost impact of physical travels but offering the capacity for communication of research at a truly global scale.

3.3. Larger Public Communication

Ultimately, research needs to go beyond the scientists and will need to be disseminated to the larger public which, through taxes, is funding a large part of it (Illes et al., 2010). The field of neuroimaging necessitates costly acquisition devices (MRI, PET, E/MEG), and has been particularly well funded, not only because of its potential for neuroscience, but also because the ideas were communicated well to the public and to funders. Communication is now largely operated and achieved by social media platforms such as Twitter, LinkedIn, Facebook, and blog platforms, to name a few. To read more about the advantages and disadvantages of social media use for scientific communication, see (Bik and Goldstein, 2013). Online resources that teach how to effectively communicate science are provided by the Alda-Kavli Center for Communicating Science[54]. To consolidate the current consensus of knowledge, Wikipedia is probably the best resource; offering an introduction to functional magnetic resonance imaging through the consensus writing of many researchers (for example, see the Wikipedia article for fMRI).

Last, but certainly not least, web based education platforms are also re-inventing how training is performed in neuroimaging. The standard in person courses are now often replaced by on-line material (see ReproNim[55], Coursera online courses[56],

---

[47]https://www.openanatomy.org/
[48]https://stenci.la/
[49]https://web.hypothes.is/
[50]https://www.labroots.com/virtual-event/neuroscience-2018
[51]https://medium.com/@cMadan/on-the-benefits-of-twitter-5af59158e4e2

[52]http://www.brainhack.org/global2018/
[53]https://www.miccai.org/
[54]https://www.aldacenter.org/aklc
[55]www.repronim.org
[56]https://www.coursera.org/learn/functional-mri

EdX online courses[57]), and a series of YouTube videos by Jeanette Mumford[58]) and also Dirk Ostwald[59] as examples, amongst many good online materials. This allows laboratories to give some inverted classroom type of training by considering that formal lectures can be taken on-line but exercises or projects are best solved or supervised with direct interactions. It should be noted that on-line question and answer forums such as Neurostars[60] (with a tagging system similar to stackoverflow[61]), NITRC, and software tool email lists, are also key for the training of young researchers and boost efficiency. For a review of scientific web communcation tools, see **Figure 2**.

3.4. Pitfalls

There are both limits and dangers associated with relying too much - possibly almost fully - on web technologies and browser enabled applications and workflows for research. Web communication does not necessarily allow the level of in depth interactions that are required to discuss a specific research question. In person meetings can be necessary both to organize projects and to advance the understanding of our scientific questions. In our experience, in person meetings are better at providing decision structures and at building trust, which are both necessary for the management of scientific projects.

Some of the dangers associated with the use of social media could also propagate to the scientific arena. For instance, while social media may be a great medium for quickly accessing or publicizing articles, it may also focus the attention on a specific cluster of the scientific community. This in turn may create research networks that are less permeable to different ideas, like a scientific echo chamber (Kim et al., 2017).

The immediate access to non -or poorly- peer-reviewed works may also amplify incorrect results that would not stand scrutiny under peer review. Take for example, a paper posted on the preprint server arXiv called "Automated Inference on Criminality using Face Images," which received a lot of criticism from the scientific community[62]. Even though it was not peer reviewed, and as of this writing has not been published in a peer-reviewed journal, it nevertheless received a lot of alarmist press coverage. This can occur within the traditional literature, albeit at a slower pace. The neuroscience and public health communities are still contending with the spread of misinformation regarding a link between vaccines and autism (Del Vicario et al., 2016), despite the strong evidence to the contrary (Taylor et al., 2014). Scientific communication is our responsibility as scientists, not only to the scientific community but to the general public; we must be cautious of the immediacy of the web.

---

[57]https://www.edx.org/course/fundamentals-biomedical-imaging-magnetic-epflx-fndbioimg2x-0
[58]https://www.youtube.com/channel/UCZ7gF0zm35FwrFpDND6DWeA
[59]https://www.youtube.com/channel/UCQ8y5WCi5yAgDFxLmh2MJyg/videos
[60]https://neurostars.org
[61]https://stackoverflow.com
[62]https://arxiv.org/abs/1611.04135

4. CONCLUSION

The way web technologies - and the browser as the window to these - are transforming scientific activity is still evolving. It is clear that an important part of research work will be on-line for the future PhD student, whether to acquire or disseminate knowledge, conduct an experiment, and collaborate with experts. This paradigm shift is already apparent with the advent of e-conferences and the use of social media in the neuroscience community. Some researchers now mostly rely on their Twitter feeds to learn about new and interesting studies, delivering more directed and rapid content than a traditional journal's table of contents. The browser brings the potential for massive online collaboration and more effective communication, but the web is still mostly an untapped resource in the neuroimaging and neuroscience fields.

Some scholars argue that we are having a reproducibility crisis. Many neuroimaging studies are found underpowered, and have reported possibly inflated effect sizes and unstable results (Yarkoni, 2009). We believe the browser can help, by connecting users to large, documented, and shared datasets through web portals, and by providing interfaces to upload, annotate, share and publish raw and derived data. This would result in a much broader pool of data that could be investigated and lead to more stable results, such as those from meta- or a more distributed, ENIGMA-style mega- analyses. These efforts should complete the FAIR principles, moving toward "Interoperable and Reusable" data, with community-defined documentation and metadata standards.

Replicating a study is complex because computing environments are difficult to transport to other systems. Works produced with tools that are not easily transportable to the web will be harder to communicate, and potentially less reproducible or re-usable by others. The analysis of a dataset performed on a local computer and producing figures as files on a local disk will need to consider all the hurdles of local storage, computational environments, and other technological challenges, to create robust software tools that work on all computational environments. Difficult installations limit the capacity to rapidly reuse the results. The web browser can help: the same analysis developed through a Jupyter notebook interface and running on the Binder service will be re-usable at no cost of transfer on either the producer and the receiver side. Considering the cost for an individual or lab to reproduce an analysis, collaborate on it, or re-use a component of it, should be a key question when working on a research project. In many cases, web technologies are the ideal solution.

We are experiencing a data deluge. As neuroimaging studies accumulate larger datasets, we encounter many new challenges in data analysis that we did not have with smaller datasets, both in terms of our capacity to consolidate datasets originating from various cohorts acquired on different scanners, and in terms of the sheer computational power needed to process very large datasets. Browsers, by interfacing with cloud computing infrastructures, can provide us access to an almost infinite resource of compute power. Data analyses that require visual inspection are unfeasible to scale; the browser provides the
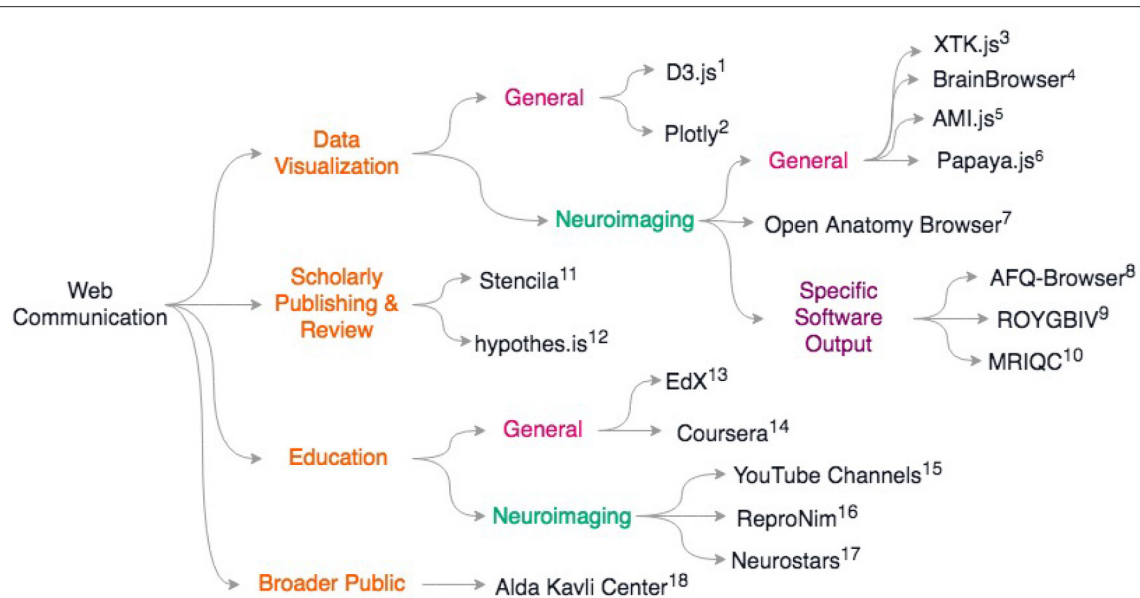
FIGURE 2 | Overview of discussed scientific communication web resources. General resources for data sharing include (1) D3.js https://d3js.org; 2) Plotly: https://plotly.com/; General neuroimaging data visualization libraries include
3) XTK (Haehn et al., 2014) 4) BrainBrowser (Sherif et al., 2015) 5) AMI.js (Bernal-Rusiel et al., 2017) 6) papaya.js https://github.com/rii-mango/papaya; 7) Open Anatomy Browser (Halle et al., 2017) Some neuroimaging packages that release associated web-viewers: 8) AFQ-Browser (Yeatman et al., 2018) 9) ROYGBIV/Mindboggle (Keshavan et al., 2017b; Klein et al., 2017) 10) MRIQC (Esteban et al., 2017) For scholarly publishing and review: 11) Stencila https://stenci.la; 12) hypothes.is https://hypothes.is; In education: 13) EdX https://www.edx.org/; 14) Coursera https://www.coursera.org/; For neuroimaging-specific courses and resources: 15) YouTube channels of Dr. Jeanette Mumford and Dr. Dirk Ostwald 16) ReproNim training modules http://www.reproducibleimaging.org/; 17) Neurostars forum https://neurostars.org; Web resources for learning how to communicate to the general public: 18) Alda-Kavli Learning Center online resources https://www.aldacenter.org/AKLC

medium to collaborate with not only other experts, but also citizen scientists. Communicating insights from high-dimensional datasets is challenging, but the browser can host interactive data visualizations that can be easily shared. As a community we need to move toward developing browser-based tools to efficiently gain insights from large neuroimaging datasets.

The browser was built under egalitarian principles of free and open information exchange[63], but scientific information is not completely free. Today, traditional scientific publishers are making unusually high profit margins and a large body of the literature is behind paywalls (Buranyi, 2017). This prevents text-mining and creates an unnecessary bottleneck to much needed meta-analyses (Van Noorden, 2012). In addition, research has become highly competitive [e.g., the famous adage, "publish or perish" (De Rond and Miller, 2005)]. Some of this competition is an impediment to the collaborative nature of research, and the community as a whole could work much more efficiently and reduce research cost if free and open principles were extended

as much as possible (respecting ethical and legal constraints). In order to advance more *efficiently* our understanding of the brain, we need to shift our scientific culture away from silos of domain expertise to a more collaborative, distributed network of information exchange; a shift from the wet lab to the web lab.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

---

[63]https://webfoundation.org/about/vision/history-of-the-web/

REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* 8:14. doi: 10.3389/fninf.2014.00014

Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., et al. (2017). An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* 4:170181. doi: 10.1038/sdata.2017.181

Badhwar, A., Kennedy, D., Poline, J., and Toro, R. (2016). Distributed collaboration: the case for the enhancement of brainspell's

interface. *GigaScience*. 5(Suppl. 1), 46. doi: 10.1186/s13742-016-0147-0

Baker, M., Shah, M., Rosenthal, D. S., Roussopoulos, M., Maniatis, P., Giuli, T. J., and Bungale, P. (2006). "A fresh look at the reliability of long-term digital storage," in *ACM SIGOPS Operating Systems Review* (New York, NY: ACM), 221–234.

Bauchner, H., Golub, R. M., and Fontanarosa, P. B. (2016). Data sharing: an ethical and scientific imperative. *JAMA* 315, 1238–1240. doi: 10.1001/jama.2016.2420

Bernal-Rusiel, J. L., Rannou, N., Gollub, R. L., Pieper, S., Murphy, S., Robertson, R., et al. (2017). Reusable client-side javascript modules for immersive web-based real-time collaborative neuroimage visualization. *Front. Neuroinform.* 11:32. doi: 10.3389/fninf.2017.00032

Berners-Lee, T. (2009). Linked Data - Design Issues. Available online at: https://www.w3.org/DesignIssues/LinkedData.html

Bik, H. M. and Goldstein, M. C. (2013). An introduction to social media for scientists. *PLoS Biol.* 11:e1001535. doi: 10.1371/journal.pbio.1001535

Brakewood, B. and Poldrack, R. A. (2013). The ethics of secondary data analysis: considering the application of belmont principles to the sharing of neuroimaging data. *Neuroimage* 82, 671–676. doi: 10.1016/j.neuroimage.2013.02.040

Buranyi, S. (2017). Is the staggeringly profitable business of scientific publishing bad for science. *The Guardian* 27:2017.

Castronova, A. M., Goodall, J. L., and Elag, M. M. (2013). Models as web services using the open geospatial consortium (ogc) web processing service (wps) standard. *Environ. Model. Soft.* 41, 72–83. doi: 10.1016/j.envsoft.2012.11.010

Cerf, V. G. (2011). Avoiding "bit rot": Long-term preservation of digital information [point of view]. *Proc. IEEE* 99, 915–916. doi: 10.1109/JPROC.2011.2124190

Chirigati, F., Rampin, R., Shasha, D., and Freire, J. (2016). "Reprozip: computational reproducibility with ease. in *Proceedings of the 2016 International Conference on Management of Data* (San Francisco, CA: ACM), 2085–2088.

Collins, F. S. and Tabak, L. (2014). NIH plans to enhance reproducibility. *Nature* 505, 612–613.

Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., et al. (2018). Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Sci. Rep.* 8:13650. doi: 10.1038/s41598-018-31911-7

Conti, A. and Liu, C. C. (2015). Bringing the lab back in: personnel composition and scientific output at the mit department of biology. *Res. Policy* 44, 1633–1644. doi: 10.1016/j.respol.2015.01.001

Contreras, J. L. (2011). Bermuda's legacy: policy, patents, and the design of the genome commons. *Minn. JL Sci. Tech.* 12:61.

Cook, I., Grange, S., and Eyre-Walker, A. (2015). Research groups: How big should they be? *PeerJ* 3:e989. doi: 10.7717/peerj.989

Das, S., Zijdenbos, A. P., Vins, D., Harlap, J., and Evans, A. C. (2012). Loris: a web-based data management system for multi-center studies. *Front. Neuroinform.* 5:37. doi: 10.3389/fninf.2011.00037

de Montjoye, Y. A., Radaelli, L., Singh, V. K., et al. (2015). Unique in the shopping mall: on the reidentifiability of credit card metadata. *Science* 347, 536–539.

De Rond, M. and Miller, A. N. (2005). Publish or perish: bane or boon of academic life? *J. Manage. Inquiry* 14, 321–329. doi: 10.1177/1056492605276850

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., et al. (2016). The spreading of misinformation online. *Proc. Nat. Acad. Sci. U.S.A.* 113, 554–559. doi: 10.1073/pnas.1517441113

Donoho, D. L. (2010). An invitation to reproducible computational research. *Biostatistics* 11, 385–388. doi: 10.1093/biostatistics/kxq028

Eickhoff, S., Nichols, T. E., Van Horn, J. D., and Turner, J. A. (2016). Sharing the wealth: neuroimaging data repositories. *Neuroimage* 124(Pt B):1065–1065. doi: 10.1016/j.neuroimage.2015.10.079

Esteban, O., Birman, D., Schaer, M., Koyejo, O. O., Poldrack, R. A., and Gorgolewski, K. J. (2017). Mriqc: advancing the automatic prediction of image quality in mri from unseen sites. *PLoS ONE* 12:e0184661. doi: 10.1371/journal.pone.0184661

Esteban, O., Blair, R. W., Nielson, D. M., Varada, J. C., Marrett, S., Thomas, A., et al. (2018). Crowdsourced mri quality metrics and expert quality annotations for training of humans and machines. *bioRxiv [Preprint]*. doi: 10.1101/420984

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Ilkay Isik, A., Erramuzpe, A., et al. (2019). FMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* 16, 111–116. doi: 10.1038/s41592-018-0235-4

Fegraus, E. H., Andelman, S., Jones, M. B., and Schildhauer, M. (2005). Maximizing the value of ecological data with structured metadata: an introduction to ecological metadata language (eml) and principles for metadata creation. *Bull. Ecol. Soc. Am.* 86, 158–168. doi: 10.1890/0012-9623(2005)86<158:MTVOED>2.0.CO;2

Fischl, B. (2012). Freesurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., and Frackowiak, R. S. J. (1994). Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Hum. Brain Mapp.* 2, 189–210. doi: 10.1002/hbm.460020402

Ganz, M., Kondermann, D., Andrulis, J., Knudsen, G. M., and Maier-Hein, L. (2017). Crowdsourcing for error detection in cortical surface delineations. *Int. J. Comp. Assis. Rad. Surg.* 12, 161–166. doi: 10.1007/s11548-016-1445-9

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., et al. (2014). Dipy, a library for the analysis of diffusion mri data. *Front. Neuroinform.* 8:8. doi: 10.3389/fninf.2014.00008

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* 5:13. doi: 10.3389/fninf.2011.00013

Gorgolewski, K., Esteban, O., Schaefer, G., Wandell, B., and Poldrack, R. (2017a). *Openneuroa Free Online Platform for Sharing and Analysis of Neuroimaging Data*. Vancouver, BC: Organization for Human Brain Mapping, 1677.

Gorgolewski, K. J., Alfaro-Almagro, F., Auer, T., Bellec, P., Capotă, M., Chakravarty, M. M., Churchill, N. W., et al. (2017b). Bids apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS Comput. Biol* 13:e1005209. doi: 10.1371/journal.pcbi.1005209

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3:160044. doi: 10.1038/sdata.2016.44

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., et al. (2015). Neurovault. org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Front. Neuroinform.* 9:8. doi: 10.3389/fninf.2015.00008

Grethe, J., Bandrowski, A., Banks, D., Condit, C., Gupta, A., Larson, S., et al. (2014). "Scicrunch: a cooperative and collaborative data and resource discovery platform for scientific communities," In *Front. Neuroinform. Conference Abstract: Neuroinformatics* (Leiden). Available online at: https://www.frontiersin.org/10.3389/conf.fninf.2014.18.00069/event_abstract

Gureckis, T. M., Martin, J., McDonnell, J., Rich, A. S., Markant, D., Coenen, A., et al. (2016). psiturk: An open-source framework for conducting replicable behavioral experiments online. *Behav. Res. Methods* 48, 829–842. doi: 10.3758/s13428-015-0642-8

Haehn, D., Rannou, N., Ahtam, B., Grant, E., and Pienaar, R. (2014). Neuroimaging in the browser using the x toolkit. *Front. Neuroinform. Conference Abstract: 5th INCF Congress of Neuroinformatics* 101. doi: 10.3389/conf.fninf.2014.08.00101

Halchenko, Y., Hanke, M., Poldrack, B., Meyer, K., Debanjum, Alteva, G., et al. (2018). datalad/datalad 0.10.0. Available online at: https://zenodo.org/record/1470783#.XFRM2cZOkng

Halle, M., Demeusy, V., and Kikinis, R. (2017). The open anatomy browser: A collaborative web-based viewer for interoperable anatomy atlases. *Front. Neuroinform.* 11:22. doi: 10.3389/fninf.2017.00022

Hayashi, S. and Pestilli, F. (2017). Reproducible neuroimaging via open cloud services: data upcycling to advance discovery in network neuroscience. Available online at: https://psyarxiv.com/y82n7/

Heuer, K., Ghosh, S., Sterling, A. R., and Toro, R. (2016). Open neuroimaging laboratory. *Res. Ideas Outcomes* 2:e9113. doi: 10.3897/rio.2.e9113

Hughes, A. J., Mornin, J. D., Biswas, S. K., Beck, L. E., Bauer, D. P., Raj, A., et al. (2018). Quanti. us: a tool for rapid, flexible, crowd-based annotation of images. *Nat. Methods* 15, 587–590. doi: 10.1038/s41592-018-0069-0

Hunter, J. and Hsu, C.-H. (2015). "Formal acknowledgement of citizen scientists' contributions via dynamic data citations," in *International Conference on Asian*

Digital Libraries (Springer), 64–75. Available online at: https://link.springer.com/chapter/10.1007/978-3-319-27974-9_7#citeas

Huppenkothen, D., Arendt, A., Hogg, D. W., Ram, K., VanderPlas, J. T., and Rokem, A. (2018). Hack weeks as a model for data science education and collaboration. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8872–8877. doi: 10.1073/pnas.1717196115

Illes, J., Moser, M. A., McCormick, J. B., Racine, E., Blakeslee, S., Caplan, A., et al. (2010). Neurotalk: improving the communication of neuroscience research. *Nat. Rev. Neurosci.* 11, 61–69. doi: 10.1038/nrn2773

Jupyter, P., Bussonnier, M., Forde, J., Freeman, J., Granger, B., Head, T., et al. (2018). Binder 2.0 - Reproducible, interactive, sharable environments for science at scale. in *Proceedings of the 17th Python in Science Conference* eds F. Akici, D. Lippa, D. Niederhut, and M. Pacer , 113–120.

Kaczmarzyk, J., Goncalves, M., Bollmann, S., Notter, M., Halchenko, Y., Mitchell, R., et al. (2018). kaczmarj/neurodocker: Version 0.4.1. Neurodocker. Available online at: https://zenodo.org/record/2530918#.XFRQ6MZOkng

Keator, D. B., Grethe, J. S., Marcus, D., Ozyurt, B., Gadde, S., Murphy, S., et al. (2008). A national human neuroimaging collaboratory enabled by the biomedical informatics research network (birn). *IEEE Trans. Inf. Technol. Biomed.* 12, 162–172. doi: 10.1109/TITB.2008.917893

Keshavan, A., Datta, E., McDonough, I. M., Madan, C. R., Jordan, K., and Henry, R. G. (2017a). Mindcontrol: a web application for brain segmentation quality control. *NeuroImage* 170, 365-372. doi: 10.1016/j.neuroimage.2017.03.055

Keshavan, A., Klein, A., and Cipollini, B. (2017b). Interactive online brain shape visualization. *Res. Ideas Outcomes* 3:e12358. doi: 10.3897/rio.3.e12358

Keshavan, A., Yeatman, J., and Rokem, A. (2018). Combining citizen science and deep learning to amplify expertise in neuroimaging. *bioRxiv [Preprint]*. doi: 10.1101/363382

Kim, J. S., Greene, M. J., Zlateski, A., Lee, K., Richardson, M., Turaga, S. C., et al. (2014). Space–time wiring specificity supports direction selectivity in the retina. *Nature* 509:331–336. doi: 10.1038/nature13240

Kim, L., West, J. D., and Stovel, K. (2017). "Echo chambers in science?" in *American Sociological Association (ASA) Annual Meeting*. Available online at: http://www.asanet.org/annual-meeting-2017

Klapwijk, E. T., van de Kamp, F., van der Meulen, M., Peters, S., Wierenga, L. M., and Wierenga, L. M. (2018). Qoala-t: a supervised-learning tool for quality control of freesurfer segmented mri data. *Neuroimage* 189, 116-129. doi: 10.1016/j.neuroimage.2019.01.014

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLoS Comput. Biol* 13:e1005350. doi: 10.1371/journal.pcbi.1005350

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., et al. (2016). "Jupyter notebooks-a publishing format for reproducible computational workflows," in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, eds F. Loizides and B. Scmidt (IOS Press), 87–90. Available online at: https://eprints.soton.ac.uk/403913/

Laird, A. R., Eickhoff, S. B., Fox, P. M., Uecker, A. M., Ray, K. L., Saenz, J. J., et al. (2011). The brainmap strategy for standardization, sharing, and meta-analysis of neuroimaging data. *BMC Res. Notes* 4:349. doi: 10.1186/1756-0500-4-349

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., et al. (2008). Galaxy zoo: morphologies derived from visual inspection of galaxies from the sloan digital sky survey. *Mon. Not. Roy. Astron. Soc.* 389, 1179–1189. doi: 10.1111/j.1365-2966.2008.13689.x

Longo, D. L. and Drazen, J. M. (2016). Data sharing. *N. Engl. J. Med.* 374, 276–277. doi: 10.1056/NEJMe1516564

Madan, C., Klein, A., Migineishvili, N., Faskowitz, J., Gorgolewski, C. F., Kong, X., et al. (2018). *Cmadan/Openmorph: 2018-August.* Available online at: https://zenodo.org/record/1342740#.XGOlKIVOlKg

Madan, C. R. (2017a). Advances in studying brain morphology: the benefits of open-access data. *Front. Hum. Neurosci.* 11:405. doi: 10.3389/fnhum.2017.00405

Madan, C. R. (2017b). *Beyond the Ivory Tower: Reciprocal Relationships Between Scientific Research, Communication, and Mentorship.* Available online at: https://f1000research.com/slides/6-755

Marelli, L. and Testa, G. (2018). Scrutinizing the EU general data protection regulation. *Science* 360, 496–498. doi: 10.1126/science.aar5419

Mennes, M., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2013). Making data sharing work: the fcp/indi experience. *Neuroimage* 82, 683–691. doi: 10.1016/j.neuroimage.2012.10.064

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., et al. (2005). The alzheimer's disease neuroimaging initiative. *Neuroimaging Clin.* 15, 869–877. doi: 10.1016/j.nic.2005.09.008

Narayanan, A. and Shmatikov, V. (2008). "Robust de-anonymization of large sparse datasets," in *Security and Privacy, 2008. SP 2008. IEEE Symposium on* (Oakland, CA: IEEE) 111–125. Available online at: https://ieeexplore.ieee.org/document/4531148

Niso, G., Rogers, C., Moreau, J. T., Chen, L.-Y., Madjar, C., Das, S., et al. (2016). Omega: the open meg archive. *Neuroimage* 124, 1182–1187. doi: 10.1016/j.neuroimage.2015.04.028

Ogawa, S., Lee, T.-M., Nayak, A. S., and Glynn, P. (1990). Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. *Mag. Res. Med.* 14, 68–78.

Poldrack, R. A., Barch, D. M., Mitchell, J., Wager, T., Wagner, A. D., Devlin, J. T., et al. (2013). Toward open sharing of task-based fmri data: the openfmri project. *Front. Neuroinform.* 7:12. doi: 10.3389/fninf.2013.00012

Poldrack, R. A. and Gorgolewski, K. J. (2014). Making big data open: data sharing in neuroimaging. *Nat. Neurosci.* 17:1510. doi: 10.1038/nn.3818

Poline, J.-B., Breeze, J. L., Ghosh, S. S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Front. Neuroinform.* 6:9. doi: 10.3389/fninf.2012.00009

Ragan-Kelley, M., Perez, F., Granger, B., Kluyver, T., Ivanov, P., Frederic, J., and Bussonnier, M. (2014). "The jupyter/ipython architecture: a unified view of computational research, from interactive exploration to communication and publication," in *AGU Fall Meeting Abstracts*.

Roskams, J. and Popović, Z. (2016). Power to the people: addressing big data challenges in neuroscience by creating a new cadre of citizen neuroscientists. *Neuron* 92, 658–664. doi: 10.1016/j.neuron.2016.10.045

Sarwate, A. D., Plis, S. M., Turner, J. A., Arbabshirani, M. R., and Calhoun, V. D. (2014). Sharing privacy-sensitive access to neuroimaging and genetics data: a review and preliminary validation. *Front. Neuroinform.* 8:35. doi: 10.3389/fninf.2014.00035

Scott, A., Courtney, W., Wood, D., De la Garza, R., Lane, S., Wang, R., King, M., Roberts, J., Turner, J. A., and Calhoun, V. D. (2011). Coins: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets. *Front. Neuroinform.* 5:33. doi: 10.3389/fninf.2011.00033

Sherif, T., Kassis, N., Rousseau, M.-É., Adalat, R., and Evans, A. C. (2015). Brainbrowser: distributed, web-based neurological data visualization. *Front. Neuroinform.* 8:89. doi: 10.3389/fninf.2014.00089

Sherif, T., Rioux, P., Rousseau, M.-E., Kassis, N., Beck, N., Adalat, R., et al. (2014). Cbrain: a web-based, distributed computing platform for collaborative neuroimaging research. *Front. Neuroinform.* 8:54. doi: 10.3389/fninf.2014.00054

Simpson, R., Page, K. R., and De Roure, D. (2014). "Zooniverse: observing the world's largest citizen science platform," in *Proceedings of the 23rd International Conference on World Wide Web* (Seoul: ACM), 1049–1054.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., et al. (2004). Advances in functional and structural mr image analysis and implementation as fsl. *Neuroimage* 23, S208–S219. doi: 10.1016/j.neuroimage.2004.07.051

Sochat, V. V., Eisenberg, I. W., Enkavi, A. Z., Li, J., Bissett, P. G., and Poldrack, R. A. (2016). The experiment factory: standardizing behavioral experiments. *Front. Psychol.* 7:610. doi: 10.3389/fpsyg.2016.00610

Taylor, L. E., Swerdfeger, A. L., and Eslick, G. D. (2014). Vaccines are not associated with autism: an evidence-based meta-analysis of case-control and cohort studies. *Vaccine* 32, 3623–3629. doi: 10.1016/j.vaccine.2014.04.085

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE* 6:e21101. doi: 10.1371/journal.pone.0021101

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The enigma consortium: large-scale collaborative

analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182. doi: 10.1007/s11682-013-9269-5

Vaccarino, A. L., Dharsee, M., Strother, S., Aldridge, D., Arnott, S. R., Behan, B., et al. (2018). Brain-code: A secure neuroinformatics platform for management, federation, sharing and analysis of multi-dimensional neuroscience data. *Front. Neuroinform.* 12:28. doi: 10.3389/fninf.2018.00028

Van Horn, J. D. and Gazzaniga, M. S. (2013). Why share data? lessons learned from the fmridc. *Neuroimage* 8, 677–682. doi: 10.1016/j.neuroimage.2012.11.010

Van Horn, J. D. and Toga, A. W. (2009). Is it time to re-prioritize neuroimaging databases and digital repositories? *Neuroimage* 47, 1720–1734. doi: 10.1016/j.neuroimage.2009.03.086

Van Noorden, R. (2012). Trouble at the text mine. *Nature* 483, 134–135. doi: 10.1038/483134a

Wang, L., Alpert, K. I., Calhoun, V. D., Cobia, D. J., Keator, D. B., King, M. D., et al. (2016). Schizconnect: mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration. *NeuroImage* 124, 1155–1167. doi: 10.1016/j.neuroimage.2015.06.065

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The fair guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Yarkoni, T. (2009). Big correlations in little studies: Inflated fmri correlations reflect low statistical power-commentary on vul et al.(2009). *Perspect. Psychol. Sci.* 4, 294–298. doi: 10.1111/j.1745-6924.2009.01127.x

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8:665. doi: 10.1038/nmeth.1635

Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A., and Feldman, H. M. (2012). Tract profiles of white matter properties: automating fiber-tract quantification. *PLoS ONE* 7:e49790. doi: 10.1371/journal.pone.0049790

Yeatman, J. D., Richie-Halford, A., Smith, J. K., Keshavan, A., and Rokem, A. (2018). A browser-based tool for visualization and analysis of diffusion mri data. *Nat. Commun.* 9:940. doi: 10.1038/s41467-018-03297-7

National Neuroinformatics Framework for Canadian Consortium on Neurodegeneration in Aging (CCNA)

Zia Mohaddes [1,2*†], Samir Das [1,2†], Rida Abou-Haidar [1,2], Mouna Safi-Harab [1,2], David Blader [1,2], Jessica Callegaro [1,2], Charlie Henri-Bellemare [1,2], Jingla-Fri Tunteng [1,2], Leigh Evans [1,2], Tara Campbell [1,2], Derek Lo [1,2], Pierre-Emmanuel Morin [3], Victor Whitehead [4], Howard Chertkow [4,5] and Alan C. Evans [1,2]

[1] McGill Centre for Integrative Neuroscience, Montreal, QC, Canada, [2] Montreal Neurological Institute, Montreal, QC, Canada, [3] Centre de recherche de l'Institut Universitaire de Gériatrie de Montréal, Montreal, QC, Canada, [4] Lady Davis Institute, Montreal, QC, Canada, [5] Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

The Canadian Institutes for Health Research (CIHR) launched the "International Collaborative Research Strategy for Alzheimer's Disease" as a signature initiative, focusing on Alzheimer's Disease (AD) and related neurodegenerative disorders (NDDs). The Canadian Consortium for Neurodegeneration and Aging (CCNA) was subsequently established to coordinate and strengthen Canadian research on AD and NDDs. To facilitate this research, CCNA uses LORIS, a modular data management system that integrates acquisition, storage, curation, and dissemination across multiple modalities. Through an unprecedented national collaboration studying various groups of dementia-related diagnoses, CCNA aims to investigate and develop proactive treatment strategies to improve disease prognosis and quality of life of those affected. However, this constitutes a unique technical undertaking, as heterogeneous data collected from sites across Canada must be uniformly organized, stored, and processed in a consistent manner. Currently clinical, neuropsychological, imaging, genomic, and biospecimen data for 509 CCNA subjects have been uploaded to LORIS. In addition, data validation is handled through a number of quality control (QC) measures such as double data entry (DDE), conflict flagging and resolution, imaging protocol checks[1], and visual imaging quality validation. Site coordinators are also notified of incidental findings found in MRI reads or biosample analyses. Data is then disseminated to CCNA researchers via a web-based Data-Querying Tool (DQT). This paper will detail the wide array of capabilities handled by LORIS for CCNA, aiming to provide the necessary neuroinformatic infrastructure for this nation-wide investigation of healthy and diseased aging.

Keywords: database, neuroimaging, infrastructure, dementia, Alzheimer's

---

[1] Canadian Dementia Imaging Protocol (CDIP).

INTRODUCTION

With 500,000 Canadians diagnosed with Alzheimer's Disease (AD), neurodegenerative diseases (NDDs) are becoming an increasing priority for Canadian society due to their significant and increasing socio-economic costs, which is estimated nationally at 15 billion Canadian dollars annually, and expected to rise to 150 billion dollars by 2038 (Fostering Alzheimer Society of Canada, 2010; Statistics Canada, 2016). As a result, conducting a nationwide study to investigate NDDs is paramount for achieving a better understanding of their etiologies, finding ways to mitigate their impact, and ultimately preventing their development. In response, an initiative spearheaded by the Canadian Institutes of Health Research (CIHR)[2] and supported by various provincial and non-governmental partners (**Appendix 1**), assembled 340 researchers from across Canada to form the Canadian Consortium on Neurodegeneration and Aging (CCNA)[3] The ultimate mandate of CCNA is to coordinate and strengthen Canadian research groups to better delineate and manage the causes, early detection, and treatment of dementia.

The primary vehicle for pursuing this mandate is the Comprehensive Assessment of Neurodegeneration and Dementia (COMPASS-ND), the signature clinical study of CCNA, which is currently collecting clinical, sensory, neuropsychological, neuroimaging, biological, and genetic data from a cohort of 1,650 individuals aged 50–90 with multiple types and severities of cognitive impairment, as well as 660 cognitively intact elderly individuals, recruited across 30 Canadian sites (**Appendix 2**). This study poses a unique and challenging technical undertaking, as it requires curation and standardization of diversified data from numerous sites across the country. To this end, CCNA has deployed LORIS[4], a web-based data management system for multi-site studies, to facilitate collection, processing, analysis, and dissemination of multi-modal data, while ensuring accuracy and completeness with numerous quality control (QC) metrics in place (Das et al., 2012, 2016). LORIS has been tailored to CCNA's needs through the customization of key features, such as (1) Bilingual forms, (2) Training Portal for the familiarization of clinical and neuropsychological measures (Campbell, 2017), (3) Study Tracker for detailed study progression, 4) Biobanking module with support for any type of biospecimen, (5) Genomic Browser hosting CPG, SNP, and CNV genomics data (Rogers, 2015), (6) Imaging Uploader for multi-modal imaging data (Mohaddes, 2015), (7) Radiological Review module for incidental finding alerts and tracking, (8) Web-based Data-Querying Tool (DQT) that enables customizable query-building and data extraction (MacFarlane, 2014), (9) Configuration module to allow end-users to customize the interface, and (10) Publications module to manage consortium-led publications. LORIS' user-friendly interface, visualization tools, and targeted workflows also conveniently connect interdisciplinary teams of researchers on one platform.

In designing the specific features of LORIS for use within CCNA, many lessons from other international initiatives have been considered. Firstly, emphasis has been placed on building features within LORIS that are FAIR (Findable, Accessible, Interoperable and Reusable) (Chertkow, under review). The driving impetus revolves around concerns of usability and accuracy of data, especially given that the field of neuroscience (among other scientific fields) has been faced with a "reproducibility crisis" (Bennet and Miller, 2010; Glatard et al., 2015; Eklund et al., 2016; Fostering reproducible fMRI Research, 2017; Gilmore et al., 2017). As data proliferates, the methods of managing this data need to be carefully considered to avoid time and resource-consuming errors related to the increased order of complexity in its handling.

Through the use of a comprehensive data management system, CCNA hopes to contribute pivotal research findings to expand our understanding of NDDs, working toward prevention, and improving the quality of life for those with dementia. This paper examines how LORIS is tailored to the specific workflows within CCNA and highlights key features that have been implemented in order to facilitate data sharing between CCNA and similar studies at a provincial level, including Ontario Neurodegenerative Disease Research Initiative (ONDRI)[5] and Consortium for the Early Identification of Alzheimer's Disease (CIMA-Q)[6]

METHODS

The CCNA infrastructure is composed of numerous elements that need to be organized, interoperable, and scalable, while conforming to the CCNA-specific cohorts, such as COMPASS-ND, the 5-years observational study aimed at recruiting 1,650 subjects (Chertkow, under review). The LORIS platform has been chosen to service this national initiative and has been set up to: (1) enable clinical sites to collect behavioral data with multi-modal QC checks, and direct feedback to data entry personnel, (2) facilitate biosample collection, (3) streamline imaging acquisitions and related QC metrics, (4) allow data sharing among researchers by performing self-administered queries, while tracking participant status and data entry progress throughout the study, (5) improve interoperability between different projects using APIs, and (6) provide comprehensive user support for the CCNA research community. As highlighted below, we will examine each of these aspects to delineate how the various multi-modal workflows have been configured in LORIS.

Behavioral Data Acquisition

Behavioral data acquisition includes the collection of numerous measures (clinical, demographic, psychometric, and neuropsychological) from subjects across study sites. Data are collected through the use of paper forms which are later digitized by entering the information into replicate online forms. Several QC steps have been implemented in order to prevent the storage and distribution of erroneous information at the

---

[2]http://www.cihr-irsc.gc.ca/e/46475.html

[3]http://ccna-ccnv.ca

[4]http://loris.ca

[5]http://ondri.ca

[6]http://www.cima-q.ca/en/home

FIGURE 1 | CCNA data-upload flow in LORIS. Step 1: Creation and registration of subjects, Step 2: Creation of timepoint and media upload, Step 3: Behavioral data entry and QC, and Step 4: Visit review and data dissemination.
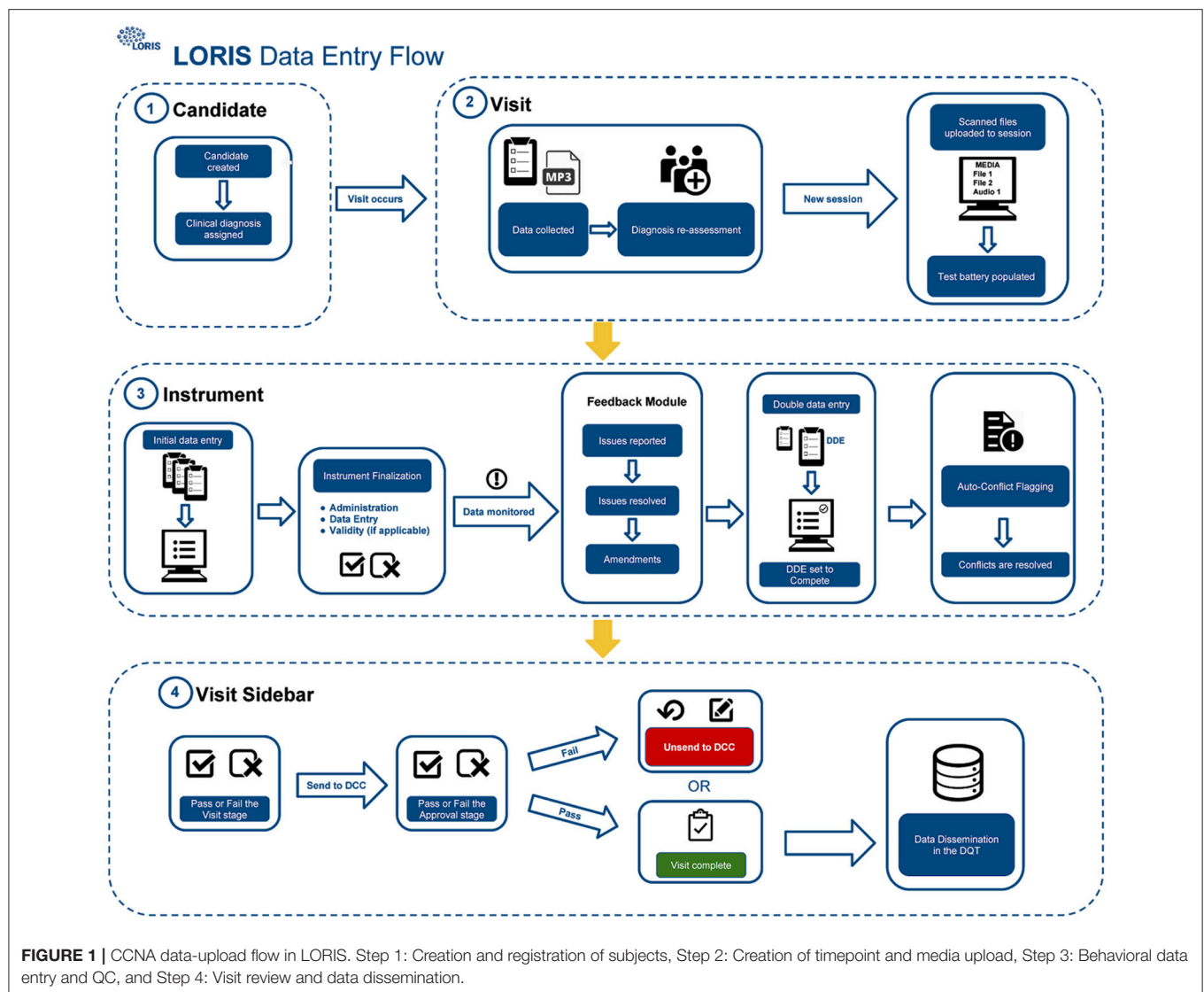
time of dissemination. Rules are also coded within the LORIS forms to ensure the completion of all required fields, avoiding accidental omission of critical data. Similarly, fields that are not contingent on the condition of a particular response are exempt to avoid redundant data. Forms are also translated in both French and English, with a unified backend storage solution that stores a singular value across languages for analysis consistency.

The procedure for uploading participant data into the database (**Figure 1**) includes four different QC checks: (1) double data entry (DDE), (2) conflict resolution, (3) monitoring, and (4) final submission validation. As described in **Figure 1**, scanned documents and audio files are uploaded to LORIS via the Media Uploader where COMPASS-ND[7] naming conventions are imposed. Subsequently, data monitoring checks take place

between initial and DDE. Once data have been validated and pertinent details have been reported by a monitor via the Behavioral Feedback tool, DDE is then completed by a second individual. In case of a discrepancy between initial and DDE results, a conflict will automatically be flagged in the Conflict Resolver module. Conflicts are then resolved when data entry personnel review and manually select the correct answer from the conflicting options, by cross-referencing the scanned original paper forms (uploaded to the Media Uploader).

Finally, once all conflicting data have been amended and behavioral feedback have been addressed, forms can be set to "Complete" and the timepoint can be submitted for final validation. This final step requires a specialist to confirm that all necessary actions were taken and that the data are correct and ready for release; only at this point will the data be frozen to avoid any further modifications, and made available for dissemination.

---

[7]"COMPASS-ND Study." CCNA. Available online at: http://ccna-ccnv.ca/en/compass-nd-study/ (Retrieved December 11, 2016).

Biosample Collection

Representing a core aspect of the data collected for the COMPASS-ND Study, biospecimen collection brings its share of challenges to the project. Biosamples collected from a subject can not be digitized and archived in the same manner as imaging and behavioral data, because the sample location, status, and availability must be easily accessible while continuously staying up to date, both in the system and in their physical storage. A Biobanking tool within LORIS was designed specifically for these purposes. Each collected specimen has a predefined storage destination (see **Table 1**) and processing plan (see **Figure 2**). CCNA draws on expertise from researchers across Canada including, but not limited to, the Toronto Genomics Center and the Canadian Biosample Repository (CBSR). As the CBSR facilities provide a biosample storage solution for CCNA's needs, LORIS adapted its sample tracking tool to streamline interactions with the CBSR database. This collaborative effort has led to the development of a strict standard for export and import of biospecimen data between CCNA sites and the CBSR database, reinforced on both ends by LORIS software and CBSR-Biobank software, respectively.

Imaging

Another crucial aspect to CCNA's COMPASS-ND project is the neuroimaging collection. The imaging workflow has been summarized in **Figure 3** and consists of (1) scan collection, (2) visual QC, (3) MRI reads, and (4) storing volumetric data analysis derived from external tools. An important step in this workflow is the automated notifications to the sites when necessary.

Scan Collection

A major aspect of a multi-site study like COMPASS-ND is the importance of timely scan upload by the acquisition sites.

TABLE 1 | Sample types, amounts, and destination.

Sample type	Amount (tubes)	Amount	Destination
Whole blood	6 ml (two 3 ml tubes)	N/A	Sent to local labs for CBC and HbA1C analysis
Serum	30 ml (three 10 ml red top tubes)	15 ml	30 aliquots → CBSR
Buffy Coat	12 ml (two 6 ml lavender top tubes)	1 ml	2 aliquots → CBSR
Plasma		6 ml	12 aliquots → CBSR
DNA	6 ml (one 6 ml lavender top tube)	N/A	Genomics Center, Toronto
RNA	2.5 ml (one 2.5 ml paxgene tube)	N/A	Genomics Center, Toronto
Urine	Up to 60 ml	6 ml	12 aliquots → CBSR
Saliva	4 ml (1 DNA genotek tube)	4 ml	8 aliquots → CBSR 3 ml → locally for glucose, cells
CSF	15 ml (two 10 ml polypropylene tubes for 12 ml plus two standard hospital tubes 1–2 ml each)	15 ml	24 aliquots → CBSR

Enabling faster QC allows for re-scan possibilities and quicker response to incidental findings. Furthermore, the COMPASS-ND central administration has set up a site reimbursement policy whereby scan costs are reimbursed only after images have been uploaded to LORIS and have adequate quality based on visual assessments. This has resulted in reducing (if not eliminating) missed scan uploads.

Another challenge in the success of this multi-site study is adherence to a common MRI protocol that is flexible enough to accommodate different scanner hardware capabilities across manufacturers (Siemens, GE, and Philips), yet strict enough to maintain a certain level of parameter uniformity across the different MRI modalities collected. The Canadian Dementia Imaging Protocol (CDIP; Duchesne, 2015; Duchesne et al., 2018) used in CIMA-Q, a preceding dementia study in Québec (Belleville, 2014), was extended to cover an increased number of scanners (7) in CIMA-Q vs. 20 in CCNA) and recruitment sites (7) in CIMA-Q vs. 35 in CCNA), thereby striking the balance between flexibility and uniformity.

Visual QC

All scans inserted into LORIS are reviewed visually by a trained team looking for artifacts (motion, scanner, intensity, and others). Newly inserted scans can be identified quickly which facilitates the visual QC feedback to be entered directly in the system. The goal of this step is to identify images within a session that pass QC standards set by the study, allowing further analysis to proceed. In addition, in failed QC cases, further information is communicated to the site to decide on the proper course of action (subject re-scan if required).

Incidental Findings

Visually QC'ed sessions are important because they trigger another important step in the imaging workflow, namely, the MRI interpretations by another specialized team. The goal is to identify any incidental findings and report back to the site coordinator for further action in a timely manner. Forms with restricted access granted to the COMPASS-ND radiology team are devised for the study, and a notification system to alert the site when this "research" read is completed. This notification system points the site coordinator to a text printable version of the report, and requires an acknowledgment of receipt of such notification by the site. In case of incidental findings, the site takes the full responsibility to follow up with the subject's physician, while LORIS facilitates physician access to the images, if desired. A detailed workflow diagram and timeline is shown below (see **Figure 4** for incidental findings procedure).

Derived Data

Although derived data of non-imaging modalities are being incorporated, the focus here is on illustrating how two types of derived imaging data, both constituting an integral part of the COMPASS-ND imaging workflow and mandate, are being integrated back to LORIS. The first is an automated hippocampal segmentation analysis based on True Positive Medical Devices

FIGURE 2 | Sample collection and distribution.
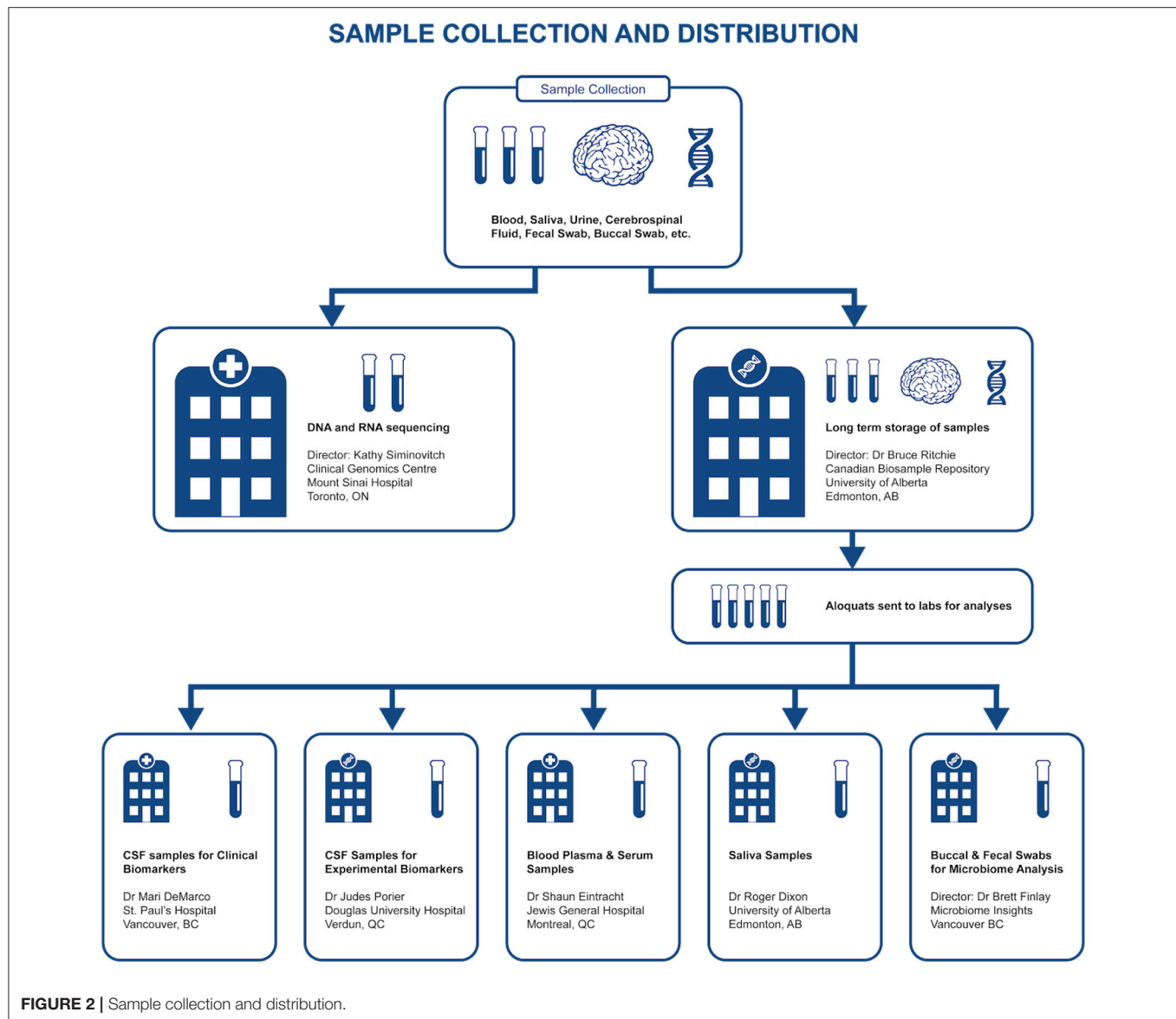
(TPMD)[8], where the goal is to perform volumetric analysis, along with z-scores on different brain regions in comparison with a control cohort from the Alzheimer's Disease Neuroimaging Initiative (ADNI[9]). LORIS is extended to facilitate this task using a set of tools that imports this data, processed externally, back onto LORIS. The larger goal of importing this data back into LORIS as behavioral forms is made possible through built-in DQT capabilities that render this information as easily queryable and accessible by researchers.

The second type of derived data is the lego phantom image processing results (Fonov et al., 2010) for scanner distortion identification and parameter extraction. These data are re-inserted into LORIS for further correction by researchers, if needed. A LORIS-CBRAIN hook was developed to read phantom images (T1W, T2W, PD, and resting state fMRI) from LORIS'

filesystem, processed on CBRAIN (Sherif et al., 2014), with processed images and files put back onto the COMPASS-ND filesystem. This used a containerized version of the Fonov processing pipeline. Automatic launching of these tasks directly from LORIS (i.e., without requiring logging into and launching the task from within CBRAIN) is currently being developed, making the process completely automated to the LORIS CCNA user.

Data-Dissemination

Data dissemination is a crucial aspect in processing and analysis. For this reason, we have incorporated several tools into LORIS to facilitate querying, download and sharing, including our DQT and Study Tracker.

The Data Query Tool
Historically, researchers required a programmer or database administrator to query the database, and disseminate particular

---

[8]http://truepositivemd.com
[9]http://adni.loni.usc.edu

FIGURE 3 | Imaging workflow.
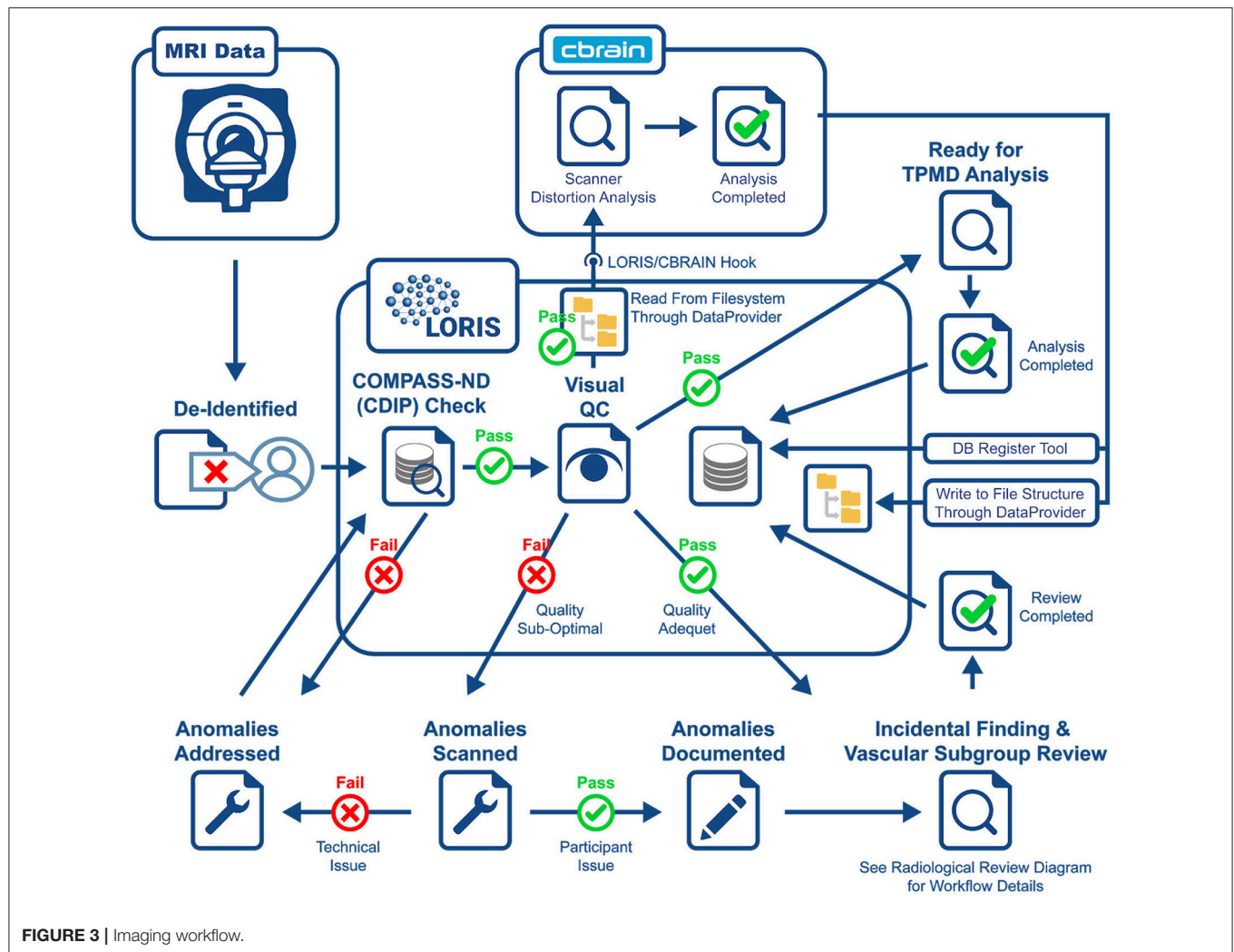
data outputs. This had potentially translated into days or weeks of delay for their investigations. Today, LORIS assigns great priority to data dissemination and enables researchers to directly query the database and easily retrieve data. The DQT allows researchers to design, execute, and save queries in a simple and intuitive manner, without having to write complex database queries. The interface allows for selection of variables, and quick download in most commonly used formats (e.g., Excel, comma separated values, and tab separated values). In addition, users can save any query (both the variables and the population) and use it in the future with new or updated data, without worrying about ambiguities and inconsistencies. Datasets can also be tagged with a version number or a timestamp such that longitudinal comparisons can be made, minimizing any ambiguity about what has been downloaded (Das et al., 2012; MacFarlane, 2014).

Study Tracker

Due to the magnitude and complexity of the COMPASS-ND project, there is substantial difficulty in monitoring the progress of data entry and visit registration among the dozens of study

sites submitting hundreds of different forms. Queries to the database to acquire this information would be considerably complex, with outputs that are convoluted and time consuming for users to parse. With this in mind, the Study Tracker module was developed, as described in **Figure 5**, to provide study administrators an efficient graphical interface to view the progression of data entry and visit registration. Using a simple color coded system, representing data entry completion relative to each timepoint's individual deadline, users are able to quickly identify data entry bottlenecks and unresolved issues for each study participant at every visit. Focusing on a specific timepoint brings up links that allow users to navigate to: (1) individual measures administered at that timepoint, (2) Conflict Resolver module to settle conflicting data entry values, and (3) open Behavioral Feedback discussion threads.

Interoperability

An important aspect of this study is being able to aggregate data from already existing nodes. In order to reach COMPASS-ND's

INCIDENTAL FINDINGS

Acknowledge reception of Radiology Report

MRI Synoptic report is made web accessible

Potentially clinically relevant findings present? — **YES** → Provide report and subject chart to study physician within 48 hours. Coordinate with Radiological team if necessary

NO

Was the cerebrovascular brain injury sufficient to classify the subject as vascular/mixed MCI or vascular/mixed dementia? — **NO** → No further action is required

YES

Any previous V-MCI or Mixed Dementia diagnosis? — **YES** → No further action is required

NO

Provide report and subject chart to study physician within two weeks. Confirm and input new diagnosis into Diagnostic Reappraisal

FIGURE 4 | Detailed diagram of procedure for incidental findings.

enrollment target, interoperability has been developed with BrainCODE[10], a platform that houses the ONDRI study as part of CCNA's mandate. There are currently about 150 subjects in ONDRI that can be shared with CCNA. As such, LORIS and BrainCODE needed to customize their API capabilities to enable a 2-way transfer of imaging data that adheres to the internal structures and desired business rules of both systems. Transfer of images from BrainCODE to LORIS is based on LORIS polling BrainCode's XNAT imaging storage system through a middleware interface (**Figure 6**). Conversely, transferring images to BrainCODE relies on a LORIS initiation of an automated message to BrainCODE's XTXGate[11] notification system. The XTXGate system receives a JSON-encoded message from LORIS with all metadata necessary to identify, and subsequently download, the raw DICOM images of the newly added MRI study using LORIS' API.

User Support

Serving as a centralized user support portal for the consortium, the Member's Portal is a forum relying on the Discourse software[12] to connect users, coordinators, and administrators to each other. The Discourse application is integrated inside LORIS to avoid confusion and the steeper learning curves associated with multiple-platform training. The portal posts are constantly and consistently monitored by moderators and support staff to ensure the content is appropriate and triaged in the appropriate categories.

RESULTS

One of the major goals of CCNA is to develop a national cohort to provide researchers with data to test and refine their hypotheses about the progression of dementia. LORIS itself is the tool through which the research community can access data in a trustworthy and robust manner. This section will highlight the

[10]http://braininstitute.ca/research-data-sharing/brain-code
[11]https://xtxgate.braincode.ca/notify

[12]https://www.discourse.org

FIGURE 5 | Study Tracker. Each row corresponds to a participant, with each circle in that row corresponding to a visit for that participant. The border color of the circle represents the status of visit registration and the fill indicates data entry completion with respect to the due date. Here the open sidebar has the visit specific view in focus, with links to individual forms as well as links to the Conflict Resolver.

data that were acquired, as well as several features that have been tailored and developed within LORIS.

Behavioral Data in Numbers

Presently, CCNA has 509 registered subjects for COMPASS-ND across 10 cohorts of various NDDs, 459 (90.17%) of which are currently active in the study, omitting subjects who have since become inactive (excluded, ineligible, etc.). **Table 2** summarizes the number of registered subjects by cohort and includes an in-depth breakdown of currently active subjects by sex. COMPASS-ND will enroll 1,650 participants with various types and severities of cognitive impairment, as well as 660 cognitively intact participants, which will be "deeply phenotyped" through data collection from numerous modalities (Chertkow, under review).

Currently clinical, neuropsychological and biospecimen data for 509 COMPASS-ND subjects have been uploaded as described in steps 1–4 in **Figure 1**. Upon successful upload, COMPASS-ND coordinators use a suite of behavioral QC modules, such as the Conflict Resolver and Feedback Module, to ensure completeness and accuracy of the data. There have been just over 30,000 conflicts flagged since the launch of COMPASS-ND, spread over ~200 subjects with over 8,100 data fields each; this number represents a 1.84% data entry error rate and 60% conflict resolution rate. Once data have passed the approval stage, they can be flagged as "ready for dissemination" and are then accessible through the DQT.

Biobank

The biospecimen collection and tracking needs for the study have been addressed by the development of a specialized tool

in LORIS. This was implemented in order to streamline and increase the level of automation for acquisition, and tracking of biosamples on site during the subject's visit. The tool is composed of seven pages, each dealing with specific tissue types or bodily fluids organized in their order of collection: blood, urine, saliva, cerebrospinal fluid (CSF), buccal and fecal samples, and a final page for extras (if there are any unused vials). Each page is fully independent and each row on the page contains information on a single sample only (see **Figure 7**); a row contains information on the time of collection, barcode ID, destination and location of the biosample. Each field can also be saved independently. Furthermore, a set of rules and validation steps automatically enforced by the software prevent loss of data integrity. These rules ensure that all barcode IDs are unique in the database, that the barcode scanned is of the correct type for the specimen, that any modification is logged with a timestamp, and that no field is left incomplete. After all the biosamples are scanned into LORIS, a shipment flag is enabled indicating that the sample is ready to be sent to the off-site biospecimen repository; this flag can only be set if all validations pass. Before the samples are shipped out and the data are exported, a last quality check consisting of all of the validations above is run by an administrator on the data. This ensures that the exported data have not been modified between the entry date and export date. Through training of the staff, in conjunction with LORIS monitoring for omissions or errors in the data entry of biosamples, the risk of inconsistencies in the database is significantly reduced.

Imaging in Numbers

A total of 467 imaging scans (364 subjects and 103 phantoms, 76 of which are for lego phantoms), conformant to the Canadian
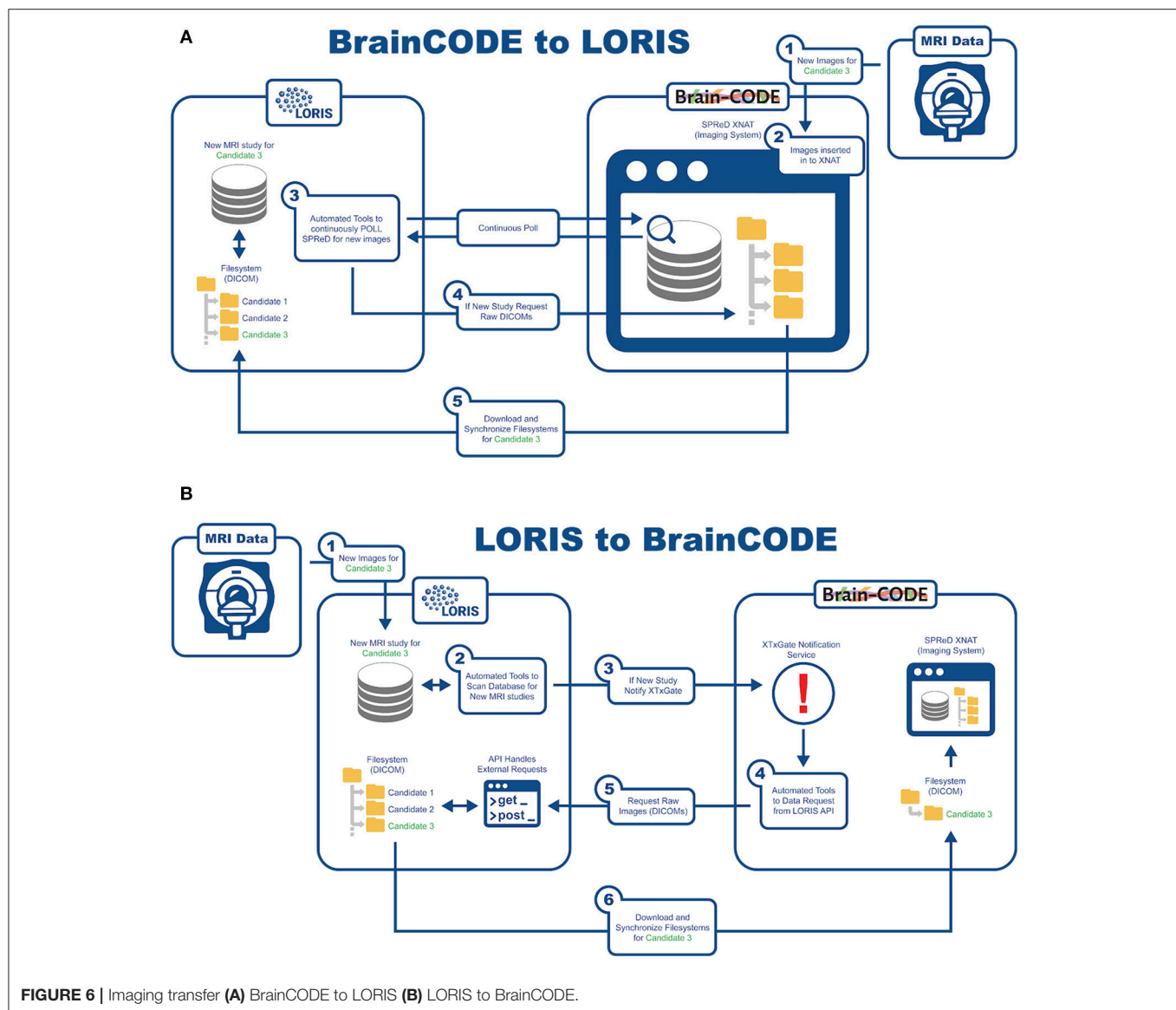
FIGURE 6 | Imaging transfer **(A)** BrainCODE to LORIS **(B)** LORIS to BrainCODE.

Dementia Imaging Protocol (CDIP), have been uploaded from 20 different sites via the Imaging Uploader (Mohaddes, 2015). Visualization and QC of image files are tailored to each modality and customized per type of review (on-site, centralized, manual, or automated). Of the 364 uploaded scans, 346 underwent visual QC on the different MRI subtype structural modalities (with the following visual QC failures: 6 T1W, 3 2D Flair, 17 T2*, and 5 dual echo PD/T2), and 331 were read for MRI incidental findings, of which 19 had potentially clinically relevant incidental findings. The remaining scans are queued to be QC'd as well, as our QC team manages to complete on average roughly a dozen QCs a week. In addition, TPMD processed data analysis for 266 subjects have been performed offline and results were imported back to LORIS, with more analyses from the remaining subjects to come. Each analysis contains specific anatomical measurements for gray and white matter volume and other quantitative biomarkers.

Follow-up scans obtained from these subjects can then be processed using TPMD, allowing for not only the course of the degenerative diseases to be tracked, but quantified with precise measurements.

Data Dissemination

The web-based DQT enables download of validated scalar data (clinical, demographic, psychometric, biosample, and neuropsychological data) linked to imaging data through a simple querying interface (see **Figure 8**). Currently, curation has been completed on the first 4 initial assessment visits (screening, clinical, neuropsychological, and MRI) including data entry feedback and conflict resolution in preparation for an upcoming data release (fall of 2018) of the first 200 CCNA participants. This release for CCNA researchers will include collected and derived measures, all made accessible through a simple query in the DQT.

TABLE 2 | Subject demographics per cohort.

Cohort	Registered subjects total	Active subjects total	Male	Female
SCI	82	76	23 (30.3%)	53 (69.7%)
MCI	186	176	86 (48.9.%)	90 (51.1%)
V-MCI	51	47	31 (66.0%)	16 (34.0%)
AD	71	55	34 (61.8%)	21 (38.2%)
Mixed	22	18	14 (77.8%)	4 (22.2%)
FTD	23	16	11 (68.7%)	5 (31.3%)
PD	21	29	18 (62.1%)	11 (37.9%)
PD-MCI	23	20	19 (86.4%)	3 (13.6%)
PDD	8	4	4 (100.0%)	0 (0.0%)
LBD	22	18	14 (77.8%)	4 (22.2%)
Total	509	459	252 (54.9%)	207 (45.1%)

SCI, Subjective Cognitive Impairment; MCI, Mild Cognitive Impairment; V-MCI, Subcortical Ischemic Vascular MCI; AD, Alzheimer's disease; Mixed, Dementia of Mixed Etiology; FTD, Frontotemporal dementia; PD, Parkinson's dementia; PD-MCI, Parkinson's dementia (PD) with MCI; PDD, Parkinson's Disease Dementia; LBD, Lewy Body disease.

It is expected that this release will provide data to 17 different CCNA teams aiming at addressing 106 proposed research projects. Example questions include:

- What distinguishes V-MCI from MCI? Mixed from AD?
- How does frailty impact cognition?
- How does sensory impairment relate to cognitive impairment?
- Does a comprehensive blood profile distinguish MCI from AD?
- Do gait metrics distinguish AD from other types of dementia?

Interoperability

The technical infrastructure for imaging interoperability between COMPASS-ND and ONDRI operates without user intervention. It consists of scan imports from ONDRI, and includes DICOM header de-identification, automatic file renaming, as well as insertion via the Imaging pipeline using customizable MRI protocol validation. In addition, the LORIS API was extended to allow sharing/downloading of raw DICOM images (along with its already existing MINC sharing capability). The addition of an on-demand notification system to inform ONDRI of any newly added DICOMs also allows for the flow of images from CCNA to ONDRI, facilitating bi-directional MRI scan exchange in a seamless manner.

User Support in Numbers

User support in a project of this scope does require significant support mechanisms. The central purpose of the Members Portal is to provide user support, similar to a web forum where discussion occurs, and the highest number of posts are prioritized. Currently there are 457 support requests from users with 437 resolved issues, 12 with fixes underway, and 8 requiring more information from the reporter.

It has been nearly 2 years since the start of CCNA's COMPASS-ND data collection, and LORIS' technology and tools have been extended to fulfill COMPASS-ND's clinical and research needs. This is further demonstrated with a comprehensive, organized, and multi-modal dataset for 200 subjects that will be disseminated to CCNA researchers in fall 2018 (see **Figures 9–11**).

DISCUSSION

A national infrastructure for data management serves many needs that are paramount to the consortium's success (Toga, 2012). Having a tested technical platform for incoming data reduces the burden on researchers to manually manage data, and also provides tools and methods for accurate and efficient data entry (Poline et al., 2012; Nichols et al., 2016).

One of the most fundamental aspects is standardization, which remains an ongoing issue for any data sharing consortium. If data are properly structured, collaborative efforts become much more efficient and reduce the future burden of cross-site analysis (Gorgolewski et al., 2016; Munafo et al., 2017). To that effect, the CCNA infrastructure has incorporated many such standardization techniques in LORIS. Behavioral measures are coded in a structured manner, with all QC checkpoints rigorously enforced, and in turn, queryable alongside individual data items. Imaging acquisitions have been harmonized, as prescribed by the CDIP protocol. Biospecimen data have been organized to conform to a standard established between CCNA and CBSR, allowing for easier collaboration and safer data transfers. Collecting data according to a standardized common protocol and organizing it in a uniform manner makes it easier to process and share, while simultaneously reducing error rates resulting from manual manipulation of data (Gorgolewski and Poldrack, 2016). A great problem in the neuroscience community has been the ability to properly reproduce findings, an issue that stems in no small part from the lack of consistency of the data (Zuo and Xing, 2014; Zuo et al., 2014; Turner et al., 2018).

A major challenge within a national network, such as CCNA is ensuring interoperability between all technical segments and institutions. This is achieved through standardization initiatives, ensuring data harmonization, and proper API documentation and development useable between existing platforms and tools. To that effect, the CCNA-LORIS system has demonstrated interoperability, in that LORIS and BrainCODE regularly exchange imaging data between the ONDRI study and the national CCNA data platform. The technical software infrastructure for bi-directional exchange of behavioral data can mirror the imaging data, as embedded API functionalities make this sharing of behavioral data possible (for example in JSON format). However, any such behavioral data exchange includes additional challenges, in large part due to the harmonization required to translate data forms and dictionaries between ONDRI and CCNA or any two projects in general (Richesson and Nadkami, 2011). An extensive data mapping exercise is currently underway in LORIS transferring behavioral data from CIMA-Q to CCNA, with a similar endeavor shortly taking place between ONDRI and CCNA. While LORIS includes a data dictionary tool to facilitate ontological harmonization for any given study, limitations persist to map ontologies across studies due to greater standardization issues within the neuroscience community.

Biosamples - Blood

Sample	Barcode	Timestamp (YYYY-MM-DD HH:MM)	Place to send	Notes
EDTA 3ml tube 1	9KJL6O6GJXOL	2018-02-21 09:59	Local lab	
EDTA 3ml tube 2	O51Y6BRL3J1A	2018-02-21 10:00	Local lab	
EDTA 6ml tube 1	MZEMF272M2K1	2018-02-21 10:01	Genomics Centre, Toronto	
PaxGene RNA 2.5ml tube	IOX2FXNV949Q	2018-02-21 10:02	Genomics Centre, Toronto	
EDTA 6ml tube 2	G5R61B9Z0ZVV	2018-02-21 10:02	--	
Plasma vial 1	0217293767	2018-02-21 10:02	Matrix box 2, A1	
Plasma vial 2	0217293741	2018-02-21 10:02	Matrix box 2, A2	
Plasma vial 3	0217292468	2018-02-21 10:02	Matrix box 2, A3	
Plasma vial 4	0217340846	2018-02-21 10:03	Matrix box 2, A4	
Plasma vial 5	0217294146	2018-02-21 10:03	Matrix box 2, A5	
Plasma vial 6	0217294133	2018-02-21 10:03	Matrix box 2, A6	300 ml
Buffy coat vial 1	0217292490	2018-02-21 10:03	Matrix box 2, B1	
EDTA 6ml tube 3	G5IKUBVT1S7M	2018-02-21 10:03	--	

FIGURE 7 | Biosamples collection forms.

PSCID	Screening_Candidate_Recruitment,consent_form_signed_date	Screening_Candidate_Recruitment,Date_screening_visit	Screening_Candidate_Recruitment,Candidate_Age_months	Screening_Candidate_Recruitment,Candidate_Age_years
BRI1080	11/20/2017	11/20/2017	854.6	71.2
BRI2365	8/21/2017	8/21/2017	958.7	79.9
BRI5370	12/19/2017	12/19/2017	892.6	74.4
BRI5562	11/23/2017	11/23/2017	834.7	69.6
BRI7822	9/29/2017	9/29/2017	933.9	77.8
BRI8484	9/26/2017	9/26/2017	996.8	83.1
CBH0140	9/11/2017	9/11/2017	875.3	72.9
CBH0964	7/4/2017	7/4/2017	775.1	64.6
CBH1151	1/10/2017	1/10/2017	977.3	81.4
CBH1392	2/22/2017	2/22/2017	858.7	71.6
CBH2070	1/5/2017	1/5/2017	980.1	81.7
CBH2320	2/8/2018	2/8/2018	981.2	81.8
CBH2718	12/21/2016	12/21/2016	866.7	72.2
CBH2948	1/25/2017	1/25/2017	884.8	73.7
CBH3549	5/16/2017	5/16/2017	722.5	60.2
CBH3593	6/15/2017	6/15/2017	867.5	72.3
CBH4380	1/18/2017	1/18/2017	933.6	77.8
CBH4630	3/10/2017	3/10/2017	782.3	65.2
CBH5568	11/28/2017	11/28/2017	842.9	70.2
CBH6319	1/30/2017	1/30/2017	939	78.3
CBH6662	1/3/2017	1/3/2017	894.1	74.5
CBH7650	1/9/2017	1/9/2017	919.3	76.6
CBH7707	2/14/2017	2/14/2017	949.4	79.1
CBH7853	3/3/2017	3/3/2017	1007.1	83.9
CBH9502	5/4/2017	5/4/2017	791.1	65.9
CBH9592	3/24/2017	3/24/2017	805.8	67.2
CBH9834	5/25/2017	5/25/2017	759.8	63.3
CSA9212	12/13/2016	12/13/2016	819.4	68.3
DMC0481	10/30/2017	10/30/2017	863	71.9
DMC1145	1/31/2017	1/31/2017	912	76
DMC5382	10/23/2017	10/23/2017	771.7	64.3

FIGURE 8 | Sample result (extracted from DQT) for the upcoming CCNA release.

Provenance capture is also an important requirement in ensuring that data are usable and reproducible. LORIS natively handles a great deal of provenance information. For example, images including any associated metadata (e.g., scanner specifications, protocols, demographic information, and image processing details) are always extracted and stored. Behavioral data, scoring, updates, corrections, and any QC results are also captured. In addition, project metadata, as well as a complete subject audit trail are always available. LORIS also leverages existing platforms to provide the necessary information required for analysis (Maumet et al., 2016), and works with standardization groups to ensure maximum provenance retention (Glatard et al., 2015; Gorgolewski et al., 2016)

Increasing sample sizes and recruitment targets can improve collaborations and result in reduced redundancy in how research resources are invested (Evans and Brain Development Cooperative Group, 2006). Nationally coordinated approaches to building infrastructure can ensure more fruitful returns to a greater number of researchers across federally-funded initiatives. Furthermore, increased population can result in a more diversified sample that considers demographic variability, which is key to generalizability of results (Zuo et al.,

FIGURE 9 | Graphs showing overall activity on the Member's Portal over the last year. Signups: accounts added which are imported automatically from the LORIS users list. Topics: number of new support requests. Posts: number of support replies from our user support team. Daily active users/monthly active users (DAU/MAU): number of members that logged in within the last day, divided by number of members that logged in within the last month. Daily Engaged Users: number of users that liked or posted something new per day. New Contributors: number of users that made their first contribution during the indicated period.

FIGURE 10 | Graph representing the number of topics (support requests) created across time between January 1st, 2016 and August 18th, 2018.

2014). These points are especially true when dealing with machine learning algorithms, which require access to large data samples that need to be organized in a consistent manner (Toga, 2012). In the absence of larger cohorts, researchers have traditionally leveraged smaller studies with non-standardized protocols for meta-analysis, often lacking adequate harmonization or normalization. Consequently, this results in significant variability and decreased confidence in findings. Momentum has accumulated toward sustaining a higher standard of data quality and volume to accelerate research discovery optimized for transparent and reliable results.

FIGURE 11 | Graph representing the longitudinal number of posts (support request replies) between January 1st, 2016 and August 18th, 2018.
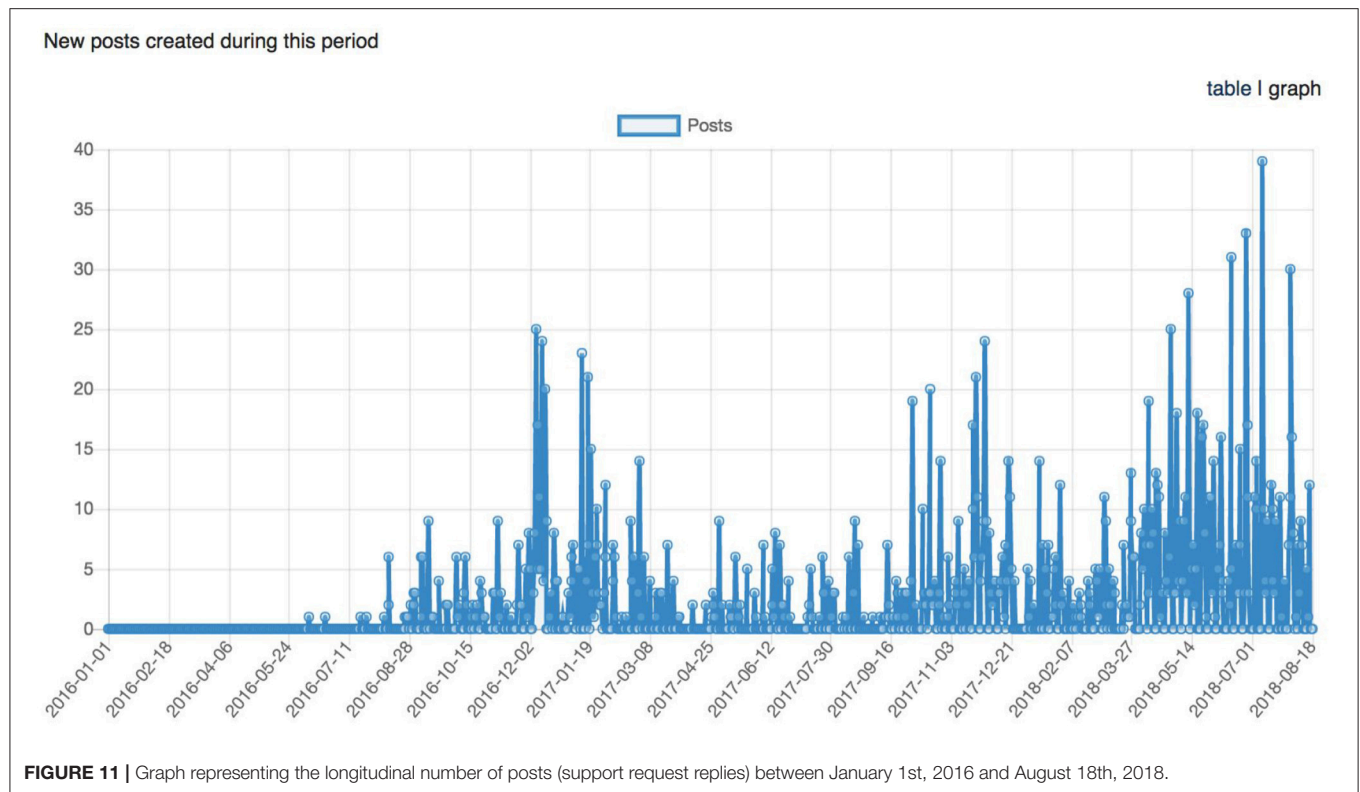
Establishing a unified methodological approach to data acquisition and processing can dramatically reduce the burden associated with calibration and harmonization (Poldrack and Poline, 2015). This setup typically requires less training as most researchers become familiar with one system for data management. Another benefit of a national infrastructure is the ability to leverage expertise of highly specialized laboratories for specific processing and analysis. Proven benefits of data sharing within a multi-site national infrastructure can be drawn from ADNI's global impact (Weiner et al., 2015). A decade since its inception, over 1000 scientific publications have been produced using this dataset (Toga and Crawford, 2015). There is a strong argument to be made that the choice of this approach has resulted not only in increased citations, but also in collaborations (Toga et al., 2017).

Despite the benefits, challenges persist in large data infrastructure and analysis (Kang et al., 2016). There is often an initial investment required in terms of effort to unify the various members of a scientific discipline, as well as a technical platform to develop (or adopt) a comprehensive infrastructure. The technical challenges can manifest themselves in a number of ways. Data management platforms need to be customizable, easily extensible, and highly interoperable. While LORIS has been designed with these features from its onset, establishing data access barriers between projects and sites continues to be improved. LORIS defines several access control for specific modules and modalities, however more granular permissions are being added (such as project-dependent access restrictions or project/site-specific permission

assignment). Exchanging data between existing sites can require establishing common data definitions and exchange protocols. Often, specialized laboratories rely on external analysis tools that generate proprietary or unstandardized output, or in the case of neuroimaging analysis tools, rely on pipelines that are either developed in-house, or placed on distributed computing resources. It then becomes imperative for the data management system to carefully: (1) interact harmoniously with external tools in terms of not only reading, but also writing the data back onto the system, and (2) integrate analyzed and derived data in a harmonized and queryable manner.

Scaling these operations across a national consortium can also be an ongoing challenge. Leveraging the full power of high performance computing environments can involve a significant learning curve both at a low-level of computing infrastructure as well as higher level issues of processing and analysis (Da Mota et al., 2014).

Issues can also result from policy requirements that might differ between sites or regions. For any study, there are specific ethical and regulatory procedures that are governed by local ethics committee, regional governing boards, and legislative privacy and ethics laws. Although LORIS is not directly governed by these procedures, there are indirect consequences which constantly present new challenges to the software. A simple example of such procedures is amendments brought to existing instruments which go through ethics approval in each region of the country simultaneously, as they are reviewed independently in each province and approved at different times. During this process, sites in different provinces administer different versions

of a single instrument, while LORIS, currently displays a single version of any instrument. Studies can certainly benefit from ethical frameworks for data sharing (Dyke et al., 2018) as these issues must be internalized when sharing data nationally (or internationally).

It should be noted that there are other systems that can curate and manage data, some of which have gained significant adoption, such as RedCap (Harris et al., 2009), a platform to create clinical and psychometric forms for web-based data entry. While this platform has been developed to simplify several aspects of data collection, it was not designed for multi-modal curation (such as imaging or genomics data). In comparison, LORIS is extensible, modular and scalable to allow for heterogeneous data types. There are numerous tools that manage parts of the curation process, but do not handle array of functionalities required for the full lifecycle of a longitudinal, multi-modal study. Other open source platforms also exist such as XNAT (Vaccarino et al., 2018) and COINS (Scott et al., 2011), each of which can handle such requirements, but have different methodologies for curation. While each of these systems could be leveraged, LORIS has recently heavily invested in open science principles and resulting data sharing capabilities.

Adhering to the *FAIR* principles (Findable, Accessible, Interoperable, and Reproducible) (Wilkinson et al., 2016) is a central tenet in building a data sharing platform (Gorgolewski et al., 2013). As CCNA continues to develop, the underpinning factors for successful data collection and sharing across a national consortia will likewise undergo enhancements, in supporting a technical infrastructure that is flexible enough to import data from other platforms, harmonize and coordinate data from different sources in a queryable manner, and seamless user-facing processes which increase the transparency and data management. To that end, CCNA has considered these concepts in its adoption of LORIS, which integrates and continues to enhance these facets in data sharing.

Future Developments

Future developments are planned in the following areas: (1) biospecimen data, (2) imaging modalities, (3) new behavioral cohorts, and (4) CCNA-specific technological advancements.

1) Biospecimens: Collection will require the following features in the CCNA data platform:

   - *Biobank:* Processed data to be imported for subject biomarkers (general health metrics, sex-related hormones, inflammation, lipid metabolism, microbiome, and oxidative stress)
   - *Genomic Browser:* Extensive genotyping[13] with summary epi/genetic data (CPG, SNP, and CNV) will be stored (and visualizable) to evaluate markers of genetic susceptibility.
   - *Brainbank:* Brain tissue pathology data for diagnosis and proteomics to be imported from Canadian ADNI Brain

Donation & Neuropathology Network (Franklin et al., 2015).

2) Imaging: Upgrades are planned to incorporate additional modalities:

   - *Electrophysiological module:* COMPASS-ND to be extended to study epilepsy. LORIS will display EEG[14], by integrating an EEG-BIDS[15] reader for standardized EEG data.
   - *Positron Emission Tomography (PET):* Extensions to support PET from Siemens High-Resolution Research Tomograph (HRRT) is currently being developed.

3) Behavioral cohorts: Clinical/psychometric data will be coded in LORIS:

   - *Functional Assessment of Vascular Reactivity (FAVR):*[16] Longitudinally investigating cognitive impairment and vascular dysfunction in AD, and small vessel disease for subjects with cerebral amyloid angiopathy, AD, and MCI.
   - *Normative Comparison & Control Group:* Cognitively intact older individuals, providing normative neuropsychological data to the COMPASS-ND battery.
   - *Sleep Study:* Identify brain mechanisms linking sleep and circadian rhythm disruption to cognitive decline and incident AD, and other dementias in older adults[17].
   - COMPASS-ND Intervention Studies:
     - *SYNERGIC* (Synchronizing Exercise Remedies in Gait & Cognition)
     - *ENGAGE* (Exploring Novel Group Activities for Geriatric Enrichment)
     - *LEAD* (Lifestyle, Exercise, And Dementia).

4) Novel Technological components: New features and technologies enhance reliability:

   - Interoperability: DataLad and Git-Annex is currently being developed to:
     - Submit metadata and images from sites to a central repository.
     - Build multi-layered user-access with prolific levels of control for sharing data.
     - Leverage metadata searching capabilities already integrated within DataLad.
     - Seamlessly link metadata with the MRI images tracked through Git-Annex[18].
     - Download images in BIDS format[19] (Gorgolewski et al., 2016).

---

[13]Using the Affymetrix UK Biobank Axiom array chip for genetic association analysis at Mt. Sinai Clinical Genomics Centre.

[14]Electroencephalography (EEG).
[15]http://bids.neuroimaging.io/
[16]https://www.ucalgary.ca/esmithresearch/projects/favr
[17]Using data collected in LORIS from the following tools: a) Actigraphy for Quantification of Sleep Architecture and Circadian Irregularity, b) WatchPAT Finger Plethysmography for Quantification of Sleep Apnea, and c) Ambulatory Polysomnography sleep recording.
[18]Scans acquired are for one CCNA and two ONDRI human phantoms. Both studies will be augmented to 3 human phantoms to measure inter-scanner variability. Data will eventually be shared openly.
[19]Brain Imaging Data Structure.

- API: Automate time consuming tasks to improve interoperability/reliability with:

  ○ Direct database queries/updates to allow external apps to use LORIS.
  ○ Enable mass uploads.
  ○ Create subject identifiers.
  ○ Extract-Transform-Load tool to import data (SQL and JSON) with user-defined rules.

- Standardization: To ensure reproducibility, several features can be improved:

  ○ Image processing containerization (e.g., tissue classification & volumetrics).
  ○ Common data elements are being leveraged increasingly.
  ○ Forms building tools[20] to abet this process.

- User Interface: Several updates will be made to optimize usability:

  ○ Dashboard will be further personalized with notifications and visualizations.
  ○ Workflow integration (intuitive sequencing of user tasks) will be customized.

CONCLUSION

CCNA's COMPASS-ND study leverages the infrastructure of LORIS, an established data management platform with the ability to harmonize, consolidate, and disseminate heterogeneous data types in a user-friendly, and robust fashion. LORIS has been fully customized to the pan-Canadian nature of CCNA, and offers the flexibility to allow for ongoing development as the study matures. This infrastructure also meets the evolving needs of the Canadian data sharing landscape, where CCNA is an exemplar of the successful efforts to consolidate data across the country to accelerate discovery in NDD research.

---

[20]Brainverse is an example.

ETHICS STATEMENT

This study was reviewed and approved by the Research Review Office of the Centre Intégré Universitaire de Santé et de Services Sociaux de Centre-Ouest-de-lÎle-de-Montréal. Participants signed an informed consent form including information on digital storage of collected data. Study is registered at ClinicalTrials.gov with the ID number NCT0340291.

AUTHOR CONTRIBUTIONS

ZM and SD contributed to the writing of this paper, contributed to the conceptualization and design of the initiative. VW, HC, and AE contributed to conception and design of the study, contributed to policy. DL contributed to the design of the figures and tables. RA-H, MS-H, and DB contributed to design and technology. JC, CH-B, J-FT, LE, TC, and P-EM contributed to the reading, revision, and approval of the submitted version of the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2018.00085/full#supplementary-material

REFERENCES

Belleville, S. (2014). Detecting early preclinical Alzheimer's Disease via cognition, neuropsychiatry, and neuroimaging: qualitative review and recommendations for testing. *J. Alzheimer Dis*. 142:4. doi: 10.3233/JAD-141470

Bennet, C. M., and Miller, M. B. (2010). How reliable are the results from functional magnetic resonance imaging? *Ann. N.Y. Acad. Sci*. 1191, 133–155. doi: 10.1111/j.1749-6632.2010.05446.x.

Campbell, T. (2017). "An online training system to enable systematic delivery of study protocols in multisite neuroimaging," in *Proceeding of: Organization of Human Brain Mapping (OHBM)*, Vancouver, BC.

Da Mota, B., Tudoran, R., Costan, A., Varoquaux, G., Brasche, G., Conrod, P., et al. (2014). Machine learning patterns for neuroimaging-genetic studies in the cloud. *Front. Neuroinform*. 8:31. doi: 10.3389/fninf.2014.00031

Das, S., Glatard, T., MacIntyre, L. C., Madjar, C., Rogers, C., Rousseau, M. E., et al. (2016). The MNI data-sharing and processing ecosystem. *Neuroimage* 124:1188–1195. doi: 10.1016/j.neuroimage.2015.08.076

Das, S., Zijdenbos, A. P., Harlap, J., Vins, D., and Evans, A. C. (2012). LORIS: a web-based data management system for multi-center studies. *Front. Neuroinform*. 5:37. doi: 10.3389/fninf.2011.00037

Duchesne, S. (2015). *The Canadian Dementia Imaging Protocol: a standardized Tool for the Study of Neurodegeneration and Aging*. Toronto, ON: ISMRM 2015.

Duchesne, S., Chouinard, I., Potvin, O., Fonov, V., Khademi, A., Bartha, R., et al. (2018). The Canadian dementia imaging protocol: Harmonizing National Cohort. *JMRI*. doi: 10.1002/jmri.26197. [Epub ahead of print].

Dyke, S. O. M., Linden, M., Lappalainen, I., Rambla de Argila, J., Carey, K., Lloyd, D., et al. (2018). Registered access: authorizing data access. 26, 1721–1731. doi: 10.1038/s41431-018-0219-y

Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U.S.A*. 113:28. doi: 10.1073/pnas.1602413113

Evans, A. C., and Brain Development Cooperative Group (2006). The NIH MRI study of normal brain development. *Neuroimage* 30:1. doi: 10.1016/j.neuroimage.2005.09.068

Fonov, V. S., Janke, A., Caramanos, Z., Arnold, D. L., and Narayanan, S., Pike et al. (2010). "Improved Precision in the Measurement of Longitudinal Global

and Regional Volumetric Changes via a Novel MRI Gradient Distortion Characterization and Correction Technique" in *International Workshop on Medical Imaging and Virtual Reality* (Berlin, Heidelberg: Springer), 324–333.

Fostering Alzheimer Society of Canada (2010). *Rising Tide: The Impact of Dementia on Canadian Society*.

Fostering reproducible fMRI Research (2017). Fostering reproducible fMRI Research. *Nat. Neurosci.* 20:298. doi: 10.1038/nn.4521.

Franklin, E. E., Perrin, R. J., Vincent, B., Baxter, M., Morris, J. C., Cairns, N. J., et al. (2015). Brain collection, standardized neuropathologic assessment, and comorbidity in ADNI participants. *Alzheimers Dement.* 11:7. doi: 10.1016/j.jalz.2015.05.010.

Gilmore, R., Diaz, M., Wyble, B., and Yarkoni, T. (2017). Progress towards openness, transparency, and reproducibility in cognitive neuroscience. *Ann. N. Y. Acad. Sci.* 1396:13325. doi: 10.1111/nyas.13325

Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., et al. (2015). Reproducibility of neuroimaging analyses across operating systems. *Front. Neuroinform.* 9:12. doi: 10.3389/fninf.2015.00012

Gorgolewski, K., Margulies, D. S., and Milham, M. P. (2013). Making data-sharing count: a publication based solution. *Front. Neurosci.* 7:9. doi: 10.3389/fnins.2013.00009

Gorgolewski, K., and Poldrack, R. (2016). A practical guide for improving transparency and reproducibility in neuroimaging research. *PLoS Biol.* 14:e1002506. doi: 10.1371/journal.pbio.1002506

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3:44. doi: 10.1038/sdata.2016.44

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42:377–81. doi: 10.1016/j.jbi.2008.08.010

Kang, J., Caffo, B., and Liu, H. (2016). Editorial: Recent Advances and Challenges on Big Data Analysis in Neuroimaging. *Front. Neurosci.* 10:505. doi: 10.3389/fnins.2016.00505

MacFarlane, D. (2014). *'Enhanced Data Querying for Neuroinformatics Databases', Proceeding of: Organization of Human Brain Mapping (OHBM)*. Hamburg.

Maumet, C., Auer, T., Bowring, A., Chen, G., Das, S., Flandin, G., et al. (2016). Sharing brain mapping statistical results with the neuroimaging data model. *Sci. Data* 3:102. doi: 10.1038/sdata.2016.102

Mohaddes, Z. (2015). *Web-Based Imaging Uploader for LORIS Web-Based Imaging Uploader for LORIS', Proceeding of: Organization of Human Brain Mapping (OHBM)*. Honolulu.

Munafo, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., et al. (2017). A manifesto for reproducible science. *Nat. Commun.* 1:1. doi: 10.1038/s41562-016-0021

Nichols, T., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke M, et al. (2016). Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20, 299–303. doi: 10.1038/nn.4500

Poldrack, R. A., and Poline, J. B. (2015). The publication and reproducibility challenges of shared data. *Trends Cogn. Sci.* 19:2. doi: 10.1016/j.tics.2014.11.008

Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Front. Neuroinform.* 6:9. doi: 10.3389/fninf.2012.00009

Richesson, R. L., and Nadkami, P. (2011). Data standards for clinical research data collection forms: current status and challenges. *J. Am. Med. Inform. Assoc.* 18, 341–346. doi: 10.1136/amiajnl-2011-000107

Rogers, C. (2015). "LORIS Neuroinformatics platform for Imaging Genetics LORIS Neuroinformatics platform for Imaging Genetics," in *Proceeding of: Organization of Human Brain Mapping (OHBM)*. Honolulu.

Scott, A., Courtney, W., Wood, D., de la Garza, R., Lane, S., King, M., et al. (2011). COINS: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets. *Front. Neuroinform.* 5:33. doi: 10.3389/fninf.2011.00033

Sherif, T., Kassis, N., Rousseau, M. É., Adalat, R., Evans, A. C., Adalat, R., et al., (2014). CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. *Front. Neuroinform.* 8:54. doi: 10.3389/fninf.2014.00054

Statistics Canada (2016). '*Census Topic: Age and Sex*'. Statistics Canada: Ottawa, ON. Available online at: http://www.census.gc.ca/census-recensement/2016/rt-td/as-eng.cfm

Toga, A. W. (2012). The clinical value of large neuroimaging data sets in Alzheimer's Disease. *Neuroimaging Clin. N. Am.* 22:101–109. doi: 10.1016/j.nic.2011.11.008

Toga, A. W., Bhatt, P., and Ashish, N. (2017). Global data sharing in Alzheimer's disease research. *Alzheimer Dis. Assoc. Disord.* 30, 160–168. doi: 10.1097/WAD.0000000000000121

Toga, A. W., and Crawford, K. L. (2015). The Alzheimer's Disease Neuroimaging Initiative informatics core: a decade in review. *Alzheimer's Dement.* 11, 832–9. doi: 10.1016/j.jalz.2015.04.004

Turner, B., Paul, E. J., Miller, M. B., and Barbey, A. K. (2018). Small sample sizes reduce the replicability of task-based fMRI studies. *Sci. Data*. 3:102. doi: 10.1038/s42003-018-0073-z

Vaccarino, A. L., Dharsee, M., Strother, S., Aldridge, D., Arnott, S. R., Behan, B., et al. (2018). Brain-CODE: a secure neuroinformatics platform for management, federation, sharing and analysis of multi-dimensional neuroscience Data. *Front. Neuroinform.* 12:28. doi: 10.3389/fninf.2018.00028

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., et al. (2015). Impact of thes Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014. *Alzheimer's Dement.* 11, 865–884. doi: 10.1016/j.jalz.2015.04.005

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:18. doi: 10.1038/sdata.2016.18

Zuo, X. N., Anderson, J. S., Bellec, P., Birn, R. M., Biswal, B. B., Blautzik, J., et al. (2014). An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci. Data* 1:140049. doi: 10.1038/sdata.2014.49

Zuo, X. N., and Xing, X. X. (2014). Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neurosci. Biobehav. Rev.* 45, 100–118. doi: 10.1016/j.neubiorev.2014.05.009

frontiers
in Neuroinformatics

Integration of *"omics"* Data and Phenotypic Data Within a Unified Extensible Multimodal Framework

Samir Das[1,2†], Xavier Lecours Boucher[1,2*†], Christine Rogers[1,2], Carolina Makowski[1,2,3], François Chouinard-Decorte[1,2], Kathleen Oros Klein[4,5], Natacha Beck[1,2], Pierre Rioux[1,2], Shawn T. Brown[1,2], Zia Mohaddes[1,2], Cole Zweber[1,2], Victoria Foing[1,2], Marie Forest[4,5], Kieran J. O'Donnell[3,4], Joanne Clark[4], Michael J. Meaney[3,4], Celia M. T. Greenwood[4,5] and Alan C. Evans[1,2]

[1] McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, Montreal, QC, Canada, [2] Montreal Neurological Institute, McGill University, Montreal, QC, Canada, [3] Douglas Hospital Research Centre, McGill University, Montreal, QC, Canada, [4] Ludmer Centre for Neuroinformatics & Mental Health, McGill University, Montreal, QC, Canada, [5] Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, QC, Canada

Analysis of "*omics*" data is often a long and segmented process, encompassing multiple stages from initial data collection to processing, quality control and visualization. The cross-modal nature of recent genomic analyses renders this process challenging to both automate and standardize; consequently, users often resort to manual interventions that compromise data reliability and reproducibility. This in turn can produce multiple versions of datasets across storage systems. As a result, scientists can lose significant time and resources trying to execute and monitor their analytical workflows and encounter difficulties sharing versioned data. In 2015, the Ludmer Centre for Neuroinformatics and Mental Health at McGill University brought together expertise from the Douglas Mental Health University Institute, the Lady Davis Institute and the Montreal Neurological Institute (MNI) to form a genetics/epigenetics working group. The objectives of this working group are to: (i) design an automated and seamless process for (epi)genetic data that consolidates heterogeneous datasets into the LORIS open-source data platform; (ii) streamline data analysis; (iii) integrate results with provenance information; and (iv) facilitate structured and versioned sharing of pipelines for optimized reproducibility using high-performance computing (HPC) environments via the CBRAIN processing portal. This article outlines the resulting *generalizable "omics" framework* and its benefits, specifically, the ability to: (i) integrate multiple types of biological and multi-modal datasets (imaging, clinical, demographics and behavioral); (ii) automate the process of launching analysis pipelines on HPC platforms; (iii) remove the bioinformatic barriers that are inherent to this process; (iv) ensure standardization and transparent sharing of processing pipelines to improve computational consistency; (v) store results in a queryable web interface; (vi) offer visualization tools to better view the data; and (vii) provide the mechanisms to ensure usability and reproducibility. This framework for workflows facilitates brain research discovery by reducing human error through automation of analysis pipelines and seamless linking of multimodal data, allowing investigators to focus on research instead of data handling.

Keywords: workflow, omics analysis, integrative neuroscience, reproducibility, database, HPC, genomics, biostatistics

INTRODUCTION

Genomic analysis and bioinformatics have undergone a technological revolution over the past few decades, one that holds great promise for scientific discovery but also poses significant big-data challenges. To increase accessibility for researchers with varying levels of informatics expertise, the "Big Data" components of "*omics*"[1] analyses need to be integrated into an automated and seamless workflow. To this end, in 2015 the Ludmer Centre for Neuroinformatics and Mental Health[2] created a genetic/epigenetic working group composed of three member institutions of McGill University: (i) the Douglas Mental Health University Institute, focusing on biological questions; (ii) the Lady Davis Institute at the Jewish General Hospital, focusing on tools for statistical analysis; and (iii) the McGill Centre for Integrative Neuroscience at the Montreal Neurological Institute (MNI), responsible for the neuroinformatics infrastructure (Das et al., 2016, 2017).

The goal of the working group is the integration of "*omics*" data into the LORIS data platform[3], a web-based open-source data and project management platform (Das et al., 2011) to streamline analysis, integrate results, and facilitate structured sharing for optimized reproducibility, using high-performance-computing (HPC) environments via CBRAIN[4] (Sherif et al., 2014), a web-based open-source platform that allows computationally intensive analyses of data by connecting researchers to HPC facilities. The pilot use-case for multimodal "omics" workflow integration focused on analysis outputs from the *Methylation450k*[5] pipeline, a functional normalization pipeline for epigenomic data from a Ludmer Centre-based study.

This article describes an extensible and adaptable framework that addresses critical gaps in integrating "omics" data with multi-modal phenotypic datasets (imaging, behavioral, clinical, demographic, . . .) using HPC and databases, while leveraging standardization and automation to provide GUI-based workflows for less technical researchers. Analysis of data, specifically genomic or imaging, can involve multiple parallel paths. These workflows typically begin with the processing of biological samples, followed by quality control and analysis using data-specific pipelines, and culminate in querying and visualization of summary data. The complexity of such analyses often requires a framework that can comprehensively integrate these steps across data modalities, an element that is currently lacking in many existing "*omics*" toolboxes and workflows (Kanwal et al., 2017).

In designing such a framework, it is also important to consider features that would simplify and strengthen effective data sharing mechanisms, especially as we enter the era of Open Science. The processing of raw data is often performed by third-party platforms, whereby the resulting files are processed using one or more bioinformatic pipelines by the host laboratories.

One of the inefficiencies of this model is that each processing step typically generates a new version of the dataset, which is often stored on a local workstation or distributed across multiple drives. As quality control and post-processing tasks remove aberrant values, additional versions can multiply across storage systems, but without having sufficient transparency in the options or environment parameters used in the execution to generate each version (Glatard et al., 2015). Not surprisingly, this also leads to ineffective data-sharing, whereby it becomes unclear which copies of the data contain the most comprehensive and accurate information, requiring researchers to sift through redundant data.

A few systems have been created, such as the *Galaxy* platform for genomic data (Afgan et al., 2016, 2018) to integrate biological data and streamline genetic analysis (Kanwal et al., 2017). Many software platforms exist for sharing workflows to capture and promote the execution of reproducible analyses, such as Jupyter notebooks[6]. While such models seek to increase reproducibility in computational biology, they do not prioritize cross-modal data integration. Importantly, the field would benefit from a structured workflow that links organized cross-sectional or longitudinal multimodal data (genetics, imaging, behavioral) with HPC platforms for analysis (Poldrack et al., 2017).

We have leveraged existing architectures to create a model that aims to abstract the complexities of multi-modal processing and analysis. This combined framework builds upon systems documented in previous publications (Das et al., 2016, 2017) and integrates additional technologies and feature-layers to support an approach that prioritizes the: (i) integration of heterogeneous biological data with multi-modal datasets (imaging, clinical, demographics and behavioral); (ii) automation in launching analysis pipelines on HPC platforms; (iii) removal of technical barriers that are inherent to this process (Pool and Esnayra, 2000); (iv) standardization and transparent sharing of processing pipelines to improve computational consistency; (v) storage of results into a queryable web interface; (vi) feature rich visualization tools; and (vii) provision of mechanisms to ensure usability and reproducibility. The result is a streamlined approach for cross-modal analysis (such as imaging genetics) that also promotes the FAIR principles (Findable, Accessible, Interoperable and Reproducible) for data sharing (Wilkinson et al., 2016). The framework presented in this article can be used by researchers interested in integrating "*omics*" data with other multimodal datasets, such as those utilized in behavioral and/or imaging genetics projects, and can be readily modified to accommodate the specific needs of other users and projects.

MATERIALS AND METHODS

The goal of this "*omics*" framework is to take individual processing and analysis tasks, including any manual steps that might already exist, and integrate them into a more automated model that leverages: (i) standardization and harmonization tools; (ii) HPC resources; and (iii) application programming

---

[1]Such as transcriptomics, proteomics, blood sugar, anthropometry, etc.
[2]http://ludmercenter.ca
[3]http://www.loris.ca
[4]http://mcin-cnim.ca/technology/cbrain/
[5]https://github.com/GreenwoodLab/methylation450KPipeline

[6]http://jupyter.org/

interface (API) interoperability for automation between the existing platforms. In this section, we describe the components of software and platforms, and recent extensions, which together support workflows for processing and transferring "*omics*" data.

The complexities of cross-modal workflows in "*omics*" analyses is a significant challenge for researchers given that such workflows are difficult to automate and require regular user intervention, support and maintenance. Tool development and integration at iterative stages of development is time-consuming and mandates thorough testing to successfully build a workflow. To this end, identifying the labor-intensive steps (file transfers, versioning, user access, etc.) of a data processing workflow and automating them is an essential priority.

Building a generalized framework by extending the MNI ecosystem's combined platform of LORIS and CBRAIN starts with populating the LORIS database with participant data for all modalities (such as behavioral, imaging and "*omics*"). For the two systems to communicate and exchange data as input or output of a given pipeline, a shared space must be defined. (This role can be served by a CBRAIN *DataProvider*, accessible

to the LORIS filesystem). That is followed by the installation of tools on CBRAIN such that they can be launched on HPCs. Finally, customizations and extensions to LORIS can support new formats of data. **Figure 1** shows the cyclical flow of data between LORIS and CBRAIN, whereby stored datasets are processed and their outputs returned as results.

A typical use-case begins with biological samples and phenotypic data collected during a subject's visit. The biological/phenotypic samples are then processed on-site or shipped to a specialized facility for genomic analysis or image capture, after which raw data files are created and made available for statistical and/or bioinformatics analysis. Files containing raw data are stored in a LORIS database and then subsets are queried, selected and sent to CBRAIN to be processed by an analysis tool. The output is returned back to LORIS for storage along with its provenance metadata from the processing task. Summary and aggregate data can be parsed and explored through various LORIS modules and then queried to create new datasets linked to provenance metadata. This model allows for iterative processing as data selections can be resubmitted from LORIS for further processing and
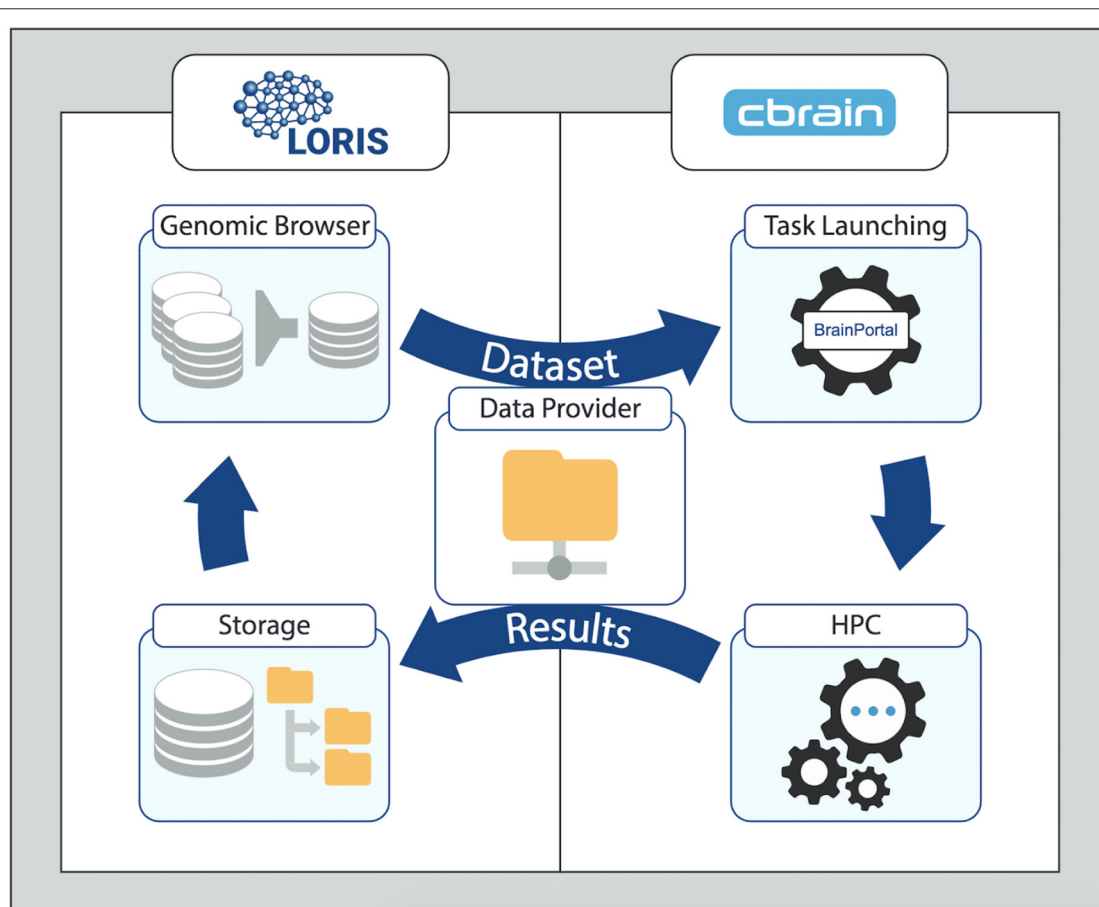
FIGURE 1 | Generalized Workflow cycle between the LORIS data-management platform and the CBRAIN processing platform. Data from LORIS (Storage) can be queried and filtered (Genomic Browser and other tools) to select a set of variables and/or files. The newly created dataset is then transferred to the CBRAIN DataProvider for processing (Task Launching) and analysis (high-performance computing, HPC). The output is synced back to LORIS with the provenance data. Results can be examined and a new iteration can begin with the added derived variables. For stepwise details of this model, please see **Figure 2** in "Results" section.
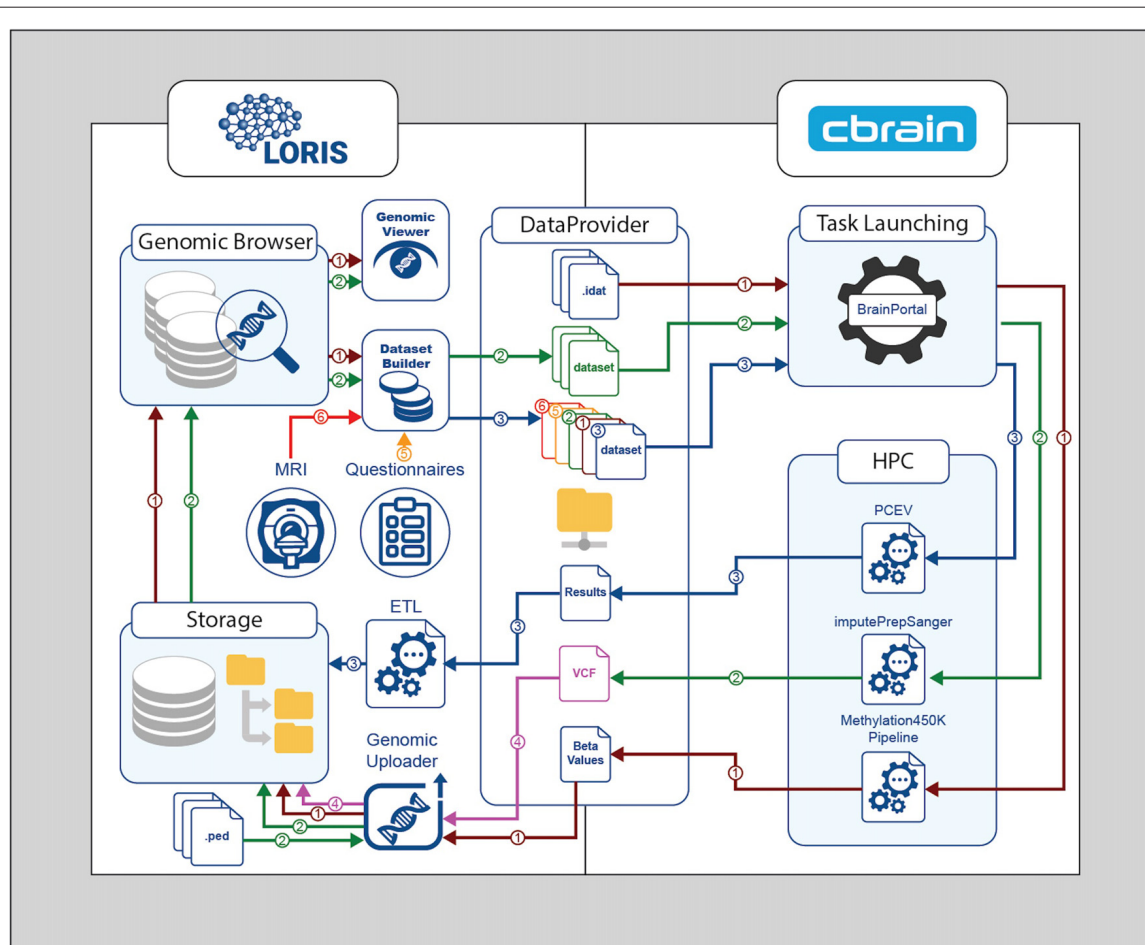
FIGURE 2 | Genomic processing cycle between LORIS and CBRAIN through the DataProvider. Methylation450K pipeline—Brown path (1): IDAT files are transferred to the DataProvider, then the methylation normalization pipeline is launched. The Beta-values output file is returned to the DataProvider, and then loaded into LORIS using the Genomic Uploader. The inserted results can be browsed or visualized in the Genomic Browser module. ImputePrepSanger pipeline—Green path (2): PLINK files are added to LORIS via the Genomic Uploader, selected in the DatasetBuilder, and sent to CBRAIN for the imputePrepSanger tool to be run. The resulting Variant Call Format (VCF) output file is stored in LORIS—Pink path (4). Statistical analysis—Blue path (3): using the DatasetBuilder module in LORIS, data from any source (Orange path (5), Red path (6)) can be packaged in a new dataset and sent to CBRAIN via the DataProvider for statistical analysis using (e.g.,) the principal component of explained variance (PCEV) pipeline.

analysis tasks via CBRAIN, with derived results returned once again into LORIS for storage and dissemination. It should be noted that a specific use-case will be demonstrated in the Results section that focuses on genomic and epigenomic data; however, similar procedures would apply for other "*omics*" data types.

To illustrate this framework with a genomic processing workflow, the relevant components of the LORIS and CBRAIN platforms (and feature extensions) are described below. Also outlined are the structural design elements facilitated by RESTful[7] API interoperability between the two systems including: (i) the data transfer mechanisms; (ii) the abstraction of data organization; and (iii) the pipeline execution flow. Key

auxiliary components and technologies interfacing with these platforms are integral to the multimodal framework, including containerization of pipelines, visualization of genomic and epigenomic data and NoSQL data storage.

LORIS Data Platform

The LORIS platform is the entry point for data in most workflows deployed on this integrated framework. LORIS can house data at various stages of the processing lifecycle, and can typically be customized with import pipelines to accept and validate files of any type. Imported files can then be parsed to extract and store any relevant values in relational database tables, which are accessed by web-facing front-end modules. For large files, the filesets themselves will be organized on the LORIS data partition, and linked by their file paths from individual database-table entries, which serve as pointers to the data location on the

---

[7]Representational State Transfer (REST) is an software architecture style compliant with Hypertext Transfer Protocol (RFC 2616) where each url is a resource that can be interact with using verbs (GET, PUT, POST, DELETE, etc.).

server. Metadata for these files can also be stored in database tables in a key-value pair format, which is also an extensible structure that accepts any data format. File paths and metadata are easily accessible via LORIS' front-end modules, through which users can peruse, filter, visualize and retrieve these datasets for download or export to other systems via the user-friendly web interface. Later in under the "LORIS Genomic Browser" section, we expand upon new "*omics*" features in LORIS.

CBRAIN

CBRAIN's web-based portal for the Compute Canada[8] network enables user-friendly deployment and execution of pipelines across the Canadian HPC grid. For LORIS to launch a data processing task[9] through CBRAIN, the interface between these systems must define the expected types and formats for both inputs and outputs.

Several key CBRAIN features support the workflow model across platforms. First, data storage and transfers are handled by a *DataProvider* (a designated file server space which connects to CBRAIN and the HPC grid), which caches and tracks data files across the HPC network. Second, CBRAIN's *ToolConfiguration* profile enables rapid setup and user-friendly re-use of a scientific tool, describing where and how it is available on the supercomputer clusters, as well as defining the cluster setup parameters (environment setup, CPUs used, queue name, etc.) and input parameters required for executing the tool.

The *ToolConfiguration* can be automatically generated in CBRAIN through a Boutiques descriptor (Glatard et al., 2018) which provides a standard JSON protocol for defining the command-line and input and output variables for pipeline execution. Typically, this initial setup needs to be configured only once, thereafter allowing for re-use of the same software setup by providing the proper input parameters. Together, the *DataProvider* and *ToolConfiguration* abstract the infrastructural complexities of data storage, transfer and processing parameters for the user while promoting transparency and reproducibility.

While CBRAIN supports the direct installation of pipelines for execution on HPC clusters, it has also introduced support for container technologies to specify the environment and package versions for optimally pre-defined execution of such pipelines.

LORIS DataProvider for CBRAIN

The *DataProvider* acts as a shared file system, such that CBRAIN and LORIS can interoperate with file-level read and write access of both the data and metadata. On the CBRAIN side, files are read from the LORIS *DataProvider* repository and made available to the HPC network. Once processing has been completed on the HPC grid, results from the pipeline execution on CBRAIN are written to the LORIS *DataProvider*, and subsequently recognized and imported back into the LORIS database and file system.

To make the file system interaction easier for LORIS' web application, a dedicated directory on the LORIS server is designated as the *DataProvider*. Both CBRAIN and LORIS can read and write to this directory, which effectively allows for communicating datasets between platforms along with accompanying metadata.

Preparation of Pipelines (Containers)

To facilitate the flexible and reproducible integration and deployment of new tools across different HPC resources, CBRAIN and other execution platforms support containerization technology such as Docker[10] and Singularity[11]. A container encapsulates the setup of the processing environment as well as any specific support packages that are needed, thereby making installation of software architecture independent, which improves reproducibility of analysis. Typically, an accompanying container description file[12] describes every step necessary to construct the container. This provides the benefit of organizing and recording each aspect of the pipeline, and facilitates transparency in defining the runtime environment in a shareable, versionable document.

Additionally, by documenting the input parameters for the pipeline, specific aspects of the pipeline run can be adjusted and tracked in a controlled manner ensuring that all other factors stay the same, such as running the same pipeline using a different R package for functional normalization. For instance, the *Methylation450k* pipeline, which provides quality control (QC) and functional normalization of the Illumina 450k beadchip array data, currently integrates the *funNorm* (Fortin et al., 2014) R package. However, the flexibility offered by container-defined plug-ins and parameters enables a user to rapidly relaunch the same pipeline on a similar R package *funtooNorm* (Oros Klein et al., 2016), providing a clearly documented trace of provenance for comparison of results between the two normalization algorithms.

Another example is the *imputePrepSanger*[13] pipeline from the Ludmer Centre. This tool prepares PLINK genotype files to be sent to the Sanger Institute's online Imputation Service[14] by performing quality control, adjusting the positions and strand alignment of PLINK files, then converting them to VCF[15] for submission to the Sanger server. The pipeline execution parameters were defined in a container on CBRAIN.

A third pipeline, *principal component of explained variance* (*PCEV*)[16], was prepared to run a dimension-reduction algorithm to explain a maximum of variance in a response vector governed by a set of covariates. Specifically, this tool can be run multiple times, using different genomic-ranges to provide

---

[8]www.computecanada.ca

[9]A task is an instance of a tool running on CBRAIN where a tool is any piece of software that take inputs and generates outputs installed on CBRAIN.

[10]Docker containers are units of processing where tool versions, an environment (OS), and sequences of operations can be reproduced on any system.

[11]Singularity is another container technology that has been privileged over Docker on HPC units served by CBRAIN.

[12]Container description files are versioned text files that contain the recipe to (re)build a given container image; they present themselves as a sequence of shell commands.

[13]https://hub.docker.com/r/eauforest/imputeprepsanger/

[14]https://www.sanger.ac.uk/science/tools/sanger-imputation-service

[15]Variant Call Format. A specification to encode genetic variations in a text file.

[16]https://github.com/GreenwoodLab/pcev_pipelineCBRAIN

a new set of methylation Beta-values and genomic variants and/or a different set of covariates from behavioral and imaging metrics.

This model can be adapted for larger workflows, enabling reproducible execution of pipelines as a generalizable concept that could be applied to many use-cases. Examples include automatically running a piece of software when new data are available, performing quality control or validation, or ensuring that users run the same tool version in the same runtime conditions throughout the lifecycle of a study.

CBRAIN/LORIS Hooks

In order for data to pass seamlessly from one system to another, communication occurs between LORIS and CBRAIN using a RESTful (web) API for requests, and the *DataProvider* for data transfer and registration. A client for the CBRAIN API written in the PHP programming language has been created using SwaggerEditor[17] with a schema[18] following OpenAPI specification v2.0, which allows LORIS to look at available files and tools on CBRAIN. This PHP client also abstracts the handling of HTTP GET and POST requests which trigger the creation of new processing tasks on the HPC grid via CBRAIN. For a newly generated dataset, LORIS starts by registering the files in CBRAIN, making it possible to run relevant tasks. The type of the tasks, their parameters, and input files are then communicated through the API to CBRAIN, which launches them.

A LORIS process running in the background monitors a CBRAIN task's status. The task progress can be followed from LORIS' Server Processes Manager module. Capture of logs from data insertion and the task's output from CBRAIN, as well as queries used to generate the new dataset, will be stored in a header file or in the database. This way, at the time of publication, all information describing provenance can be formatted in a file compliant with the Neuroimaging Data Model (NIDM; Keator et al., 2016).

LORIS Genomic Browser

The *Genomic Browser* module (Rogers et al., 2015) is the principal LORIS component for visualization, querying, validation and storage of genomic and genetic data, and is part of an open-source feature set available on GitHub. This module enables browsing of single-nucleotide polymorphisms (SNPs) and copy number variants (CNVs) data, but has been expanded for this application to allow exploration of epigenomic data using the same functionalities. Any filtered subset of data can be downloaded and exported for further analysis, in addition to being passed to the visualization utilities embedded within the module. This allows for a genomic dataset to be viewed alongside behavioral and imaging data. The system includes functionality for viewing, filtering and linking of summary genetic data [CNV, SNP and other results from genome wide association

studies (GWAS)]. Links to reference databases (UCSC genome browser[19], dbSNP) have also been added.

Genomic Uploader

Genomic data is loaded into LORIS from raw or processed files using the web interface in the *Genomic Uploader*. This rudimentary upload tool is provided to facilitate loading and linkage of data files and records in the database. In addition to maintaining a reference for uploaded files, the uploader creates relations between inserted values, their annotations, and the study subject they belong to within the file header. When the file type fits a study's expected types, user-defined scripts tailored to the genotyping platform of interest are provided. Inserted data are accessible and browsable in the module's tabs.

Profile Summary Tab

The first tab of the *Genomic Browser* is called the Profile Summary tab and provides researchers with a high-level understanding of the data types available for individual subjects as well as summary statistics. This tab displays a sortable view of this information and enables filtering by population of interest and subject metadata for available genomic datasets stored in LORIS. The number of CNVs and SNPs or methylation CpGs found for each subject can be reviewed, filtered and sorted at a glance. By applying filters based on cohort or phenotypic gender, users can view these summary statistics for a sub-population of interest.

Genomic Browser Tabs: CNV, SNP, Methylation

Other tabs of the *Genomic Browser* provide subject-specific results for each data type from various epi-genomic and -genetic analyses (e.g., for CNV, SNP, or methylation results). When pipeline outputs are imported into LORIS and matched with an expected file format, the appropriate tab is automatically populated with data that is visible to the user. Each tab enables filtering by specific genomic regions around genes of interest or shared properties.

Genomic Viewer

An additional tab within the *Genomic Browser* was added to provide advanced exploration for epigenomic data, with genomic data aligning these points along the genome in superimposed tracks. This visualization technique is found in many domain-specific softwares and was developed for LORIS using *React.js*[20] components for each track to dynamically render as page elements. Interactive display features are also created using *D3.js*[21] visualization libraries for HTML5 canvas and SVG image generation. These combined technology layers can efficiently manage large volumes of data.

In our example implementation, the *Methylation450k* normalization pipeline produced a single output file containing Beta-values for all samples across all probes which were uploaded as a batch into LORIS via CBRAIN. Upon loading
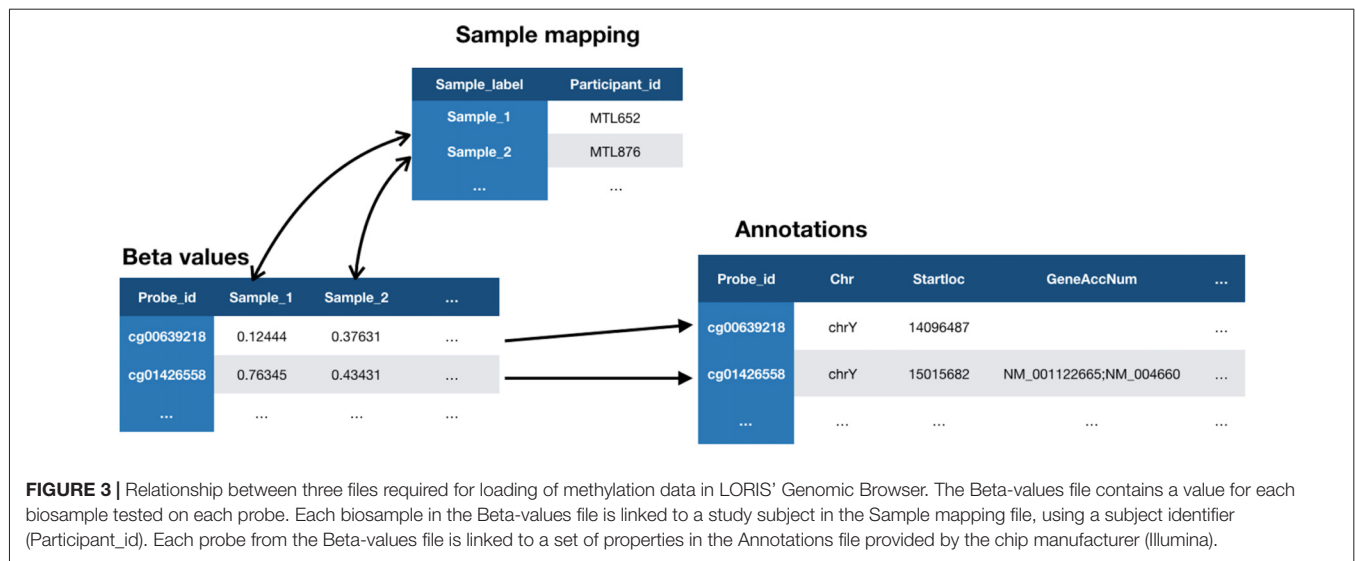
---

[17]https://editor.swagger.io/
[18]https://github.com/aces/cbrain/blob/master/BrainPortal/public/swagger/cbrain-4.5.1-swagger.yaml

[19]https://genome.ucsc.edu/
[20]https://reactjs.org/
[21]https://d3js.org/

FIGURE 3 | Relationship between three files required for loading of methylation data in LORIS' Genomic Browser. The Beta-values file contains a value for each biosample tested on each probe. Each biosample in the Beta-values file is linked to a study subject in the Sample mapping file, using a subject identifier (Participant_id). Each probe from the Beta-values file is linked to a set of properties in the Annotations file provided by the chip manufacturer (Illumina).

Beta-values[22] into LORIS, each probe must be associated with an annotation record provided by the manufacturer of the array (Illumina). These annotation records are stored in the *genomic_cpg_annotation* database table which is populated using a script[23] provided in the LORIS codebase. Each probe is then linked to a sample ID and a corresponding subject in the database. A mapping file is used in this process to link each sample to the subject ID.

The MySQL database contains paths to the three files (**Figure 3**) that comprise the dataset: the Beta-values file, the sample mapping file, and an annotation file. Once registered in the database, any type of biological data can be linked to behavioral and imaging data for each subject using their subject ID. The relationship between subjects and their biological data records is defined at the sample level, allowing for metrics from duplicate biosamples to be linked to the same subject. Once this link has been established, visualization tools within the *Genomic Browser* are used to look at available data for regions of interest on the genome. The SNP and CpG locations are aligned with histone marks or CpG islands, providing additional information about genomic features and regulatory interactions in the same locus.

Building Cross-Modal Queries

Within LORIS, a prototype *DatasetBuilder* module allows users to create new datasets by joining filtered genomic data with phenotypic data and/or imaging files queried from the Data Querying Tool (DQT; MacFarlane et al., 2014), to rapidly handle large datasets on the scale of genomic results, and provide that data to the user-facing frontend.

Both the DQT and the *DatasetBuilder* are built upon CouchDB, a file-based NoSQL database that provides a REST

API for querying and filtering prebuilt data views. The views are generated by applying MapReduce[24] algorithms, where each document is transformed using a mapping function and then summarized by the reducer function to create an indexed set of key-value pairs.

The *DatasetBuilder* processes an HTTP request issued for a specified *genomic_range* or DNA chip probe identifier, and retrieves all data records corresponding to the indexed range. For each record returned, a filter function identifies the samples of interest and extracts the Beta-values for display in the module. The subject IDs corresponding to these records are identified and a request is made to run an existing query saved in the DQT to select other phenotypic variables of interest (e.g., demographics, behavioral measures, etc.). The phenotypic datasets returned by the DQT are then joined with the biosample subject data to produce a combined dataset of fields across all modalities. These results are exported as CSV files to the CBRAIN *DataProvider* for further processing.

RESULTS

To demonstrate this framework for "omics" workflows, a specific "use-case" implementation from the Ludmer Centre working group is discussed, which includes genotyping, methylation assessments and typical phenotypic data (age, sex, etc.). The data was collected and derived from human subjects participating in a longitudinal study conducted by Ludmer researchers at the Douglas Mental Health University Institute in Montreal. The *Methylation450k* pipeline was run on the study dataset, and the outputs transferred via CBRAIN to LORIS. Using the *Genomic Browser* in LORIS, users could then query, select visualize and download data across phenotypic and epigenomic datasets. Further containers were created for additional pipelines such as *PCEV*, and installed and launched on the HPC grid

---

[22]Beta-values represent levels of DNA methylation at a given probe (CpG) and range from 0 to 1, representing 0%–100% DNA methylation at a given site.
[23]https://github.com/aces/Loris/blob/master/modules/genomic_browser/tools/HumanMethylation450k_annotations_to_sql.py

[24]Category of functions that split a problem into parallelizable parts so it can run on multiple threads and/or distributed computers.

via CBRAIN. The output of each task is transferred to the DataProvider and can then be loaded in the database, where it is linked to the provenance history of the task parameters and inputs.

Throughout this example, end-users seeking to reproduce, review, and use the data and metadata have the ability to use this complex pipeline with little technical knowledge through transparently accessible computing, negating the need to focus on: (i) transferring files across servers and clusters; (ii) managing versions; (iii) controlling user access; (iv) connecting with HPC units; (v) launching tasks; (vi) tracking progress; and (vii) capturing processing status, parameters and results. Once the outputs are stored and accessible in the main data platform, users can explore their data across modalities using additional web-based tools.

Loading Raw Files Into the Relational Database

In a typical implementation of a workflow in this framework, raw data is imported into the LORIS data system and stored or linked in its relational tables. For the Ludmer Centre's pilot implementation, data on 328 subjects from the Maternal Adversity, Vulnerability and Neurodevelopment study (MAVAN; O'Donnell et al., 2014) were processed and stored in LORIS. Data collected and stored on these subjects included questionnaires, demographic and phenotypic information and imaging scans.

Biosamples from each subject were collected, stored, and then processed by a third-party genotyping facility. The resulting IDAT files were run on the the *Methylation450k* pipeline and then transferred via a project-specific *DataProvider* to CBRAIN.

This output was stored on CBRAIN as a large (CSV) matrix of 328 columns (samples) and 450,000 rows (probes) of Beta-values. This file was transferred to the LORIS server via SFTP and its contents were loaded into LORIS along with the Illumina annotation records. The *Genomic Uploader* module in LORIS was used to do this, creating a bio-sample record that associated over 450,000 values with each subject in LORIS. As a result, more than 147,600,000 values were stored in the *genomic_cpg* table.

In parallel, SNP data from these processed biosamples were transferred in the form of PLINK files (.PED and .MAP format) from a private FTP site to the LORIS server. These data points were transformed via PLINK commands and loaded into the LORIS database. SNP annotations were taken from the *dbSNP*[25] resource database to build filters on individual SNP values in the *Genomic Browser*.

Selection, Filtering and Visualization Within the LORIS Data Platform

With several modalities of data for the population now stored in LORIS, the *Genomic Browser* and *Genomic Viewer* were used to select and filter variables of interest across data types. With the *DatasetBuilder*, new datasets were then defined by joining across other modalities, and can serve as input for later processing tasks to be launched on the HPC grid via CBRAIN.

Genomic Browser

For researchers, a key feature is linking cross-modal data using a simple interface with querying, visualization, and

---

[25]https://www.ncbi.nlm.nih.gov/projects/SNP/

FIGURE 4 | LORIS Genomic Browser: Profiles tab. Filter applied to search for subjects based on Site, Gender, Subproject, External ID and the availability of genomic data. In the table, detailed subject data can be accessed by clicking on the link that appears on each item.
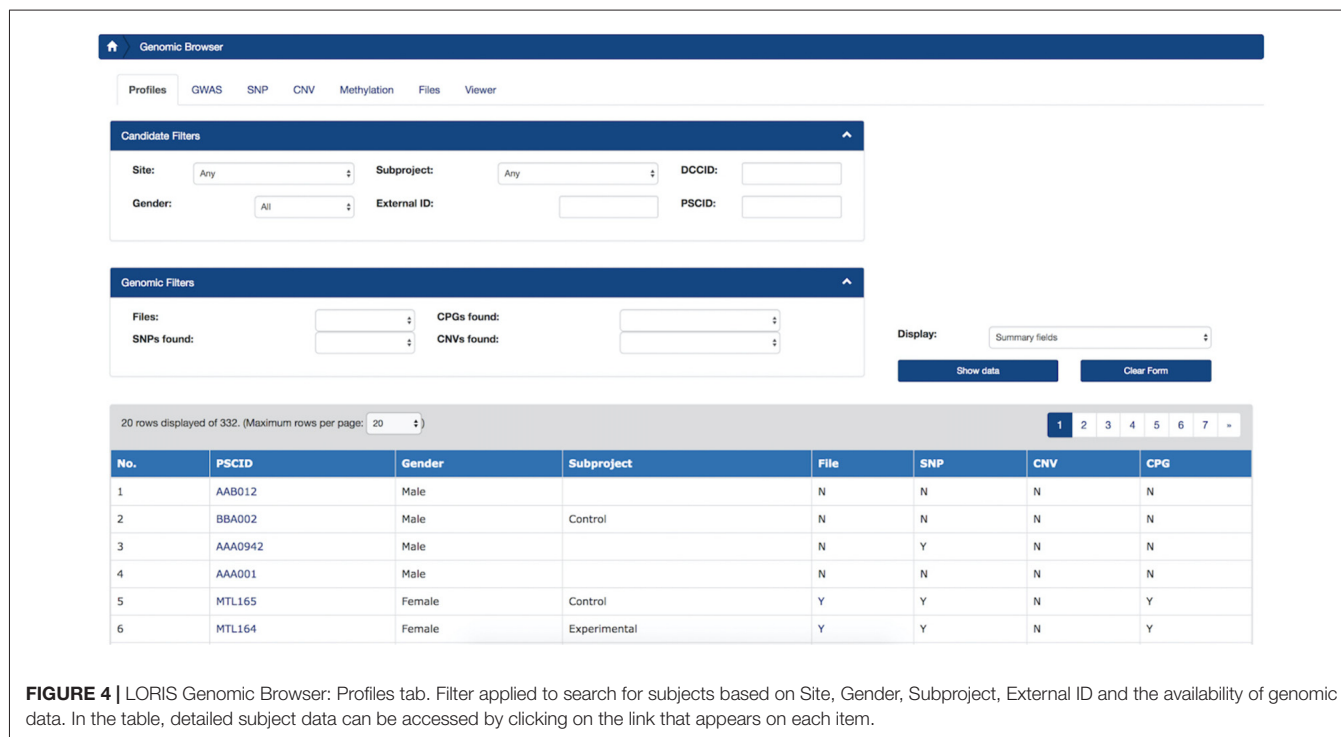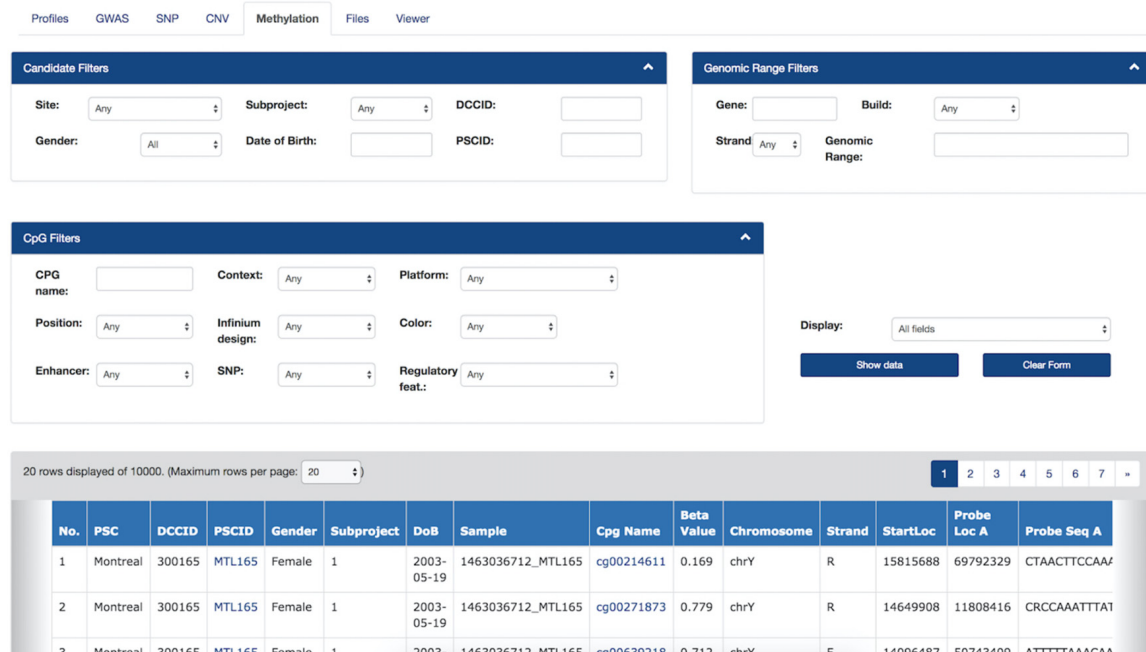
FIGURE 5 | Filters and Methylation Beta-values in the Genomic Browser. Filters are applied on subject information, genomic range and the probe's annotations. The filtered data view can be downloaded as a CSV file. Hyperlinks on each "CpG Name" column cell will bring the user to the online UCSC genome browser[10], which provides detailed information about a given CpG from the most recent human genome build version.

search capabilities. The *Genomic Browser* (**Figures 4**, **5**) enabled filtering values by their annotations, such that genomic data was uploaded and imported into LORIS, and then analyzed and visualized.

Genomic Viewer

For each subject's methylation data, the *Genomic Viewer* tab (**Figure 6**) displayed detailed genomic information. In this tab, users could view aggregated CpG Beta-value distributions
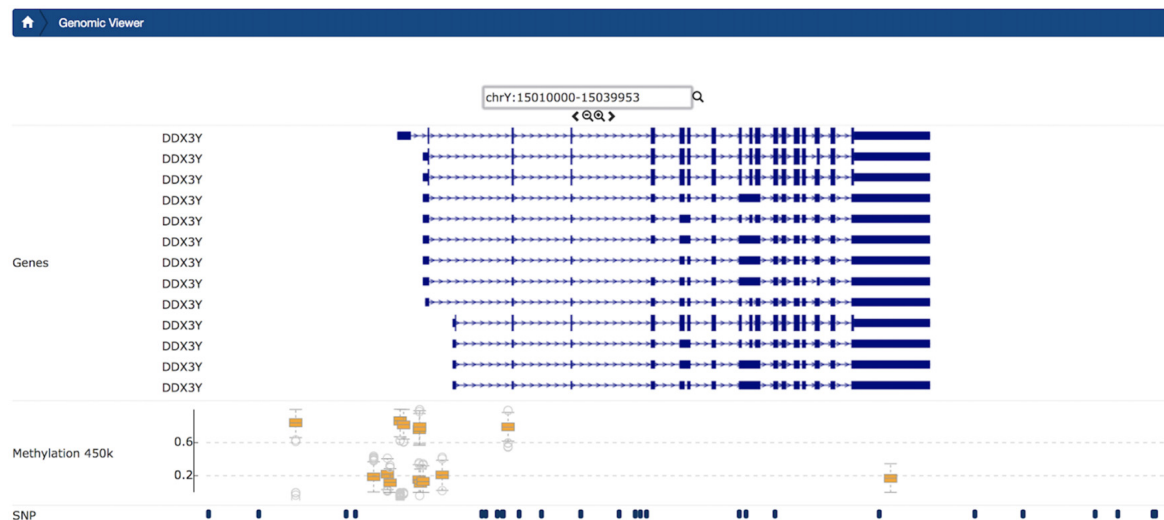
FIGURE 6 | Example Genomic Viewer shows the context for single-nucleotide polymorphisms (SNPs) and CpGs in a small region of CpGs. Visualized context includes features from external sources, for chromosome Y from position 15010000 to 15039953. The upper section of the visualization plot presents the transcripts of gene DDX3Y with 5′UTR, as well as exons and transcription direction dynamically queried from the UCSC Genome browser. In the middle track, box plot distributions show Beta-values for each CpG. In the lowest track, in this view, users can view SNP and CpG positions stored in LORIS.
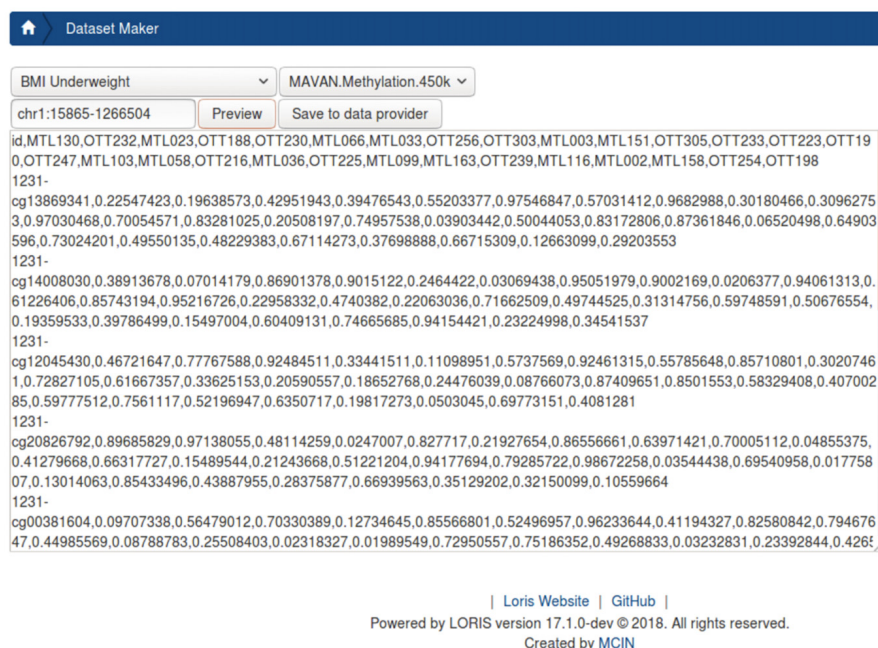
FIGURE 7 | Prototype DatasetBuilder module. The preview panel displays all records returned from jointly querying the database, using the "BMI underweight" pre-built query stored in the data querying tool (DQT) module. This is joined with all subject-samples on which CpGs were found on chromosome 1 between position 15865 and 1266504 from the Methylation450k dataset Beta-values.

visually aligned with SNP data alongside salient gene features for a given range on the genome. This module complemented more sophisticated and domain-specific tools by providing an intuitive web-accessible exploration utility directly within the context of the database, aligning all data points for all subjects of interest on the genome. The ability to "zoom in" on the genome, to better contextualize the measurement of interest, facilitates

understanding of the data within a unified platform. Additional "tracks" from the UCSC Genome Browser are dynamically displayed to provide context for displayed CpGs and SNPs.

DatasetBuilder

Once genomic data have been filtered and collated, the *DatasetBuilder* (**Figure 7**) allows users to aggregate phenotypic,

FIGURE 8 | Prototype of LORIS Imaging Browser with PhantomPipeline processing launch capability using a single button. A user can click on the "Launch" button, under the "PhantomPipeline" column to initiate transfer of the scan dataset to CBRAIN to begin execution of the task.

FIGURE 9 | View of task (PhantomPipeline) running on CBRAIN web portal, launched from LORIS Imaging Browser module in **Figure 8**. The task was launched automatically through CBRAIN's application programming interface (API), but can also be viewed and monitored interactively this way.
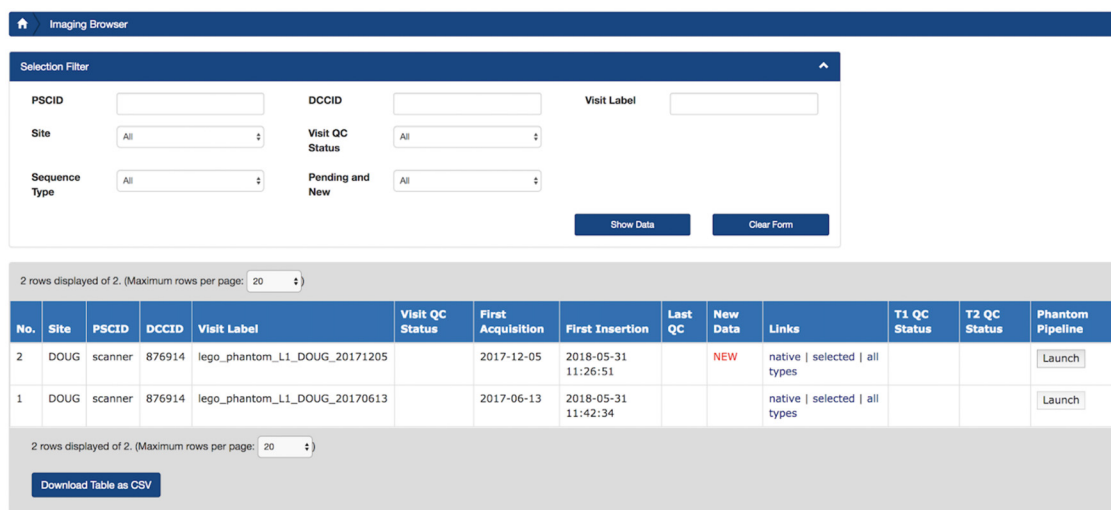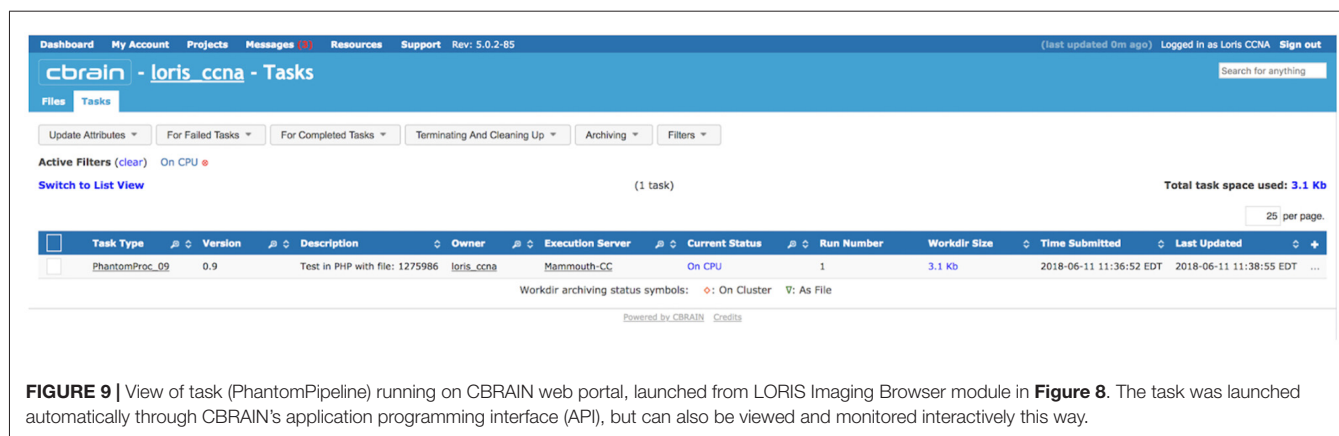
imaging, and other modalities of data for a range of variables across all subjects. A custom dataset can be filtered for specific genomic regions of interest. An intuitive interface design leads users through a process of selecting a genomic fileset, targeting ranges of interest on the genome, and then cross-joining these results by subject ID based on a pre-constructed query across other modalities. The results are saved on the *DataProvider* directory file structure, ensuring that they are available to CBRAIN.

CBRAIN Execution of Containerized Tools

Several pipelines have been made available through CBRAIN for the MAVAN study, such as the *Methylation450K* and *imputePrepSanger*[26] *PCEV*, all described and running in containers. Once installed on CBRAIN and freely available to the community, users can launch these pipelines for their project easily on a number of available HPC resources without any need for additional installation or setup.

The above-mentioned pipelines are spawned as tasks on HPC clusters, where they process data accessed via the *DataProvider*. The output formats described for the pipeline are predefined and remain consistent. These pipelines can be updated on CBRAIN with new versions which may include updates to data format definitions.

Recent work on both LORIS and CBRAIN allows for task creation to spawn processes on CBRAIN where each instance is logged in the LORIS database. Provided an existing tool is registered on CBRAIN and the *DataProvider* is set up, LORIS can register files on CBRAIN and launch an analysis process on them using CBRAIN's RESTful API. Once files are registered on a *DataProvider*, they are recognized by CBRAIN, and transferred to HPC units without any user intervention.

Applications of Additional Pipelines for Derived Data

After pre-processing datasets using containerized pipelines on CBRAIN, additional pipelines can be executed on selected datasets from LORIS in a similar manner. Populations and fields of interest are identified, the datasets are sent to CBRAIN, and

---

[26] https://hub.docker.com/r/eauforest/imputeprepsanger

then a particular container-defined pipeline can be launched. All of these steps can be customized in order to enable execution from the LORIS front-end. Derived datasets from pipeline runs can be generated and returned to LORIS in a similar manner. As mentioned above, users also have the flexibility to re-run desired pipelines with altered parameters in subsequent stages to compare the results within or between pipelines.

Beyond the Ludmer Centre pilot project, applications of this model have been tested on neuroimaging datasets for the Canadian Consortium for Neuroimaging in Aging (CCNA, Mohaddes et al., 2018, this issue). Derived data from MRI lego phantom processing (Fonov et al., 2010) plays a key role in identifying and correcting scanner distortion on scans collected across the CCNA network. LORIS' Imaging Browser (**Figure 8**) is being customized to support automatic launching of the *PhantomPipeline* (Fonov et al., 2010) for execution through CBRAIN (**Figure 9**).

A key advantage of this framework is reproducibility of results, facilitated by detailed provenance capture (logs and parameter definitions from each processing step), as well as container technology (Merkel, 2014) to encapsulate the software environment used for processing and enabling rapid re-deployment.

DISCUSSION

This article focuses on the integration of "*omics*" data with phenotypic data to describe a novel framework for multimodal workflows. One of the key advantages of this model is the variety of functions and tasks covered within a single access-controlled system, such as enhanced monitoring of tasks, provenance tracking and storage of results and visualization features. Improving setup time for installation and re-deployment of containerized pipelines, and abstraction of HPC execution complexities also serve to remove constraints on researchers embarking on the computational learning curve. That being said, the most important aspect of a generalizable framework is to streamline processing and analysis through automation and standardization. Our use-case concretely exemplifies

those steps through: (i) containerizing the *Methylation450K* and *ImputePrepSanger* pipelines in CBRAIN; (ii) launching and relaunching analysis from LORIS using APIs; and (iii) returning results to the *Genomic Browser* module in a structured manner.

Another important element to consider is that in many research environments, workflows are typically processed without the benefits of automated tools or computational infrastructure leading to inefficiencies, disorganization and with time, unmaintained datasets (Siebra et al., 2012). This has become increasingly evident in collaborations that require data sharing, scaling, or re-analysis. As such, we have leveraged established infrastructure to remove or abstract the complexities of data management from the end-user. This is of particular importance given that not all researchers have the time, interest, or expertise to manage the technical aspects of pipeline design and implementation of HPC execution on large datasets. The benefits of organized and curated datasets (Van Horn and Toga, 2009; Kanwal et al., 2017; Nichols et al., 2017; NIH Data Sharing Policy) have been reinforced through the generalizable framework described in this article. While it is true that there are a plethora of software tools and platforms that seek to reduce the technical burden on researchers, not all of them incorporate the full array of best practices necessary for ensuring reproducibility and accuracy in scientific analysis. Our main focus has been to leverage those missing pieces, namely standardization, provenance capabilities, interoperability between systems (such as HPCs) and enhance them with multimodal capabilities and effective visualization of data.

The ability to cross-link *-omic* output with phenotypic and imaging datasets is becoming an increasingly important factor in analysis. Cross-modal linking enables centralized sharing of richer study datasets within a network of investigators, establishing common dataset versions among researchers, and reducing the diffusion of multiple versions of similar datasets. In environments where computational infrastructure is lacking, a great deal of time is typically spent manually organizing datasets in spreadsheets and linking multi-modal data (Calabria et al., 2015). The *Genomic Browser* we describe provides an at-a-glance view of the available data for each participant within LORIS. It also provides a transparent and reproducible capability for visualizing genomic data by enabling filtering and querying across all available data types on shared properties and specific genomic regions around genes of interest. All of these features are graphically displayable on the *Genomic Viewer*. At the same time, the *DatasetBuilder* assembles multimodal datasets to run on processing pipelines in an automated and reproducible manner to significantly improve reliability of data outputs and traceability of targeted datasets. Looking towards a broader use-case, integration of genetics with other data types in a single platform can facilitate validation of genotypic vs. phenotypic characteristics. Basic validations of reported/phenotypic sex compared to genomic sex in a population and comparing reported ethnicity to genomic population markers are common

examples. Such functions, which consider participant-specific phenotypes, allow for multi-level data integration, which are lacking in many existing online informatic resources e.g., GTEx.

Pursuant to utilizing an established data management platform, the benefits of standardization are an important topic and become evident in the execution of pipelines. A key example is how standardizing software installation through container technology reduces potential errors in the configuration and deployment of such pipelines. At the same time, it enhances portability to other platforms, irrespective of the operating systems (Roure et al., 2011; Cito et al., 2016; Sochat et al., 2017), while ensuring the pipelines are consistently executed across networks and research applications. This standardized execution and storage model can be generalized and scaled to larger, more complex workflows and multimodal data types ranging from other kinds of biological "*omics*" data (transcriptomics, proteomics, blood sugar, anthropometry) to behavioral, imaging and electrophysiological data, among others (Zhao et al., 2008). Beyond the example of the *Methylation450k* pipeline, this framework can be used to run any other processing task supported in CBRAIN, yet launched through LORIS. Currently, development is underway to use Galaxy to design additional workflows, and further optimize the *PCEV*[27] pipeline. This pipeline is however only one amongst many other analysis methods that can be used in imaging genetics (Vilor-Tejedor et al., 2018).

Provenance also remains an important issue in any kind of analysis, especially in a multi-modal and multi-software environment, such as the generalizable workflow proposed in this article. To ensure complete accessibility of provenance information:

- task log details from CBRAIN's internal records are communicated to LORIS with each set of returned results and made queryable via the LORIS front-end.
- standard file formats (e.g., JSON, XML, TSV) are used for the re-insertion process for derived data, as well as metadata to facilitate integration into LORIS with minimal interface development.
- quality control results are stored alongside raw and processed outputs which improves usability.
- increasing adoption of *Boutiques* descriptors (Glatard et al., 2018) as a framework for sharing and defining task creation on HPC resources will support standardization and transparency in neuroinformatic analyses.

The ultimate aim is to produce results and maintain provenance information that is compatible with emerging neuroimaging standards (e.g., the NIDM, Keator et al., 2016).

Interoperability between systems and datasets has become a requirement for sharing and collaboration in numerous fields involving many complex analytics, such as machine learning algorithms which are a rising interest in the field of imaging

---

[27] https://github.com/GreenwoodLab/pcev_pipelineCBRAIN

genetics. Making use of APIs that can seamlessly operate from one environment to another is a key consideration in our model. Linking to other systems to share data, or simply for reference pointers (e.g., links to the UCSC Genome Browser), is an important step in data harmonization (Zaveri, 2017). Developing APIs that are streamlined across platforms and easily fulfill community standards and workflow requirements provides an important asset for interoperability in large-scale consortia and open data initiatives (Poline et al., 2012; Poldrack et al., 2013; Van Horn and Toga, 2014; Craddock et al., 2016; Das et al., 2017).

One key advantage of this infrastructure is "Privacy by Design" which uses several mechanisms from acquisition to dissemination to ensure privacy, such as anonymous identifiers that link epigenetic data to a subject record, encryption methods to secure data transfers, specific anonymization techniques and other best practices (Cavoukian, 2009). This method largely removes the need to store personally identifying information (e.g., research participants and patient names) further mitigating the risk of re-identification. This facilitates sharing of other available data elements with a detailed provenance history when publishing analyses of genomic data through LORIS, where permissible, and in compliance with ethical regulations. Rendering these datasets non-identifiable is an active research area, giving rise to masking algorithms, which may be of interest to data-sharing initiatives.

Another major challenge in analysis is reproducibility. This becomes particularly evident in workflows that span different domains such as imaging and genetics (Nekrutenko and Taylor, 2012). In its process design and technical implementation, this generalizable framework aims to adhere to the *FAIR* (Wilkinson et al., 2016) data principles. In our workflow, inputs and outputs of each processing task are available to platform members alongside provenance information from container descriptions and pipeline execution logs, and each step of the workflow can be re-run locally or on other systems. Using the open-source constituent tools of this workflow, capturing the same outputs in the same manner from a reproduction of this workflow provides a powerful means to directly compare each aspect of an analysis that has been re-run.

Through the development of this combined framework and across several infrastructure initiatives, best practices have emerged. These have been articulated in **Appendices 1** and **2** as guidelines summarizing both the principles and practical recommendations for implementations of this framework.

Future extensions of this infrastructure, based on user feedback, will add richer features and more seamless automation at several stages. As a result, a number of features will be developed and improved:

- Streamlining the data-loading processes in LORIS via the release of open-source tools will facilitate easier adoption of this framework for other "*omics*" workflows.
- Integrating formats from other platforms will expand the scope of this technology.
- Address scaling challenges through increasing use of NoSQL schema-less databasing to flexibly handle increasing volumes

of genomic data and its significant variability across data types and structures.

- *Boutiques* descriptors for CBRAIN to generalize LORIS task-launching capabilities and ease the development burden of deploying new pipelines.
- A well-defined API using the OpenAPI[28] standard, registered on SmartAPI[29], to facilitate the creation of specialized tools to interact with LORIS programmatically.
- Interoperability with data discovery platforms like *DataLad*[30], to support querying, packaging and return of LORIS-hosted data into BIDS[31]-formatted data objects. Adding enhanced support for API endpoints will support these operations.
- Encapsulating the *Genomic Viewer* into a Javascript module would help portability across platforms.

While these components will fulfill the vision for a fully robust feature-set in LORIS and CBRAIN, further developments, documentation, unit tests and integration tests will be important to include beyond the prototyping stage, to ensure the resulting combined framework does not amass technical debt for future workflows.

CONCLUSION

The goal of this article is to present a novel framework that can facilitate brain research discovery by reducing human error through the automation of analysis pipelines and seamless linking of multimodal data workflows. The described framework for "*omics*" workflows integrates multi-modal data support in a mature databasing system with analysis on HPC platforms, with a wide array of capabilities including provenance tracking, a well-defined processing environment, visualization, querying and links with other existing genomics databases. Ultimately, this framework aims to create an optimally user-friendly experience to allow researchers to focus on scientific aims rather than the obstacles that otherwise occur with complex data handling.

AUTHOR CONTRIBUTIONS

SD, XLB, CR, MF, NB and CG contributed to the conception and design of the generalized workflow and wrote sections of the manuscript. FC-D, CM, PR, SB, VF and CZ wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

---

[28]https://github.com/OAI/OpenAPI-Specification
[29]https://smart-api.info/
[30]https://github.com/DataLad/DataLad
[31]https://github.com/INCF/BIDS; BIDS is an emerging neuroimaging data standard for describing and sharing data and metadata.

ACKNOWLEDGMENTS

REFERENCES

Afgan, E., Baker, D., Batut, B., van de Beek, M., Bouvier, D., Čech, M., et al. (2018). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., et al. (2016). The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research.* 44, W3–W10. doi: 10.1093/nar/gkw343

Calabria, A., Spinozzi, G., Benedicenti, F., Tenderini, E., and Montini, E. (2015). adLIMS: a customized open source software that allows bridging clinical and basic molecular research studies. *BMC Bioinform.* 16:S5. doi: 10.1186/1471-2105-16-S9-S5

Cavoukian, A. (2009). Privacy by design: the 7 foundational principles. *PrivacyByDesign.* Available online at: http://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

Cito, J., Ferme, V., and Gall, H. C. (2016). "Using docker containers to improve reproducibility in software and web engineering research," in *Presented at the International Conference on Web Engineering (ICWE) 2016*, Lugano, 609–612.

Craddock, R. C., Margulies, D. S., Bellec, P., Nichols, B. N., Alcauter, S., Barrios, F. A., et al. (2016). Brainhack: a collaborative workshop for the open neuroscience community. *GigaScience* 5:16. doi: 10.1186/s13742-016-0121-x

Das, S., Glatard, T., MacIntyre, L. C., Madjar, C., Rogers, C., Rousseau, M. E., et al. (2016). The MNI data-sharing and processing ecosystem. *Neuroimage* 124, 1188–1195. doi: 10.1016/j.neuroimage.2015.08.076

Das, S., Glatard, T., Rogers, C., Saigle, J., Paiva, S., MacIntyre, L., et al. (2017). Cyberinfrastructure for open science at the montreal neurological institute. *Front. Neuroinform* 10:53. doi: 10.3389/fninf.2016.00053

Das, S., Zijdenbos, A. P., Harlap, J., Vins, D., and Evans, A. C. (2011). LORIS: a web-based data management system for multi-center studies. *Front. Neuroinform.* 5:37. doi: 10.3389/fninf.2011.00037

Fonov, V. S., Janke, A., Caramanos, Z., Arnold, D. L., Narayanan, S., Pike, G. B., et al. (2010). "Improved precision in the measurement of longitudinal global and regional volumetric changes via a novel MRI gradient distortion characterization and correction technique," in *Medical Imaging and Augmented Reality MIAR 2010. Lecture Notes in Computer Science*, (vol 6326) eds H. Liao, P. J. Edwards, X. Pan, Y. Fan, and G. Z. Yang (Berlin, Heidelberg: Springer), 324–333. doi: 10.1007/978-3-642-15699-1_34

Fortin, J.-P., Labbe, A., Lemire, M., Zanke, B. W., Hudson, T. J., Fertig, E. J., et al. (2014). Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* 15:503. doi: 10.1186/s13059-014-0503-2

Glatard, T., Kiar, G., Aumentado-Armstrong, T., Beck, N., Bellec, P., Bernard, R., et al. (2018). Boutiques: a flexible framework to integrate command-line applications in computing platforms. *Gigascience* 7:5. doi: 10.1093/gigascience/giy016

Glatard, T., Lewis, L. B., Ferreira da Silva, R., Adalat, R., Beck, N., Lepage, C., et al. (2015). Reproducibility of neuroimaging analyses across operating systems. *Front. Neuroinform* 9:12. doi: 10.3389/fninf.2015.00012

Kanwal, S., Khan, F. Z., Lonie, A., and Sinnott, R. O. (2017). Investigating reproducibility and tracking provenance—A genomic workflow case study. *BMC Bioinformatics* 18:337. doi: 10.1186/s12859-017-1747-0

Keator, D. B., Helmer, K., Ghosh, S. S., Auer, T., Maumet, C., Das, S., et al. (2016). "Describing assessments and experiment metadata with the neuroimaging data model (NIDM)," in *Front. Neuroinform.* (Conference Abstract: Neuroinformatics 2016, Reading, UK) .

MacFarlane, D., Das, S., Kostopolous, P., Kat, J., Rogers, C., Makowski, C., et al. (2014). "Enhanced data querying for neuroinformatics databases," in *Paper Presented at the 2014 Annual Meeting of the Organization for Human Brain Mapping*, (Hamburg, Germany).

Merkel, D. (2014). Docker: lightweight linux containers for consistent development and deployment. *Linux J.* 239. Available online at: http://dl.acm.org/citation.cfm?id=2600239.2600241

Mohaddes, Z., Das, S., Abou-Haidar, A., Safi-Harab, M., Blader, D., Callegaro, J., et al. (2018). National neuroinformatics framework for canadian consortium on neurodegeneration in aging (CCNA). *Front. Neuroinf.* doi: 10.3389/fninf.2018.00085

Nekrutenko, A., and Taylor, J. (2012). Next-generation sequencing data interpretation: enhancing reproducibility and accessibility. *Nat. Rev. Genet.* 13, 667–672. doi: 10.1038/nrg3305

Nichols, T. E., Das, S., Eickhoff, S. B., Evans, A. C., Glatard, T., Hanke, M., et al. (2017). Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* 20, 299–303. doi: 10.1038/nn.4500

O'Donnell, K. A., Gaudreau, H., Colalillo, S., Steiner, M., Atkinson, L., Moss, E., et al. (2014). The maternal adversity, vulnerability and neurodevelopment project: theory and methodology. *Can. J. Psychiatry* 59, 497–508. doi: 10.1177/070674371405900906

Oros Klein, K., Grinek, S., Bernatsky, S., Bouchard, L., Ciampi, A., Colmegna, I., et al. (2016). FuntooNorm: an R package for normalization of DNA methylation data when there are multiple cell or tissue types. *Bioinformatics* 32, 593–595. doi: 10.1093/bioinformatics/btv615

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., and Munafò, M. R. (2017). Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* 18, 115–126. doi: 10.1038/nrn.2016.167

Poldrack, R. A., Barch, D. M., Mitchell, J. P., Wagner, T. D., Wagner, A. D., Devlin, J. T., et al. (2013). Toward open sharing of task-based fMRI data: the openfMRI project. *Front. Neuroinform.* 7:12. doi: 10.3389/fninf.2013.00012

Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Front. Neuroinform.* 6:9. doi: 10.3389/fninf.2012.00009

Pool, R., Esnayra, J., and National Research Council (US) Board on Biology. (2000). Bioinformatics - Converting Data to Knowledge: Workshop Summary. Washington, DC: National Academies Press (US).

Rogers, C., Das, S., Mohaddes, Z., MacFarlane, D., Kostopoulos, P., Portales-Casamar, E., et al. (2015). "LORIS neuroinformatics platform for Imaging Genetics," in *Paper Presented at the 2015 Annual Meeting of the Organization for Human Brain Mapping*, (Honolulu, USA).

Roure, D. D., Belhajjam, K., Missier, P., Gómez-Pérez, J. M., Palma, R., Ruiz, J. E., et al. (2011). "Towards the preservation of scientific workflows," in *Paper Presented at iPRES 2011—8th International Conference on Preservation of Digital Objects*, Singapore.

Sherif, T., Kassis, N., Rousseau, M. E., Adalat, R., Rioux, P., Evans, A. C., et al. (2014). CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. *Front. Neuroinform.* 8:54. doi: 10.3389/fninf.2014.00054

Siebra, C. S. A., Tonin, G. S., Silva, F. Q. B., Oliveira, R. G., Junior, A. L. O. C., Miranda, R. C. G., et al. (2012). "Managing technical debt in practice: an industrial report", in *ESEM '12 Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (New York, NY: ACM), 247–250. doi: 10.1145/2372251.2372297

Sochat, V. V., Prybol, C. J., Kurtzer, G. M., and Antoniewski, C. (2017). Enhancing reproducibility in scientific computing: metrics and registry for singularity containers. *PLoS One* 12:e0188511. doi: 10.1371/journal.pone.0188511

Van Horn, J. D., and Toga, A. W. (2009). Is it time to re-prioritize neuroimaging databases and digital repositories? *Neuroimage* 47, 1720–1734. doi: 10.1016/j.neuroimage.2009.03.086

Van Horn, J. D., and Toga, A. W. (2014). Human neuroimaging as a "big data" science. *Brain Imaging Behav.* 8, 323–331. doi: 10.1007/s11682-013-9255-y.

Vilor-Tejedor, N., Alemany, S., Cáceres, A., Bustamante, M., Pujol, J., Jordi Sunyer, J., et al. (2018). Strategies for integrated analysis in imaging genetics

studies. *Neurosci. Biobehav. Rev.* 93, 57–70. doi: 10.1016/j.neubiorev.2018.06.013

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Zaveri, A. (2017). "SmartAPI: towards a more intelligent network of web APIs," in *The Semantic Web. ESWC 2017. Lecture Notes in Computer Science*, (vol 10250) eds E. Blomqvist, D. Maynard, A. Gangemi, R. Hoekstra, P. Hitzler, and O. Hartig (Cham: Springer), 154–169.

Zhao, Y., Raicu, I., and Foster, I. (2008). "Scientific workflow systems for 21st century, new bottle or new wine?" in *Paper Presented at the 2008 Congress on Services—SERVICES'08. IEEE*, Honolulu, HI, 467–471. doi: 10.1109/SERVICES-1.2008.79

Preface to Appendices: The authors have developed and recommend community-supported best practices for adhering to *FAIR* principles in both the development of infrastructure and its implementation in practice.

APPENDIX 1

Best Practices checklist for technology design and development for *FAIR Multimodal* Framework integration:

Findable:

– Recognized by SmartAPI registry
– Use DOI for datasets and link them to LORIS API endpoint
– Softwares openly available via public repositories, and free (as in Freedom[32])
– Connect datasets using technologies like DataLad
– Publish and cite others

Accessible:

– Must be governed by the study and subject consent
– Ensure protection of https but no institution-restricting firewall (hospitals)
– Sustainable infrastructure plan to backup, support and maintain the server and storage. This includes software security upgrades to keep data safe as well as accessible

Interoperable:

– API endpoints for datasets + version (GitAnnex/DataLad)
– Converters for BIDS—import/export of datasets
– Provide guidance where possible on use with other standards such as NIDM

Reusable:

– Container technologies
– Boutiques descriptors

---

[32]https://www.gnu.org/philosophy/free-sw.en.html

– Software pipelines coded in a way that it can be reused
– Reuse of analysis results: ensure execution parameters and Provenance information is readily accessible *and* shareable with the data (so they can be exported together)

APPENDIX 2

Implementation Guidelines for *FAIR Multimodal* Workflow integration:

– Store raw data at an early stage; this facilitates the linkage between subjects and the provenance "trail"
– Use a centralized database system with user authentication for full auditing and data organization
– Describe entities, agents and activities with *PROV*[33] family vocabulary.
– Use https and ssh as protocols of communication
– Use a network with known geographical location of nodes (HPC, Storage, Hubs)
– Design pipelines with checkpoints where the state of the data can be reused with alternative paths (forks)
– Use container technology to describe the execution environment setup
– Publish software (web-apps, pipelines, containers) in a public hub with appropriate licenses—ideally free (as in freedom)
– Document software for users but also future contributors, including "contribution guidances"
– Describe pipeline requirements: space, CPUs, nodes, formats (Kanwal et al., 2017)
– Cite all used software and datasets
– Publish datasets on sustainable platforms
– Describe your datasets with common terms and standard community-supported formats (JSON-LD, RDF)

---

[33]https://www.w3.org/TR/prov-overview/ PROV is a family of standards for interoperable interchange of provenance information.

Check for updates

Brain-CODE: A Secure Neuroinformatics Platform for Management, Federation, Sharing and Analysis of Multi-Dimensional Neuroscience Data

Anthony L. Vaccarino[1,2]*, Moyez Dharsee[2], Stephen Strother[2,3], Don Aldridge[4], Stephen R. Arnott[2,3], Brendan Behan[1], Costas Dafnas[4], Fan Dong[2,3], Kenneth Edgecombe[4], Rachad El-Badrawi[2], Khaled El-Emam[5], Tom Gee[2,3], Susan G. Evans[2], Mojib Javadi[2], Francis Jeanson[1], Shannon Lefaivre[1], Kristen Lutz[2], F. Chris MacPhee[4], Jordan Mikkelsen[2], Tom Mikkelsen[1], Nicholas Mirotchnick[1], Tanya Schmah[6], Christa M. Studzinski[1], Donald T. Stuss[1,3,7], Elizabeth Theriault[1] and Kenneth R. Evans[2]

[1] Ontario Brain Institute, Toronto, ON, Canada, [2] Indoc Research, Toronto, ON, Canada, [3] Rotman Research Institute, Toronto, ON, Canada, [4] Centre for Advanced Computing, Kingston, ON, Canada, [5] Privacy Analytics Inc., Ottawa, ON, Canada, [6] Department of Mathematics and Statistics, University of Ottawa, Ottawa, ON, Canada, [7] Departments of Psychology and Medicine, University of Toronto, Toronto, ON, Canada

Historically, research databases have existed in isolation with no practical avenue for sharing or pooling medical data into high dimensional datasets that can be efficiently compared across databases. To address this challenge, the Ontario Brain Institute's "Brain-CODE" is a large-scale neuroinformatics platform designed to support the collection, storage, federation, sharing and analysis of different data types across several brain disorders, as a means to understand common underlying causes of brain dysfunction and develop novel approaches to treatment. By providing researchers access to aggregated datasets that they otherwise could not obtain independently, Brain-CODE incentivizes data sharing and collaboration and facilitates analyses both within and across disorders and across a wide array of data types, including clinical, neuroimaging and molecular. The Brain-CODE system architecture provides the technical capabilities to support (1) consolidated data management to securely capture, monitor and curate data, (2) privacy and security best-practices, and (3) interoperable and extensible systems that support harmonization, integration, and query across diverse data modalities and linkages to external data sources. Brain-CODE currently supports collaborative research networks focused on various brain conditions, including neurodevelopmental disorders, cerebral palsy, neurodegenerative diseases, epilepsy and mood disorders. These programs are generating large volumes of data that are integrated within Brain-CODE to support scientific inquiry and analytics across multiple brain disorders and modalities. By providing access to very large datasets on patients with different brain disorders and enabling linkages to provincial, national and international databases, Brain-CODE will help to generate new hypotheses about the biological bases of brain disorders, and ultimately promote new discoveries to improve patient care.

Keywords: Brain-CODE, neuroinformatics, big data, electronic data capture, open data

INTRODUCTION

The principles of data sharing as a catalyst for scientific discovery are widely recognized by international organizations such as the National Institutes of Health (2003), Wellcome Trust (2010) and Canadian Institutes of Health Research (2013). Historically, however, research databases have existed in isolation with no practical avenue for sharing or pooling medical data into high dimensional "big" datasets that can be efficiently compared across databases. Databases have their own sets of data standards, software and processes, thus limiting their ability to synthesize and share data with one another. To address this challenge, the Ontario Brain Institute (OBI) created Brain-CODE – an extensible, neuroinformatics platform designed to support curation, sharing and analysis of different data types across several brain disorders[1]. Brain-CODE allows researchers to collaborate and work more efficiently to understand the biological basis of brain disorders.

Ontario Brain Institute supports collaborative research networks focused on various brain conditions, including neurodevelopmental disorders[2], cerebral palsy[3], epilepsy[4], mood disorders[5], and neurodegenerative diseases[6] (**Figure 1**). The creation of these programs has resulted in a "big data" opportunity to support the development of innovative, impactful diagnostics and treatments for brain disorders (Stuss, 2014, 2015). By providing researchers access to aggregated datasets that they otherwise could not obtain independently, Brain-CODE incentivizes data sharing and collaboration, and facilitates analyses both within and across disorders and across an array of data types, including clinical, neuroimaging, and molecular. By collecting data elements across disorders Brain-CODE enables deep phenotyping across data modalities within a brain disorder, as well as investigations across disorders. Moreover, linkages with provincial, national and international databases will allow scientists, clinicians, and industry to work together in powerful new ways to better understand common underlying causes of brain dysfunction and develop novel approaches to treatment.

Using the FAIR Data Principles as guidance, Brain-CODE is being developed to support the principles of data being Findable, Accessible, Interoperable, and Reusable (FAIR, Wilkinson et al., 2016). The Brain-CODE system architecture provides the technical capabilities to support:

- Consolidated data management to securely capture, monitor and curate data.
- Privacy and security best-practices.
- Interoperable and extensible federation systems that supports harmonization, integration and query across diverse data modalities and linkages to external data sources.

---

[1]www.braincode.ca

[2]www.pond-network.ca

[3]cpnet.canchild.ca

[4]www.eplink.ca

[5]www.canbind.ca

[6]www.ondri.ca

Brain-CODE Design Principles
Interoperability and Standardization to Support Data Integration and Collaboration

The types of data being collected in modern research are increasingly diverse, from larger numbers of sources and patient populations, and involving highly specialized technologies, from genomics and imaging, to wearable devices and surveys delivered via mobile apps. There is also a growing need to link and query data collected within a given research study with data stored in disparate other locations and formats, such as public data repositories, health administration data holdings, electronic medical records, and legacy databases. As a result, researchers deploy a broad range of tools to collect, process and analyze their data, but the lack of interoperability of these platforms serves as a barrier to data sharing and collaboration. Establishing standard software would address these issues; however, available platforms each have their unique advantages and there is a significant cost for researchers in time and effort to move to new platforms. This complex set of data integration needs cannot be addressed using inflexible systems working in isolation nor by the development of a "one size fits all" platform. Rather, to support this level of data integration, interoperability must be a core requirement. This approach differs from most other data platforms in which data are combined at the data analyses stage. Interoperability, however, enables large-scale data aggregation and federation of systems and data across multiple data types, allowing novel discoveries and analyses to be conducted. Moreover, allowing researchers to decide which system to use ensures greater researcher uptake, which facilitates collaboration and data sharing within and across the broader research community.

From its inception, Brain-CODE architecture was designed with interoperability in mind, such that it could support the integration and analysis of large volumes of complex data from diverse sources. With this approach each platform can maintain its autonomy while still integrating into a much larger whole. This can be a challenge as databases are often stored in individual "silos" with their own sets of data standards, software and processes which limit their ability to interact with one another. Interoperability, therefore, requires the development of pipelines and processes between existing platforms; software to allow efficient and seamless exchange of data and information between systems and technologies, including application programming interfaces (APIs) to allow data flow between applications. In addition, rigorous standardization processes are required that govern how information is recorded and exchanged in order to define and format the vast array of clinical, neuroimaging and molecular data, and to optimize federation by ensuring that data in one system is understood by another. Effort must, therefore, be devoted to creating standards across studies, including common data elements (CDEs) (i.e., the same endpoints applied to multiple studies), common ontologies (i.e., utilizing common nomenclature and format across studies), as well as standard processes and procedures related to the collection of data.
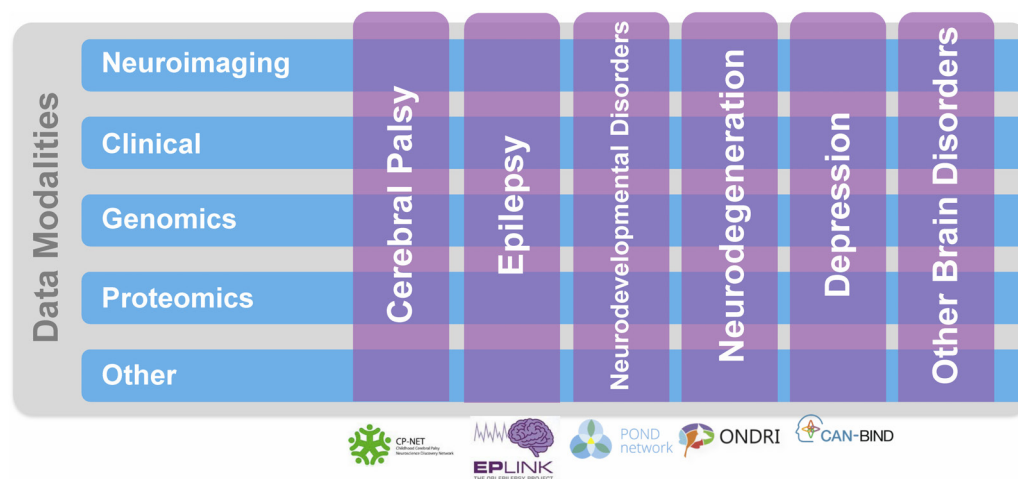
FIGURE 1 | Ontario Brain Institute (OBI) Programs. These programs take a different approach to research that spans many disciplines and brings together a diverse group of stakeholders including researchers, clinicians, industry partners, and patients and their advocates. The programs collect various types of data, including genetic, molecular, imaging, and behavioral, which are stored on Brain-CODE. Figure adapted from Stuss (2014).

An Extensible Design to Accommodate Expanded or Modified Functionality

Since not all functionality can be determined upfront, extensibility of the system must also be considered a core design principle to accommodate new and expanded functionality without impacting existing features. This approach allows the integration of users' programs and third-party software into the system, as well as allowing for customization and enhancement of existing systems. Choice of technologies used and how the databases are built is critical and the use of commercial software can limit the ability to extend functionality, as these are typically built for a specific purpose and source codes are often not available. Extensibility is less of an issue when using open source software, as the code is published and can be modified. Where possible, therefore, Brain-CODE infrastructure is built using open-source tools.

Privacy and Security

Brain-CODE was designed with best-practice privacy strategies at the forefront to enable secure capture of sensitive participant data in a manner that abides by ethical principles and government legislation while fostering data sharing and linking opportunities. As such, privacy and security features have been robustly incorporated into the foundation of Brain-CODE's infrastructure, and are reinforced by guidelines and safeguards that ensure participant data security.

Federation and linking with other databases involves the implementation of high-security data transfer infrastructure. These include encryption and de-identification tools to protect participant data and enhanced validation certificates to guarantee authenticity of outward-facing software applications, as well as administrative, physical and technical safeguards and security processes that are aligned with Code of Federal Regulations Title 21, Part 11 standards (CFR Title 21, Part 11, 2017). As a result, OBI has been named a "Privacy by Design" Ambassador

by the Office of the Information and Privacy Commissioner of Ontario (Cavoukian, 2011). This designation refers to the mitigation of privacy and security risks through a proactive and preventative approach to research data management by embedding privacy and security measures directly into the design of systems and practices. Working with a team of experts, OBI has developed clear and comprehensive policies and guidelines on data privacy and governance[7]. These documents outline how data are collected, stored, and accessed by Brain-CODE users.

DATA LIFE-CYCLE

Researchers collect sensitive participant data in the form of clinical assessments, interventional studies, and brain imaging, cognitive and sensory-motor measures, as well as biological samples for proteomic and (epi-)genetic analyses. Personal Health Information (PHI) must be carefully handled in accordance with the Personal Health Information Protection Act, 2004, S.O. 2004, c. 3, Sched. A (Personal Health Information Protection Act [PHIPA], 2004) from a governance and contract perspective, as informed by principles in ISO 27001 for information management. To maximize the data sharing and analytics capacity of Brain-CODE, while enabling the secure collection of PHI, processes were developed to permit functional separation of sensitive data while being complemented by granular access controls to ensure that data are only available to Brain-CODE users who are authorized to access it (**Figure 2**).

Data Capture and Curation

Brain-CODE provides a virtual laboratory environment where researchers (data producers) can upload, download, manage,

---

[7]www.braincode.ca/content/governance

FIGURE 2 | Data life cycle within Brain-CODE.
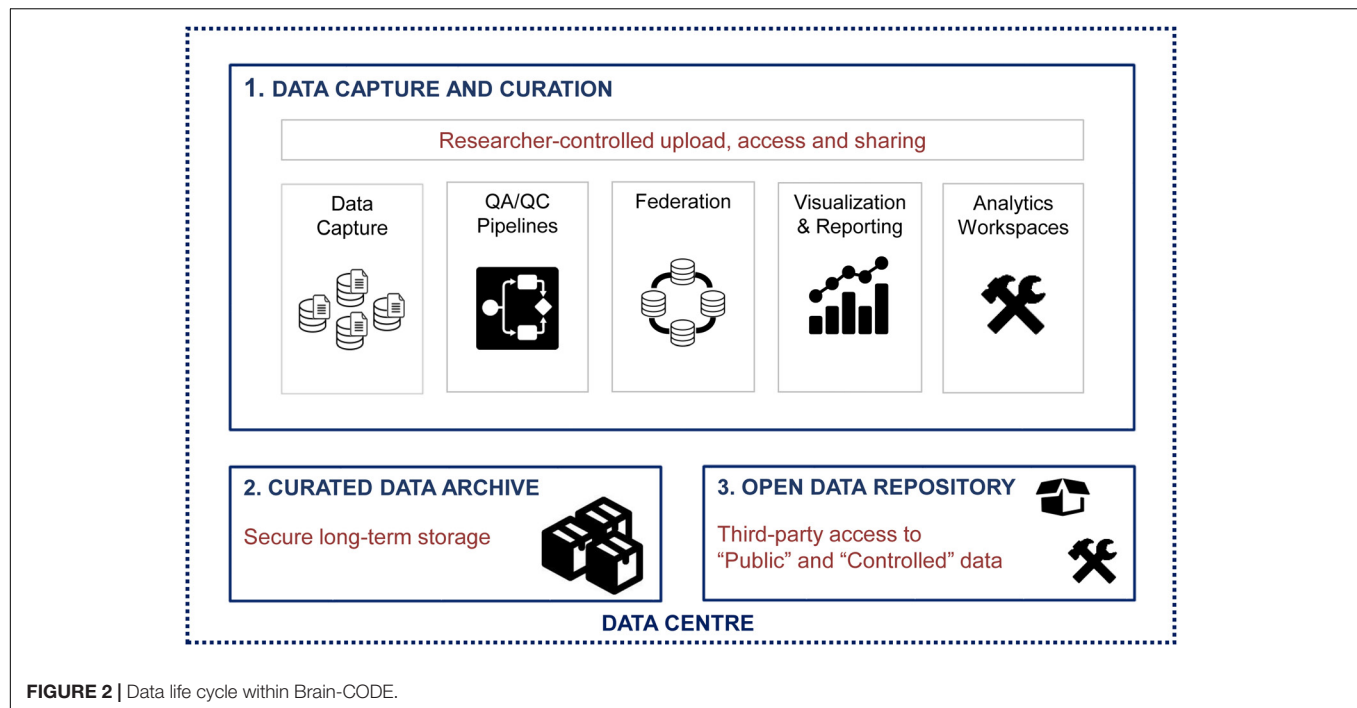
curate and share their own research datasets with direct study collaborators. Based on Research Ethics Board (REB) approval and participant informed consent, data uploaded to Brain-CODE may include PHI. Before any data are uploaded to Brain-CODE, institutions enter into a Participation Agreement with OBI, whereby the institution and affiliated researchers agree to make use of the platform in a manner that abides by OBI's Informatics Governance Policy, Platform Terms of Use and applicable privacy laws, and particularly institutional REBs. The participating institutions also grant OBI a non-exclusive license to share de-identified study datasets in the future, following an exclusivity period. An exclusivity plan is established between OBI and the researchers; during the period of exclusivity, data access remains exclusive to data producers and their direct collaborators. Before, during and after the exclusivity period, data producers and direct collaborators continue to have full access to their data, including access to a suite of analytical tools and workspaces, to enable data cleaning, curation and analysis required by studies.

Curated Data Archive

Following an exclusivity period, curated datasets are versioned for long-term secure storage. These data are labeled as either "Controlled Data" or "Public Data." Controlled Data are datasets that have been de-identified. These Controlled datasets are made available to third-party Brain-CODE users by request, and can be augmented through links to external databases in a secure environment. Public Data are either basic science datasets (i.e., from animal model studies), metadata, or human datasets that did not previously contain PHI. Public datasets can be shared directly with Brain-CODE users, without requiring an access request.

Open Data Repository

One of the goals of Brain-CODE is to release high quality research data to researchers outside of OBI. The "Open Data" interface was developed for third-party users to browse information about Controlled and Public datasets and access data releases (see **Figures 2**, **3**). While Public Data can be accessed directly, Controlled Data requires users to submit data access requests. Data access requests are reviewed by Brain-CODE's Data Access Committee (DAC) which is composed of researchers, neuroinformatics experts, and OBI staff. The DAC makes a recommendation to the Informatics Steering Committee, which makes final decisions related to data access. Once a request is approved, third-party users must provide proof of REB approval and enter into a Data Use Agreement with OBI before being granted access to the data for retrieval and analyses. The de-identified dataset can be exported to a workspace environment available upon request to any registered Brain-CODE user, allowing access to high performance computing resources and analysis tools. The access request, review and approval process is streamlined within the Brain-CODE portal to ensure a timely turnaround of 10 days from access request to granting data access.

CONSOLIDATED DATA MANAGEMENT

Within the Brain-CODE Portal, data capture is consolidated with a diverse set of electronic data capture (EDC) tools for various data modalities including clinical, imaging, and 'omics that allow researchers to securely upload, store and manage research data electronically[8] (**Table 1**). The Brain-CODE platform was

---

[8]www.braincode.ca

FIGURE 3 | Data exploration and release dashboard. Open data release of High Resolution MRI of Mouse Models Related to Autism. Available at www.braincode.ca

developed to allow incorporation of new data capture systems as required by the various research teams (**Figure 4**). In addition to providing a single point of access to data management tools, the Brain-CODE Portal features project management dashboards, private file repositories and discussion for a that researchers can use to facilitate sharing and collaboration.

As with most data repositories, naming conventions standards are key. Not only do these features enable a given subject's data to be linked across the corresponding data stored on separate data capture systems (e.g., that same subject's clinical data stored on REDCap system with imaging stored on XNAT), but such standardization ensures that automated quality assurance (QA) and quality control (QC) pipelines can be successfully applied to the data. The naming format used across Brain-CODE programs conform to the general format of PPPTT_HHH_SSSS, where PPP is the program code, TT is the study code, HHH is the site code and SSSS is the subject number. The four-digit subject number is typically assigned by the subject co-coordinator. A fifth Subject ID digit can be employed if deemed necessary.

Clinical Data Management

A core objective of Brain-CODE is to organize, standardize, and integrate the various forms of clinical information collected from OBI-funded and partner research programs. Traditionally, data have been collected on paper but there is a growing trend both in industry and academic research settings toward EDC for some forms of data (Food and Drug Administration [FDA], 2013). However, many academic research teams lack the necessary infrastructure and specialized skills to use and maintain a clinical data management system. To alleviate this situation, multiple web-based clinical data management software packages are deployed and hosted in Brain-CODE to allow researchers to

remotely access these tools and integrate them into daily research practice.

The two primary clinical data management systems used to collect demographic and clinical data are REDCap (Research Electronic Data Capture[9]) and OpenClinica[10]. REDCap is a web-based application developed by a multi-institutional consortium led by Vanderbilt University specifically to support data capture for academic research studies. The software is freely available under the conditions of an end-user license agreement, and has been designed to be very simple to configure, use and maintain. As such, REDCap has grown into a very popular solution within the research community. REDCap is designed to comply with the United States' Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations, but is currently not CFR Title 21, Part 11 compliant. OpenClinica is developed and maintained by OpenClinica LLC, in both an open source Community Edition as well as a commercially licensed Enterprise Edition, the latter providing training and technical support. The Enterprise Edition is currently deployed in Brain-CODE; both a development/test and a production instance are installed. OpenClinica LLC fully maintains the deployment, including installation validation, database backup configuration, OS updates, software patches and upgrades, and technical support. OpenClinica is a fully featured, web-based system that supports multi-site clinical trials and clinical data management. The software is compliant with HIPAA and CFR Title 21, Part 11, providing the required electronic signature and audit trail functionality for use in clinical trials requiring FDA regulatory approval. Additional clinical data capture systems can be deployed as required.

---

[9]projectredcap.org

[10]www.openclinica.com

TABLE 1 | Data modalities currently collected in Brain-CODE.

Modality
Demographics
Patient-reported outcomes
Clinician-reported outcomes
Cognitive assessments
Structural MRI
Functional MRI
Diffusion tensor imaging (DTI)
Spectral MRI
Behavioral outcome files (timing, events)
Investigators notes
Electroencephalogram (EEG)
Electrocardiograph (ECG)
Pulse plethysmograph (PPG)
Respiratory
Magnetoencephalogram (MEG)
Ocular computed tomography (OCT)
Fundal photography
Eye tracking
Pupil metrics
Gait track data
Accelerometers
Force plate
Audio files
Video files
Pathology images
Imaging manual QC
fBIRN fMRI imaging metrics
OHIP numbers
Genotyping
ONDRISeq
SNP and expression arrays
GWAS
Sequencing (NGS)
Proteomics
Absorbance based assays (i.e., ELISA, etc.)

Clinical Data Standardization

Brain-CODE includes multidisciplinary collaborative research networks across multiple brain disorders. Given the different research aims, study designs and technologies used across research programs, establishment of a minimum set of clearly defined and standardized assessments across studies is essential to facilitate data sharing and integration, and to conduct meaningful analyses across disorders. Indeed, these data must be sufficiently comparable to allow any levels of data integration, and in the absence of common measures and data standards it is difficult to compare the results from one study to another. From a data integration perspective, CDEs and other standardized variables represent shared attributes between different data models that can significantly enhance the implementation of the federated database by reducing the semantic and syntactic heterogeneities between constituent databases. Therefore, in an effort to optimize the ability to

aggregate and analyze data within Brain-CODE, CDEs were developed to provide standard definitions and formats so that investigators collect data consistently across studies and programs.

Using the framework of the National Institute of Neurological Disorders and Stroke (NINDS) CDE Project as guidance (Grinnon et al., 2012), a Delphi consensus-based methodology (Dalkey and Helmer, 1963; Hsu and Sandford, 2007) was used to identify core demographic and clinical variables to be collected across all participating OBI research programs. The CDEs include standardized assessments across the life-span of quality of life, medical and psychiatric co-morbidities, as well as clinical outcome measures of depression, anxiety, and sleep (**Table 2**). There was also agreement that when possible, the measure should be patient-reported, brief and easy to administer, widely used and validated, and available in the public domain. In addition, where possible the Clinical Data Interchange Standards Consortium (CDISC) standards are applied to define data collection fields, formatting, and terminology (Souza et al., 2007). This reduces variability in data collection and ultimately facilitates comparisons across disorders, merging of datasets and meta-analyses.

Clinical Data Quality Assurance and Control

Prior to data collection, clinical databases are validated to ensure adherence to data standards, compliance with the Brain-CODE CDEs, potential governance and privacy issues, and database quality. Identifying fields are compared against the language used in their ethics submission for compliance. Validation can also identify errors or missing data points in the database before data entry begins. Project validation involves a thorough review of a project's variable naming, field naming, item coding, field validation and case report form equivalence through data entry, the data dictionary and data exports. This process is partially automated against a library of existing scales where possible. For novel forms, the digital version of the form is compared to the paper form and scoring manual as well.

Once collected, data cleaning and curation is typically supported within the clinical EDC system. REDCap users have the option to use REDCap's API to extract data directly into a Brain-CODE workspace, allowing users to extract, subset and analyze their data, entirely within Brain-CODE's secure environment. By extracting the data directly into a workspace, researchers avoid any errors potentially introduced by spreadsheet software, or through encoding conversion issues. For large collaborative studies having a centralized way for multiple users to run outlier analysis scripts in the same environment can help save data analysis resources required to reconfigure pipelines between different users' institutional and personal computers. After the data are exported from the EDC system they will typically be manually reviewed against source documentation, or run through a curation pipeline to detect any outlying erroneous or aberrant data points. Those data points are then reviewed and if appropriate corrected in the source data, or noted as true outliers by the study

FIGURE 4 | Brain-CODE system architecture.
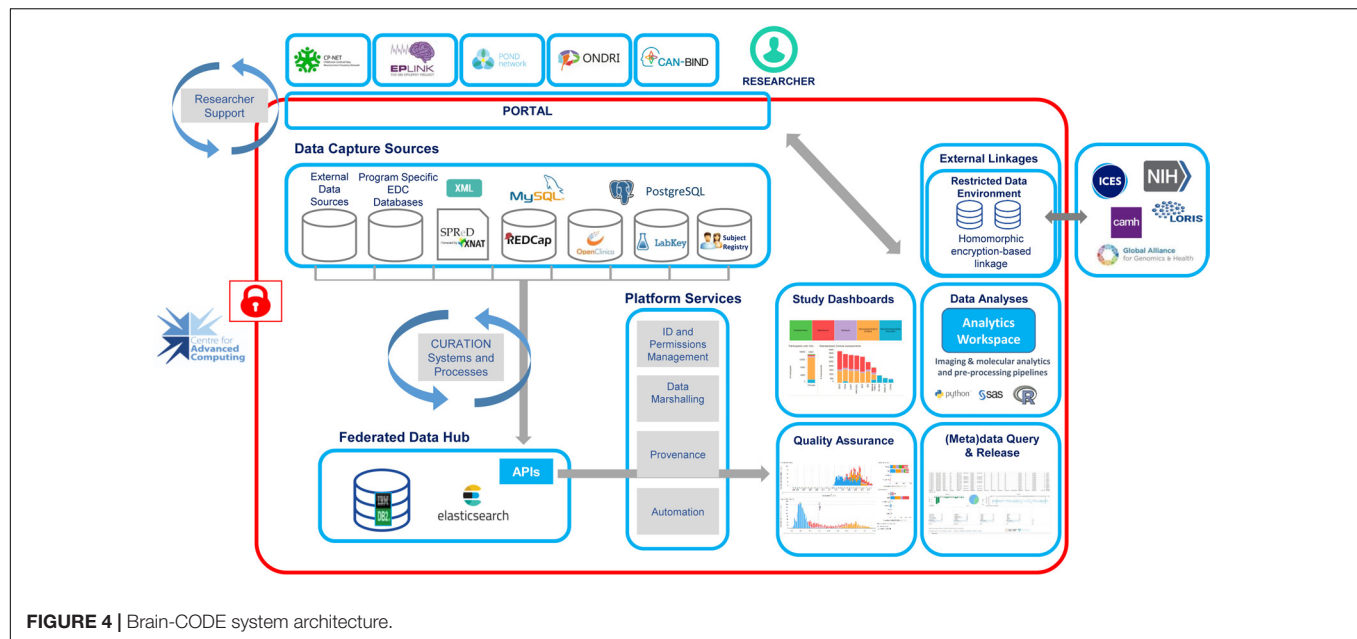
teams in the data capture system itself alongside the raw data.

Imaging Data Management

Many of the studies hosted on Brain-CODE collect various forms of medical imaging, with a particular focus on Magnetic Resonance Imaging (MRI). Although many different scanners are used across the various research sites, all the scanners provide data in the Digital Imaging and Communications in Medicine

TABLE 2 | Summary of Brain-CODE core demographic and clinical CDEs.

Domain	Sub-domain	Brain-CODE CDE
Patient Characteristics	Demographic	Sex, DOB, Handedness, Ethnicity
	SES	Education, Marital Status, Occupation, Income
Physical and Mental Health	Quality of Life	WHO Quality of Life -Short Version[1] KINDL-R[2,3]
	Activities of Daily Living	Sheehan Disability Scale[1]
	Medical Comorbidity	NINDS Medical History[1,2,3]
	Psychiatric Comorbidity	Brief Symptom Inventory[1,2]
Clinical Endpoints	Depression	Quick Inventory of Depressive Symptomatology[1] Revised Children's Anxiety and Depression Scale[2,3]
	Anxiety	Generalized Anxiety Disorder-7[1] Revised Children's Anxiety and Depression Scale[2,3]
	Sleep	Pittsburgh Sleep Quality Index[1,2] Children's Sleep Habits Questionnaire[3]

[1] Adult; [2] adolescent; [3] child.

(DICOM) format[11]. The open source XNAT (eXtensible Neuroimaging Archive Toolkit) project by the Neuroinformatics Research Group at Washington University in St. Louis (Marcus et al., 2007) is used within Brain-CODE to gather, organize, query, and control access to MRI and related data. In addition to DICOM data, XNAT at Brain-CODE is also used to organize and assemble other large binary datasets, including magnetic encephalography (MEG), electrocardiography (ECG), electroencephalography (EEG), ocular computed tomography (SD-OCT), fundal photography, accelerometer and instrumented gait tracking data. Several forms of data that are required to interpret the scans are also included, such as output from stimulus presentation systems such as E-Prime® and even simple scans of hand-written notes taken during sessions.

Data are uploaded via a secure web page into XNAT, either via manual transfer or through bulk upload via scripts. Using sophisticated DICOM interpretation capabilities, XNAT organizes the input files into appropriate sessions, which can be confirmed by the user or the upload script. Once in place, the system generates visual thumbnails, as well as populates a PostgreSQL database with metadata from the DICOM headers as well other environmental sources. These data are made available for searching, retrieval of metadata, and download, under a well-defined authorization structure.

Imaging Standards

The Brain-CODE's XNAT file structure is hierarchically organized, with Project ID folders (i.e., PPPTT_HHH, see naming convention standards in section "Consolidated Data Management" above) occupying the highest level and containing the brain imaging data for that particular project. Within the Project ID folders are a series of subject folders (i.e.,

---

[11] www.dicomstandard.org

PPPTT_HHH_SSSS), each containing Session ID folders of brain imaging data from one or more distinct testing sessions from that subject. A Session ID always begins with a Subject ID, followed by an underscore and a_2-digit Visit ID, and then "_SE" and a 2-digit session number (i.e., 'PPPTT_HHH_SSSS_02_SE01_MR'), as well as an optional "part code" which, if present, is a single lower case letter used to identify and link sessions that were broken up or spread out over time (e.g., intervening days as is the case with MR rescan requests). The session number is then followed by an underscore and a Modality code, which is a string of 2–4 characters indicating the imaging/recording modality.

Anyone requested by the program manager can be given read-only access to a Project folder, while only program manager-approved users who have also taken and passed an upload training tutorial on a non-production version of XNAT can be given upload access to such project(s). In order to reduce the chance of upload errors, XNAT uploaders are given the opportunity to review files and correct any issues at a pre-archive stage. Once archived, however, only Brain-CODE administration staff is allowed to amend files, and only at the written request of uploaders. To ensure provenance and prevent accidental data loss, data are not actually deleted, but session names have the suffix '_deleted' added to them so that the files can be excluded from eventual curation. The one exception to this non-deletion rule pertains to uploaded data that violate ethical restrictions.

Imaging Quality Assurance and Control

Imaging data undergo multiple QA/QC steps as well as curation. Some forms have built-in support via XNAT, such as the manual QC reports, while others represent custom extensions to the basic system. Such custom extensions were started in the Stroke Patient Recovery Research Database (SPReD) (Gee et al., 2010) and extended within Brain-CODE so that the neuroimaging component is referred to as SPReD powered by XNAT, which we will simply refer to as XNAT. The extensive back-end API supported via a representational state transfer (REST) interface allows many manual and automated pipelines to be connected to XNAT, providing automated image transformation, conversion, evaluation and process coordination.

Multi-site brain-imaging studies offer many unique challenges compared to traditional single-site research approaches (see Farzan et al., 2017). Many of these become apparent when reviewing the QA and QC measures that are undertaken for imaging data on Brain-CODE. There are several QA and QC pipelines that are employed on Brain-CODE's XNAT. Due to its DICOM format, many of these pipelines cater to MR data.

SPReD/XNAT naming consistency QC

The naming consistency pipeline is a Python executable script that every night iterates through the data uploaded to SPReD/XNAT in Brain-CODE and checks whether the names of the uploaded files comply with the naming convention described above. In case a non-compliant name is found, the data uploader is notified by e-mail within 24 h, if the naming problem persists more than 7 days, the Program Manager is notified weekly by e-mail until the problem is corrected.

Scan acquisition protocol QC (pipeline operational for scanning sites)

The scan acquisition QC pipeline compares the parameters for all scans within an MRI session from a particular scanning site against a reference protocol defined by the relevant program. The protocols are configured on a project-by-project and scanner-by-scanner basis for each scanning site. The protocol defines a set of pulse sequences that should exist within the session, along with a set of values for the acquisition parameters for each sequence. Each parameter has an upper and lower value against which the actual scan parameters are evaluated. Within 24 h of a failure occurring for any parameter the Program and Brain-CODE neuroimaging managers are notified by e-mail and s/he will contact and work with the scanning site to try to ascertain and correct the cause of the failure. Protocol adherence is aggregated and displayed (**Figure 5**).

Manual/visual QC

It is strongly recommended to every program that they institute a manual visual inspection of all data uploaded to SPReD. The criteria for assessment is based on the Qualitative Quality Control Manual by Massachusetts General Hospital (2013). Results of the manual QC check are recorded in SPReD/XNAT and may be viewed and retrieved from the records of each scan session. If any acquisition fails manual QC the results are discussed with the scanning site within 48 h of the initial patient scan.
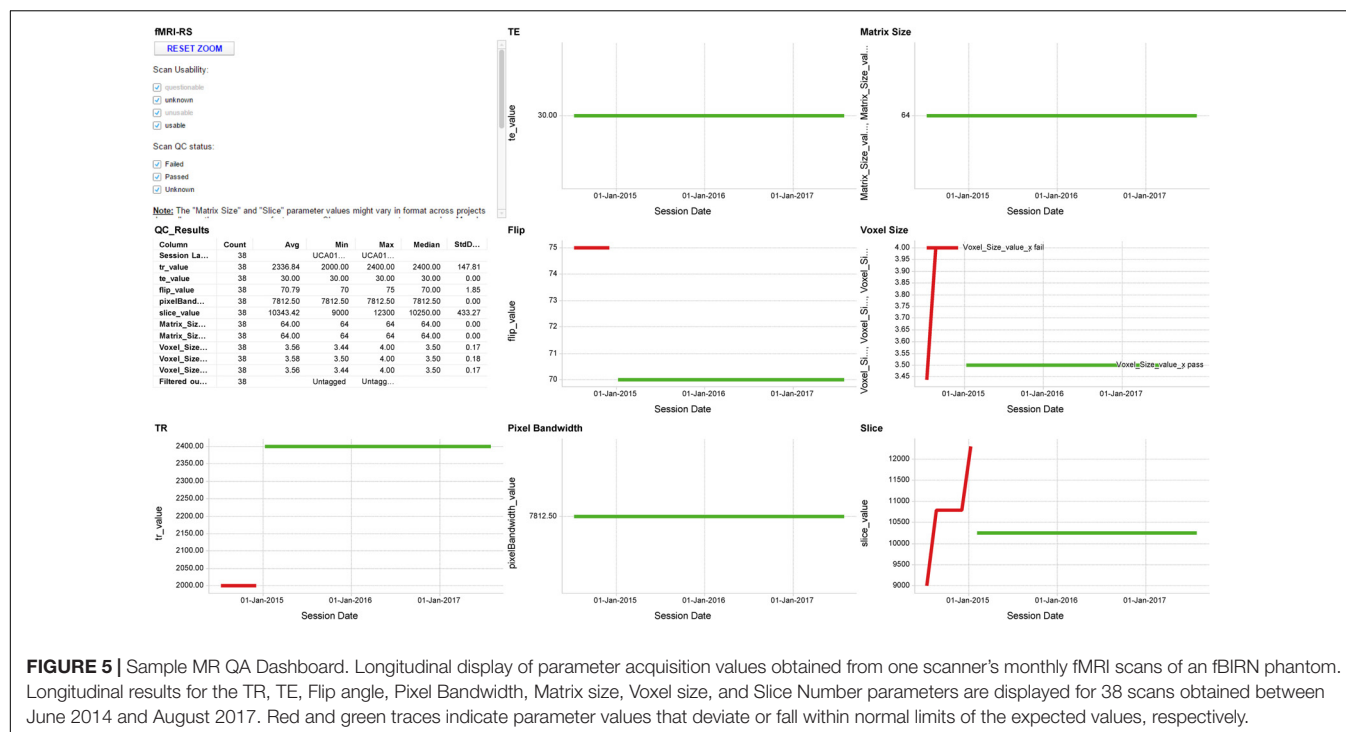
fMRI QA pipeline for fBIRN phantom

The goal of the fBIRN phantom and pipeline software from the Biomedical Informatics Research Network is to provide QA tools for tracking functional MRI (fMRI) imaging performance (Friedman and Glover, 2006). OBI scanning sites have an fBIRN phantom purchased for them by OBI. These phantoms are scanned on a monthly basis and uploaded to XNAT. The fBIRN QA pipeline is then automatically run on these data within 24 h of upload, and a full QA report is generated and stored within the session. The phantom and QA procedures are more formally described in Friedman and Glover (2006), and Glover et al. (2012). Tools for tracking these QA results over time and notification thresholds for scanning sites have been developed using dashboards visualizations. Currently a site is notified if any derived phantom parameter differs from its mean by more than 3 Standard Deviations, based on all previous values acquired to date.

DTI QA pipeline for fBIRN phantom

The utility of the fBIRN spherical gel phantom has been extended to monitoring the performance of DTI acquisitions (Chavez et al., 2018). As is the case for the fMRI QA results, tools for tracking these DTI QA results over time and notification thresholds for scanning sites are available as dashboards.

fBIRN fMRI human QC pipeline

A goal of the Biomedical Informatics Research Network is to provide QC tools for tracking functional MRI imaging performance. A full QC report (index.html) for every fMRI scan generated by running the fBIRN phantom and the fBIRN human pipeline software packages on human data is available through

FIGURE 5 | Sample MR QA Dashboard. Longitudinal display of parameter acquisition values obtained from one scanner's monthly fMRI scans of an fBIRN phantom. Longitudinal results for the TR, TE, Flip angle, Pixel Bandwidth, Matrix size, Voxel size, and Slice Number parameters are displayed for 38 scans obtained between June 2014 and August 2017. Red and green traces indicate parameter values that deviate or fall within normal limits of the expected values, respectively.

the Brain-CODE XNAT file manager in the scan's session folder. Tools for tracking these QC results over time and notification thresholds for scanning sites are available as dashboards.

LEGO phantom QA/QC pipeline
The LEGO phantom and associated pipeline are designed to measure and correct for magnetic field gradient induced geometric distortion, and thereby reduce measurement variability of morphometric measurements from high-resolution T1 MRI scans. The pipeline procedure and its impact on morphometric measurements in neurodegeneraton are described in Caramanos et al. (2010).

MRI registration QC pipeline
The MRI registration pipeline automatically registers (non-linear warping with ANTS[12]) every new high-resolution T1 MRI structural scan to a template and then automatically measures signal-to-noise (SNR) and contrast-to-noise (CNR) in gray matter. The pipeline also includes white matter and automatically measured volumes of interest using the MNI152 registration template and the LPBA40 segmentation atlas (Shattuck et al., 2008).

DICOM header de-identification pipeline
Brain-CODE also employs a number of security pipelines for imaging data. The de-identification pipeline is configured to remove or replace a set of fields within the header of MRI DICOM files and employs a fixed set of fields to be cleared or modified. The appropriate set of fields needs to be reviewed by the users, as they may vary somewhat between projects, between scanners and even between scanner software revision levels.

Defacer pipeline
The Deface DICOM pipeline removes facial features from a DICOM-format T1 image, and produces a defaced DICOM image that is identical to the original in all other respects. It is based on the mri_deface tool released with FreeSurfer and described in Bischoff-Grethe et al. (2007). The output of mri_deface is in Neuroimaging Informatics Technology Initiative (NIfTI) format. The pipeline converts this to DICOM, using the original DICOM file set and the tools mri_convert and analyze2dcm.

Virus pipeline
All new files in the SPReD/XNAT database are scanned for viruses every 24 h.

'Omics and Molecular Data Management
Many of the participating studies collect various molecular and 'omics data as biomarkers for diagnosis and prognosis of disease (Lam et al., 2016; Farhan et al., 2017). Ultimately, Brain-CODE federates these various molecular data modalities with the clinical and imaging data also being collected in these studies, enabling integrated query and analysis of these complex datasets. Brain-CODE currently utilizes the LabKey Server Community Edition, an open source web server developed by the LabKey Corporation[13]. LabKey provides an array of features crucial in efficient management and organization of molecular data from sample tracking, to file archiving to tabularization of finalized datasets. LabKey provides both technology/assay-specific as well as customizable data schemas, making it a flexible and scalable solution for dealing with the large variety of data types being

---

[12]http://sourceforge.net/projects/advants/

[13]www.labkey.com

collected by the Brain-CODE-supported studies. Additionally, LabKey provides a suite of intuitive collaboration features, making it more efficient for investigators across multiple sites to coordinate biological samples, processing and analysis of data.

The installation of LabKey within Brain-CODE provides researchers from multiple labs with a centralized location for the collection and tracking of sample information, raw data files, processed data and associated metadata, including protocol and experimental details, QA/QC and processing metadata of samples and resulting data. Projects are set up to ensure that all these components are appropriately integrated, making it easy to obtain query-based data cuts of processed data and raw data files. Additionally, where possible, final processed data points are structured into a Postgres database which enables more granular and in-depth integrated queries of the molecular datasets with other data modalities. This provides a challenge as 'omics datasets expand in size and complexity, requiring scalable query solutions that can be integrated into existing systems.

'Omics and Molecular Standards

Centralized management of 'omics and molecular data introduce a unique set of challenges including a very diverse set of data modalities, large and ever-growing datasets and files, and harmonization with prominent 'omics databases, like the Gene Expression Omnibus (GEO), GenBank, Sequence Reach Archives, and existing standards [i.e., Minimum Information About a Microarray Experiment (MIAME), Minimum Information about a high-throughput nucleotide SEQuencing Experiment (MINSEQE), Global Alliance for Genomics and Health (GA4GH) and others]. Brain-CODE takes advantage of existing standards and workflows to ensure a thorough capture of all data and associated metadata, while harmonizing with the upload processes of prominent 'omics databases. This in turn makes future submission of data prospectively collected on Brain-CODE simpler for the data producer.

Data Query and Visualization

Several levels of query access are possible on the Brain-CODE system (see **Figure 4**). At the project level, researchers may query their own data within the applicable Brain-CODE data collection platform(s). Post-federation, the Brain-CODE data warehouse structure allows for flexibility in query methodology using either traditional relational database approaches (Structured Query Language, SQL) or unstructured methods such as Lucene via ElasticSearch[14]. This approach allows for future scalability as additional studies and data collection platforms are added to the Brain-CODE system. Metadata compiled for each study are stored in the Brain-CODE system and provides additional context to the data tables.

At the federation level, raw and/or curated federated datasets appropriate to the stage of the Brain-CODE data life-cycle are compiled for exposure to end users. Data visualization and query tools such as TIBCO Spotfire[15] are employed to display and

permit query of aggregated datasets across platforms and, if appropriate, permit users to access and download data tables. Alternative query tools can also be implemented to accommodate different data modalities. Security is ensured by means of user-based access controls at all levels of the data query system.

Brain-CODE currently utilizes Spotfire to develop comprehensive administrative and analytic dashboards, providing unified views on integrated datasets stored in the platform's federation system (**Figure 6**). This takes advantage of the continuous data federation across multiple data sources, allowing near real-time interaction with cross-project, multi-modal datasets. Administrative dashboards allow the Brain-CODE team to monitor the status of all studies on the Brain-CODE platform, describe and quantify data table properties, and apply global QC methods to ensure data quality across all studies and platforms. Project dashboards are configured to provide researchers with fully customizable views of the status of their studies (e.g., recruitment rates, participant profiles), ongoing QC and edit checks (e.g., missing data, protocol violations), and the ability to track ethics and informed consent restrictions. Data exploration and query dashboard interfaces enable permission-based sharing of data, both within study teams and with collaborators, and the broader research community.

Analytics Workspace

Research groups utilizing Brain-CODE present highly variable computational needs during the data curation and analysis stages of their studies. Some are self-sufficient in their capacity to process large volumes of raw data such as MR images or DNA sequences, or to apply machine learning tools on high-dimensional datasets. For example, core sequencing labs used by some research groups have access to their own bioinformatics pipelines, server clusters and expertise required to conduct whole-genome variant detection, differential RNA quantification, or other analysis. Other groups are less equipped, wish to supplement their resources, or prefer to avoid the cost and risk associated with the transfer of large datasets and choose instead to carry out their computations where the data are already aggregated.

To this end, researchers can access a Brain-CODE analytics workspace, a secure environment with dedicated computing resources and necessary software to allow for specialized data processing and analyses. The term "workspace" is used broadly. It can be a cluster of Linux virtual machines (VMs) running the Slurm job scheduler for batch processing; a single Windows VM with SAS or SPSS installed; or an RStudio shared project accessed by data scientists from multiple locations. The analytics workspaces ensure the data are kept securely within the platform to satisfy any privacy and REB requirements while providing easy access to both the data and required resources.

Subject Registry

When researchers enter or upload a dataset for a given participant, a standard Brain-CODE subject ID is assigned. A unique index of projects and Subject IDs is maintained in

---

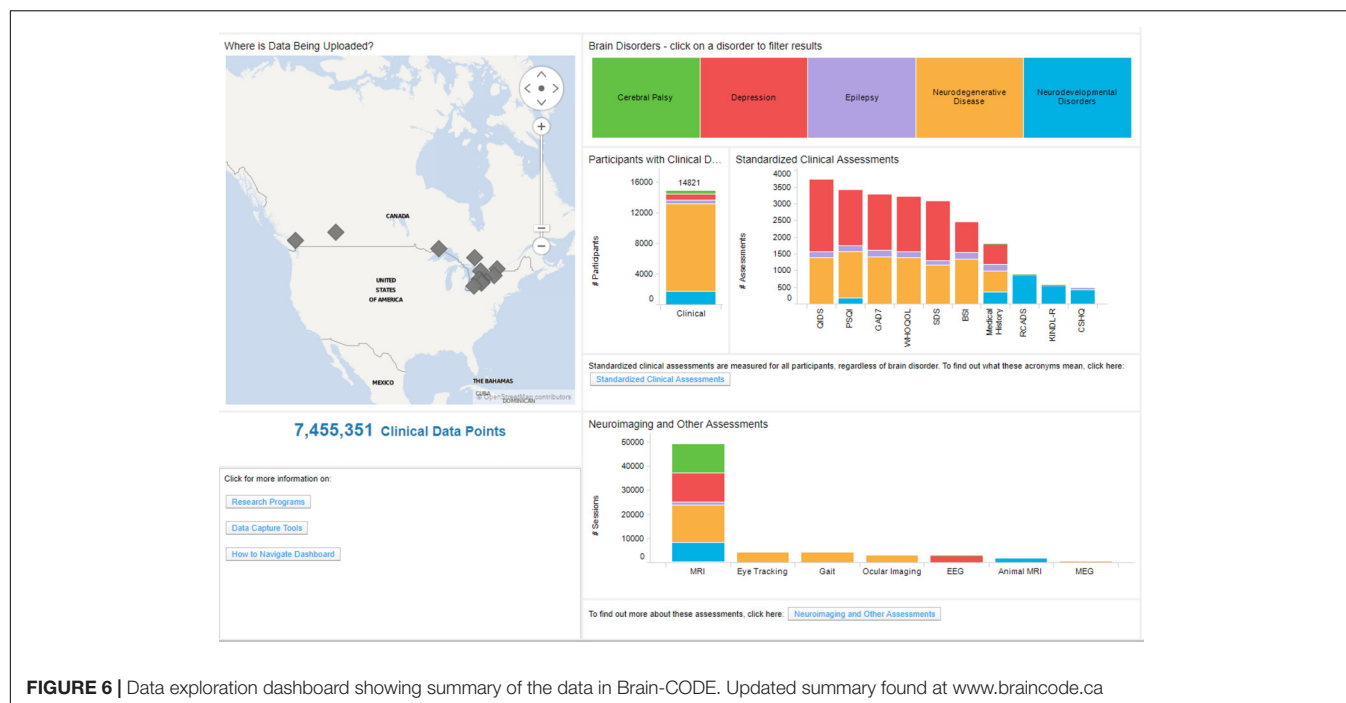[14]www.elastic.co

[15]spotfire.tibco.com

FIGURE 6 | Data exploration dashboard showing summary of the data in Brain-CODE. Updated summary found at www.braincode.ca

the Brain-CODE "Subject Registry," which regularly collects all subject identifiers from the domain-specific databases and provides QC functionality. This critical integration between each database system and the Subject Registry is implemented through a "Reporter" application which extracts necessary information from the database (e.g., the Subject ID) and reports the information to the Subject Registry over REST-based web services.

The Subject Registry also provides functionality for encryption of PHI that can be used to link participant level information across databases, such as a health plan or medical record numbers. Encryption is performed within the user's web browser, and the original value of the element never leaves the research site; only the ciphertext is transmitted and stored in the Subject Registry. Furthermore, the private key required for decryption is maintained by a third-party and is not known to Brain-CODE. The encryption algorithm has a particular homomorphic property which allows mathematical operations and comparisons to be applied to the encrypted data itself, i.e., without the need for decryption. These encryption capabilities not only provide robust safeguards against re-identification of sensitive data, they also enable secure data integration. For example, using a common identifier such as the Ontario Health Insurance Plan number, research data stored in Brain-CODE can be securely linked with administrative health databases such as the Institute for Clinical and Evaluative Sciences without requiring either party to disclose PHI.

Privacy and Security
PHI and De-Identification

To protect the privacy and confidentiality of individuals and security of data held in Brain-CODE, OBI has adopted a Privacy-by-Design approach to creating and implementing protective measures. This policy is specific to Brain-CODE and is based on the 10 Canadian Standards Association (CSA) Privacy Principles (Canadian Standards Association [CSA], 1996). To ensure that privacy is not compromised, direct identifiers that provide an explicit link to a study participant and can identify an individual (i.e., health card number) are removed (or encrypted) to the extent possible. Nonetheless, Brain-CODE may include personal health information that has been collected for the purposes of the research study and analyses (i.e., date of birth). When such information is required and informed consent has been obtained, only researchers involved in the study will have access to it in a firewalled and secure environment. Prior to disclosure to third parties, direct identifiers are removed (or encrypted) to the extent possible.

Ethics Tracking and Monitoring

Brain-CODE operates based upon informed participant consent, meaning that institutional REB approvals and associated informed consents govern which data can be collected, uploaded, de-identified, and shared on Brain-CODE. This information is tracked in a centralized Brain-CODE Ethics Tracking Database, which contains information on the sensitivity of datasets and sharing permissions. The information in the Ethics Tracking Database is linked to each participant via the Subject Registry which allows the tracking and management of data permissions on a participant-by-participant basis.

DATA FEDERATION AND LINKING

By design, research data stored in Brain-CODE are distributed over multiple distinct database applications, each with a unique

underlying data model geared toward the capture of a subset of data modalities. There may be multiple systems in place to support a given modality. For example, clinical assessment data are captured in OpenClinica for some studies, and in REDCap for others. The choice of a clinical data management system for a given study is left to individuals involved in the study and to Brain-CODE personnel providing study support, who collectively take into consideration various factors such as regulatory requirements, training implications, specific features, etc. The same reasoning applies to data capture for other modalities, such as neuroimaging or molecular. While this approach provides maximum flexibility to researchers, allows use of best-of-breed systems developed by domain experts, and enables the platform as a whole to adapt and evolve according to changing needs, it does entail the technical challenge of systematic aggregation of data stored amongst several heterogeneous systems.

To make it possible to search, query, and extract these distributed data, Brain-CODE employs a hybrid "federated data hub" model whereby relevant data from each data source are harmonized and aggregated into one or more repositories (see **Figure 4**). APIs allow cross-system, and hence cross-modality query of federated data for diverse purposes by downstream systems, such as curation pipelines, interactive dashboards, search interfaces, and linkages with data systems external to Brain-CODE.

In its current implementation, federated data sources include: OpenClinica Enterprise^{TM}; REDCap^{TM}; Medidata RAVE^{TM}; LimeSurvey^{TM}; Subject Registry; XNAT; LabKey^{TM}; LORIS (Das et al., 2012). The federated repository is implemented with a combination of IBM InfoSphere Federation Server[16], which provides functions for extracting and staging source data into a DB2 relational database system, and Elasticsearch[17], which provides functions to store data without the need of a pre-defined data model, and functions to index these data for very rapid searching. Query APIs consist of database-level functions and REST-based web services. Automated pipelines are implemented to extract data from source systems and ingest them into the repository. These pipelines execute at varying frequencies for different data types, depending on downstream data consumption needs; generally, federated data are refreshed daily.

Data records stored in the federated repository are associated with metadata. For participant records, these metadata include identifiers which point to the research project, data collection site, and participant associated with the data. Additional participant-related metadata include data sharing permissions derived from informed consent forms and institutional ethics review. These metadata provide a basis for access control implemented in downstream systems. This allows permission-controlled access by researchers to the data they collect from their own studies, as well as data collected across research programs. By federating data from multiple sources and data types, Brain-CODE provides researchers with unprecedented

tools for combining, accessing and analyzing data in novel and powerful ways.

Linkages With External Databases

To augment and complement data in Brain-CODE for enriched analysis and enhanced data outcomes, the system is also used to support linkages with data holdings external to Brain-CODE, such as public data repositories, health administration data holdings, electronic medical records, and legacy databases (see **Figure 4**). For example, a federation of clinical and neuroimaging data has recently been implemented between the Brain-CODE and the LORIS database hosted at McGill University (Das et al., 2012), initially to support data exchange between the OBI-funded Ontario Neurodegeneration Disease Research Initiative program[18] and the Canadian Consortium on Neurodegeneration in Aging[19]. The aim of this project is to ensure that researchers using both platforms can exchange data in an interoperable fashion, with minimal interference to their workflows. This has also laid the groundwork for a recently funded Brain Canada Platform Support Grant, the Canadian Open Neuroscience Platform (CONP), designed to bring together existing Canadian neuroscience platforms, initiatives and networks, and allow them to link, leverage, enhance and expand to form an integrated network. Both LORIS and Brain-CODE platforms will be actively involved in the creation of the CONP. In addition, the system is being extended to enable linkages with other partners, including linking of single-subject data with administrative health data holdings at the Institute for Clinical Evaluative Sciences (Institute for Clinical and Evaluative Services [ICES], 2017), and at the cohort-level with the National Institute of Mental Health Data Archive (Ontario Brain Institute [OBI], 2015).

Other Brain-CODE Deployments

Where possible, Brain-CODE infrastructure was built using open-source tools, which lends itself to replication at other institutions. As discussed elsewhere in this special issue (Rotenberg et al., in review)[20], the Brain-CODE infrastructure has been installed as the central informatics platform for servicing the Krembil Centre for Neuroinformatics at the Centre for Addiction and Mental Health (CAMH). With common software packages installed and similar standardization procedures in place, the groundwork has been laid for other institutions to benefit from this integrative data analytics approach.

DATA CENTER

The computational infrastructure for Brain-CODE is provided and maintained by the Centre for Advanced Computing (CAC) at Queen's University, in Kingston, Canada[21]. CAC is a member of

---

[16]www.ibm.com/analytics/information-server

[17]www.elastic.co

[18]www.ondri.ca

[19]ccna-ccnv.ca

[20]Rotenberg, D., Chang, Q., Potapova, N., Wang, A., Hon, M., Sanches, M., et al., The CAMH Neuroinformatics Platform: a hospital-focused Brain-CODE implementation. Submitted to Frontiers in Neuroinformatics.

[21]https://cac.queensu.ca

the regional Compute Ontario consortium, and affiliated with the Compute Canada national network. The CAC currently supports over 800 research teams across Canada, including academic and industry organizations. Reliable high-speed connectivity with major computing and academic centers is enabled regionally by redundant CAC links to the Ontario Research and Innovation Optical Network (ORION) private fiber optic network, and nationally and internationally through the CANARIE high-speed national backbone. Security best practices including administrative, technical and physical safeguards, and rigorous enforcement of information security policies and procedures, ensure that the platform can satisfy the most stringent regulatory requirements pertaining to the storage and use of sensitive data.

The Brain-CODE deployment at CAC provides a robust, scalable, high performance computing platform that can satisfy long-term processing and storage requirements of multiple large scale research programs, while enabling secure and seamless open access data sharing and analysis, which includes a combined processing performance of 5 TFLOPS (Gee et al., 2010). As usage and requirements of Brain-CODE grow, additional hardware resources can be allocated for increased data storage, specialized data processing, added demand for federation, and intensive concurrent analytical tasks. Brain-CODE public-facing applications and internal systems, including databases, pipelines, and various data handling services, are all deployed with containerization and virtualization technologies (e.g., Docker), allowing optimal use of processor and memory resources while streamlining system maintenance, and enabling the platform to be readily scaled or redeployed into new environments.

DISCUSSION

Ontario Brain Institute supports multidisciplinary collaborative research networks from across Canada focusing on various brain conditions. These programs generate large volumes of data that are integrated within Brain-CODE to support scientific inquiry and analytics across multiple brain disorders and modalities, including clinical, imaging, and 'omics data. By providing access to very large datasets on patients with different neurological disorders and enabling linkages to provincial, national and international databases, Brain-CODE will generate new hypotheses about brain disorders and underlying causes, and ultimately promote new discoveries to improve patient care. As of March 18, 2018, Brain-CODE supports the acquisition, storage and analysis of multi-dimensional data from over 40 Canadian institutions, supporting more than 600 users in over 100 studies and contains data from more than 17,000 study participants and 1,500 animal subjects[22] (see **Figure 6**). These research programs are continually adding data and new programs are being added.

In addition to OBI-supported programs, Brain-CODE also supports the collection, storage and sharing of data from other studies as well. Depending on the requirements of the programs, these data can be collected within the current instance of Brain-CODE with appropriate access control provided to the researchers. Alternately, a Brain-CODE instance can be located within separate servers at the CAC or installed within a separate data center altogether, as is the case with the CAMH instance of Brain-CODE. To facilitate sharing of these data with OBI-sponsored programs, all studies are encouraged to incorporate Brain-CODE CDEs into their protocols, which are made publically available on the Brain-CODE portal[23]. Furthermore, as many granting agencies and journals now require that research data be available for re-use by others, Brain-CODE also provides the infrastructure to support the upload and sharing of data collected outside of Brain-CODE, which can be made publically available or with restricted access to specified persons. Although Brain-CODE does not currently support "regulatory-compliant" clinical trials, plans are well underway to ensure that both the infrastructure and processes are in place to support regulatory-complaint clinical trials, including support of 21-CRF Part 11 compliant EDC systems (i.e., OpenClinica Enterprise) and development and adherence to Standard Operating Procedures, which have been adopted from N2 Network of Networks[24].

One of the key goals of OBI is to support a collaborative approach to neuroscience as a mechanism to bring researchers together to maximize their collective impact (Stuss, 2015; Stuss et al., 2015). To help track the impact of OBI-supported initiatives in fostering collaborations among Ontario's neuroscience community, an "Atlas of Ontario Neuroscience" was developed to explore the growing collaborations both at the individual and institutional level[25]. For example, the "People Connection Map" shows collaborations OBI has fostered through Brain-CODE and other OBI-supported initiatives. It is expected that Brain-CODE, as a centralized informatics platform that supports the management, federation, sharing and analysis of multidimensional neuroscience data, will continue to strengthen and expand these collaborations not only within Ontario but also across the international neuroscience community.

AUTHOR CONTRIBUTIONS

All authors contributed to the development of Brain-CODE and commented on/revised the manuscript at all stages. AV wrote the first draft of the paper and prepared the manuscript.

FUNDING

[22]www.braincode.ca

[23]https://www.braincode.ca/content/getting-started

[24]http://n2canada.ca

[25]http://axon.braininstitute.ca/index.html

REFERENCES

Bischoff-Grethe, A., Ozyurt, I. B., Busa, E., Quinn, B. T., Fennema-Notestine, C., Clark, C. P., et al. (2007). A technique for the deidentification of structural brain MR images. *Hum. Brain Mapp.* 28, 892–903. doi: 10.1002/hbm.20312

Canadian Institutes of Health Research (2013). *CIHR Open Access Policy*. Available at: http://www.cihr-irsc.gc.ca/e/46068.html

Canadian Standards Association [CSA] (1996). *Model for the Protection of Personal Information*. Available at: http://cmcweb.ca/eic/site/cmc-cmc.nsf/eng/fe00076.html

Caramanos, Z., Fonov, V. S., Francis, S. J., Narayanan, S., Pike, G. B., Collins, D. L., et al. (2010). Gradient distortions in MRI: characterizing and correcting for their effects on SIENA-generated measures of brain volume change. *Neuroimage* 49, 1601–1611. doi: 10.1016/j.neuroimage.2009.08.008

Cavoukian, A. (2011). *Privacy by Design: The 7 Foundational Principles*. Toronto, ON: Information and Privacy Commissioner of Ontario.

CFR Title 21, Part 11 (2017). *U.S. Food and Drug Administration*. Available at www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11

Chavez, S., Viviano, J., Zamyadi, M., Kingsley, P. B., Kochunov, P., Strother, S., et al. (2018). A novel DTI-QA tool: automated metric extraction exploiting the sphericity of an agar filled phantom. *Magn. Reson. Imaging* 46, 28–39. doi: 10.1016/j.mri.2017.07.027

Dalkey, N. C., and Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Manag. Sci.* 9, 458–467. doi: 10.1287/mnsc.9.3.458

Das, S., Zijdenbos, A. P., Harlap, J., Vins, D., and Evans, A. C. (2012). LORIS: a web-based data management system for multi-center studies. *Front. Neuroinformatics* 5:37. doi: 10.3389/fninf.2011.00037

Farhan, S. M., Bartha, R., Black, S. E., Corbett, D., Finger, E., Freedman, M., et al. (2017). The ontario neurodegenerative disease research initiative (ONDRI). *Can. J. Neurol. Sci.* 44, 196–202. doi: 10.1017/cjn.2016.415

Farzan, F., Atluri, S., Frehlich, M., Dhami, P., Kleffner, K., and Price, R. (2017). Standardization of electroencephalography for multi-site, multi-platform and multi-investigator studies: insights from the Canadian biomarker integration network in depression. *Sci. Rep.* 7:7473. doi: 10.1038/s41598-017-07613-x

Food and Drug Administration [FDA] (2013). *Electronic Source Data in Clinical Investigations (Guidance for Industry)*. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf

Friedman, L., and Glover, G. H. (2006). Report on a multicenter fMRI quality assurance protocol. *J. Magn. Reson. Imaging* 23, 827–839. doi: 10.1002/jmri.20583

Gee, T., Kenny, S., Price, C., Seghier, M. L., Small, S. L., Leff, A., et al. (2010). Data warehousing methods and processing infrastructure for brain recovery research. *Arch. Ital. Biol.* 148, 207–217.

Glover, G. H., Mueller, B. A., Turner, J. A., van Erp, T. G., Liu, T. T., and Greve, D. N. (2012). Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *J. Magn. Reson. Imaging* 36, 39–54. doi: 10.1002/jmri.23572

Grinnon, S. T., Miller, K., Marler, J. R., Lu, Y., Stout, A., Odenkirchen, J., et al. (2012). National institute of neurological disorders and stroke common data element project - approach and methods. *Clin. Trials* 9, 322–329. doi: 10.1177/1740774512438980

Hsu, C. C., and Sandford, B. A. (2007). The delphi technique: making sense of consensus. *Pract. Assess. Res. Eval.* 12, 1–8.

Institute for Clinical and Evaluative Services [ICES] (2017). *The Ontario Brain Institute and the Institute for Clinical Evaluative Sciences Pool their Strengths to make an Impact on Brain Disorders in Ontario*. Available at: https://www.ices.on.ca/Newsroom/News-Releases/2017/OBI-and-ICES-pool-their-strengths-to-make-an-impact-on-brain-disorders-in-Ontario

Lam, R. W., Milev, R., Rotzinger, S., Andreazza, A. C., Blier, P., Brenner, C., et al. (2016). Discovering biomarkers for antidepressant response: protocol from the Canadian biomarker integration network in depression (CAN-BIND) and clinical characteristics of the first patient cohort. *BMC Psychiatry* 16:105. doi: 10.1186/s12888-016-0785-x

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. doi: 10.1385/NI:5:1:11

Massachusetts General Hospital (2013). *Qualitative Quality Control Manual*. Available at: http://cbs.fas.harvard.edu/usr/mcmains/CBS_MRI_Qualitative_Quality_Control_Manual.pdf

National Institutes of Health (2003). *NIH Data Sharing Policy and Implementation Guidance*. Available at: http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

Ontario Brain Institute [OBI] (2015). *New Canada-U.S. Partnership Paves the Way for International Collaborations in Autism Research*. Available at: https://www.autismspeaks.ca/news/news-press-releases/new-canada-u-s-partnership-paves-the-way-for-international-collaborations-in-autism-research/

Personal Health Information Protection Act [PHIPA] (2004). *S.O. 2004, c. 3, Sched. A*. Available at: https://www.ontario.ca/laws/statute/04p03

Shattuck, D. W., Chiang, M. C., Barysheva, M., McMahon, K. L., de Zubicaray, G. I., Meredith, M., et al. (2008). Visualization tools for high angular resolution diffusion imaging. *Med. Image Comput. Comput. Assist. Interv.* 11, 298–305.

Souza, T., Kush, R., and Evans, J. P. (2007). Global clinical data interchange standards are here! Drug discovery today. *Drug Discov. Today* 12, 174–181. doi: 10.1016/j.drudis.2006.12.012

Stuss, D. T. (2014). The Ontario Brain Institute: completing the circle. *Can. J. Neurol. Sci.* 41, 1–11. doi: 10.1017/cjn.2014.36

Stuss, D. T. (2015). From silos to systems: an integrated approach to neuroscience innovation. *Nat. Rev. Drug Discov.* 14, 295–296. doi: 10.1038/nrd4615

Stuss, D. T., Amiri, S., Rossor, M., Johnson, R., and Khachaturian, Z. (2015). "How we can work together on research and health big data: strategies to ensure value and success," in *Dementia Research and Care: Can Big Data Help?* eds G. Anderson and J. Oderkirk (Paris: OECD Publishing), 61–74. doi: 10.1787/9789264228429-en

Wellcome Trust (2010). *Policy on Data Management and Sharing*. Available at: http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Check for
updates

The CAMH Neuroinformatics Platform: A Hospital-Focused Brain-CODE Implementation

David J. Rotenberg[1]*, Qing Chang[1], Natalia Potapova[1], Andy Wang[1], Marcia Hon[1], Marcos Sanches[1,2], Nikola Bogetic[1], Nathan Frias[3], Tommy Liu[3], Brendan Behan[4], Rachad El-Badrawi[5], Stephen C. Strother[6,7], Susan G. Evans[5], Jordan Mikkelsen[5], Tom Gee[5,6], Fan Dong[5,6], Stephen R. Arnott[5,6], Shuai Laing[5,6], Moyez Dharsee[5], Anthony L. Vaccarino[4,5], Mojib Javadi[5], Kenneth R. Evans[5] and Damian Jankowicz[1]

[1]Krembil Center for Neuroinformatics, Center for Addiction and Mental Health (CAMH), Toronto, ON, Canada, [2]Dalla Lana School of Public Health, Toronto, ON, Canada, [3]Business Intelligence, Center for Addiction and Mental Health (CAMH), Toronto, ON, Canada, [4]Ontario Brain Institute, Toronto, ON, Canada, [5]Indoc Research, Toronto, ON, Canada, [6]Rotman Research Institute, Toronto, ON, Canada, [7]Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

Investigations of mental illness have been enriched by the advent and maturation of neuroimaging technologies and the rapid pace and increased affordability of molecular sequencing techniques, however, the increased volume, variety and velocity of research data, presents a considerable technical and analytic challenge to curate, federate and interpret. Aggregation of high-dimensional datasets across brain disorders can increase sample sizes and may help identify underlying causes of brain dysfunction, however, additional barriers exist for effective data harmonization and integration for their combined use in research. To help realize the potential of multi-modal data integration for the study of mental illness, the Centre for Addiction and Mental Health (CAMH) constructed a centralized data capture, visualization and analytics environment—the *CAMH Neuroinformatics Platform*—based on the Ontario Brain Institute (OBI) Brain-CODE architecture, towards the curation of a standardized, consolidated psychiatric hospital-wide research dataset, directly coupled to high performance computing resources.

Keywords: neuroinformatics, collaborative brain science, medical informatics, XNAT, LabKey

INTRODUCTION

Mental illness affects one in three individuals in their lifetimes (Smetanin et al., 2011), and is the leading cause of disability in Canada (Lim et al., 2008; Mental Health Commission of Canada, 2014; Whiteford et al., 2015) exerting an economic burden estimated at $51 billion per year, including health care costs, lost productivity and reductions in health-related quality of life (Lim et al., 2008; Smetanin et al., 2011). Investigations of mental illness have been enriched by the advent and maturation of neuroimaging technologies and the rapid pace and increased affordability of molecular sequencing techniques (Lynch, 2003; Linden, 2012; Factors Study, 2013; Fu and Costafreda, 2013; Schreiber et al., 2013; Mayberg, 2014; Etkin, 2014; Power et al., 2016; Altman et al., 2016).

While these tools can independently provide powerful insights into the brain's structure and function, directed integration of complementary information holds considerable promise to accelerate discovery and identify cross-modal biomarkers for stratification, diagnosis and treatment of mental illness (Potkin et al., 2014; Mufford et al., 2017).

This increased volume, variety and velocity (Bellazzi, 2014; Lee and Yoon, 2017) of research data, presents a considerable technical and analytic challenge to curate, federate and interpret, requiring the adoption of clear standardizations and aligned infrastructure to coordinate data within and across studies. Neuroinformatics has emerged as a discipline in response to these needs and the progressive evolution of computational psychiatry.

To help realize the potential of multi-modal data towards the study of mental illness, the Center for Addiction and Mental Health (CAMH) constructed a centralized data capture, visualization and analytics environment—the *CAMH Neuroinformatics Platform*—based on the Ontario Brain Institute's (OBI) Brain-CODE platform, enabling the curation of a standardized, consolidated psychiatric hospital-wide research dataset, directly connected to high performance computing resources.

The CAMH Neuroinformatics platform was developed to support core capabilities for institutional researchers:

- Provide a research data management platform that can accommodate and federate the varied research data collected at an academic teaching hospital.
- Provide value to researchers through data visualization, quality reports and intuitive query interfaces.
- Accelerate analytics, by bringing organized data structures and compute power together in an integrated environment.
- Establish a standardized framework, to facilitate cross-institutional data integration.

This article centers on the recent implementation of the CAMH Neuroinformatics Platform, a hospital-focused adoption of the OBI's Brain-CODE model to enable organization of site-wide multi-modal research data to accelerate discovery in mental health. The manuscript addresses the utility and flexibility of Brain-CODE as applied to a hospital environment, and the extensibility of the model, as demonstrated by further developments, including the federation of anonymized clinical records and coupling to unified compute resources.

MATERIALS AND METHODS

To develop a centralized data management and analytics environment, CAMH approached the OBI to review the design elements of the Brain-CODE platform for large-scale multi-dimensional provincial data management, guided by the FAIR data principles (Jeanson et al., 2014, 2016; Wilkinson et al., 2016; Vaccarino et al., 2018). The Brain-CODE model met core criteria appropriate for translation to a research hospital environment.

Flexible
Brain-CODE adopted data capture and organization systems to support the vast array of data types found in brain science. This was essential to meet the requirements posed by the considerable variety of research data collected at CAMH, including magnetic resonance imaging (MRI), positron emission tomography (PET), computed tomography (CT), electroencephalography (EEG), genetics, epigenetics and

proteomics. The systems were also extensible to adapt custom data types and structures. This flexibility extended through the choice of technologies, each of which allow for considerable customization, and open integration with other systems, including the addition of other databases, such as in the case of electronic medical record (eMR) datasets (CERNER), administrative data (such as the Institute for clinical evaluate sciences, ICES), population health and economics data.

Scalable
The Brain-CODE platform was demonstrated to be highly scalable as applied to province-wide neuroscience studies supported through the OBI. This scalability met the requirements to aggregate data across hospital research programs and to facilitate national and international multi-site studies. The platform needed to be capable of handling the hundreds of active studies CAMH supports and the thousands of closed/archived projects of historical data.

Secure
Brain-CODE was developed with a "privacy by design" approach, embedding security into each layer of implementation based on the 10 Canadian Standards Association (CSA) Privacy Principles[1]. This aligned with the requirements of a hospital environment, where security of research and clinical data are paramount. Granular and defined access levels, built around the structure of research endeavors, provided a solid framework for secure access.

Accessible
The individual applications and interfaces are highly accessible to the research community. The web-based tools are intuitive and well-suited for data collection in each domain (imaging, molecular, clinical), and require limited training to reach a sufficient level of comfort for systems adoption and can be made accessible securely within the hospital network, through centralized two-factor authentication.

Research Domain Databases

The Neuroinformatics Platform consists of open-source domain-specific database systems, federated through a DB2 back-end to provide subject-by-subject records. Each database interface is designed for a particular data-type, e.g., imaging, molecular, clinical, allowing for intuitive data entry and handling (**Figure 1**).

REDCap[2] is used to capture behavioral and clinical assessments, including harmonized common data elements (CDEs) and self-report surveys (Harris et al., 2009). The CAMH instance of REDCap was validated in collaboration with the internal research ethics board (REB) and IT Security teams, to enable usage in regulated clinical trials in compliance with Health Canada.

XNAT[3] (adapted as SPReD[4]) is used to store and organize medical imaging data, including MRI, CT/PET and EEG. MRI

---

[1]https://www.csagroup.org/codes-standards/health-safety/
[2]http://project-redcap.org
[3]http://www.xnat.org
[4]https://sites.google.com/a/research.baycrest.org/informatics/spred

FIGURE 1 | Overview of the Center for Addiction and Mental Health (CAMH) Neuroinformatics Platform. Data sources include XNAT (imaging), LabKey (molecular), REDCap (electronic case report forms, eCRFs) and case of electronic medical record (eMR) case of electronic medical record (eMR) datasets (CERNER; electronic health records, eHRs) which are federated into a central DB2 database. Federated datasets are available to compute resources (compute and Hadoop clusters) and easily accessible through dashboards and software notebooks through the Neuroinformatics Portal.

data are stored in both their original DICOM and derived formats, including NiFTI, MINC and ANALYZE, automatically generated through pre-processing pipelines.

LabKey[5] is used for the coordination and storage of biological specimens and molecular data, including genetics, epigenetics and proteomics. This system supports both raw data storage and direct tabularization of results.

The databases support both original source data, derived values (e.g., quality assessments and final results) and pre-processed datasets (e.g., artifact correction).

All subject data are collected with informed consent, under a study-specific REB protocol. Authentication has been harmonized through the hospital-wide active directory system and within each sub-system, rights are limited depending on user-role to maintain security and to separate projects based on REB study protocol. All changes to user access require submission of an auditable electronic form, which requires principle investigator sign-off. This extends to visualization dashboards and individual table access for analytics (Clinical data access has additional constraints, described in the section specific to clinical record data).

In the current phase, external access can be provided to researchers who are named collaborators on the REB study protocol. Access requires confidentiality agreements and a centrally administered institutional account.

Data Federation

Multi-modal datasets are federated using the IBM InfoSphere Federation Server[6], which provides a thin, virtual data definition layer that allows seamless communication with data sources. A flexible API backend utilizes this federation capability to provide subject-oriented, de-normalized mart-like data tables, within a DB2 database environment. Data are linked, by unique standardized research participant IDs, across each source system, to generate a subject-level, profile for each individual.

Visualization and Query Interface

Visualization and federated query interfaces are provided through TIBCO Spotfire[7]. Dynamic dashboards, refreshed daily, provide federated data views across data sources. These data views are served to specific research teams, defined by their study protocols and data requirements.

---

[5]https://www.labkey.com

[6]http://www-03.ibm.com/software/products/en/ibminfofedeserv
[7]http://spotfire.tibco.com/

Dashboards provide visualizations that can be constructed from any data or metadata in the source systems (XNAT, REDCap, LabKey and CERNER). Filters can be applied directly through interactive selection, or a variable-by-variable query interface, to refine cohorts for data export to compute cluster environments or local processing centers.

Statistical packages included with the dashboard implementation allow for clustering, regression and stratification of datasets, presenting an initial layer of rapid exploration and visualization, prior to offloading to dedicated compute resources for further investigation.

Neuroinformatics Portal

Access to each of the data entry tools, dashboards and analytics applications are coordinated through a central *Neuroinformatics Portal* (**Figure 2**). This primarily web-based design of the Neuroinformatics Platform provides a consolidated gateway for CAMH researchers to interact with their data.

Central Subject Registry

A central ledger of all participants entered into the platform is supported by the *Subject Registry* (Vaccarino et al., 2018). As a core component of this tool, medical record numbers (MRNs) or health card numbers can be encrypted on entry, allowing for the identification of common participants across studies. As participants can be identified across studies, visits and encounters, the subject registry facilitates longitudinal dataset linkages and simplified hospital-wide research participant review and oversight.

The Neuroinformatics Platform operates based upon informed participant consent, meaning that institutional REB approvals and associated informed consents govern what data can be collected, uploaded, de-identified and shared.
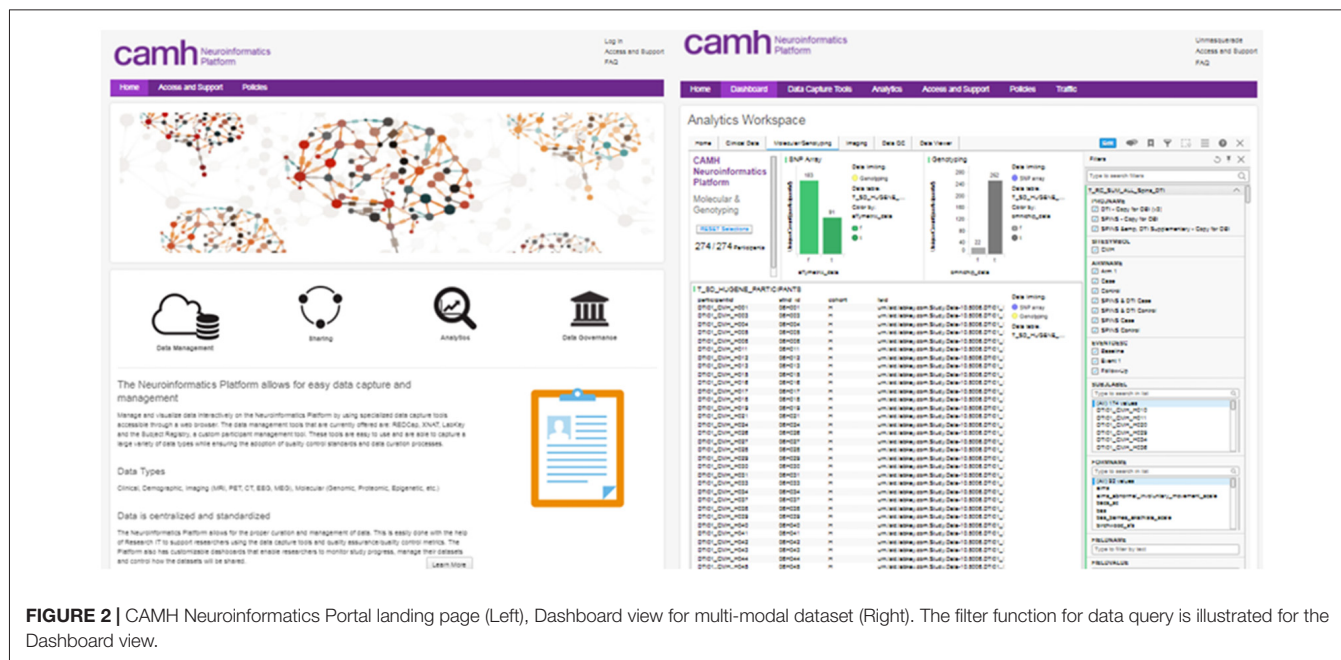
This information is tracked in an Ethics Tracking Database, (supported through a validated REDCap instance) which contains information on the sensitivity of datasets and sharing permissions. The information in the Ethics Tracking Database is linked to each participant via the Subject Registry which allows the tracking and management of data permissions on a participant-by-participant basis.

Quality Assurance

Prompt and reproducible metrics of data quality are essential to ensuring the integrity of research data. This is supported through the Neuroinformatics Platform in the implementation of quality control and quality assurance (QC/QA) scripts launched for new data entry into data collection systems, and the presentation of data quality dashboards.

QC scripts and summary dashboards are a core component of the XNAT implementation. Automated QC scripts are initiated on a nightly basis, with computation coordinated through the CAMH compute cluster. These include naming convention checkers, scan protocol checkers and both human and phantom QC/QA. Functional MRI data quality is assessed using phantom and human implementations of the fBIRN pipeline from the Biomedical Informatics Research Network (Friedman and Glover, 2006; Glover et al., 2012). Structural data, specifically T1 scans are evaluated through an MRI registration pipeline that automatically registers (non-linear warping with ANTS[8] every new high-resolution T1 MRI structural scan to a template and then automatically measures signal-to-noise (SNR) and contrast-to-noise (CNR) in gray matter. The pipeline also includes white matter measures and automatically measures volumes of interest using the MNI152 registration template and the LPBA40 segmentation atlas (Shattuck et al., 2008).
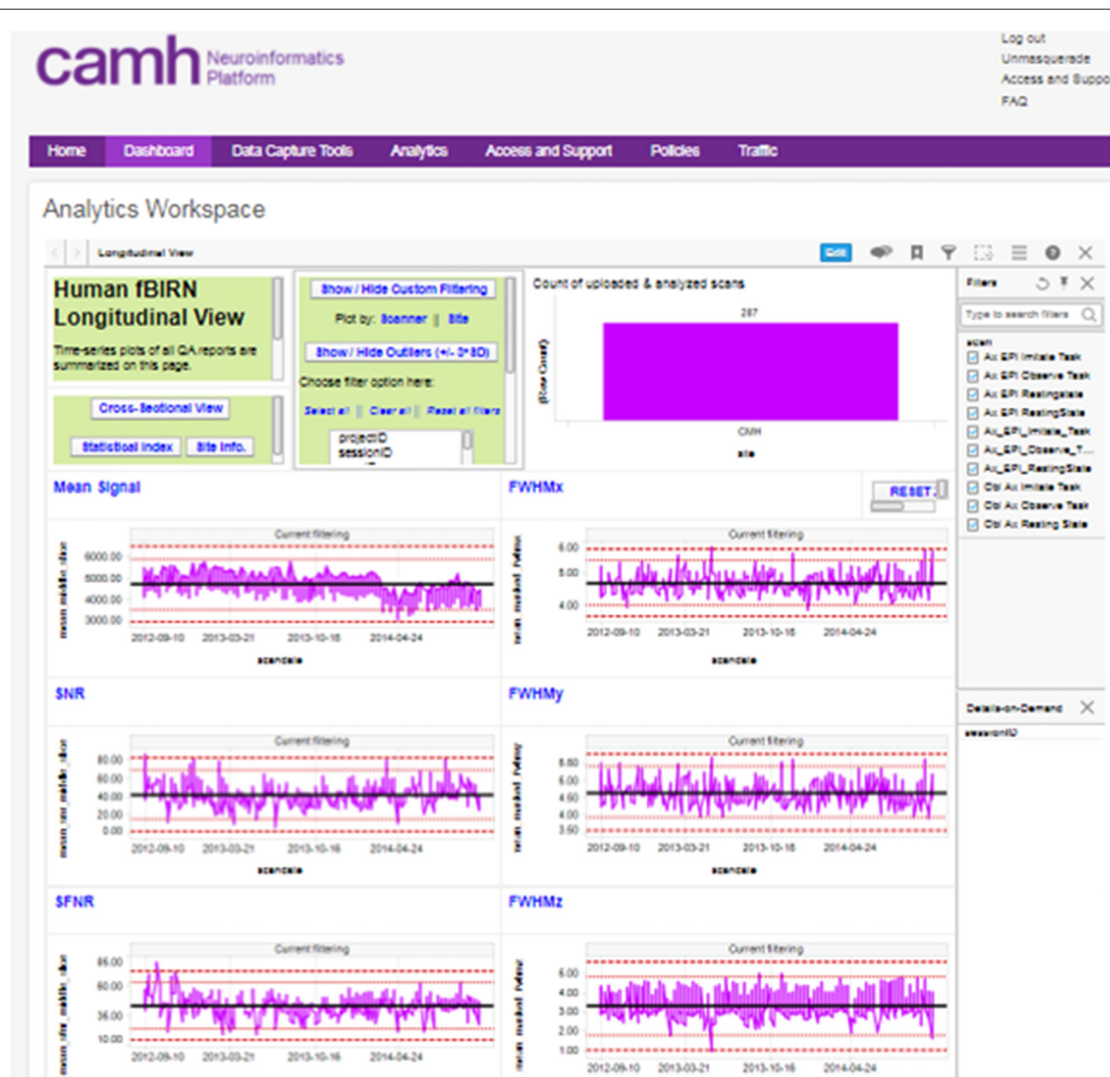
---

[8]https://sourceforge.net/projects/advants/

FIGURE 2 | CAMH Neuroinformatics Portal landing page (Left), Dashboard view for multi-modal dataset (Right). The filter function for data query is illustrated for the Dashboard view.

FIGURE 3 | Example, "global" longitudinal quality assurance and quality control (QA/QC) dashboard for functional MRI (fMRI) data.

The reports generated by these scripts are captured and associated with the subject/imaging sessions in XNAT, and are further aggregated into interactive dashboards visible to each research group, with both cross-sectional and longitudinal views across the study (**Figure 3**).

A "global" imaging data quality dashboard also provides a full view of all data entered into the Neuroinformatics platform. This assists with the evaluation of overall site performance, long-term trending and detection of outlier data.

Any number of pipelines can be added to these workflows to support additional QC or pre-processing steps on neuroimaging datasets that can be executed on secure local compute resources.
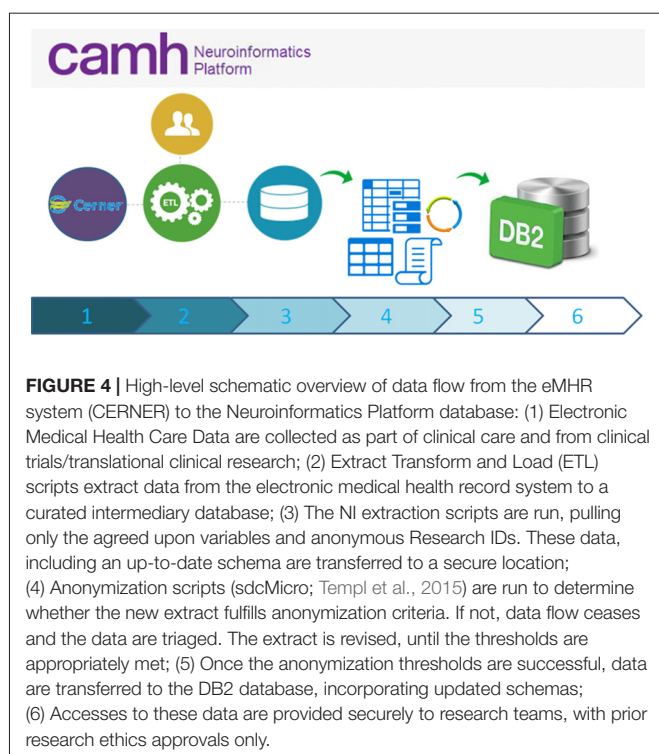
XNAT—Anonymization

In additional to anonymization of clinical data discussed in the following sections, de-identification of imaging data is also handled through automated pipelines (Li, 2011). A DICOM header de-identification pipeline is applied to remove or replace fields within the DICOM files. The fields to be modified are configurable and are evaluated on a project-by-project basis, dependent on REB protocol and in co-ordination with the CAMH privacy office. High-resolution structural MRI scans have been demonstrated to allow for the reconstruction of facial features and identification of individuals (Schimke et al., 2011). To support anonymization of imaging data a defacing pipeline based on the MRI_deface tool (FreeSurfer; Bischoff-Grethe et al., 2007) can be applied to data to remove facial features from T1 images. In combination these pipelines can reduce the likelihood of re-identification of imaging datasets.

Clinical Datasets

Electronic Medical Health Records

CAMH is a "HIMMS EMRAM Stage 7" hospital with highly coordinated electronic medical health records

FIGURE 4 | High-level schematic overview of data flow from the eMHR system (CERNER) to the Neuroinformatics Platform database: (1) Electronic Medical Health Care Data are collected as part of clinical care and from clinical trials/translational clinical research; (2) Extract Transform and Load (ETL) scripts extract data from the electronic medical health record system to a curated intermediary database; (3) The NI extraction scripts are run, pulling only the agreed upon variables and anonymous Research IDs. These data, including an up-to-date schema are transferred to a secure location; (4) Anonymization scripts (sdcMicro; Templ et al., 2015) are run to determine whether the new extract fulfills anonymization criteria. If not, data flow ceases and the data are triaged. The extract is revised, until the thresholds are appropriately met; (5) Once the anonymization thresholds are successful, data are transferred to the DB2 database, incorporating updated schemas; (6) Accesses to these data are provided securely to research teams, with prior research ethics approvals only.

systems (CERNER) deployed to clinicians as I-CARE[9]. These records are of significant interest to researchers, both as independent sources of information related to patient prognosis, progression and outcomes, as well as when combined with research data, such as medical imaging and molecular expression.

Clinical datasets are provisioned to researchers through two methods: (1) anonymized aggregate data for review by internal researchers; and (2) data cuts specific to a REB approved study, including retrospective chart review, restricted only to those named members on the study protocol and in agreement with identifiers included when and if allowed by the REB.

Coordinated data extracts of the hospital electronic medical health record system, are staged through the federation server, and then imported into the DB2 data-lake (**Figure 4**). These records, including demographics, laboratory results and pharmacological information, are linked to extended research datasets, securely bridging clinical and research domains.

Anonymization

The capability to ensure anonymization is essential to the use of clinical data in a research environment. Three primary methods are applied to clinical data prior to exposure to research systems: direct identifier removal, k-anonymity and l-diversity (using the sdcMicro software package; Templ et al., 2015).

Direct identifiers, such as name, address, phone number, date of birth, as well as IDs (such as medical record and health card numbers) are isolated and removed. These variables are masked (i.e., cells are nullified or the columns are removed entirely

---

[9]www.cerner.com

from the table) in the standard extract for the Neuroinformatics Platform.

Anonymous "Research IDs," following the CAMH research naming convention, are generated in-place of other internal IDs tied to identifiable information. The clinical team retains secure mappings, to recover information if re-identification is required.

Variables that pose an identification risk, alone or in combination with others, including Gender, Age Group, Local Health Integration Network (LHIN) and Major Program are considered Key Variables. To enforce k-anonymity (Samarati and Sweeney, 1998; El Emam et al., 2009) the datasets are processed for unique values or unique combinations of up to three variables, which if identified are nulled.

Confidentiality is breached if a set of subjects with the same combination of (up to 3) key variables has the same diagnosis. In these cases subjects have their key variables nulled, to enforce l-diversity, while guaranteeing a minimum loss of information (Machanavajjhala et al., 2007).

After the application of k-anonymity and l-diversity algorithms, risk measures related to the probability of identification are calculated, to help ensure low risk of disclosure and monitor the disclosure risk changes over time.

These metrics are calculated for each subject in two ways: (i) "Disclosure Risk" for a given subject is calculated as 1 divided by the number of subjects with the same combination of key variables. It will be 1 if the subject has a unique combination of key variables, considered unacceptable; and (ii) "Sample Frequency on Subsets," is calculated using the Special Unique Detection Algorithm (SUDA2). A Data Intrusion Simulation (DIS) score is derived for each subject based on considerations of how unique the combination of key variables is (with higher weight for combination of fewer variables).

The output of this process is an anonymized dataset and a report that highlights the changes made to the original data and summaries of the risk measures of anonymity.

If the risk probability for re-identification exceeds established thresholds, further processing will cease and the data will remain in the staging area. The dataset is adjusted in coordination with clinical teams until the re-identification risk is reduced to within the set parameters.

Cohort Explorer

The anonymized medical record data are utilized to provide a cohort explorer for study feasibility evaluation and statistical power calculations (**Figure 5**). This follows a similar model to Informatics for Integrating Biology and the Bedside (i2b2; Murphy et al., 2006), by providing a layer of access to explore cohorts across the breadth of the clinical records systems. The clinical data can be further combined with research data from the other source databases through the common DB2 backend.

As the anonymization process can reduce the amount of information available, the aggregate cohort explorer is intended primarily as an overview to identify study feasibility. Further variables do continue to be added to the aggregate clinical extract, to make these data more valuable for analysis. Where further information is required, detailed extracts are provisioned in
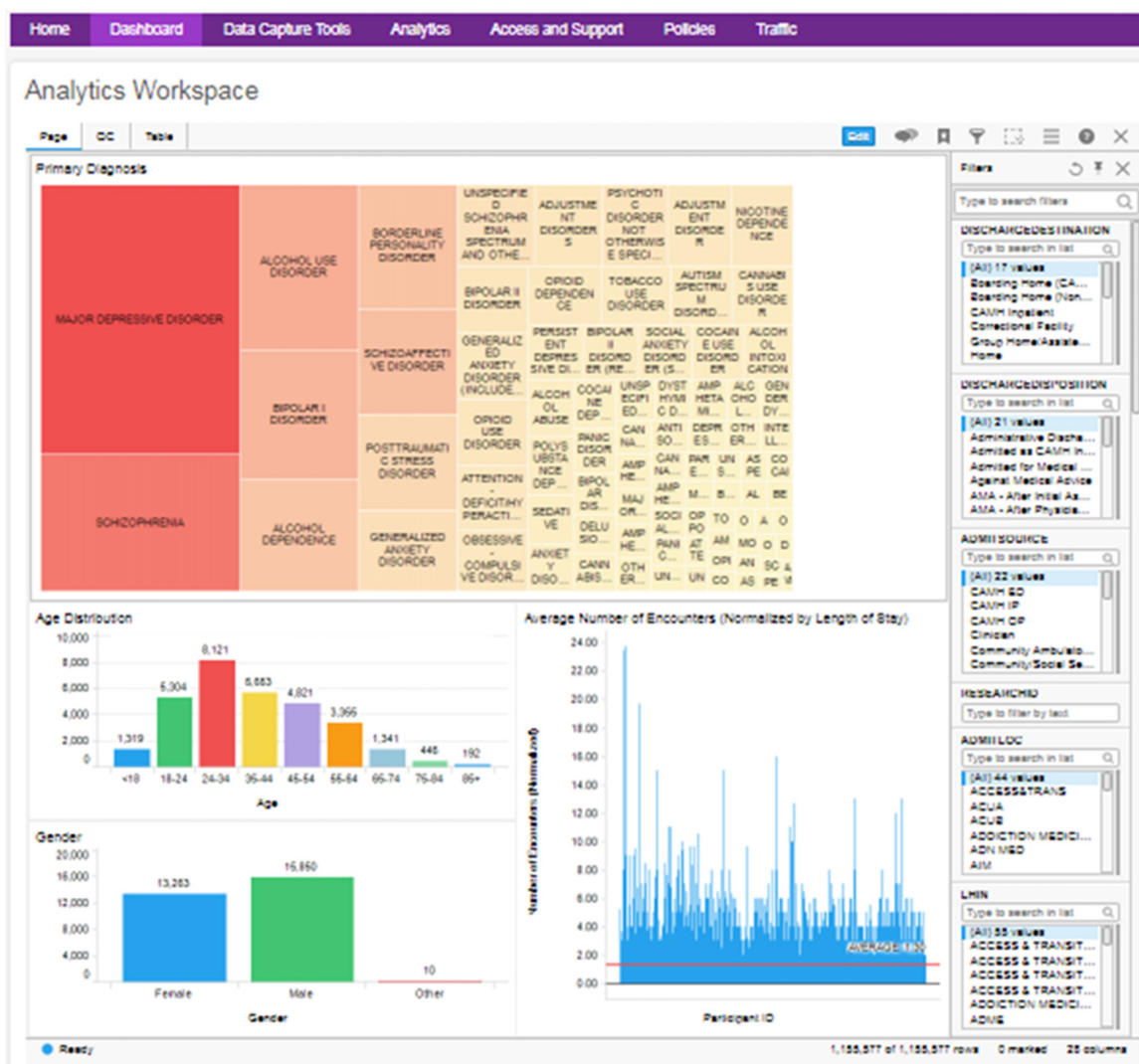
FIGURE 5 | Example clinical data cohort explorer dashboard, with visualizations of diagnosis, age, gender and average encounter (filterable by diagnosis). QC data and full table views are also made available.

alignment with a specific REB protocol, and are anonymized as far as possible, to limit identifiers to those prescribed by the REB.

Analytics

Compute Cluster

The scale and complexity of medical imaging and molecular datasets necessitates substantial compute capabilities for the pre-processing, QC measures and post-processing. The Neuroinformatics Platform was designed with full connectivity to a local high-performance compute cluster to handle computationally demanding tasks (**Figure 6**).

Automated scripts initiated from the source databases (e.g., XNAT and LabKey) are issued to the local compute infrastructure, on dedicated secure queues.

Researchers are able to access their datasets, via queries and data pointers directly from the compute clusters. The

architecture adopted, minimizes data transfers, and includes a tightly connected network on a unified VLAN, at 10 GB bandwidth, between all Neuroinformatics platform resources.
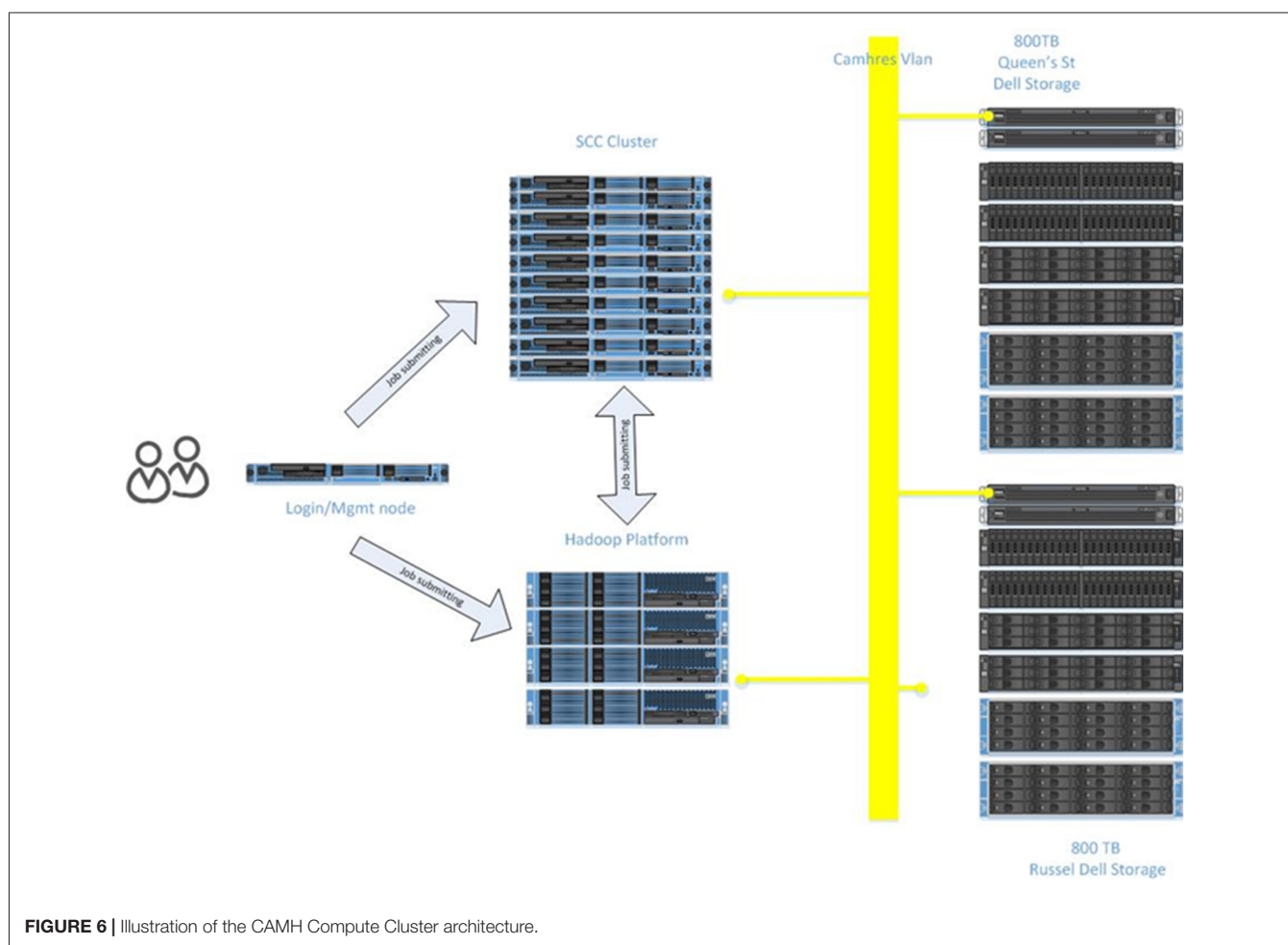
Hadoop Analytics Environment

To enable analysis of increasingly large datasets, otherwise intractable to conventional approaches, the Neuroinformatics Platform was implemented alongside dedicated Hadoop infrastructure[10]. The DB2 database is imported in full to a *HIVE 2.0*[11] framework, utilizing *SQOOP*[12], with secured permissions enforced on a column-by-column level. Researcher's datasets are directly accessible to the active workspace to apply pipelines and processing frameworks.

---

[10]http://hadoop.apache.org
[11]http://hive.apache.org/
[12]https://hortonworks.com/apache/sqoop/

FIGURE 6 | Illustration of the CAMH Compute Cluster architecture.

Notebook Interfaces

To further the accessibility and web-based design of Brain-CODE, notebooks for Python (Jupyter[13]) and R (RStudio[14]), common languages in computational psychiatry, are accessible through the central *Neuroinformatics Portal*. These notebooks can process code on either a classical compute cluster, or dedicated Hadoop environment, leveraging SparkR[15] and PySpark[16] to seamlessly execute pre-developed code, without recoding in native MapReduce.

Data Center

The infrastructure to support the functions of the Neuroinformatics Platform is maintained locally at CAMH across three secure data centers. The Neuroinformatics Platform adopted a design philosophy to ensure no "single point of failure." Each server includes redundant components, network connections, RAID storage configurations and hot-spares.

Each database application (XNAT, LabKey, Spotfire and DB2) is provisioned with a dedicated development and production server, physically separated between the primary data centers for high availability and disaster recovery purposes.

Similar to the OBI, CAMH has adopted a primarily virtualized architecture, using Oracle VM (OVM[17]). While there are some limitations in performance as a result of virtualization, this approach provides substantial operational benefits, notably: (a) flexible deployment; (b) efficient snapshots for backup; and (c) simplified fail-over procedures to initialize replicated VMs. The virtual machines are distributed to a cluster of computers, through OVM, such that they can be dynamically deployed/re-deployed as required in case of hardware failure (**Figure 7**).

Data storage and backup functions are supported through a 1.9 PB high performance storage system. Replication at the file-level is conducted on an hourly basis, between the primary and secondary storage sites, maintaining concurrent mirrors of all raw and processed data (MRI, EEG, PET, etc.). Point-in-time snapshots are taken each day, and retained up to 1-month, such that accidental deletions or modifications can be rolled back for up to 30-days. Daily extracts of system configurations are included in the file-level replication.
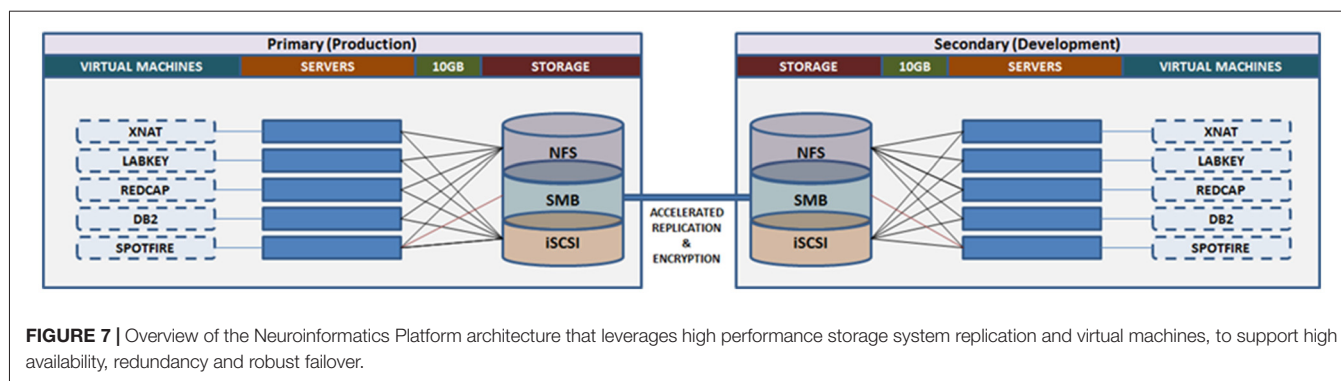
---

[13]http://jupyter.org/
[14]https://www.rstudio.com/
[15]https://spark.apache.org/docs/latest/sparkr.html
[16]http://spark.apache.org/docs/2.1.0/api/python/pyspark.html

[17]http://www.oracle.com/technetwork/server-storage/virtualbox/overview/index.html

FIGURE 7 | Overview of the Neuroinformatics Platform architecture that leverages high performance storage system replication and virtual machines, to support high availability, redundancy and robust failover.

The Neuroinformatics platform virtual machines are stored on a separate file system, accessed via Internet Small Computer Systems Interface (iSCSI), on the central storage system. This allows for block-level replication of the entire virtual machine environment between primary and secondary sites. Automated scripts allow for the preparation and launch of replicated virtual machines, (either the production or development frameworks), which can resume access of the research data from the file-level replica. Both replication channels are further accelerated using specialized hardware, and encrypted point-to-point.

The research storage systems, Neuroinformatics platform and high performance compute environments are interconnected by 10 GB optical fiber, under a single harmonized research VLAN. This interconnect provides high bandwidth and low latency to synchronize research data across applications and analytics systems. The compute infrastructure includes a Hadoop deployment (HortonWorks), a GPU node for machine learning applications, and 45 high memory (128–256 GB RAM) compute nodes, providing over 1,000 available processing cores.

This implementation of the Brain-CODE model on new hardware architecture demonstrates the flexibility of the design, and that it can be deployed under differing data center conditions.

RESULTS

The Neuroinformatics platform has provided a key component of technological infrastructure that affords researchers with a standardized framework for data organization and analytics, accessible through a centralized portal. The system, based on the OBI Brain-CODE framework, has been able to support and federate the varied research data types collected at CAMH.

At the time of writing, the CAMH Neuroinformatics Platform supports 38 distinct research projects, spanning each of the hospital's primary research programs, with 3,61,777 total participant records (including medical records), and anticipated growth of 30,000 records per year (**Table 1A**). The total datasets span 20 TB and adoption across the hospital has been strong, with the web-based access model allowing for simplified study management and data transfer.

Supported studies range multiple disorders and cross-lifespan populations including, Pediatric, Geriatric, Neurodegenerative (Alzheimer's, Parkinson's), Depression, Bipolar Disorder,

TABLE 1 | Summary table of data currently stored in the Center for Addiction and Mental Health (CAMH) Neuroinformatics platform.

(A) Neuroinformatics platform data summaries.

Primary database	Number of Participants
XNAT—Medical Imaging	2,878
REDCap—Assessments	13,514
LabKey—Molecular	15,385
eMHR—Clinical	330,000
Total	361,777

(B) Neuroimaging summary.

Modality	Scans
DTI	2277
EEG	1837
T1	2600
T2	4322
fMRI	22108
Total	33144

Number of primary records stored in each database, XNAT, REDCap, LabKey and from clinical records, Summary of Neuroimaging data types currently stored in XNAT.

Psychosis, Autism, Schizophrenia and Addictions (Alcohol, Nicotine). Data types include MRI: *Functional, Structural and Diffusion* (**Table 1B**), PET, EEG, Whole Genome Sequencing, Methylation, Chip Sequencing, MicroArray Sequencing and RNA Sequencing.

Each study varies in the data types that are required for collection and management. While not all studies include data across each domain (e.g., studies with molecular and assessment data, or imaging data only), several studies collect extensive phenotypic data incorporating medical imaging, molecular, assessment and clinical data for each participant.

In particular, the Social Processes Initiative in Neurobiology of the Schizophrenia(s) (SPINS[18]; $d = 109$) and Preventing Alzheimer's Dementia With Cognitive Remediation Plus Transcranial Direct Current Stimulation in Mild Cognitive Impairment and Depression (PACt-MD[19]). These studies collect biological samples, neuroimaging data (with the inclusion of EEG data for PACt-MD) and extensive clinical and assessment data. The complex data collected by these studies are well supported

[18]http://camhstudies.ca/cgi-bin/ver2/findCAMHstudy_study.php?
[19]https://sunnybrook.ca/research/content/?page = sri-groups-nppc-proj-7

by the CAMH Neuroinformatics platform as the system can accommodate the diverse data types and combine records through federation: SPINS (LabKey—274, REDCap—174, XNAT—319), PACt-MD (LabKey—230, REDCap—212, XNATtextemdash217).

Tight coupling with computing environments supporting classic parallel clusters and Hadoop frameworks, avoids intermediary data transfer and storage, staging an environment for rapid data exploration at-scale. The analytics environments supporting the platform have run a total of 2,50,000 parallel jobs, spanning QC, pre and post-processing workloads. The use of web-based "notebook" interfaces has simplified access to computational resources and abstracted complexities of queue management from the user.

Federated records can be served securely to researchers through interactive dashboards, functionally refined to suit the requirements of each study. Dynamic query and filter functions embedded within the platform have enabled researchers to quickly identify cohorts and data sub-sets, greatly enhancing data accessibility, and shifting time spent on "collating data" to scientific interpretation.

The development of the Neuroinformatics platform establishes the first phase of hospital-wide data integration by providing a consistent framework for data organization and management.

DISCUSSION

Sophisticated systems are required to handle the increasing variety and scale of neuropsychiatric research data. These challenges are well-known to the neuroscience community, which have driven the development of several concurrent approaches to manage complex datasets including, FBIRN FIRE, COINS, LORIS, NeuroLOG, i2b2 and the Human Brain Project Medical Informatics Platform (Amorim et al., 2016).

Comparisons to Similar Approaches

The Function Biomedical Informatics Research Network (FBRIN) and Federated Informatics Research Environment (FIRE; Keator et al., 2015) are a set of open-source integrated tools for multi-side or multi-study neuroimaging studies that includes many critical components such as central authentication, online clinical data entry forms and the Human Imaging Database[20] for data management. FIRE also includes the FBIRN image processing stream[21]. This is a valuable open-source resource for functional MRI studies and shares several similarities with the CAMH deployment, including imaging and clinical assessment data collection, a centralized database and coupling to compute for processing pipelines (both including components of FBIRN QA). The two systems also share querying interfaces with URLs pointing to image data for staging downstream analyses. The Brain-CODE instance includes additional data sources, and has been extended for use with other neuroimaging data types, such as DTI.

The Collaborative Informatics and Neuroimaging Suite COINS[22] (Scott et al., 2011) is based on an open-source model that includes web-based tools to manage studies, subjects, imaging, clinical data, and other assessments, including a standard metadata model and powerful query interface. It acts as an institutional data repository that enables secure data sharing with a focus on PHI considerations. While there are advantages to the COINS deployment, as compared to XNAT as a standalone implementation, such as longitudinal tracking and standardized meta-data and data structures, the Brain-CODE model incorporates strict standardization, including naming conventions for longitudinal studies and enhanced query through the federation system.

The Longitudinal Online Research and Imaging System (LORIS; Das et al., 2016) is an extensible web-based data management system that supports multiple data types, including imaging, clinical, behavior and genetics. The system includes capabilities to store, process and disseminate datasets and is used for a variety of multi-site studies with instances used worldwide.

It shares many conceptual components of Brain-CODE and the CAMH implementations, and provides valuable insight into the challenges of managing longitudinal research data. Compatibility between Brain-CODE and LORIS (Vaccarino et al., 2018) using the underlying federation model has been achieved to bridge these two systems towards data integration for specific studies.

NeuroLOG (Batrancourt et al., 2014) provides a middleware data management layer, to share heterogeneous and distributed neuroimaging data using a federated approach. Shared information can be captured through a multi-layer ontology and federation schema to harmonize heterogeneous data. This shares some components of the federation approach used in Brain-CODE, through standardization approaches and centralized federate schema. The challenge of combining retrospective heterogeneous datasets from legacy databases, still presents a challenge that may be addressed through the use of mappable data models and semantic database frameworks, discussed in relation to future work.

i2b2 is an open-source system developed to provide tools for clinical investigators to integrate medical records and clinical research data (Murphy et al., 2010). This provides similar functionality to the eMHR and research data integration provided through the CAMH instance of Brain-CODE, including a query tool to search applicable datasets, and are access restricted based on REB review. The i2b2 implementation also has two primary methods of exposure of medical record data: an anonymized dataset of researcher review and restricted matched sets of patients and controls based on study-specific requirements. The i2b2 platform uses ontologies to standardize data, and can link to diverse databases to access other data streams and connections to compute resources are supported. This system does lack the visualization capabilities afforded by Spotfire, and would rely on the source systems for QC.

---

[20]www.nitrc.org/projects/hid
[21]http://www.nmr.mgh.harvard.edu/~greve/fbirn/fips/

[22]http://coins.mrn.org

The Human Brain Projects' Medical Informatics Platform can provide support for hospital clinical data to be uploaded and maintained locally for analysis (without leaving the originating institution), and also view aggregated data for large-scale analyses of clinical data across hospitals (Galili et al., 2014). The CAMH Neuroinformatics platform approach is more similar to the i2b2 model, with data not yet federated in aggregate with other institutions. Secure aggregates are made available for internal use, however, the inclusion of data models and ontologies, coupled with anonymization, can allow for more broad clinical data integration.

In the context of the current environment of Neuroinformatics approaches, the Brain-CODE model as implemented at CAMH and its extension through local resources represents a unique application with several advantages suited to the hospital-focused use-case.

The Brain-CODE model utilizes open-source databases for imaging, molecular data and assessment data, leveraging the specialization of those tools to their data type(s). This supports a highly diverse range of modalities, as required by CAMH research programs. This also allows for new systems to be added, or replaced, as the Neuroinformatics field evolves. The underlying federation model has also been demonstrated to be flexible combining data from multiple internal and external data sources, such as eMHR data at CAMH.

The Neuroinformatics platform combines many of the key components of comparable systems, with flexibility to extend additional capabilities, to enrich the existing datasets and move towards institutional data integration.

Limitations

There are several limitations to the implemented system, from a user perspective, repository perspective and the data federation approach.

Development of QC and pre-processing pipelines still requires substantial coding and subject matter expertise. Technical teams are available to assist researchers in implementing their pipelines under the existing frameworks (XNAT, LabKey), however, considerable knowledge of coding is still required to ensure that these analyses work seamlessly.

Work was done to allow for direct data download after querying federated study records. While this has been successfully implemented for imaging data from XNAT, the system can only provide tabularized molecular data from LabKey and has not yet been built to pull raw data in bulk through the query interface.

Many scripts and tools rely on standardized naming conventions for MRI scans, which have been shown to vary considerably between studies. While re-naming can be performed during data import, and look-up tables established to accommodate cases where re-naming is not possible, further effort is required to generalize the system to better handle varied conventions, particularly when considering inclusion of externals sources. The authors are also aware of the importance of provenance and maintaining full information about the sequence that was performed for data generation, which may preclude

re-naming. Additional efforts are underway institutionally to standardize acquisitions.

As discussed in sections "Electronic Medical Health Records" and "Cohort Explorer" there are two methods that clinical data extracts can be made available: (a) as an anonymized aggregate; (b) a more complete extract dependent on REB approval for chart review. The anonymization framework for the clinical data is by design, conservative and results in a reduction of information available in the output records that make these data less useful to investigators. Ongoing efforts include adding additional variables to the aggregated extracts to provide further information of interest, while maintaining anonymization criteria.

A primary limitation of the current iteration of the Neuroinformatics platform is that while data are federated on a subject-by-subject level, they are not "integrated" across studies. These limitations exist for legal, ethical and technical reasons. Foremost patient consent and approved REB protocols are not generalized for data sharing. There are further technical limitations imposed by the initial federation software layer. It is a key component of current and future directions to implement an interoperability system, through Blue Brain Nexus[23]) supporting permutable data models and detailed provenance. Blue Brain Nexus was designed to fully support the FAIR data model, and is currently being implemented within the Neuroinformatics Platform to allow for findability, interoperability, accessibility and reproducibility. Through the development of standardized and consistent data model(s) that incorporate data sharing options and the technology of Nexus, will support the aggregation of different data sources for the purpose to increase study sample sizes and enrich a growing institutional dataset.

CONCLUSION

The CAMH Neuroinformatics Platform represents a unique application of the Brain-CODE model in a hospital setting, enabling data management and federation between research and clinical domains, in support of treatment units and study centers.

The CAMH Neuroinformatics Platform supports individual study data management and lays the foundations to facilitate hospital-wide dataset federation, through the application of data standardization and CDEs[24]. Maximizing statistical power is challenging in individual studies, however, integration of related data through participatory consortia such as, ENIGMA (Kelly et al., 2018), ADNI (Yao et al., 2017), HCP (Van Essen et al., 2013), bioCADDIE (Cohen et al., 2017) demonstrate that more expansive datasets can be established for analysis. Thorough data integration requires the adoption of data models, ontologies and semantic description frameworks, to map between existing data and optimally coordinate future data collection and institutional developments of harmonized consent models. These capabilities are critical to the development of large-scale datasets from across diverse studies and the formulation of longitudinal datasets. The extensibility of the OBI Brain-CODE model allows

---

[23]https://github.com/BlueBrain/nexus
[24]https://www.braincode.ca/content/getting-started#toc-2

these developments to be applied effectively at the individual domain-database level and the intermediary and federation layers.

Further expansion of the Neuroinformatics Platform will be focused on establishing a core integration layer that will ensure data remain "live," in a searchable, accessible and interconnected format, under the FAIR data principles. Provenance will also be a cornerstone of future initiatives, embedded into the platform, to provide clear descriptors of data origins, processing pipelines and derivations, and to coordinate authorship in accordance with applicable data trajectories.

The implemented model of primarily open-source tools represents a crucial component of research infrastructure, which can be replicated at institutions of varying size to approach "Big Data" and multi-modal investigations. The Neuroinformatics Platform at CAMH will continue to accumulate multi-dimensional medical imaging, molecular and clinical data to further expand a rich dataset for large-scale studies to further our understanding of the etiology, progression and treatment of psychiatric illness.

AUTHOR CONTRIBUTIONS

Contributions to the development of the CAMH Neuroinformatics Platform: DR: led implementation of CAMH Neuroinformatics Platform. QC: administration for core research storage system and hardware. NP: REDCap development and integration. AW: CAMH cluster and compute administration, management of virtual infrastructure. MH: Neuroinformatics Platform administration support. MS: biostatistics support and clinical data anonymization. NB: cohort explorer clinical dashboard developer. NF: business intelligence lead for clinical data management. TL: data warehouse lead for clinical data extraction. BB: OBI project management support. RE-B: data federation development. SS: development of SPReD. SE: dashboard and visualization development. JM: molecular Data and Subject Registry support. TG: implementation project manager. FD: imaging database development. SA: quality control scripts and dashboard development. SL: imaging database development. MD: indoc lead for implementation. AV: clinical database development, common data elements. MJ: molecular database and dashboard development. DJ: institutional project lead. All authors have approved the manuscript and agree with submission to Frontiers in Neuroscience.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Altman, R. B., Prabhu, S., Sidow, A., Zook, J. M., Goldfeder, R., Litwack, D., et al. (2016). A research roadmap for next-generation sequencing informatics. *Sci. Transl. Med.* 8:335ps10. doi: 10.1126/scitranslmed.aaf7314

Amorim, R. C., Castro, J. A., Rocha, J., and Ribeiro, C. (2016). A comparison of research data management platforms: architecture, flexible metadata and interoperability. *Univers. Access Inf. Soc.* 16, 851–862. doi: 10.1007/s10209-016-0475-y

Batrancourt, B., Dojat, M., Gibaud, B., and Kassel, G. (2014). A multilayer ontology of instruments for neurological, behavioral and cognitive assessments. *Neuroinformatics* 13, 93–110. doi: 10.1007/s12021-014-9244-3

Bellazzi, R. (2014). Big data and biomedical informatics: a challenging opportunity. *Yearb. Med. Inform.* 9, 8–13. doi: 10.15265/iy-2014-0024

Bischoff-Grethe, A., Ozyurt, I. B., Busa, E., Quinn, B. T., Fennema-Notestine, C., Clark, C. P., et al. (2007). A technique for the deidentification of structural brain MR images. *Hum. Brain Mapp.* 28, 892–903. doi: 10.1002/hbm.20312

Cohen, T., Roberts, K., Gururaj, A. E., Chen, X., Pournejati, S., Alter, G., et al. (2017). A publicly available benchmark for biomedical dataset retrieval: the reference standard for the 2016 bioCADDIE dataset retrieval challenge. *Database* 2017:bax061. doi: 10.1093/database/bax061

Das, S., Glatard, T., MacIntyre, L. C., Madjar, C., Rogers, C., Rousseau, M. E., et al. (2016). The MNI data-sharing and processing ecosystem. *Neuroimage* 124, 1188–1195. doi: 10.1016/j.neuroimage.2015.08.076

El Emam, K., Dankar, F. K., Issa, R., Jonker, E., Amyot, D., Cogo, E., et al. (2009). A globally optimal k-anonymity method for the de-identification of health data. *J. Am. Med. Inform. Assoc.* 16, 670–682. doi: 10.1197/jamia.m3144

Etkin, A. (2014). Neuroimaging and the future of personalized treatment in psychiatry. *Depress. Anxiety* 31, 899–901. doi: 10.1002/da.22325

Factors Study. (2013). Data retrieved. Available online at: http://www.healthdata.org/data-visualization/gbd-compare

Friedman, L., and Glover, G. H. (2006). Report on a multicenter fMRI quality assurance protocol. *J. Magn. Reson. Imaging* 23, 827–839. doi: 10.1002/jmri.20583

Fu, C. H., and Costafreda, S. G. (2013). Neuroimaging-based biomarkers in psychiatry: clinical opportunities of a paradigm shift. *Can. J. Psychiatry* 58, 499–508. doi: 10.1177/070674371305800904

Galili, T., Mitelpunkt, A., Shachar, N., Marcus-Kalish, M., and Benjamini, Y. (2014). "Categorize, cluster, and classify: a 3-C strategy for scientific discovery in the medical informatics platform of the human brain project," in *Discovery Science. DS 2014. Lecture Notes in Computer Science*, eds S. Džeroski, P. Panov, D. Kocev and L. Todorovski (Cham: Springer), 73–86.

Glover, G. H., Mueller, B. A., Turner, J. A., van Erp, T. G. M., Liu, T. T., Greve, D. N., et al. (2012). Function biomedical informatics research network recommendations for prospective multicenter functional MRI studies. *J. Magn. Reson. Imaging* 36, 39–54. doi: 10.1002/jmri.23572

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., and Conde, J. G. (2009). Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inform.* 42, 377–381. doi: 10.1016/j.jbi.2008.08.010

Jeanson, F., Amiri, S., Pisterzi, L. F., Javadi, M., Pong, J., Evans, K. R., et al. (2014). Brain-CODE: a large-scale neuroinformatics platform for deep and broad data. *Front. Neuroinform.* 8:18. doi: 10.3389/conf.fninf.2014.18.00018

Jeanson, F., Popovich, C., Behan, B., El-Badrawi, R., Strother, S. C., Gee, T., et al. (2016). From integration to visualization of multisite brain data on brain-CODE. *Front. Neuroinform.* 10:52. doi: 10.3389/conf.fninf.2016.20.00052

Keator, D. B., van Erp, T. G. M., Turner, J. A., Glover, G. H., Mueller, B. A., Liu, T. T., et al. (2015). The function biomedical informatics research network data repository. *Neuroimage* 124, 1074–1079. doi: 10.1016/j.neuroimage.2015.09.003

Kelly, S., Jahanshad, N., Zalesky, A., Kochunov, P., Agartz, I., Alloza, C., et al. (2018). Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Mol. Psychiatry* 23, 1261–1269. doi: 10.1038/mp.2017.170

Lee, C. H., and Yoon, H.-J. (2017). Medical big data: promise and challenges. *Kidney Res. Clin. Pract.* 36, 3–11. doi: 10.23876/j.krcp.2017.36.1.3

Li, Y. (2011). "Preserving privacy in structural neuroimages," in *Data and Applications Security and Privacy XXV. DBSec 2011. Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer.

Lim, K.-L., Jacobs, P., Ohinmaa, A., Schopflocher, D., and Dewa, C. S. (2008). A new population-based measure of the burden of mental illness in Canada. *Chronic Dis. Can.* 28, 92–98. Available online at: https://www.canada.ca/content/dam/phac-aspc/migration/phac-aspc/publicat/hpcdp-pspmc/28-3/pdf/cdic28-3-2eng.pdf

Linden, D. E. (2012). The challenges and promise of neuroimaging in psychiatry. *Neuron* 73, 8–22. doi: 10.1016/j.neuron.2011.12.014

Lynch, C. A. (2003). Institutional repositories: essential infrastructure for scholarship in the digital age. *Portal Libr. Acad.* 3 327–336. doi: 10.1353/pla.2003.0039

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). L-diversity: privacy beyond K-anonymity. *ACM Trans. Knowl. Discov. Data.* 3:52. doi: 10.1145/1217299.1217302

Mayberg, H. S. (2014). Neuroimaging and psychiatry: the long road from bench to bedside. *Hastings Cent. Rep.* 44, S31–S36. doi: 10.1002/hast.296

Mental Health Commission of Canada. (2014). Why investing in mental health will contribute to Canada's economic prosperity and to the sustainability of our health care system. Available online at: https://www.mentalhealthcommission.ca/English

Mufford, M. S., Stein, D. J., Dalvie, S., Groenewold, N. A., Thompson, P. M., and Jahanshad, N. (2017). Neuroimaging genomics in psychiatry-a translational approach. *Genome Med.* 9:102. doi: 10.1186/s13073-017-0496-z

Murphy, S. N., Mendis, M. E., Berkowitz, D. A., Kohane, I., and Chueh, H. (2006). Integration of clinical and genetic data in the i2b2 architecture. *AMIA Annu. Symp. Proc.* 2006:1040.

Murphy, S. N., Weber, G., Mendis, M., Gainer, V., Chueh, H. C., Churchill, S., et al. (2010). Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J. Am. Med. Inform. Assoc.* 17, 124–130. doi: 10.1136/jamia.2009.000893

Potkin, S. G., Macciardi, F., and van Erp, T. G. (2014). Imaging genetics approaches to identify mechanisms in severe mental illness. *Biol. Psychiatry* 76, 436–437. doi: 10.1016/j.biopsych.2014.07.020

Power, B. D., Nguyen, T., Hayhow, B., and Looi, J. (2016). Neuroimaging in psychiatry: an update on neuroimaging in the clinical setting. *Australas. Psychiatry* 24, 157–163. doi: 10.1177/1039856215618525

Samarati, P., and Sweeney, L. (1998). Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. *Harv. Data Priv. Lab* 12:2017.

Schimke, N., Kuehler, M., and Hale, J. (2011). "Preserving privacy in structural neuroimages," in *Data and Applications Security and Privacy XXV. DBSec 2011. Lecture Notes in Computer Science*, ed. Y. Li (Berlin, Heidelberg: Springer), 301–308.

Schreiber, M., Dorschner, M., and Tsuang, D. (2013). Next-generation sequencing in schizophrenia and other neuropsychiatric disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 162B, 671–678. doi: 10.1002/ajmg.b.32156

Scott, A., Courtney, W., Wood, D., de la Garza, R., Lane, S., King, M., et al. (2011). COINS: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets. *Front. Neuroinform.* 5:33. doi: 10.3389/fninf.2011.00033

Shattuck, D. W., Chiang, M.-C., Barysheva, M., McMahon, K. L., de Zubicaray, G. I., Meredith, M., et al. (2008). Visualization tools for high angular resolution diffusion imaging. *Med. Image Comput. Comput. Assist. Interv.* 11, 298–305.

Smetanin, P., Stiff, D., Briante, C., Ahmad, S., and Khan, M. (2011). *The Life and Economic Impact of Major Mental Illnesses in Canada: 2011–2041. Prepared for the Mental Health Commission of Canada.* Toronto, ON: RiskAnalytica.

Templ, M., Kowarik, A., and Meindl, B. (2015). Statistical disclosure control for micro-data using the R package sdcMicro. *J. Stat. Softw.* 67:4. doi: 10.18637/jss.v067.i04

Vaccarino, A. L., Dharsee, M., Strother, S., Arnott, S., Behan, B., Dafnas, C., et al. (2018). Brain-CODE: a secure neuroinformatics platform for management, federation, sharing and analysis of multi-dimensional neuroscience data. *Front. Neuroinform.* 12:28. doi: 10.3389/fninf.2018.00028

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn human connectome project: an overview. WU-Minn HCP consortium. *Neuroimage* 80, 62–79. doi: 10.1016/j.neuroimage.2013.05.041

Whiteford, H. A., Ferrari, A. J., Degenhardt, L., Feigin, V., and Vos, T. (2015). The Global burden of mental, neurological and substance use disorders: an analysis from the global burden of disease study 2010. *PLoS One* 10:e0116820. doi: 10.1371/journal.pone.0116820

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18

Yao, X., Yan, J., Ginda, M., Börner, K., Saykin, A. J., and Shen, L. (2017). Mapping longitudinal scientific progress, collaboration and impact of the Alzheimer's disease neuroimaging initiative. *PLoS One* 12:e0186095. doi: 10.1371/journal.pone.0186095

frontiers
in Neuroinformatics

APPIAN: Automated Pipeline for PET Image Analysis

Thomas Funck[1,2]*, Kevin Larcher[3], Paule-Joanne Toussaint[1], Alan C. Evans[1,3,4] and Alexander Thiel[2,4]

[1] Montreal Neurological Institute, McGill University, Montreal, QC, Canada, [2] Jewish General Hospital and Lady Davis Institute for Medical Research, Montreal, QC, Canada, [3] Biospective, Inc., Montreal, QC, Canada, [4] Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

APPIAN is an automated pipeline for user-friendly and reproducible analysis of positron emission tomography (PET) images with the aim of automating all processing steps up to the statistical analysis of measures derived from the final output images. The three primary processing steps are coregistration of PET images to T1-weighted magnetic resonance (MR) images, partial-volume correction (PVC), and quantification with tracer kinetic modeling. While there are alternate open-source PET pipelines, none offers all of the features necessary for making automated PET analysis as reliably, flexibly and easily extendible as possible. To this end, a novel method for automated quality control (QC) has been designed to facilitate reliable, reproducible research by helping users verify that each processing stage has been performed as expected. Additionally, a web browser-based GUI has been implemented to allow both the 3D visualization of the output images, as well as plots describing the quantitative results of the analyses performed by the pipeline. APPIAN also uses flexible region of interest (ROI) definition— with both volumetric and, optionally, surface-based ROI—to allow users to analyze data from a wide variety of experimental paradigms, e.g., longitudinal lesion studies, large cross-sectional population studies, multi-factorial experimental designs, etc. Finally, APPIAN is designed to be modular so that users can easily test new algorithms for PVC or quantification or add entirely new analyses to the basic pipeline. We validate the accuracy of APPIAN against the Monte-Carlo simulated SORTEO database and show that, after PVC, APPIAN recovers radiotracer concentrations within 93–100% accuracy.

Keywords: open science, automation, pipeline, software, quality control, PET

INTRODUCTION

The increasing availability of large brain imaging data sets makes automated analysis essential. Not only is automated analysis important for saving time, but it also increases the reproducibility of research. No existing post-reconstruction positron emission tomography (PET) software package satisfies all the needs of researchers, specifically code that is free, open-source, language agnostic, easily extendible, deployable on web platforms as well as locally, and including all necessary processing steps prior to statistical analysis. We therefore present APPIAN (Automated Pipeline for PET Image Analysis) a new open-source pipeline based on NiPype (Gorgolewski et al., 2011) for performing automated PET data analysis. The starting point for APPIAN are reconstructed PET images on which all necessary processing steps are performed to obtain quantitative measures from the original PET images (**Figure 1**). In conjunction with the reconstructed PET image, APPIAN

FIGURE 1 | APPIAN performs all processing steps necessary to obtain quantitative parameters from reconstructed PET images. Flexible definition of ROI allows use of APPIAN for a wide variety of experimental designs. Integrated QC helps ensure that the pipeline performs as expected.
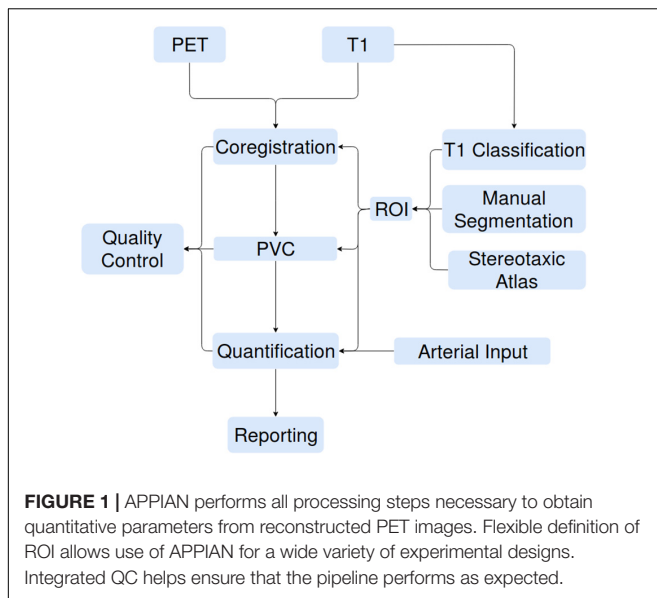
uses T1-weighted MR images to define regions of interest (ROI) that are used at multiple processing stages. Briefly, APPIAN (1) coregisters the T1 MR image with the PET image, (2) defines ROI necessary for later processing steps, (3) performs partial-volume correction (PVC), (4) calculates quantitative parameters, (5) produces a report of the results, and finally, (6) performs QC on the results (see **Figure 1** for a schema of APPIAN, and Discussion section for a detailed description of the pipeline, complete with flowchart).

MATERIALS AND METHODS

Pipeline Overview
Coregistration
Positron emission tomography images are coregistered to the corresponding non-uniformity corrected (Sled et al., 1998) T1 MR-images using a six parameter linear fitting algorithm that minimizes normalized mutual information. The algorithm is based on minctracc[1] and proceeds hierarchically by performing iterative coregistration at progressively finer spatial scales (Collins et al., 1994). Coregistration is performed in two stages, the first using a binary mask for the PET and the T1 MR images, respectively, to obtain a coarse coregistration. This is followed by a second registration step to refine the initial fit between the PET and T1 MR images without the use of the binary images.

MR Image Processing
T1 structural preprocessing is performed if the user does not provide a binary brain mask volume and a transformation file that maps the T1 MR image into stereotaxic space. If these inputs are not provided, APPIAN will automatically coregister the T1 MR image to stereotaxic space. By default, the stereotaxic space is defined on the ICBM 152 6th generation non-linear brain

---

[1]https://github.com/BIC-MNI/minc-toolkit-v2

atlas (Mazziotta et al., 2001), but users can provide their own stereotaxic template if desired. Coregistration is performed using an iterative implementation of minctracc (Collins et al., 1994). Brain tissue extraction is performed in stereotaxic space using BEaST (Eskildsen et al., 2012). In addition, tissue segmentation can also be performed on the normalized T1 MR image. Currently, only ANTs Atropos package (Avants et al., 2011) has been implemented for T1 tissue segmentation but this can be extended based on user needs.

Regions of Interest
Regions of interest have an important role in three of the processing steps in APPIAN: PVC, quantification, and reporting of results. ROIs are used in PVC algorithms to define anatomical constraints. When no arterial input is available for quantification, a reference ROI is placed in a brain region devoid of specific tracer binding. Finally, when reporting results from APPIAN, ROIs are needed to define the brain areas from which average parameters are calculated for final statistical analysis. ROIs for each of these processing steps can be defined from one of three sources. The simplest ROI are those derived from a classification of the T1 MR image, e.g., using ANIMAL (Mazziotta et al., 2001), prior to using APPIAN. Users can also use tissue classification software implemented in APPIAN to classify their T1 MR images, thereby eliminating the need to run a strictly MR image-based pipeline prior to using APPIAN.

Regions of interest can also be defined on a stereotaxic atlas, e.g., AAL (Tzourio-Mazoyer et al., 2002), with a corresponding template image. In this case, the template image is non-linearly coregistered to the T1 MR image in native space, and subsequently aligned to the native PET space of the subject. Finally, it is frequently necessary to manually define ROI on each individual MR image, for instance when segmenting focal brain pathologies such as a tumor or ischemic infarct. This option is also implemented in APPIAN.

Partial-Volume Correction
In PET, partial-volume effects result from the presence of multiple tissue types within a single voxel and the blurring of the true radiotracer concentrations. PVC of PET images is thus necessary to accurately recover the true radiotracer distribution and, for example, differentiate between true neuronal loss from cortical thinning. Several methods have been proposed to perform PVC, many of which are implemented in PETPVC (Thomas et al., 2016). In addition, we have also implemented idSURF (Funck et al., 2014), a voxel-wise iterative deconvolution that uses anatomically constrained smoothing to control for noise amplification while limiting the amount of spill-over between distinct anatomical regions. APPIAN thus allows the user to select the appropriate PVC method based on their needs and their data. If the desired PVC method is not implemented in APPIAN, it can be easily included in the pipeline by creating a file describing the inputs and outputs of the method.

Quantification
In PET images, quantitative biological or physiological parameters—such as non-displaceable binding potential or

cerebral blood flow—are often calculated from the measured temporal change of tissue radiotracer concentration, so-called time activity curves (TACs), within voxels or ROIs. Many models exist for performing quantification depending on the type of radiotracer, parameter of interest, and time frames acquired. The quantification methods available in APPIAN are from the Turku PET Centre tools (Oikonen, 2017). Currently, the implemented models are: the Logan Plot (Logan et al., 1990), Patlak–Gjedde Plot (Gjedde, 1982; Patlak et al., 1983), Simplified Reference Tissue Model (Gunn et al., 1997), and standardized uptake value (Sokoloff et al., 1977). APPIAN implements both voxel-based and ROI-based quantification methods. It can also process arterial input functions as well as input functions from reference regions devoid of specific binding. Arterial inputs are in the ".dft" format described by the Turku PET Centre[2].

Results Report

The ROI defined in "MR Image Processing" section are used to calculate regional mean values for the parameter of interest from the output images after coregistration, PVC and quantification processing steps. Additionally, if cortical surface meshes are provided by the user, the output images can be interpolated on these meshes and be used to derive surface-based parameter estimates. Regional mean parameter values are saved in wide format '.csv' files in the so-called 'vertical format' (i.e., the output measure from each subject and each region is saved in a single column). This standardized data format simplifies subsequent analysis with statistical software, such as R (R Core Team, 2016) or scikit-learn (Pedregosa et al., 2001).

APPIAN also calculates group-level descriptive statistics obtained from the output images. The group-level statistics that are provided exploit the BIDS naming convention which requires that file names include the subject ID, the task or condition, and the scanning session. APPIAN thus provides users with summary statistics for the subjects, tasks, and sessions. Descriptive statistics are plotted and displayed in a web browser-based GUI to allow simple and easy visualization of the results.

Quality Control and Visualization

APPIAN includes both visual and automated quality control. Visual quality control is facilitated by the incorporation of BrainBrowser–a 3D/4D brain volume viewer (Sherif et al., 2015)– in the web browser-based GUI (**Figure 2**). This makes it possible to visualize the output images of the coregistration, PVC and quantification processing stages without the need for additional software.

While visual inspection remains the gold-standard method for verifying the accuracy of PET coregistration (Ge et al., 1994; Andersson et al., 1995; Alpert et al., 1996; Mutic et al., 2001; DeLorenzo et al., 2009), automated QC can be useful in guiding the user to potentially failed processing steps. The first stage of the automated QC is to define a QC metric that quantifies the performance of a given processing step. For example, in the case of PET-MRI coregistration the relevant QC metric is the

similarity metric that quantifies the joint-dependence of spatial signal intensity distribution of the PET and MR images. By itself a single metric is insufficient to determine whether the processing step has been performed correctly. However, by calculating the distribution of several QC metrics for all subjects, it is possible to identify potential anomalies. Kernel density estimation is used to calculate the probability of observing a given QC metric under the empirical distribution of the entire set of QC metrics. The results are displayed in an interactive plot in the web browser-based dashboard (**Figure 3**).

File Formats

Input files for APPIAN are organized following the Brain Imaging Data Structure (BIDS) specifications (Gorgolewski et al., 2016), which uses the Nifti format. In addition, APPIAN also supports input files in the MINC file format (Vincent et al., 2016), which are also organized according to the BIDS specifications but with the MINC file extension.

High Performance Computing

APPIAN is optimized for high performance computing in two ways. APPIAN is distributed in a Docker container[3] that contains all the software necessary to run APPIAN on any computing platform supporting such containers (i.e., where Docker or Singularity has been installed). APPIAN can therefore be run identically across a wide variety of computing environments. This not only facilitates the reproducibility of results, but also allows APPIAN to be deployed simultaneously across multiple computing nodes to analyze subjects in parallel. Additionally, APPIAN supports multithread processing via NiPype and can therefore be run in parallel on multiple CPUs on a given computing platform, e.g., a personal workstation or a processing node on a server.

APPIAN also follows the specification of the BIDS apps in being capable of running subject-level and group-level analyses independently. This means that an instance of APPIAN can be run for each subject in parallel across the available computing resources. Once the individual processing steps have been completed and stored in the same location, the group-level analyses can then be run, e.g., automated QC and reporting of group-level descriptive statistics. Thus, a given data set can be processed with APPIAN at different times and on different computing platforms.

The ability to process large data sets in an easy, fast, and reproducible manner is essential, particularly in cases where parameters for a given algorithm need to be optimized or where the performance of different algorithms at a given processing stage is being compared.

Accuracy of APPIAN

The accuracy of the APPIAN pipeline was evaluated using the SORTEO Monte-Carlo simulated PET data set (Reilhac et al., 2005). These data consist of 15 subjects with a real T1 MR image segmented into anatomical defined ROIs derived from these images. From each of these anatomically segmented images,

---

[2]http://www.turkupetcentre.net/petanalysis/format_tpc_dft.html
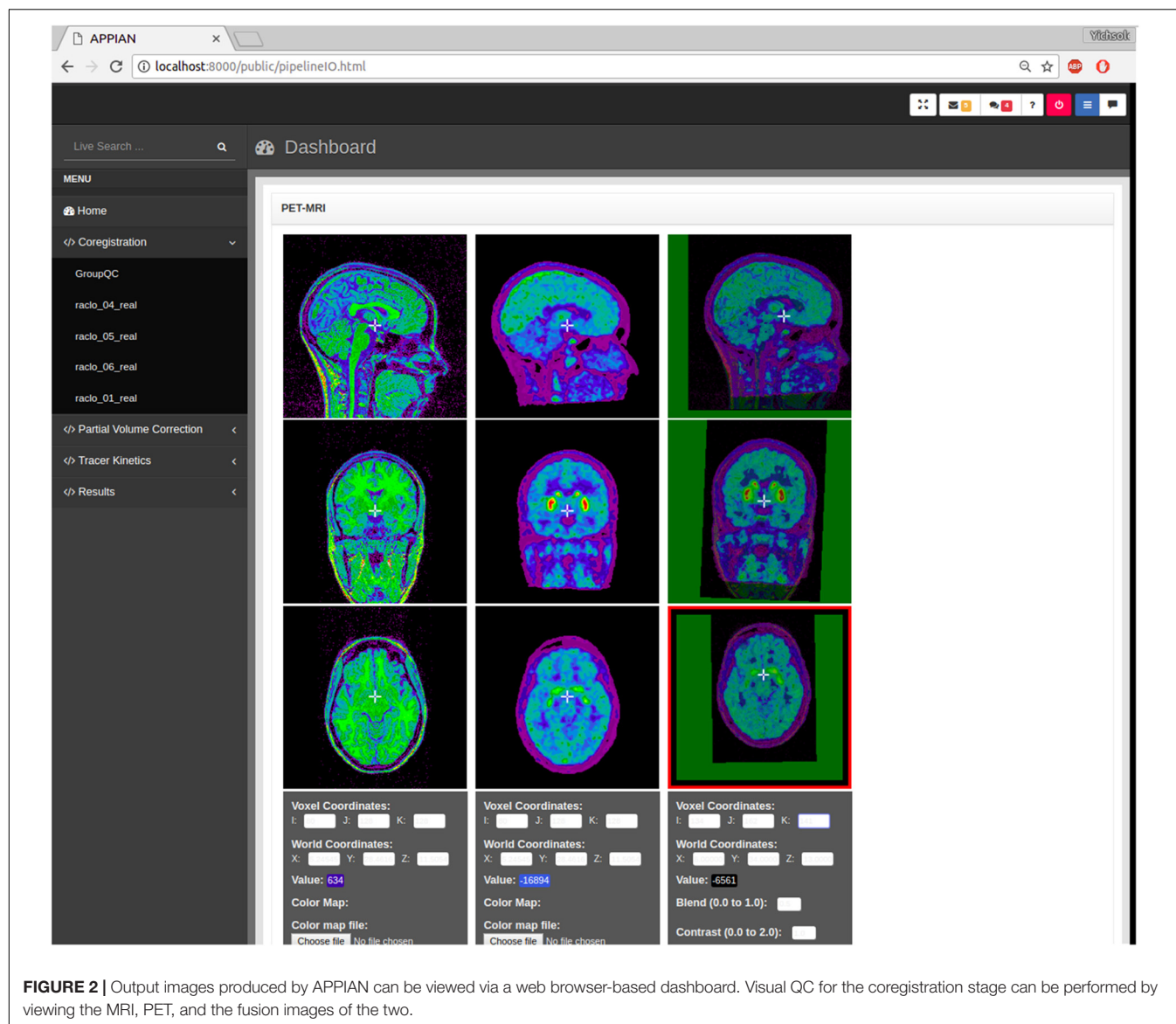
[3]https://www.docker.com/

FIGURE 2 | Output images produced by APPIAN can be viewed via a web browser-based dashboard. Visual QC for the coregistration stage can be performed by viewing the MRI, PET, and the fusion images of the two.

three sets of simulated PET images were produced by assigning empirically derived TACs of radiotracer concentrations of [11-C]-raclopride (RCL), [18-F]-fluorodeoxyglucose (FDG), and [18-F]-fluorodopa (FDOPA) into each segmented ROI. The PET images were simulated using the SORTEO Monte-Carlo PET simulator for the Siemens ECAT HR+ scanner (Adam et al., 1997).

Magnetic resonance images were processed using CIVET. CIVET uses the non-parametric N3 method to correct MR field non-uniformity (Sled et al., 1998). The MR image is then transformed to MNI stereotaxic space of the ICBM 152 6th generation non-linear brain atlas (Mazziotta et al., 2001), using a 12 parameter affine transformation (Collins et al., 1994). Spatially normalized images are then segmented into gross anatomical regions with ANIMAL (Collins and Evans, 1997). Thus all ROI images used in the subsequent analysis were derived using CIVET prior to running APPIAN.

The accuracy of the APPIAN was verified by comparing the results of the three central processing stages (coregistration, PVC, quantification) to the true radiotracer concentration TACs or the parametric values derived from them. For the coregistration and PVC stages, the integral of the TAC recovered from the processed images was compared to the integral of the true radiotracer concentration TACs. Parameter values were obtained by calculating the Ki, BPnd, and SUVR for the FDOPA, RCL, and FDG images, respectively, and compared to the same values calculated from the true radiotracer concentration TACs.

The accuracy for each processing stage was calculated by dividing the results from APPIAN by the true radiotracer concentration or parametric values. This calculation was performed for a specific ROI for each radiotracer: cortical GM for FDG, the putamen for FDOPA, and the caudate nucleus for RCL. PVC was performed using the GTM method with a point spread function of 6.5 mm full-width
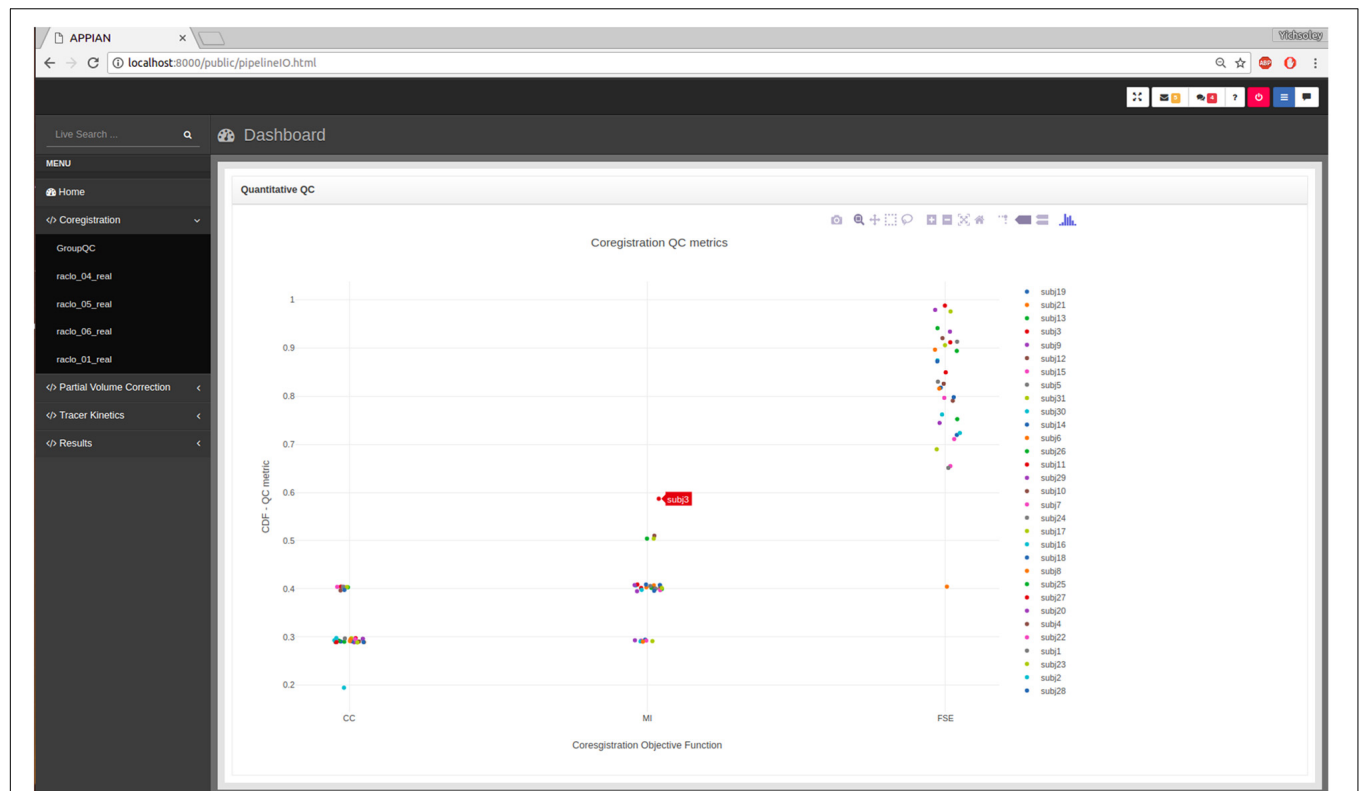
FIGURE 3 | Output from automated quality control (QC) allows users to assess the performance of major processing steps at a glance. Here the automated QC metrics for the coregistration processing stage are shown: CC, cross-correlation; MI, mutual information; FSE, feature-space entropy.

TABLE 1 | Accuracy is measured as the ratio of recovered to true radiotracer concentration or parameter value. APPIAN accurately recovers radiotracer concentrations and tracer kinetic parameters from the SORTEO simulated PET images.

Radiotracer	ROI	PVE	Analysis	Metric	Accuracy
FDG	GM	Uncorrected	Coregistration	integral	0.66 ± 0.006
FDG	GM	Corrected	PVC	integral	0.93 ± 0.025
FDG	GM	Corrected	Quantification	SUVR	0.94 ± 0.048
FDOPA	Putamen	Uncorrected	Coregistration	integral	0.69 ± 0.03
FDOPA	Putamen	Corrected	PVC	integral	1 ± 0.055
FDOPA	Putamen	Corrected	Quantification	Ki	0.83 ± 0.238
RCL	Caudate Nucleus	Uncorrected	Coregistration	integral	0.77 ± 0.016
RCL	Caudate Nucleus	Corrected	PVC	integral	1.05 ± 0.035
RCL	Caudate Nucleus	Corrected	Quantification	BPnd	1.03 ± 0.042

half-maximum (Rousset et al., 1998). The cerebellum was used as a reference region for the calculation of parametric values in the quantification stage.

RESULTS

APPIAN was able to recover accurate values at each major processing stage (**Table 1**), see **Figure 4** for illustrative example from one subject. The recovered values for the coregistration and PVC were the integral of the regional TACs. For the quantification stage the recovered values were the parametric

values as described in section "Accuracy of APPIAN". The accuracy of the coregistration stage was between 0.66 and 0.77, which represented an underestimation of the radiotracer distribution due to partial-volume effects. The accuracy was significantly improved by PVC, ranging between 0.93 and 1.05. The effect of PVC on the uncorrected radioactivity concentration for each radiotracer is shown in **Figure 5**. The PVC led to a slight overestimation in the caudate nucleus with RCL, but near perfect accuracy in the putamen with FDOPA. The final output parametric values were very accurate for RCL (1.02) and FDG (0.94), and lower in the case of FDOPA (0.83).
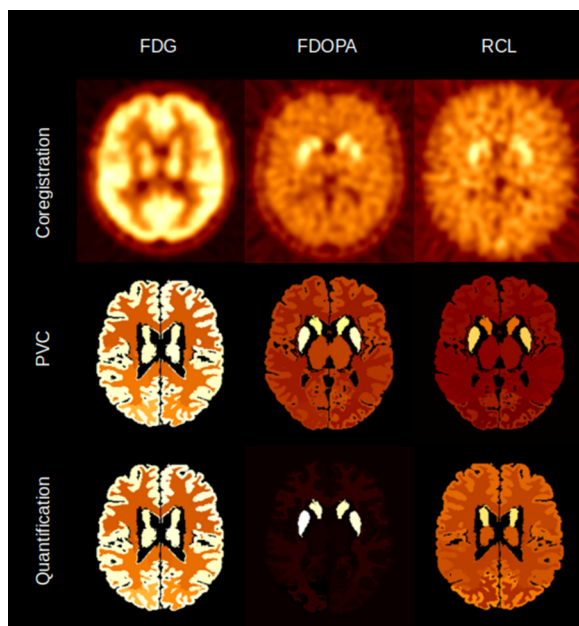
FIGURE 4 | Illustrative example of the image volumes produced by APPIAN for the three major processing stages for FDG, FDOPA, and RCL.

DISCUSSION

Accuracy of APPIAN

APPIAN recovered accurate values for each of the three major processing steps on the SORTEO simulated PET data set. Not surprisingly, the accuracy of the recovered parameters was initially low (0.65–0.77), because of partial-volume effects. This improved significantly after PVC with the GTM method (0.93–1.05). For RCL and FDG, the parametric values resulting from the quantification processing stage maintained a similar level of accuracy to that of the PVC radiotracer concentrations. This was not the case with FDOPA where the accuracy decreased from 1 to 0.83. The decrease in accuracy was

due to noise in the radiotracer concentrations that were measured in the caudate nucleus, which led to errors in the calculation of the integrals used by the Patlak plot to determine Ki.

For each radiotracer, the validation of APPIAN's accuracy was performed with differing ROI and using different methods for calculating parametric values. These differences mean that it is not possible to quantitatively compare APPIAN's accuracy for each radiotracer. The choice of ROI and algorithms for deriving parametric values were chosen to reflect analysis procedures that are widely used by researchers for each of the three radiotracers. It should be noted that the cerebellum is not typically used as a reference region for calculating SUVR or Ki for FDG and FDOPA, respectively. However, while the specific location of the reference region is of utmost importance when performing true PET quantification, it is not relevant for verifying the computational accuracy of the algorithms in the APPIAN pipeline.

Comparison to Existing Pipelines

Several PET processing pipelines have been presented in recent years. We here briefly describe them to highlight their relative strengths (**Table 2**) and discuss how APPIAN compares to these. There are other PET pipelines that carry out at least three of the six steps performed by APPIAN, they are: PMOD (Mikolajczyk et al., 1998), CapAIBL (Bourgeat et al., 2015), MIAKAT (Gunn et al., 2016), Pypes (Savio et al., 2017), and NiftyPET (Markiewicz et al., 2017).

PMOD

PMOD (Mikolajczyk et al., 1998) is the gold-standard software for quantification of PET images and is distributed in modules that perform specific aspects of PET analysis. PKIN includes an exhaustive list of quantification models and preprocessing methods for blood and plasma activity curves for analyzing regional PET data, while PXMOD performs the same analyses at the pixel level. PMOD also has modules that perform analysis and PVC (PBAS), and image registration (PFUS). All these modules can be used interactively using a graphical user
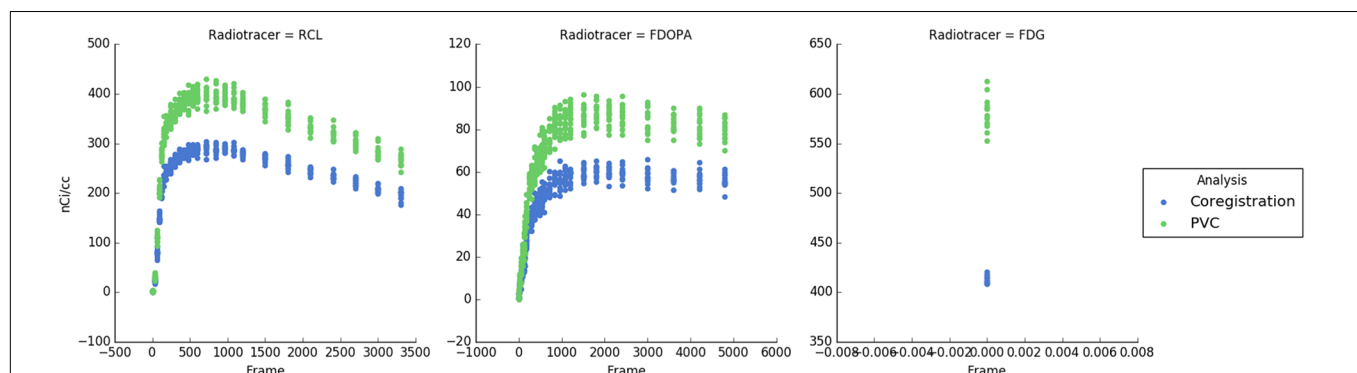

FIGURE 5 | Time-activity curves for each subject and each radiotracer. Blue points indicate the uncorrected PET radioactivity concentration after PET-MRI coregistration and green points show radioactivity concentration after PVC with the GTM method. PVC corrects for spill-over of radiotracer distribution and increases the measured radioactivity concentration.

TABLE 2 | Many different PET processing software exist with various features.

Feature	MIAKAT	PMOD	Pypes	CapAIBL	NiftyPET	APPIAN
Cost	Free	2,970–14,850$	Free	Free	Free	Free
Open-source	Yes	No	Yes	No	Yes	Yes
Language	MATLAB	Java	Agnostic*	C++	Python	Agnostic*
Quantification	Yes	Yes	No	SUVR	No	Yes
PVC	No	No	Yes	No	Yes	Yes
Structural imaging	Yes	Optional	Yes	No	Yes	Required
Cloud-based processing	No	DICOM server	No	Yes	Maybe	Yes
Local processing	Yes	Yes	Yes	No	Yes	Yes
Visualization	GUI	GUI	Result plots	3D surfaces	No	Dashboard
Surface-based	No	No	No	Yes	No	Yes
Reconstruction	No	No	No	No	Yes	No

*APPIAN attempts to provide all post-reconstruction tools needed for PET research. *Agnostic: these packages are written in Python but support software written in any language as long as it can run on the command line. Here, some of the most established and more recent pipelines are compared to APPIAN.*

interface (GUI) but can also be linked together in a pipeline to automate the analysis of large data sets. A particularly useful feature is the option to add a QC step after each processing stage. PMOD thus includes all the preprocessing and analysis methods needed for automated PET analysis. As a commercial software solution however, the PMOD code is not open-source and thus imposes limitations on the user community with respect to flexible development and implementation of new image processing and analytical methods.

CapAIBL

CapAIBL (Bourgeat et al., 2015) is a surface-based PET processing pipeline that is available through an online platform. It spatially normalizes PET images to cortical surface templates for the surface-based analysis and visualization of PET data without the need for structural imaging. Cortical surfaces are derived from a standardized template, thus subcortical structures such as the basal ganglia are not included in the analysis. A purely surface-based approach is also limited to images from structurally intact brains and may thus be difficult to apply to datasets with focal brain lesions. Nonetheless, CapAIBL provides a highly original method for performing automated PET analysis that is useful for the study of the cerebral cortex in cases where no structural image has been acquired alongside the PET image. Dore et al. (2016) have shown a close correspondence in PET quantification across a wide range of radiotracers with coregistered PET and MR images and using CapAIBL, i.e., without coregistration.

Pypes

A recent multi-modal pipeline, Pypes (Savio et al., 2017), combines PET analysis with structural, diffusion, and functional MR images. This pipeline is free, open-source, and it is also written using NiPype (Gorgolewski et al., 2011). Pypes leverages several brain imaging software packages–including SPM12 (Ashburner, 2012), FSL (Jenkinson et al., 2012), and AFNI (Cox, 2012)–to provide multi-modal workflows. While Pypes does incorporate PVC, it does not incorporate tracer kinetic analysis, flexible ROI definition, or automated QC.

MIAKAT

MIAKAT (Mikolajczyk et al., 1998) is the most complete, open-source PET processing pipeline. In addition to featuring many tracer-kinetic models, MIAKAT also includes motion-correction; a feature that is not currently implemented in APPIAN. One of MIAKAT's most important features is its user-friendly GUI. This makes MIAKAT easy to use for users not familiar with the command-line interface. In addition to analyzing PET images, MIAKAT also includes the option to include structural images which are used to define regions of interest (ROI). MIAKAT has been recently extended for use on non-brain PET image analysis and for application to species other than humans (Searle and Gunn, 2017).

One limitation of MIAKAT is that it does not include PVC, although this could potentially be added to the pipeline. More importantly, it is built using MATLAB, which restricts MIAKAT to a single, proprietary language with licensing restrictions.

NiftyPET

NiftyPET is another open-source, Python-based PET processing pipeline that implements Graphical Processing Unit-processing for massively parallel processing (Markiewicz et al., 2017). It is the only PET processing pipeline to reconstruct PET images from sinograms and to perform PVC (Yang et al., 1995). It should be noted that the authors of NiftyPET use the term "quantification" to refer to quantification of radioactivity concentrations, whereas this term is here used to refer to the quantification of underlying biological or physiological parameters. NiftyPET therefore does not include parametric quantification.

APPIAN

There are a wide variety of PET pipelines presently available, each satisfying a different niche. APPIAN provides a highly flexible framework for processing large PET data sets, see **Figure 6** for a detailed flowchart of APPIAN. One important feature is that APPIAN allows the user to define ROI from a variety of sources and is therefore compatible with a wide variety of experimental designs. Whereas lesion studies frequently use a binary lesion image defined on each subject's respective structural

FIGURE 6 | Flowchart of the modules implemented in APPIAN. Green boxes indicate mandatory inputs, blue boxes indicate optional inputs, and tan boxes indicate the primary quantitative outputs of the pipeline.
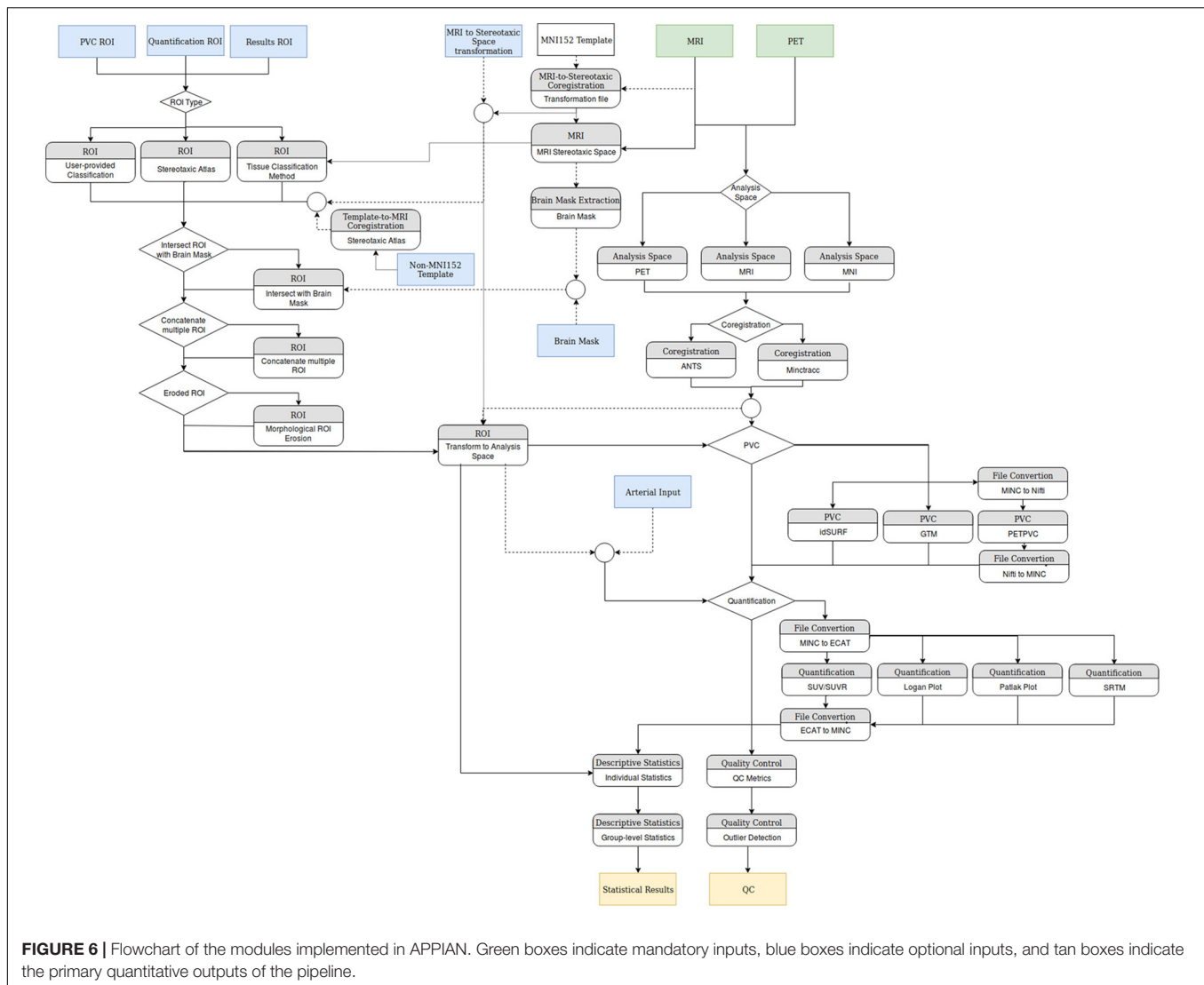
image in its native coordinate space, it may be necessary for some studies (e.g., investigating lesion effects on functional systems as in aphasia post stroke) to use a common brain atlas in MNI-space. On the other hand, PET studies of, e.g., microglial inflammation may identify ROI based on the subjects' respective tracer binding pattern in PET images in their native space. Quantification of PET images also requires users to be able to use either ROI to define a reference region without specific binding of the radiotracer or TAC measured from arterial blood samples. APPIAN is therefore suited for a wide variety of experimental contexts because of its flexible system for ROI definition.

APPIAN is also modular and easily extendable so that users can either test new algorithms, e.g., a new PVC method, or add entirely new analyses to the pipeline. Moreover APPIAN, like Pypes, is written with NiPype and can thus use any program that can be run in a Bash shell environment. Users therefore do not need to rewrite their software in, e.g., Python if they wish to implement it in APPIAN. Also, given that

descriptive statistics for ROI are automatically generated in the reporting stage, it is easy to extend APPIAN to perform sophisticated group-wise analyses. For example, investigators interested in implementing graph theoretical analyses can append their analysis to the group-level processing and input the descriptive statistics that are collected at the reports stage to their analysis.

Finally, APPIAN implements automated and visual QC to facilitate the analysis of large data sets. This is essential because as multiple processing stages are linked together into increasingly sophisticated pipelines, it is important that users be able to easily and reliably confirm that each processing stage has been performed correctly.

Using APPIAN

APPIAN is available for both local use and cloud-based use. The source code for APPIAN is freely available[4]. While the code-base

---

[4]www.github.com/APPIAN-PET/APPIAN

will be maintained by the authors, we hope to create a community of developers to support the project in the long-term. Changes to APPIAN will be validated against the open CIMBI PET data[5] (Knudsen et al., 2016). APPIAN is provided via a Docker (see footnote 3) image and can be easily downloaded from Docker hub under tffunck/appian:latest. Cloud-based APPIAN is available via the CBRAIN platform[6].

CONCLUSION

APPIAN is a novel PET processing pipeline that seeks to automate the processing of reconstructed PET images for a wide variety of experimental designs. It is therefore flexible and easily extendable. In order to ensure that each processing step is performed as expected, visual and automated QC are implemented. Our results on Monte-Carlo simulated PET data have shown that APPIAN accurately recovers radiotracer concentration and parametric values. Future work will focus on

---

[5] https://openneuro.org/datasets/ds001421
[6] portal.cbrain.mcgill.ca

increasing the sensitivity of the automated QC and implementing more algorithms for coregistration, PVC, and quantification.

AUTHOR CONTRIBUTIONS

TF is the primary author of the manuscript, developed the APPIAN code. KL developed the APPIAN code. P-JT is the advisor for designing APPIAN, edited the manuscript, and ongoing development of new PET quantification models. AT (principal investigator) and AE (co-principal investigator) provided conceptual guidance and edited the manuscript.

FUNDING

REFERENCES

Adam, L. E., Zaers, J., and Ostertag, H. (1997). Performance evaluation of the whole-body PET scanner ECAT EXACT HR/sup+ following the IEC standard. *IEEE Trans. Nucl. Sci.* 44, 299–302. doi: 10.1109/23.596983

Alpert, N. M., Berdichevsky, D., Levin, Z., Morris, E. D., and Fischman, A. J. (1996). Improved methods for image registration. *Neuroimage* 3, 10–18. doi: 10.1006/NIMG.1996.0002

Andersson, J. L., Sundin, A., and Valind, S. (1995). A method for coregistration of PET and MR brain images. *J. Nucl. Med.* 36, 1307–1315.

Ashburner, J. (2012). SPM: a history. *Neuroimage* 62, 791–800. doi: 10.1016/j.neuroimage.2011.10.025

Avants, B. B., Tustison, N. J., Wu, J., Cook, P. A., and Gee, J. C. (2011). An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics* 9, 381–400. doi: 10.1007/s12021-011-9109-y

Bourgeat, P., Dore, V., Fripp, J., Villemagne, V. L., Rowe, C. C., and Salvado, O. (2015). "Computational analysis of PET by AIBL (CapAIBL): a cloud-based processing pipeline for the quantification of PET images", in *Proceedings of the International Society for Optics and Photonics,* eds S. Ourselin and M. A. Styner (Washington DC: SPIE). doi: 10.1117/12.2082492

Collins, D. L., and Evans, A. C. (1997). Animal: validation and applications of nonlinear registration-based segmentation. *Int. J. Pattern Recognit. Artif. Intell.* 11, 1271–1294. doi: 10.1142/S0218001497000597

Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3D inter-subject registration of MR volumetric data in standardized talairach space. *J. Comput. Assist. Tomogr.* 10, 192–205.

Cox, R. W. (2012). AFNI: what a long strange trip its been. *Neuroimage* 62, 743–747. doi: 10.1016/j.neuroimage.2011.08.056

DeLorenzo, C., Klein, A., Mikhno, A., Gray, N., Zanderigo, F., Mann, J. J., et al. (2009). "A new method for assessing PET-MRI coregistration," in *Proceedings of SPIE International Society for Optics and Photonics*, eds J. P. W. Pluim, B. M. Dawant (Washington DC: SPIE) doi: 10.1117/12.812170

Dore, V., Bourgeat, P., Villemagne, V. L., Fripp, J., Macaulay, L., Masters, C. L., et al. (2016). "CapAIBL: automated reporting of cortical PET quantification without need of MRI on brain surface using a patch-based method," in *Proceedings of the: Patch-Based Techniques in Medical Imaging*, eds G. Wu, P. Coupé, Y. Zhan, B. Munsell and D. Rueckert (Cham: Springer International Publishing), 109–116. doi: 10.1007/978-3-319-47118-1_14

Eskildsen, S. F., Coupé, P., Fonov, V., Manjón, J. V., Leung, K. K., Guizard, N., et al. (2012). BEaST: brain extraction based on nonlocal segmentation technique. *Neuroimage* 59, 2362–2373. doi: 10.1016/j.neuroimage.2011.09.012

Funck, T., Paquette, C., Evans, A., and Thiel, A. (2014). Surface-based partial-volume correction for high-resolution PET. *Neuroimage* 102(Pt 2), 674–87. doi: 10.1016/j.neuroimage.2014.08.037

Ge, Y., Fitzpatrick, J. M., Votaw, J. R., Gadamsetty, S., Maciunas, R. J., Kessler, R. M., et al. (1994). Retrospective registration of pet and mr brain images: an algorithm and its stereotactic validation. *J. Comput. Assist. Tomogr.* 18, 800–810. doi: 10.1097/00004728-199409000-00021

Gjedde, A. (1982). Calculation of cerebral glucose phosphorylation from brain uptake of glucose analogs in vivo: a re-examination. *Brain Res.* 257, 237–274. doi: 10.1016/0165-0173(82)90018-2

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., et al. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* 3:160044. doi: 10.1038/sdata.2016.44

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front. Neuroinform.* 5:13. doi: 10.3389/fninf.2011.00013

Gunn, R. N., Coello, C., and Searle, G. E. (2016). Molecular imaging and kinetic analysis toolbox (MIAKAT) - a quantitative software package for the analysis of PET neuroimaging data. *J. Nucl. Med.* 57:1928.

Gunn, R. N., Lammertsma, A. A., Hume, S. P., and Cunningham, V. J. (1997). Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. *Neuroimage* 6, 279–287. doi: 10.1006/nimg.1997.0303

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Knudsen, G. M., Jensen, P. S., Erritzoe, D., Baaré W. F. C., Ettrup, A., Fisher P. M., et al. (2016). The center for integrated molecular brain imaging (Cimbi) database. *Neuroimage* 124(Pt B), 1213–1219. doi: 10.1016/j.neuroimage.2015.04.025

Logan, J., Fowler, J. S., Volkow, N. D., Wolf, A. P., Dewey, S. L., Schlyer, D. J., et al. (1990). Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-11C-methyl]-(-)-cocaine PET studies in human subjects. *J. Cereb. Blood Flow Metab.* 10, 740–747. doi: 10.1038/jcbfm.1990.127

Markiewicz, P. J., Ehrhardt, M. J., Erlandsson, K., Noonan, P. J., Barnes, A., Schott, J. M., et al. (2017). NiftyPET: a high-throughput software platform for high quantitative accuracy and precision PET imaging and analysis. *Neuroinformatics* 16, 95–115. doi: 10.1007/s12021-017-9352-y

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., et al. (2001). A probabilistic atlas and reference system for the human brain: international

consortium for brain mapping (ICBM). *Philos. Trans. R. Soc. Biol. Sci.* 356, 1293–1322. doi: 10.1098/rstb.2001.0915

Mikolajczyk, K., Szabatin, M., Rudnicki, P., Grodzki, M., and Burger, C. (1998). A JAVA environment for medical image data analysis: initial application for brain PET quantitation. *Med. Inform.* 23, 207–214. doi: 10.3109/14639239809001400

Mutic, S., Dempsey, J. F., Bosch, W. R., Low, D. A., Drzymala, R. E., Chao, K. S. C., et al. (2001). Multimodality image registration quality assurance for conformal three-dimensional treatment planning. *Int. J. Radiat. Oncol. Biol. Phys.* 51, 255–260. doi: 10.1016/S0360-3016(01)01659-5

Oikonen, V. (2017). *Software for Positron Emission Tomography.* Available at: https://gitlab.utu.fi/vesoik/tpcclib/tags

Patlak, C. S., Blasberg, R. G., and Fenstermacher, J. D. (1983). Graphical evaluation of blood-to-brain transfer constants from multiple-time uptake data. *J. Cereb. Blood Flow Metab.* 3, 1–7. doi: 10.1038/jcbfm.1983.1

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2001). Scikit-learn: machine learning in python. *J. Machine Learn. Res.* 12, 2825–2830.

R Core Team (2016). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Reilhac, A., Batan, G., Michel, C., Grova, C., Tohka, J., Louis Collins, D., et al. (2005). PET-SORTEO: validation and development of database of simulated PET volumes. *IEEE Trans. Nucl. Sci.* 52, 1321–1328. doi: 10.1109/TNS.2005.858242

Rousset, O. G., Ma, Y., and Evans, A.C. (1998). Correction for partial volume effects in PET: principle and validation. *J. Nucl. Med.* 39, 904–911.

Savio, A. M., Schutte, M., Graña, M., and Yakushev, I. (2017). Pypes: workflows for processing multimodal neuroimaging data. *Front. Neuroinform.* 11:25. doi: 10.3389/fninf.2017.00025

Searle, G., and Gunn, R. (2017). Extension of the MIAKAT analysis software package to non-brain and pre-clinical PET analysis. *J. Nucl. Med.* 58(Suppl. 1):1305.

Sherif, T., Kassis, N., Rousseau, M. Ã., Adalat, R., and Evans, A. C. (2015). BrainBrowser: distributed, web-based neurological data visualization. *Front. Neuroinform.* 8:89. doi: 10.3389/fninf.2014.0008920

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97. doi: 10.1109/42.668698

Sokoloff, L., Reivich, M., Kennedy, C., Des, R. M. H., Patlak, C. S., Pettigrew, K. D., et al. (1977). The [14C] deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *J. Neurochem.* 28, 897–916. doi: 10.1111/j.1471-4159.1977.tb10649.x

Thomas, B. A., Cuplov, V., Bousse, A., Mendes, A., Thielemans, K., Hutton, B. F., et al. (2016). PETPVC: a toolbox for performing partial volume correction techniques in positron emission tomography. *Phys. Med. Biol.* 61, 7975–7993. doi: 10.1088/0031-9155/61/22/7975

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. 15, 273–289. doi: 10.1006/nimg.2001.0978

Vincent, R. D., Neelin, P., Khalili-Mahani, N., Janke, A. L., Fonov, V. S., Robbins, S. M., et al. (2016). MINC 2.0: a flexible format for multi-modal images. *Front. Neuroinform.* 10:35. doi: 10.3389/fninf.2016.00035

Yang, J., Mega, M., Huang, S. C., Lin, K. P., Toga, A., Small, G., et al. (1995). "Investigation of partial volume correction methods for brain FDG-PET studies [MRI application]," in *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference Record* (Francisco, CA: IEEE). doi: 10.1109/nssmic.1995.501907

Pipeline for Analyzing Lesions After Stroke (PALS)

Kaori L. Ito[1], Amit Kumar[1], Artemis Zavaliangos-Petropulu[1,2], Steven C. Cramer[3] and Sook-Lei Liew[1,2]*

[1] Neural Plasticity and Neurorehabilitation Laboratory, University of Southern California, Los Angeles, CA, United States, [2] Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, United States, [3] Department of Neurology, University of California, Irvine, Irvine, CA, United States

Lesion analyses are critical for drawing insights about stroke injury and recovery, and their importance is underscored by growing efforts to collect and combine stroke neuroimaging data across research sites. However, while there are numerous processing pipelines for neuroimaging data in general, few can be smoothly applied to stroke data due to complications analyzing the lesioned region. As researchers often use their own tools or manual methods for stroke MRI analysis, this could lead to greater errors and difficulty replicating findings over time and across sites. Rigorous analysis protocols and quality control pipelines are thus urgently needed for stroke neuroimaging. To this end, we created the Pipeline for Analyzing Lesions after Stroke (PALS; DOI: https://doi.org/10.5281/zenodo.1266980), a scalable and user-friendly toolbox to facilitate and ensure quality in stroke research specifically using T1-weighted MRIs. The PALS toolbox offers four modules integrated into a single pipeline, including (1) reorientation to radiological convention, (2) lesion correction for healthy white matter voxels, (3) lesion load calculation, and (4) visual quality control. In the present paper, we discuss each module and provide validation and example cases of our toolbox using multi-site data. Importantly, we also show that lesion correction with PALS significantly improves similarity between manual lesion segmentations by different tracers ($z = 3.43$, $p = 0.0018$). PALS can be found online at https://github.com/npnl/PALS. Future work will expand the PALS capabilities to include multimodal stroke imaging. We hope PALS will be a useful tool for the stroke neuroimaging community and foster new clinical insights.

Keywords: stroke, big data, lesion analysis, lesion load, MRI imaging, neuroimaging, stroke recovery

INTRODUCTION

Characterizing the relationship between brain structure and function is an important step in identifying and targeting biomarkers of recovery after stroke (Dimyan and Cohen, 2011). As stroke is heterogeneous in both its anatomical and clinical presentation, it is often difficult to draw generalizable inferences with typical sample sizes. Moreover, many stroke research groups have traditionally operated in silos (Hachinski et al., 2010). This poses a problem for scientific reproducibility, as different research groups have various in-house analytic processes and pipelines that are often not transparent (Gorgolewski and Poldrack, 2016). In recent years, big data approaches have emerged and been embraced in the neuroimaging field (Milham, 2012). This offers new hope for discovery of otherwise difficult-to-detect neural patterns that hold promise for promoting advanced therapeutic techniques (Feldmann and Liebeskind, 2014; Huang et al., 2016). While promising in their potential to overcome the problem of heterogeneity in stroke research, big

data approaches to research come with their own challenges, especially with respect to combining data across sites, and managing and analyzing such large quantities of data (Van Horn and Toga, 2014). Particularly for the analysis of data from persons with stroke, there is a pressing need for the development of reproducible image processing and analysis pipelines that properly incorporate the lesion to promote collaborative efforts in the analysis of large stroke datasets.

The semiautomatic brain region extraction (SABRE) pipeline is one such example of an image processing pipeline for lesion analysis that has been made open-source (Dade et al., 2004). The SABRE pipeline integrates currently existing software, such as FSL and ANTs to allow for volumetric profile of regionalized tissue and lesion classes, while emphasizing quality control (Avants et al., 2009; Jenkinson et al., 2012). However, the SABRE pipeline was not specifically developed for stroke MRIs, and requires multi-modal inputs, which are not commonly available for research on chronic stroke.

To this end, we created the Pipeline for Analyzing Lesions after Stroke (PALS; DOI: https://doi.org/10.5281/zenodo.1266980), an open-source analysis pipeline with a graphical user interface (GUI) to facilitate reproducible analyses across stroke research sites using a single modality—a T1-weighted MRI, which is the most commonly available for chronic stroke research. Our goal is to improve the standardization and analysis of stroke lesions and to encourage collaboration across stroke research groups by creating a flexible, scalable, user-friendly toolbox for researchers. PALS has four modules integrated into a single analysis pipeline (**Figure 1**, bolded text): (1) reorientation of image files to the standard radiological convention, (2) lesion correction for healthy white matter, which removes voxels in the lesion mask that are within a normal intensity range of white matter, (3) lesion load calculation, which calculates the number of voxels that are overlapping between the lesion and a specified region of interest, and (4), visual quality control (QC), which creates HTML pages with screenshots of lesion segmentations and intermediary outputs to promote visual inspection of data at each analysis step. Notably, researchers should use a method of their choice to generate the initial lesion masks for their dataset before using PALS. We provide a comprehensive review of all existing automated lesion segmentation methods (Ito, Kim, and Liew, under review), and note that the gold standard is still manual lesion segmentation. However, once lesion masks are generated, whether through automated or manual methods, the PALS pipeline will facilitate quality control and additional analyses using the lesion masks.

The rationale for each step was informed by both existing literature as well as current attempts to combine stroke data collected across multiple sites (Liew et al., 2018). In this report, we will first review the rationale for each of these features, then discuss the implementation of the features, and finally present results from using the toolbox on multi-site data. The compiled toolbox, source code, and instructions can be freely accessed at our Github repository[1].

---

[1]https://github.com/npnl/PALS

MAIN FEATURES: RATIONALE

PALS features a GUI-based navigation system for ease of use (**Figure 2**). Any combinations of the four modules (reorientation to radiological convention, lesion correction, lesion load calculation, and visual quality control) can be selected and the entire pipeline will run automatically.

Reorientation to Radiological Convention

Inconsistent orientation of images within a dataset is a common and serious issue in image processing. Neurological and radiological orientations are both widely used conventions for storing image information (Brett et al., 2017). Whereas the neurological convention stores a patient's left side on the left part of the image, the radiological convention stores left side information on the right side of the image. The convention in which image information is stored can vary between scanners or even acquisition parameters, such that some images are stored in the radiological convention, and others are stored in the neurological convention. Moreover, commonly used neuroimaging processing tools display and store information in different ways, which can lead to orientation inconsistencies. For example, FSL and FSLeyes by default displays images in radiological convention (Jenkinson et al., 2012), the SPM display utility by default displays images in neurological convention (Penny et al., 2011), and MRIcron allows users to switch between the orientations (Rorden and Brett, 2000). If image labels are inconsistent or incorrect, analyses may be negatively impacted since one may be incorrectly flipping the two sides of the brain (Duff, 2015). This is particularly problematic for stroke neuroimaging research, as one may mislabel the hemisphere of the stroke lesion. As such, image orientation needs to be carefully considered especially for large collaborative efforts, when data has been collected from multiple sites. We thus built a simple, optional module to convert all image inputs to the radiological convention prior to performing any subsequent step to harmonize data across sites. We recommend use of this module with all datasets.

Lesion Correction for Healthy White Matter Voxels

While many automated approaches have been developed for lesion segmentation, manual segmentation remains the gold standard for tissue labeling and continues to be the benchmark for automated approaches (Fiez et al., 2000; Maier et al., 2017). Yet, depending on the size and location of the lesion, manual lesion segmentation could be a highly time- and labor-intensive process. This becomes particularly challenging for large, multi-site collaborative efforts, as having larger datasets places an increasing demand on skilled manual labor. As such, multiple individuals are often trained to perform lesion segmentations to distribute the heavy labor demands. However, the wide variability in lesion characteristics as well as inter-subjective differences in the way that lesions are defined may introduce potential inconsistencies in the manual lesion segmentation process (Fiez et al., 2000). Lesion correction for healthy white matter voxels
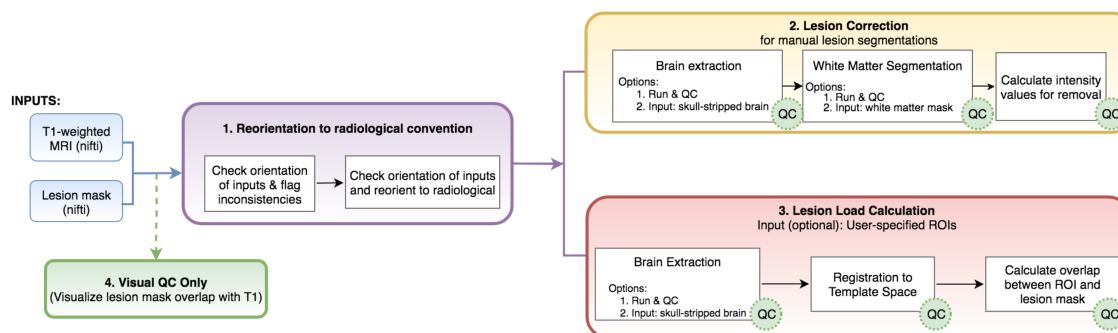
FIGURE 1 | Analysis pipeline. PALS takes in a minimum of two inputs (in blue): a T1-weighted MRI and a lesion mask file and has four main modules: (1) reorientation to radiological convention, (2) lesion correction, (3) lesion load calculation, and (4) visual QC. Users can choose to perform any or all of the main modules. White boxes indicate processing steps used in the pipeline. Green "QC" circles indicate that PALS will create a quality control page for that processing step. ROIs, regions of interest.
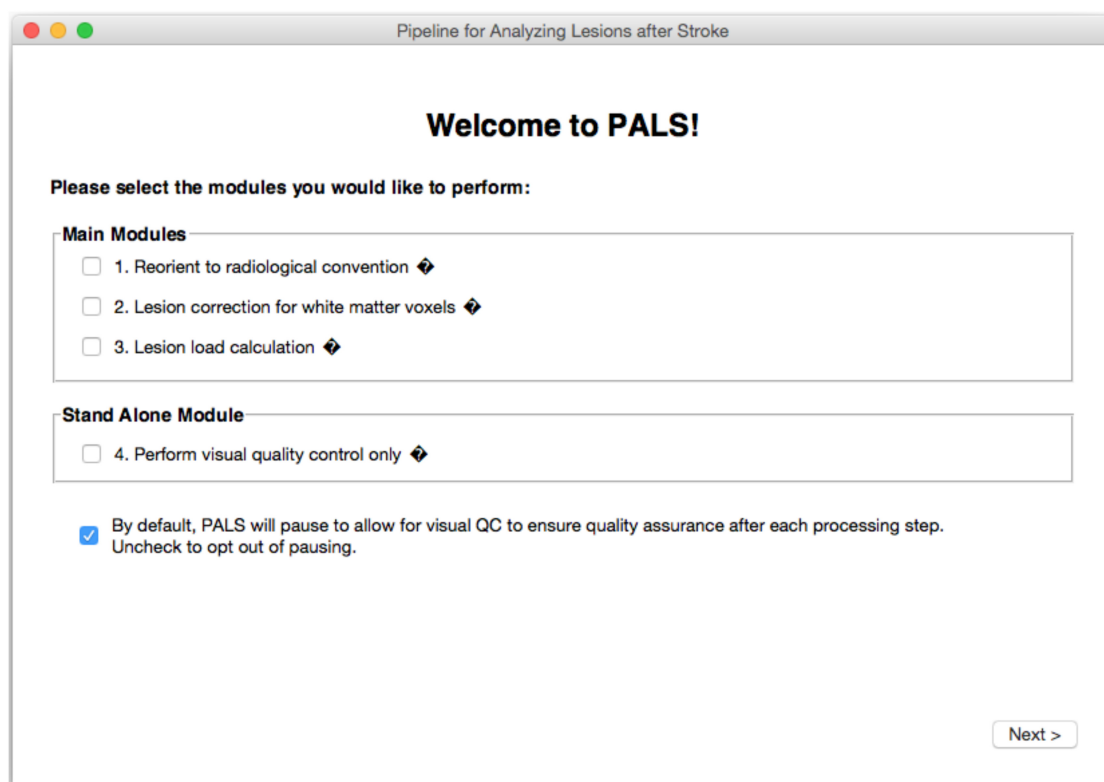

FIGURE 2 | PALS Interface. PALS has a simple and intuitive graphical user interface (GUI) which allows back and forth navigation and easy selection of modules. Additionally, the interface has tool tip icons (indicated by small black diamonds with question marks), which provide a brief helper text about each function when the user hovers over the tool tip icon.

is one method proposed to decrease subjective differences in the manual definition of lesions (Riley et al., 2011). The lesion correction aims to correct for intact white matter voxels that may have been inadvertently included in a manually segmented lesion mask. This is done by removing voxels in the lesion mask that were within the intensity range of a healthy white matter mask. We previously created a semi-automated toolbox to address this (SRQL toolbox; Ito et al., 2017). However, it

required manual delineation of a white matter mask for each subject. Here, we integrated an updated version of the SRQL toolbox as an optional lesion correction module that improves on the SRQL toolbox by taking advantage of automated white matter segmentation in FSL. We note that we recommend use of the lesion correction module only on manually segmented lesions, and not on automated segmentations, as evidenced in our validation work below. Furthermore, careful visual inspection of

white matter segmentation masks should be completed prior to using this module.

Lesion Load Calculation

Currently, one of the main goals of stroke research is to identify biomarkers for recovery, which can help identify patient subgroups and predict which treatments would be most beneficial for different patient subgroups (González, 2006; Cramer, 2010; Stinear, 2017). Studying the anatomy and precise location of stroke lesions is one potential avenue for drawing clinically meaningful inferences about recovery. Specifically, the structural integrity of white matter motor pathways, which has been measured as the overlap of the lesion with a corticospinal (CST) tract template, has been associated with motor performance (Zhu et al., 2010; Riley et al., 2011; Stinear, 2017), and it has been suggested that good recovery of motor function is largely reflective of spontaneous processes that involve the ipsilesional motor pathway (Byblow et al., 2015). In fact, it has been shown that both initial motor impairment and long-term motor outcome are dependent on the extent of CST damage, and the extent of white matter damage had greater predictive value than lesion volume (Puig et al., 2011; Feng et al., 2015). The extent of CST damage has been developed into an imaging biomarker as the weighted CST lesion load, which is calculated by overlaying lesion maps from anatomical MRIs with a canonical, atlas-based CST tract (Riley et al., 2011). Here, we built a module to calculate the CST lesion load using T1w MRIs, and validate use of our module against a similar lesion load calculator (Riley et al., 2011). However, as it is likely that other motor and non-motor regions in the brain may also be predictive of motor or cognitive recovery (Crafton et al., 2003; Rondina et al., 2017), we have extended the lesion load module to analyze lesion overlap with corticospinal tract or other cortical and subcortical structures and tracts, based on regions of interest from the FreeSurfer software and sensorimotor area tract template (S-MATT; Archer et al., 2017) packages, respectively.

Visual Quality Control

To analyze large quantities of data efficiently, most neuroimaging processing steps are now automated. Yet the presence of a stroke lesion substantially increases the susceptibility to image preprocessing errors (Andersen et al., 2010; Siegel et al., 2017). The accuracy of each image processing step, including but not limited to lesion segmentation, brain extraction, and normalization, could impact subsequent downstream processing and analyses. Therefore, visual inspection of automated output is imperative for lesion analyses. To this end, we encourage visual inspection of data for quality data assurance by integrating the creation of quality control review pages for each preprocessing step that PALS requires. PALS is designed to pause after each intermediary step and ask the user to inspect the data and provide manual input on whether each subject's output passes visual inspection (which can be marked in a checkbox under each individual). From there, PALS will only perform subsequent analyses on subjects that pass the visual inspection. If, however, users wish to run all subjects through the entire pipeline without

pausing, they are given the option to do so, but are highly encouraged to visually inspect all analyses steps after completion.

For users who simply wish to efficiently visualize lesion masks and do not wish to run other modules, PALS also offers the visual quality control feature as a stand-alone tool.

BASIC STRUCTURE OF PALS DIRECTORIES

PALS requires the user to specify the path to an *Input Directory* and an empty *Output Directory* (**Figure 3**).

Inputs

The *Input Directory* must contain separate *Subject Directories* for each subject. Each *Subject Directory* must at minimum contain: the subject's T1-weighted anatomical image in NifTI format, and one or more corresponding lesion masks, also in NifTI format. Importantly, all inputs should be in valid NifTI format and have the same image dimensions within each subject. T1 anatomical images for all subjects must contain the same T1 image identifier (e.g., T1 images for the first and second subject should be subj1_T1.nii.gz and subj2_T1.nii.gz, respectively); similarly lesion masks for all subjects must contain the same lesion mask identifier (e.g., subj1_Lesion.nii.gz and subj2_Lesion.nii.gz). If any subject has multiple lesions, each additional lesion mask must contain the lesion identifier, appended by the index, beginning with one for each additional lesion (e.g., subj1_Lesion1.nii.gz; see blue boxes in **Figure 3**).

Additionally, if the user chooses to run the Lesion Correction and/or Lesion Load Calculation modules, they are given the option to include the following files in each *Subject Directory*: a brain mask file (NifTI) and a white matter segmentation file (NifTI). If these steps have already been performed, brain extraction and white matter segmentation can be skipped during subsequent analyses. One caveat of this is that the same option must be implemented for all subjects in a given analysis pipeline. That is, the user cannot choose to skip brain extraction for only one subject; they would have to skip the step and provide their own brain mask files for all subjects.

If the user has already performed FreeSurfer cortical and subcortical segmentation for each subject, they may use subject-specific ROIs derived from FreeSurfer for lesion load calculation. If so, the user will be required to provide a (1) T1.mgz and (2) aparc + aseg.mgz parcellation and segmentation volume file from FreeSurfer outputs in each *Subject Directory*. The same caveat of pursuing the same option for all subjects applies.

Outputs

To encourage reproducible analysis, PALS also automatically creates time-stamped log files indicating selected options, inputs, and all processing steps each time it is run. These log files can be found in the source directory for PALS under the logs directory. This directory will only be created after the first run of PALS.

The general structure of the *Output Directory* will look similar to that of the *Input Directory*, with a separate directory created for each subject. Each new *Subject Directory* will contain
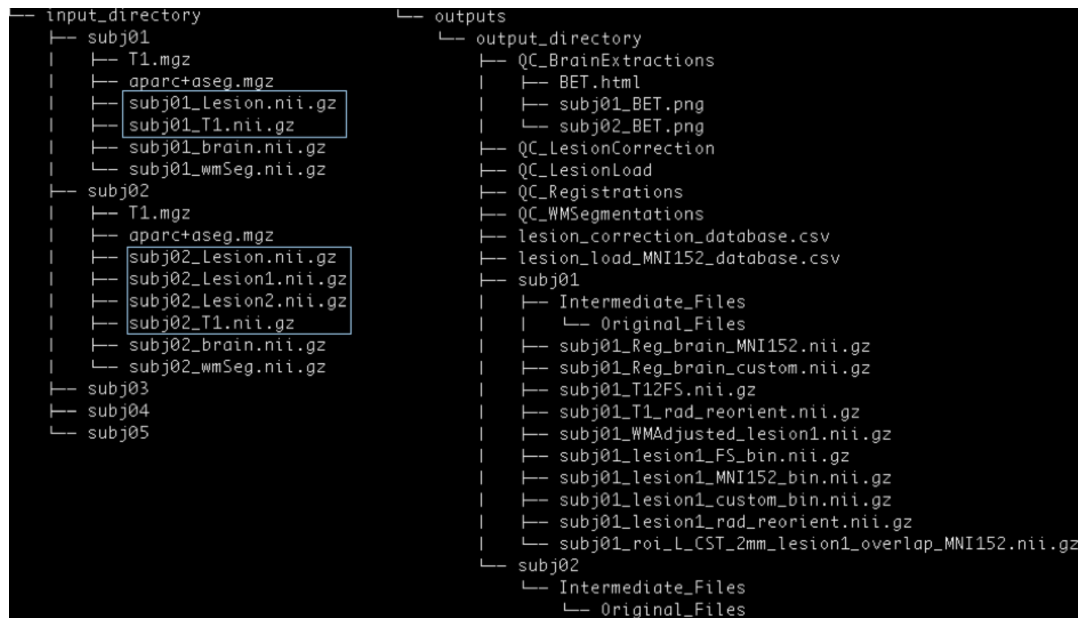
FIGURE 3 | PALS data structure. The user is expected to provide an input directory with subdirectories as shown above. Files in blue boxes are necessary inputs. The user is also expected to provide the path to an output directory, and PALS will create all other directories and files under the output directory.

the final outputs of the selected modules (e.g., white matter intensity adjusted lesion masks for the lesion correction module), and a subdirectory called *Intermediate_Files*, in which outputs from intermediary processing steps will be stored. Within the *Intermediate_Files* directory will be an *Original_Files* directory, which will contain a copy of all input files for the subject. Please see our github page[2] for a detailed description of each output file.

The *Output Directory* will additionally contain separate *QC* directories for each intermediary step taken (e.g., *QC_BrainExtractions* for the brain extraction step). These *QC* files will contain screenshots for each subject, and a single HTML page for manual visual control.

Finally, if the lesion correction and/or lesion load modules are selected, the *Output Directory* will also contain CSV files with information on the lesion (e.g., number of voxels removed during lesion correction, and percentage of lesion-ROI overlap per subject).

IMPLEMENTATION

Dependencies

PALS was built in Mac OSX on Python 2.7 and requires pre-installation of FSL. Separate installation of FSLeyes is necessary only if a version of FSL older than 5.0.10 is installed. FreeSurfer installation is necessary only if the user desires to use subject-specific FreeSurfer segmentations for the lesion load calculation module (see more information on lesion load calculation below).

PALS is compatible with Unix and Mac OS operating systems. For first-time users, PALS will ask users for the directory path to

FSL binaries. While we note that only 9 MB of space is needed for PALS installation (not including its dependencies), the total amount of space used for outputs created by the program will vary widely depending on the operations and number of subjects selected. Minimally, we recommend that 54 MB is allocated per subject, assuming only one ROI is selected for lesion load calculation, to run all operations.

Modules

Reorienting to Radiological Convention

The purpose of the reorient to radiological module is to make sure that lesion masks are in the same convention as the anatomical brain file, since some software used to create lesion masks may flip the orientation of the lesion file. Additionally, this module attempts to homogenize the orientation of files across subjects, especially when combining data across sites. Importantly, this module assumes that the conversion from DICOM to NifTI format was performed correctly. There should be no errors in data storage and no missing information in the NifTI header.

The reorientation module first checks the orientation of the T1 anatomical and lesion mask images. If they are already in the radiological convention as indicated by the image header, the image convention is conserved. If both T1 and lesion mask images are found to be in the neurological convention, the image data and image header for both the T1 anatomical image and associated lesion masks are changed to the radiological convention, using FSL commands *fslswapdim* and *fslorient*, respectively[3]. If, however, the T1 and lesion masks are not in the same convention, PALS flags the subject and does not perform

subsequent analyses on that subject. We recommend that the user perform a thorough check of all flagged subjects to verify that image orientations are correct by running FSL command *fslorient* on flagged images.

If the user has also provided additional optional inputs, such as a skull-stripped brain and/or white matter mask, those images are also reoriented to the radiological convention if they are not already. Finally, the FSL command *fslreorient2std* is applied on all images to reorient images to match the orientation of a standard T1-weighted template image (MNI152).

Lesion Correction for Healthy White Matter Voxels

The basic steps that SRQL, the original toolbox we created for lesion correction, implements for white matter lesion correction are outlined elsewhere (Ito et al., 2017). However, as several steps have been modified and updated for PALS, we describe the steps in detail here.

First, the intensity of each subject's T1 structural image is scaled to a range within 0 to 255 (intensity normalization). Skull stripping using FSL's Brain Extraction Toolbox (BET) and automated white matter segmentation using FSL's Automated Segmentation Tool (FAST) are then performed (Smith, 2002). The user is given an option to skip the skull-stripping and segmentation steps if he or she specifies that these steps have already been performed. If skull-stripping and/or white matter segmentation are performed, PALS will create a quality control page and the program will pause for the user to perform a visual inspection of each brain extraction/white matter segmentation.

Next, intensity normalized values from the T1 image (step 1) are projected onto the white matter segmentation as well as the binarized lesion mask, and the mean white matter intensity value is calculated from the white matter segmentation.

To calculate the upper and lower bounds for white matter intensity removal, the percent intensity for removal is first specified by the user. A default value of 5% is built into the toolbox. The specified percentage for removal is then converted to a 0 to 255 scale and divided by 2. This value is added to and subtracted from the mean white matter intensity value, such that:

$$\text{Intensity values to be removed} = \text{mean} \pm \frac{(255*\text{specified percentage \%})}{2}$$

Following this calculation, any voxels with intensity values within this range in the T1-projected lesion mask are removed, thereby removing voxels in the lesion mask that are within the specified intensity range of healthy white matter for that individual. As the last step, the white matter adjusted lesion mask file is binarized as a final lesion mask.

After lesion correction has been completed for all subjects, a CSV file containing information about the number of voxels removed for each subject's corrected lesion is created along with a quality control page for visual inspection of the effect of lesion correction on the lesion. The impact of using the lesion correction module is reported in validation (section IV), where we show that lesion correction decreases inter-individual variability on

manual segmentations, but does not improve upon automated segmentations.

Lesion Load Calculation

The lesion load calculation module computes the amount of lesion-ROI overlap with minimal input from the user. Notably, the user does not need to register or reslice regions of interest (ROI) to native space prior to using the lesion load module in PALS—PALS automatically normalizes all native space lesion masks and anatomical files to match that of the ROI. We offer several options for selecting ROIs in calculating lesion load based on commonly-used conventions for ROI analysis (Poldrack, 2007). (1) PALS comes with a set of default anatomical ROIs, all of which have been converted to standard 2 mm MNI152 space, including bilateral corticospinal tract ROIs (Riley et al., 2011), FreeSurfer subcortical and cortical ROIs (Fischl, 2012), and sensorimotor area tract ROIs (S-MATT; Archer et al., 2017). (2) We also allow calculation of lesion load using subject-specific FreeSurfer cortical and subcortical segmentations, if the user indicates that they have already performed FreeSurfer and have FreeSurfer-derived aparc + aseg.mgz and T1.mgz files for each subject. (3) Lastly, we give users the option of providing their own regions of interest to calculate lesion load. This option requires that the user also provides the standard space template of the regions of interest so that PALS can convert subject files to the ROI space.

After the ROIs are specified by the user, all ROIs are binarized, and lesion masks and T1 images are registered to the ROI space, whether it is MNI152 (default ROIs), FreeSurfer space, or user-defined. At this point, the program will pause again and have the user perform a visual inspection of the registrations to confirm that the normalization looks appropriate. Lesion masks are also binarized (such that only voxels within the lesions have a value of 1 and all other voxels have a value of 0), and summed with the voxel values of each binarized ROI mask using the FSL command *fslmaths*, so that regions that are overlapping between the lesion and ROI have a value of 2. Next, to obtain the mask of the lesion-ROI overlap, a threshold is applied to the combined lesion-ROI mask, such that anything below a value of 2 is zeroed. Finally, the lesion-ROI overlap mask is used to calculate the total percentage of overlap between the lesion and ROI, calculated as:

$$\text{Percentage overlap} = \frac{(\text{overlap volume between lesion and ROI})}{\text{ROI volume}}$$

The percentage overlap between the lesion and ROI is then saved into a CSV file containing lesion load information for all subjects, and a quality control page is created for visual inspection of lesion load performance.

Quality Control Webpages

PALS uses the FSL module *fsleyes*[4] to render screenshots of each subject's brain extraction overlaid on its corresponding T1-weighted image. The screenshots will display the overlays along the three orthogonal planes. These screenshots are concatenated
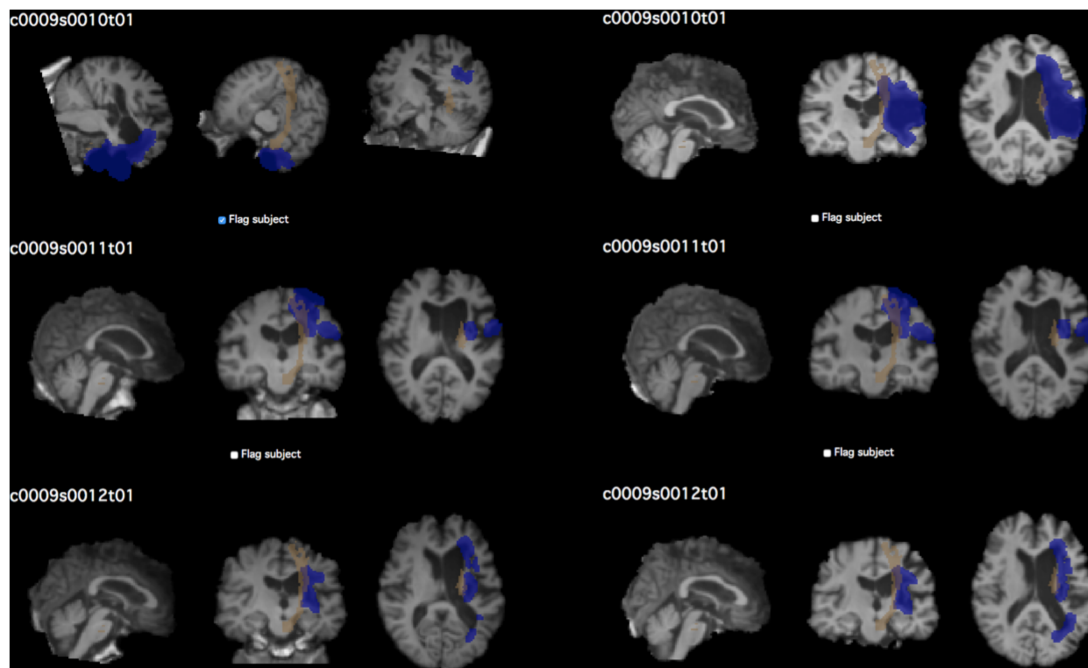
---

[4]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLeyes

FIGURE 4 | Example of quality control page. The QC page shown above was used to assess the quality of CST lesion load calculations in the test dataset. Each subject's screenshot is displayed with a checkbox below for flagging subjects that do not pass visual inspection. Left, prior to performing QC; right, after correcting brain extraction. The first subject in particular demonstrates the vast difference that quality inspection makes. Brown shows the CST ROI; blue shows the lesion mask of each subject.

into a single HTML page for review. Below each screenshot is a checkbox for the user to indicate whether the subject's brain extraction passes visual inspection.

The same process was used to create quality control pages for each subject's white matter segmentation mask and registered brain masks. For lesion masks, both the final white matter adjusted mask as well as the original manually traced lesion mask were overlaid onto the T1-image for comparison of lesions before and after lesion correction. For lesion load calculations, the lesion mask is overlaid on the selected region of interest for calculating the lesion load (see **Figure 4**).

VALIDATION

Reorient to Radiological

Here, we validated that this tool performs each function correctly. For this evaluation, we simply wanted to test as many cases as possible, and used a combined dataset of 355 MRIs and lesion masks from 12 research sites [11 from the Anatomical Tracings of Lesions after Stroke (ATLAS) database, and one additional from a collaborator; Liew et al., 2018]. We checked that PALS correctly flagged all 30 subjects whose lesion mask and anatomical T1 files had mismatching orientations. For all other images that were not flagged, the module corrected identified images in the neurological orientation and transformed them to radiological orientation (see **Supplementary Table S1**).

We next simulated cases to confirm that PALS also correctly identified subjects with mismatched orientations in optional inputs (e.g., where a brain mask and/or white matter mask file are provided by the user in addition to the necessary T1 and lesion mask files). Additionally, we created a case in which all but one subject contained the additional optional inputs. We checked that PALS was able to correctly identify when orientations of inputs were mismatched for any subjects, and overrides the user input to skip brain extraction and/or white matter segmentation if a subject is missing those inputs (**Table 1**). For additional simulation cases, see **Supplementary Tables S2–S4**.

Lesion Correction for Healthy White Matter Voxels

Inter-Rater Reliability

For this module, we tested whether PALS could improve inter-rater reliability on five manually segmented lesion masks (**Figure 5**). Ten trained research assistants manually traced stroke lesions from five separate brains with lesions of different sizes (Liew et al., 2018). For each stroke brain, we calculated a dice correlation coefficient (DC) for each pair among manual tracings by 10 different trained individuals to evaluate agreement between all raters. The dice correlation coefficient is a measure of similarity between two images, and is defined as:

$$DC = 2 * \frac{|X \cap Y|}{|X| + |Y|}$$

TABLE 1 | Validation for reorient to radiological module.

Cases	Input orientations	Output orientations
Case 1	Lesion: neurological	Radiological
	T1: neurological	Radiological
	Brain: neurological	Radiological
	WM: neurological	Radiological
Case 2	Lesion: radiological	Radiological
	T1: radiological	Radiological
	Brain: radiological	Radiological
	WM: radiological	Radiological
Case 3	Lesion: neurological	Flagged
	T1: neurological	
	Brain: radiological	
	WM: neurological	
Case 4	Lesion: neurological	Flagged
	T1: neurological	
	Brain: radiological	
	WM: radiological	
Case 5	Lesion: neurological	Flagged
	T1: radiological	
	Brain: radiological	
	WM: radiological	
Case 6: Optional input missing	Lesion: neurological	Radiological
	T1: neurological	Radiological
	Brain: missing	Brain extraction set to run on all subjects
	WM: missing	WM segmentation set to run for all subjects

Simulated cases including subjects with T1, lesion mask (Lesion), brain mask (Brain), and white matter segmentation (WM) inputs with varying orientations. PALS correctly flagged cases in which orientations of input files were mismatched.

where DC ranges between 0 (no overlap) and 1 (complete overlap), and X represents voxels in the first lesion volume, and Y represents voxels in the second lesion volume. For each stroke brain, the average of all DC values from all 45 pairwise comparisons of manual segmentations was calculated as a mean inter-rater DC score, and then mean inter-rater DC scores across the five stroke brains was again averaged for an overall inter-rater DC score. We then ran both SRQL, our previous version of lesion correction, and PALS-lesion correction on all manual segmentations, using the default value of 5% lesion white matter intensity removal, to compare which performed better, and recalculated the overall inter-rater dice coefficient score on white matter adjusted lesion masks (**Table 2**).

We next performed a one-way repeated measures ANOVA on the mean inter-rater DC scores averaged over the five lesions to determine whether there were any differences between inter-rater scores without any correction, with lesion correction from SRQL, and with the new PALS lesion correction. Mauchly's test indicated that the assumption of sphericity had been violated ($p < 0.001$), therefore Greenhouse–Geisser corrected tests are reported ($\varepsilon = 0.514$). We found a significant difference among

the inter-rater DC scores ($F = 5.91$, $p = 0.0183$); Tukey *post hoc* comparisons with Bonferroni correction showed that inter-rater DC scores, the average number of voxels overlapping between manual segmentations of the tracers (see above for description), were significantly higher after lesion correction with PALS compared to lesion masks without any adjustment ($z = 3.43$, $p = 0.0018$); other pairwise comparisons did not reach significance ($p > 0.18$). In other words, the lesion correction module in PALS significantly improved the similarity between manual tracings across the 10 tracers. We thus recommend using the PALS lesion correction module when analyzing manually traced lesions.

Automated vs. Manual Lesion Segmentations

We were also interested in assessing whether lesion correction could improve similarity between automated segmentations and manual segmentations, the latter considered the gold standard for lesion segmentation. For this evaluation, we used 90 stroke T1-weighted MRIs from the publicly-available ATLAS database (Liew et al., 2018). The ATLAS database consists of chronic stroke (>6 months) MRIs obtained across 11 research groups worldwide, and also includes manually segmented lesion masks for each MRI, created by a team of trained individuals (for further information on the full lesion dataset and labeling protocol, see Liew et al., 2018). The 90 brains included for this evaluation consisted of 34 cortical, 54 subcortical, and 2 cerebellar lesions on both left ($n = 36$) and right ($n = 54$) hemispheres. Lesion volume ranged from 386 to 164,300 mm^3 ($M = 31,578.41$, $SD = 38,582.13$) based on manual segmentations.

We used the manually segmented lesions included in the ATLAS database as our gold standard of manually traced lesion masks. We then used the lesion identification with neighborhood data analysis (LINDA) approach to automatically segment the 90 stroke T1-weighted MRIs (Pustina et al., 2016). Finally, we calculated the dice DC between each automated segmentation and manually traced lesion and obtained an average DC of 0.58 ± 0.25 (range 0.006 to 0.88).

We note that a DC value of 0.58 is relatively low considering that DC ranges between 0 and 1. However, given that limitations still exist with performance of automated lesion segmentation algorithms, particularly for single-modality data and for data that have been pooled together from different sites, such as the ATLAS database, an average DC of 0.58 is fairly standard (Ito, Kim, and Liew, *under review;* for a representative example, see **Figure 6**).

Removing White Matter From Manual Tracings

We next performed lesion correction on manually traced lesions, using the default 5% white matter intensity removal, and re-calculated DC to determine whether lesion correction improved similarity between the manual and automated segmentations. We found that lesion correction on manual lesions made no difference on similarity between manual and automated lesions (average DC before and after correction: 0.58 ± 0.24; $t = 1.59$, $p = 0.11$).

Removing White Matter From Automated Segmentations

Finally, we assessed whether lesion correction on automated segmentations could improve similarity to manually traced
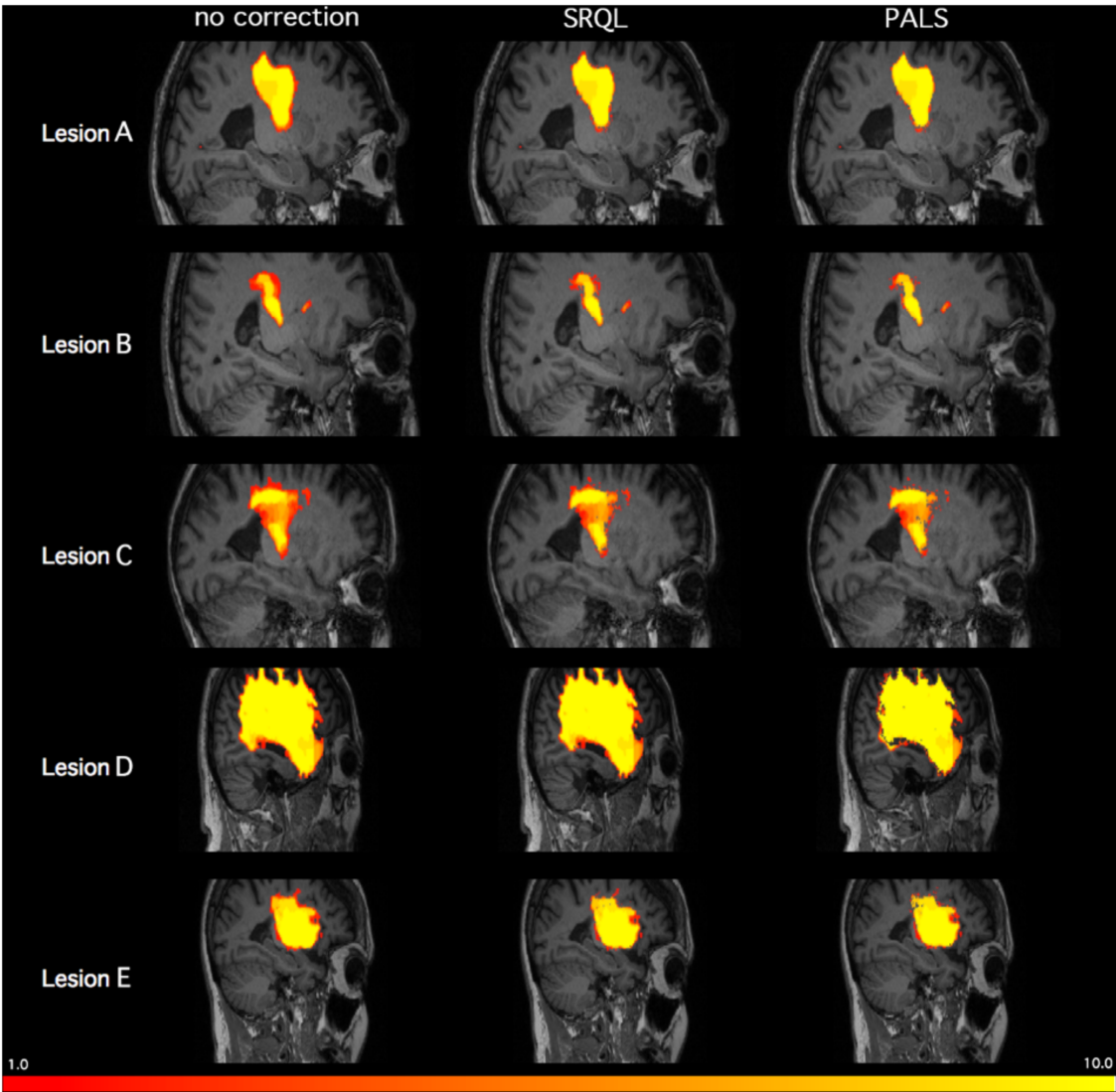
FIGURE 5 | Inter-tracer heat maps of five stroke lesions with and without lesion correction. Brighter colors indicate greater overlap between tracers; as shown above in regions with less bright colors and greater dark red, PALS improves similarity between inter-rater tracings, particularly when lesion correction is included.

TABLE 2 | Inter-rater Dice Correlation Coefficient values with and without lesion correction.

	Lesion A	Lesion B	Lesion C	Lesion D	Lesion E	Average
No correction	0.84 ± 0.006	0.62 ± 0.03	0.66 ± 0.02	0.93 ± 0.002	0.81 ± 0.01	0.77 ± 0.009
SRQL	0.87 ± 0.006	0.63 ± 0.03	0.69 ± 0.01	0.93 ± 0.002	0.82 ± 0.01	0.79 ± 0.009
PALS-LesionCorr	0.87 ± 0.006	0.65 ± 0.03	0.72 ± 0.014	0.94 ± 0.002	0.84 ± 0.009	0.80 ± 0.009

Average dice coefficient values (mean ± standard deviation) for manual tracings across 10 trained individuals. 1, perfectly overlapping; 0, no overlap. PALS-Lesion correction consistently improved the dice coefficient across all lesions compared to no correction or the earlier SRQL toolbox.

lesions. We thus applied lesion correction using default values on lesions automatically segmented using LINDA, and calculated DC between manual segmentations (without lesion correction) and white matter corrected automated segmentations for the 90 brain lesions. Here, we found that lesion corrections did not improve similarity between manual and automated lesions and actually significantly decreased similarity by a small amount (average DC before: 0.58 ± 0.03; average DC after correction: 0.57 ± 0.24; $t = 2.58$, $p = 0.01$).

Lesion Load Calculation and Quality Control

We validated our lesion load calculation module with a CST lesion load calculation tool implemented by a separate research
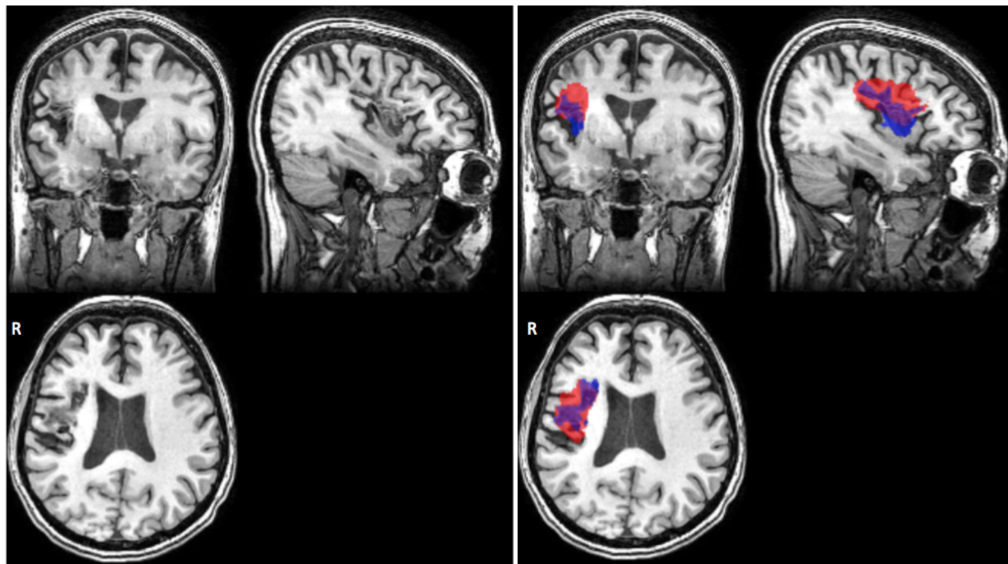
FIGURE 6 | Representative case of automated versus manual lesion segmentation. Left, an individual's T1w anatomical MRI; Right, the manual lesion mask in blue overlaid on automated lesion mask produced by the LINDA algorithm, in red (Pustina et al., 2016). DC, 0.57 for this lesion.

group (Riley et al., 2011). As their group divided up their CST ROI into 16 longitudinal strings to obtain the lesion-to-CST percentage overlap (see Riley et al., 2011 for methods), we also tested the PALS lesion load calculation module with identical ROI input, courtesy of Riley et al. (2011).

For this evaluation, we implemented both lesion load calculation tools on 122 brains from the ATLAS dataset. These brains were made up of 40 cortical, 67 subcortical, 12 brainstem, and 3 cerebellar strokes. To validate that the lesion load tool correctly assesses the hemisphere of the lesion, we included both left ($n = 70$) and right ($n = 40$) hemisphere lesions (and 12 brainstem lesions). Lesion volume ranged from 27 to 62,460 mm^3 ($M = 7,391.48$ mm^3, SD $= 9,060.82$ mm^3).

We used the left CST ROI, and found a strong significant correlation between the PALS method and the previously described method from Riley et al. (2011) (PALS average CST lesion load percentage: 44.96 ± 44.70%; Riley CST lesion load: 43.75 ± 44.30%; $r = 0.87$, $p < 0.0001$). We also verified that the CST lesion load percentage was equal to 0% on all right hemisphere lesions. However, the correlation was lower than expected. Using our QC tool, we visually inspected the quality of the intermediary outputs created by PALS and identified seven cortical stroke brains that performed poorly on brain extraction and registration. We cleaned up the brain extractions using additional features in FSL's *BET* (e.g., bias field and neck cleanup), and reran these brains through the PALS pipeline, feeding in the cleaned-up brain extractions (**Figure 4**). As expected, this substantially improved registration. We then re-calculated the CST lesion load as well as the correlation between the values obtained through PALS and the method described above from Riley et al. (2011). Doing so resulted in a stronger correlation between the two lesion load calculation tools (PALS average CST lesion load percentage: 47.33 ± 45.24%; $r = 0.96$, $p < 0.0001$). This

demonstrates the importance of performing a thorough quality inspection on each processing step and overall confirms that our tool accurately calculates lesion overlap in accordance with previous work.

We additionally assessed whether the accuracy of lesion load calculation differed between cortical and subcortical lesions. As such, we split these 122 validation cases by category (40 cortical, 67 subcortical, excluding brainstem and cerebellar lesions). We then calculated the Pearson's correlation coefficient by stroke category, to assess how well the PALS method compares to the method implemented by Riley et al. (2011). For cortical strokes, we obtained a correlation coefficient of $r = 0.73$, $p < 0.0001$; 95% CI [0.54, 0.85]. However, after correcting for image processing errors that occurred for the seven brains mentioned above, we obtained the following values for cortical strokes: $r = 0.98$, $p < 0.0001$; 95% CI [0.97, 0.99]. For subcortical strokes, we obtained a correlation coefficient of $r = 0.95$, $p < 0.0001$; 95% CI [0.93, 0.97]. Again, this demonstrates the susceptibility of larger, cortical strokes to image processing errors and highlights the importance of quality control.

DISCUSSION

Despite the recent surge of interest in big data neuroimaging, the infrastructure and image processing pipelines necessary to support it, particularly for stroke lesion analysis, are still severely lacking. To this end, we created an open-source toolbox with a user-friendly GUI to help standardize stringent stroke lesion analyses. A detailed manual and source code can be downloaded from our github repository[5].

---

[5]https://github.com/npnl/PALS

To demonstrate some of the key features of the toolbox, we validated use with multi-site data. We demonstrated that PALS successfully harmonizes data to be in the same orientation convention across sites. We also showed that PALS increases inter-rater reliability of manual tracings: applying the lesion correction module in PALS significantly increased similarity between manually segmented lesions compared to no lesion correction and our previous version of lesion correction from the SRQL toolbox. However, we found that similarity between manual segmentations and automated segmentations, in cases where groups might try to use manual segmentations for a subset of the data and automated segmentation in another subset of data, did not improve when applying PALS lesion correction on either the manual segmentations or the automated segmentations. A likely explanation for this is that the automated segmentations algorithm we used (LINDA; Pustina et al., 2016) already included a tissue classification step in the derivation of features, which would prevent white matter voxels to be classified as lesion tissue. We thus recommend that research groups do not mix different lesion segmentation methods (e.g., a subset manually and a subset with an automated algorithm) for the PALS lesion correction module, but rather use lesion correction only for datasets with all manual lesion segmentations. This is because applying white matter intensity removal to human errors in manual segmentations would provide a systematic way to remove voxels within the designated healthy white matter intensity range that might be missed due to human bias (Riley et al., 2011). Finally, we also showed that PALS lesion load calculation module is comparable to another CST lesion load calculator implemented by a different research group.

Limitations and Future Directions

PALS was created to respond to the need for reliable image processing pipelines for collaborative efforts in stroke neuroimaging. PALS integrates multiple functions into a single analysis pipeline to facilitate lesion analysis and quality control. However, the PALS toolbox has a few limitations. First, as PALS was created to address the need for lesion analysis software that takes a single modality, we have only tested the toolbox on T1w MRI data. We hope to expand these tools for other types of multimodal stroke imaging, such as T2 or FLAIR sequences, in the future. However, we will plan to retain the option for using

a single channel input so that users will not be required to have multi-modal data to use PALS. Additionally, in the reorient to radiological module, the PALS toolbox makes the assumption that the input files are in valid NifTI format, which requires proper user input.

We plan to continue to refine our software in the future based on feedback and comments from users[6], and hope to expand these tools for other multimodal stroke imaging data types. We hope our toolbox will be useful to clinicians and researchers, and foster greater collaboration leading to the discovery of new clinical insights.

AUTHOR CONTRIBUTIONS

KI implemented the toolbox, tested the toolbox, and drafted the manuscript. AK implemented the toolbox. AZ-P tested the toolbox and revised the manuscript. SC contributed to the conceptualization of toolbox modules and revised the manuscript. S-LL conceptualized the toolbox, tested the toolbox, and revised the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2018.00063/full#supplementary-material

---

[6] https://github.com/npnl/PALS/issues

REFERENCES

Andersen, S. M., Rapcsak, S. Z., and Beeson, P. M. (2010). Cost function masking during normalization of brains with focal lesions: still a necessity? *Neuroimage* 53, 78–84. doi: 10.1016/j.neuroimage.2010.06.003

Archer, D. B., Vaillancourt, D. E., and Coombes, S. A. (2017). A template and probabilistic atlas of the human sensorimotor tracts using diffusion MRI. *Cereb. Cortex* 28, 1685–1699. doi: 10.1093/cercor/bhx066

Avants, B. B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTS). *Insight J.* 2, 1–35.

Brett, M., Hanke, M., Larson, E., and Markiewicz, C. (2017). *Radiological vs Neurological Conventions*. Avilable at: http://nipy.org/nibabel/neuro_radio_conventions.html#radiological-vs-neurological-conventions

Byblow, W. D., Stinear, C. M., Barber, P. A., Petoe, M. A., and Ackerley, S. J. (2015). Proportional recovery after stroke depends on corticomotor integrity. *Ann. Neurol.* 78, 848–859. doi: 10.1002/ana.24472

Crafton, K. R., Mark, A. N., and Cramer, S. C. (2003). Improved understanding of cortical injury by incorporating measures of functional anatomy. *Brain* 126, 1650–1659. doi: 10.1093/brain/awg159

Cramer, S. C. (2010). Stratifying patients with stroke in trials that target brain repair. *Stroke* 41(10 Suppl. 1), 114–117. doi: 10.1161/STROKEAHA.110.595165

Dade, L. A., Gao, F. Q., Kovacevic, N., Roy, P., Rockel, C., O'Toole, C. M., et al. (2004). Semiautomatic brain region extraction: a method of parcelling brain regions from structural magnetic resonance images. *Neuroimage* 22, 1492–1502. doi: 10.1016/j.neuroimage.2004.03.023

Dimyan, M. A., and Cohen, L. G. (2011). Neuroplasticity in the context of motor rehabilitation after stroke. *Nat. Rev. Neurosci.* 7, 76–85. doi: 10.1038/nrneurol. 2010.200

Duff, E. (2015). *Orientation Explained.* Available at: https://fsl.fmrib.ox.ac.uk

Feldmann, E., and Liebeskind, D. (2014). Developing precision stroke imaging. *Front. Neurol.* 5:29. doi: 10.3389/fneur.2014.00029

Feng, W., Wang, J., Chhatbar, P., Doughty, C., Landsittel, D., Lioutas, V., et al. (2015). Corticospinal tract lesion load - a potential imaging biomarker for stroke motor outcomes. *Ann. Neurol.* 78, 860–870. doi: 10.1002/ana.24510. Corticospinal

Fiez, J. A., Damasio, H., and Grabowski, T. J. (2000). Lesion segmentation and manual warping to a reference brain: intra- and interobserver reliability. *Hum. Brain Mapp.* 9, 192–211. doi: 10.1002/(SICI)1097-0193(200004)9:4<192::AID-HBM2>3.0.CO;2-Y

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage. 2012.01.021

González, R. G. (2006). Imaging-guided acute stroke therapy: from "time is brain" to "physiology is brain. *Am. J. Neuroradiol.* 27, 728–735.

Gorgolewski, K. J., and Poldrack, R. A. (2016). A Practical guide for improving transparency and reproducibility in neuroimaging. *PLoS Biol.* 14:e1002506. doi: 10.1371/journal.pbio.1002506

Hachinski, V., Donnan, G. A., Gorelick, P. B., Hacke, W., Cramer, S. C., and Kaste, M. (2010). Editorials stroke?: working toward a prioritized world agenda. *Int. J. Stroke* 5, 238–256. doi: 10.1111/j.1747-4949.2010.00442.x

Huang, B. E., Mulyasasmita, W., and Rajagopal, G. (2016). The path from big data to precision medicine. *Expert Rev. Precis. Med. Drug Dev.* 1, 129–143. doi: 10.1080/23808993.2016.1157686

Ito, K. L., Anglin, J., Kim, H., and Liew, S.-L. (2017). Semi-automated robust quantification of lesions (SRQL) toolbox. *Res. Ideas Outcomes* 3:e12259. doi: 10.3897/rio.3.e12259

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09. 015

Liew, S.-L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., et al. (2018). A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. *Sci. Data* 5:180011. doi: 10.1038/sdata.2018.11

Maier, O., Menze, B. H., Gablentz, J., Von Der Häni, L., Heinrich, M. P., Liebrand, M., et al. (2017). ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Med. Image Anal.* 35, 250–269. doi: 10.1016/j.media.2016.07.009

Milham, M. P. (2012). Open neuroscience solutions for the connectome-wide association era. *Neuron* 73, 214–218. doi: 10.1016/j.neuron.2011.11.004

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical Parametric Mapping: the Analysis of Functional Brain Images.* New York, NY: Elsevier.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Soc. Cogn. Affect. Neurosci.* 2, 67–70. doi: 10.1093/scan/nsm006

Puig, J., Pedraza, S., Blasco, G., Daunis-i-Estadella, J., Prados, F., Remollo, S., et al. (2011). Acute damage to the posterior limb of the internal capsule on diffusion tensor tractography as an early imaging predictor of motor outcome after stroke. *Am. J. Neuroradiol.* 32, 857–863. doi: 10.3174/ajnr.1 A2400

Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., and Avants, B. (2016). Automated segmentation of chronic stroke lesions using LINDA: lesion identification with neighborhood data analysis. *Hum. Brain Mapp.* 37, 1405–1421. doi: 10.1002/hbm.23110

Riley, J. D., Le, V., Der-Yeghiaian, L., See, J., Newton, J. M., Ward, N. S., et al. (2011). Anatomy of stroke injury predicts gains from therapy. *Stroke* 42, 421–426. doi: 10.1161/STROKEAHA.110.599340

Rondina, J. M., Park, C. H., and Ward, N. S. (2017). Brain regions important for recovery after severe post-stroke upper limb paresis. *J. Neurol. Neurosurg. Psychiatry* 88, 737–743. doi: 10.1136/jnnp-2016-315030

Rorden, C., and Brett, M. (2000). Stereotaxic display of brain lesions. *Behav. Neurol.* 12, 191–200. doi: 10.1155/2000/421719

Siegel, J. S., Shulman, G. L., and Corbetta, M. (2017). Measuring functional connectivity in stroke: approaches and considerations. *J. Cereb. Blood Flow Metab.* 37, 2665–2678. doi: 10.1177/0271678X17709198

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155. doi: 10.1002/hbm.10062

Stinear, C. M. (2017). Prediction of motor recovery after stroke: advances in biomarkers. *Lancet Neurol.* 16, 826–836. doi: 10.1016/S1474-4422(17) 30283-1

Van Horn, J. D., and Toga, A. W. (2014). Human neuroimaging as a "Big Data" science. *Brain Imaging Behav.* 8, 323–331. doi: 10.1007/s11682-01 3-9255-y

Zhu, L. L., Lindenberg, R., Alexander, M. P., and Schlaug, G. (2010). Lesion load of the corticospinal tract predicts motor impairment in chronic stroke. *Stroke* 41, 910–915. doi: 10.1161/STROKEAHA.109.577023

Check for updates

Diffusion MRI Indices and Their Relation to Cognitive Impairment in Brain Aging: The Updated Multi-protocol Approach in ADNI3

Artemis Zavaliangos-Petropulu[1†], Talia M. Nir[1†], Sophia I. Thomopoulos[1], Robert I. Reid[2], Matt A. Bernstein[3], Bret Borowski[3], Clifford R. Jack Jr.[3], Michael W. Weiner[4], Neda Jahanshad[1], Paul M. Thompson[1*] and the Alzheimer's Disease Neuroimaging Initiative (ADNI)[‡]

[1]Imaging Genetics Center, Mark & Mary Stevens Neuroimaging & Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, United States, [2]Department of Information Technology, Mayo Clinic and Foundation, Rochester, MN, United States, [3]Department of Radiology, Mayo Clinic, Rochester, MN, United States, [4]Department of Radiology, School of Medicine, University of California, San Francisco, San Francisco, CA, United States

Brain imaging with diffusion-weighted MRI (dMRI) is sensitive to microstructural white matter (WM) changes associated with brain aging and neurodegeneration. In its third phase, the Alzheimer's Disease Neuroimaging Initiative (ADNI3) is collecting data across multiple sites and scanners using different dMRI acquisition protocols, to better understand disease effects. It is vital to understand when data can be pooled across scanners, and how the choice of dMRI protocol affects the sensitivity of extracted measures to differences in clinical impairment. Here, we analyzed ADNI3 data from 317 participants (mean age: 75.4 ± 7.9 years; 143 men/174 women), who were each scanned at one of 47 sites with one of six dMRI protocols using scanners from three different manufacturers. We computed four standard diffusion tensor imaging (DTI) indices including fractional anisotropy (FA^{DTI}) and mean, radial, and axial diffusivity, and one FA index based on the tensor distribution function (FA^{TDF}), in 24 bilaterally averaged WM regions of interest. We found that protocol differences significantly affected dMRI indices, in particular FA^{DTI}. We ranked the diffusion indices for their strength of association with four clinical assessments. In addition to diagnosis, we evaluated cognitive impairment as indexed by three commonly used screening tools for detecting dementia and AD: the AD Assessment Scale (ADAS-cog), the Mini-Mental State Examination (MMSE), and the Clinical Dementia Rating scale sum-of-boxes (CDR-sob). Using a nested random-effects regression model to account for protocol and site, we found that across all dMRI indices and clinical measures, the hippocampal-cingulum and fornix *(crus)/stria terminalis* regions most consistently showed strong associations with clinical impairment. Overall, the greatest effect sizes were detected in the hippocampal-cingulum (CGH) and uncinate fasciculus (UNC) for associations between axial or mean diffusivity and CDR-sob. FA^{TDF} detected robust widespread associations with

clinical measures, while FA^{DTI} was the weakest of the five indices for detecting associations. Ultimately, we were able to successfully pool dMRI data from multiple acquisition protocols from ADNI3 and detect consistent and robust associations with clinical impairment and age.

Keywords: Alzheimer's disease, ADNI3, white matter, DTI, multi-site, harmonization, TDF, ComBat

INTRODUCTION

Alzheimer's disease (AD) is the most common type of dementia, affecting approximately 10% of the population over age 65 (Alzheimer's Association, 2018). As life expectancy increases, there is an ever-increasing need for sensitive biomarkers of AD—to better understand the disease, and to serve as surrogate markers of disease burden for use in treatment and prevention trials. The Alzheimer's Disease Neuroimaging Initiatve (ADNI) is an ongoing large-scale, multi-center, longitudinal study designed to improve methods for clinical trials by identifying brain imaging, clinical, cognitive, and molecular biomarkers of AD and aging. Now in its third phase (ADNI3), ADNI continues to incorporate newer technologies as they become established (Jack et al., 2015); data from ADNI, collected at participating sites across the U.S. and Canada, is publicly available and has been used in a diverse range of publications (Veitch et al., 2019).

ADNI's second phase (ADNI2) introduced to the initiative the use of diffusion-weighted MRI (dMRI) as an additional approach for tracking AD progression (Jack et al., 2015). dMRI has since been used in numerous studies to understand the effects of AD on white matter (WM) microstructure and brain connectivity (Daianu et al., 2013a,b; Nir et al., 2013; Prasad et al., 2013). Some of these approaches use scalar dMRI measures to evaluate microstructural WM changes not detectable with anatomical T1-weighted images (Giulietti et al., 2018), while others use tractography and graph-theory analysis to study abnormalities in structural brain networks (Nir et al., 2015; Hu et al., 2016; Maggipinto et al., 2017; Sulaimany et al., 2017; Powell et al., 2018; Sanchez-Rodriguez et al., 2018). In aggregate, these studies point to WM abnormalities in AD, which may play a key role in early pathogenesis and diagnosis (Sachdev et al., 2013).

ADNI2 acquired dMRI data with one acquisition protocol from approximately one third of enrolled participants at the subset of ADNI sites that used 3 tesla General Electric (GE) scanners. To ensure that dMRI could be collected from all enrolled participants, ADNI3 developed new dMRI protocols for all GE, Siemens and Philips scanners used across ADNI sites. Now, data is being acquired with seven different dMRI acquisition protocols (see "Materials and Methods" section for details[1]). ADNI3 began in October 2016, and has already acquired data from over 300 participants. dMRI spatial resolution was improved between ADNI2 and ADNI3 by reducing the voxel size from 2.7 × 2.7 × 2.7 mm to 2.0 × 2.0 × 2.0 mm.

---

[1]http://adni.loni.usc.edu/methods/documents/mri-protocols/

While voxel size (i.e., spatial resolution) remains consistent across all seven ADNI3 protocols, angular resolution (the number of gradient directions) varies across protocols to accommodate scanner restrictions and to ensure that the multi-modal scanning session is completed in under 60 min. Although many large-scale multi-site DTI studies have obtained consistent results even when acquisition protocols across sites are not harmonized in advance (Jahanshad et al., 2013; Kochunov et al., 2014; Acheson et al., 2017; Kelly et al., 2018), differences in dMRI acquisition parameters, including vendor, voxel size, and angular resolution, are known to affect derived dMRI measures (Alexander et al., 2001; Cercignani et al., 2003; Zhan et al., 2010; Zhu et al., 2011). As a result, improved harmonization of multi-site diffusion data is of great interest (Grech-Sollars et al., 2015; Pohl et al., 2016; Palacios et al., 2017). For example, ComBat—originally developed to model and remove batch effects from genomic microarray data (Johnson et al., 2007)—was one of the most effective methods for harmonizing DTI measures in a recent comparison of such techniques (Fortin et al., 2017).

Here, we tested whether standard diffusion tensor imaging (DTI)-derived anisotropy and diffusivity indices, calculated from multiple imaging protocols in ADNI3, can be pooled and harmonized to show robust associations with age and four clinical assessments. In addition to diagnosis, cognitive impairment was assessed with three commonly used screening tools for detecting dementia and AD: the Alzheimer's Disease Assessment Scale (ADAS-cog; Rosen et al., 1984), the Mini-Mental State Examination (MMSE; Folstein et al., 1975), and the Clinical Dementia Rating scale sum-of-boxes (CDR-sob; Berg, 1988). For the rest of the article we refer to these tools as "cognitive measures". In addition to standard DTI indices—fractional anisotropy (FA^{DTI}), mean diffusivity (MD^{DTI}), radial diffusivity (RD^{DTI}), and axial diffusivity (AxD^{DTI})—we also evaluated a modified measure of FA, derived from the tensor distribution function (FA^{TDF}; Leow et al., 2009) which can be more sensitive to neurodegenerative disease-related WM abnormalities than FA^{DTI} across high- and low-angular resolution dMRI (Nir et al., 2017). The TDF model addresses well-established limitations of the standard single-tensor diffusion model—which cannot resolve complex profiles of WM architecture such as crossing or mixing fibers, present in up to 90% of WM voxels (Tournier et al., 2004; Descoteaux et al., 2007, 2009; Jeurissen et al., 2013).

In 24 WM regions of interest (ROIs), we ranked these five anisotropy and diffusivity indices, in terms of their strength of association with key clinical measures, to identify dMRI

indices that may help understand and track AD progression. We hypothesized that the diffusion indices from ADNI2 (Nir et al., 2013, 2017) would still be associated with clinical measures of disease burden in ADNI3—despite the variation in protocols. We hypothesized that when data were pooled across ADNI3 protocols: (1) higher diffusivity and lower anisotropy in the temporal lobe WM would be most sensitive to cognitive impairment, with highest effect sizes for associations with CDR-sob; and (2) FA^{TDF} would detect associations with clinical impairment with larger effect sizes than FA^{DTI}.

MATERIALS AND METHODS

ADNI Participants

Baseline MRI, DTI, diagnosis, demographics, and cognitive measures were downloaded from the ADNI database[2]. This analysis was performed when data collection for ADNI3 was still ongoing (May 2018), and reflects the data available on April 30, 2018. Of the 381 participants scanned to date, 55 were excluded after quality assurance: this included ensuring complete clinical and demographic information, and image-level quality control (removing scans with severe motion, missing volumes, or corrupt files). To ensure sufficient statistical power to assess differences in data collected with different protocols, we evaluated only those protocols with complete available data for at least 10 participants at the time of download; we did not assess protocol GE36, for which scans from 9 of 12 participants passed quality assurance. Details on excluded participants are outlined in **Supplementary Table S1**.

Three-hundred and seventeen remaining participants—from 47 scanning sites—were included in the analysis (mean age: 75.4 ± 7.9 years; 143 men, 174 women; **Table 1**): 211 were elderly cognitively normal (CN) controls (mean age: 74.5 ± 7.3 years; 84 men, 127 women), 84 were diagnosed with mild cognitive impairment (MCI); mean age: 76.3 ± 8.1 years; 48 men, 36 women), and 22 were diagnosed with AD (mean age: 80.6 ± 10.5 years; 11 men, 11 women). We note that two of the ADNI2 diagnostic categories—CN and significant memory concern (SMC)—are combined and identified as CN in ADNI3.

2https://ida.loni.usc.edu/

ADNI2's early and late MCI categories are combined and identified as MCI in ADNI3.

Clinical Assessments

In addition to diagnosis, we indexed cognitive impairment using total scores from commonly used screening tools for detecting dementia and AD (**Table 1**): the Alzheimer's Disease Assessment Scale 13 (ADAS-cog), the Mini-Mental Status Examination (MMSE), and the Clinical Dementia Rating scale sum-of-boxes (CDR-sob). We refer to these tools as "cognitive measures", but recognize the limitations of these assessments as proxy measures of specific cognitive abilities (Balsis et al., 2015). The ADAS-cog is frequently used in pharmaceutical trials, with scores ranging from 0 to 70; higher scores represent more severe cognitive dysfunction (Rosen et al., 1984). MMSE is more often used by clinicians and researchers in assessing cognitive aging. Scores for MMSE range from 0 to 30; lower scores typically indicate greater cognitive dysfunction (Folstein et al., 1975). CDR-sob is used primarily in clinical trials and in clinical practice for evaluating disease severity including the mild and early symptomatic stages of dementia. It is calculated based on the sum of severity ratings in six domains ("boxes")—memory, orientation, judgment and problem solving, community affairs, home and hobbies, and personal care. Scores range from 0 (no dementia) to 3 (severe dementia; Berg, 1988). These evaluations are among the measures used in diagnosing ADNI participants. Not all cognitive measures were available for every participant (MMSE, $N = 315$; CDR-sob, $N = 316$, and ADAS-cog, $N = 278$; **Supplementary Table S2** lists these by protocol).

Diffusion MRI Acquisition Protocols

ADNI3 incorporated dMRI protocols for 3 tesla Siemens, Philips, and GE scanners. ADNI2, the first phase of ADNI to include diffusion MRI, only prescribed dMRI protocols for GE scanners. The available scanners span a wide range of software capabilities, such as support (or the lack of it) for custom diffusion gradient tables and/or simultaneous multi-slice acceleration. Including additional scanners while staying in a 7–10-min scan duration resulted in data acquired with seven different acquisition protocols—of which six had

TABLE 1 | Demographic and clinical measures for participants in Alzheimer's Disease Neuroimaging Initiative (ADNI3), subdivided by diffusion-weighted MRI (dMRI) protocol.

Protocols			Demographics		Clinical assessments					
					Diagnosis			Cognitive measures[+]		
Name	Total N	Sites	Age (years)	Male	CN	MCI	AD	MMSE*	CDR-sob*	ADAS-cog*
GE54	65	8	76.7 ± 7.3	32 (49.2%)	45	16	4	28.50 ± 3.26	0.78 ± 1.81	11.75 ± 6.81
P33	24	3	78.1 ± 7.1	13 (54.2%)	17	4	3	28.75 ± 2.03	1.31 ± 2.84	13.32 ± 6.76
P36	19	4	75.3 ± 6.6	7 (36.8%)	12	7	0	28.21 ± 2.39	0.76 ± 1.35	12.63 ± 5.12
S31	36	9	72.8 ± 8.6	15 (41.7%)	21	10	5	28.31 ± 2.77	0.79 ± 1.35	11.54 ± 5.25
S55	153	18	75.0 ± 8.4	66 (43.1%)	100	43	10	27.94 ± 3.28	0.95 ± 2.05	11.96 ± 5.65
S127	20	5	75.3 ± 5.4	10 (50.0%)	16	4	0	28.80 ± 1.70	0.33 ± 0.75	10.27 ± 2.83
TOTAL	317	47	75.4 ± 7.9	143 (45.1%)	211	84	22	28.23 ± 3.01	0.87 ± 1.91	11.89 ± 5.78

*We report the average age, Mini-Mental State Examination (MMSE), Clinical Dementia Rating scale sum-of-boxes (CDR-sob), and AD Assessment scale 13 (ADAS-cog) measures, and their standard deviations. *Data not available for all participants: MMSE N = 315; CDR-sob N = 316 and ADAS-cog N = 278. [+]We recognize the limitations of these assessments as proxy measures of specific cognitive abilities (Balsis et al., 2015).*

TABLE 2 | ADNI diffusion MRI acquisition protocols.

	Name	Scanner	Protocol	b_0 volumes	DWI volumes	Total volumes	Time (min)	Total N
ADNI3	GE36	GE	Basic Widebore 25x	4 $b = 0$ s/mm^2	32 $b = 1,000$ s/mm^2	36	9:52	—
	GE54	GE	Basic 25x	6 $b = 0$ s/mm^2	48 $b = 1,000$ s/mm^2	54	7:09	65
	P33	Philips	Basic Widebore	1 $b = 0$ s/mm^2	32 $b = 1,000$ s/mm^2	33	7:32	24
	P36	Philips	Basic Widebore R3	1 $b = 0$ s/mm^2	32 $b = 1,000$ s/mm^2	36	6:54	19
				3 $b = 2$ s/mm^2				
	P54	Philips	Basic R5	1 $b = 0$ s/mm^2	48 $b = 1,000$ s/mm^2	54	8:05	—
				5 $b = 2$ s/mm^2				
	S31	Siemens	Basic VB17	1 $b = 0$ s/mm^2	30 $b = 1,000$ s/mm^2	31	7:02	36
	S55	Siemens	Basic Skyra E11 and Prisma D13	7 $b = 0$ s/mm^2	48 $b = 1,000$ s/mm^2	55	9:18	153
	S127	Siemens	Advanced Prisma VE11C	13 $b = 0$ s/mm^2	48 $b = 1,000$ s/mm^2	61	7:25*	20
ADNI2	G46	GE	Discovery MR750 and MR750w, Signa HDx and HDxt	5 $b = 0$ s/mm^2	41 $b = 1,000$ s/mm^2	46	7:00–10:00	59

*Reflects the time to acquire the full multi-shell protocol (127 volumes), not the single-shell subset.

sufficient sample sizes to be evaluated here. Protocols varied in the number of diffusion weighted imaging (DWI) directions (i.e., angular resolution), and the number of non-diffusion sensitized gradients (b_0 images), which serve as a reference to assess diffusion-related decay of the MR signal. Voxel size across all ADNI3 protocols was 2.0 × 2.0 × 2.0 mm and 2.7 × 2.7 × 2.7 mm in ADNI2. **Table 2** summarizes the different protocols.

There is currently one multi-shell multiband protocol for Siemens Advanced Prisma scanners (S127). As ADNI3 is still in its early stages, GE and Philips protocols for multi-shell acquisition have not yet been finalized, so only 20 multi-shell scans were available for analysis at the time of writing. Here our goal was to evaluate single-shell dMRI indices across protocols, so we used a subsample of the 127 DWI volumes from the S127 multi-shell protocol to include only 13 $b = 0$ and 48 $b = 1,000$ s/mm^2 DWI volumes (removing 6 $b = 500$ s/mm^2 and 60 $b = 2,000$ s/mm^2 volumes).

The Philips Basic Widebore R3 protocol (P36) included three $b = 2$ s/mm^2 volumes and one $b = 0$ s/mm^2, because Philips scanners cannot acquire more than one $b = 0$ s/mm^2. The Philips Basic Widebore (P33) was not a prescribed protocol, but rather acquired from Philips sites with a software version less than 5.0 that could not acquire the $b = 2$ s/mm^2 volumes.

dMRI Preprocessing and Scalar Indices

All DWI were preprocessed using the ADNI2 DTI analysis protocol as in Nir et al. (2013). Briefly, we corrected for head motion and eddy current distortion, removed extra-cerebral tissue, and registered each participant's DWI to the respective T1-weighted brain to correct for echo planar imaging (EPI) distortion. Details of the preprocessing steps may be found here[3]. All DWI and T1-weighted images were visually checked for quality assurance.

Scalar dMRI indices were derived from two reconstruction models: the single-tensor model (DTI; Basser et al., 1994) and the tensor distribution function (TDF; Leow et al., 2009). From the single-tensor model, FA^{DTI}, AxD^{DTI}, MD^{DTI}, and RD^{DTI} scalar maps were generated. In contrast to DTI, the TDF represents

the diffusion profile as a probabilistic mixture of tensors that optimally explain the observed diffusion data, allowing for the reconstruction of multiple underlying fibers per voxel, together with a distribution of weights, from which the TDF-derived form of FA (FA^{TDF}) was calculated (Nir et al., 2017).

White Matter Tract Atlas ROI Summary Measures

ROI measures were generated as reported previously (Nir et al., 2013). Briefly, the FA image from the Johns Hopkins University single-subject Eve atlas (JHU-DTI-SS[4]) was registered to each participant's corrected FA image using an inverse consistent mutual information based registration (Leow et al., 2007); the transformation was then applied to the atlas WM parcellation map (WMPM) ROI labels (Mori et al., 2008) using nearest neighbor interpolation. Mean anisotropy and diffusivity indices were extracted from 24 WM ROIs total (**Table 3**): 22 ROIs averaged bilaterally, the full corpus callosum, and a summary across all ROIs (full WM).

Comparing the ADNI2 and ADNI3 Protocols in Cognitively Normal Participants

Sample Sizes for the ADNI2 and ADNI3 Cognitively Normal Participants

We evaluated the six ADNI3 protocols and the ADNI2 protocol using scans from CN participants only. Of 85 CN participants in ADNI2 with dMRI, 30 rolled over to ADNI3. To avoid duplication, and boost the number of scans available for each protocol, we did not include all these roll-over participants in the ADNI3 group. Twenty-six CN roll-over participants were included in the ADNI3 group. Four CN roll-over participants were scanned with the S55 protocol, and due to the larger sample size already available for that protocol ($N = 156$), we included these four in the ADNI2 group. In total, 59 out of 85 ADNI2 CN participants were included in the ADNI2 group and the remaining 26 were kept in the ADNI3 group for a total of 207 ADNI3 CN participants (see **Supplementary Table S3** for CN demographics by ADNI phase and protocol).

---

[3]https://adni.bitbucket.io/reference/docs/DTIROI/DTI-ADNI_Methods-Thompson-Oct2012.pdf

[4]http://cmrm.med.jhmi.edu/cmrm/atlas/human_data/file/AtlasExplanation2.htm

TABLE 3 | The following 24 regions of interest (ROIs) from the Johns Hopkins University (JHU) atlas (Mori et al., 2008) were analyzed.

CST	Corticospinal tract	SLF	Superior longitudinal fasciculus
CP	Cerebral peduncle	SFO	Superior fronto-occipital fasciculus
ALIC	Anterior limb of internal capsule	IFO	Inferior fronto-occipital fasciculus
PLIC	Posterior limb of internal capsule	SS	Sagittal *stratum*
RLIC	Retrolenticular part of internal capsule	EC	External capsule
PTR	Posterior thalamic radiation	UNC	Uncinate fasciculus
ACR	Anterior *corona radiata*	GCC	*Genu* of corpus callosum
SCR	Superior *corona radiata*	BCC	Body of corpus callosum
PCR	Posterior *corona radiata*	SCC	Splenium of corpus callosum
CGC	Cingulum (cingulate gyrus)	CC	Full corpus callosum
CGH	Cingulum (hippocampal bundle)	TAP	*Tapetum*
Fx/ST	Fornix *(crus) / stria terminalis*	Full WM	Full white matter

Assessing Age Effects

In CN participants, multivariate random-effects linear regressions were used to assess whether dMRI indices from each ADNI protocol individually were associated with age, controlling for sex and age*sex interactions as fixed variables, and acquisition site as a random variable. dMRI indices for the CN group were subsequently pooled across ADNI3 protocols ($N = 207$), or ADNI3 and ADNI2 protocols ($N = 266$) and tested for associations with age using an analogous model, but with protocol and acquisition site as nested random variables (e.g., eight sites used protocol GE54, and three sites used protocol P33, so the acquisition site grouping variable is nested within the protocol grouping variable). We used the false discovery rate (FDR) procedure to correct for multiple comparisons ($q = 0.05$; Benjamini and Hochberg, 1995) across the 24 ROIs assessed for each dMRI index. Regions that survive a more stringent Bonferroni correction at an alpha of 0.05 ($p \leq 0.05/24 = 0.0021$) are also shown in the Supplements.

Effect of Protocol on dMRI Indices

In CN participants, we tested for significant differences in dMRI indices between the seven ADNI protocols using analyses of covariance (ANCOVAs), adjusting for age, sex, and age*sex interactions as fixed variables, and acquisition site as a random variable. For each dMRI index, we used FDR to correct for multiple comparisons across the 24 ROIs assessed. Pairwise tests were performed to directly compare protocols. In total, there were 504 tests per dMRI index: 24 ROIs * 21 pairs of protocol comparisons (protocol 1 vs. 2, protocol 1 vs. 3, etc). As before, we used FDR to account for multiple comparisons.

dMRI Harmonization With ComBat

ComBat uses an empirical Bayes framework to reduce unwanted variation in multi-site data due to differences in acquisition protocol, while preserving the desired biological variation in the data (Fortin et al., 2017). In the CN participants from ADNI2 and ADNI3, we ran ComBat on each of the dMRI indices, including age, sex, age*sex, and information from all 24 ROIs to inform the statistical properties of the protocol effects. Random-effects regressions tested for dMRI microstructural associations with age, covarying for sex and age*sex as fixed variables and site as a random variable; ANCOVAs and pairwise tests of dMRI differences between protocols were repeated for the harmonized ROI data.

Clinical Assessments and Their Relation to Pooled ADNI3 Diffusion Indices

Multivariate random-effects linear regressions were used to test associations between five dMRI indices in each of the 24 WM ROIs and the three cognitive measures (ADAS, MMSE, CDR-sob), and with diagnosis. Due to the limited available sample size of AD participants ($N = 22$), and their uneven distribution across the acquisition protocols tested here, we compared only groups of people with CN and MCI diagnoses. Age, sex, and age*sex interactions were controlled for as fixed effects, and the protocol and acquisition site were modeled as nested random variables. FDR was again used to correct for 24 ROI tests ($q = 0.05$; Benjamini and Hochberg, 1995). Bonferroni corrections ($p \leq 0.05/24 = 0.0021$) are available in the Supplements. Effect sizes for associations were determined using the d-value standardized coefficient (Rosenthal and Rosnow, 1991).

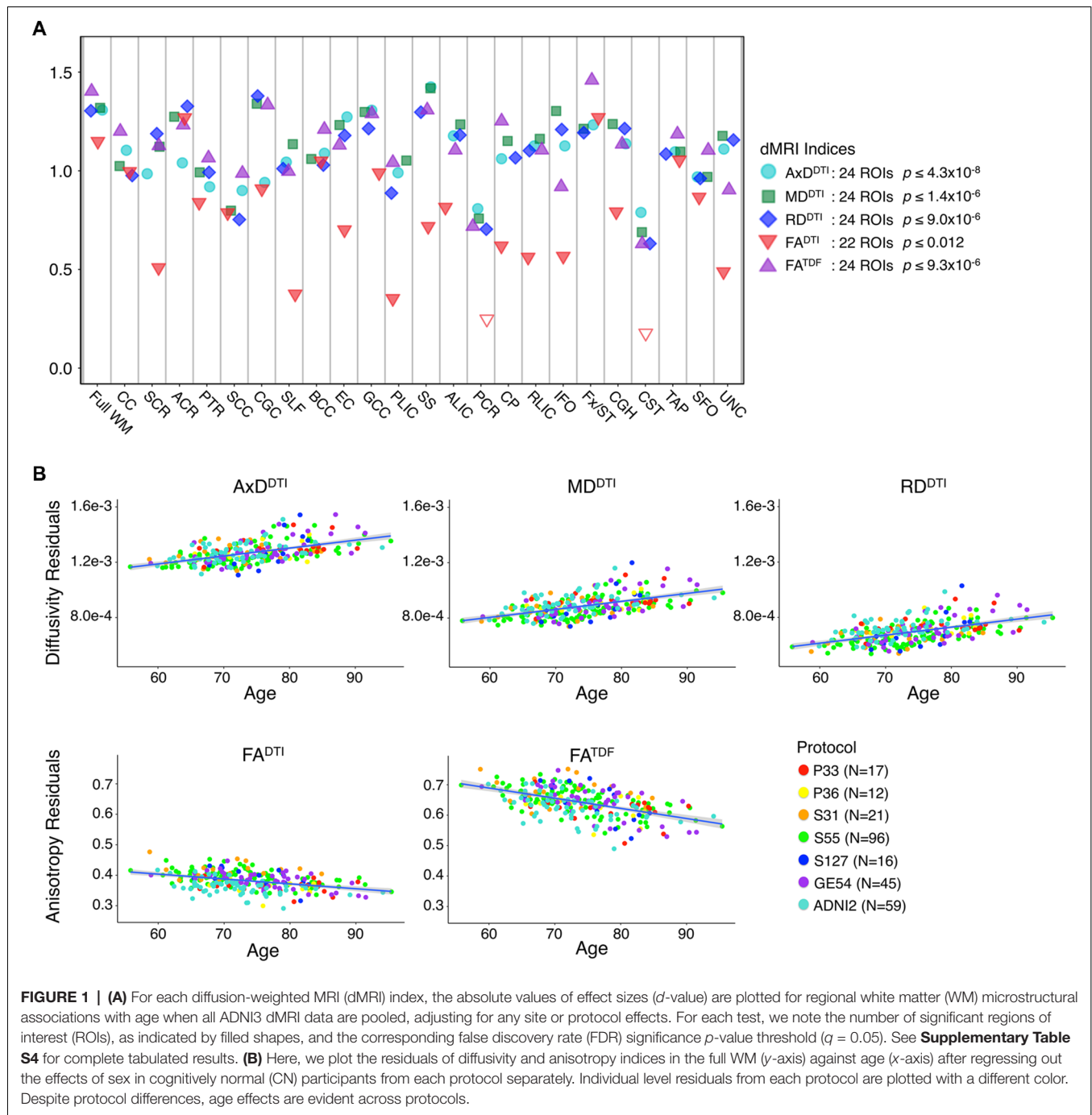$$d = \frac{(2 * Tvalue)}{\sqrt{Degrees\ of\ Freedom}}$$

RESULTS

ADNI2 and ADNI3 Protocols in Cognitively Normal Participants

Age Effects in Cognitively Normal Participants From ADNI2 and ADNI3 Protocols

When data were pooled across ADNI2 and ADNI3, significant associations with age were detected throughout the WM. **Figure 1A** shows effect sizes for ROIs significantly associated with age after FDR multiple comparisons correction (tabulated results and more stringent Bonferroni thresholds are shown in **Supplementary Table S4**). Lower FA^{TDF} and higher diffusivity indices were significantly associated with older age in all 24 ROIs. For FA^{DTI}, 22 ROIs were significantly associated with age. The largest effect size was detected with FA^{TDF} in the fornix *(crus)/stria terminalis* (Fx/ST; $d = -1.459$; $p = 5.07 \times 10^{-21}$). The Fx/ST, *genu* of corpus callosum (GCC) and full WM consistently showed one of the 10 largest effect sizes across dMRI indices.

The mean ages of the CN participants assessed in the two phases of ADNI were significantly different ($p = 0.049$; ADNI2 mean age: 72.4 ± 6.6 years; ADNI3 mean age: 74.5 ± 7.4 years; demographics in **Supplementary Table S3**).

FIGURE 1 | (A) For each diffusion-weighted MRI (dMRI) index, the absolute values of effect sizes (*d*-value) are plotted for regional white matter (WM) microstructural associations with age when all ADNI3 dMRI data are pooled, adjusting for any site or protocol effects. For each test, we note the number of significant regions of interest (ROIs), as indicated by filled shapes, and the corresponding false discovery rate (FDR) significance *p*-value threshold (*q* = 0.05). See **Supplementary Table S4** for complete tabulated results. **(B)** Here, we plot the residuals of diffusivity and anisotropy indices in the full WM (*y*-axis) against age (*x*-axis) after regressing out the effects of sex in cognitively normal (CN) participants from each protocol separately. Individual level residuals from each protocol are plotted with a different color. Despite protocol differences, age effects are evident across protocols.

Pairwise tests comparing the mean age of CN participants scanned with each protocol also showed significant differences between those scanned with S31 and two other protocols: GE54 and S31 (*p* = 0.026); P33 and S31 (*p* = 0.0037). Due to differences in age and sample size between protocols and phases, effect sizes could not be directly compared (Button et al., 2013), but the directions of associations with age were largely consistent for ADNI2 and ADNI3 phases separately, and each ADNI3 protocol (**Figures 1**, **2**). Each ADNI protocol showed directionally consistent associations in more than 89%

of tests (24 ROIs * 5 dMRI indices), except for P36 which was consistent in 81%, but had the smallest sample size (*N* = 12; **Figure 2B**; **Supplementary Tables S5–S11**). FA^{TDF} and all three diffusivity indices were consistent in ≥96% of tests (24 ROIs * 8 protocols/phases), while FA^{DTI} was only consistent in 88% of tests. Most of the associations detected in the unexpected direction for each protocol were driven by FA^{DTI}. None of the associations in the unexpected direction were significant after multiple comparisons correction, and only two had a *p* ≤ 0.05.
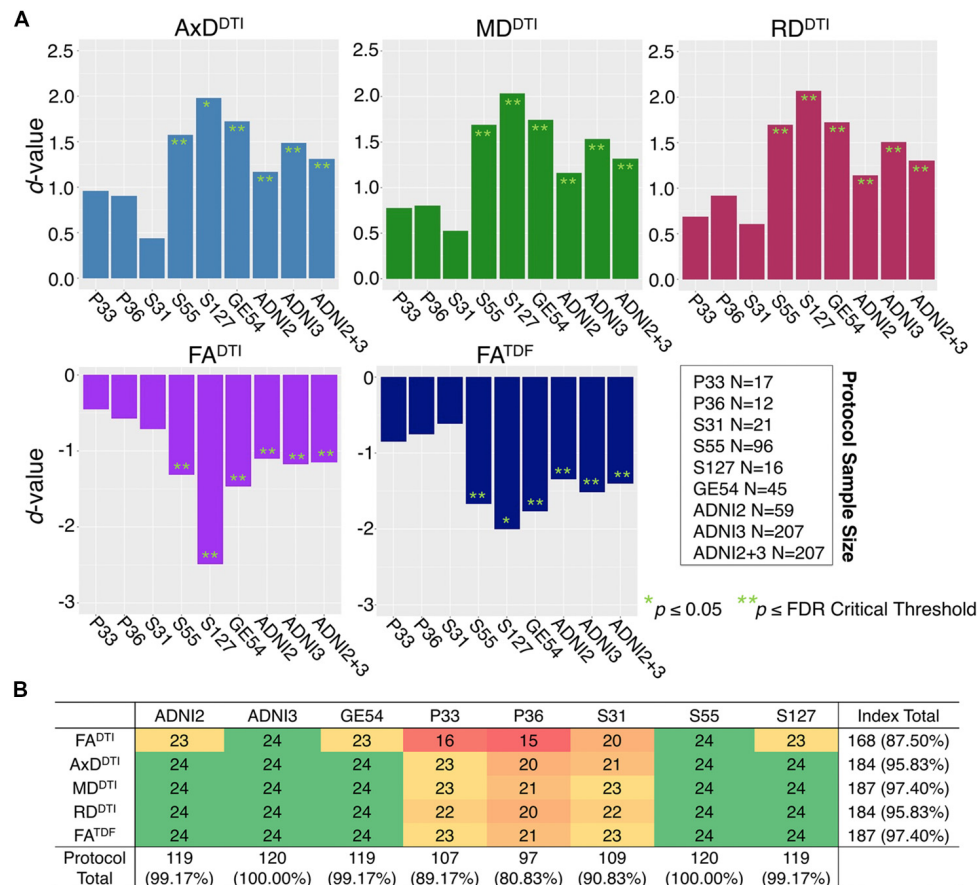
FIGURE 2 | (A) Effect sizes (d-value) for each ADNI protocol and phase show the direction of dMRI associations with age in the full WM are consistent. Due to differences in age and sample size between protocols and phases, effect sizes could not be directly compared. **(B)** For each protocol and phase, the number of ROIs (out of 24), that show the expected association direction, regardless of significance, are reported for each dMRI index; consistent associations were detected across tests, except for protocol P36 which had the smallest sample size, and FA^{DTI}, which showed the smallest effect sizes and fewest significant associations across protocols when pooled.

Figure 2 shows consistent associations in the full WM by protocol. As demographic and sample size variability between protocols affect detected effect sizes, we also evaluated full WM dMRI associations with age in an age- and sex-matched subset of 12 participants from each protocol (total $N = 84$; demographics in **Supplementary Table S3**). A comparison of the effect sizes between protocols suggests that the protocols with greatest total number of diffusion-weighted ($b = 1,000$ s/mm^2) and non-diffusion sensitized (b_0) gradients may detect larger effects (S127 followed by S55; **Supplementary Figure S1**).

Effect of Protocol on dMRI Indices From Cognitively Normal Controls

The influence of dMRI acquisition protocol on mean values of the diffusion indices is evident in boxplots of dMRI indices in the full WM for each protocol (**Figure 3**). When modeling the mean full WM values for each diffusion index, the residuals of the statistical model become closer to 0 after fitting the effect of protocol and site (nested as a random variable with age, sex,

and age*sex interactions as fixed effects) than when we plot the residuals of just age, sex, and age*sex interactions (**Figure 3**).

ANCOVAs and pairwise tests for each ROI suggest there are significant differences between protocols for all 5 dMRI indices across most ROIs (**Figure 4**). ANCOVAs revealed significant protocol differences for 22 ROIs for FA^{DTI} and FA^{TDF}, with the highest overall effect size detected in the anterior limb of the internal capsule (ALIC) and the external capsule (EC) for FA^{DTI} (ALIC: $d = 0.648$; EC: $d = 0.652$). AxD^{DTI} had the smallest effect size, overall, in the splenium of the corpus callosum (SCC; $d = 0.106$), and only 13 ROIs showed significant AxD^{DTI} differences between protocols.

In pairwise analyses, AxD^{DTI} was the most stable index across protocols, as significant protocol differences were detected in only 20.6% of pairwise tests (24 ROIs * 21 pairwise tests), compared to FA^{DTI}, the most variable index, which showed significant protocol differences in 81.9% of tests (**Figure 4B**). ADNI2 was the most divergent protocol across dMRI indices, showing differences in 36.3% of tests.
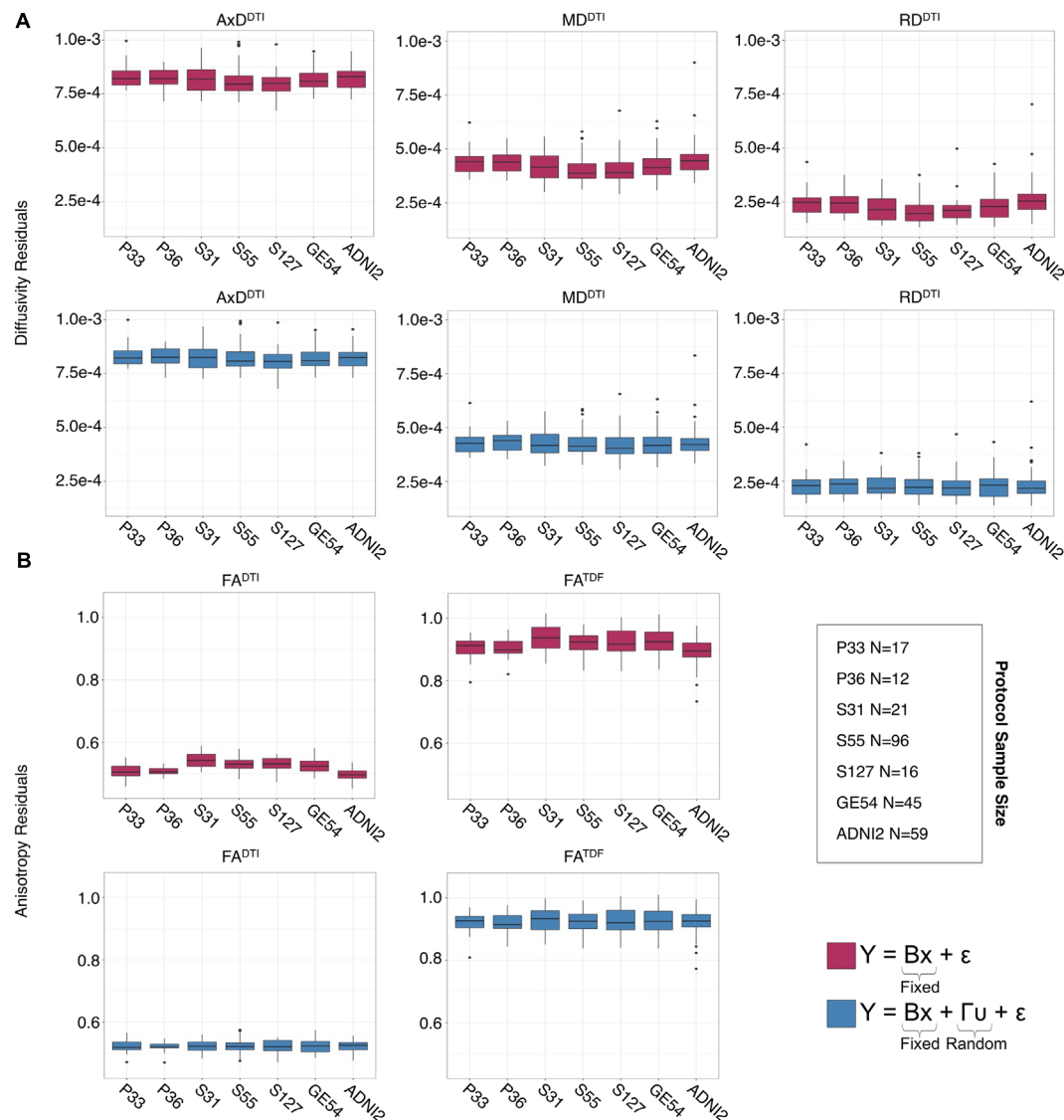
FIGURE 3 | Full WM mean **(A)** AxD^{DTI}, MD^{DTI}, and RD^{DTI} and **(B)** FA^{DTI} and FA^{TDF} residuals for each protocol, after fitting effects of age, sex, and age*sex interactions, are plotted here in the top rows (*red*). Protocol has an effect on anisotropy and diffusivity measures. The lower panels (*blue*) show residuals after additionally fitting protocol and site as nested random-effects, after which the residuals across protocols are closer to 0.

Diffusion MRI Harmonization With ComBat

After using ComBat to harmonize dMRI indices across protocols, ANCOVAs revealed that significant protocol differences in dMRI indices were all but eliminated across ROIs (**Supplementary Figure S2A**); significant protocol differences were detected only in the CST, for each of the dMRI indices. The number of pairwise tests for which each protocol showed significant differences in dMRI indices decreased by 93.8% with ComBat (**Supplementary Figure S2B**).

After harmonization, we still detected significant associations between age and dMRI indices from ADNI2 and ADNI3 pooled in the same number of ROIs (**Supplementary Table S12**). ComBat correction did not significantly change effect sizes, while correcting for effects of protocol (**Supplementary Figure S3**). In

Figure 5 we show effect sizes before and after harmonization with ComBat in the full WM, Fx/ST, and GCC, the three ROIs that consistently showed one of the 10 largest effect sizes for associations with age across all five diffusion indices (for changes by protocol see **Supplementary Figures S4–S6**). As harmonization with ComBat did not improve or change results found with random-effect linear regressions, we proceeded to test clinical associations without applying the ComBat transformation.

Cognitive Measure Associations With Pooled ADNI3 dMRI Indices

Pooling data across ADNI3, we detected significant associations between all three cognitive measures and regional dMRI
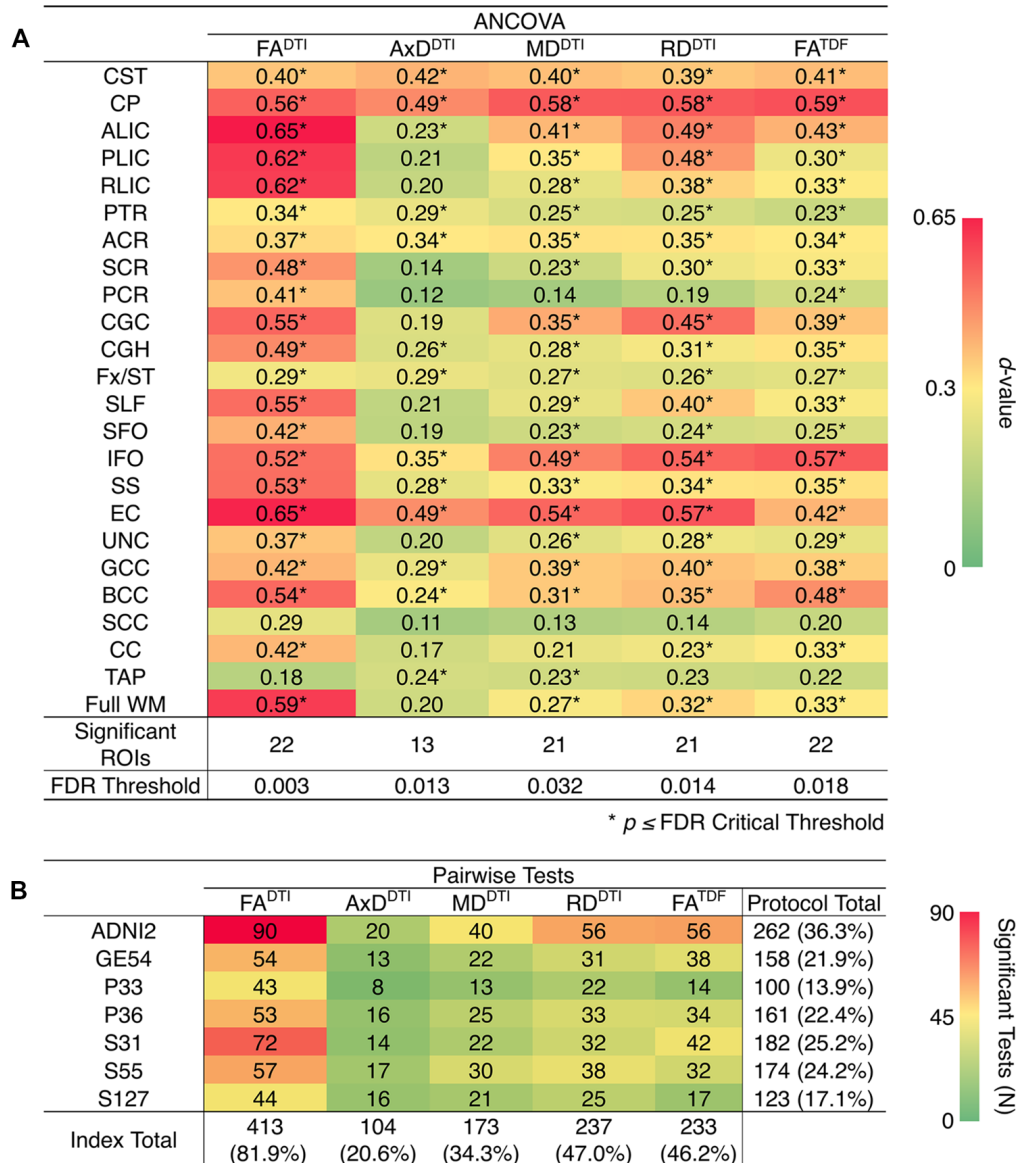
FIGURE 4 | (A) d-values from the analyses of covariance (ANCOVAs) assessing differences in dMRI indices between protocols, for each of the 24 ROIs; FA[DTI] showed the greatest significant differences (largest d-values; dark red) between protocols and AxD[DTI] the fewest (dark green). **(B)** We report the number of times each protocol and each dMRI index showed significant differences in pairwise tests between protocols (out of 504 tests per index and 720 tests per protocol); AxD[DTI] was the most stable dMRI index across protocols, while FA[DTI] was the least stable.

indices throughout the WM. Greater cognitive impairment was associated with lower anisotropy and higher diffusivity. **Figures 6A–C** shows effect sizes for ROIs significantly associated with each cognitive measure after FDR multiple comparisons correction (for tabulated results and more stringent Bonferroni corrections, please see **Supplementary Tables S13–S15**). Across tests (5 dMRI indices * 3 cognitive measures), the hippocampal-cingulum (CGH), fornix *(crus)/stria terminalis* region (Fx/ST), and the full WM consistently showed one of the 10 largest effect sizes (see **Supplementary Figures S7–S9** for associations with indices in the CGH, Fx/ST, and full WM, by protocol).

In 14 of 15 tests, the CGH consistently showed one of the top two largest effect sizes (CGH FA[DTI] association with CDR-sob was the third largest), along with the uncinate fasciculus (UNC), which was top two in 12 of 15 tests (while significant, cognitive associations with UNC FA[DTI] never showed one of the largest effect sizes).

FA[DTI] showed significant associations in the fewest ROIs: 55 out of 72 tests (76.4%; 24 ROIs * 3 cognitive measures) were significant. FA[TDF] showed more widespread associations with cognitive measures throughout WM ROIs: 69 out of 72 tests (94.4%) were significant. Effect sizes were consistently
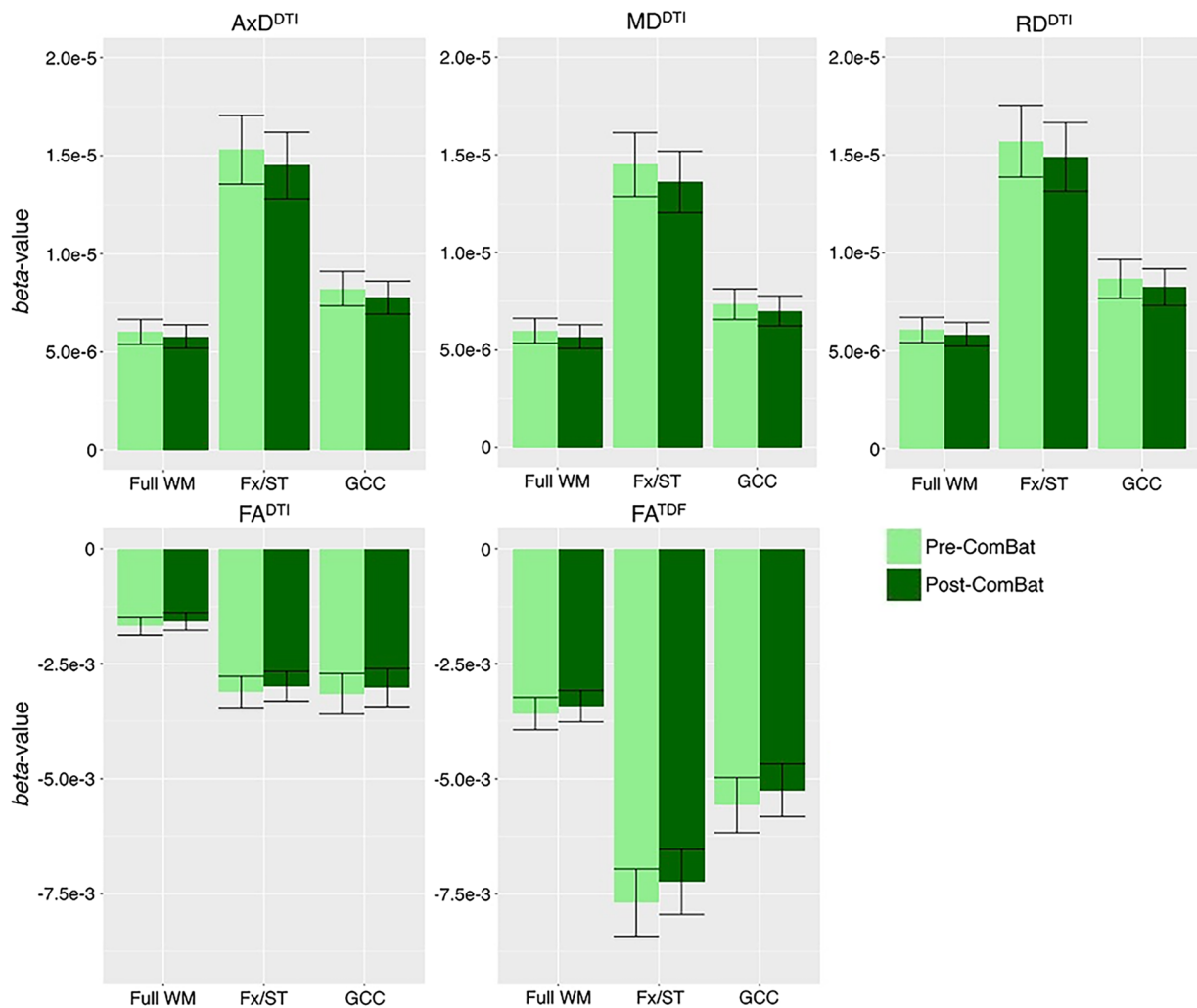
FIGURE 5 | *Beta*-values and error bars representing standard error from the association between each diffusion index and age in CN participants, before and after ComBat harmonization. We show the three ROIs that consistently showed one of the 10 largest effect sizes for associations with age across all five diffusion indices (see **Supplementary Figure S3** for all ROIs). Compared to pre-ComBat analyses, effect sizes are marginally different across indices, but still within the standard error.

lower for FA^{DTI} than for the other dMRI indices, across all three cognitive measures; the largest FA^{DTI} effect sizes were most consistently found in the Fx/ST, followed by the CGH or the GCC. The strongest FA^{DTI} association overall was in the Fx/ST with CDR-sob ($d = -0.681$, $p = 7.01 \times 10^{-8}$). Compared to FA^{DTI}, FA^{TDF} showed larger effect sizes; across cognitive tests, the strongest FA^{TDF} associations were detected in the UNC with CDR-sob ($d = -1.244$; $p = 1.39 \times 10^{-20}$), followed by the CGH ($d = -1.213$; $p = 8.86 \times 10^{-20}$). CDR-sob effect sizes for FA^{DTI} and FA^{TDF} in the CGH, UNC, Fx/ST, and full WM are depicted by protocol in **Supplementary Figure S10**, revealing consistently larger effect sizes for FA^{TDF} across protocols.

Cognitive associations with all of the diffusivity indices were widespread: significant associations were detected in 207 out of 216 tests (95.8%; 24 ROIs * 3 cognitive measures * 3 diffusivity indices). Regional measures of AxD^{DTI} consistently showed the

largest effect sizes across all cognitive measures (CDR-sob and the UNC: $d = 1.344$, $p = 3.13 \times 10^{-23}$; MMSE and the CGH: $d = -1.178$, $p = 7.87 \times 10^{-19}$; ADAS-cog and the UNC: $d = 1.048$, $p = 1.09 \times 10^{-13}$).

Of the three cognitive measures, CDR-sob associations showed the largest effect sizes across dMRI indices (in the UNC followed by the CGH for all indices except FA^{DTI}); the largest effect sizes across all tests were detected with AxD^{DTI} (UNC: $d = 1.344$) and MD^{DTI} (UNC: $d = 1.342$, $p = 3.47 \times 10^{-23}$). **Figure 7** shows the distribution of the effect sizes for CDR-sob throughout the brain. Temporal lobe regions (UNC, CGH, IFO, SS) frequently showed greatest effect sizes (for ADAS-cog and MMSE figures, see **Supplementary Figures S11, S12**). Effect size was not correlated with ROI size (**Supplementary Figure S13**), consistent with prior studies of other disorders (Kelly et al., 2018).

FIGURE 6 | For each dMRI index, the absolute values of effect sizes (*d*-value) are plotted for regional WM microstructural associations with clinical measures. Lower anisotropy and higher diffusivity were significantly associated with **(A)** higher CDR-sob, **(B)** lower MMSE, **(C)** higher ADAS-cog, and **(D)** an MCI diagnosis, when all ADNI3 dMRI data are pooled, adjusting for any site or protocol effects. For each test, we note the number of significant ROIs, as indicated by filled shapes, and the corresponding FDR significance *p*-value threshold (*q* = 0.05). See **Supplementary Tables S13–S16** for complete tabulated results.

CN vs. MCI Diagnosis Associations With Pooled ADNI3 dMRI Indices

For each diffusion index, **Figure 6D** shows the significant regional effect sizes for differences between CN and MCI participants. Widespread diffusivity differences were detected, with significantly higher diffusivity in MCI participants in 21 out of 24 ROIs (**Supplementary Table S16** and **Supplementary Figure S14**). Only three regions showed significantly lower FA^{DTI} in MCI participants—Fx/ST ($d = -0.460$; $p = 3.89 \times 10^{-4}$), CGH ($d = -0.410$; $p = 1.53 \times 10^{-3}$), and the posterior thalamic radiation (PTR; $d = 0.367$; $p = 4.55 \times 10^{-3}$). On the other hand, FA^{TDF} was significant in 20 out of 24 ROIs, similar to diffusivity indices. FA^{TDF} and diffusivity indices in the CGH showed the largest effect sizes overall (AxD^{DTI} $d = 0.681$; $p = 2.26 \times 10^{-7}$, MD^{DTI} $d = 0.700$; $p = 1.15 \times 10^{-7}$; RD^{DTI} $d = 0.679$; $p = 2.41 \times 10^{-7}$; FA^{TDF} $d = -0.622$; $p = 2.00 \times 10^{-6}$).

For all three cognitive measures, and in the comparison between CN and MCI participants, the CGH and Fx/ST were the only regions that survived multiple comparisons correction

across all dMRI indices. The Fx/ST always had the largest effect size in FA^{DTI} tests. The UNC showed either the first or second largest effect size (alternating with CGH) across diffusivity indices and FA^{TDF} tests, but was significant only for cognitive measure associations with FA^{DTI} (i.e., three of four clinical tests).

DISCUSSION

This study has three main findings: (1) when data were pooled from the six available diffusion MRI protocols used in ADNI3, anisotropy and diffusivity indices showed robust associations with MCI diagnosis, and with three common cognitive measures: MMSE, ADAS-cog, and CDR-sob; (2) when using a higher-order diffusion model, the derived measure of anisotropy (FA^{TDF}) showed stronger and more widespread associations with clinical impairment than the standard DTI anisotropy measure (FA^{DTI}); and (3) despite significant differences in protocols, for each dMRI index, we were able to detect consistent associations with
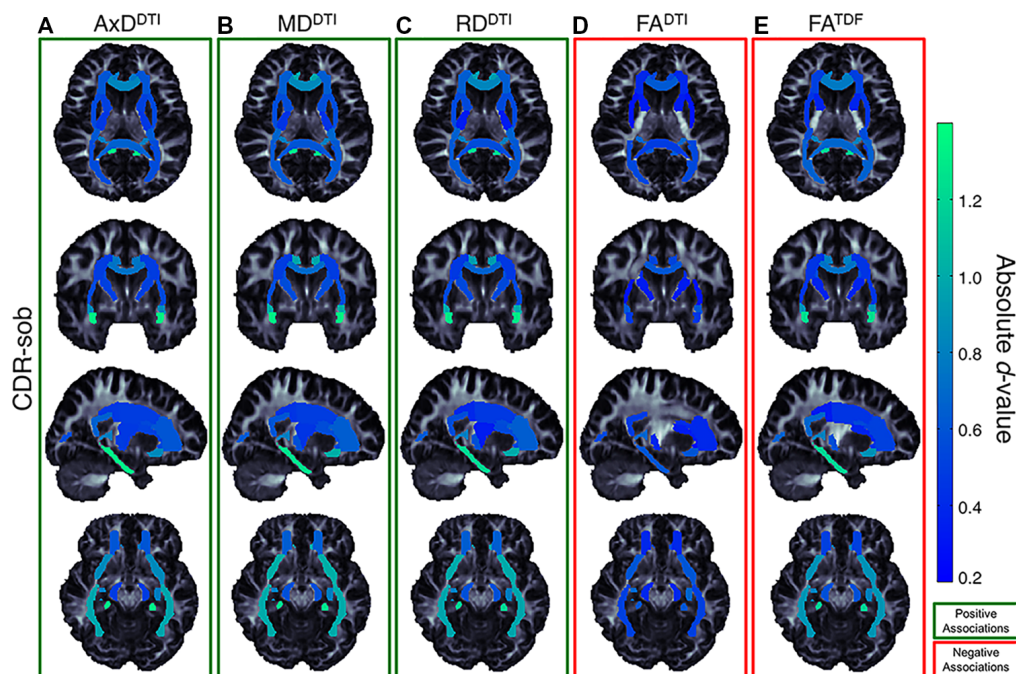
FIGURE 7 | Effect size (absolute *d*-value) maps of WM regions that show significant associations with CDR-sob—the cognitive measure with the largest effect sizes—reveal widespread associations throughout the WM, with particularly strong associations in the temporal lobes (SS, IFO, UNC, and CGH; *light green* regions show the largest effect sizes). As expected, positive associations were detected between CDR-sob and **(A)** AxD^{DTI} (FDR critical threshold $p = 1.78 \times 10^{-4}$; **B**) MD^{DTI} (FDR critical threshold $p = 3.64 \times 10^{-4}$) and **(C)** RD^{DTI} (FDR critical threshold $p = 6.92 \times 10^{-3}$); higher diffusivity was associated with greater cognitive impairment. Lower **(D)** FA^{DTI} (FDR critical threshold $p = 0.025$) and **(E)** FA^{TDF} (FDR critical threshold $p = 7.73 \times 10^{-3}$) were also associated with greater impairment, but FA^{DTI} associations were detected in fewer regions with weaker effect sizes compared to FA^{TDF}.

clinical measures in ADNI3 participants, and age in ADNI2 and ADNI3 CN participants.

Accumulation of amyloid plaques and neurofibrillary tangles (NFTs) in the brain (Braak and Braak, 1991, 1996; Frank et al., 2003; Shaw et al., 2007) can directly impact WM (Lee et al., 2004; Roth et al., 2005), promoting myelin degeneration and axonal loss (Braak and Braak, 1996; Kneynsberg et al., 2017). While many factors drive anisotropy and diffusivity measures from DTI, higher anisotropy values may indicate, in part, more coherent intact axons, while lower anisotropy and higher diffusivity may reflect factors such as axonal injury and demyelination, among other factors (Beaulieu, 2002; Song et al., 2003, 2005; Harsan et al., 2006; Le Bihan and Johansen-Berg, 2012; Kantarci et al., 2017; Moore et al., 2018). In this article, lower anisotropy values and higher diffusivity values were correlated with clinical impairment most strongly in the hippocampal-cingulum and uncinate. Along with the full WM, reflecting global WM effects, the largest effect sizes were most frequently detected in the hippocampal-cingulum and fornix *(crus)/stria terminalis*, WM bundles connecting hippocampal and parahippocampal regions to the rest of the brain, consistent with patterns of AD pathology. The histopathological validity of these findings has been supported, specifically in a recent study that compared NFT stages in ante-mortem MRI and post-mortem tissue; elevated MD^{DTI} and lower FA^{DTI} significantly correlated with higher postmortem NFT stage, particularly in the

crus of the fornix, the ventral cingulum tracts, the precuneus, and entorhinal WM (Kantarci et al., 2017).

The participants recruited for ADNI3 tend to be younger and healthier, on average, than those in ADNI2, as they were recruited with the intention of studying the transition from CN to AD (Jack et al., 2015). With few AD patients enrolled so far in ADNI3, the primary focus of this article was to assess three cognitive assessments (ADAS-cog, CDR-sob, and MMSE), and to compare CN to MCI participants. MCI is now the focus of intense research; it is essential to find ways to clinically categorize the transitional stages between normal aging and AD to evaluate targeted treatments, as pathophysiological mechanisms may differ or change throughout the course of AD (Mueller et al., 2005). As in our prior analysis of ADNI2 (Nir et al., 2013), FA^{DTI} was the least sensitive DTI measure. In ADNI3, AxD^{DTI} and MD^{DTI} showed the largest effect sizes. Lower FA^{DTI} and higher MD^{DTI} are most frequently reported in studies of AD (Kavcic et al., 2008; Clerx et al., 2012; Nir et al., 2013; Maggipinto et al., 2017; Mayo et al., 2017), but AxD^{DTI} may be more sensitive to unspecific microscopic cellular loss earlier in the disease (O'Dwyer et al., 2011), perhaps making it more sensitive in the healthier participants of the ADNI3 dataset. Similarly, in ADNI2, AxD^{DTI} was the most sensitive to differences between CN and MCI diagnosis (Nir et al., 2013).

Among the three cognitive assessments, CDR-sob showed the strongest correlations with dMRI indices, in line with prior

ADNI brain imaging studies (Hua et al., 2009; Nir et al., 2013). The largest of these effects were found in temporal WM tracts including the hippocampal-cingulum, uncinate, sagittal stratum, and inferior fronto-occipital fasciculus. These are all regions that show early degenerative changes in MCI and AD (Mielke et al., 2009; Nir et al., 2013; Maggipinto et al., 2017; Powell et al., 2018). While associations with clinical impairment were detected throughout the WM, the region that most frequently showed the lowest effect sizes and was significant in only 3 of the 20 clinical tests, was the corticospinal tract (CST). However, the CST ROI from the JHU WMPM atlas is limited to a small region in the inferior portion of the brain and has been shown to be the least reliable and reproducible ROI (Jahanshad et al., 2013; Acheson et al., 2017), suggesting alternate approaches, such as tractography-based evaluations (Jin et al., 2017), or the use of the probabilistic JHU atlas (Hua et al., 2008), may be more appropriate for studying the CST. Our analysis focused on WM microstructure, but future work assessing tract geometry and properties of anatomical brain networks using tractography may reveal more detailed information. The validation and harmonization of tractography methods and derived network metrics is a vast field of research with active ongoing work (Maier-Hein et al., 2017).

DTI is widely recognized as a useful tool for studying neurodegenerative disorders such as AD (Oishi et al., 2011; Müller and Kassubek, 2013; Abhinav et al., 2014; Acosta-Cabronero and Nestor, 2014; Maggipinto et al., 2017). However, at the spatial resolutions now used, a single voxel typically captures partial volumes of different tissue compartments–e.g., the intra- and extra-cellular compartments, the vascular compartment, the CSF and myelin; each affects water diffusion and the MR signal. The DTI model cannot differentiate these components or even crossing fibers (Tuch et al., 2002; Jbabdi et al., 2010), which are estimated to occur in up to 90% of WM voxels at the typical dMRI resolution (Descoteaux et al., 2009; Jeurissen et al., 2013). In healthy tissue with crossing fibers, the DTI model may show low FA. FA^{DTI} may paradoxically appear to increase in regions where crossing fibers deteriorate in neurodegenerative diseases such as AD (Douaud et al., 2011). FA^{TDF} addresses this limitation even in low angular resolution data (Nir et al., 2017). Here, compared to FA^{DTI}, FA^{TDF} showed more widespread associations with cognitive measures and diagnosis throughout WM ROIs: FA^{TDF} was significant in 89 of the 96 tests (92.7%; 24 ROIs * 4 clinical tests), while FA^{DTI} was only significant in 58 (60.4%). The greatest difference was seen for diagnostic associations (CN vs. MCI): FA^{TDF} was significant in 20 out of 24 ROIs while FA^{DTI} was only significant in three. FA^{TDF} also showed stronger effect sizes across the protocols, suggesting that tensor limitations have likely confounded previous diffusion studies of cognitive decline that have found little or no effects with FA (Acosta-Cabronero et al., 2010). Recently proposed biophysical models of brain tissue may help to relate diffusion signals directly to underlying microstructure and different tissue compartments (Harms et al., 2017). We may be able to further disentangle questions of orientation coherence (dispersing and "kissing" fibers), fiber diameter, fiber density, membrane permeability, and

myelination, which all influence classic anisotropy and diffusivity measures derived from DTI. Several AD studies have already used multi-shell protocols to compute diffusion indices from models that do not assume mono-exponential decay, such as diffusion kurtosis imaging (DKI; Jensen et al., 2005; Chen et al., 2017; Cheng et al., 2018; Wang M.-L. et al., 2018), and multi-compartment models such as neurite orientation dispersion and density imaging (NODDI; Zhang et al., 2012; Colgan et al., 2016; Slattery et al., 2017; Parker et al., 2018). To date, approximately 20 participants in ADNI have been scanned with multi-shell diffusion protocols; in a future report, we will relate multi-shell measures to those examined here.

Large-scale, multi-site neuroimaging studies can increase the power of statistical analyses and establish greater confidence and generalizability for findings. Most multi-site neuroimaging studies are susceptible to variability across sites. Variability in dMRI studies is due in part to heterogeneity in acquisition protocols, scanning parameters, and scanner manufacturers (Zhu et al., 2009, 2011, 2018). Anisotropy and diffusivity maps are affected by angular and spatial resolution (Alexander et al., 2001; Kim et al., 2006; Zhan et al., 2010), the number of DWI directions (Giannelli et al., 2009), and the number of acquired b-values (Correia et al., 2009). All five dMRI indices were significantly different between protocols; AxD^{DTI} was the most stable index, while FA^{DTI} was the least stable, reflective of their performance in detecting associations with cognitive measures. ADNI2 was the most divergent protocol across dMRI indices, perhaps due to the larger voxel size in ADNI2 (2.7 mm^3 vs. 2.0 mm^3 isotropic voxels used in ADNI3). This is consistent with the notion that DTI measures vary with voxel size due to partial voluming (Zhan et al., 2013). Despite differences in protocols, the directions of associations were consistent across protocols.

ADNI3 extends dMRI acquisitions across scanner manufacturers and platforms to maximize the number of participants scanned with dMRI; this makes it necessary to account for site-related heterogeneities and confounds in analytical models where data are pooled. Multi-site dMRI studies are becoming increasingly common, and new data harmonization methods to adjust for site and acquisition protocol are being developed and tested. A thorough investigation of dMRI harmonization methods is now possible with ADNI3, one of the few publicly available multi-site datasets acquired with multiple protocols. As regional dMRI measures are available for download as part of the ADNI database, we highlight two ways that the data may be pooled across sites: (1) performing statistical analyses with nested random-effects models to account for site and acquisition protocol differences; and (2) harmonizing the derived regional measures before aggregating the data across sites. In a preliminary analysis, we showed that one harmonization method performed on these regional measures, ComBat, reduced cross-site differences in dMRI indices, while preserving biological relationships with age in CN controls. The only region where differences remained after ComBat, was the CST, the ROI with the weakest associations with clinical measures, and previously identified as least reliable (Acheson et al., 2017). In Fortin et al. (2017), compared to other harmonization methods, ComBat increased the number

of voxels where significant associations between age and FA^{DTI} or MD^{DTI} were detected. Here, the number of significant ROIs and the magnitude of effect sizes were comparable for ComBat and nested random-effects model approaches. This discrepancy between our findings and that of Fortin et al. (2017), may be due to several differences between studies: (1) ADNI3 includes more sites and protocols; (2) in contrast to the number of voxels, the number of ROIs is far less than the number of participants; and (3) the age effects in the elderly populations tested here are stronger than the effects tested in adolescents in Fortin et al. (2017). When effects are more readily detected, one harmonization approach may not be more advantageous than others. In addition to exploring additional harmonization techniques, future work should evaluate voxel-wise ComBat approaches and the effects of harmonization beyond CN participants (i.e., across the entire ADNI cohort).

In addition to ComBat, a number of harmonization approaches have recently been proposed at various stages of analysis (Tax et al., 2018; Zhu et al., 2018). Site differences can be accounted for at the time of overall group inference, such as with the random-effects regression level correction used here, or by using a meta-analysis approach in lieu of pooling data (Thompson et al., 2014). The data may also be transformed prior to multi-site group-level statistics. Some methods, such as ComBat and RAVEL, use the distribution of derived features, such as diffusivity and anisotropy measures (Fortin et al., 2016, 2017). Alternatively, several proposed methods use information from the raw image to adjust for acquisition variability (Zhu et al., 2018). For example, Kochunov et al. (2018) calculated the signal to noise ratio for each protocol and include it in their regression models. Mirzaalian et al. (2018) use voxel-wise spherical harmonic residual networks to derive local correction parameters. Finding the best method to harmonize dMRI data is an active topic at "hackathons" and technical challenges; in 2017 and 2018, the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) hosted a computational diffusion MRI challenge to explore approaches for data harmonization. With so many available approaches, the preliminary random-effects regression and ComBat results from this article serve as a first step towards future work establishing robust approaches for combining data in ADNI3 and other multi-site studies.

The current study is limited in that the sample sizes and sample demographics available for each protocol vary, complicating direct comparison of the protocols (Button et al., 2013). A matched comparison might be possible if a group of participants or a phantom were scanned using every protocol. Even so, separating protocol differences from differences in scanner manufacturer is difficult. We also could not directly compare all diagnostic groups in ADNI3, as few participants with AD were scanned.

A more complete picture of brain changes in aging and AD would include imaging metrics from other modalities, such as perfusion imaging, resting state functional MRI (Wang et al., 2017), and radiotracer methods such as FDG-PET (Popuri et al., 2018), or amyloid- and tau-sensitive PET (Grothe et al., 2017; Phillips et al., 2018). Genetic and other "omics"

data could be analyzed as well, and may help to predict diagnostic classification and brain aging, when combined with other neuroimaging markers (Ding et al., 2018; Kauppi et al., 2018). While these data are all being collected as part of ADNI3 and other studies of brain aging, our focus here was on the variety of available dMRI measures, calculated using different protocols. With this in mind, the optimal dMRI indices to include in a multimodal study may be those that contribute the greatest independent information beyond that available from anatomical MRI and other standard imaging modalities. Multivariate methods—such as machine learning (Zhou et al., 2017; Wang X. et al., 2018) and even deep learning (Liu et al., 2017)—may also help to extract and capitalize on features that predict clinical decline beyond those studied here.

In addition to providing a roadmap for the new ADNI3 dMRI data, these preliminary analyses show that despite differences in the updated dMRI protocols, diffusion indices can be pooled to detect WM microstructural differences associated with aging and AD.

AUTHOR CONTRIBUTIONS

RR, MB, BB, CJ, PT and MW designed the ADNI3 diffusion MRI study. ST, AZ-P and TN performed the image analysis. TN, AZ-P, NJ and PT conceived and designed the image analysis study. TN, AZ-P, ST, NJ and PT drafted the manuscript. All authors contributed to interpreting the results and critically revised the manuscript for intellectual content.

FUNDING

the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Samples from the National Centralized Repository for AD and Related Dementias (NCRAD), which receives government support under a cooperative agreement grant (U24 AG21886) awarded by the NIA, were used in this study.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2019.00002/full#supplementary-material

REFERENCES

Abhinav, K., Yeh, F.-C., Pathak, S., Suski, V., Lacomis, D., Friedlander, R. M., et al. (2014). Advanced diffusion MRI fiber tracking in neurosurgical and neurodegenerative disorders and neuroanatomical studies: a review. *Biochim. Biophys. Acta* 1842, 2286–2297. doi: 10.1016/j.bbadis.2014.08.002

Acheson, A., Wijtenburg, S. A., Rowland, L. M., Winkler, A., Mathias, C. W., Hong, L. E., et al. (2017). Reproducibility of tract-based white matter microstructural measures using the ENIGMA-DTI protocol. *Brain Behav.* 7:e00615. doi: 10.1002/brb3.615

Acosta-Cabronero, J., and Nestor, P. J. (2014). Diffusion tensor imaging in Alzheimer's disease: insights into the limbic-diencephalic network and methodological considerations. *Front. Aging Neurosci.* 6:266. doi: 10.3389/fnagi.2014.00266

Acosta-Cabronero, J., Williams, G. B., Pengas, G., and Nestor, P. J. (2010). Absolute diffusivities define the landscape of white matter degeneration in Alzheimer's disease. *Brain* 133, 529–539. doi: 10.1093/brain/awp257

Alexander, A. L., Hasan, K. M., Lazar, M., Tsuruda, J. S., and Parker, D. L. (2001). Analysis of partial volume effects in diffusion-tensor MRI. *Magn. Reson. Med.* 45, 770–780. doi: 10.1002/mrm.1105

Alzheimer's Association. (2018). 2018 Alzheimer's disease facts and figures. *Alzheimers Dement.* 14, 367–429. doi: 10.1016/j.jalz.2018.02.001

Balsis, S., Benge, J. F., Lowe, D. A., Geraci, L., and Doody, R. S. (2015). How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? *Clin. Neuropsychol.* 29, 1002–1009. doi: 10.1080/13854046.2015.1119312

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267. doi: 10.1016/s0006-3495(94)80775-1

Beaulieu, C. (2002). The basis of anisotropic water diffusion in the nervous system—a technical review. *NMR Biomed.* 15, 435–455. doi: 10.1002/nbm.782

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Berg, L. (1988). Clinical dementia rating (CDR). *Psychopharmacol. Bull.* 24, 637–639.

Braak, H., and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol.* 82, 239–259. doi: 10.1007/bf00308809

Braak, H., and Braak, E. (1996). Evolution of the neuropathology of Alzheimer's disease. *Acta Neurol. Scand. Suppl.* 94, 3–12. doi: 10.1111/j.1600-0404.1996.tb05866.x

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376. doi: 10.1038/nrn3475

Cercignani, M., Bammer, R., Sormani, M. P., Fazekas, F., and Filippi, M. (2003). Inter-sequence and inter-imaging unit variability of diffusion tensor MR imaging histogram-derived metrics of the brain in healthy volunteers. *Am. J. Neuroradiol.* 24, 638–643. Available online at: http://www.ajnr.org/content/24/4/638.long

Chen, Y., Sha, M., Zhao, X., Ma, J., Ni, H., Gao, W., et al. (2017). Automated detection of pathologic white matter alterations in Alzheimer's disease using

combined diffusivity and kurtosis method. *Psychiatry Res. Neuroimaging* 264, 35–45. doi: 10.1016/j.pscychresns.2017.04.004

Cheng, J.-X., Zhang, H.-Y., Peng, Z.-K., Xu, Y., Tang, H., Wu, J.-T., et al. (2018). Divergent topological networks in Alzheimer's disease: a diffusion kurtosis imaging analysis. *Transl. Neurodegener.* 7:10. doi: 10.1186/s40035-018-0115-y

Clerx, L., Visser, P. J., Verhey, F., and Aalten, P. (2012). New MRI markers for Alzheimer's disease: a meta-analysis of diffusion tensor imaging and a comparison with Medial temporal lobe measurements. *J. Alzheimers Dis.* 29, 405–429. doi: 10.3233/JAD-2011-110797

Colgan, N., Siow, B., O'Callaghan, J. M., Harrison, I. F., Wells, J. A., Holmes, H. E., et al. (2016). Application of neurite orientation dispersion and density imaging (NODDI) to a tau pathology model of Alzheimer's disease. *Neuroimage* 125, 739–744. doi: 10.1016/j.neuroimage.2015.10.043

Correia, M. M., Carpenter, T. A., and Williams, G. B. (2009). Looking for the optimal DTI acquisition scheme given a maximum scan time: are more b-values a waste of time? *Magn. Reson. Imaging* 27, 163–175. doi: 10.1016/j.mri.2008.06.011

Daianu, M., Dennis, E. L., Jahanshad, N., Nir, T. M., Toga, A. W., Jack, C. R., et al. (2013a). "Alzheimer's disease disrupts rich club organization in brain connectivity networks," in *Proceedings of the 2013 IEEE 10th International Symposium on Biomedical Imaging* (San Francisco, CA: IEEE), 266–269.

Daianu, M., Jahanshad, N., Nir, T. M., Toga, A. W., Jack, C. R. Jr., Weiner, M. W., et al. (2013b). Breakdown of brain connectivity between normal aging and Alzheimer's disease: a structural k-core network analysis. *Brain Connect.* 3, 407–422. doi: 10.1089/brain.2012.0137

Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R. (2007). Regularized, fast, and robust analytical Q-ball imaging. *Magn. Reson. Med.* 58, 497–510. doi: 10.1002/mrm.21277

Descoteaux, M., Deriche, R., Knösche, T., and Anwander, A. (2009). Deterministic and probabilistic tractography based on complex fiber orientation distributions. *IEEE Trans. Med. Imaging* 28, 269–286. doi: 10.1109/tmi.2008.2004424

Ding, L., Zhu, A. H., Saremi, A., Faskowitz, J. I., Håberg, A., Thompson, P. M., et al. (2018). "Voxelwise meta-analysis of brain structural associations with genome-wide polygenic risk for Alzheimer's disease," in *Proceedings of the 14th International Symposium on Medical Information Processing and Analysis*, Mazatlán, Mexico. doi: 10.1117/12.2506660

Douaud, G., Jbabdi, S., Behrens, T. E. J., Menke, R. A., Gass, A., Monsch, A. U., et al. (2011). DTI measures in crossing-fibre areas: increased diffusion anisotropy reveals early white matter alteration in MCI and mild Alzheimer's disease. *Neuroimage* 55, 880–890. doi: 10.1016/j.neuroimage.2010.12.008

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-0

Fortin, J.-P., Parker, D., Tunç, B., Watanabe, T., Elliott, M. A., Ruparel, K., et al. (2017). Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161, 149–170. doi: 10.1016/j.neuroimage.2017.08.047

Fortin, J.-P., Sweeney, E. M., Muschelli, J., Crainiceanu, C. M., and Shinohara, R. T. (2016). Removing inter-subject technical variability in magnetic resonance imaging studies. *Neuroimage* 132, 198–212. doi: 10.1016/j.neuroimage.2016.02.036

Frank, R. A., Galasko, D., Hampel, H., Hardy, J., de Leon, M. J., Mehta, P. D., et al. (2003). Biological markers for therapeutic trials in Alzheimer's disease: proceedings of the biological markers working group; NIA initiative on neuroimaging in Alzheimer's disease. *Neurobiol. Aging* 24, 521–536. doi: 10.1016/S0197-4580(03)00002-2

Giannelli, M., Cosottini, M., Michelassi, M. C., Lazzarotti, G., Belmonte, G., Bartolozzi, C., et al. (2009). Dependence of brain DTI maps of fractional anisotropy and mean diffusivity on the number of diffusion weighting directions. *J. Appl. Clin. Med. Phys.* 11, 176–190. doi: 10.1120/jacmp.v11i1.2927

Giulietti, G., Torso, M., Serra, L., Spanò, B., Marra, C., Caltagirone, C., et al. (2018). Whole brain white matter histogram analysis of diffusion tensor imaging data detects microstructural damage in mild cognitive impairment and Alzheimer's disease patients. *J. Magn. Reson. Imaging* 48, 767–779. doi: 10.1002/jmri.25947

Grech-Sollars, M., Hales, P. W., Miyazaki, K., Raschke, F., Rodriguez, D., Wilson, M., et al. (2015). Multi-centre reproducibility of diffusion MRI parameters for clinical sequences in the brain. *NMR Biomed.* 28, 468–485. doi: 10.1002/nbm.3269

Grothe, M. J., Barthel, H., Sepulcre, J., Dyrba, M., Sabri, O., and Teipel, S. J. (2017). *In vivo* staging of regional amyloid deposition. *Neurology* 89, 2031–2038. doi: 10.1212/WNL.0000000000004643

Harms, R. L., Fritz, F. J., Tobisch, A., Goebel, R., and Roebroeck, A. (2017). Robust and fast nonlinear optimization of diffusion MRI microstructure models. *Neuroimage* 155, 82–96. doi: 10.1016/j.neuroimage.2017.04.064

Harsan, L. A., Poulet, P., Guignard, B., Steibel, J., Parizel, N., de Sousa, P. L., et al. (2006). Brain dysmyelination and recovery assessment by noninvasive *in vivo* diffusion tensor magnetic resonance imaging. *J. Neurosci. Res.* 83, 392–402. doi: 10.1002/jnr.20742

Hu, C., Hua, X., Ying, J., Thompson, P. M., Fakhri, G. E., and Li, Q. (2016). Localizing sources of brain disease progression with network diffusion model. *IEEE J. Sel. Top. Signal Process.* 10, 1214–1225. doi: 10.1109/jstsp.2016.2601695

Hua, X., Lee, S., Yanovsky, I., Leow, A. D., Chou, Y.-Y., Ho, A. J., et al. (2009). Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. *Neuroimage* 48, 668–681. doi: 10.1016/j.neuroimage.2009.07.011

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., et al. (2008). Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage* 39, 336–347. doi: 10.1016/j.neuroimage.2007.07.053

Jack, C. R. Jr., Barnes, J., Bernstein, M. A., Borowski, B. J., Brewer, J., Clegg, S., et al. (2015). Magnetic resonance imaging in Alzheimer's disease neuroimaging initiative 2. *Alzheimers Dement.* 11, 740–756. doi: 10.1016/j.jalz.2015.05.002

Jahanshad, N., Kochunov, P. V., Sprooten, E., Mandl, R. C., Nichols, T. E., Almasy, L., et al. (2013). Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. *Neuroimage* 81, 455–469. doi: 10.1016/j.neuroimage.2013.04.061

Jbabdi, S., Behrens, T. E. J., and Smith, S. M. (2010). Crossing fibres in tract-based spatial statistics. *Neuroimage* 49, 249–256. doi: 10.1016/j.neuroimage.2009.08.039

Jensen, J. H., Helpern, J. A., Ramani, A., Lu, H., and Kaczynski, K. (2005). Diffusional kurtosis imaging: the quantification of non-Gaussian water diffusion by means of magnetic resonance imaging. *Magn. Reson. Med.* 53, 1432–1440. doi: 10.1002/mrm.20508

Jeurissen, B., Leemans, A., Tournier, J.-D., Jones, D. K., and Sijbers, J. (2013). Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Hum. Brain Mapp.* 34, 2747–2766. doi: 10.1002/hbm.22099

Jin, Y., Huang, C., Daianu, M., Zhan, L., Dennis, E. L., Reid, R. I., et al. (2017). 3D tract-specific local and global analysis of white matter integrity in Alzheimer's disease. *Hum. Brain Mapp.* 38, 1191–1207. doi: 10.1002/hbm.23448

Johnson, W. E., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi: 10.1093/biostatistics/kxj037

Kantarci, K., Murray, M. E., Schwarz, C. G., Reid, R. I., Przybelski, S. A., Lesnick, T., et al. (2017). White-matter integrity on DTI and the pathologic staging of Alzheimer's disease. *Neurobiol. Aging* 56, 172–179. doi: 10.1016/j.neurobiolaging.2017.04.024

Kauppi, K., Fan, C. C., McEvoy, L. K., Holland, D., Tan, C. H., Chen, C.-H., et al. (2018). Combining polygenic hazard score with volumetric MRI and cognitive measures improves prediction of progression from mild cognitive impairment to Alzheimer's disease. *Front. Neurosci.* 12:260. doi: 10.3389/fnins.2018.00260

Kavcic, V., Ni, H., Zhu, T., Zhong, J., and Duffy, C. J. (2008). White matter integrity linked to functional impairments in aging and early Alzheimer's disease. *Alzheimers Dement.* 4, 381–389. doi: 10.1016/j.jalz.2008.07.001

Kelly, S., Jahanshad, N., Zalesky, A., Kochunov, P., Agartz, I., Alloza, C., et al. (2018). Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA schizophrenia DTI working group. *Mol. Psychiatry* 23, 1261–1269. doi: 10.1038/mp.2017.170

Kim, M., Ronen, I., Ugurbil, K., and Kim, D.-S. (2006). Spatial resolution dependence of DTI tractography in human occipito-callosal region. *Neuroimage* 32, 1243–1249. doi: 10.1016/j.neuroimage.2006.06.006

Kneynsberg, A., Combs, B., Christensen, K., Morfini, G., and Kanaan, N. M. (2017). Axonal degeneration in tauopathies: disease relevance and underlying mechanisms. *Front. Neurosci.* 11:572. doi: 10.3389/fnins.2017.00572

Kochunov, P., Dickie, E. W., Viviano, J. D., Turner, J., Kingsley, P. B., Jahanshad, N., et al. (2018). Integration of routine QA data into mega-analysis may improve quality and sensitivity of multisite diffusion tensor imaging studies. *Hum. Brain Mapp.* 39, 1015–1023. doi: 10.1002/hbm.23900

Kochunov, P., Jahanshad, N., Sprooten, E., Nichols, T. E., Mandl, R. C., Almasy, L., et al. (2014). Multi-site study of additive genetic effects on fractional anisotropy of cerebral white matter: comparing meta and mega-analytical approaches for data pooling. *Neuroimage* 95, 136–150. doi: 10.1016/j.neuroimage.2014.03.033

Le Bihan, D., and Johansen-Berg, H. (2012). Diffusion MRI at 25: exploring brain tissue structure and function. *Neuroimage* 61, 324–341. doi: 10.1016/j.neuroimage.2011.11.006

Lee, J.-T., Xu, J., Lee, J.-M., Ku, G., Han, X., Yang, D.-I., et al. (2004). Amyloid-β peptide induces oligodendrocyte death by activating the neutral sphingomyelinase-ceramide pathway. *J. Cell Biol.* 164, 123–131. doi: 10.1083/jcb.200307017

Leow, A. D., Yanovsky, I., Chiang, M. C., Lee, A. D., Klunder, A. D., Lu, A., et al. (2007). Statistical properties of Jacobian maps and inverse-consistent deformations in non-linear image registration. *IEEE Trans. Med. Imaging* 26, 822–832. doi: 10.1109/TMI.2007.892646

Leow, A. D., Zhu, S., Zhan, L., McMahon, K., de Zubicaray, G. I., Meredith, M., et al. (2009). The tensor distribution function. *Magn. Reson. Med.* 61, 205–214. doi: 10.1002/mrm.21852

Liu, M., Zhang, J., Adeli, E., and Shen, D. (2017). Deep multi-task multi-channel learning for joint classification and regression of brain status. *Med. Image Comput. Comput. Assist. Interv.* 10435, 3–11. doi: 10.1007/978-3-319-66179-7_1

Maggipinto, T., Bellotti, R., Amoroso, N., Diacono, D., Donvito, G., Lella, E., et al. (2017). DTI measurements for Alzheimer's classification. *Phys. Med. Biol.* 62, 2361–2375. doi: 10.1088/1361-6560/aa5dbe

Maier-Hein, K. H., Neher, P. F., Houde, J. C., Côté, M. A., Garyfallidis, E., Zhong, J., et al. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nat. Commun.* 8:1349. doi: 10.1038/s41467-017-01285-x

Mayo, C. D., Mazerolle, E. L., Ritchie, L., Fisk, J. D., and Gawryluk, J. R. (2017). Longitudinal changes in microstructural white matter metrics in Alzheimer's disease. *Neuroimage Clin.* 13, 330–338. doi: 10.1016/j.nicl.2016.12.012

Mielke, M. M., Kozauer, N. A., Chan, K. C. G., George, M., Toroney, J., Zerrate, M., et al. (2009). Regionally-specific diffusion tensor imaging in mild cognitive impairment and Alzheimer's disease. *Neuroimage* 46, 47–55. doi: 10.1016/j.neuroimage.2009.01.054

Mirzaalian, H., Ning, L., Savadjiev, P., Pasternak, O., Bouix, S., Michailovich, O., et al. (2018). Multi-site harmonization of diffusion MRI data in a registration framework. *Brain Imaging Behav.* 12, 284–295. doi: 10.1007/s11682-016-9670-y

Moore, E. E., Hohman, T. J., Badami, F. S., Pechman, K. R., Osborn, K. E., Acosta, L. M. Y., et al. (2018). Neurofilament relates to white matter microstructure in older adults. *Neurobiol. Aging* 70, 233–241. doi: 10.1016/j.neurobiolaging.2018.06.023

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., et al. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage* 40, 570–582. doi: 10.1016/j.neuroimage.2007.12.035

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., et al. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's disease neuroimaging initiative (ADNI). *Alzheimers Dement.* 1, 55–66. doi: 10.1016/j.jalz.2005.06.003

Müller, H.-P., and Kassubek, J. (2013). Diffusion tensor magnetic resonance imaging in the analysis of neurodegenerative diseases. *J. Vis. Exp.* 77:e50427. doi: 10.3791/50427

Nir, T. M., Jahanshad, N., Toga, A. W., Bernstein, M. A., Jack, C. R., Weiner, M. W., et al. (2015). Connectivity network measures predict volumetric atrophy in mild cognitive impairment. *Neurobiol. Aging* 36, S113–S120. doi: 10.1016/j.neurobiolaging.2014.04.038

Nir, T. M., Jahanshad, N., Villalon-Reina, J. E., Isaev, D., Zavaliangos-Petropulu, A., Zhan, L., et al. (2017). Fractional anisotropy derived from the diffusion tensor distribution function boosts power to detect Alzheimer's disease deficits. *Magn. Reson. Med.* 78, 2322–2333. doi: 10.1002/mrm.26623

Nir, T. M., Jahanshad, N., Villalon-Reina, J. E., Toga, A. W., Jack, C. R., Weiner, M. W., et al. (2013). Effectiveness of regional DTI measures in distinguishing Alzheimer's disease, MCI, and normal aging. *Neuroimage Clin.* 3, 180–195. doi: 10.1016/j.nicl.2013.07.006

O'Dwyer, L., Lamberton, F., Bokde, A. L. W., Ewers, M., Faluyi, Y. O., Tanner, C., et al. (2011). Multiple indices of diffusion identifies white matter damage in mild cognitive impairment and Alzheimer's disease. *PLoS One* 6:e21745. doi: 10.1371/journal.pone.0021745

Oishi, K., Mielke, M. M., Albert, M., Lyketsos, C. G., and Mori, S. (2011). DTI analyses and clinical applications in Alzheimer's disease. *J. Alzheimers Dis.* 26, 287–296. doi: 10.3233/JAD-2011-0007

Palacios, E. M., Martin, A. J., Boss, M. A., Ezekiel, F., Chang, Y. S., Yuh, E. L., et al. (2017). Toward precision and reproducibility of diffusion tensor imaging: a multicenter diffusion phantom and traveling volunteer study. *Am. J. Neuroradiol.* 38, 537–545. doi: 10.3174/ajnr.a5025

Parker, T. D., Slattery, C. F., Zhang, J., Nicholas, J. M., Paterson, R. W., Foulkes, A. J. M., et al. (2018). Cortical microstructure in young onset alzheimer's disease using neurite orientation dispersion and density imaging. *Hum. Brain Mapp.* 39, 3005–3017. doi: 10.1002/hbm.24056

Phillips, J. S., Das, S. R., McMillan, C. T., Irwin, D. J., Roll, E. E., Da Re, F., et al. (2018). Tau PET imaging predicts cognition in atypical variants of Alzheimer's disease. *Hum. Brain Mapp.* 39, 691–708. doi: 10.1002/hbm.23874

Pohl, K. M., Sullivan, E. V., Rohlfing, T., Chu, W., Kwon, D., Nichols, B. N., et al. (2016). Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study. *Neuroimage* 130, 194–213. doi: 10.1016/j.neuroimage.2016. 01.061

Popuri, K., Balachandar, R., Alpert, K., Lu, D., Bhalla, M., Mackenzie, I. R., et al. (2018). Development and validation of a novel dementia of Alzheimer's type (DAT) score based on metabolism FDG-PET imaging. *Neuroimage Clin.* 18, 802–813. doi: 10.1016/j.nicl.2018.03.007

Powell, F., Tosun, D., Sadeghi, R., Weiner, M., Raj, A., and Alzheimer's Disease Neuroimaging Initiative. (2018). Preserved structural network organization mediates pathology spread in Alzheimer's disease spectrum despite loss of white matter tract integrity. *J. Alzheimers Dis.* 65, 747–764. doi: 10.3233/jad-170798

Prasad, G., Nir, T. M., Toga, A. W., and Thompson, P. M. (2013). Tractography density and network measures in Alzheimer's disease. *Proc. IEEE Int. Symp. Biomed. Imaging* 2013, 692–695. doi: 10.1109/isbi.2013.6556569

Rosen, W. G., Mohs, R. C., and Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *Am. J. Psychiatry* 141, 1356–1364. doi: 10.1176/ajp.141. 11.1356

Rosenthal, R., and Rosnow, R. L. (1991). *Essentials of Behavioral Research: Methods and Data Analysis.* New York, NY: McGraw-Hill.

Roth, A. D., Ramírez, G., Alarcón, R., and Von Bernhardi, R. (2005). Oligodendrocytes damage in Alzheimer's disease: β amyloid toxicity and inflammation. *Biol. Res.* 38, 381–387. doi: 10.4067/s0716-97602005000400011

Sachdev, P. S., Zhuang, L., Braidy, N., and Wen, W. (2013). Is Alzheimer's a disease of the white matter? *Curr. Opin. Psychiatry* 26, 244–251. doi: 10.1097/YCO. 0b013e32835ed6e8

Sanchez-Rodriguez, L. M., Iturria-Medina, Y., Baines, E. A., Mallo, S. C., Dousty, M., Sotero, R. C., et al. (2018). Design of optimal nonlinear network controllers for Alzheimer's disease. *PLoS Comput. Biol.* 14:e1006136. doi: 10.1371/journal.pcbi.1006136

Shaw, L. M., Korecka, M., Clark, C. M., Lee, V. M.-Y., and Trojanowski, J. Q. (2007). Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics. *Nat. Rev. Drug Discov.* 6, 295–303. doi: 10.1038/nrd2176

Slattery, C. F., Zhang, J., Paterson, R. W., Foulkes, A. J. M., Carton, A., Macpherson, K., et al. (2017). ApoE influences regional white-matter axonal density loss in Alzheimer's disease. *Neurobiol. Aging* 57, 8–17. doi: 10.1016/j. neurobiolaging.2017.04.021

Song, S.-K., Sun, S.-W., Ju, W.-K., Lin, S.-J., Cross, A. H., and Neufeld, A. H. (2003). Diffusion tensor imaging detects and differentiates axon and myelin degeneration in mouse optic nerve after retinal ischemia. *Neuroimage* 20, 1714–1722. doi: 10.1016/j.neuroimage.2003.07.005

Song, S.-K., Yoshino, J., Le, T. Q., Lin, S.-J., Sun, S.-W., Cross, A. H., et al. (2005). Demyelination increases radial diffusivity in corpus callosum of mouse brain. *Neuroimage* 26, 132–140. doi: 10.1016/j.neuroimage.2005.01.028

Sulaimany, S., Khansari, M., Zarrineh, P., Daianu, M., Jahanshad, N., Thompson, P. M., et al. (2017). Predicting brain network changes in Alzheimer's disease with link prediction algorithms. *Mol. Biosyst.* 13, 725–735. doi: 10.1039/c6mb00815a

Tax, C. M. W., Grussu, F., Kaden, E., Ning, L., Rudrapatna, U., Evans, J., et al. (2018). Cross-vendor and Cross-protocol harmonisation of diffusion MRI data: a comparative study. *Proc. Jt. Annu. Meet.* 471. Available online at: https://www.researchgate.net/publication/325737885_Cross-vendor_and_ Cross-protocol_harmonisation_of_diffusion_MRI_data_a_comparative_ study

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182. doi: 10.1007/s11682-013-9269-5

Tournier, J.-D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage* 23, 1176–1185. doi: 10.1016/j.neuroimage.2004.07.037

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. doi: 10.1002/mrm.10268

Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2019). Understanding disease progression and improving Alzheimer's disease clinical trials: recent highlights from the Alzheimer's disease neuroimaging initiative. *Alzheimers Dement.* 15, 106–152. doi: 10.1016/j.jalz. 2018.08.005

Wang, B., Niu, Y., Miao, L., Cao, R., Yan, P., Guo, H., et al. (2017). Decreased complexity in Alzheimer's disease: resting-state fMRI evidence of brain entropy mapping. *Front. Aging Neurosci.* 9:378. doi: 10.3389/fnagi.2017.00378

Wang, M.-L., Wei, X.-E., Fu, J.-L., Li, W., Yu, M.-M., Li, P.-Y., et al. (2018). Subcortical nuclei in Alzheimer's disease: a volumetric and diffusion kurtosis imaging study. *Acta Radiol.* 59, 1365–1371. doi: 10.1177/0284185118758122

Wang, X., Zhen, X., Li, Q., Shen, D., and Huang, H. (2018). Cognitive assessment prediction in Alzheimer's disease by multi-layer multi-target regression. *Neuroinformatics* 16, 285–294. doi: 10.1007/s12021-018-9381-1

Zavaliangos-Petropulu, A., Nir, T. M., Thomopoulos, S. I., Jahanshad, N., Reid, R. I., Bernstein, M. A., et al. (2018). "Ranking diffusion tensor measures of brain aging and Alzheimer's disease," in *Proceedings of the 14th International Symposium on Medical Information Processing and Analysis*, Mazatlán, Mexico. doi: 10.1117/12.2506694

Zhan, L., Jahanshad, N., Ennis, D. B., Jin, Y., Bernstein, M. A., Borowski, B. J., et al. (2013). Angular versus spatial resolution trade-offs for diffusion imaging under time constraints. *Hum. Brain Mapp.* 34, 2688–2706. doi: 10.1002/hbm. 22094

Zhan, L., Leow, A. D., Jahanshad, N., Chiang, M.-C., Barysheva, M., Lee, A. D., et al. (2010). How does angular resolution affect diffusion imaging measures? *Neuroimage* 49, 1357–1371. doi: 10.1016/j.neuroimage.2009. 09.057

Zhang, H., Schneider, T., Wheeler-Kingshott, C. A., and Alexander, D. C. (2012). NODDI: practical *in vivo* neurite orientation dispersion and density imaging of the human brain. *Neuroimage* 61, 1000–1016. doi: 10.1016/j.neuroimage.2012. 03.072

Zhou, T., Thung, K.-H., Zhu, X., and Shen, D. (2017). Feature learning and fusion of multimodality neuroimaging and genetic data for multi-status dementia diagnosis. *Mach. Learn. Med. Imaging* 10541, 132–140. doi: 10.1007/978-3-319-67389-9_16

Zhu, T., Hu, R., Qiu, X., Taylor, M., Tso, Y., Yiannoutsos, C., et al. (2011). Quantification of accuracy and precision of multi-center DTI measurements: a diffusion phantom and human brain study. *Neuroimage* 56, 1398–1411. doi: 10.1016/j.neuroimage.2011.02.010

Zhu, T., Liu, X., Gaugh, M. D., Connelly, P. R., Ni, H., Ekholm, S., et al. (2009). Evaluation of measurement uncertainties in human diffusion tensor imaging (DTI)-derived parameters and optimization of clinical DTI protocols with a wild bootstrap analysis. *J. Magn. Reson. Imaging* 29, 422–435. doi: 10.1002/jmri. 21647

Zhu, A. H., Moyer, D. C., Nir, T. M., Thompson, P. M., and Jahanshad, N. (2018). "Challenges and opportunities in diffusion MRI data harmonization," in *Computational Diffusion MRI and Brain Connectivity. Mathematics and Visualization* (Cham: Springer). doi: 10.1007/978-3-030-05831-9

Check for updates

An Empirical Comparison of Meta- and Mega-Analysis With Data From the ENIGMA Obsessive-Compulsive Disorder Working Group

Premika S. W. Boedhoe [1,2], Martijn W. Heymans [3], Lianne Schmaal [4,5], Yoshinari Abe [6], Pino Alonso [7,8,9], Stephanie H. Ameis [10,11], Alan Anticevic [12], Paul D. Arnold [13,14], Marcelo C. Batistuzzo [15], Francesco Benedetti [16], Jan C. Beucke [17], Irene Bollettini [16], Anushree Bose [18], Silvia Brem [19], Anna Calvo [20], Rosa Calvo [8,21], Yuqi Cheng [22], Kang Ik K. Cho [23], Valentina Ciullo [24,25], Sara Dallaspezia [16], Damiaan Denys [26,27], Jamie D. Feusner [28], Kate D. Fitzgerald [29], Jean-Paul Fouche [30], Egill A. Fridgeirsson [26], Patricia Gruner [12], Gregory L. Hanna [29], Derrek P. Hibar [31], Marcelo Q. Hoexter [15], Hao Hu [32], Chaim Huyser [33,34], Neda Jahanshad [35], Anthony James [36], Norbert Kathmann [17], Christian Kaufmann [17], Kathrin Koch [37,38], Jun Soo Kwon [39,40], Luisa Lazaro [8,21,41,42], Christine Lochner [43], Rachel Marsh [44,45], Ignacio Martínez-Zalacaín [7], David Mataix-Cols [46], José M. Menchón [7,8,9], Luciano Minuzzi [47], Astrid Morer [8,21,41], Takashi Nakamae [6], Tomohiro Nakao [48], Janardhanan C. Narayanaswamy [18], Seiji Nishida [6], Erika L. Nurmi [28], Joseph O'Neill [28], John Piacentini [28], Fabrizio Piras [24], Federica Piras [24], Y. C. Janardhan Reddy [18], Tim J. Reess [37,38], Yuki Sakai [6,49], Joao R. Sato [50], H. Blair Simpson [44,51], Noam Soreni [52], Carles Soriano-Mas [7,8,53], Gianfranco Spalletta [24,54], Michael C. Stevens [55,56], Philip R. Szeszko [57,58], David F. Tolin [55,59], Guido A. van Wingen [26], Ganesan Venkatasubramanian [18], Susanne Walitza [19], Zhen Wang [32,60], Je-Yeon Yun [35,39], ENIGMA-OCD Working-Group [†], Paul M. Thompson [31], Dan J. Stein [30], Odile A. van den Heuvel [1,2*‡] and Jos W. R. Twisk [3*‡]

[1] Department of Psychiatry, Amsterdam University Medical Centers (UMC), Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam, Netherlands, [2] Department of Anatomy and Neurosciences, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Amsterdam Neuroscience, Amsterdam, Netherlands, [3] Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam University Medical Centers, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, [4] Orygen, The National Centre of Excellence in Youth Mental Health, Melbourne, VIC, Australia, [5] Centre for Youth Mental Health, The University of Melbourne, Melbourne, VIC, Australia, [6] Department of Psychiatry, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan, [7] Department of Psychiatry, Bellvitge University Hospital, Bellvitge Biomedical Research Institute-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain, [8] Centro de Investigación Biomèdica en Red de Salud Mental (CIBERSAM), Barcelona, Spain, [9] Department of Clinical Sciences, University of Barcelona, Barcelona, Spain, [10] Department of Psychiatry, Faculty of Medicine, The Centre for Addiction and Mental Health, The Margaret and Wallace McCain Centre for Child, Youth and Family Mental Health, Campbell Family Mental Health Research Institute, University of Toronto, Toronto, ON, Canada, [11] The Hospital for Sick Children, Centre for Brain and Mental Health, Toronto, ON, Canada, [12] Department of Psychiatry, Yale University School of Medicine, New Haven, CT, United States, [13] Mathison Centre for Mental Health Research and Education, Cumming School of Medicine, Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada, [14] Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada, [15] Departamento de Psiquiatria, Faculdade de Medicina, Instituto de Psiquiatria, Universidade de São Paulo, São Paulo, Brazil, [16] Division of Neuroscience, Psychiatry and Clinical Psychobiology, Scientific Institute Ospedale San Raffaele, Milan, Italy, [17] Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany, [18] Obsessive-Compulsive Disorder (OCD) Clinic Department of Psychiatry National Institute of Mental Health and Neurosciences, Bangalore, India, [19] Department of Child and Adolescent Psychiatry and Psychotherapy, Psychiatric Hospital, University of Zurich, Zurich, Switzerland, [20] Magnetic Resonance Image Core Facility, IDIBAPS (Institut d'Investigacions Biomèdiques August Pi i Sunyer), Barcelona, Spain, [21] Department of Child and Adolescent Psychiatry and Psychology, Hospital Clínic Universitari, Institute of Neurosciences, Barcelona, Spain, [22] Department of Psychiatry, First Affiliated Hospital of Kunming Medical University, Kunming, China, [23] Institute of Human Behavioral Medicine, SNU-MRC, Seoul, South Korea, [24] Laboratory of Neuropsychiatry, Department of Clinical and Behavioral Neurology, IRCCS Santa Lucia Foundation, Rome, Italy, [25] Neurosciences, Psychology, Drug Research and Child Health (NEUROFARBA), University of Florence, Florence, Italy

[26] Department of Psychiatry, Amsterdam Neuroscience, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [27] Netherlands Institute for Neuroscience, Royal Netherlands Academy of Arts and Sciences, Amsterdam, Netherlands, [28] Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA, United States, [29] Department of Psychiatry, University of Michigan, Ann Arbor, MI, United States, [30] MRC Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry, University of Cape Town, Cape Town, South Africa, [31] Imaging Genetics Center, Keck School of Medicine of the University of Southern California, Mark and Mary Stevens Neuroimaging and Informatics Institute, Marina del Rey, CA, United States, [32] Shanghai Mental Health Center Shanghai Jiao Tong University School of Medicine, Shanghai, China, [33] De Bascule, Academic Center for Child and Adolescent Psychiatry, Amsterdam, Netherlands, [34] Department of Child and Adolescent Psychiatry, Amsterdam University Medical Centers, University of Amsterdam, Amsterdam, Netherlands, [35] Yeongeon Student Support Center, Seoul National University College of Medicine, Seoul, South Korea, [36] Department of Psychiatry, Oxford University, Oxford, United Kingdom, [37] Department of Neuroradiology, Klinikum rechts der Isar, Technische Universität München, Munich, Germany, [38] TUM-Neuroimaging Center (TUM-NIC) of Klinikum rechts der Isar, Technische Universität München, Munich, Germany, [39] Department of Psychiatry, Seoul National University College of Medicine, Seoul, South Korea, [40] Department of Brain and Cognitive Sciences, Seoul National University College of Natural Sciences, Seoul, South Korea, [41] Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain, [42] Department of Medicine, University of Barcelona, Barcelona, Spain, [43] SU/UCT MRC Unit on Anxiety and Stress Disorders, Department of Psychiatry, University of Stellenbosch, Stellenbosch, South Africa, [44] Columbia University Medical College, Columbia University, New York, NY, United States, [45] The New York State Psychiatric Institute, New York, NY, United States, [46] Department of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet, Stockholm, Sweden, [47] Mood Disorders Clinic, St. Joseph's HealthCare, Hamilton, ON, Canada, [48] Department of Neuropsychiatry, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan, [49] ATR Brain Information Communication Research Laboratory Group, Kyoto, Japan, [50] Center for Mathematics, Computing and Cognition, Universidade Federal do ABC, Santo Andre, Brazil, [51] Center for OCD and Related Disorders, New York State Psychiatric Institute, New York, NY, United States, [52] Anxiety Treatment and Research Center, St. Joseph's HealthCare, Hamilton, ON, Canada, [53] Department of Psychobiology and Methodology of Health Sciences, Universitat Autònoma de Barcelona, Barcelona, Spain, [54] Beth K. and Stuart C. Yudofsky Division of Neuropsychiatry, Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, United States, [55] Yale University School of Medicine, New Haven, CT, United States, [56] Clinical Neuroscience and Development Laboratory, Olin Neuropsychiatry Research Center, Hartford, CT, United States, [57] Icahn School of Medicine at Mount Sinai, New York, NY, United States, [58] James J. Peters VA Medical Center, Bronx, NY, United States, [59] Institute of Living/Hartford Hospital, Hartford, CT, United States, [60] Shanghai Key Laboratory of Psychotic Disorders, Shanghai, China

Objective: Brain imaging communities focusing on different diseases have increasingly started to collaborate and to pool data to perform well-powered meta- and mega-analyses. Some methodologists claim that a one-stage individual-participant data (IPD) mega-analysis can be superior to a two-stage aggregated data meta-analysis, since more detailed computations can be performed in a mega-analysis. Before definitive conclusions regarding the performance of either method can be drawn, it is necessary to critically evaluate the methodology of, and results obtained by, meta- and mega-analyses.

Methods: Here, we compare the inverse variance weighted random-effect meta-analysis model with a multiple linear regression mega-analysis model, as well as with a linear mixed-effects random-intercept mega-analysis model, using data from 38 cohorts including 3,665 participants of the ENIGMA-OCD consortium. We assessed the effect sizes and standard errors, and the fit of the models, to evaluate the performance of the different methods.

Results: The mega-analytical models showed lower standard errors and narrower confidence intervals than the meta-analysis. Similar standard errors and confidence intervals were found for the linear regression and linear mixed-effects random-intercept models. Moreover, the linear mixed-effects random-intercept models showed better fit indices compared to linear regression mega-analytical models.

Conclusions: Our findings indicate that results obtained by meta- and mega-analysis differ, in favor of the latter. In multi-center studies with a moderate amount of variation between cohorts, a linear mixed-effects random-intercept mega-analytical framework appears to be the better approach to investigate structural neuroimaging data.

Keywords: neuroimaging, MRI, IPD meta-analysis, mega-analysis, linear mixed-effect models

INTRODUCTION

Data pooling across individual studies has the potential to significantly accelerate progress in brain imaging (Van Horn et al., 2001), as demonstrated by large-scale neuroimaging initiatives, such as the ENIGMA (Enhanced NeuroImaging Genetics through Meta-Analysis) consortium (Thompson et al., 2014). The most immediate advantage of data pooling is increased power due to the larger number of subjects available for analysis. Data pooling across multiple centers worldwide can also lead to a more heterogeneous and potentially representative participant sample. Large-scale studies are well-powered to distinguish consistent, generalizable findings from false positives that emerge from smaller-sampled studies. The participation of many experts may also lead to a more balanced interpretation, wider endorsement of the conclusions by others, and greater dissemination of results (Stewart, 1995).

An aggregate data meta-analysis is the most conventional approach, where summary results, such as effect size estimates, standard errors, and confidence intervals, are extracted from primary published studies and then synthesized to estimate the overall effect for all the studies combined (de Bakker et al., 2008). This approach is relatively quick and inexpensive, but often prone to selective reporting in primary studies, publication bias, low power to detect interaction effects and lack of harmonization of data processing and analysis methods among the included studies. To overcome these issues, collaborative groups are increasingly collating individual-participant data (IPD) from multiple studies to jointly analyze the individual-level data in a meta-analysis of IPD (Stewart, 1995). The IPD approach allows standardization of processing protocols and statistical analyses, culminating in study results not provided by the individual publications. This approach also allows modeling of interaction effects within the studies. Given these advantages, the IPD approach is currently the gold standard.

There are two competing statistical approaches for IPD meta-analysis: a two-stage or a one-stage approach (Thomas et al., 2014). In the two-stage approach, the first step includes analyzing the IPD from each study separately, to obtain aggregate (summary) data (e.g., effect size estimates and confidence intervals). The second step includes using standard meta-analytical techniques, such as a random effects meta-analysis model. The alternative one-stage approach analyzes all IPD in one statistical model while accounting for clustering among patients in the same study, to estimate an overall effect. Throughout this manuscript, the one-stage IPD approach is referred to as *mega-analysis*, while the two-stage approach is referred to as *meta-analysis*.

Some methodologists claim that a mega-analysis can be superior to meta-analysis. The comprehensive evaluation of missing data and greater flexibility in the control of confounders at the level of individual patients and specific studies are significant advantages of a mega-analytical approach. Mega-analyses have also been recommended as they avoid the assumptions of within-study normality and known within-study variances, which are especially problematic with smaller samples (Debray et al., 2013). Despite these advantages, mega-analysis

requires homogeneous data sets and the establishment of a common centralized database. The latter criterion is time-consuming since cleaning, checking, and re-formatting the various data sets adds to the time and costs of performing mega-analyses. Obtaining IPD may also be challenging and limited by the terms of the informed consent or other data sharing constraints within each study. These are the main reasons why researchers often prefer meta-analysis using summary statistics. Additionally, meta-analysis allows for analyses of individual studies to account for local population substructure and study-specific covariates that may be better dealt with within each study. While each method has its own advantages and limitations, researchers still debate which method is superior for tackling different types of questions [see (Stewart and Tierney, 2002; Burke et al., 2017) for reviews on advantages and disadvantages of each approach].

Brain imaging communities focusing on different diseases have started collaborating to perform well-powered meta- and mega-analyses. In the largest studies to date on the neural correlates of OCD, the authors of the ENIGMA-OCD consortium (Boedhoe et al., 2017a, 2018) conducted a mega-analysis, pooling individual participant-level data from more than 25 research institutes worldwide, as well as a meta-analysis by combining summary statistic results from the independent sites. The meta- and mega-analyses revealed comparable findings of subcortical abnormalities in OCD (Boedhoe et al., 2017a), but the mega-analytical approach seemed more sensitive for detecting subtle cortical abnormalities (Boedhoe et al., 2018). Before definitive conclusions regarding the performance of either method can be drawn, it is necessary to critically evaluate the results obtained by various approaches for meta- and mega-analyses.

Herein, we use data from the ENIGMA-OCD consortium to compare results obtained by meta- and mega-analyses. Specifically, we applied the inverse variance weighted random-effect meta-analysis model and the multiple linear regression mega-analysis model as used in the aforementioned studies (Boedhoe et al., 2017a, 2018). In addition, we compared findings from these models to those detected with a linear mixed-effects random-intercept mega-analytical model. Effect sizes and standard error estimates, and (where possible) model fit were used to evaluate which of the methods performs best.

METHODS

Samples

The ENIGMA-OCD working group includes 38 data sets from 27 international research institutes with neuroimaging and clinical data from OCD patients and typically developing healthy control subjects, including both children and adults (Boedhoe et al., 2018). We defined adults as individuals aged ≥18 years and children as individuals aged <18 years. The split at the age of 18 followed from a natural selection of the age ranges used in these samples, as most samples used the age of 18 years as a cut-off for inclusion. Because our previous findings and the literature suggest differential effects between pediatric and adult samples, we performed separate analyses for adult and pediatric data [for demographics and further details on the samples, see (Boedhoe

et al., 2018)]. In total, we analyzed data from 3,665 participants including 1,905 OCD patients (407 children and 1,498 adults) and 1,760 control participants (324 children and 1,436 adults). All local institutional review boards permitted the use of measures extracted from the coded data for analyses.

Image Acquisition and Processing

Structural T1-weighted brain MRI scans were acquired and processed locally. For image acquisition parameters of each site, please see (Boedhoe et al., 2018). All cortical parcellations were performed with the fully automated segmentation software FreeSurfer, version 5.3 (Fischl, 2012), following standardized ENIGMA protocols to harmonize analyses and quality control procedures across multiple sites (see http://enigma.usc.edu/protocols/imaging-protocols/). Segmentations of 68 (34 left and 34 right) cortical gray matter regions based on the Desikan-Killiany atlas (Desikan et al., 2006) and two whole-hemisphere measures were visually inspected and statistically evaluated for outliers [see (Boedhoe et al., 2018) for further details on quality checking].

Statistical Framework

We examined differences between OCD patients and controls across samples by performing (1) an inverse variance weighted random-effects meta-analysis model; (2) a multiple linear regression mega-analysis model; and (3) a linear mixed-effects random-intercept mega-analysis model. Each of the 70 cortical regions of interest (68 regions and two whole-hemisphere averages) served as the outcome measure and a binary indicator of diagnosis as the predictor of interest. In the meta-analysis, all cortical thickness models were adjusted for age and sex (Im et al., 2008; Westlye et al., 2010), and all cortical surface area models were corrected for age, sex, and intracranial volume (Barnes et al., 2010; Ikram et al., 2012). In the mega-analysis all models were also adjusted for scanning center (cohort). The two mega-analytical frameworks are similar, but the models account differently for clustering of data within cohorts; linear regression with a dummy variable for each cohort and linear mixed-effects models (more efficiently) with only one variance parameter. Finally, all models were fit using the restricted maximum likelihood method [REML (Harville, 1977)].

The meta- and mega-analysis encompass intrinsically different statistics, including differences in approaches for dealing with missing data. E.g., the mega-analysis estimates one restricted maximum likelihood over the entire data set. This estimation contains information of each of the other cohorts. The first stage of the meta-analysis includes the estimation of a restricted maximum likelihood per cohort, making this method more vulnerable to missing outcome data. Therefore, we descriptively compared the meta- and mega-analyses by examining the confidence intervals and standard error estimates for the effect sizes assessed. In addition, the Bayesian information criterion (BIC) were used to evaluate which of the mega-analytical models performs better. A lower BIC indicates a better model fit. Throughout the manuscript, we report $p < 0.001$.

Meta-Analysis

We analyzed the IPD from each study to obtain aggregated summary data. Effect size estimates were calculated using Cohen's d, computed from the t-statistic of the diagnosis indicator variable from the regression models [(Nakagawa and Cuthill, 2007), *equation 10*]. All regression models and effect size estimates were fitted at each site separately. A final Cohen's d effect size estimate was obtained using an inverse variance-weighted random-effect meta-analysis model in R (metafor package, version 1.9-118). This meta-analytic framework enabled us to combine data from multiple sites and take the sample size of each cohort into account by weighing individual effect size estimates for the inverse variance per cohort.

Mega-Analysis

We pooled all IPD in one statistical model to perform mega-analyses and fitted the following models:

Linear Regression

The linear regression model included cohorts as dummy variables. Effect size estimates were calculated using the Cohen's d metric computed from the t-statistic of the diagnosis indicator variable from the regression models [(Nakagawa and Cuthill, 2007), *equation 10*].

Linear Mixed-Effects Model – Random-Intercept

Linear mixed-effects models are extensions of linear regression models and efficiently account for clustering of data within cohorts. By adding a random-intercept for cohort, the adjustment for the clustering of data within cohorts is performed with only one (variance) parameter, which reduces the number of estimated parameters (rather than estimating the intercept of each dummy variable separately as in the linear regression model described above). We used *lme4* (linear mixed-effects analysis) package in R to perform the analyses. Effect size estimates were calculated using the Cohen's d metric computed from the t-values from the mixed-effects model [(Nakagawa and Cuthill, 2007), *equation 22*].

RESULTS

The results of the meta-analysis and linear regression mega-analysis have been published previously (Boedhoe et al., 2018). In this paper, we added the linear mixed-effects random-intercept mega-analysis and statistically compared the various approaches.

Meta-Analysis

No significant differences ($p < 0.001$) in cortical thickness were observed in adult OCD patients ($N = 1,498$) compared to healthy controls ($N = 1,436$) (**Supplementary Table S1**). The meta-analysis did reveal a lower surface area of the transverse temporal cortex (Cohen's d −0.17) in OCD patients (**Supplementary Table S2**). No group differences in cortical thickness or surface area were observed in children with OCD ($N = 407$) compared to control children ($N = 324$) (**Supplementary Tables S3, S4**).

Mega-Analysis

Both the linear regression (Cohen's d −0.14) and the linear mixed-effects random-intercept (Cohen's d −0.11) models revealed significantly lower cortical thickness in bilateral inferior parietal cortices in adult OCD patients ($N = 1,498$) compared to healthy controls ($N = 1,436$) (**Supplementary Table S5**). Both models also showed significantly lower surface area (Cohen's d −0.16) in the left transverse temporal cortex in OCD patients (**Supplementary Table S6**).

Both the linear regression (Cohen's d between −0.24 and −0.31) and the linear mixed-effects random-intercept (Cohen's d between −0.20 and −0.28) models revealed significantly thinner cortices in pediatric OCD patients ($N = 407$) compared with control children ($N = 324$) in the right superior parietal, left inferior parietal, and left lateral occipital cortices (**Supplementary Tables S7**). Neither model revealed significant group differences in cortical surface area (**Supplementary Tables S8**).

Comparing Meta- and Mega-Analysis
Effect Sizes

When looking at the magnitude and order of effect sizes we see the same pattern resulting from the meta-analysis and linear regression mega-analysis in both the pediatric (**Supplementary Tables S3, S7**) and adult (**Supplementary Tables S1, S5**) datasets, i.e., the magnitude and direction of effect of the effect sizes derived from the meta-analysis and linear regression mega-analysis were highly similar. The linear mixed-effects random-intercept mega-analysis also showed a similar pattern of results, but slightly smaller effect sizes (**Table 1** and **Supplementary Tables S5, S7**).

Standard Error and 95% Confidence Intervals

Overall, linear regression and linear mixed-effects random-intercept models showed lower standard errors and narrower confidence intervals than the meta-analysis. Similar standard errors and confidence intervals were found for the different mega-analysis models (**Table 1** and **Supplementary Tables S1–S8**).

Goodness-of-Fit

The linear mixed-effects random-intercept models showed lower BIC values compared to the linear regression mega-analysis (**Table 1** and **Supplementary Tables S9–S12**).

DISCUSSION

The aim of this study was to evaluate different statistical methods for large-scale multi-center neuroimaging analyses. We empirically evaluated whether a meta-analysis provides results comparable to a mega-analysis and which analytical framework performs better. Clinical interpretation of the results can be found elsewhere (Boedhoe et al., 2017b, 2018). Although effect sizes were similar for the meta-analysis and linear regression mega-analysis, lower standard errors and narrower confidence intervals of both mega-analytical approaches compared to the meta-analysis suggest better performance of the mega-analytical

approach over the meta-analytical approach. While the meta-analysis failed to detect cortical thickness differences in both the adult and pediatric samples, it did support the findings of the mega-analyses at a less stringent significance threshold ($p < 0.05$ uncorrected). As a second aim, we investigated which mega-analytical framework was superior. The BIC values indicated a better model fit of the linear mixed-effects random-intercept model compared to the linear regression mega-analytical model.

Whereas, the linear regression model showed similar standard errors and confidence intervals to the linear mixed-effects random-intercept model, the latter fitted the data better. The effect sizes of the linear regression model appeared to be higher than those of the linear mixed-effects models, possibly indicating an overestimation of the effect of diagnosis. Indeed fixed-effects analyses (comparable to the linear regression models in our case) are reported to produce biased estimates or inflated type I error rates when pooled data includes cohorts with a small number of patients (Agresti and Hartzel, 2000; Kahan and Morris, 2012). Mathew and Nordstorm (2010) also suggested that a mega-analysis (one-stage approach) with a random intercept term might be slightly more precise than a meta-analysis (two-stage approach), which has a distinct intercept term per study (Mathew and Nordstorm, 2010). Taken together, our results suggest that the linear mixed-effects random-intercept mega-analysis model is the better approach for analyzing cortical gray matter data in a multi-center neuroimaging study.

We also explored (data not shown) a linear mixed-effects random-intercept and random-slope mega-analytical approach, since the various cohorts might have shown differences in effects of diagnosis related to clinical heterogeneity between patient samples. However, for most of the regions of interest the model did not converge. These computational difficulties and convergence problems have been reported before (Debray et al., 2013). As a result, effect sizes, confidence intervals, standard errors, and BIC values could not be estimated accurately. Indeed previous literature has demonstrated that mega-analyses may produce downwardly biased coefficient estimates when an incorrect model is specified, for instance when random effects are wrongly assumed (Dutton, 2010). Note that including a random slope in the linear mixed-effects model might be valuable when large variance is present in the data between cohorts. Therefore, we recommend the following strategy: (1) run a mixed-effects model with a random-intercept to correct for clustering of participants within cohorts; (2) add a random-slope to correct for potential variance in effects between cohorts; (3) and perform a likelihood-ratio test to statistically compare both models. If the likelihood-ratio test is significant i.e., there is a better fit of the random-intercept random-slope model, this model is preferred over the random-intercept only model. If the likelihood-ratio test is not significant i.e., there is a better fit of the random-intercept only model, this model is preferred over the random-intercept random-slope model.

Olkin and Sampson (1998) showed that for comparing treatments with respect to a continuous outcome in clinical trials, meta-analysis is equivalent to mega-analysis if the treatment effects and error variances are constant across trials. The equivalence has been extended even if the error variances

TABLE 1 | Effect size, confidence interval, and standard error estimates, and BIC values of the main findings.

			Cohen's d	Standard Error	95% CI	BIC
ADULT						
Left inferior parietal cortex		meta-analysis	−0.13	0.053	−0.237 to −0.029	–
		mega-analysis LR*	−0.14	0.038	−0.211 to −0.063	9,693
	cortical thickness	mega-analysis LMEri*	−0.11	0.038	−0.214 to −0.065	9,638
Right inferior parietal cortex		meta-analysis	−0.13	0.046	−0.221 to −0.042	–
		mega-analysis LR*	−0.14	0.038	−0.211 to −0.064	9,799
		mega-analysis LMEri*	−0.10	0.038	−0.213 to −0.066	9,750
Left transverse temporal cortex	cortical surface area	meta-analysis*	−0.17	0.038	−0.243 to −0.095	–
		mega-analysis LR*	−0.16	0.037	−0.238 to −0.092	33,368
		mega-analysis LMEri*	−0.16	0.037	−0.240 to −0.095	33,263
PEDIATRIC						
Right superior parietal cortex		meta-analysis	−0.27	0.138	−0.540 to −0.001	–
		mega-analysis LR*	−0.27	0.075	−0.416 to −0.121	2,645
		mega-analysis LMEri*	−0.21	0.075	−0.415 to −0.119	2,634
Left inferior parietal cortex	cortical thickness	meta-analysis	−0.31	0.144	−0.593 to −0.027	–
		mega-analysis LR*	−0.31	0.077	−0.457 to −0.154	2,634
		mega-analysis LMEri*	−0.28	0.077	−0.455 to −0.151	2,610
Left lateral occipital cortex		meta-analysis	−0.26	0.095	−0.445 to −0.071	–
		mega-analysis LR*	−0.26	0.075	−0.404 to −0.109	2,464
		mega-analysis LMEri*	−0.23	0.075	−0.401 to −0.106	2,444

LR, linear regression; LMEri, linear mixed-effects random-intercept model; CI, confidence interval; BIC= Bayesian information criterion.
*Indicates significant group difference at a threshold of p < 0.001.

are different across trials (Mathew and Nordstrom, 1999). Lin and Zeng theoretically and empirically showed asymptotic equivalence between meta- and mega-analyses when the effect sizes are the same for all studies (Lin and Zeng, 2010a,b). The different cohorts in our study did not all show similar effect sizes and error variances, possibly explaining why we did not find the meta- and mega-analyses to be equivalent. In practice, effect sizes and error variances vary across studies more often than not. Moreover, these authors (Lin and Zeng, 2010a) focused on a fixed-effects meta-analysis rather than a random-effects meta-analysis which is carried out in the current study. A fixed-effect model only takes into account the random error within cohorts, whereas the random-effect model also takes into account the random error between cohorts (Borenstein et al., 2010). Not taking into account the random error between different cohorts in neuroimaging data, for example, may lead to potentially misleading conclusions. More comprehensive simulation studies may be performed to assess theoretical differences in the results of meta- and mega-analyses. Such simulation studies covering various scenarios regarding varying effect sizes and error variances would strengthen our findings.

Conclusions of meta-analyses are often used to guide health care policy and to make decisions regarding the management of individual patients. Thus, it is important that the conclusions of meta-analyses are valid. Although the two approaches (meta- and mega-analysis) often produce similar results, sometimes clinical and/or statistical conclusions are affected (Burke et al., 2017). We agree with Burke et al. (2017) and Debray et al. (2013) that when planning IPD analyses in a multi-center

setting, the choice and implementation of a mega-analysis (one-stage approach) or meta-analysis (two-stage approach) method should be pre-specified, as occasionally they lead to different conclusions. Standardized statistical guidelines addressing the best approach, such as those mentioned in Burke et al. (2017), would be beneficial in this area. For example, meta-analysis (two-stage approach) or mega-analysis (one-stage approach) may be more suitable, depending on outcome types (continuous, binary of time-to-event). In a multi-center study including multiple small sample cohorts, a mega-analysis (one-stage approach) is preferred, as it avoids the use of approximate normal sampling distributions, known within-study variances, and continuity corrections that plague mega-analysis (two-stage approach) with an inverse variance weighting. Additionally, any mega-analysis (one-stage approach) should account for the clustering of participants within cohorts, ideally by including a random-intercept term for cohort. If the effect sizes of the separate studies are expected to vary greatly, it should be investigated whether adding a random-slope to the model is beneficial. For further details about choosing an appropriate method for a multi-center study we recommend Burke et al. (2017).

To our knowledge, this is the first report investigating the utility of meta- vs. mega-analyses for multi-center structural neuroimaging data. The validity of our findings is limited to cortical gray matter measures. Therefore, they may not be generalized to all other brain measures. Nevertheless, our findings show that in the case of cross-sectional structural neuroimaging data a mega-analysis performs better than a meta-analysis. In a multi-center study with a moderate amount

of variation between cohorts, a linear mixed-effects random-intercept mega-analytical framework seems to be the better approach to investigate structural neuroimaging data. We urge researchers worldwide to join forces by sharing data with the goal of elucidating biomedical problems that no group could address alone.

ETHICS STATEMENT

All subjects gave written informed consent in accordance with the Declaration of Helsinki. All local institutional review boards permitted the use of measures extracted from the coded data for meta- and mega-analysis.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2018.00102/full#supplementary-material

REFERENCES

Agresti, A., and Hartzel, J. (2000). Strategies for comparing treatments on a binary response with multi centre data. *Stat. Med.* 19, 1115–1139. doi: 10.1002/(SICI)1097-0258(20000430)19:83.0.CO;2-X

Barnes, J., Ridgway, G. R., Bartlett, J., Henley, S. M., Lehmann, M., Hobbs, N., et al. (2010). Head size, age and gender adjustment in MRI studies: a necessary nuisance? *Neuroimage* 53, 1244–1255. doi: 10.1016/j.neuroimage.2010.06.025

Boedhoe, P. S., Schmaal, L., Abe, Y., Alonso, P., Ameis, S. H., Anticevic, A., et al. (2018). Cortical abnormalities associated with pediatric and adult obsessive-compulsive disorder: findings from the ENIGMA Obsessive-Compulsive Disorder Working Group. *Am. J. Psychiatry* 175, 453–462. doi: 10.1176/appi.ajp.2017.17050485

Boedhoe, P. S., Schmaal, L., Abe, Y., Ameis, S. H., Arnold, P. D., Batistuzzo, M. C., et al. (2017a). Distinct subcortical volume alterations in pediatric and adult OCD: a worldwide meta- and mega-analysis. *Am. J. Psychiatry* 174, 60–69. doi: 10.1176/appi.ajp.2016.16020201

Boedhoe, P. S., Schmaal, L., Mataix-Cols, D., Jahanshad, N., ENIGMA OCD Working Group, Thompson, P. M., et al. (2017b). Association and causation

in brain imaging in the case of OCD: response to McKay et al. *Am. J. Psychiatry* 174, 597–599. doi: 10.1176/appi.ajp.2017.17010019r

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res. Synth. Methods* 1, 97–111. doi: 10.1002/jrsm.12

Burke, D. L., Ensor, J., and Riley, R. D. (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat. Med.* 36, 855–875. doi: 10.1002/sim.7141

de Bakker, P. I., Ferreira, M. A., Jia, X., Neale, B. M., Raychaudhuri, S., and Voight, B. F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* 17, R122–R128. doi: 10.1093/hmg/ddn288

Debray, T. P., Moons, K. G., Abo-Zaid, G. M., Koffijberg, H., and Riley, R. D. (2013). Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PLoS ONE* 8:e60650. doi: 10.1371/journal.pone.0060650

Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021

Dutton, M. T. (2010). *Individual Patient-Level Data Meta-Analysis: A Comparison of Methods for the Diverse Population Collaboration Data Set*. Ph.D. thesis, Florida State University.

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* 72, 320–338. doi: 10.1080/01621459.1977.10480998

Ikram, M. A., Fornage, M., Smith, A. V., Seshadri, S., Schmidt, R., Debette, S., et al. (2012). Common variants at 6q22 and 17q21 are associated with intracranial volume. *Nat. Genet.* 44, 539–544. doi: 10.1038/ng0612-732c

Im, K., Lee, J.-M., Lyttelton, O., Kim, S. H., Evans, A. C., and Kim, S. I. (2008). Brain size and cortical structure in the adult human brain. *Cereb. Cortex* 18, 2181–2191. doi: 10.1093/cercor/bhm244

Kahan, B. C., and Morris, T. P. (2012). Improper analysis of trials randomised using stratified blocks or minimisation. *Stat. Med.* 31, 328–340. doi: 10.1002/sim.4431

Lin, D. Y., and Zeng, D. (2010a). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika* 97, 321–332. doi: 10.1093/biomet/asq006

Lin, D. Y., and Zeng, D. (2010b). Meta-analysis of genome-wide association studies: no efficiency gain in using individual participant data. *Genet. Epidemiol.* 34, 60–66. doi: 10.1002/gepi.20435

Mathew, T., and Nordstorm, K. (2010). Comparison of one-step and two-step meta-analysis models using individual patient data. *Biometric. J.* 52, 271–287. doi: 10.1002/bimj.200900143

Mathew, T., and Nordstrom, K. (1999). On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics* 55, 1221–1223. doi: 10.1111/j.0006-341X.1999.01221.x

Nakagawa, S., and Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* 82, 591–605. doi: 10.1111/j.1469-185X.2007.00027.x

Olkin, I., and Sampson, A. (1998). Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics* 54, 317–322. doi: 10.2307/2534018

Stewart, L. A. (1995). Practical methodology of meta-analyses (overviews) using updated individual patient data. *Stat. Med.* 14, 2057–2079. doi: 10.1002/sim.4780141902

Stewart, L. A., and Tierney, J. F. (2002). To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval. Health Prof.* 25, 76–97. doi: 10.1177/0163278702025001006

Thomas, D., Radji, S., and Benedetti, A. (2014). Systematic review of methods for individual patient data meta- analysis with binary outcomes. *BMC Med. Res. Methodol.* 14:79. doi: 10.1186/1471-2288-14-79

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The ENIGMA Consortium: large-scale collaborative

analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182. doi: 10.1007/s11682-013-9269-5

Van Horn, J. D., Grethe, J. S., Kostelec, P., Woodward, J. B., Aslam, J. A., Rus, D., et al. (2001). The Functional Magnetic Resonance Imaging Data Center (fMRIDC): the challenges and rewards of large-scale databasing of neuroimaging studies. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 356, 1323–1339. doi: 10.1098/rstb.2001.0916

Westlye, L. T., Walhovd, K. B., Dale, A. M., Bjørnerud, A., Due-Tønnessen, P., Engvig, A., et al. (2010). Differentiating maturational and aging-related changes of the cerebral cortex by use of thickness and signal intensity. *Neuroimage* 52, 172–185. doi: 10.1016/j.neuroimage.2010.03.056

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.

Citation: Boedhoe PSW, Heymans MW, Schmaal L, Abe Y, Alonso P, Ameis SH, Anticevic A, Arnold PD, Batistuzzo MC, Benedetti F, Beucke JC, Bollettini I, Bose A, Brem S, Calvo A, Calvo R, Cheng Y, Cho KIK, Ciullo V, Dallaspezia S, Denys D, Feusner JD, Fitzgerald KD, Fouche J-P, Fridgeirsson EA, Gruner P, Hanna GL, Hibar DP, Hoexter MQ, Hu H, Huyser C, Jahanshad N, James A, Kathmann N, Kaufmann C, Koch K, Kwon JS, Lazaro L, Lochner C, Marsh R, Martínez-Zalacaín I, Mataix-Cols D, Menchón JM, Minuzzi L, Morer A, Nakamae T, Nakao T, Narayanaswamy JC, Nishida S, Nurmi EL, O'Neill J, Piacentini J, Piras F, Piras F, Reddy YCJ, Reess TJ, Sakai Y, Sato JR, Simpson HB, Soreni N, Soriano-Mas C, Spalletta G, Stevens MC, Szeszko PR, Tolin DF, van Wingen GA, Venkatasubramanian G, Walitza S, Wang Z, Yun J-Y, ENIGMA-OCD Working-Group, Thompson PM, Stein DJ, van den Heuvel OA and Twisk JWR (2019) An Empirical Comparison of Meta- and Mega-Analysis With Data From the ENIGMA Obsessive-Compulsive Disorder Working Group. Front. Neuroinform. 12:102. doi: 10.3389/fninf.2018.00102

Homogenizing Estimates of Heritability Among SOLAR-Eclipse, OpenMx, APACE, and FPHI Software Packages in Neuroimaging Data

Peter Kochunov[1*†], Binish Patel[1†], Habib Ganjgahi[2†], Brian Donohue[1], Meghann Ryan[1], Elliot L. Hong[1†], Xu Chen[3], Bhim Adhikari[1], Neda Jahanshad[4], Paul M. Thompson[4], Dennis Van't Ent[5], Anouk den Braber[5], Eco J. C. de Geus[5], Rachel M. Brouwer[5], Dorret I. Boomsma[5], Hilleke E. Hulshoff Pol[6], Greig I. de Zubicaray[7], Katie L. McMahon[8], Nicholas G. Martin[9], Margaret J. Wright[9,10] and Thomas E. Nichols[11]

[1]Maryland Psychiatric Research Center, Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, United States, [2]Department of Statistics, University of Oxford, Oxford, United Kingdom, [3]Department of Cognitive Neuroscience, Maastricht University, Maastricht, Netherlands, [4]Imaging Genetics Center, Keck School of Medicine of USC, Marina del Rey, CA, United States, [5]Department of Biological Psychology, VU University, Amsterdam, Netherlands, [6]Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, Netherlands, [7]Faculty of Health, and Institute of Health and Biomedical Innovation, Queensland University of Technology (QUT), Brisbane, QLD, Australia, [8]Centre for Advanced Imaging, University of Queensland, Brisbane, QLD, Australia, [9]QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia, [10]Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia, [11]Big Data Institute, University of Oxford, Oxford, United Kingdom

Imaging genetic analyses use heritability calculations to measure the fraction of phenotypic variance attributable to additive genetic factors. We tested the agreement between heritability estimates provided by four methods that are used for heritability estimates in neuroimaging traits. SOLAR-Eclipse and OpenMx use iterative maximum likelihood estimation (MLE) methods. Accelerated Permutation inference for ACE (APACE) and fast permutation heritability inference (FPHI), employ fast, non-iterative approximation-based methods. We performed this evaluation in a simulated twin-sibling pedigree and phenotypes and in diffusion tensor imaging (DTI) data from three twin-sibling cohorts, the human connectome project (HCP), netherlands twin register (NTR) and BrainSCALE projects provided as a part of the enhancing neuro imaging genetics analysis (ENIGMA) consortium. We observed that heritability estimate may differ depending on the underlying method and dataset. The heritability estimates from the two MLE approaches provided excellent agreement in both simulated and imaging data. The heritability estimates for two approximation approaches showed reduced heritability estimates in datasets with deviations from data normality. We propose a data homogenization approach (implemented in solar-eclipse; www.solar-eclipse-genetics.org) to improve the convergence of heritability estimates across different methods. The homogenization steps include consistent regression of any nuisance covariates and enforcing normality on the trait data using inverse Gaussian transformation. Under these conditions,

the heritability estimates for simulated and DTI phenotypes produced converging heritability estimates regardless of the method. Thus, using these simple suggestions may help new heritability studies to provide outcomes that are comparable regardless of software package.

INTRODUCTION

Reproducibility is the cornerstone of scientific research. Recent reports on low reproducibility in biomedical research are raising concerns that have to be addressed within the scientific community (Ioannidis, 2014). The emerging field of imaging genetics is not immune to these challenges[1]. Imaging genetics applies modern statistical genetics methods to quantitative phenotypes extracted from high dimensional neuroimaging modalities and has to address replication challenges in both imaging and genetic domains (Thompson et al., 2010). Challenges in replication include low statistical power, complexity of analysis, large number of dependent variables, statistical complexity, and differences in the analysis approaches and software (Meyer-Lindenberg et al., 2008; Collins and Tabak, 2014). All these challenges apply to imaging genetics studies. Imaging genetic studies look for factors that typically explain a small proportion of variance (<1%) and may require a large sample sizes (N = 1,000–100,000) to be statistically powerful (Thompson et al., 2014). Imaging genetic studies employ complex analyses involving both imaging and genetic specialized analysis software (Meyer-Lindenberg et al., 2008). We tested the agreement between heritability estimates provided by four methods that are used for heritability estimates in neuroimaging traits. We demonstrated that the heritability estimates may vary by method and sample and propose a way to homogenize the outcomes.

The incomplete description of methods and low statistical power are the two chief factors that are likely contributing to the lack of reproducibility in imaging genetics studies (Collins and Tabak, 2014). Imaging genetic studies combine methods from both imaging and genetic disciplines. These studies require software for extraction of imaging phenotypes and software for genetic analyses of imaging traits, each having individual operating characteristics. For example, the outputs of imaging and genetic software may differ between versions of the same analysis software and even with the same version of software on different operating systems (Gronenschild et al., 2012). Imaging genetic analyses may also suffer from low power because the contribution from common variations in genome to phenotypic variability is typically small (~0.1%), thus requiring large samples to achieve significance and obtain reproducible results (Flint and Munafò, 2013). This further underscores the need for a careful study of the potential biases among different software analysis tools. These methodological biases may lead to challenges to replicate imaging genetic findings if in-kind imaging or genetic software is used during replication.

To address method-related biases, large consortia such as enhancing neuro imaging genetic meta analyses (ENIGMA) have developed standardized multi-site phenotype extraction and genetic analyses pipelines. In this manuscript, we consider the impact of analysis method for the estimation of heritability. We compared four approaches: two commonly used genetic analysis packages (SOLAR-Eclipse and OpenMx), and two recently developed accelerated heritability estimation methods [accelerated permutation inference for ACE (APACE), and fast permutation heritability inference (FPHI)]. These packages use the same variance component model and definition of heritability, but use different numerical methods and data preprocessing steps to calculate the proportion of variance attributed to additive genetic factors. We performed this study to (A) analyze if heritability estimates derived by the four packages' analyses are comparable to one another and; (B) develop a homogenization approach that minimizes the variability in heritability estimates across the four packages.

We performed these analyses using two datasets: a simulated—with known additive genetic contribution and an experimental—consisting of fractional anisotropy (FA) measurements collected in twins and siblings by three independent studies. FA is the most commonly analyzed scalar parameter extracted from diffusion tensor imaging (DTI; Basser et al., 1994; Basser and Pierpaoli, 1996) and is a sensitive index of fiber coherence, myelination levels, and axonal integrity (Thomason and Thompson, 2011). FA values are under a strong genetic control (Geng et al., 2012; Jahanshad et al., 2013; Shen et al., 2014). Individual differences in FA values are predictive of cognitive performance (Kochunov et al., 2016, 2017) and it is a promising phenotype for multiple neuropsychological disorder including schizophrenia (Friedman et al., 2008; Perez-Iglesias et al., 2010; Alba-Ferrara and de Erausquin, 2013; Kochunov et al., 2013; Mandl et al., 2013; Nazeri et al., 2013). All experimental data were processed using the harmonization protocol previously developed by ENIGMA and provided on-line at http://enigma.ini.usc.edu/ongoing/dti-working-group/). This included the use of the ENIGMA protocol for following the QA/QC steps for each site, registration to the ENIGMA-DTI target, extraction of white matter skeleton, followed by extraction of tract-average FA values.

MATERIALS AND METHODS

Heritability Estimation Methods

We evaluated the agreement in quantification of the Additive genetic and Environmental, AE, components of the phenotypic variance in simulated and imaging genetic datasets among

---

[1]http://www.biorxiv.org/content/early/2017/02/20/107987

four heritability calculation methods. SOLAR-Eclipse[2] and OpenMx[3] use the iterative maximum likelihood estimation (MLE) approach to fit quantitative genetics variance components models. The iterative MLE approach is used to determine the parameters that maximize the compatibility between the fitted model and the data. It is a versatile computational approach that produces estimates that are optimally precise asymptotically (Almasy and Blangero, 1998; Blangero et al., 2001). SOLAR-Eclipse is an extensive and flexible imaging genetics analysis software package. SOLAR-Eclipse functions include calculation of heritability, genetic correlation, linkage and genome-wide association analysis (Almasy and Blangero, 1998; Blangero et al., 2001). SOLAR-Eclipse polygenic function uses MLE to perform genetic analyses in the pedigrees of arbitrary size and complexity, including twin-siblings and complex multigenerational family designs. SOLAR-Eclipse is frequently used in imaging genetic studies especially in the multi-site analyses that aggregate measurements across multiple datasets using meta and mega-analyses (Jahanshad et al., 2013; Kochunov et al., 2014, 2015). OpenMx is an extensive and flexible structural equation modeling and path analysis library for [R] software (Boker et al., 2011). OpenMX is frequently used by imaging genetic studies to calculate heritability and genetic correlation in twin-siblings pedigrees (Jahanshad et al., 2010; Bootsman et al., 2016). Like SOLAR-Eclipse, OpenMx uses an iterative MLE method for calculation of heritability parameters.

APACE model and FPHI use statistical approximations to estimate heritability values. APACE uses a regression approach based on the squared differences of twin pairs, a variant of a U-statistic (Chen et al., 2013; Chen, 2014), while FPHI starts with the same likelihood as used in SOLAR-Eclipse but uses a single-step, rather than iterative, optimization (Ganjgahi et al., 2015). This overcomes the main limitation of the MLE-based software: long computational times. The iterative MLE heritability calculations in SOLAR-Eclipse and OpenMx can take ∼1 s per trait in a pedigree of 1,000 subjects. Therefore, MLE-based heritability analyses require access to large computational clusters to perform imaging genetic analyses that involve 10^4–10^6 voxel-wise traits. The non-iterative estimates from APACE and FPHI offer appreciable (∼10^3) gains in computational efficiency. This allows performing voxel-wise heritability analyses on a single workstation. While APACE is only intended for twin or twin-plus-sibling designs, FPHI can use any kinship structure, like SOLAR-Eclipse.

The four software packages were used to compare additive genetic contribution (heritability) in simulated and experimental data using twin family study designs. For experimental data we used DTI acquisitions from three different studies. The human connectome project (HCP; Van Essen et al., 2012), is a large-scale international collaboration aimed at elucidating the genetic and environmental sources of normal variability within the structural and functional connections of the human brain. The other two twin and sibling datasets were drawn

from the ENIGMA project, specifically from the ENIGMA-DTI workgroup whose focus is the analysis of DTI data. The first of these is the netherlands twin register (NTR) that collected DTI data in normally developing adolescent twins and siblings. And the other ENIGMA-DTI source is the Brain Structure and Cognition: an Adolescent Longitudinal Twin Study into Genetic Etiology (BrainSCALE). The BrainSCALE dataset collected DTI data in young adult twins and siblings. Subjects for NTR and BrainSCALE datasets were recruited from the same twin register in Netherlands.

We compare heritability estimates for tract-wise average FA values using ENIGMA-DTI, HCP, and simulated data. FA is a widely used quantitative measure of white matter microstructure (Basser et al., 1994; Basser and Pierpaoli, 1996) calculated from the diffusion tensor model of water diffusion (Thomason and Thompson, 2011). Studies suggest FA is an important biomarker in clinical studies, since it is a sensitive index of white matter integrity in Alzheimer's disease (Clerx et al., 2012; Teipel et al., 2012), general cognitive function (Penke et al., 2010a,b), and several neurological and psychiatric disorders (Sprooten et al., 2011; Barysheva et al., 2012; Carballedo et al., 2012; Kochunov et al., 2013; Mandl et al., 2013). Overall, our goal was to determine if additive genetic contribution (heritability) is comparable between software packages regardless of the variation in the twin-sibling cohort data. Our hypothesis was that estimates of heritability would be consistent amongst the cohorts, irrespective of the variability in cohort data and software package.

Simulated Data

A simulated N = 1,000 person twin-sibling pedigree with 250 monozygotic (MZ) twins, 250 dizygotic (DZ) twins, and 500 founders (not included in the phenotype file) was created using SOLAR-Eclipse simulate function. SOLAR-Eclipse simulation functionality was also used to produce a data set of 10,000 traits with heritability estimates varied uniformly between 0 and 100%. All simulated traits had normal distribution and did not include effects of covariates.

Experimental Data

Human Connectome Project (HCP)

- Subjects: the cohort contained 481 (194/287 M/F; average age 29.1 ± 3.5) healthy participants of the HCP for whom the scans and data were released in June 2014 (humanconnectome.org) after passing the HCP quality control and assurance standards (Marcus et al., 2013). The participants in the HCP study were recruited from the Missouri Family and Twin Registry, a large population-based study (Van Essen et al., 2012). This release included 117 twin pairs (57 MZ and 60 DZ pairs), and 246 of their siblings. The full set of inclusion and exclusion criteria is detailed elsewhere (Van Essen et al., 2012).

- Imaging: diffusion data was collected at Washington University in St. Louis using a customized Siemens Magnetom Connectome 3-Tesla scanner with a 100 mT/m maximum gradient strength and a 32-channel head coil. Details on the scanner, image acquisition and reconstruction are provided

in Ugurbil et al. (2013)[4]. Diffusion data were collected using a single-shot, single refocusing spin-echo, echo-planar imaging sequence with 1.25 mm isotropic spatial resolution (TE/TR = 89.5/5520 ms, FOV = 210 × 180 mm). Three gradient tables of 90 diffusion-weighted directions and six $b = 0$ images each, were collected with right-to-left and left-to-right phase encoding polarities for each of the three diffusion weightings ($b = 1,000, 2,000,$ and $3,000$ s/mm^2). The total imaging time for collection of diffusion data was approximately 1 h.

Netherlands Twin Register (NTR)

- Subjects: the cohort consisted of 246 adults (93/153 M/F; average age 33.9 ± 10.1, range 19–57), recruited from the NTR and consisted of 72 MZ pairs, 48 DZ pairs, and six siblings. Exclusion criteria consisted of having any metal material in the head, having a pacemaker, a history of any major medical conditions or psychiatric illness (den Braber et al., 2008, 2011, 2012).
- Imaging: DTI data were collected on a 3-Tesla Philips Intera MR scanner (32 diffusion-weighted volumes with different non-collinear diffusion directions with b-factor = 1,000 s/mm^2 and one b-factor = 0 s/mm^2 image, flip angle = 90 degrees; 38 axial slices of 3.0 mm; no slice gap; voxel size, 2.0 × 2.0 × 3.0 mm; FOV = 230 mm; TE = 94 ms; TR = 4,863 ms; no cardiac gating; and total scan duration = 185 s).

Brain Structure and Cognition: An Adolescent Longitudinal Twin Study into Genetic Etiology (BrainSCALE)

- Subjects: the sample comprised of 199 children (100M/99F; average age 9.2 ± 0.1, range 9.0–9.6). It included 42 MZ and 57 DZ twin pairs that were recruited from families participating in the BrainSCALE cohort (van Soelen et al., 2012) that were recruited from the NTR (van Beijsterveldt et al., 2013). Exclusion criteria consisted of having any metal material in the head, having a pacemaker, a known history of any major medical condition or psychiatric illness. Zygosity was determined based on DNA polymorphisms, using 8–11 highly polymorphic di-, tri- and tetranucleotide genetic markers and confirmed by genome-wide single nucleotide polymorphism data.
- Imaging: DTI data were collected on a 1.5 Philips Achieva MR scanner (32 diffusion-weighted volumes with different non-collinear diffusion directions with b-factor = 1,000 s/mm^2 and eight diffusion-unweighted volumes with b-factor = 0 s/mm^2; parallel imaging SENSE factor = 2.5; flip angle = 90 degrees; 60 slices of 2.5 mm; no slice gap; 96 × 96 acquisition matrix; reconstruction matrix 128 × 128; FOV = 240 mm; TE = 88 ms; TR = 9,822 ms; two repetitions; no cardiac gating; and total scan duration = 296 s). More information may be found in Brouwer et al. (2010, 2012).

---

[4]https://www.humanconnectome.org/documentation/S500/HCP_S500_Release_Reference_Manual.pdf

ENIGMA-DTI Processing

We used ENIGMA-DTI protocol to extract whole-brain and tract-wise average FA values for experimental datasets. These protocols are detailed elsewhere (Jahanshad et al., 2013) and are available online at http://enigma.ini.usc.edu/protocols/dti-protocols/. In brief, FA images from all subjects were non-linearly registered to the ENIGMA-DTI target FA image using FSL's FNIRT (Smith et al., 2006). This target was created as a minimal deformation target based on images from the participating studies as previously described (Kochunov et al., 2002; Jahanshad et al., 2013). The data were then processed using FSL's tract-based spatial statistics (TBSS) analytic method (Smith et al., 2006) modified to project individual FA values onto the ENIGMA-DTI skeleton mask. After extracting the skeletonized white matter and the projection of individual FA values, ENIGMA tract-wise regions of interest (ROIs), derived from the Johns Hopkins University (JHU) white matter parcellation atlas available as a part of FSL, were transferred to extract the mean FA across the full skeleton and average FA values for major white matter tracts. The protocol, target brain, ENIGMA-DTI skeleton mask, source code and executables are all publicly available[5]. This protocol was shown to provide highly replicable measurements based on test-rest analyses in human subjects (Acheson et al., 2017; McGuire et al., 2017).

Inverse Normal Transformation

Multivariate quantitative trait models are sensitive to outliers, skewness, kurtosis and other deviations from normal distribution. Therefore, we consider the use of a rank-based inverse normal transformation to ensure the normal distribution in quantitative traits. For each phenotype, rank values are replaced with the expected ranked values of a standard normal distribution with the same number of observations. While it cannot ensure multivariate normality, it does ensure that each univariate distribution is normal and thus reduces the impact of outliers; for more discussion on this transformation see (Beasley et al., 2009). We implemented inverse normalization in SOLAR-Eclipse as the "polyclass_normalize" functions. This function produces inverse normalized residuals for the trait after regression of all covariates. The output from this function was used for the secondary analyses of the imaging data where we first analyze the raw data and then compare our results after the application of the inverse normal transformation to the residual data.

Heritability Analysis

Heritability analyses were performed in the simulated and FA traits. Heritability (h^2) is the proportion of the total phenotypic variance (σ_P^2) that can be explained by the genetic effects of genes (σ_g^2),

$$h^2 = s_g^2/s_P^2 \tag{1}$$

MLE Based Analysis

SOLAR-Eclipse and OpenMX employ MLE based variance decomposition approach that is an extension of the strategy

---

[5]https://www.nitrc.org/projects/enigma_dti

developed by Amos (1994). The multivariate normal covariance matrix Ω for a pedigree of individuals is given by

$$\Omega = 2 \cdot \Phi \cdot s_g^2 + I \cdot s_e^2 \qquad (2)$$

where Φ is the kinship matrix representing the pair-wise kinship coefficients among related individuals, σ_e^2 is the variance due to individual-specific environmental effects, and I is an identity matrix (under the assumption that all environmental effects are uncorrelated among family members). Narrow sense heritability is defined as the fraction of phenotypic variance σ_P^2 attributable to additive genetic factors. In twin designs a third variance parameter is can be identified and may be added to the model, σ_c^2, for the common environment shared by twins and siblings growing up in the same family. This three-parameter model is known as the ACE model, while the two-parameter model (Equation 2) is referred to as the AE model.

The variance parameters are estimated by comparing the observed phenotypic covariance matrix with the covariance matrix predicted by kinship (Almasy and Blangero, 1998). Significance of heritability is tested by comparing the likelihood of the model in which σ_g^2 is constrained to zero with that of a model in which σ_g^2 is estimated. Twice the difference between the $\log_e$ likelihoods of these models yields a test statistic, which is asymptotically distributed as a 1/2:1/2 mixture of a X^2 variable with 1 degree-of-freedom and a point mass at zero.

The Accelerated Permutation for the ACE Model (APACE)

APACE[6] uses an approximation technique developed originally for animal genetics studies (Grimes and Harvey, 1980) and is based on the result that squared differences of pair's of subjects' data reflect their covariance. Thus, the squared differences among the DZ, MZ and unrelated subjects can be entered into a linear regression model to estimate the variance parameters (Grimes and Harvey, 1980). The speed advantage of APACE over MLE approaches allows a permutation analysis to compute familywise error corrected P-values for voxel-wise imaging measures.

Fast Permutation Heritability Inference (FPHI)

SOLAR-Eclipse's iterative MLE approach is accelerated by the use of a data transformation based on the eigenvectors of the kinship matrix Φ (Blangero et al., 2013). This transformation converts the dependent data from related subjects into data that is independent but has heterogeneous-variance. SOLAR-Eclipse uses this simplified model to obtain iterative MLE estimates using linear regressions. The FPHI approach uses the same likelihood and data transformation, but then performs just a single step estimation to produce an asymptotically unbiased estimate (Ganjgahi et al., 2015). The FPNI technique is implemented SOLAR-Eclipse as the CPU and graphics processing unit (GPU) functions. The CPU version of FPHI provides a significant (10^3) computational acceleration relative to the iterative MLE estimation in SOLAR-Eclipse, while the graphics processing unit (GPU) version further improves this performance ($\sim 10^6$) vs. iterative MLE approach.

[6]http://warwick.ac.uk/tenichols/software/APACE

All analyses with imaging data were conducted with age, sex, age^2, age $\times$ sex, and age$^2 \times$ sex included as covariates.

RESULTS

Heritability Analyses—Simulated

Figure 1 shows the scatter plots of four methods using a simulated dataset of heritability values distributed between 0 and 1. The two ML-based methods (SOLAR-Eclipse and OpenMX) showed an excellent agreement ($r = 0.999$, slope $= 1.000$, intercept $= 0.000$) with the expected heritability values and with each other (**Figure 1**). We quantified bias as estimated h^2 minus true h^2 and "average spread" as the absolute bias divided by true value (i.e., |estimated $-$ true|/true). In the simulated dataset, the two ML-based methods show zero bias (absolute value bias $<10^{-6}$) and the average spread in heritability estimates of 1.2%. The APACE and FPHI methods showed excellent overall agreement with expected values (APACE: absolute value of bias $= 10^{-5}$, $r = 0.997$, slope $= 0.997$, intercept $= 0.005$; FPHI: absolute value of bias $= 10^{-6}$, $r = 0.998$, slope $= 0.999$, intercept $= 0.001$). APACE showed significantly higher average spread than the FPHI method: 3.7 vs. 2.2% ($p = 10^{-10}$).
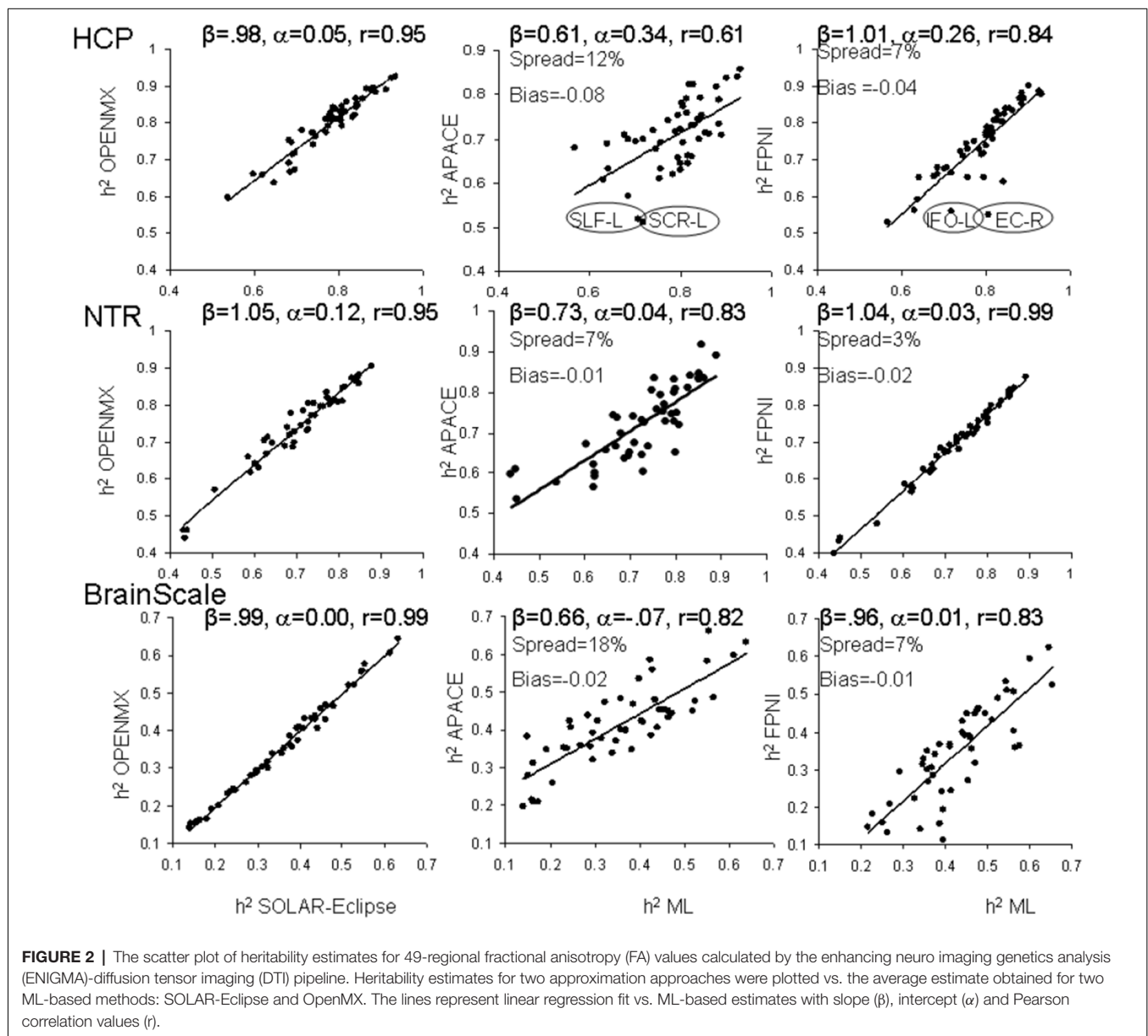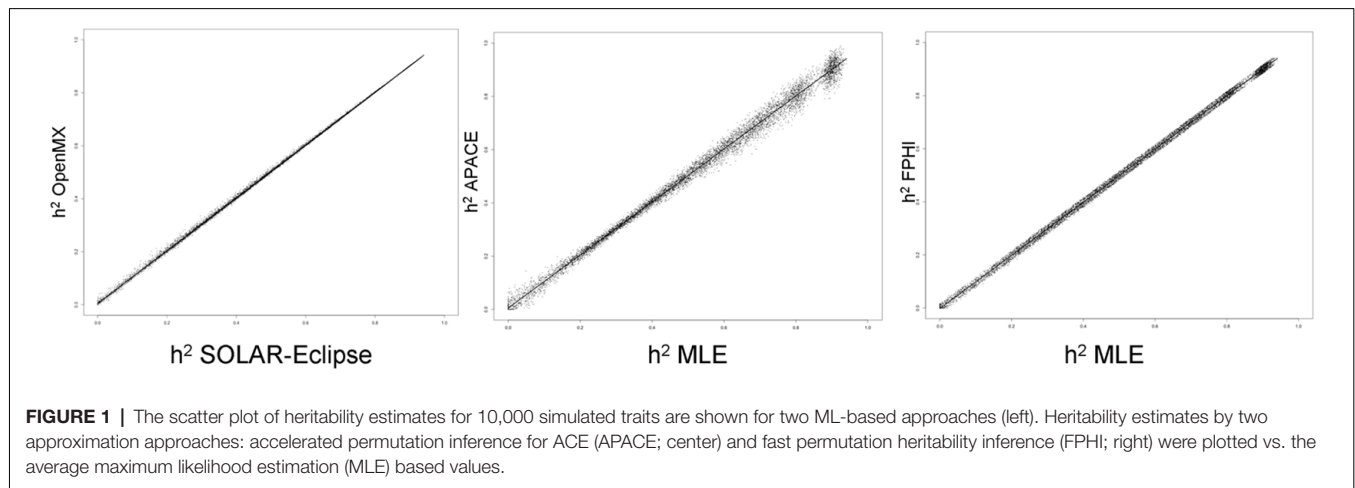
Heritability Analyses—Diffusion Data

The heritability analyses were performed in FA data for 49 tracts in HCP, NTR and BrainScale cohorts using age, sex, age^2, age $\times$ sex, and age$^2 \times$ sex as covariates. The two ML-based method showed excellent agreement in all three datasets (**Figure 2**). The best agreement was observed in BrainScale data ($r = 0.99$, slope $= 0.99$, intercept $= 0.001$). The least agreement ($\sim$5% average spread) between two ML-based approaches was observed in HCP ($r = 0.95$, slope $= 1.05$, intercept $= 0.121$). Intermediate results were observed in the NTR dataset ($r = 0.98$, slope $= 0.98$, intercept $= 0.055$). Hence, we averaged the heritability values produced by the two ML methods to create a "ground truth" reference for the two approximation methods.

The heritability estimates provided by the approximation approaches were more variable among three cohorts (**Figure 2**). The FPHI showed better accuracy in variance in slopes ($\beta = 0.97$–1.04) and intercepts ($\alpha = 0.01$–0.26) vs. APACE ($\beta = 0.61$–0.73 and intercepts $\alpha = -0.07$–0.34; **Figure 2**). Both FPNI and APACE showed a modest negative bias. The highest bias was seen for the HCP cohort (-0.08 and -0.04 for APACE and FPNI, respectively). The bias in NTR and BrainSCALE cohorts was small (-0.01 and -0.02). The spread for FPHI was about half that for APACE (6% vs. 12% for FPHI and APACE, respectively).

Heritability Analyses—Normalized Diffusion Data

Next, heritability estimates were calculated on the residual data after inverse normal transformation (**Figure 3**). Trait normalization improved agreement among the ML-based methods ($r = 0.96$–0.99, slope $= 0.99$–1.00, intercept $= 0.00$–0.02; **Figure 3**).

FIGURE 1 | The scatter plot of heritability estimates for 10,000 simulated traits are shown for two ML-based approaches (left). Heritability estimates by two approximation approaches: accelerated permutation inference for ACE (APACE; center) and fast permutation heritability inference (FPHI; right) were plotted vs. the average maximum likelihood estimation (MLE) based values.

FIGURE 2 | The scatter plot of heritability estimates for 49-regional fractional anisotropy (FA) values calculated by the enhancing neuro imaging genetics analysis (ENIGMA)-diffusion tensor imaging (DTI) pipeline. Heritability estimates for two approximation approaches were plotted vs. the average estimate obtained for two ML-based methods: SOLAR-Eclipse and OpenMX. The lines represent linear regression fit vs. ML-based estimates with slope (β), intercept (α) and Pearson correlation values (r).

FIGURE 3 | The scatter plot of heritability estimates for 49-regional FA values calculated by ENIGMA-DTI pipeline and then normalized using the trait normalization function in SOLAR-Eclipse. Heritability estimates for two approximation approaches were plotted vs. the average estimate obtained for two ML-based methods: SOLAR-Eclipse and OpenMX. The lines represent linear regression fit vs. ML-based estimates with slope (β), intercept (α) and Pearson correlation values (r).

Trait normalization brought improvements in the agreement between the estimates by two approximation approaches and the average ML-based estimation (**Figure 3**). APACE method showed improvements in slope (β = 0.75–1.01), intercept (α = −0.05–0.23) and correlation coefficients (r = 0.76–0.95), in all three cohorts. For FPHI, the improvements were more subtle and were mainly observed as decrease in bias and spread. The bias for APACE increased for NTR cohort (from −0.01 to 0.08). Both approximation methods showed a 50% improvement in the percentage spread vs. the average ML-based estimate, yet, the % spread for FPHI remained about half that for APACE (4% vs. 7.6% for FPHI and APACE, respectively).

Analysis of the Disagreement

We tested the normality of the distribution of the neuroimaging traits using the Shapiro–Wilk method, focusing on the HCP dataset because it had the largest number of subjects. We observed that four traits: the anterior limb of internal capsule-left (ALIC-L), uncinate fasciculus-right (UNC-R), external capsule-right (EC-R) and superior corona radiate-left (CR-L), failed the

null hypothesis for normal distribution ($W > 0.94$, $p < 0.05$; **Figure 2**). However, there was no significant correlation between the deviation from normality or the heritability values for any of the four methods (all $r < 0.20$, all $p > 0.4$). Furthermore, some traits that visibly contributed to dispersion of heritability values, for example the inferior fronto-occipital tract-left (IFO-L) and superior corona-radiata-right (SCR-L; **Figure 2**), passed the Shapiro–Wilk test ($p > 0.10$). The histograms for SLF-L and SCR-L showed only modest kurtosis (kurtosis = −0.2 and 0.15 for SLF-L and SCR-L), but visibly varied from the normal distribution (**Figure 4**). The histograms for IFO-L varied visibly from a normal distribution despite having low kurtosis (0.13), while EC-R had high kurtosis (7.7; **Figure 4**).

DISCUSSION

We conducted a careful evaluation of four quantitative genetic approaches used by imaging genetic studies to measure heritability—the proportion of variance attributable to the additive genetic factors. Two of the methods (SOLAR-Eclipse
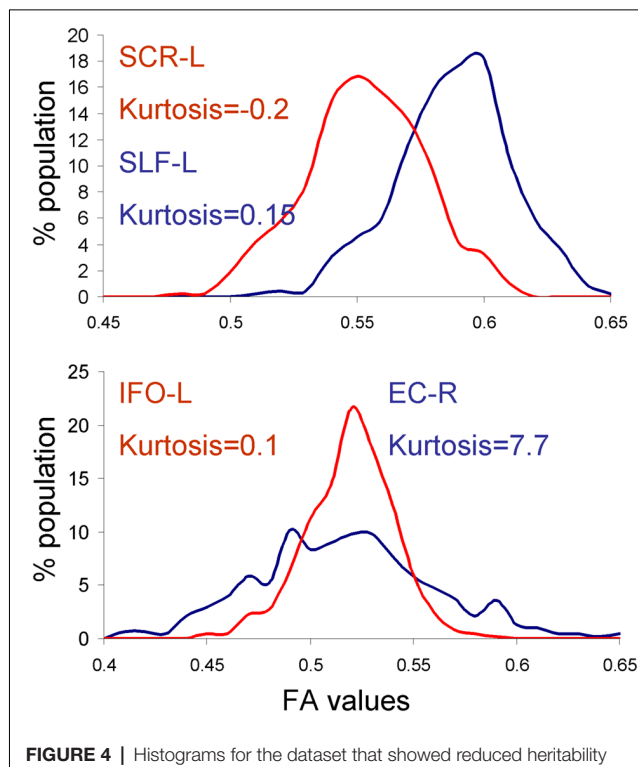
FIGURE 4 | Histograms for the dataset that showed reduced heritability estimates for fast vs. MLE based heritability estimation approaches. APACE showed reduced heritability estimates in superior corona-radiata-right (SCR-L) and SLF-L tracts in the human connectome project (HCP) cohort due to deviations from normal distribution (top panel). FPHI showed reduced heritability estimates in the external capsule-right (EC-R) and inferior fronto-occipital tract-left (IFO-L) tracts in the HCP cohort due to the high kurtosis and non-Gaussian shape of the histograms for EC-R and IFO-L, respectively.

and OpenMX) used an iterative MLE approach. Two methods (APCE and FPNI) were developed specifically to accelerate (by 10^{3-6}) voxel-wise imaging genetics analyses using fast approximation approaches. We performed the evaluation in a simulated dataset and imaging data from three independent datasets. In the simulated data, we observed an excellent agreement between all heritability estimate approaches. The two MLE approaches accurately replicated the expected heritability values, with the unity slope and near zero intercept and measurement bias. The two approximation techniques likewise showed excellent agreement in the simulated data, with only slight spread (2.2% and 3.7% for FPNI and APACE, respectively). In neuroimaging data, the two MLE approaches produced consistent estimates of heritability for all cohorts. We used the average MLE as the reference measures for approximation techniques because the true additive genetic contribution is unknown (Parisi et al., 2014). In the neuroimaging data, the approximation methods showed deviations from MLE values that varied by the dataset and method. The approximation methods showed the best consistency for NTR and the lowest consistency in the HCP data. *Post hoc* analyses attempted to identify the sources of the dispersion based on the underlying distribution in imaging data. The heritability values were not significantly correlated with Shapiro-Wilk's W-value for any

method or dataset (all $r < 0.2$). However, the traits with high dispersion in heritability estimates did show deviations from normality in the underlying dataset. The heritability estimates produced by the FPHI approach were generally closest to that produced by MLE estimates. The agreement among all methods was significantly improved following data normalization approach that ensured normality for quantitative traits. This data normalization approach is now available as a part of SOLAR-Eclipse distribution.

Imaging genetics is a field that combines imaging and genetics—the two disciplines that have greatly advanced neuroscience in recent years. The replication challenges are not unique to this new field and require concerted efforts to address them. The main replication challenges that imaging genetics faces are the complexity of the methods and the low statistical power (Meyer-Lindenberg et al., 2008; Collins and Tabak, 2014). Genetic factors may explain a small proportion of variance that require a sample sizes that are challenging to collect in a single study ($N = 1,000–100,000$; Stein et al., 2010, 2012; Thompson et al., 2014). Yet, imaging genetics approaches have many advantages that should help in overcoming this challenge. Modern MRI offers phenotypic measurements that provide more detailed and quantitative descriptions than disorder diagnostic status or clinical symptoms. Modern MRI phenotypes offer high precision and reproducibility with the inter-session, scan-rescan variability of many common imaging measurements in the range of 1%–5% (Agartz et al., 2001; Kim et al., 2005; Lerch and Evans, 2005; Kochunov and Duff Davis, 2009; Acheson et al., 2017). Therefore, the solution to statistical power is meta-analyses that combine data across multiple studies.

ENIGMA, Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) and other multi-study initiatives aim to overcome the challenge of limited power by performing meta-analytical analyses. In these initiatives, phenotypic and genetic analyses are performed by individual sites and meta-analytical aggregation is used to derive the overall estimates of genetic effects. The main challenge in this approach is overcoming the diversity and complexity of analytical and statistical approaches that may lead to variance in phenotype extractions and estimation of effect sizes (Meyer-Lindenberg et al., 2008; Collins and Tabak, 2014). This complexity exists on both imaging and genetic sides where the difference in analysis software and even versions of software may lead to varying results (Gronenschild et al., 2012). On the phenotype extraction side, ENIGMA provides the standardized pipeline for extraction of homogenized neuroimaging phenotypes across the sites (Jahanshad et al., 2013). Here, we demonstrate the need of homogenized treatments of the traits to avoid erroneous variances at the meta-analytical state.

In our evaluations, we observed excellent agreement between estimates produced by the two MLE-based approaches that were the corner stone of imaging genetic research in the past. The main disadvantage of MLE approaches is the long calculation times associated with the iterative maximization of the likelihood. In imaging genetic studies, up to a million voxel-based imaging traits may be analyzed (Stein et al., 2011), making MLE approaches less practical. Voxel-wise analyses require a

permutation-based correction for multiple comparisons because standard multiple comparison approaches are deemed to be too conservative for voxel-wise traits (Nichols and Hayasaka, 2003). Therefore, there is a need for fast and accurate methods to estimate genetic variance where the calculations can be repeated with 10^{5-6} permutations to derive cluster-based significance on the voxel-wise levels. We measured the performance of two such methods (APACE and FPNI) that use approximation to obtain fast inference of genetic variance.

APACE and FPNI use data transformation and approximation fits to accelerate the calculation of genetic parameters. APACE uses a squared difference in phenotype values between pairs of related and unrelated subjects to derive the fraction of variance contributable to the additive genetic variance. This approach is appropriate for twins and siblings pedigree. FPNI uses the eigenvalue decomposition followed by a single step approximation to calculate genetic variance in pedigrees of any complexities. The approximation approaches demonstrated an excellent performance in the simulated dataset where the trait data was normally distributed. However, their performance in the imaging data was less uniform, likely due to sensitivity to noise and violations of the normality assumption.

The two MLE approaches appeared to provide more stable estimates of heritability in datasets with noise and the non-normally distributed traits, while these deviations had a greater impact on the heritability estimates produced by the approximation approaches. In the cases where the trait's distribution deviated from normality, the heritability values calculated by the approximation techniques deviated from those calculated by ML-based approaches. However, the correlation between heritability values and the deviation of normality (Shapiro-Wilk's W) was not significant. We explored four cases of visible outliers. Some traits (ALIC-L, UNC-R, EC-R and CR-L) failed assumptions for normality, but other outliers passed normality according to Shapiro-Wilk's test. We concluded that approximation approaches may be more sensitive to the noise and deviation from data normality and may produce biased heritability estimates even in traits whose distributions pass the standard tests for normality.

We found that the use of inverse normal transformation improved the agreement between ML and approximation-based approaches and resolved the outlier heritability estimates observed in uncorrected data. The inverse normal transformation did not alter the pattern of ML-based estimates: high correlation ($r > 0.95$) was observed for averaged ML-estimates before and after inverse normal transformation. Enforcing normality upon data reduced the dispersion in h^2 values and improved the average spread for the approximation approaches. This was especially noticeable for FPNI approach where the correlations with ML-estimates became high ($r > 0.97$) for all cohorts.

LIMITATION

The ML estimations were used as the reference to compare the performance of approximation-based approaches in the simulated and imaging data. The two ML approaches produced convergent heritability estimates in both simulated and imaging datasets. However, this does not constitute the "ground truth" especially in imaging datasets where ML approaches may be biased despite convergence.

CONCLUSION

We have conducted a careful comparison of four heritability estimation methods for imaging data. Based on "ground-truth" simulations, four packages can produce low-bias, low-variance heritability estimates, with ML-based methods understandably performing slightly better than the approximation methods. In real data, the approximation methods exhibit more variability relative to the ML-based methods, but this variability was reduced with the use of a rank-based inverse normal transformation, suggesting that this may be an important tool to maximize inter-method reliability.

ETHICS STATEMENT

This study performed secondary data analyses in anonymized human subjects.

AUTHOR CONTRIBUTIONS

PK, BP, HG, BD, MR, XC, NJ, PT and TN designed experiment, performed analyses, and wrote the manuscript. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

REFERENCES

Acheson, A., Wijtenburg, S., Rowland, L., Winkler, A., Mathias, C. W., Hong, L., et al. (2017). Reproducibility of tract-based white matter microstructural measures using the ENIGMA-DTI protocol. *Genes Brain Behav.* 7:e00615. doi: 10.1002/brb3.615

Agartz, I., Okuguwa, G., Nordström, M., Greitz, D., Magnotta, V., and Sedvall, G. (2001). Reliability and reproducibility of brain tissue volumetry from segmented MR scans. *Eur. Arch. Psychiatry Clin. Neurosci.* 251, 255–261. doi: 10.1016/s1053-8119(01)92358-4

Alba-Ferrara, L. M., and de Erausquin, G. A. (2013). What does anisotropy measure? Insights from increased and decreased anisotropy in selective fiber tracts in schizophrenia. *Front. Integr. Neurosci.* 7:9. doi: 10.3389/fnint.2013.00009

Almasy, L., and Blangero, J. (1998). Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* 62, 1198–1211. doi: 10.1086/301844

Amos, C. I. (1994). Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* 54, 535–543.

Barysheva, M., Jahanshad, N., Foland-Ross, L., Altshuler, L. L., and Thompson, P. M. (2012). White matter microstructural abnormalities in bipolar disorder: a whole brain diffusion tensor imaging study. *Neuroimage Clin.* 2, 558–568. doi: 10.1016/j.nicl.2013.03.016

Basser, P. J., Mattiello, J., and LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267. doi: 10.1016/S0006-3495(94)80775-1

Basser, P. J., and Pierpaoli, C. (1996). Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *J. Magn. Reson. B* 111, 209–219. doi: 10.1006/jmrb.1996.0086

Beasley, T. M., Erickson, S., and Allison, D. B. (2009). Rank-based inverse normal transformations are increasingly used, but are they merited? *Behav. Genet.* 39, 580–595. doi: 10.1007/s10519-009-9281-0

Blangero, J., Diego, V. P., Dyer, T. D., Almeida, M., Peralta, J., Kent, J. W. Jr., et al. (2013). A kernel of truth: statistical advances in polygenic variance component models for complex human pedigrees. *Adv. Genet.* 81, 1–31. doi: 10.1016/B978-0-12-407677-8.00001-4

Blangero, J., Williams, J. T., and Almasy, L. (2001). Variance component methods for detecting complex trait loci. *Adv. Genet.* 42, 151–181. doi: 10.1016/s0065-2660(01)42021-9

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., et al. (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika* 76, 306–317. doi: 10.1007/s11336-010-9200-6

Bootsman, F., Brouwer, R. M., Schnack, H. G., Kemner, S. M., Hillegers, M. H. J., Sarkisyan, G., et al. (2016). A study of genetic and environmental contributions to structural brain changes over time in twins concordant and discordant for bipolar disorder. *J. Psychiatr. Res.* 79, 116–124. doi: 10.1016/j.jpsychires.2016.04.011

Brouwer, R. M., Mandl, R. C., Peper, J. S., van Baal, G. C., Kahn, R. S., Boomsma, D. I., et al. (2010). Heritability of DTI and MTR in nine-year-old children. *Neuroimage* 53, 1085–1092. doi: 10.1016/j.neuroimage.2010.03.017

Brouwer, R. M., Mandl, R. C., Schnack, H. G., van Soelen, I. L., van Baal, G. C., Peper, J. S., et al. (2012). White matter development in early puberty: a longitudinal volumetric and diffusion tensor imaging twin study. *PLoS One* 7:e32316. doi: 10.1371/journal.pone.0032316

Carballedo, A., Amico, F., Ugwu, I., Fagan, A. J., Fahey, C., Morris, D., et al. (2012). Reduced fractional anisotropy in the uncinate fasciculus in patients with major depression carrying the met-allele of the Val66Met brain-derived neurotrophic factor genotype. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 159B, 537–548. doi: 10.1002/ajmg.b.32060

Chen, X. (2014). *Accelerated Estimation and Inference for Heritability of fMRI Data.* Dissertation for PhD thesis. University of Warwick.

Chen, X., Blokland, G. A., Strike, L., and Nichols, T. (2013). Voxelwise and cluster-based heritability inferences of fMRI data. *Proceedings of the 19th Annual Meeting for the Organization for Human Brain Mapping*, 16–20.

Clerx, L., Visser, P. J., Verhey, F., and Aalten, P. (2012). New MRI markers for Alzheimer's disease: a meta-analysis of diffusion tensor imaging and a comparison with medial temporal lobe measurements. *J. Alzheimers Dis.* 29, 405–429. doi: 10.3233/jad-2011-110797

Collins, F. S., and Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature* 505, 612–613. doi: 10.1038/505612a

den Braber, A., van't Ent, D., Blokland, G. A., van Grootheest, D. S., Cath, D. C., Veltman, D. J., et al. (2008). An fMRI study in monozygotic twins discordant for obsessive-compulsive symptoms. *Biol. Psychol.* 79, 91–102. doi: 10.1016/j.biopsycho.2008.01.010

den Braber, A., van 't Ent, D., Boomsma, D. I., Cath, D. C., Veltman, D. J., Thompson, P. M., et al. (2011). White matter differences in monozygotic twins discordant or concordant for obsessive-compulsive symptoms: a combined diffusion tensor imaging/voxel-based morphometry study. *Biol. Psychiatry* 70, 969–977. doi: 10.1016/j.biopsych.2011.03.029

den Braber, A., van 't Ent, D., Cath, D. C., Veltman, D. J., Boomsma, D. I., and de Geus, E. J. (2012). Brain activation during response interference in twins discordant or concordant for obsessive compulsive symptoms. *Twin Res. Hum. Genet.* 15, 372–383. doi: 10.1017/thg.2012.2

Friedman, J. I., Tang, C., Carpenter, D., Buchsbaum, M., Schmeidler, J., Flanagan, L., et al. (2008). Diffusion tensor imaging findings in first-episode and chronic schizophrenia patients. *Am. J. Psychiatry.* 165, 1024–1032. doi: 10.1176/appi.ajp.2008.07101640

Flint, J., and Munafò, M. R. (2013). Candidate and non-candidate genes in behavior genetics. *Curr. Opin. Neurobiol.* 23, 57–61. doi: 10.1016/j.conb.2012.07.005

Ganjgahi, H., Winkler, A. M., Glahn, D. C., Blangero, J., Kochunov, P., and Nichols, T. E. (2015). Fast and powerful heritability inference for family-based neuroimaging studies. *Neuroimage* 115, 256–268. doi: 10.1016/j.neuroimage.2015.03.005

Geng, X., Prom-Wormley, E. C., Perez, J., Kubarych, T., Styner, M., Lin, W., et al. (2012). White matter heritability using diffusion tensor imaging in neonatal brains. *Twin Res. Hum. Genet.* 15, 336–350. doi: 10.1017/thg.2012.14

Grimes, L., and Harvey, W. R. (1980). Estimation of genetic variances and covariances using symmetric differences squared. *J. Anim. Sci.* 50, 634–644. doi: 10.2527/jas1980.504634x

Gronenschild, E. H., Habets, P., Jacobs, H. I., Mengelers, R., Rozendaal, N., van Os, J., et al. (2012). The effects of FreeSurfer version, workstation type, and Macintosh operating system version on anatomical volume and cortical thickness measurements. *PLoS One* 7:e38234. doi: 10.1371/journal.pone.0038234

Ioannidis, J. P. A. (2014). How to make more published research true. *PLoS Med.* 11:e1001747. doi: 10.1371/journal.pmed.1001747

Jahanshad, N., Kochunov, P., Sprooten, E., Mandl, R. C., Nichols, T. E., Almassy, L., et al. (2013). Multi-site genetic analysis of diffusion images and voxelwise heritability analysis: a pilot project of the ENIGMA-DTI working group. *Neuroimage* 81, 455–469. doi: 10.1016/j.neuroimage.2013.04.061

Jahanshad, N., Lee, A. D., Barysheva, M., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., et al. (2010). Genetic influences on brain asymmetry: a DTI study of 374 twins and siblings. *Neuroimage* 52, 455–469. doi: 10.1016/j.neuroimage.2010.04.236

Kim, J. S., Singh, V., Lee, J. K., Lerch, J., Ad-Dab'bagh, Y., MacDonald, D., et al. (2005). Automated 3-D extraction and evaluation of the inner and outer cortical surfaces using a Laplacian map and partial volume effect classification. *Neuroimage* 27, 210–221. doi: 10.1016/j.neuroimage.2005.03.036

Kochunov, P., Coyle, T., Rowland, L., Jahanshad, N., Thompson, P., Kelly, S., et al. (2017). White matter and core cognitive deficits in schizophrenia. *JAMA Psychiatry* 74, 958–966. doi: 10.1001/jamapsychiatry.2017.2228

Kochunov, P., and Duff Davis, M. (2009). Development of structural MR brain imaging protocols to study genetics and maturation. *Methods* 50, 136–146. doi: 10.1016/j.ymeth.2009.08.002

Kochunov, P., Glahn, D. C., Rowland, L. M., Olvera, R., Wincker, P., Yang, D., et al. (2013). Testing the hypothesis of accelerated cerebral white matter aging in schizophrenia and major depression. *Biol. Psychiatry* 73, 482–491. doi: 10.1016/j.biopsych.2012.10.002

Kochunov, P., Jahanshad, N., Marcus, D., Winkler, A., Sprooten, E., Nichols, T. E., et al. (2015). Heritability of fractional anisotropy in human white matter: a comparison of human connectome project and ENIGMA-DTI data. *Neuroimage* 111, 300–311. doi: 10.1016/j.neuroimage.2015.02.050

Kochunov, P., Jahanshad, N., Sprooten, E., Nichols, T. E., Mandl, R. C., Almasy, L., et al. (2014). Multi-site study of additive genetic effects on fractional anisotropy

of cerebral white matter: comparing meta and megaanalytical approaches for data pooling. *Neuroimage* 95C, 136–150. doi: 10.1016/j.neuroimage.2014. 03.033

Kochunov, P., Lancaster, J., Thompson, P., Toga, A. W., Brewer, P., Hardies, J., et al. (2002). An optimized individual target brain in the Talairach coordinate system. *Neuroimage* 17, 922–927. doi: 10.1016/s1053-8119(02)91084-0

Kochunov, P., Thompson, P. M., Winkler, A., Morrissey, M., Fu, M., Coyle, T. R., et al. (2016). The common genetic influence over processing speed and white matter microstructure: evidence from the old order amish and human connectome projects. *Neuroimage* 125, 189–197. doi: 10.1016/j.neuroimage. 2015.10.050

Lerch, J. P., and Evans, A. C. (2005). Cortical thickness analysis examined through power analysis and a population simulation. *Neuroimage* 24, 163–173. doi: 10.1016/j.neuroimage.2004.07.045

Mandl, R. C., Rais, M., van Baal, G. C., van Haren, N. E., Cahn, W., Kahn, R. S., et al. (2013). Altered white matter connectivity in never-medicated patients with schizophrenia. *Hum. Brain Mapp.* 34, 2353–2365. doi: 10.1002/hbm. 22075

Marcus, D. S., Harms, M. P., Snyder, A. Z., Jenkinson, M., Wilson, J. A., Glasser, M. F., et al. (2013). Human connectome project informatics: quality control, database services, and data visualization. *Neuroimage* 80, 202–219. doi: 10.1016/j.neuroimage.2013.05.077

McGuire, S. A., Wijtenburg, S. A., Sherman, P. M., Rowland, L. M., Ryan, M., Sladky, J. H., et al. (2017). Reproducibility of quantitative structural and physiological MRI measurements. *Brain Behav.* 7:e00759. doi: 10.1002/ brb3.759

Meyer-Lindenberg, A., Nicodemus, K. K., Egan, M. F., Callicott, J. H., Mattay, V., and Weinberger, D. R. (2008). False positives in imaging genetics. *Neuroimage* 40, 655–661. doi: 10.1016/j.neuroimage.2007.11.058

Nazeri, A., Chakravarty, M. M., Felsky, D., Lobaugh, N. J., Rajji, T. K., Mulsant, B. H., et al. (2013). Alterations of superficial white matter in schizophrenia and relationship to cognitive performance. *Neuropsychopharmacology* 38, 1954–1962. doi: 10.1038/npp.2013.93

Nichols, T., and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.* 12, 419–446. doi: 10.1191/0962280203sm341ra

Parisi, F., Strino, F., Nadler, B., and Kluger, Y. (2014). Ranking and combining multiple predictors without labeled data. *Proc. Natl. Acad. Sci. U S A* 111, 1253–1258. doi: 10.1073/pnas.1219097111

Penke, L., Muñoz Maniega, S., Houlihan, L. M., Murray, C., Gow, A. J., Clayden, J. D., et al. (2010a). White matter integrity in the splenium of the corpus callosum is related to successful cognitive aging and partly mediates the protective effect of an ancestral polymorphism in *ADRB2*. *Behav. Genet.* 40, 146–156. doi: 10.1007/s10519-009-9318-4

Penke, L., Muñoz Maniega, S., Murray, C., Gow, A. J., Hernandez, M. C., Clayden, J. D., et al. (2010b). A general factor of brain white matter integrity predicts information processing speed in healthy older people. *J. Neurosci.* 30, 7569–7574. doi: 10.1523/JNEUROSCI.1553-10.2010

Pérez-Iglesias, R., Tordesillas-Gutiérrez, D., Barker, G. J., McGuire, P. K., Roiz-Santiañez, R., Mata, I., et al. (2010). White matter defects in first episode psychosis patients: a voxelwise analysis of diffusion tensor imaging. *Neuroimage* 49, 199–204. doi: 10.1016/j.neuroimage.2009.07.016

Shen, K.-K., Rose, S., Fripp, J., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., et al. (2014). Investigating brain connectivity heritability in a twin study using diffusion imaging data. *Neuroimage* 100, 628–641. doi: 10.1016/j.neuroimage. 2014.06.041

Smith, S. M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T. E., Mackay, C. E., et al. (2006). Tract-based spatial statistics: voxelwise analysis

of multi-subject diffusion data. *Neuroimage* 31, 1487–1505. doi: 10.1016/j. neuroimage.2006.02.024

Sprooten, E., Sussmann, J. E., Clugston, A., Peel, A., McKirdy, J., Moorhead, T. W., et al. (2011). White matter integrity in individuals at high genetic risk of bipolar disorder. *Biol. Psychiatry* 70, 350–356. doi: 10.1016/j.biopsych.2011.01.021

Stein, J. L., Hua, X., Lee, S., Ho, A. J., Leow, A. D., Toga, A. W., et al. (2011). Voxelwise genome-wide association study (vGWAS). *Neuroimage* 53, 1160–1174. doi: 10.1016/j.neuroimage.2010.02.032

Stein, J. L., Hua, X., Morra, J. H., Lee, S., Hibar, D. P., Ho, A. J., et al. (2010). Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in Alzheimer's disease. *Neuroimage* 51, 542–554. doi: 10.1016/j.neuroimage.2010.02.068

Stein, J. L., Medland, S. E., Vasquez, A. A., Hibar, D. P., Senstad, R. E., Winkler, A. M., et al. (2012). Identification of common variants associated with human hippocampal and intracranial volumes. *Nat. Genet.* 44, 552–561. doi: 10.1038/ng.2250

Teipel, S. J., Wegrzyn, M., Meindl, T., Frisoni, G., Bokde, A. L., Fellgiebel, A., et al. (2012). Anatomical MRI and DTI in the diagnosis of Alzheimer's disease: a european multicenter study. *J. Alzheimers Dis.* 31, S33–S47. doi: 10.3233/jad-2012-112118

Thomason, M. E., and Thompson, P. M. (2011). Diffusion imaging, white matter, and psychopathology. *Annu. Rev. Clin. Psychol.* 7, 63–85. doi: 10.1146/annurev-clinpsy-032210-104507

Thompson, P. M., Martin, N. G., and Wright, M. J. (2010). Imaging genomics. *Curr. Opin. Neurol.* 23, 368–373. doi: 10.1097/WCO.0b013e32833b764c

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182. doi: 10.1007/s11682-013-9269-5

Ugurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., et al. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the human connectome project. *Neuroimage* 80, 80–104. doi: 10.1016/j.neuroimage.2013.05.012

van Beijsterveldt, C. E., Groen-Blokhuis, M., Hottenga, J. J., Franic, S., Hudziak, J. J., Lamb, D., et al. (2013). The young netherlands twin register (YNTR): longitudinal twin and family studies in over 70,000 children. *Twin Res. Hum. Genet.* 16, 252–267. doi: 10.1017/thg.2012.118

Van Essen, D. C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T. E., Bucholz, R., et al. (2012). The human connectome project: a data acquisition perspective. *Neuroimage* 62, 2222–2231. doi: 10.1016/j.neuroimage.2012.02.018

van Soelen, I. L., Brouwer, R. M., Peper, J. S., van Leeuwen, M., Koenis, M. M., van Beijsterveldt, T. C., et al. (2012). Brain SCALE: brain structure and cognition: an adolescent longitudinal twin study into the genetic etiology of individual differences. *Twin Res. Hum. Genet.* 15, 453–467. doi: 10.1017/thg.2012.4

Decentralized Analysis of Brain Imaging Data: Voxel-Based Morphometry and Dynamic Functional Network Connectivity

Harshvardhan Gazula[1]*, Bradley T. Baker[1,2]*, Eswar Damaraju[1,3], Sergey M. Plis[1], Sandeep R. Panta[1], Rogers F. Silva[1] and Vince D. Calhoun[1,3]

[1] The Mind Research Network, Albuquerque, NM, United States, [2] Department of Computer Science, The University of New Mexico, Albuquerque, NM, United States, [3] Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM, United States

In the field of neuroimaging, there is a growing interest in developing collaborative frameworks that enable researchers to address challenging questions about the human brain by leveraging data across multiple sites all over the world. Additionally, efforts are also being directed at developing algorithms that enable collaborative analysis and feature learning from multiple sites without requiring the often large data to be centrally located. In this paper, we propose two new decentralized algorithms: (1) A decentralized regression algorithm for performing a voxel-based morphometry analysis on structural magnetic resonance imaging (MRI) data and, (2) A decentralized dynamic functional network connectivity algorithm which includes decentralized group ICA and sliding-window analysis of functional MRI data. We compare results against those obtained from their pooled (or centralized) counterparts on the same data i.e., as if they are at one site. Results produced by the decentralized algorithms are similar to the pooled-case and showcase the potential of performing multi-voxel and multivariate analyses of data located at multiple sites. Such approaches enable many more collaborative and comparative analysis in the context of large-scale neuroimaging studies.

Keywords: decentralized algorithms, COINSTAC, VBM, dFNC, multi-shot

1. INTRODUCTION

In the current times, innovation and discovery are often underpinned by the size of data at one's disposal and this has led to a paradigm shift in scientific research increasing the emphasis on collaborative data-sharing (Cragin et al., 2010; Tenopir et al., 2011). This growing significance of data-sharing is more evident in the field of neuroscience where, in the past few years, there has been a proliferation of efforts (Poldrack et al., 2013) toward enabling researchers to leverage data across multiple sites. In part, this is due to the fact that collecting neuroimaging data is expensive as well as time consuming (Landis et al., 2016) and aggregating or sharing data across various sites provides researchers with an opportunity to uncover important findings that are beyond the scope of the original study (Poldrack et al., 2013). In addition to making predictions more certain by increasing the sample size (Button et al., 2013), sharing data ensures reliability and validity of the results, and safeguards against data fabrication and falsification (Tenopir et al., 2011; Ming et al., 2017).

As mentioned previously, data-specific collaborative efforts include either aggregating the data via a centralized data sharing repository or sharing data via agreement based collaborations, or data usage agreement (DUA) in other words (Thompson et al., 2014, 2017). However, each methodology has its own set of barriers. For example, policy or proprietary restrictions or data re-identification concerns (Sweeney, 2002; Shringarpure and Bustamante, 2015) might hinder data sharing whereas DUAs might take months to complete and even if one comes through, there is no guarantee of the utility of the data until the planned analysis is performed (Baker et al., 2015; Ming et al., 2017). Other significant challenges include the storage and computational resources needed which could prove costly as the volume of the data shared goes up.

Frameworks such as ENIGMA (Thompson et al., 2014, 2017) to some extent bypass the need for DUAs by performing a centrally coordinated analysis at each local site. This enables potentially large data at each local site to stay put allowing a greater level of control as well as privacy. Another framework called ViPAR (Carter et al., 2015) tries to go one step further by, relying on open-source technologies, completely isolating the data at the local site but only pooling them via transfer to perform automated statistical analyses. This repeated pooling of data becomes cumbersome as the number of sites or the size of the data at each site goes up and ENIGMA (Thompson et al., 2014, 2017; Hibar et al., 2015; van Erp et al., 2016) addresses this issue by pooling local statistical results for further analysis, also known as, meta-analysis (Adams et al., 2016). However, the heterogeneity among the local analyses caused by adopting various data collection mechanisms or preprocessing methods can lead to inaccurate meta-analysis findings.

Plis et al. (2016), proposed a web-based framework titled Collaborative Informatics and Neuroimaging Suite Toolkit for Anonymous Computation (COINSTAC) to address the aforementioned issues. COINSTAC provides a platform to analyze data stored locally across multiple organizations without the need for pooling the data at any point during the analysis. It is intended to be an ultimate one-stop shop by which researchers can build any statistical or machine learning model collaboratively in a decentralized fashion. This framework implements a message passing infrastructure that will allow large scale analysis of decentralized data with results on par with those that would have been obtained if the data were in one place. Since, there is no pooling of data it also preserves the privacy of individual datasets.

Some of the decentralized computations discussed in the literature so far include decentralized regression (Plis et al., 2016), joint independent component analysis (Baker et al., 2015), decentralized independent vector analysis (Wojtalewicz et al., 2017), decentralized neural networks (Lewis et al., 2017), decentralized stochastic neighbor embedding (Saha et al., 2017) and many more. To our knowledge, most of these algorithms have been tested on synthetic data. In this work we present two new decentralized algorithms that are widely used in a centralized manner in the imaging community and demonstrate their utility on real world brain imaging data.

Regression, is widely used in neuroimaging studies as it enables one to regress certain covariates, for example- age, diagnosis, gender or treatment response, to study their effects on the structure and function of various brain regions. Some examples of regression related studies in this field include (Fennema-Notestine et al., 2007) where regression was used as a validity test in examining the aggregation of structural imaging across different datasets. In addition, the very successful ENIGMA studies are mostly using regression analyses for a small number of variables. Roshchupkin et al. (2016) presented a framework titled HASE (high-dimensional association analyses) that is capable of analyzing high-dimensional data at full resolution, yielding exact association statistics. While singleshot and multishot regression have been presented previously (Plis et al., 2016), their treatment was cursory in nature without any actual consideration of the appropriate gradient descent scheme or the validity of the methods on real datasets both of which have been presented in this work.

In this paper, in addition to improving the single-shot and multi-shot regression we also present a new variant of decentralized regression- "decentralized regression with normal equation" and extend this work to operate on voxels in an MRI image, in order to implement a voxel-based morphometry (VBM) study in a decentralized framework (Ashburner and Friston, 2000). We implement and evaluate the proposed decentralized VBM approach on the publicly available MIND Clinical Imaging Consortium (MCIC) dataset (available via the COINS data exchange at https://coins.mrn.org and contrast the results obtained with those from pooled/centralized regression to validate the proof-of-concept.

Another widely utilized method in neuroimaging analysis is dynamic functional network connectivity (dFNC) (Sakoglu et al., 2010; Allen et al., 2014). dFNC is an analysis pipeline for functional magnetic resonance imaging (fMRI) data, which allows for the identification and analysis of networks of co-activating brain states. In contrast to static approaches (Smith et al., 2009), which take the mean connectivity over time-points, dFNC uses clustering of time varying connectivity estimates computed from sliding-windows taken over subject time-courses, thus becoming desirable in experiments where network connectivity is highly dynamic in the time dimension, for example in experiments which utilize resting-state fMRI (Deco et al., 2013; Damaraju et al., 2014).

Importantly, dFNC is focused on time-courses of networks extracted from a group independent component analysis (ICA), which is a widely used approach for estimating functional brain networks (Calhoun and Adali, 2012) and as such to implement dFNC we needed to also implement a decentralized group ICA approach.

For collaborative neuroimaging applications, a decentralized version of dFNC is desirable for many of the same reasons as regression, and currently, no such decentralized version exists. Unlike regression, however, the dFNC pipeline consists of multiple, distinct stages, all of which require decentralization. In this paper, we present an initial version of decentralized dFNC by providing decentralized approaches to both the group spatial independent component analysis (ICA) and K-Means

clustering steps in the pipeline, which, along with additional preprocessing steps including sliding window correlation, can be implemented together to perform decentralized dFNC. Our resulting methods, dgICA, and ddFNC via dK-Means, provide dynamic connectivity results consistent with established pooled approaches in the literature, thus representing an important step toward more exhaustive analysis of the decentralized approaches to the dFNC pipeline. Our contributions in this paper can thus be summarized as follows.

1. Development of decentralized regression with normal equation, improvement of single-shot and multi-shot regression and their validation on structural MRI data
2. Presentation of a decentralized dynamic functional network connectivity analysis pipeline and its evaluation on functional MRI data

2. METHODS

2.1. Decentralized VBM (i.e., Voxelwise Decentralized Regression)

Statistical analysis plays a key role in the field of neuroimaging studies. Researchers would often want to characterize the effect of various factors such as age, gender, disease condition, etc., on the composition of brain tissue at various regions of the brain. Voxel-based morphometry (VBM) (Ashburner and Friston, 2000) is one such approach that facilitates a comprehensive comparison, via generalized linear modeling, of voxel-wise gray matter concentration between different groups, for example. To enable such statistical assessment on data present at various sites, it is important to develop decentralized tools. In this section, we first provide a brief overview of decentralized regression algorithms (the building blocks of decentralized VBM which is essentially voxel-wise regression) along with some notation.

The goal of decentralized regression is to fit a linear equation (given by Equation 1) relating the covariates at S different sites to the corresponding responses. Assume each site j has data set $\mathcal{D}_j = \{(\mathbf{x}_i, y_i) : i \in \{1, 2, \ldots, s_j\}\}$ where $\mathbf{x}_{i,j} \in R^d$ is a d-dimensional vector of real-values features, and $y_j \in$ is a response. We consider fitting the model in Equation 2 where $\mathbf{w}$ is given as $[\mathbf{w}; b]$ and $\mathbf{x}$ as $[\mathbf{x}; 1]$

$$y \approx \mathbf{w}^\top \mathbf{x} + b \tag{1}$$
$$\mathbf{y} \approx \mathbf{w}^\top \mathbf{x} \tag{2}$$

The vector of regression parameters/weights $\mathbf{w}$ is found by minimizing the sum of the squared error given in Equation (3)

$$F(\mathbf{w}) = \sum_{j=1}^{S} \sum_{i=1}^{s_j} (y_i - \mathbf{w}^\top \mathbf{x}_{i,j})^2 \tag{3}$$

The regression objective function is a linearly separable function, that can be written as sum of a local objective function calculated at each local site as follows:

$$F(\mathbf{w}) = \sum_{j=1}^{S} F_j(\mathbf{w}) \tag{4}$$

where

$$F_j(\mathbf{w}) = \sum_{i=1}^{s_j} (y_i - \mathbf{w}^\top \mathbf{x}_{i,j}) \tag{5}$$

A central aggregator (AGG) is assumed whose role is to compute the global minimizer $\hat{\mathbf{w}}$ of $F(\mathbf{w})$.

2.1.1. Single-Shot Regression
In one approach to solve the decentralized regression problem, termed the *single-shot* regression (Plis et al., 2016), each site j finds the minimizer $\hat{\mathbf{w}}_j$ of the local objective function $F_j(\mathbf{w})$. This is the same as solving the regression problem at each local site. Once the regression model at each site is fit, the weights are sent to the central aggregator (AGG) where they are aggregated (weighted average) to find the global minimizer or can be used separately to perform a meta-analysis similar to those performed in ENIGMA (using a manual spreadsheet-based approach however) (Turner et al., 2013; van Erp et al., 2016). The pseudocode to perform single-shot decentralized regression (Plis et al., 2016), with a slight modification, is presented here again for completeness.

Algorithm 1 Single-shot Regression
Require: Data $\mathcal{D}_j$ at site j for sites $j = 1, 2, \ldots, S$, where $|\mathcal{D}_j| = s_j \forall j$
1: **for** $j = 1$ to S **do**
2:     $\hat{\mathbf{w}}_j = \mathrm{argmin}_\mathbf{w} F_j(\mathbf{w})$.
3:         Node j sends $\hat{\mathbf{w}}_j$ to AGG.
4: **end for**
5: AGG computes $\hat{\mathbf{w}} = \frac{1}{\sum_{j=1}^{S} s_j} \sum_{j=1}^{S} s_j \hat{\mathbf{w}}_j$ and return $\hat{\mathbf{w}}$

2.1.2. Decentralized Regression With Normal Equation
One limitation of single-shot regression is that the "site" level covariates cannot be included at each local site as this leads to collinearity issues. This issue can be offset by utilizing a decentralized version of the analytical solution to the linear regression problem. For a standard regression problem of the form given in Equation (2), the analytical solution is given as

$$\hat{\mathbf{w}} = (\mathbf{x}^\top \mathbf{x})^{-1} \mathbf{x}^\top \mathbf{y} \tag{6}$$

Assuming that the augmented data matrix $\mathbf{x}$ is made up of data from different local sites, i.e.,

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_S \end{bmatrix} \tag{7}$$

it's easy to see that $\hat{\mathbf{w}}$ can be written as

$$\hat{\mathbf{w}} = \left(\begin{bmatrix} \mathbf{x}_1^\top & \cdots & \mathbf{x}_S^\top \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_S \end{bmatrix} \right)^{-1} \times$$

$$\begin{bmatrix} \mathbf{x}_1^\top & \cdots & \mathbf{x}_S^\top \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_s \end{bmatrix} \quad (8)$$

$$\hat{\mathbf{w}} = \left(\sum_{j=1}^{S} \mathbf{x}_j^T \mathbf{x}_j \right)^{-1} \times \left(\sum_{j=1}^{S} \mathbf{x}_j^T \mathbf{y}_j \right) \quad (9)$$

The above variant of the analytical solution to a regression model shows that even if the data resides in different locations, fitting a global model in the presence of site covariates delivers results that are exactly similar to the pooled case.

Algorithm 2 Decentralized Regression with Normal Equation

Require: Data $\mathcal{D}_j$ at site j for sites $j = 1, 2, \ldots, S$, where $|\mathcal{D}_j| = s_j \forall j$

1: **for** $j = 1$ to S **do**
2:     Compute $Cov(X_j) = \mathbf{x}_j^\top \mathbf{x}_j$
3:     Compute $\mathbf{x}_j^\top \mathbf{y}_j$
4:     Node j sends $Cov(X_j)$ and $\mathbf{x}_j^\top \mathbf{y}_j$ to AGG.
5: **end for**
6: AGG computes

$$\hat{\mathbf{w}} \leftarrow \left(\sum_{j=1}^{S} Cov(X_j) \right)^{-1} \left(\sum_{j=1}^{S} \mathbf{x}_j^\top \mathbf{y}_j \right) \text{ and return}$$
$\hat{\mathbf{w}}$

2.1.3. Multi-Shot Regression

Decentralized regression with a normal equation is a nice mathematical formulation which produces results that are exactly the same as those from the pooled regression. However, one of the biggest drawback of the analytical form of regression is it becomes computationally expensive to evaluate the inverse of $\mathbf{x}^\top \mathbf{x}$ as the number of features in the dataset ($\mathcal{D}$) increases. While in a neuroimaging setting there might not be as many covariates to make it computationally expensive, it is indeed a challenge while working with datasets where the cardinality of the feature set is usually large (especially in machine learning). One can overcome this drawback by implementing an optimization method in a way that entails the local sites and *AGG* having to communicate iteratively. This is a type of distributed gradient descent and such a regression is termed "*multi-shot*" regression (Plis et al., 2016).

For a regression model of the form given in Equation 5, the gradient update equation (given a learning rate η) is given as

$$\hat{\mathbf{w}}_{t+1} = \hat{\mathbf{w}}_t - \eta \cdot \nabla F_j(\hat{\mathbf{w}}) \quad (10)$$

Algorithm 3 Multi-shot Regression

Require: Data $\mathcal{D}_j$ at site j for sites $j = 1, 2, \ldots, S$, where $|\mathcal{D}_j| = s_j \forall j$
Require: Step size η (Suggested default: 0.001)
Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates (Suggested defaults: 0.9 and 0.999, respectively)
Require: Small constant δ used for numerical stabilization (Suggested default: 10^{-8})
Require: $\hat{w}_{t-1} \leftarrow \mathbf{0}$ (Initial parameter vector), $m_0 \leftarrow \mathbf{0}$ (Initialize 1st moment vector), $v_0 \leftarrow \mathbf{0}$ (Initialize 2nd moment vector), $t \leftarrow 0$ (Initialize timestep), tolerance **Tol** ▷ at AGG

1: **while True do**
2:     **for** $j = 1$ to S **do**
3:         AGG sends $\hat{\mathbf{w}}_{t-1}$ to node j
4:         Node j computes $\nabla F_j(\hat{\mathbf{w}}_{t-1})$
5:         Node j sends $\nabla F_j(\hat{\mathbf{w}}_{t-1})$ to AGG.
6:     **end for**
7:     AGG computes $\nabla F_c \leftarrow \sum_{j=1}^{S} \nabla F_j(\hat{\mathbf{w}}_{t-1})$ ▷ aggregate gradient
8:     $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \nabla F_c$ ▷ update biased first moment estimate
9:     $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot \nabla F_c^2$ ▷ update biased second moment estimate
10:    $\hat{m}_t \leftarrow m_t/(1 - \beta_1^t)$ ▷ Compute bias-corrected first moment estimate
11:    $\hat{v}_t \leftarrow v_t/(1 - \beta_2^t)$ ▷ Compute bias-corrected second moment estimate
12:    AGG computes $\hat{\mathbf{w}}_t \leftarrow \hat{\mathbf{w}}_{t-1} - \eta \cdot \hat{m}_t/(\sqrt{\hat{v}_t} + \delta)$ ▷ Update parameters
13:    **if** $||\hat{\mathbf{w}}_t - \hat{\mathbf{w}}_{t-1}||_2 \leq$ **Tol then**
14:        **break**
15:    **end if**
16:    $\hat{\mathbf{w}}_{t-1} \leftarrow \hat{\mathbf{w}}_t$
17: **end while**
18: **return** $\hat{\mathbf{w}}_t$ as $\hat{\mathbf{w}}$ ▷ Resulting parameters

---

where

$$\nabla F_j(\hat{\mathbf{w}}) = \sum_{i=1}^{s_j} (y_i - \hat{\mathbf{w}}^\top \mathbf{x}_{i,j}) \mathbf{x}_{i,j} \quad (11)$$

In multi-shot regression, at every time step the *AGG* sends the value of $\hat{w_{t-1}}$ to each of the local sites which then compute their local gradients $\nabla F_j(w_t)$ and send them back to the *AGG* where it sums up all the local gradients in order to update the parameter vector $\hat{\mathbf{w}}_t$. The need to sum up all the local gradients is explained as follows:

From Equation (4), $F(\hat{\mathbf{w}}) = \sum_{j=1}^{S} F_j(\hat{\mathbf{w}})$

$$\therefore \nabla F(\hat{\mathbf{w}}) = \sum_{j=1}^{S} \nabla F_j(\hat{\mathbf{w}}) \quad (12)$$

To illustrate this using an example, suppose there are 3 sites ($S = 3$) with s_1, s_2 and s_3 number of samples, respectively, at each site. The global objective function $F(\hat{\mathbf{w}})$ can be easily written

as the sum of objective functions from each site (this because the objective function is linear) as follows:

$$
\begin{aligned}
F(\hat{\mathbf{w}}) &= \sum_{j=1}^{s_1+s_2+s_3} (y_j - \hat{\mathbf{w}}^\top x_j)^2 \\
&= \sum_{j=1}^{s_1} (y_j - \hat{\mathbf{w}}^\top x_j)^2 + \sum_{j=1}^{s_2} (y_j - \hat{\mathbf{w}}^\top x_j)^2 \\
&\quad + \sum_{j=1}^{s_3} (y_j - \hat{\mathbf{w}}^\top x_j)^2 \\
&= \sum_{j=1}^{s_1} F_1(\hat{\mathbf{w}}) + \sum_{j=1}^{s_2} F_2(\hat{\mathbf{w}}) + \sum_{j=1}^{s_3} F_3(\hat{\mathbf{w}}) \\
\therefore\ \nabla F(\hat{\mathbf{w}}) &= \sum_{j=1}^{s_1} \nabla F_1(\hat{\mathbf{w}}) + \sum_{j=1}^{s_2} \nabla F_2(\hat{\mathbf{w}}) + \sum_{j=1}^{s_3} \nabla F_3(\hat{\mathbf{w}}) \quad (13)
\end{aligned}
$$

From Equation (13), it should be easy to see that the aggregated gradient is just a sum of the gradients from each site. On the other hand, if the mean sum of squared errors is preferred i.e., $F(\hat{\mathbf{w}}) = \frac{1}{m} \sum_{j=1}^{m} (y_j - \hat{\mathbf{w}}^\top x_j)^2$, which mathematically has the same minimizer as $\sum_{j=1}^{m} (y_j - \hat{\mathbf{w}}^\top x_j)^2$ since $F(\hat{\mathbf{w}})$ is convex, it can be shown that the aggregated gradient is a weighted average of the gradients from the local sites:

$$
\begin{aligned}
F(\hat{\mathbf{w}}) &= \frac{1}{s_1+s_2+s_3} \sum_{j=1}^{s_1+s_2+s_3} F_j(\hat{\mathbf{w}}) \\
&= \frac{1}{s_1+s_2+s_3} \left(\frac{s_1}{s_1} \sum_{j=1}^{s_1} F_j(\hat{\mathbf{w}}) + \frac{s_2}{s_2} \sum_{j=1}^{s_2} F_j(\hat{\mathbf{w}}) \right. \\
&\quad \left. + \frac{s_3}{s_3} \sum_{j=1}^{s_3} F_j(\hat{\mathbf{w}}) \right) \\
&= \frac{1}{s_1+s_2+s_3} (s_1 F_1(\hat{\mathbf{w}}) + s_2 F_2(\hat{\mathbf{w}}) + s_3 F_3(\hat{\mathbf{w}})) \\
\therefore\ \nabla F(\hat{\mathbf{w}}) &= \frac{1}{s_1+s_2+s_3} (s_1 \nabla F_j(\hat{\mathbf{w}}) + s_2 \nabla F_j(\hat{\mathbf{w}}) + s_3 \nabla F_j(\hat{\mathbf{w}}))
\end{aligned}
$$

$$(14)$$

Algorithm 3 shows the steps involved in multi-shot regression. In order to update the parameters (here, $\hat{\mathbf{w}}$), any off-the-shelf optimization scheme, for example, gradient descent, adagrad (Duchi et al., 2011), adadelta (Zeiler, 2012), momentum gradient descent (Rumelhart et al., 1986), nesterov accelerated gradient descent (Nesterov et al., 1983), Adam (Kingma and Ba, 2014) could have been used. The choice of scheme adopted could depend on the data being analyzed, Moreover, additional considerations have to be given to the stopping criterion tolerance, the number of iterations, the choice of learning rate and any other additional hyper-parameters depending on the scheme utilized. In some cases, the choice of optimization scheme can result in an analysis which could take minutes, days or years to arrive. In our tests, we found out that the Adam optimization scheme performs extremely well on the real dataset and hence has been adopted to perform the multi-shot regression.

2.1.4. Other Statistics

In addition to generating the weights of the covariates (regression parameters), one would also be interested in determining the overall model performance given by goodness-of-fit or the coefficient of determination (R^2) as well as the statistical significance of each weight parameter (t-value or p-value).

As demonstrated in Algorithm 4 (Ming et al., 2017), determining R^2 entails calculating the sum-square-of-errors (SSE) as well as total sum of squares (SST) which are evaluated at each local site and then aggregated at the global site to evaluate R^2 given by $1 - SSE/SST$. An intermediary step before the calculation of SST is the calculation of the global $\bar{\mathbf{y}}$ which is determined by taking a weighted average of the local $\bar{\mathbf{y}}_j$ weighted on the size of data at each local site.

Algorithm 4 Decentralized R^2 calculation

Require: Data $\mathcal{D}_j$ at site j for sites $j = 1, 2, \ldots, S$, where $|\mathcal{D}_j| = s_j \forall j$

1: AGG sends $\hat{\mathbf{w}}$ to each local site.
2: **for** $j = 1$ to S **do**
3:      Node j computes $\bar{\mathbf{y}}_j = \frac{1}{s_j} \sum_{i=1}^{s_j} \mathbf{y}_i$
4:      Node j sends $\bar{\mathbf{y}}_j$ and s_j to AGG.
5: **end for**
6: AGG computes $\bar{\mathbf{y}} = \frac{\sum_{j=1}^{S} (s_j \cdot \bar{\mathbf{y}}_j)}{\sum_{j=1}^{S} s_j}$ ▷ global mean
7: AGG sends $\bar{\mathbf{y}}$ to local sites
8: **for** $j = 1$ to S **do**
9:      $SST_j = \sum_{i=1}^{s_j} (y_i - \bar{\mathbf{y}})^2$
10:     $\hat{\mathbf{y}}_j = \hat{\mathbf{w}} \cdot \mathbf{x}_j$
11:     $SSE_j = \sum_{i=1}^{s_j} (\mathbf{y}_i - \hat{\mathbf{y}}_j)^2$
12:     Node j sends SST_j and SSE_j to AGG
13: **end for**
14: AGG computes $SST \leftarrow \sum_{j=1}^{S} SST_j$,
       $SSE \leftarrow \sum_{j=1}^{S} SSE_j$,
       $R^2 \leftarrow 1 - \frac{SSE}{SST}$
15: **return** R^2

Algorithm 5 (Ming et al., 2017) details the steps involved in calculating the t-values (and therefore p-values) of each regression parameter. Assuming the weight vector has been calculated using either the single-shot or multi-shot regression, the global weight vector ($\hat{\mathbf{w}}$) is sent to each of the local sites where the local covariance matrix as well as the sum-square-of-errors is calculated and sent back along with the data size to the aggregator (AGG) which then utilizes that information to calculate the t-values for each parameter (or coefficient). Once, the t-values have been calculated, the corresponding two-tailed p-values can be deduced using any publicly available *distributions* library.

Algorithm 5 Decentralized t-value calculation

Require: Data $\mathcal{D}_j$ at site j for sites $j = 1, 2, \ldots, S$, where $|\mathcal{D}_j| = s_j \forall j$
 1: AGG sends $\hat{\mathbf{w}}$ to each local site.
 2: **for** $j = 1$ to S **do**
 3:     $\hat{\mathbf{y}}_j = \hat{\mathbf{w}} \cdot \mathbf{x}_j$
 4:     $SSE_j = \sum_{i=1}^{s_j} (y_i - \hat{\mathbf{y}}_j)^2$
 5:     $Cov(\mathbf{x}_j) = \mathbf{x}_j^\top \mathbf{x}_j$
 6:     Node j sends SSE_j, $Cov(X_j)$ and s_j to AGG.
 7: **end for**
 8: AGG computes $Cov(\mathbf{x}) \leftarrow \sum_{j=1}^{S} Cov(\mathbf{x}_j)$,
        $MSE \leftarrow \frac{1}{\sum_{j=1}^{S} s_j} \sum_{j=1}^{S} SSE_j$,
        $SE(W) \leftarrow \sqrt{diag(MSE \cdot Cov(\mathbf{x})^{-1})}$,
        $t \leftarrow \hat{w}/SE(W))$
 9: **return** t

2.1.5. Bandwidth and Complexity

For singleshot regression, each site communicates a local weight vector $\hat{\mathbf{w}}_j$ of size $(d + 1)$ to the aggregator in addition to the cardinality of the dataset at each site $|\mathcal{D}_j| = s_j$, a scalar. Once all the information is aggregated, a weighted average of the local $\hat{\mathbf{w}}_j$s with the weights being s_j performed to get the global weight vector $\hat{\mathbf{w}}$. Assuming $s_j > d$ and that the normal equation is used to get the local weight vectors $\hat{\mathbf{w}}_j$s, the computational complexity is $\mathcal{O}(d^2 s_j)$ whereas the computational complexity of calculating the weighted average at the AGG is $\mathcal{O}(d)$.

In the case of decentralized regression with normal equation, the first step (at each site) includes the calculation of $\mathbf{x}^\top \mathbf{x}$ (at $\mathcal{O}(d^2 s_j)$) and $\mathbf{x}^\top \mathbf{y}$ (at $\mathcal{O}(d s_j)$) with an overall complexity of $\mathcal{O}(d^2 s_j)$. A total information of $\sum_{j=1}^{S} \{s_j \times [(d+1)^2 + (d+1)]\}$ is communicated to the AGG where they are aggregated (as shown in Algorithm 2) to obtain the global weight vector $\hat{\mathbf{w}}$ at $\mathcal{O}(d^3)$.

Contrary to where the computation starts in the case of singleshot or DRNE, the computation/communication starts from the AGG in multishot regression. The AGG initializes the $\hat{\mathbf{w}}$ and communicates the $(d + 1)$-sized vector to each of the S sites. At every iteration, each site j then calculates the gradient vector ($\mathcal{O}(d)$) and sends it back to the AGG which again means the communication $S \times (d + 1)$ accounting for S sites. At the AGG, steps 7 though 12 (refer to Algorithm 3) are performed at an order of $\mathcal{O}(d)$ which are again sent back to each of the local sites, implying a communication of $S \times d$, for the next iteration of the gradient descent.

The above treatment of communication bandwidth and complexity is subject to certain considerations *viz.*, the number of covariates, the number of samples at each site, the optimization scheme used in the calculation of $\mathbf{x}^\top \mathbf{x}$, the stopping criterion, etc.

2.2. Decentralized dFNC

In this section, we briefly present our initial work toward performing dynamic functional network connectivity (dFNC) analysis in a decentralized framework. As mentioned earlier, dFNC is a multi-step pipeline finds common states in subject fMRI time-courses (TCs), and is often done by clustering a

sliding window over subject time-courses, as is done (e.g., Allen et al., 2014; Damaraju et al., 2014). Thus, we present methods for decentralized spatial ICA along with decentralized K-Means clustering. Our presentation here is by no means a rigorous take on dFNC which we save for future work.

2.2.1. Decentralized Group Spatial ICA

Following preprocessing, the first step in the dFNC pipeline includes group ICA (Calhoun et al., 2001). Since we are dealing with fMRI data, suppose that we now have data $\mathbf{X} \in \mathbb{R}^{d \times N}$, where d is the voxel-space of the data (in brain voxels), and N is the total number of time-points across all subjects in the network. In linear spatial ICA, we model each individual subject as a mixture of r many statistically independent spatial maps, $\mathbf{A} \in \mathbb{R}^{d \times r}$, and their time-courses, $\mathbf{S} \in \mathbb{R}^{r \times N_i}$, where N_i is the length of the time-course belonging to subject i. In the decentralized case, we can model the global data set $\mathbf{X}$ as the column-wise concatenation of s sites in the temporal dimension, where each site is modeled as a set of subjects concatenated in the temporal dimension:

$$\mathbf{X} = [\mathbf{A}_1 \mathbf{S}_1 \ \mathbf{A}_2 \mathbf{S}_2 \ \cdots \ \mathbf{A}_s \mathbf{S}_s] \in \mathbb{R}^{d \times N}.$$

Our goal is to learn a global unmixing matrix, $\mathbf{W}$, such that $\mathbf{XW} \approx \hat{\mathbf{A}}$, where $\hat{\mathbf{A}} \in \mathbb{R}^{d \times r}$ is a set of unmixed spatially independent components. To this end, we perform a decentralized group independent component analysis (dgICA). Our method consists first of the two-stage GlobalPCA procedure utilized in Baker et al. (2015). In this procedure, each site first performs subject-specific LocalPCA dimension-reduction and whitening to a common k principal components in the temporal dimension. A decentralized, second stage, then produces a global set of r spatial eigenvectors, $\mathbf{V} \in \mathbb{R}^{r \times d}$. As outlined in Baker et al. (2015), this second stage has sites pass locally-reduced eigenvectors to other sites in a peer-to-peer scheme, where upon receiving a set of eigenvectors, a site then stacks them in the column dimension, and performs a further reduction of the stacked matrix, which is then passed to the next peer in the network. This process iterates until the global eigenvectors reach some aggregator (AGG), or otherwise terminal site in the network.

Algorithm 6 Decentralized group ICA algorithm (dgICA)

Require: s sites with data $\{\mathbf{X}_i \in \mathbb{R}^{d \times N_i} : i = 1, 2, \ldots, s\}$, intended final rank r, local site rank $k_2 \geq r$, local subject rank k_1.
 1: **for all** sites $i = 1, 2, \ldots, s$ **do**
 2:     Perform LocalPCA (Baker et al., 2015) on each site $\rightarrow$ k_1 eigen-vectors for each subject.
 3:     Perform LocalPCA (Baker et al., 2015) on concatenated subjects $\rightarrow k_2$ eigenvectors at each site.
 4:     Reduce local data set to $\mathbf{X}_{i,red} \in \mathbb{R}^{d \times k_2}$
 5: **end for**
 6: Perform GlobalPCA (Baker et al., 2015) to obtain r global eigenvectors, $\mathbf{V}$, at the aggregator.
 7: On the aggregator, perform ICA to obtain global unmixing matrix, $\mathbf{W}$.

The aggregator site then performs whitening on these resulting eigenvectors, and runs a local ICA algorithm, such as infomax ICA (Bell and Sejnowski, 1995), to produce the spatial unmixing matrix, $\mathbf{W}$. The global spatial eigenvectors, $\mathbf{V}$, are then unmixed to produce $\hat{\mathbf{A}}$ by computing $\hat{\mathbf{A}} \approx \mathbf{VW}$, which is shared across the decentralized network. Each site then uses this unmixing matrix to produce individual time-courses for each i-th subject by computing $\mathbf{A}_i \approx \mathbf{X}_i^T \mathbf{S}$. Each site can then perform spatio-temporal regression back reconstruction approach (Calhoun et al., 2001; Erhardt et al., 2011) to produce subject-specific spatial maps.

2.2.2. Decentralized Clustering
In order to perform dFNC in a decentralized paradigm, we first require a notion of decentralized clustering. Following the precedent of previous work in dFNC, we focus first on decentralized K-Means optimization, for which there exist a number of pre-established methods for decentralization. A number of methods utilize some manner of weighted centroid averaging, where each site in the network broadcasts updated centroids to an aggregator node which then computes the merged centroids, and rebroadcasts them to the local sites (Forman and Zhang, 2000; Dhillon and Modha, 2000; Jagannathan and Wright, 2005), though completely peer-to-peer approaches have also been proposed (Datta et al., 2006, 2009), as well as methods robust to asynchronous updates (Di Fatta et al., 2013). Though we have not found any methods which do this, methods which compute K-Means via gradient descent (Bottou, 2010) are also amenable to decentralization (Yuan et al., 2016). For simplicity's sake, we take the approach of centroid-averaging outlined in Dhillon and Modha (2000), and leave rigorous presentation and comparison of the remaining methods as future work.

To perform clustering for distributed dFNC, we first have each site separate its subjects into sliding-window time-courses, where the window length is fixed across the decentralized network. Additionally, initial clustering was performed on a subset of windows from each subject, corresponding to windows of maximal variability in correlation across component pairs. To obtain these exemplars, each site computes variance of dynamic

---

Algorithm 7 Decentralized dFNC algorithm (ddFNC)

Require: s sites with data $\{\mathbf{X}_i \in \mathbb{R}^{d \times N_i} : i = 1, 2, \ldots, s\}$, win-size t, number of clusters k.

1: dgICA $\rightarrow \mathbf{W}$, global unmixing matrix, broadcast to sites.
2: **for all** sites $= i = 1, 2, \ldots, s$ **do**
3:     Back-reconstruct subject TCs
4:     Using sliding window of size t, obtain $r \times r$ covariance matrices.
5:     Obtain exemplar covariance matrices (Damaraju et al., 2014).
6: **end for**
7: Run K-Means on exemplar covariance matrices to obtain k initial centroids, C_0.
8: Run K-Means with initial clusters C_0 to obtain k centroids C, and clustering assignment for each instance, L.

---

connectivity across all pairs of components at each window. We then select windows corresponding to local maxima in this variance time-course. This resulted in an average of 8 exemplar windows per subject. We then perform decentralized K-Means on the exemplars to obtain a set of centroids, which are shared across the decentralized network, which we feed into a second stage of K-Means clustering.

For the second stage of decentralized clustering, at each iteration, each site computes updated centroids according to Dhillon and Modha (2000), which corresponds to a local K-Means update. These local centroids are then sent to the aggregator node, which computes the weighted average of these updated centroids, and re-broadcasts the updated global centroids until convergence.

2.2.3. Bandwidth and Complexity
To compute the communication and complexity for ddFNC, we separately analyse the novel component algorithms of dgICA and dK-Means.

For decentralized group ICA, the communication of the algorithm is closely related to the communication of GlobalPCA. In the GlobalPCA algorithm given in Baker et al. (2015), each site communicates a $d \times r$ matrix of eigenvectors to the subsequent site until the aggregator is reached. After the aggregator performs ICA to obtain the global unmixing matrix, $\mathbf{W}$, this matrix is broadcast to all other sites in the network. Thus, for a single, non-aggregator site, the total communication for dgICA is exactly $d \times r + r^2$. At the aggregator, the total communication is exactly $d \times r + r^2 \times s$ if the unmixing matrix is broadcast directly to each node. Of course, this cost could be mitigated by following a peer to peer communication scheme, and having other non-aggregator sites broadcast the unmixing matrix as well.

Next, we can compute the overall complexity of dgICA as the total complexity of local site operations. Consider an individual site, i, with m subjects, where the concatenated matrix is given as $\mathbf{X}_i \in \mathbb{R}^{d \times N_i}$. In general, the complexity of SVD on the $N_i \times N_i$ covariance matrix is $\mathcal{O}(N_i^3)$, though this can be improved upon by using iterative methods, such as the MATLAB *svds* function. Thus, the complexity for the two-stage LocalPCA computation on one site is $\mathcal{O}(2N_i^3)$. The per-site complexity for GlobalPCA is given as the complexity of a SVD computed on a $d \times d$ covariance matrix, which is created by concatenating the k_2 eigenvectors from the previous site; i.e., the per-site complexity for GlobalPCA is $\mathcal{O}(d^3)$. Finally, the complexity of ICA is exactly equal to the number of ICA iterations, $\mathcal{J}$, which depends heavily on the choice of ICA algorithm, and hyper-parameter selection (see Bell and Sejnowski, 1995 for more details on the complexity of Infomax, for example). Thus, the total per-site complexity for dgICA is $\mathcal{O}(N_i^3 + d_i^3)$ for non-aggregator nodes, and $\mathcal{O}(N_i^3 + d_i^3 + \mathcal{J})$ on the aggregator node. The overall runtime of dgICA is thus dependent on the computational resources available at each site, as well as the computational resources and ICA parameters chosen by the aggregator site.

Prior to performing K-Means, each site i computes $N_{i,j} - w$ windowed time-courses of length w on each subject j, computing the rank r covariance matrix for those windows. Thus, if there are

m_i subjects at site i, the local complexity is $\mathcal{O}(m_i(N - w)r^3)$ for this operation. No inter-site communication occurs during this process.

For decentralized K-Means, the communication between sites depends on the number of "K-Means Iterations," $\mathcal{J}$, i.e., the number of iterations required for the centroids to stabilize. $\mathcal{J}$ depends heavily on the initial centroids, the distance metric used, the distribution of the global data set, and other factors which make it difficult to compute exactly for arbitrary data. In each iteration of decentralized K-Means, we communicate k many centroids of size $\mathbb{R}^{r^2}$, for an average communication of $r^2 \cdot k \cdot \mathcal{J}$ from the sites to the aggregator. The aggregator, then, performs a total of $r^2 \cdot k \cdot \mathcal{J} \cdot s$ communication (Dhillon and Modha, 2000), which again, could be mitigated by passing centroids to intermediate sites, provided those sites can be trusted with the centroid information.

The time complexity of decentralized K-Means is described in Dhillon and Modha (2000). At each site, the distance and centroid recalculation computations come out to per-site complexity of $\mathcal{O}((3kr^2 + M_ik + M_ir^2 + kr^2) \cdot \mathcal{J})$ (Dhillon and Modha, 2000), where M_i is the number of instances at site i. The total number of computations consists of the sum of these site-wise complexities, and the centroid-averaging step with a complexity of $\mathcal{O}(kr^2)$, for a total of $\mathcal{O}((3kr^2 + Mk + Mr^2 + kr^2) \cdot \mathcal{J})$, where M is the total number of data instances in the decentralized network.

Since dK-Means is computed twice for full ddFNC, once on the exemplars, and once on the global set of subject windows, the complete complexity of the clustering stage of the algorithm is given as the dK-Means complexity for $M = \sum E_i$ added to the dK-Means complexity for $M = \sum m_i$, i.e., $\mathcal{O}((3kr^2 + (\sum E_i + \sum m_i)(k + r^2) + kr^2) \cdot \mathcal{J} + kr^2)$.

The overall site-wise complexity and communication for ddFNC is just the sum of the site-wise communication and complexities for each of the stages described here. In the paradigm described here, the communication and complexity on the aggregator is generally more demanding than that on the individual sites, which makes sense for cases where the aggregator has sufficient and reliable network and hardware resources. In cases where this is not necessarily true, some of the aggregation tasks can be distributed to other sites in the network, thus reducing communication and complexity on the final aggregator. In the dgICA algorithm, performing ICA on the aggregator may become a bottleneck if the aggregator does not have sufficient computational resources to perform a standard run of ICA; however, this problem could be mitigated by performing a hardware check on sites in the consortium, and assigning the role of aggregator dynamically based on availability of computational resources. For more discussion of the particularities of network communication and other issues which may arise in decentralized frameworks like the one used for ddFNC, see Plis et al. (2016).

3. DATA

3.1. Structural MRI for Decentralized VBM

As part of validating the proof-of-concept, we applied decentralized VBM to brain structure data collected on chronic schizophrenic patients and healthy controls. Specifically, the data comes from the Mind Clinical Imaging Consortium (MCIC) collection- a publicly accessible, on-line data repository containing curated anatomical and functional MRI, in addition to other data, collected from individuals with and without a schizophrenia spectrum disorder (Gollub et al., 2013) and available via the COINS data exchange https://coins.mrn.org (Scott et al., 2011).

Although more information about the MCIC can be found in Gollub et al. (2013), here we will report numbers for the final data used in this study as some subjects were excluded during the preprocessing phase. The final cohort for whom data are available includes 146 patients and 160 controls with site distribution as follows: Site B (IA) 40 patients/67 controls; Site D (MGH) 32/23; Site C (UMN) 32/26; Site A (UNM) 42/44, respectively. All subjects provided informed consent to participate in the study that was approved by the human research committees at each of the sites.

Briefly, T1-weighted structural MRI (sMRI) images were acquired with the following scan parameters: $TR = 2,530\,ms$ for 3 T, $TR = 12\,ms$ for 1.5 T; $TE = 3.79\,ms$ for 3 T, $TE = 4.76\,ms$ for 1.5 T; $FA = 7°$ for 3 T, $FA = 20°$ for 1.5 T; $TI = 1100\,ms$ for 3 T; $Bandwidth = 181$ for 3 T, $Bandwidth = 110$ for 1.5 T; $voxelsize = 0.625 \times 0.625\,mm$; slice thickness 1.5 mm; $FOV = 16 - 18cm$.

The T1-weighted sMRI data were preprocessed using the Statistical Parametric Mapping software using unified segmentation (Ashburner and Friston, 2005), in which image registration, bias correction and tissue classification were performed using a single integrated algorithm resulting in individual brains segmented into gray matter, white matter and cerebrospinal fluid and nonlinearly warped to the Montreal Neurological Institute (MNI) standard space. The resulting gray matter concentration (GMC) images were re-sliced to $2 \times 2 \times 2mm$, resulting in $91 \times 109 \times 91$ voxels. Although one can obtain both modulated (Jacobian corrected) and unmodulated gray matter segmentations, in this study, we use unmodulated GMC maps to test our regression models.

To test the decentralized regression on the MCIC data described in the previous paragraph, we regress the age, diagnosis, gender and the site covariates on the voxel intensities ($\sim$600,000 voxels). All the decentralized computations discussed here have been performed on a single machine.

3.2. Functional MRI for dFNC

To evaluate ddFNC , we utilize imaging data from Damaraju et al. (2014) collected from 163 healthy controls (117 males, 46 females; mean age: 36.9 years) and 151 age- and gender matched patients with schizophrenia (114 males, 37 females; mean age: 37.8 years), for a total of 314 subjects.

The scans were collected during an eyes closed resting fMRI protocol at 7 different sites across United States and pass data quality control (see **Supplementary Material**). Informed and written consent was obtained from each participant prior to scanning in accordance with the Internal Review Boards of corresponding institutions (Keator et al., 2016). A total of 162 brain-volumes of echo planar imaging BOLD fMRI data were collected with a temporal resolution of 2 s on 3-Tesla scanners.

Imaging data for six of the seven sites was collected on a 3T Siemens Tim Trio System and on a 3T General Electric Discovery MR750 scanner at one site. Resting state fMRI scans were acquired using a standard gradient-echo echo planar imaging paradigm: FOV of 220 × 220 mm (64 × 64 matrix), TR = 2 s, TE = 30 ms, FA = 770, 162 volumes, 32 sequential ascending axial slices of 4 mm thickness and 1 mm skip. Subjects had their eyes closed during the resting state scan. Data preprocessing for dgICA was performed according to the preprocessing steps in Damaraju et al. (2014).

3.3. ddFNC Experimental Parameters

We verify that ddFNC can generate sensible dFNC clusters by replicating the centroids produced in Damaraju et al. (2014). We run both pooled and decentralized versions of our algorithm, and compare our results directly with the results provided by the authors of Damaraju et al. (2014). We thus closely follow the experimental procedure in Damaraju et al. (2014), with some of the additional post-processing omitted for simplicity. To evaluate

the success of our pipeline, we run a simple experiment where we implement the ddFNC pipeline end-to-end on the data, simulating 314 subjects being evenly shared over 2 decentralized sites.

We set a window-length of 22 time-points (44 s), for a total of 140 windows per subject. For dgICA, we first estimate 120 subject-specific principal components locally, and reduce each subject to 120 points in the temporal dimension. Subjects are then concatenated temporally on each site, and we use the GlobalPCA algorithm in Baker et al. (2015) to estimate 100

TABLE 1 | Correlation between *SSE* from pooled, single-shot and multi-shot regression.

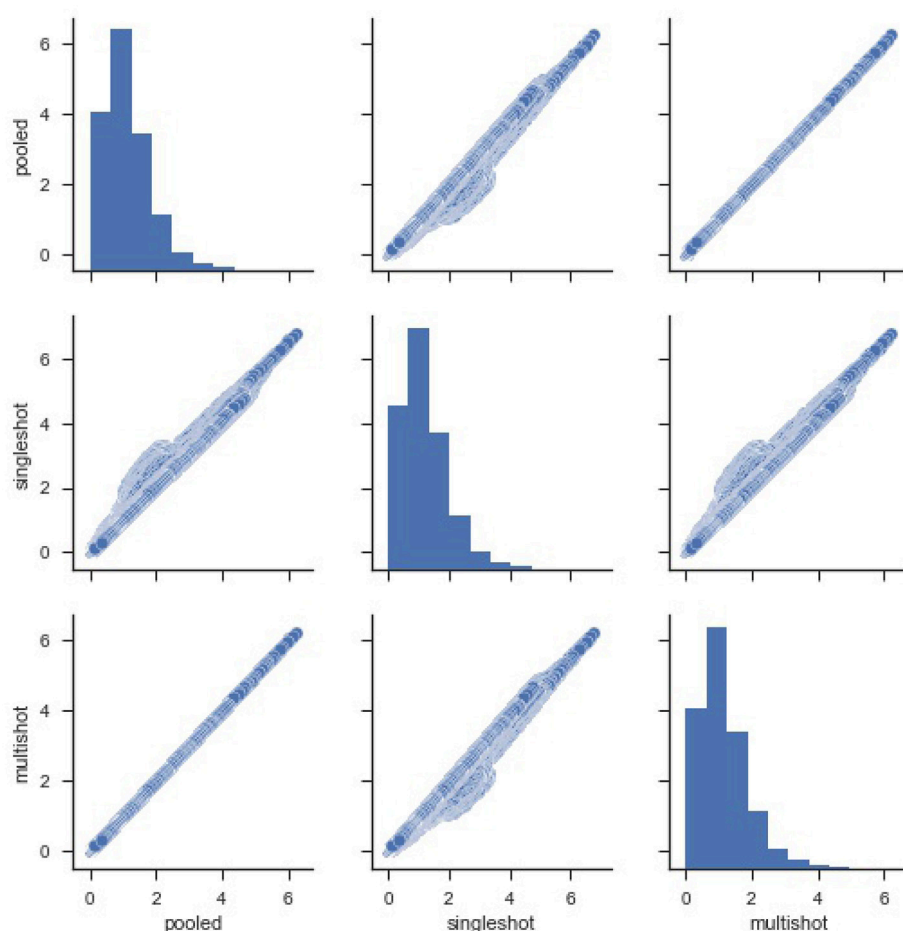|	Pooled	Single-shot	Multi-shot
Pooled	1.000000	0.992905	1.000000
Single-shot	0.992905	1.000000	0.992905
Multi-shot	1.000000	0.992905	1.000000

FIGURE 1 | Pairwise plot of Sum Square of Errors (SSE) from pooled, single-shot and multi-shot regression. Although the distribution plot looks similar across the three regressions, the pooled regression vs. multi-shot regression scatter plot demonstrates how identical they are to each other.The scatter plot of pooled regression vs. single-shot regression demonstrates that the SSE values obtained from singles-shot regression are on the higher side compared to the values from pooled regression.

spatial components, and perform whitening. We then use local infomax ICA (Bell and Sejnowski, 1995) on the aggregator to estimate the unmixing matrix **W**, and estimate 100 spatially independent components, **Â**. We then broadcast **Â** back to the local sites, and each site computes subject-specific time-courses.

After spatial ICA, we have each site perform a set of additional post-processing steps prior to decentralized dFNC. First, we select 47 components from the initial 100, by computing components which are most highly correlated with the components from Damaraju et al. (2014). We then have each site drop the first 2 points from each subject, regress subject head movement parameters with 6 rigid body estimates, their derivatives and squares (total of 24 parameters). Additionally, any spikes identified are interpolated using 3rd order spline fits to good neighboring data, where spikes are defined as any points exceeding mean (FD) + 2.5 *std(FD) , where FD is framewise displacement [interpolating 0 to 9 points (mean, sd: 3, 1.76)].

For clustering, we forgo a separate elbow-criterion estimation, and use the optimal number of clusters from Damaraju et al. (2014), setting $k = 5$. For the exemplar stage of clustering, we evaluate 200 runs where we initialize centroids uniformly randomly from local data, and then run dK-Means using the cluster averaging strategy in Dhillon and Modha (2000). For our distance measure, we use scikit-learn (Pedregosa et al., 2011) to compute the correlation distance between covariance matrices following the methods in Damaraju et al. (2014). To keep our implementation simple, unlike Damaraju et al. (2014), we do not utilize graphical LASSO to estimate the covariance matrix, and thus do not optimize for any regularization

parameters. Additionally, we do not perform additional Fisher-Z transformations or perform additional regularization using a previously computed static dFNC result. Future implementations may also utilize a decentralized static functional network connectivity (sFNC) algorithm as preprocessing, as is done for the pooled case in Damaraju et al. (2014). Finally, for the second stage of dK-Means, we initialize using the centroids from the run with the highest silhouette score, computed using the scikit-learn python toolbox (Pedregosa et al., 2011), again running dK-Means to convergence. After computing the centroids, we use the correlation distance and the Hungarian matching algorithm (Kuhn, 1955) to match both plotted spatial components from dgICA and the resulting centroids from dK-Means.

4. RESULTS

4.1. Decentralized VBM Results

For starters, in order to compare the efficacy of each regression (single-shot and multi-shot) against the pooled case, we present a simple pairwise plot of the *SSE* of the regression performed on every voxel, **Figure 1**. In mathematical terms, the *SSE* represents lowest objective function value that could be attained from the regression model. It can be seen from **Figure 1** that the *SSE* from multi-shot and pooled/centralized regression lie perfectly along a diagonal indicating the parameters obtained from them are identical. This can also be verified from **Table 1** showing the correlation between the different *SSE*s. Please note that results from the decentralized regression with normal equation were not presented as it has been mathematically shown to be equivalent to that of a pooled regression.
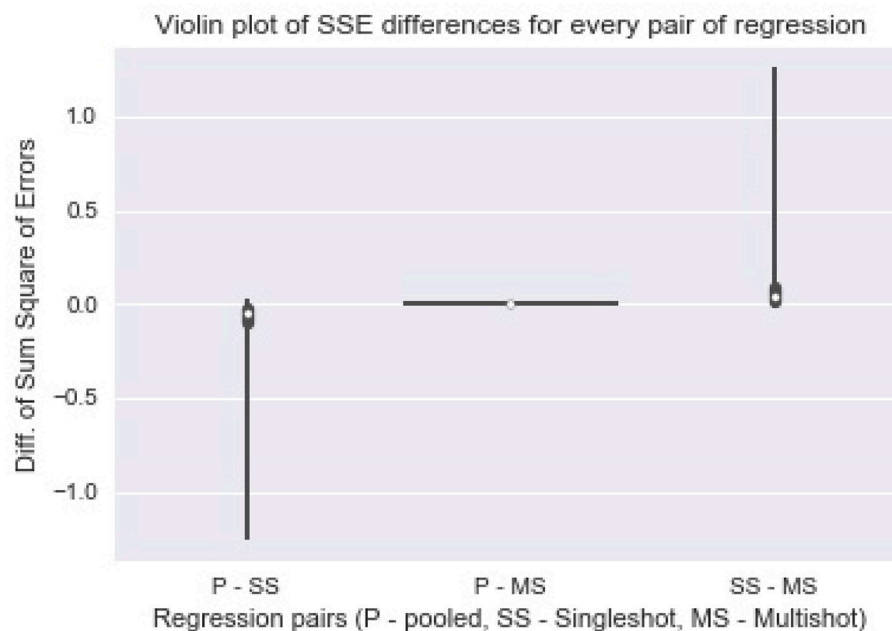

FIGURE 2 | Violin plot of Sum Square of Error differences between every pair of regression. The plot of differences in SSE from pooled regression and multi-shot regression (P-MS) centered around 0 demonstrates how identical the results from the two regressions are. On the other hand, the SSE values from single-shot regression are higher compared to those from the pooled regression.

It can be seen that the correlation between *SSE* from the centralized regression and multi-shot is 1. On the other hand, it can also be noticed that the *SSE* correlations between single-shot and pooled or single-shot and multi-shot are slightly lower than perfect correlation. The single-shot approach can be considered to be similar to a meta-analysis, whereas the multi-shot approach is basically a mega-analysis (i.e., equivalent to the pooled analysis).

Figure 2 shows a violin (distribution) plot of the difference in *SSE* from every pair of regression. Evidently, the differences in *SSE* between pooled and multi-shot regression are centered around 0. To reinforce our notion that the multi-shot is superior to single-shot we take a look at the R^2 values from the different regressions and compare. It can be seen from **Figure 3** that the R^2 values from multi-shot and pooled regression align perfectly along a diagonal (correlation = 1, refer to **Table 2**) or have exactly

the same distribution, whereas those from single-shot are all over the place.

As noted earlier, in addition to evaluating the regression model parameters, researchers will also be interested in understanding the statistical significance of the various parameter estimates. **Figures 4–6** show the statistical significance of each covariate (age, diagnosis and gender), from both

TABLE 2 | Correlation between R^2 from pooled, single-shot and multi-shot regression.

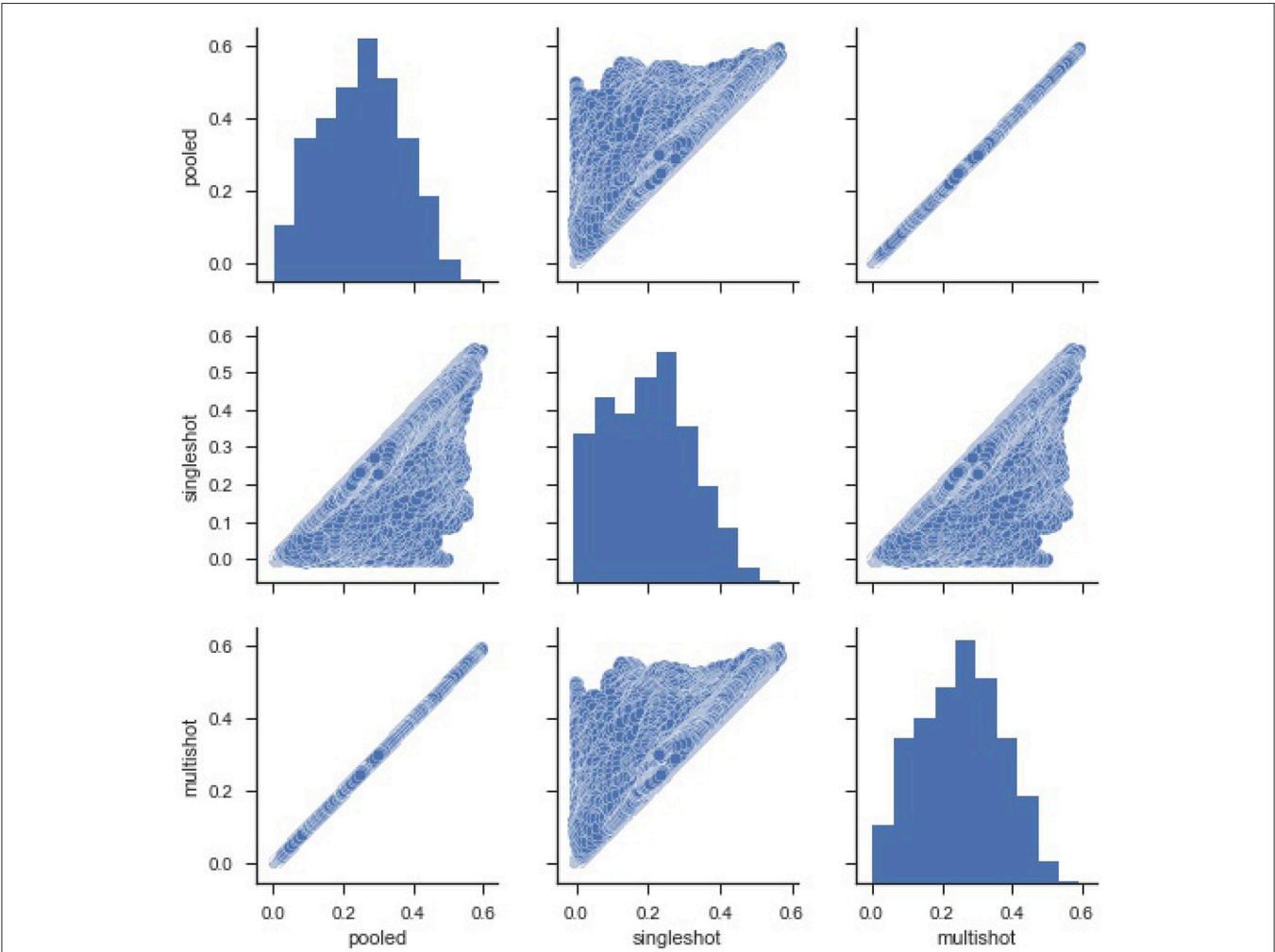|	Pooled	Single-shot	Multi-shot
Pooled	1.000000	0.906662	1.000000
Single-shot	0.906662	1.000000	0.906662
Multi-shot	1.000000	0.906662	1.000000

FIGURE 3 | Pairwise scatter plots of Coefficient of Determination R^2 from the three types of regression. It can be seen again that the R^2 values for the regressions from multi-shot regression and pooled regression are exactly equal. The R^2 values from single-shot regression are less than their corresponding values from pooled regression or multi-shot regression because the model being fit in single-shot has fewer covariates (Note, one of the limitations of the single-shot is that the site specific covariates could not be included as it introduces collinearity).
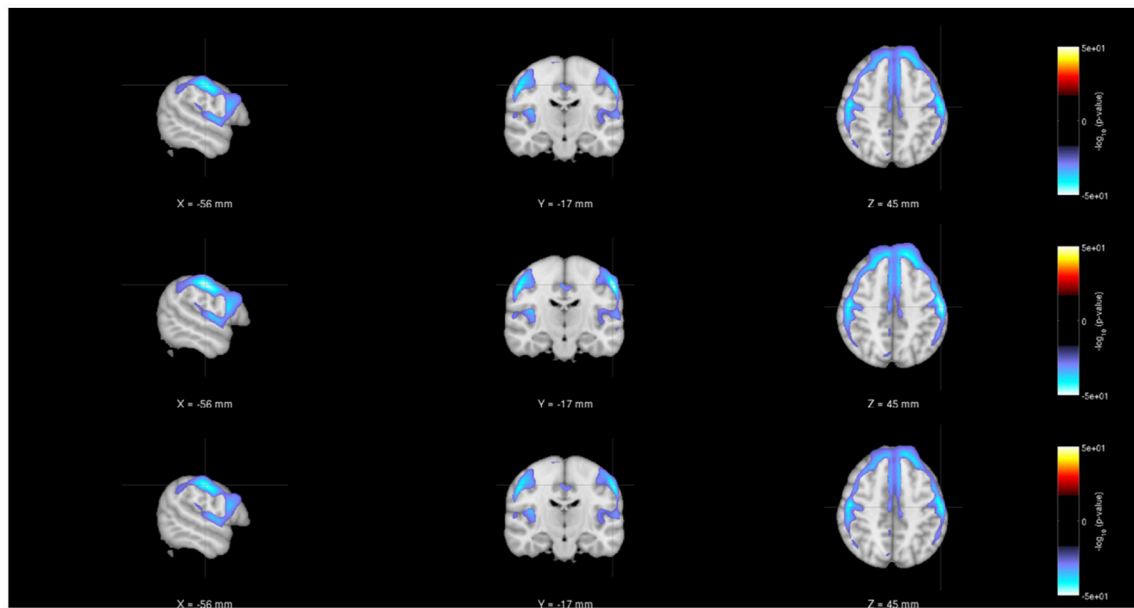
FIGURE 4 | Rendered images of voxel-wise significance values ($-log_{10}p$-value $\times$ sign(t)) for the covariate "Age" from pooled regression **(Top)** and single-shot regression **(Center)**, and multi-shot regression **(Bottom)** overlaid on MNI average template. One could see that the regions with expected gray matter decrease as age increases are similar from all kinds of regression. Although the single-shot regression uses fewer covariates, the similarity of the rendered images with those of pooled regression or multi-shot regression indicate the relative weight or orientation of the corresponding β coefficient will be similar to those from pooled/multi-shot regression.
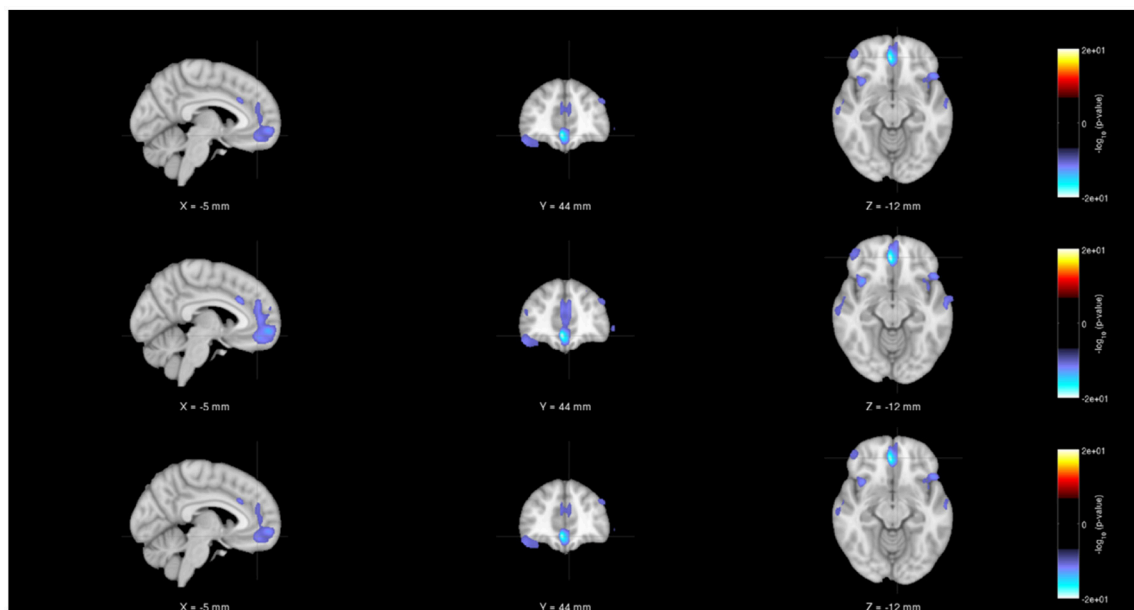

FIGURE 5 | Rendered images of voxel-wise significance values ($-log_{10}p$-value $\times$ sign(t)) for the covariate "Diagnosis" from pooled regression **(Top)** and single-shot regression **(Center)** and multi-shot regression **(Bottom)** overlaid on MNI average template. Regardless of the type of regression performed, the images indicate that in the medial frontal and bilateral temporal lobe/insula there is a significant gray matter density reduction for schizophrenic patients compared to the same regions of the healthy subjects.
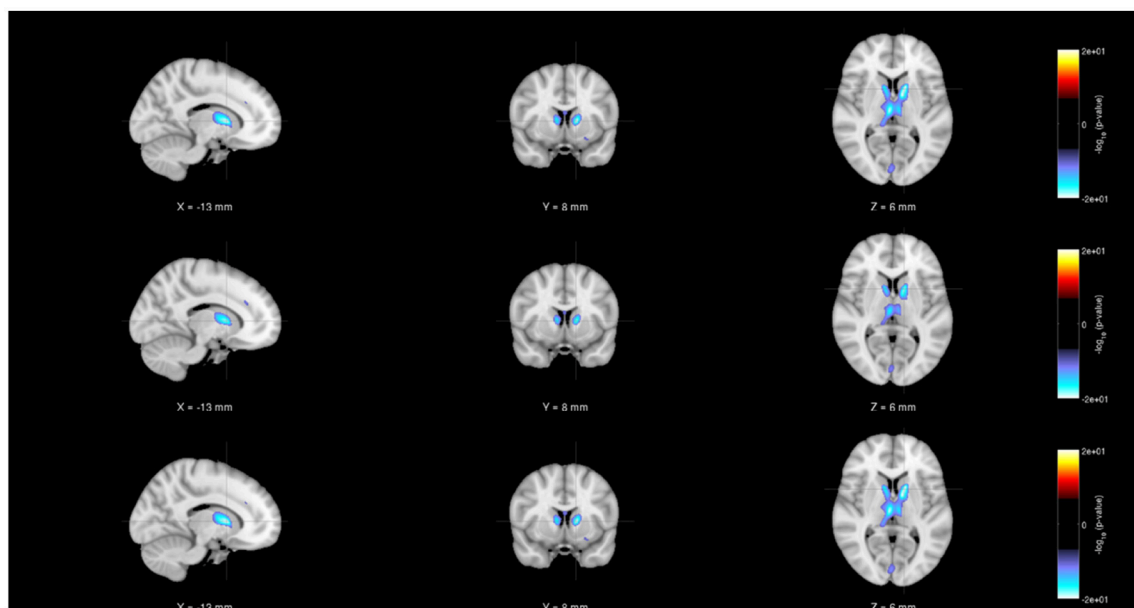
FIGURE 6 | Rendered images of voxel-wise significance values ($-log_{10}p$-value $\times$ sign(t)) for the covariate "Gender" from pooled regression **(Top)** and single-shot regression **(Center)** and multi-shot regression **(Bottom)** overlaid on MNI average template. It can be seen from all the three rendered images that there is a significant amount of gray matter reduction in the sub-cortical regions for males. Since we are using unmodulated gray matter maps, these sex differences could be due to changes in brain volumes.
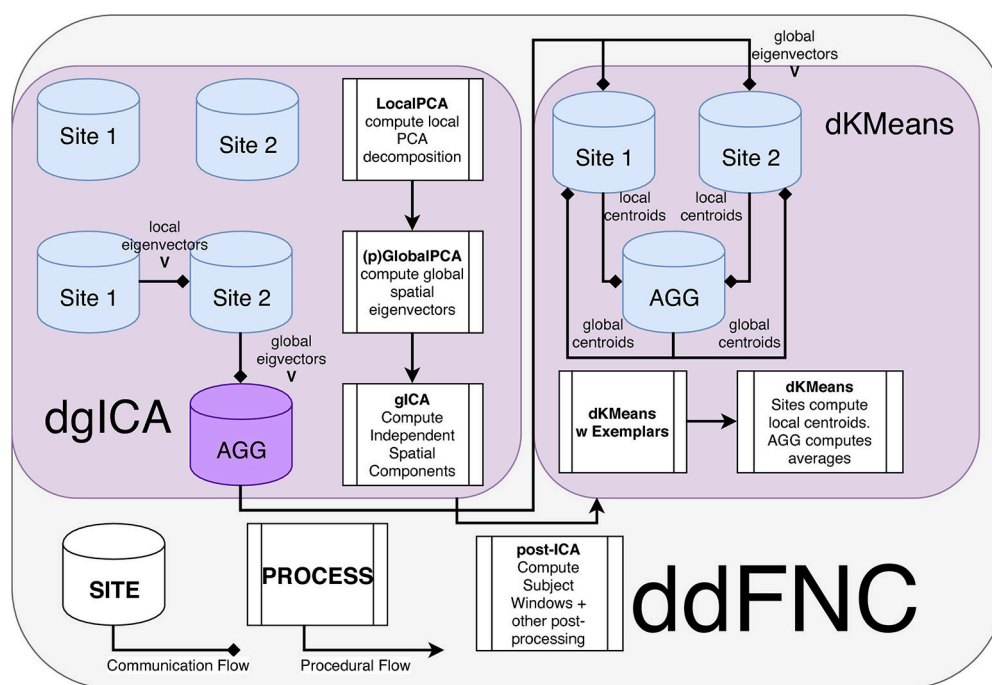

FIGURE 7 | Flowchart of the ddFNC procedure e.g., with 2 sites. To perform dgICA, sites first locally compute subject-specific LocalPCA to reduce the temporal dimension, and then use the GlobalPCA procedure from Baker et al. (2015) to compute global spatial eigenvectors, which are then sent to the aggregator. The aggregator then performs ICA on the global spatial eigenvectors, using InfoMax ICA (Bell and Sejnowski, 1995) for example, and passes the resulting spatial components back to local sites. The dK-Means procedure then iteratively computes global centroids using the procedure outlined in Dhillon and Modha (2000), first computing centroids from subject exemplar dFNC windows, and then using these centroids to initialize clustering over all subject windows.

centralized and decentralized regressions performed against each voxel, plotted on an MNI brain template. **Figure 4** shows the brain images with the $-(log_{10}p\text{-val} \times sign(t))$-values for the weight parameter corresponding to "Age." It is notable to see that the results from the multi-shot regression have a perfect correlation to those from the pooled version. Moreover, the observations show the expected decrease in gray matter concentration as age increases. **Figures 5**, **6** show the rendered images for $-log_{10}p$-values for the "Diagnosis" and "Gender" covariate, respectively.

4.2. ddFNC Results

A summary of the complete steps in the decentralized dFNC pipeline is given in **Figure 7**. In **Figure 8**, we plot some examples of the components estimated from decentralized spatial ICA in comparison with the spatial components from Damaraju et al. (2014), after performing Hungarian matching between the estimated spatial maps. We also plot the correlation of the components from our ICA implementation in comparison to the

components from Damaraju et al. (2014). Indeed, the estimated components are highly correlated with the results from Damaraju et al. (2014), for all 100 estimated components, as well for the 47 selected neurological components from Damaraju et al. (2014), indicating that dgICA is able to produce results comparable to the pooled case. We include additional spatial maps for all 47 estimated spatial components in the **Supplementary Material**.

In **Figure 9**, we plot the centroids from Damaraju et al. (2014) (**Figure 9A**), as well as the centroids estimated using decentralized dFNC (**Figure 9B**). Indeed, the centroids found using ddFNC prove similar to the centroids found in Damaraju et al. (2014), with centroids 2 and 3 being the closest matches under correlation distance.

5. DISCUSSION

The results described in the previous section demonstrate the fidelity of decentralized regression and decentralized dynamic function network connectivity in analyzing neuroimaging data.
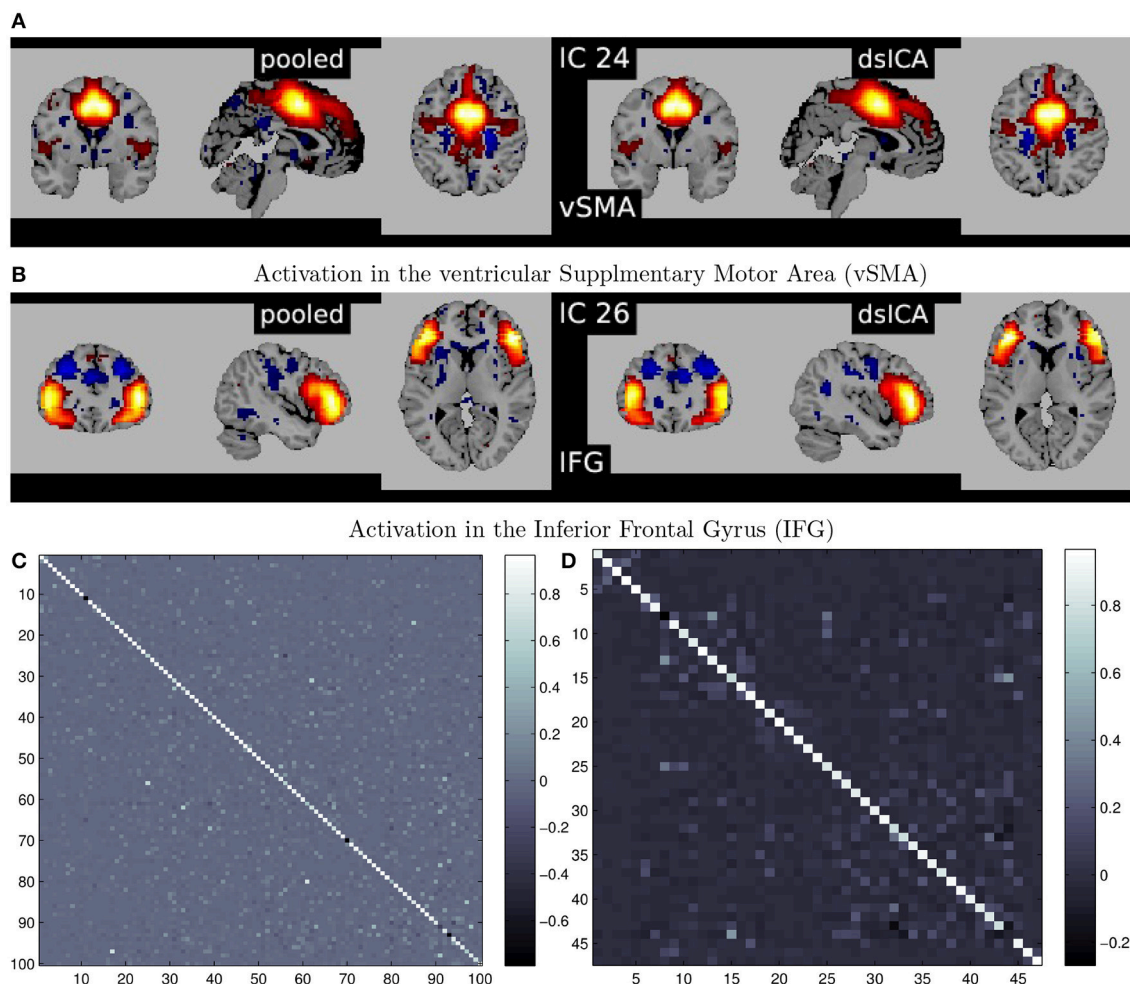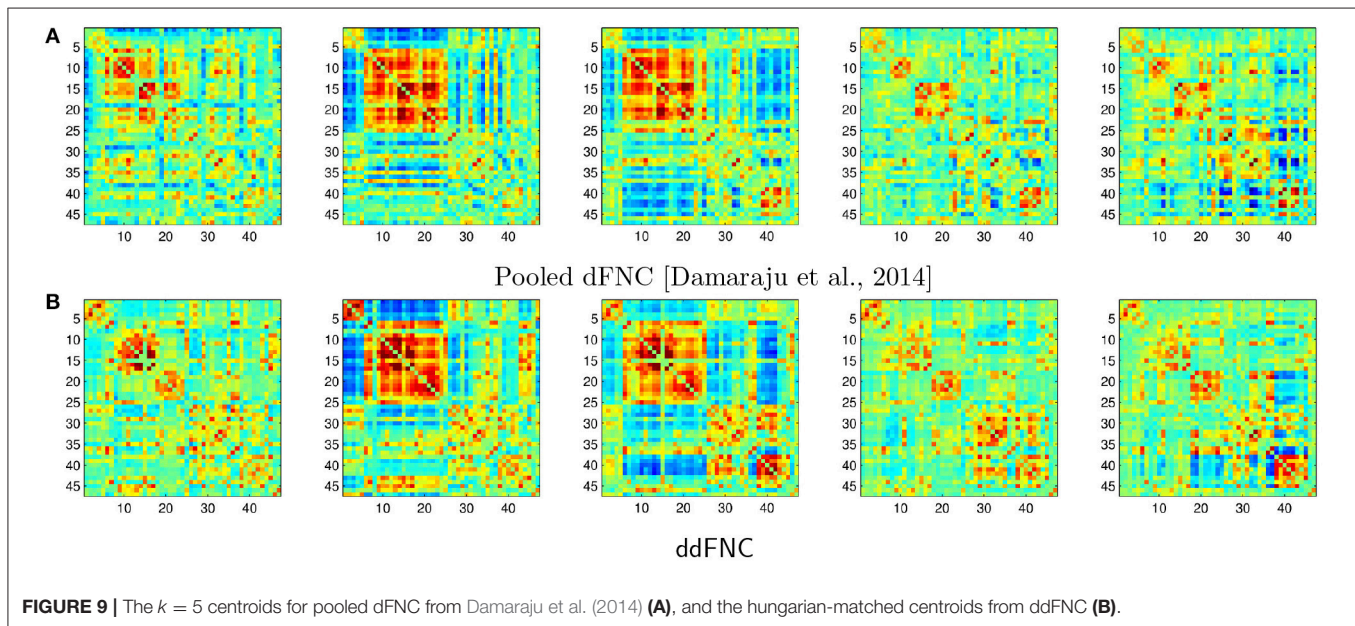

FIGURE 8 | (A,B) Illustrate examples of matched spatial maps from dgICA and pooled ICA. **(C,D)** Show the correlation of the components between pooled spatial ICA and dgICA after hungarian matching. **(C)** Shows correlation between all 100 components, and **(D)** Shows correlation between the 47 neurological components selected in Damaraju et al. (2014).

FIGURE 9 | The $k = 5$ centroids for pooled dFNC from Damaraju et al. (2014) **(A)**, and the hungarian-matched centroids from ddFNC **(B)**.

Although single-shot regression is simple and easy to implement, it limits our ability to incorporate site covariates and thus might not be extremely helpful. The decentralized regression with normal equation and multi-shot regression are superior to single-shot regression because not only do they allow incorporating site related variables but also give exact results as the pooled regression. The linearity and convexity of the regression objective function made this possible and thus are an excellent alternative to perform regression on multi-site datasets.

In terms of the regression objective function, either the sum of squared errors or mean sum of squared errors can be used in practice. However, it's mathematically convenient to use sum of squared errors which subsequently entails (at the *AGG*) a simple addition of the gradients ($\mathcal{O}(1)$) instead of a weighted average of the gradients ($\mathcal{O}(n)$). Added to that, we also showed how the sample size at the local sites has no bearing on the final results.

On a more practical note, the need for multi-shot regression might not arise often in a neuroimaging setting where the number of covariates is usally small. In such cases, the decentralized regression with normal equation will suffice. However, in decentralized settings where the number of covariates is usually large (machine learning/big data) the multi-shot regression comes to the fore. From a computational time standpoint, and as discussed in the computational complexity section, it should be obvious that the multi-shot regression takes more time to complete than the decentralized regression with normal equation as it involves iteratively passing the gradients between the local nodes and the *AGG*. It is worth mentioning that although the decentralized regression algorithms demonstrated here pertain to a simple linear regression model, these algorithms can easily be extended to more complex models with polynomial terms or interaction terms as well as to ridge regression, lasso regression, and elastic net regression.

Regarding ddFNC, we plan on performing a more robust analysis, going into the future, as a stand-alone algorithm,

particularly with respect to different variations on the dK-Means optimization and initialization, or with differing versions of ICA on the aggregator (*AGG*) node, such as fastICA (Koldovský et al., 2006), Entropy Bound Minimization (Li and Adali, 2010), and others. Additionally, the possibility of performing a decentralized static FNC either as a preprocessing step to ddFNC or a separate analysis is attractive. One other avenue worth exploring with ddFNC is the flow of information across the decentralized network. In particular, since the GlobalPCA step in dgICA already makes the procedure partially peer-to-peer, it makes sense to explore adding this functionality to the dK-Means methods to preserve this peer-to-peer structure. Finally, we plan to evaluate privacy-sensitive versions of ddFNC, utilizing differential-privacy or other privacy measures as a way to perform these analyses with some assurance of per-subject privacy in the decentralized network.

Finally, we note that the decentralization of algorithms in a neuroimaging setting emphasizes the importance of analysis on data present at multiple sites, the decentralization discussed herewith is no different from other decentralized algorithms discussed elsewhere in literature. The *AGG* is not really a master node *per se* but in fact one of the local sites itself. The term *AGG* was introduced to separate all the other local sites from that site where the results are accumulated.

6. CONCLUSION

In this paper, we presented a simple case study of how voxel-based morphometry and dynamic functional network connectivity analysis can be performed on multi-site data without the need for pooling data at a central site. The study shows that both the decentralized voxel-based morphometry as well as the decentralized dynamic functional network connectivity yield results that are comparable to its pooled counterparts guaranteeing a virtual pooled analysis effect by a chain of computation and communication process. Other advantages

of such a decentralized platform include data privacy and support for large data. In conclusion, the results presented here strongly encourage the use of decentralized algorithms in large neuroimaging studies over systems that are optimized for large-scale centralized data processing.

ETHICS STATEMENT

For the MCIC data, all subjects provided informed consent to participate in the study that was approved by the human research committees at each of the sites (UNM HRRC #03-429; UMinn IRB #0404M59124; MGH IRB# 2004P001360; UIowa IRB #1998010017). In addition to the informed consent, all patients successfully completed a questionnaire verifying that they understood the study procedures.

For fBIRN data, all subjects provided informed consent to participate in the study that was approved by the human research committees of each of the participating institutes in the fBIRN data repository.

AUTHOR CONTRIBUTIONS

HG implemented the decentralized regression algorithms on structural MRI data and wrote the regression part of the paper. BB implemented the decentralized dynamic functional network connectivity pipeline on functional MRI data and wrote that part of the paper. ED contributed immensely to the analysis as well as interpretation of the results from both decentralized regression and decentralized dFNC pipeline. SRP contributed to the brain imaging data preprocessing pipeline. SMP proposed the decentralized data analysis system and led the algorithm development effort. RS helped formulate the decentralized regression with normal equation and development of decentralized spatial ICA. VC led the team and formed the vision.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf.2018.00055/full#supplementary-material

REFERENCES

Adams, H. H., Adams, H., Launer, L. J., Seshadri, S., Schmidt, R., Bis, J. C., et al. (2016). Partial derivatives meta-analysis: pooled analyses when individual participant data cannot be shared. bioRxiv 038893. doi: 10.1101/038893

Allen, E. A., Damaraju, E., Plis, S. M., Erhardt, E. B., Eichele, T., and Calhoun, V. D. (2014). Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676. doi: 10.1093/cercor/bhs352

Ashburner, J., and Friston, K. J. (2000). Voxel-based morphometry–the methods. *Neuroimage* 11, 805–821. doi: 10.1006/nimg.2000.0582

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *NeuroImage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Baker, B. T., Silva, R. F., Calhoun, V. D., Sarwate, A. D., and Plis, S. M. (2015). "Large scale collaboration with autonomy: Decentralized data ICA," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, (Boston, MA: IEEE), 1–6.

Bell, A. J., and Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* 7, 1129–1159. doi: 10.1162/neco.1995.7.6.1129

Bottou, L. (2010). "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010* (Paris: Springer), 177–186.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14:365. doi: 10.1038/nrn3475

Calhoun, V. D., and Adali, T. (2012). Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. *IEEE Rev. Biomed. Eng.* 5, 60–73. doi: 10.1109/RBME.2012.2211076

Calhoun, V. D., Adali, T., Pearlson, G. D., and Pekar, J. (2001). A method for making group inferences from functional mri data using independent component analysis. *Hum. Brain Mapp.* 14, 140–151. doi: 10.1002/hbm.1048

Carter, K. W., Francis, R. W., Carter, K., Francis, R., Bresnahan, M., Gissler, M., et al. (2015). Vipar: a software platform for the virtual pooling and analysis of research data. *Int. J. Epidemiol.* 45, 408–416. doi: 10.1093/ije/dyv193

Cragin, M. H., Palmer, C. L., Carlson, J. R., and Witt, M. (2010). Data sharing, small science and institutional repositories. *Philos. Trans. R. Soc. Lond. A Math. Phys. Eng. Sci.* 368, 4023–4038. doi: 10.1098/rsta.2010.0165

Damaraju, E., Allen, E. A., Belger, A., Ford, J., McEwen, S., Mathalon, D., et al. (2014). Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia. *NeuroImage* 5, 298–308. doi: 10.1016/j.nicl.2014.07.003

Datta, S., Giannella, C., and Kargupta, H. (2006). "K-means clustering over a large, dynamic network," in *Proceedings of the 2006 SIAM International Conference on Data Mining* (SIAM), 153–164. doi: 10.1137/1.9781611972764.14

Datta, S., Giannella, C., and Kargupta, H. (2009). Approximate distributed k-means clustering over a peer-to-peer network. *IEEE Trans. Knowl. Data Eng.* 21, 1372–1388. doi: 10.1109/TKDE.2008.222

Deco, G., Ponce-Alvarez, A., Mantini, D., Romani, G. L., Hagmann, P., and Corbetta, M. (2013). Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations. *J. Neurosci.* 33, 11239–11252. doi: 10.1523/JNEUROSCI.1091-13.2013

Dhillon, I. S., and Modha, D. S. (2000). "A data-clustering algorithm on distributed memory multiprocessors," in *Large-Scale Parallel Data Mining*, Workshop on Large-Scale Parallel KDD Systems, SIGKDD (Berlin; Heidelberg: Springer), 245–260.

Di Fatta, G., Blasa, F., Cafiero, S., and Fortino, G. (2013). Fault tolerant decentralised k-means clustering for asynchronous large-scale networks. *J. Paral. Distribut. Comput.* 73, 317–329. doi: 10.1016/j.jpdc.2012.09.009

Duchi, J., Hazan, E., and Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* 12, 2121–2159.

Erhardt, E. B., Rachakonda, S., Bedrick, E. J., Allen, E. A., Adali, T., and Calhoun, V. D. (2011). Comparison of multi-subject ica methods for analysis of fMRI data. *Hum. Brain Mapp.* 32, 2075–2095. doi: 10.1002/hbm.21170

Fennema-Notestine, C., Gamst, A. C., Quinn, B. T., Pacheco, J., Jernigan, T. L., Thal, L., et al. (2007). Feasibility of multi-site clinical structural neuroimaging studies of aging using legacy data. *Neuroinformatics* 5, 235–245. doi: 10.1007/s12021-007-9003-9

Forman, G., and Zhang, B. (2000). Distributed data clustering can be efficient and exact. *ACM SIGKDD Explor. Newsl.* 2, 34–38. doi: 10.1145/380995.381010

Gollub, R. L., Shoemaker, J. M., King, M. D., White, T., Ehrlich, S., Sponheim, S. R., et al. (2013). The mcic collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Neuroinformatics* 11, 367–388. doi: 10.1007/s12021-013-9184-3

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., et al. (2015). Common genetic variants influence human subcortical brain structures. *Nature* 520, 224. doi: 10.1038/nature14101

Jagannathan, G., and Wright, R. N. (2005). "Privacy-preserving distributed k-means clustering over arbitrarily partitioned data," in *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, KDD'05* (Chicago, IL: ACM), 593–599.

Keator, D. B., van Erp, T. G., Turner, J. A., Glover, G. H., Mueller, B. A., Liu, T. T., et al. (2016). The function biomedical informatics research network data repository. *Neuroimage* 124, 1074–1079. doi: 10.1016/j.neuroimage.2015.09.003

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. arXiv:1412.6980.

Koldovský, Z., Tichavský, P., and Oja, E. (2006). Efficient variant of algorithm fastica for independent component analysis attaining the cramér-rao lower bound. *IEEE Trans. Neural Netw.* 17, 1265–1277. doi: 10.1109/TNN.2006.875991

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Res. Logist. Q.* 2, 83–97. doi: 10.1002/nav.3800020109

Landis, D., Courtney, W., Dieringer, C., Kelly, R., King, M., Miller, B., et al. (2016). Coins data exchange: an open platform for compiling, curating, and disseminating neuroimaging data. *NeuroImage* 124, 1084–1088. doi: 10.1016/j.neuroimage.2015.05.049

Lewis, N., Plis, S., and Calhoun, V. (2017). "Cooperative learning: Decentralized data neural network," in *2017 International Joint Conference on Neural Networks (IJCNN)* (Anchorage, AK), 324–331.

Li, X.-L., and Adali, T. (2010). Independent component analysis by entropy bound minimization. *IEEE Trans. Signal Process.* 58, 5151–5164. doi: 10.1109/TSP.2010.2055859

Ming, J., Verner, E., Sarwate, A., Kelly, R., Reed, C., Kahleck, T., et al. (2017). Coinstac: decentralizing the future of brain imaging analysis. *F1000Res.* 6:1512. doi: 10.12688/f1000research.12353.1

Nesterov, Y. (1983). A method for unconstrained convex minimization problem with the rate of convergence O(1/k^2). *Dokl. AN USSR* 269, 543–547.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830.

Plis, S. M., Sarwate, A. D., Wood, D., Dieringer, C., Landis, D., Reed, C., et al. (2016). Coinstac: a privacy enabled model and prototype for leveraging and processing decentralized brain imaging data. *Front. Neurosci.* 10:365. doi: 10.3389/fnins.2016.00365

Poldrack, R. A., Barch, D. M., Mitchell, J., Wager, T., Wagner, A. D., Devlin, J. T., et al. (2013). Toward open sharing of task-based fmri data: the openfmri project. *Front. Neuroinform.* 7:12. doi: 10.3389/fninf.2013.00012

Roshchupkin, G. V., Adams, H., Vernooij, M. W., Hofman, A., Van Duijn, C., Ikram, M. A., et al. (2016). Hase: framework for efficient high-dimensional association analyses. *Sci. Rep.* 6:36076. doi: 10.1038/srep36076

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature* 323:533. doi: 10.1038/323533a0

Saha, D. K., Calhoun, V. D., Panta, S. R., and Plis, S. M. (2017). "See without looking: joint visualization of sensitive multi-site datasets," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI'2017)* (Melbourne, VIC), 2672–2678.

Sakoglu, Ü., Pearlson, G. D., Kiehl, K. A., Wang, Y. M., Michael, A. M., and Calhoun, V. D. (2010). A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magn. Reson. Mater. Phys. Biol. Med.* 23, 351–366. doi: 10.1007/s10334-010-0197-8

Scott, A., Courtney, W., Wood, D., de la Garza, R., Lane, S., Wang, R., et al. (2011). Coins: an innovative informatics and neuroimaging tool suite built for large heterogeneous datasets. *Front. Neuroinform.* 5:33. doi: 10.3389/fninf.2011.00033

Shringarpure, S. S., and Bustamante, C. D. (2015). Privacy risks from genomic data-sharing beacons. *Am. J. Hum. Genet.* 97, 631–646. doi: 10.1016/j.ajhg.2015.09.010

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13040–13045. doi: 10.1073/pnas.0905267106

Sweeney, L. (2002). k-anonymity: a model for protecting privacy. *Int. J. Uncert. Fuzziness Knowl. Based Syst.* 10, 557–570. doi: 10.1142/S0218488502001648

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE* 6:e21101. doi: 10.1371/journal.pone.0021101

Thompson, P. M., Andreassen, O. A., Arias-Vasquez, A., Bearden, C. E., Boedhoe, P. S., Brouwer, R. M., et al. (2017). Enigma and the individual: predicting factors that affect the brain in 35 countries worldwide. *Neuroimage* 145, 389–408. doi: 10.1016/j.neuroimage.2015.11.057

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., et al. (2014). The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* 8, 153–182. doi: 10.1007/s11682-013-9269-5

Turner, J. A., Damaraju, E., Van Erp, T. G., Mathalon, D. H., Ford, J. M., Voyvodic, J., et al. (2013). A multi-site resting state fmri study on the amplitude of low frequency fluctuations in schizophrenia. *Front. Neurosci.* 7:137. doi: 10.3389/fnins.2013.00137

van Erp, T. G., Hibar, D. P., Rasmussen, J. M., Glahn, D. C., Pearlson, G. D., Andreassen, O. A., et al. (2016). Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the enigma consortium. *Mol. Psychiatry* 21:547. doi: 10.1038/mp.2015.63

Wojtalewicz, N. P., Silva, R. F., Calhoun, V. D., Sarwate, A. D., and Plis, S. M. (2017). "Decentralized independent vector analysis," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (New Orleans, LA: IEEE), 826–830.

Yuan, K., Ling, Q., and Yin, W. (2016). On the convergence of decentralized gradient descent. *SIAM J. Optim.* 26, 1835–1854. doi: 10.1137/130943170

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. arXiv[preprint] arXiv:1212.5701.

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.

Multi-Template Mesiotemporal Lobe Segmentation: Effects of Surface and Volume Feature Modeling

Hosung Kim[1,2], Benoit Caldairou[1], Andrea Bernasconi[1] and Neda Bernasconi[1]*

[1] Neuroimaging of Epilepsy Laboratory, McConnell Brain Imaging Center, Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, Canada, [2] Laboratory of Neuro Imaging, Department of Neurology, Stevens Neuroimaging and Informatics Institute, University of Southern California, Los Angeles, CA, United States

Numerous neurological disorders are associated with atrophy of mesiotemporal lobe structures, including the hippocampus (HP), amygdala (AM), and entorhinal cortex (EC). Accurate segmentation of these structures is, therefore, necessary for understanding the disease process and patient management. Recent multiple-template segmentation algorithms have shown excellent performance in HP segmentation. Purely surface-based methods precisely describe structural boundary but their performance likely depends on a large template library, as segmentation suffers when the boundaries of template and individual MRI are not well aligned while volume-based methods are less dependent. So far only few algorithms attempted segmentation of entire mesiotemporal structures including the parahippocampus. We compared performance of surface- and volume-based approaches in segmenting the three mesiotemporal structures and assess the effects of different environments (i.e., size of templates, under pathology). We also proposed an algorithm that combined surface- with volume-derived similarity measures for optimal template selection. To further improve the method, we introduced two new modules: (1) a non-linear registration that is driven by volume-based intensities and features sampled on deformable template surfaces; (2) a shape averaging based on regional weighting using multi-scale global-to-local icosahedron sampling. Compared to manual segmentations, our approach, namely *HybridMulti* showed high accuracy in 40 healthy controls (mean Dice index for HP/AM/EC = 89.7/89.3/82.9%) and 135 patients with temporal lobe epilepsy (88.7/89.0/82.6%). This accuracy was comparable across two different datasets of 1.5T and 3T MRI. It resulted in the best performance among tested multi-template methods that were either based on volume or surface data alone in terms of accuracy and sensitivity to detect atrophy related to epilepsy. Moreover, unlike purely surface-based multi-template segmentation, *HybridMulti* could maintain accurate performance even with a 50% template library size.

Keywords: label fusion, multiatlas segmentation, surface feature modeling, medial temporal lobe (MTL), epilepsy, temporal Lobe

INTRODUCTION

Mesiotemporal lobe (MTL) structures, such as the hippocampus (HP), amygdala (AM), and entorhinal cortex (EC), undergo marked morphological changes in numerous neurological and neuropsychiatric conditions (Wang et al., 2010; Cavedo et al., 2011; Bernhardt et al., 2013; Shi et al., 2013; Joo et al., 2014; Maccotta et al., 2015; Arnone et al., 2016). MRI volumetry has been the most commonly employed technique to assess MTL pathology *in vivo* (Goncharova et al., 2001; Bernasconi et al., 2003). In temporal lobe epilepsy (TLE), the most common surgically-amenable epilepsy in adults, manual MRI volumetry allows defining the side of mesiotemporal atrophy in up to 70–90% of patients (Schramm and Clusmann, 2008), and thereby help identifying the surgical target.

Manual MTL volumetry is a labor-intensive task with high demands on neuroanatomical expertise. Although existing automatic segmentation algorithms produce excellent segmentation results for HP and AM in healthy controls (Collins and Pruessner, 2010), their performance in TLE is challenged by the combined effects of atrophy and positional abnormalities (Kim et al., 2012a). Only a relatively small number of studies have attempted segmentation of the entire MTL regions including parahippocampal gyrus (PHG) (Heckemann et al., 2006; Keihaninejad et al., 2012). A study (Hu et al., 2014) specifically segmented the EC, a PHG subregion considered a core epileptogenic zone in TLE (Bernasconi et al., 2003) with suboptimal accuracy (Dice index=73%), likely due to challenges imposed by its complex and variable shape.

Volume-based multi-template and label fusion approaches have been designed to account for shape complexity and anatomical variability by selecting a subset of templates from a large library that best describes the target structure (Collins and Pruessner, 2010; Khan et al., 2011). More recently, our previously proposed surface-based SurfMulti method automatically segmented HP using vertex-wise texture and shape sampling (Kim et al., 2012b), demonstrating improved performances compared to purely volumetric techniques (Collins and Pruessner, 2010). However, performance of purely surface-based approaches likely depends on the availability of a large library, as it may be negatively impacted when the boundaries of the template and individual MRI are not well aligned. The label fusion in volume-based approaches has become sophisticated using local weighted averaging (Artaechevarria et al., 2009; Coupé et al., 2011; Eskildsen et al., 2012; Wang et al., 2013; Awate and Whitaker, 2014). These approaches have demonstrated the improvement of segmentation.

MICCAI Grand Challenge on Multiatlas Labeling (Landman and Warfield, 2012) systemically evaluated various multi-template approaches for the segmentation of numerous brain structures but the parahippocampal gyrus. A total of 25 algorithms that were trained by 15 atlases were tested on 20 images. The performance for the hippocampus and the amygdala ranged 82–87 and 75–83% in mean Dice similarity index, respectively. Among the methods that were evaluated, the ones that displayed higher accuracy were the joint label fusion technique that used a joint probability of selected atlases to correct for the bias due to the inclusion of similar atlases in the template library or the training-set (Wang et al., 2013) and the Non-Local STAPLE algorithm that combined Staple method with the non-local means estimator (Asman and Landman, 2013).

The current work aimed at segmenting simultaneously HP, AM, and EC using a large template library ($n = 175$) which included shape and volume variants in relation to TLE ($n = 135$). We tested well-established volume-based and surface-based approaches as well as looked for a possibility of the combined approach. The proposed algorithm, *HybridMulti*, combined surface-based with volume-based similarity measures for optimal template selection. The SurfMulti was based on the linear alignment between the template and individual MRI. Volume-based approaches (Asman and Landman, 2013; Wang et al., 2013) rely also on the accuracy of the linear and non-linear registration. To improve alignment, we introduced a non-linear registration step that incorporates a novel hybrid cost function based on surface and volume. Our algorithm furthermore included a new multi-level feature weighting for shape averaging. We compared MTL segmentation of *HybridMulti* to our previous SurfMulti (Kim et al., 2012b) and two volume-based approaches with/without local weighted averaging (Collins and Pruessner, 2010; Wang et al., 2013); evaluations also took into account the influence of template library size on segmentation performance.

METHODS

HybridMulti includes a "template library construction" where the algorithm learns image features using a training-set and an "automatic segmentation" step where the algorithm segments MTL structures for an individual test MRI (**Figure 1**). Training set consists of MR images and manual labels of controls and patients (**Figure 1A**). Labels are converted into surface meshes using spherical harmonics and point distribution model (SPHARM-PDM) that ensure shape-inherent point-wise correspondences across subjects (Styner et al., 2004, 2006b). Each surface is mapped onto its corresponding MRI. In the beginning of the segmentation step, the pair of each template image and its MTL surface are mapped on the test image. As the test image does not have its own surface, the surface features extracted on the test image are from the surface of each template. By comparing the features extracted from each template and those from the test image, Surface- with volume-derived similarity measures for optimal template selection are then computed to select an optimal subset n_a (**Figure 1B**-1). Next, a non-linear registration that is driven by volume-based intensities and features sampled on evolving template surfaces is performed to improve alignment between each template in the subset n_a and the individual MRI (**Figure 1B**-2). The motivation of using this hybrid registration was to improve the boundary fitting by weighting the features extracted using deformable surfaces as well as to use a consistent similarity measurement in all the steps. After choosing a smaller subset n_b, templates are then averaged using adaptive weighting combined with local averaging, which creates the final segmentation (**Figure 1B**-3,4). The test image's features are updated during the series of the
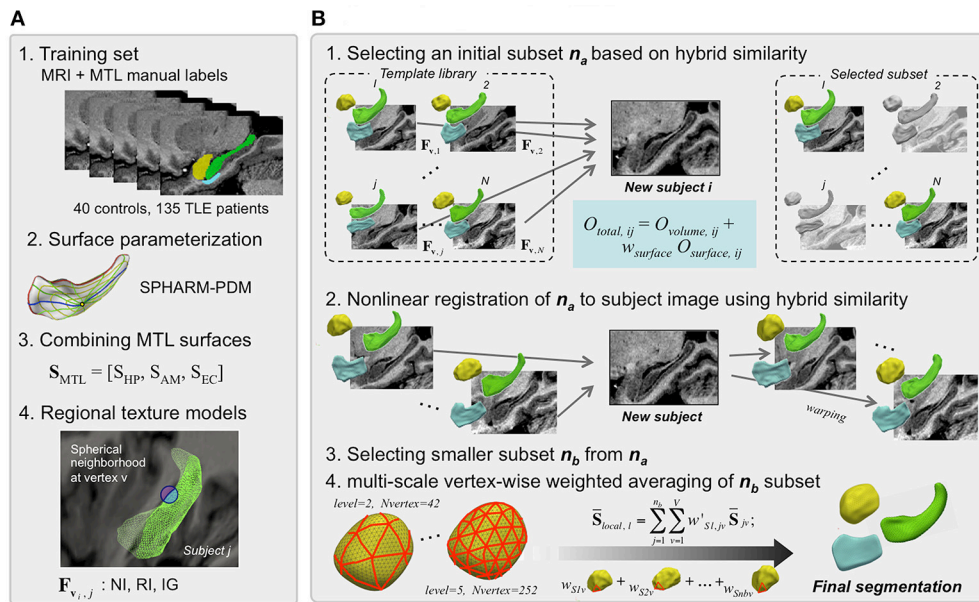
FIGURE 1 | HybridMulti automatic hippocampal segmentation steps. Flowchart of the proposed algorithm (in **A**, steps 2 and 4 are illustrated only for the HP). The segmentation procedure consists mainly of two: template library construction and automated segmentation of mesiotemporal structures. **(A)** Template library construction. **(B)** Automatic segmentation of MTL structures.

steps including template selection, non-linear registration and weighted averaging as the image and the surface deform. In this manner, the similarity of the deformable surface and the target MTL border is expected to increase and the surface gets a similar shape to the true MTL boundary.

Template Library Construction (Figure 1A)

Prior to the subsequent procedures, all MR images in the training-set and the test-set are spatially normalized by registering them into MNI ICBM 152 space. We create a template library that aggregates surface-based regional texture models of HP, AM, and EC as a joint representation of the three MTL structures.

Manually delineated labels of each MTL structure [linearly registered to MNI ICBM-152 space (Collins et al., 1994)] are converted into surface meshes and parameterized using the spherical harmonics and uniform icosahedron-subdivision model (SPHARM-PDM) that guarantees shape-inherent vertex-wise correspondence across subjects (Styner et al., 2006a). MTL surfaces are treated as one concatenated surface, $\mathbf{S}_{MTL} = [S_{HP}, S_{AM}, S_{EC}]$.

Each surface $\mathbf{S}_{MTL}$ is mapped to its corresponding MRI. At a given surface vertex $\boldsymbol{v}$, we define three spherical neighborhoods of 3, 5, and 7 mm radius. These spheres are subdivided into an inner region (IR) and outer region (OR) with respect to the surface boundary, where we compute the following texture features (Kim et al., 2012b): *i) Normalized intensity (NI)*: the ratio between mean intensity and intensity standard deviation for each of IR/OR to capture regional tissue homogeneity. We defined $NI_{IR, i} = \mu_{IR, i} / SD_{IR, i}$ and $NI_{OR, i} = \mu_{OR, j} / SD_{OR, i}$; *ii) Relative*

intensity (RI): the ratio of mean intensity between IR and OR to assess the contrast between IR and OR voxels. RI was defined as $RI_i = 2 \times (\mu_{OR, i} - \mu_{IR, i}) / (\mu_{OR, i} + \mu_{IR, i})$; *iii) Intensity gradient (IG)*: the 1st derivative of intensity along x-, y-, and z-directions to capture edge information was summarized into the magnitude as $IG = \sqrt{g_x^2 + g_y^2 + g_z^2} = \sqrt{\frac{\partial I}{\partial x} + \frac{\partial I}{\partial y} + \frac{\partial I}{\partial z}}$. $[x\ y\ z]$ is a voxel location and I is an image.

These texture features comprises a set of "true" feature vectors (3 normalized intensity + 3 relatively intensity + 3 gradients = 9 features), $\mathbf{F}_{v,j}$ extracted at v-th vertex on the jth (1 … j … N) surface template. Previously we demonstrated that each feature almost equally contributed to the segmentation accuracy and observed the optimal result using all the features. Notably, we did not use the shape features proposed in our previous surface-based framework (Kim et al., 2012b), which was used to constrain the shape deformation in the Automatic segmentation step. The deformation in the current study is instead governed directly by a volume-based non-linear registration (see section Boundary-Weighted Non-linear Registration of Template Subset to Test MRI).

Automatic Segmentation (Figure 1B)
Initial Template Subset Selection

From the template library, we first select a subset of candidates that are most similar to the test image. To that end, we compute the hybrid similarity O_{total} that combined surface-based ($O_{surface}$) and volume-based (O_{volume}) similarity term between each template j and the test MRI i using:

$$O_{total,ij} = O_{volume,ij} + w_{surface}O_{surface,ij} \tag{1}$$

$w_{surface}$ is a weighting constant. The surface-based similarity $O_{surface}$ is defined as:

$$O_{surface,ij} = -\sum_{\mathbf{v}} \frac{\left\| \mathbf{F}_{\mathbf{v},j} - \hat{\mathbf{F}}_{\mathbf{v},ij} \right\|}{\sqrt{\frac{1}{N} \sum_{k=1}^{N} \left(\mathbf{F}_{\mathbf{v},k} - \overline{\mathbf{F}}_{\mathbf{v}} \right)^2}}, \quad \overline{\mathbf{F}}_{\mathbf{v}} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{F}_{\mathbf{v},k} \quad (2)$$

$O_{surface}$ is calculated across all surface vertices $\mathbf{v}$. It represents a normalized similarity between *true* features extracted from the jth $(1 \ldots j \ldots N)$ template $(\mathbf{F}_{\mathbf{v},j})$ and *estimated* features extracted from the test MRI i $(\hat{\mathbf{F}}_{\mathbf{v},i})$. O_{volume} can be any similarity function including the cross-correlation or the normalized mutual information (NMI) that quantifies statistical intensity distribution dependency of two images A and B (Studholme et al., 1999). The computation of cross-correlation is generally faster while the NMI is more robust in similarity of multi-modal images compared to each other. For computational efficiency, we compute O_{volume} within a mask defined by dilating the current template label three times. The number of selected templates (n_a) was empirically determined to maximize O_{total} (see section Parameter Selection).

Boundary-Weighted Non-linear Registration of Template Subset to Test MRI

Each template MRI is non-linearly registered to the test MRI to increase shape similarity. To estimate the deformation field from a template $\mathbf{T}$ to the test MRI $\mathbf{I}$, a "conventional" non-linear registration iteratively matches intensity features by maximizing a volume-based similarity function $O_{vol, reg}$. Accordingly, the deformation field d is estimated as:

$$\vec{d} = \arg\max_{\vec{d}} O_{vo, reg}(\mathbf{T} + \vec{d}, \mathbf{I}) + O_{smooth} \quad (3)$$

O_{smooth} is a smoothness term to constrain the estimated deformation. We employed a type of freeform deformation models defined in Collins et al. (1995). To improve the registration accuracy, we increase the weight of voxels on and nearby the target boundary by incorporating a similarity measure derived from the template surface evolving during the registration with the original volume similarity. Let $\mathbf{S}_{MTL, T}$ be the *true* template surface on the original MRI and $\mathbf{S}_{MTL, S}$ an *estimated* template mapped onto the test MRI. We define $\mathbf{S}_{MTL, S}$ by deforming $\mathbf{S}_{MTL, T}$ using the deformation field estimated at the current iteration. A surface-based feature similarity measure between $\mathbf{S}_{MTL, T}$ and $\mathbf{S}_{MTL, S}$ is defined as:

$$O_{surf, reg} = -\frac{\sum_{\mathbf{v}} (\mathbf{F}_{\mathbf{v}, T} - \overline{\mathbf{F}}_{\mathbf{v}, T})(\mathbf{F}_{\mathbf{v}, \hat{s}} - \overline{\mathbf{F}}_{\mathbf{v}, \hat{s}})}{\sqrt{\sum_{\mathbf{v}} (\mathbf{F}_{\mathbf{v}, T} - \overline{\mathbf{F}}_{\mathbf{v}, T})^2} \sqrt{\sum_{\mathbf{v}} (\mathbf{F}_{\mathbf{v}, \hat{s}} - \overline{\mathbf{F}}_{\mathbf{v}, \hat{s}})^2}},$$
$$\mathbf{F}_{\mathbf{v}} = (\mu_{\mathbf{v}, OR} - \mu_{\mathbf{v}, IR})/(\mu_{\mathbf{v}, OR} + \mu_{\mathbf{v}, IR}) \quad (4)$$

where $\mathbf{v}$ is a vertex on surfaces $\mathbf{S}$; $\mathbf{F}_{\mathbf{v}}$ is the relative intensity defined in **2.1**. Therefore, $O_{surf, reg}$ is a correlation coefficient between feature $\mathbf{F}_{\mathbf{v}, T}$ extracted on $\mathbf{S}_{MTL, T}$ and feature $\mathbf{F}_{\mathbf{v}, \hat{s}}$

extracted on $\mathbf{S}_{MTL, \hat{s}}$. To estimate the deformation field, we redefine the Equation (3) as:

$$\vec{d} = \arg\max_{\vec{d}} O_{hybrid, reg}(\mathbf{T} + \vec{d}, \mathbf{I}) + O_{smooth},$$
$$O_{hybrid, reg} = O_{vol, reg} + w_{surf, reg} O_{surf, reg} \quad (5)$$

$O_{vol, reg}$ is the correlation coefficient over a volume of interest (here, a geometric union of all MTL template labels in the library, subsequently dilated 5 times for more extensive spatial coverage) as in Collins and Pruessner (2010). A larger weight $w_{surf, reg}$ moves $\mathbf{S}_{MTL, S}$ more rapidly to areas presenting with feature characteristics similar to those on the surface of the template image. Finally, Equation (5) is optimized using a derivative-free 3D Nelder-Mead Simplex approach (Lagarias et al., 1998) as also known as the simplex method, is a commonly applied approach. This method is applied to non-linear optimization problems for which derivatives may not be known and is robust against the local minima problem. This function has been used as the standard optimization method in the non-linear registration algorithm (Collins et al., 1995) we adopted in the current paper.

Subset Restriction and Global Weighed Averaging

The non-linear registration in the previous section (Boundary-weighted Non-linear Registration of Template Subset to Test MRI) is applied to decrease shape variability and to increase similarity between the template-subset and test image. From the initially selected n_a template-subset $(n_a < N)$, we choose an even smaller subset of the n_b most similar templates $(n_b < n_a < N)$ based on Equation (1), increasing computational efficiency in subsequent steps. We determine n_b empirically, which will be evaluated in the section Parameter Optimization.

Optimal global weights for these n_b templates are calculated using the similarity function Equation (2) as in Kim et al. (2012b). Let $\mathbf{w}_S$ and $\mathbf{w}_F$ be $n_b \times 1$ weight vectors for optimal surfaces and features. We then define $\overline{\mathbf{S}}$ as the average surface of the n_b template-subset as:

$$\overline{\mathbf{S}}_{global} = \sum_{j=1}^{n_b} w_{F,j} \mathbf{F}_{v,j}; \quad \mathbf{w}_F = [w_{F,1}, w_{F,2}, \ldots,]; \quad \sum w_{F,j} = 1 \quad (6)$$

Analogously, we define the weighted mean and SD of features at a given vertex $\mathbf{v}_i$ by:

$$\overline{\mathbf{F}}_{\mathbf{v}} = \sum_{j=1}^{n_b} w_{F,j} \mathbf{F}_{v,j}; \quad \mathbf{w}_F = [w_{F,1}, w_{F,2}, \ldots, w_{F,n_b}]; \quad \sum w_{F,j} = 1; \quad (7)$$

$$\sigma_{F,v} = \sqrt{\sum_{j=1}^{n_b} w_{F,j} \left(\mathbf{F}_{v,j} - \overline{\mathbf{F}}_{\mathbf{v}} \right)^2} \quad (8)$$

Similarity from Equation (2) can be formulated for the template-subset n_b:

$$O_{subset} = -\sum_{v} \frac{\left\| \overline{\mathbf{F}}_{v} - \hat{\mathbf{F}}_{v, \bar{s}} \right\|}{\sigma_{F,v}} \quad (9)$$

$\hat{\mathbf{F}}_{v,\ \bar{s}}$ is the *estimated* feature-set computed on the averaged surface $\bar{\mathbf{S}}$ mapped on the test image. In the above formulas, weights are determined by maximizing the similarity between the n_b template-subset and test image.

$$\mathbf{w} = [\mathbf{w}_s \ \ \mathbf{w}_F] = \arg_w \max O_{subset} \tag{10}$$

We initialized all components of $\mathbf{w}_S$ and $\mathbf{w}_F$ to 1/n. The cost function O_{subset} is optimized using the multivariate derivative-free Nelder-Mead approach (Lagarias et al., 1998).

Multi-Level Local Weighted Averaging

To incorporate a local weighting to Equations (5–9), the resulting surface $\bar{\mathbf{S}}$ in Equation (10) is resampled through icosahedron-subdivision (Styner et al., 2006b), first at the coarsest level $l = l_0$. We determine weights at each sampling vertex, and interpolate these weights to vertices at the next finer level l_1. Let $\mathbf{w}_S\ l$ be a $n_b \ \square \ m$ weight matrix: m is the number of vertices at level l. We compute $\mathbf{w'}_S\ l$, (a $n_b \ \square \ V$ vector) by interpolating $\mathbf{w}_{S,\ l}$ to all vertices v [1, 2, ...,V] of the original surface [[Inline Image]] ($V > m$). For interpolation, we use the Fast Spherical Linear Interpolation (Shoemake, 1985). We define the locally weighted average surface as:

$$\bar{S}_{local,l} = \sum_{j=1}^{nb} \sum_{v=1}^{v} w'_{sl,jv} \, \bar{S}_{jv}; \ \sum_{j=1}^{nb} \sum_{v=1}^{v} w'_{sl,jv} = V \tag{11}$$

The similarity function at the level l was defined as:

$$O_{subset,l} = -\sum_i \frac{\left\| \bar{\mathbf{F}}_{\mathbf{v_i}} - \hat{\mathbf{F}}_{\mathbf{v_i}, \bar{S}_{local}} \right\|}{\sigma_F}; \ \ \mathbf{w}_S\ l = \arg\ max_w\ O_{subset} \tag{12}$$

To achieve the final segmentation of all three MTL structures, we optimized $\mathbf{w}_S\ l$ using the Nelder-Mead method while increasing subdivision level $l=[l \mid l_0,\ l_1,...,\ l_{max}]$. The algorithm stops when **Equation (11)** stops increasing or l reaches preset l_{max} to prevent from an extensive computation. The proposed multi-level approach using different subdivisions is mainly for coarse-to-fine spatial fitting and the use of this strategy avoids the introduction of a constraint term preventing from local minima while the surface shape gets finer. In the current study, we set the coarsest level ($l_0 = 2$) where 42 equally distributed vertices are sampled; the finest level l_{max} is determined empirically (See section MRI Acquisition).

EXPERIMENTS AND RESULTS

Experiments

Subjects

Our training-set included 40 healthy controls (18 men; mean $\pm$ SD age = 33 $\pm$ 12 years) and 135 drug-resistant TLE patients (61 men; mean $\pm$ SD age = 37 $\pm$ 11 years). TLE diagnosis and lateralization of the side of the seizure focus into left TLE (n =

65) and right TLE ($n = 70$) were determined by a comprehensive evaluation including video-EEG recordings and MRI. The Ethics Committee of the Montreal Neurological Institute and Hospital approved the study and written informed consent was obtained from all participants.

MRI Acquisition

MR images were acquired on a 1.5 Tesla Phillips Gyroscan using a T1-weighted FFE sequence (TR = 18 ms; TE = 10 ms; NEX = 1; flip angle = 30°; matrix size = 256 $\square$ 256; FOV = 256 mm; slice thickness = 1 mm), yielding 1 mm-isotropic voxels. Images underwent intensity non-uniformity correction (Sled et al., 1998). Intensities were normalized and images were linearly registered to the MNI ICBM-152 template (Collins et al., 1994). MTL structures were manually segmented by an expert using the protocol described in Bernasconi et al. (2003). Based on z-score normalization with respect to volumes in controls, 81 (60%) patients showed hippocampal atrophy (i.e., a z-score below −2) ipsilateral to the seizure focus.

We also acquired 3T T1-weighted images on Siemens Trio Tim scanner using a 32-channel phased-array head coil. T1-weighted images were acquired using 3DMPRAGE with 1 mm isotropic voxels (TR = 3,000 ms, TE = 4.32 ms, TI = 1,500 ms, flip angle = 7°, matrix size = 336×384, FOV = 201 × 229 mm). This data was used to evaluate whether the algorithm consistently selected the same or similar parameter values for different dataset. The 3T dataset included 39 healthy controls and 84 drug-resistant TLE patients who were further classified into left TLE (n = 38) and right TLE ($n = 46$).
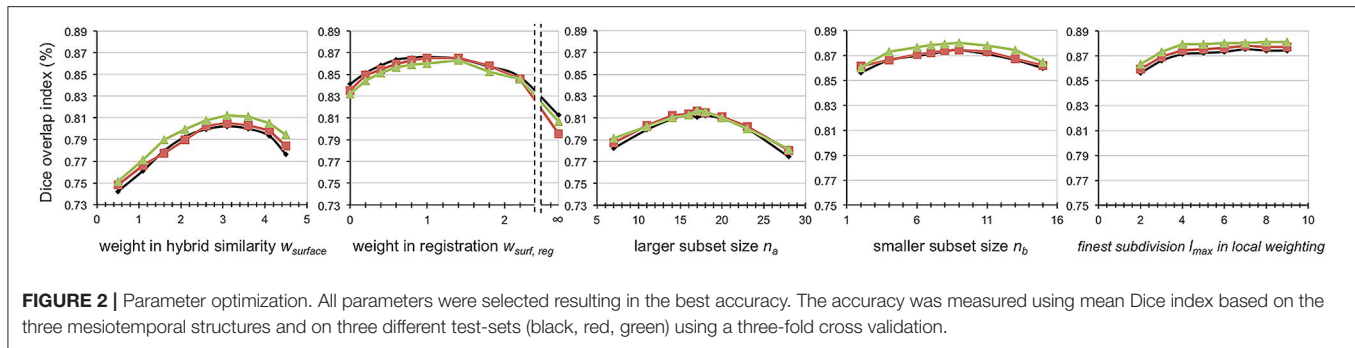
Evaluation Metrics

To quantify the accuracy of automated segmentations, we computed the Dice similarity index:$D = 2xv(M \cap A)/(v(M) + v(A))$, where M/A are the voxels comprising manual/automated labels; "M n A" are voxels in the intersection of M and A; v ($\cdot$) is the volume operator.

Parameter Optimization

Based on maximal Dice overlap index between automated and manual labels, the following parameters were chosen empirically: weight of surface-based similarity $w_{surface}$ to select the optimal subset as in Equation (1); weight of surface-based similarity $w_{surf,\ reg}$ used in non-linear registration; size of initial template-subset n_a; size of final template-subset n_b; and finest subdivision l_{max} in local weighting. We validated *HybridMulti* using a three-fold cross-validation where we subdivided our data into 3 sets with an almost equal sized sample ($n = 58,58,59$) and merged two sets among them to create a training-set and used the remaining set as a test-set while we balanced the proportion of controls ($\sim$25%) and patients ($\sim$75%) per set. The optimal parameters that resulted in most accurate segmentation were selected for each training-set. We segmented the test-set based on their corresponding training-set and the parameters. We repeated this process three times while all the three sets were tested.

Performance at Each Segmentation Stage

Segmentation accuracy was evaluated at the following stages: *i)* initial n_a template-subset selection; *ii)* non-linear registration; *iii)*

FIGURE 2 | Parameter optimization. All parameters were selected resulting in the best accuracy. The accuracy was measured using mean Dice index based on the three mesiotemporal structures and on three different test-sets (black, red, green) using a three-fold cross validation.

final n_b template-subset selection; *iv)* global and local weighted averaging. We compared accuracy at each stage to that of the previous stage using paired *t*-tests.

Comparison With State-of-the-Art Multi-Template Approaches

We compared Dice indices between *HybridMulti*, and SurfMulti (Kim et al., 2012b), or a volume-based multitemplate approach (VolMulti) based on non-weighted averaging (Collins and Pruessner, 2010) or a volume-based approach (JointFusion) based on local-weighted averaging (Wang et al., 2013) in controls and each patient group using Student's *t*-tests. The parameters for each algorithm were selected empirically (VolMulti: size of subset = 15; JointFusion: search area r_s = 3 x 3 x 3, patch size r_p = 3 x 3 x 3, β = 2) which resulted in the best accuracy using a leave-one-out approach.

Detection of Mesiotemporal Atrophy Related to the Epileptic Focus

We assessed the ability of each automatic algorithm to detect each structure's atrophy in TLE groups relative to controls by computing Cohen's d ([mean volume controls—mean volume TLE] / pooled SD) that measures the effect size of a between-group difference, and calculated the significance of the observed effect using *t*-tests.

Impact of Template Library Size on Segmentation Accuracy

Keeping proportions of controls and patients constant, we randomly selected 40 subjects as a test-set. We then created the template library by selecting randomly from the rest of data, with various sizes: n = 88 (1/2), n = 58 (1/3), n = 44 (1/4), and n = 35 (1/5) of its original size. We repeated this process 20 times to avoid a possible bias. We evaluated automated segmentation accuracy at these smaller template library sizes.

Significances of all statistical tests were adjusted for multiple comparisons using Bonferroni-correction.

Results
Parameter Selection
The parameters resulting in the best segmentation accuracy were selected at very similar values between the 3 test-sets when using a three-fold cross validation. The proposed HybridMulti achieved maximal accuracy with the following parameters: $w_{surface}$ = 3.1,

$w_{surf,reg}$ = 1.1, n_a = 17, and n_b = 8 (average between the 3 test-sets; **Figure 2**). Use of the cross-correlation or NMI as the similarity function did not make a difference in segmentation accuracy. We thus used the cross-correlation as it was faster to compute. We also found that the local weighting using the finest subdivision l_{max} larger than 5 (producing 252 sampling vertices) maintained the segmentation accuracy without a further improvement. Thus, we chose l_{max} = 5 as a larger l_{max} increased the computational time. JointFusion yielded best results with the following parameters: beta = 0.5; rp = 3; rs = 3. SurfMulti used n = 10 for the optimal subset whereas VolMulti used n = 14. All the algorithms were tested on a same computing environment (Linux workstation, 1 CPU, 2.30 Ghz, 8 GB RAM). Average computation times per individual hemisphere were 20 or 25 min for HybridMulti (O_{volume} = cross-correlation or NMI, respectively; step-wise: initial subset selection: 1 min; non-linear registration: 10 [cross-correlation] or 15 [NMI] min; smaller subset selection: 0.5 min; global weighting: 3 min; Local weighting: 5.5 min); 15 min VolMulti; 15 min JointFusion; 13 min SurfMulti.

When performing the same evaluation on 3T dataset, we found the parameters yielding the maximal accuracy were selected at very similar values: $w_{surface}$ = 3.2, $w_{surf,reg}$ = 1.2, n_a = 17, n_b = 8, and l_{max} = 6.

Segmentation Accuracy in Different Steps
Accuracy of HybridMulti was improved gradually from the initial selection step and the highest accuracy was achieved at the final local weighted averaging (**Figure 3**).

Highest improvement was found at the boundary-weighted non-linear registration step for all structures (mean Dice = +4.8%, *p* < 0.0001). Moreover, the proposed non-linear registration that included a surface-term outperformed the original volume-based registration (Collins et al., 1995) (+2.5%, *p* < 0.001). Inclusion of local weighted averaging also significantly improved segmentation of EC (0.7%) and (HP: 0.3%) compared to the global weighting (*p* < 0.05).

Performance Comparison Between Algorithms
For all MTL structures, HybridMulti consistently outperformed SurfMulti and VolMulti in patients and controls (*p* < 0.001, **Table 1**), which was equally significant for 1.5T and 3T data (**Table 2**). HybridMulti also showed a superior accuracy in TLE patients compared to JointFusion as higher Dice indices
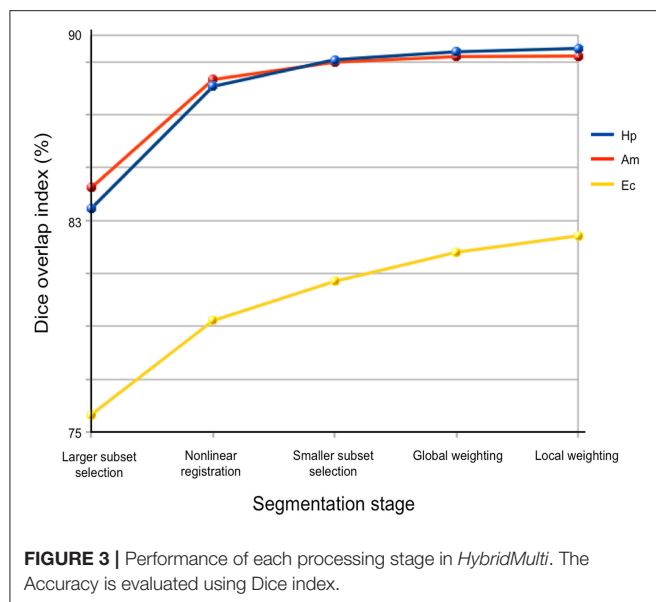
FIGURE 3 | Performance of each processing stage in *HybridMulti*. The Accuracy is evaluated using Dice index.

TABLE 1 | Segmentation accuracy using a three-fold cross validation (% mean ± SD of Dice similarity index).

	VolMulti	SurfMulti	JointFusion	*HybridMulti*
CONTROLS				
HP	84.4 ± 4.1*	86.8 ± 2.7*	89.5 ± 1.5	89.7 ± 2.1
AM	83.8 ± 4.3*	86.3 ± 3.6*	88.4 ± 2.8	89.3 ± 2.5
EC	75.8 ± 6.8*	79.3 ± 4.6*	78.4 ± 4.3*	82.9 ± 3.7
IPSILATERAL				
HP	80.3 ± 5.4*	86.1 ± 3.4*	87.3 ± 3.8*	88.7 ± 2.5
AM	83.2 ± 4.2*	85.2 ± 3.9*	88.0 ± 2.9	89.0 ± 2.6
EC	75.2 ± 8.1*	77.7 ± 5.2*	78.0 ± 6.5*	82.6 ± 3.8
CONTRALATERAL				
HP	84.0 ± 4.4*	86.9 ± 3.1*	88.8 ± 3.0	89.4 ± 2.3
AM	83.8 ± 4.2*	85.8 ± 3.7*	88.3 ± 2.8*	89.2 ± 2.7
EC	76.2 ± 7.4*	78.8 ± 5.4*	78.6 ± 4.8*	82.7 ± 4.1

*Ipsilateral/Contralateral refers to the epileptogenic hemisphere. Decreased performance relative to HybridMulti - *: significant after Bonfferoni correction ($p < 0.05/36 = 0.0013$).*

TABLE 2 | Segmentation accuracy for a smaller set of 3T data (controls: $n = 39$; TLE: $n = 84$) using a three-fold cross validation (% mean ± SD of Dice similarity index).

	VolMulti	SurfMulti	JointFusion	*HybridMulti*
CONTROLS				
HP	85.6 ± 3.6*	87.3 ± 2.5*	89.7 ± 1.4	89.8 ± 1.8
AM	84.3 ± 3.9*	86.4 ± 3.1*	88.5 ± 2.4	89.4 ± 2.3
EC	77.3 ± 6.4*	80.2 ± 4.6*	79.1 ± 4.1*	83.1 ± 3.3
IPSILATERAL				
HP	82.3 ± 5.4*	86.1 ± 2.9*	87.0 ± 3.6*	88.4 ± 2.3
AM	83.2 ± 4.2*	84.9 ± 3.8*	87.6 ± 2.7*	88.9 ± 2.4
EC	76.4 ± 8.2*	78.1 ± 4.9*	78.8 ± 6.4*	82.6 ± 3.5
CONTRALATERAL				
HP	84.2 ± 4.3*	87.7 ± 2.7*	88.8 ± 3.0	89.5 ± 1.9
AM	84.5 ± 4.2*	85.9 ± 3.4*	88.3 ± 2.8*	89.3 ± 2.3
EC	77.2 ± 7.1*	78.7 ± 5.1*	79.7 ± 4.7*	82.4 ± 4.2

*Ipsilateral/Contralateral refers to the epileptogenic hemisphere. Decreased performance relative to HybridMulti - *: significant after Bonfferoni correction ($p < 0.05/36 = 0.0013$).*

were found in HP and EC ipsilaterally and in AM and EC contralaterally ($p < 0.05$). HybridMulti also segmented EC in healthy controls more accurately than JointFusion ($p < 0.001$). This pattern of difference between HybridMulti and JointFusion was similar in 3T data (**Table 2**).

For the 3T data, even using a smaller dataset, we found that all the methods resulted in accuracy comparable to the larger 1.5T dataset, with generally decreased SDs. A separate test that segmented 3T dataset using the 1.5T training-set showed the result where we found overall a slight drop down in the accuracy and a larger SD (Controls: HP = 89.5 ± 2.4; AM = 89.0 ± 2.9; EC = 82.8 ± 4.4; TLE-ipsilateral: HP = 88.5 ± 2.8; AM = 89.1 ± 3.2; EC = 82.5 ± 4.9; TLE-contralateral: HP = 89.2 ± 2.6; AM = 89.1 ± 2.8; EC = 82.5 ± 5.2) compared to when using a smaller-set of the same field strength training data. This

suggests that using a lower field training-set to segment a higher field strength data results in slightly decreased accuracy due to a different tissue-contrast.

Examples for 1.5T are shown in **Figure 4** and those for 3T in Supplementary Figure 1.

Ability of Automated Methods to Detect Atrophy Related to the Epileptic Focus

Group-wise comparisons identified hippocampal atrophy ipsilateral to the seizure focus in TLE patients irrespective of the method, i.e., manual or automated ($p < 0.05$, **Table 3**). The effect sizes of atrophy detected using algorithms were all large (Cohen's d > 0.8). HybridMulti and JointFusion, nevertheless, detected an effect size of atrophy closest to manual volumetry (Cohen's d: manual = 1.67; HybridMulti = 1.57; JointFusion = 1.56).

Manual and *HybridMulti* segmentation also detected a large effect size of ipsilateral EC atrophy, which was significant compared to controls ($t > 3.2$, $p < 0.05$).

Impact of Template Library Size on Segmentation Accuracy

Reducing the template library size from N ($n = 175$) to N/5 ($n = 35$) showed that the accuracy of EC segmentation declined fastest compared to HP and AM, consistently in all algorithms tested. Size of the library had a lower influence on segmentation accuracy of HybridMulti, and volume-based approaches (JointFusion, VolMulti) than SurfMulti. Indeed linear model analysis of an interaction term between "segmentation method" and "size of the library" revealed a faster decline in Dice index for SurfMulti than for the other three methods ($p < 0.001$). HybridMulti and JointFusion, on the other hand, resulted in a similar accuracy when reducing the template library size from N to N/4 across all MTL structures (mean Dice decrease < 1%, $p < 0.1$, **Figure 5**). In HP and EC, reducing the library size from N/4 to N/5 influenced the accuracy
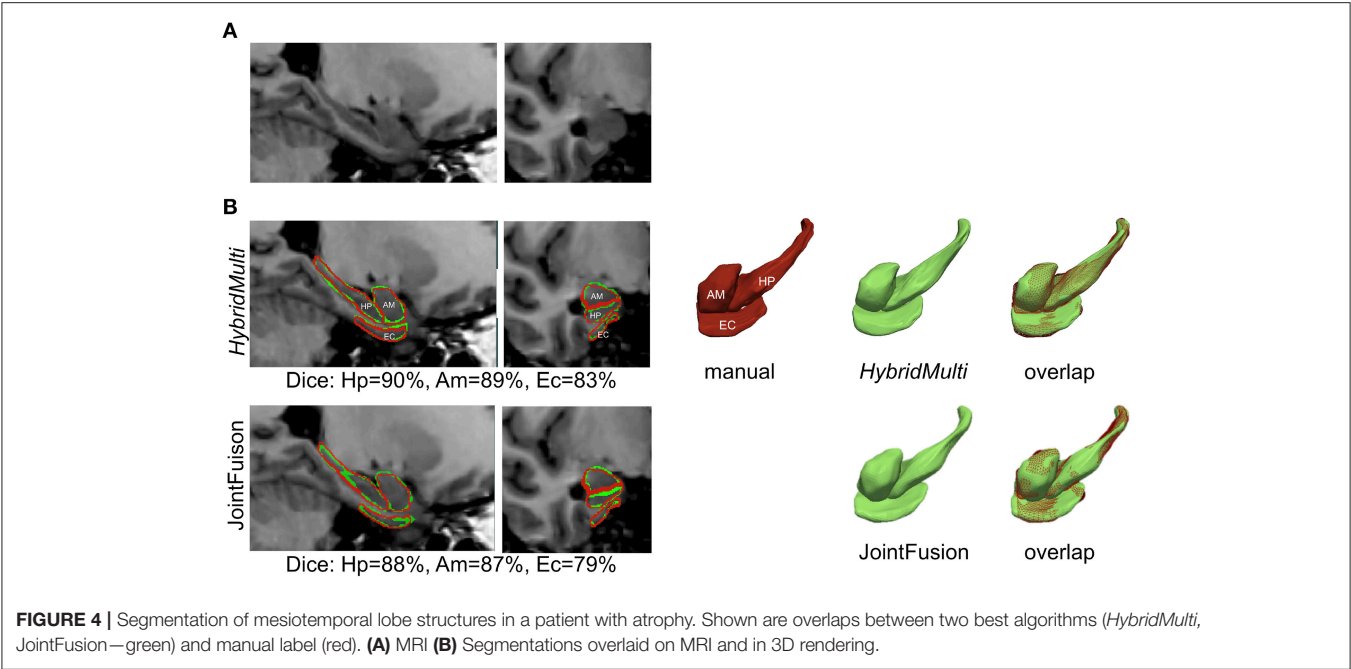
FIGURE 4 | Segmentation of mesiotemporal lobe structures in a patient with atrophy. Shown are overlaps between two best algorithms (*HybridMulti*, JointFusion—green) and manual label (red). **(A)** MRI **(B)** Segmentations overlaid on MRI and in 3D rendering.

TABLE 3 | Group differences between patients and controls.

	Manual	*HybridMulti*	JointFusion	SurfMulti	Volmulti
HP					
Ipsilateral	**−2.29 ± 1.85** (1.67)	**−2.09 ± 1.97** (1.58)	**−2.13 ± 2.37** (1.56)	**−1.69 ± 1.56** (1.39)	**−1.32 ± 1.62** (1.32)
Contralateral	−0.51 ± 1.70 (0.33)	−0.37 ± 1.53 (0.24)	−0.28 ± 1.78 (0.19)	−0.08 ± 1.34 (0.05)	−0.05 ± 1.45 (0.03)
AM					
Ipsilateral	−0.10 ± 1.45 (0.10)	−0.11 ± 1.41 (0.11)	0.39 ± 1.43 (−0.18)	0.05 ± 1.26 (−0.01)	0.35 ± 1.84 (−0.13)
Contralateral	0.20 ± 1.32 (−0.15)	0.27 ± 1.25 (−0.16)	0.45 ± 1.38 (−0.22)	0.15 ± 1.11 (−0.08)	−0.05 ± 1.71 (−0.02)
EC					
Ipsilateral	**−1.49 ± 1.19** (1.11)	**−0.98 ± 0.92** (0.82)	−0.52 ± 0.94 (0.45)	−0.65 ± 1.22 (0.46)	−0.35 ± 2.06 (0.18)
Contralateral	−0.69 ± 1.41 (0.52)	−0.37 ± 0.91 (0.29)	−0.18 ± 1.02 (0.16)	−0.15 ± 1.62 (0.09)	−0.05 ± 1.91 (0.03)

Mesiotemporal volume in mean z-scores ± SD and effect sizes for atrophy shown in brackets (Cohen's d index; 0.2 indicates a small, 0.5 medium, and >0.8 large effect); group-wise significances in volumes (bold) are adjusted for multiple comparisons using Bonferroni correction.

more significantly for HybridMulti than JointFusion ($p < 0.01$). However, the accuracy of HybridMulti was higher than that of JointFusion in all structures (mean Dice difference—HP: 0.3%; AM: 0.1%; E: 1%).

DISCUSSION AND CONCLUSION

We propose *HybridMulti*, an algorithm that combines surface- and volume-based similarity to automatically segment key regions in the mesiotemporal lobe (i.e., HP, AM, and EC). In controls and TLE patients alike, segmentation accuracy was excellent, with Dice indices above 88% for HP and AM and above 82% for EC. In particular, the proposed

method outperformed previous multi-template approaches in pathological MTL structures, as its overlap to manual delineation and its sensitivity to detect atrophy were superior. Reducing template library showed that our method is reliable in even case of a small size of training-set.

Our algorithm was compared to three recently proposed multi-template approaches: volume-based approaches—JointFusion (Wang et al., 2013), VolMulti (Collins and Pruessner, 2010), and a purely surface-based framework—SurfMulti (Kim et al., 2012b). Improved segmentation accuracy of *HybridMulti* relative to these algorithms likely results from modeling both volume- and surface-derived features to select the optimal template subset and to improve the alignment between these templates and the test MRI prior to surface-shape
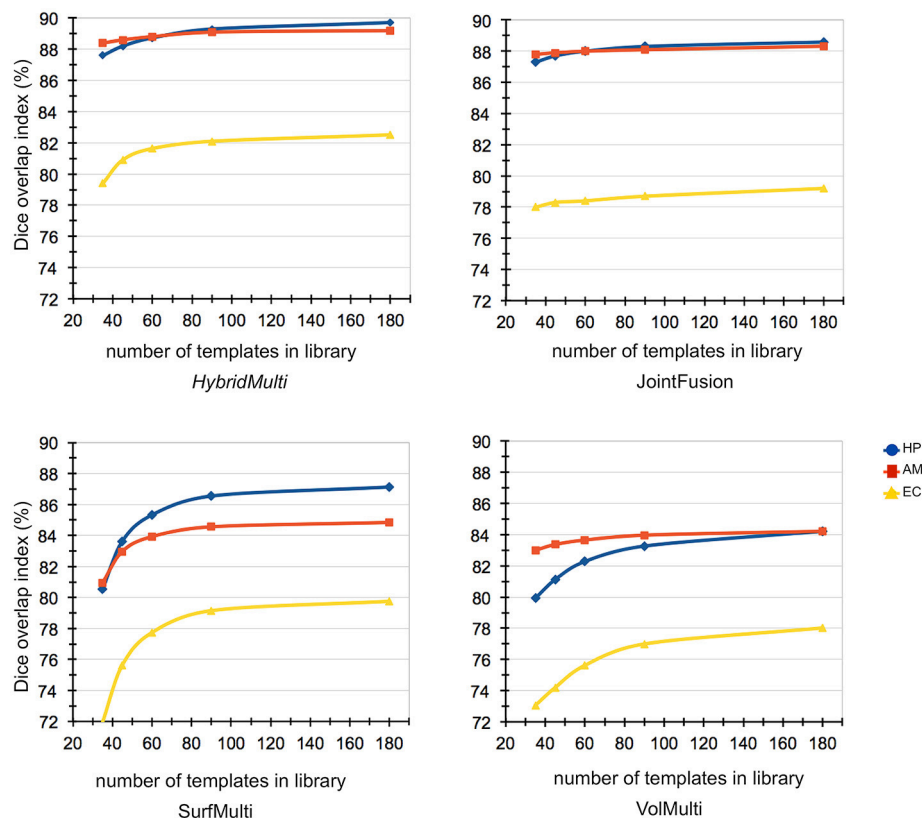
FIGURE 5 | Impact of template library size on automated segmentations. Reducing the template library size from N ($n = 175$) to N/5 ($n = 35$) showed that the accuracy of EC segmentation declined fastest compared to HP and AM, consistently in all algorithms tested. Size of the library had a lower influence on segmentation accuracy of *HybridMulti*, and volume-based approaches (JointFusion, VolMulti) than SurfMulti.

averaging. Noticeably, our approach did not only sequentially apply a volumetric non-linear registration prior to the surface-based segmentation; instead, surface features were integrated with volume data-term into a unified cost function governing the non-linear registration, an approach yielding additional increases in accuracy.

In addition to absolute gain in segmentation accuracy, the proposed *HybridMulti* algorithm demonstrated robust segmentation for our two separate data-sets when the size of the template library was reduced, an important challenge for purely surface-based approaches as shown in our analysis. Indeed, volume-based approaches were inclined to maintain its original accuracy at the largest template library when reducing the size of the library. At the smallest size that was tested in our study ($n = 35$), the accuracy of JointFusion and *HybridMulti* was almost equal in all MTL structures. This informs us to an interesting aspect of feature modeling where local features modeled nearby the structure's boundary may be individually very specific and become powerful with construction of a large training-set. On the other hand, features collected within a "relatively large" volume of interest may include redundant information in a large database but may provide supplementary characteristics of the target structure in case of using a limited size of template library.

In our hybrid approach, tuning of the weight between surface- and volume-features according to the size of a given template library can possibly improve the segmentation accuracy.

Our EC segmentation in the current work (>82%) outperformed a previous study that reported a Dice index of 73% (Hu et al., 2014), and another study that segmented the whole parahippocampal gyrus with a similar degree of accuracy (Heckemann et al., 2006). The performance of *HybridMulti* was also superior to JointFusion and SurfMulti in the current evaluation. Nevertheless, our EC segmentation accuracy was still lower than that of HP and AM, which approached 90%. It is likely that intensity-based segmentation is challenged by the highly variable morphology of the collateral sulcus that defines the border of EC. Also, the posterior end of EC is defined with an external anatomical landmark. Use of a smaller size of template library also showed a faster decline of accuracy in EC than other MTL structures. In the literature (Bernasconi et al., 2001; Pruessner et al., 2002), multiple landmarks were borrowed to address for lack of intensity contrast when defining the border of EC. For example, the medial and lateral boundaries, which meet the same GM structures such as the subiculum of the hippocampus and the collateral sulcus, cannot be defined by the tissue contrast but by landmarks such as a location with a

high angular shape. A human expert may intuitively identify such landmarks whereas the features used in our algorithm do not necessarily take into account them. The suboptimal modeling of these landmarks in our approach is likely the source of inaccuracy in segmentation. This faster decline in accuracy was consistently observed in all algorithms tested. Future works might, therefore, benefit from the incorporation of sulco-gyral shape patterns such as sulcal depth, curvature or spatial relationship with surrounding structures other than HP and AM.

A new multi-scale weighting strategy improved EC and HP segmentation. In particular, the improvement of EC segmentation was higher. This was in line with a previous finding that such a technique mainly improve the segmentation of structures presenting highly variable morphology (Artaechevarria et al., 2009).

The proposed algorithm and JointFusion detected largest effect sizes of atrophy in HP ipsilateral to the epileptic focus and resulted in the most sensitivity to detect hippocampal atrophy among algorithms. Only *HybridMulti* identified EC atrophy among algorithm even if the accuracy yet is to reach human expert's exquisiteness. Our results suggest that the proposed approach may have the potential for clinical utility in the presurgical evaluation of temporal lobe epilepsy.

Varying the parameters for HybridMulti (i.e., the weights for surface-term in the similarity measure and the registration, and the number of templates in the subset) yielded different segmentation accuracy. We determined these parameters in empirical fashion for optimal segmentation performance. We observed that almost same parameter setting were determined for achievement of the best results on both 1.5T and 3T. In a further analysis, we found that these parameters did not differ between segmentation of the three MTL structures. This suggests that the parameters optimized in our study, albeit done empirically, may be generally applicable to segmentation of other datasets or other brain structures. A more thorough analysis is demanded to establish the generalization of the parameters.

For 3T dataset, all the methods resulted in accuracy comparable to the larger 1.5T dataset, with generally decreased SDs. This likely explains that reliable segmentation can be achieved on 3T images where the higher tissue contrast and clearer structural boundaries seen.

As the initial selection was not optimal and we did not like to miss templates which can be potentially useful, we defined a relatively lager subset whereas we set a smaller sample in the subsequent selection with a deformable registration. Our empirical selection of parameters indeed found better segmentation performance was obtained using a larger subset in the initial selection (best performance at $n = 17$) and a smaller set in the latter selection ($n = 8$). The vertex-wise correspondence between individual surface templates defined through SPHARM-PDM ensures the same topology across templates. When we averaged the template shapes, we performed a vertex-wise averaging method that averages the location of a given vertex of the correspondence between templates. The integrity of the topology was not corrupted after this averaging as the same observation is found in a similar process of shape averaging such as in construction of cortical surface template (Styner et al., 2004; Lyttelton et al., 2007).

To determine the number of templates with the best performance, it would be ideal if we observed a plateau occurring after the continuous hiking in Dice index value from the minimum number of templates to test with (**Figure 5**). However, no plateau with an on-going climbing pattern was found in EC, which make difficult to determine when the best performance takes place. The best performance might have been identified if we tested with more templates. This is our limitation as collecting a sufficiently large dataset is often a long-term procedure in the inpatient epilepsy monitoring unit. Thus, it was unrealistic for us to include more data in the study. Alternatively, the very slow increase in Dice index observed at the test with 90+ templates likely explains the increase of the templates would not gain a very significant improvement of the current method. There have been studies dealing with the size of the template library using statistical models (Awate et al., 2012; Awate and Whitaker, 2014).

We did not explore the possible selection of too many similar templates in the subset. A previous study (Wang et al., 2013) investigated this using a joint label fusion technique that address for the covariance of the image appearance between any pair of two templates in the training-set. Generalization of the proposed method across different subcortical structures (e.g., ventricles, striatum, or thalamic nucleus) would be also interesting to enable their morphometry analysis, in particular with regard to size, shape, and variability. We are also working on to extend our current framework to segmentation of the subregions of MTL structures such as hippocampal subfields. The deep learning algorithm using convolutional neural networks (CNN) has been more widely applied in recent works for the medical image segmentation (Kamnitsas et al., 2017; Bao and Chung, 2018; Dolz et al., 2018). Augmentation of our relatively large set of our MRI data and manual annotations could meet the requirement for the training of the CNNs, which can be a proper future extension of our work. We are currently taking steps to make our tools available, including obtaining proper institutional ethics approval, with the plan to ultimately upload the software and training set to a public domain, such as the Neuroimaging Informatics Tools and Resources Clearinghouse (http://www.nitrc.org/).

AUTHOR CONTRIBUTIONS

HK implemented the study design and the algorithm. Performed the evaluation. Wrote and edited the draft. BC tested and re-tested the data using a conventional method to compare. AB provided the MRI data and patient clinical data. Edited the manuscript. NB manually labeled the MTL structures edited the manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fninf. 2018.00039/full#supplementary-material

REFERENCES

Arnone, D., Job, D., Selvaraj, S., Abe, O., Amico, F., Cheng, Y., et al. (2016). Computational meta-analysis of statistical parametric maps in major depression. *Hum. Brain Mapp.* 37, 1393–1404. doi: 10.1002/hbm.23108

Artaechevarria, X., Munoz-Barrutia, A., and Ortiz-de-Solorzano, C. (2009). Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE Trans. Med. Imaging* 28, 1266–1277. doi: 10.1109/TMI.2009.2014372

Asman, A. J., and Landman, B. A. (2013). Non-local statistical label fusion for multi-atlas segmentation. *Med. Image Anal.* 17, 194–208. doi: 10.1016/j.media.2012.10.002

Awate, S. P., and Whitaker, R. T. (2014). Multiatlas segmentation as nonparametric regression. *IEEE Trans. Med. Imaging* 33, 1803–1817. doi: 10.1109/TMI.2014.2321281

Awate, S. P., Zhu, P., and Whitaker, R. T. (2012). How many templates does it take for a good segmentation?: Error analysis in multiatlas segmentation as a function of database size. *Med. Image Comput. Comput. Assist. Interv.* 7509, 103–114.

Bao, S. Q., and Chung, A. C. S. (2018). Multi-scale structured CNN with label consistency for brain MR image segmentation. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* 6, 113–117. doi: 10.1080/21681163.2016.1182072

Bernasconi, N., Bernasconi, A., Caramanos, Z., Antel, S. B., Andermann, F., and Arnold, D. L. (2003). Mesial temporal damage in temporal lobe epilepsy: a volumetric MRI study of the hippocampus, amygdala and parahippocampal region. *Brain* 126, 462–469. doi: 10.1093/brain/awg034

Bernasconi, N., Bernasconi, A., Caramanos, Z., Dubeau, F., Richardson, J., Andermann, F., et al. (2001). Entorhinal cortex atrophy in epilepsy patients exhibiting normal hippocampal volumes. *Neurology* 56, 1335–1339. doi: 10.1212/WNL.56.10.1335

Bernhardt, B. C., Kim, H., and Bernasconi, N. (2013). Patterns of subregional mesiotemporal disease progression in temporal lobe epilepsy. *Neurology* 81, 1840–1847. doi: 10.1212/01.wnl.0000436069.20513.92

Cavedo, E., Boccardi, M., Ganzola, R., Canu, E., Beltramello, A., Caltagirone, C., et al. (2011). Local amygdala structural differences with 3T MRI in patients with Alzheimer disease. *Neurology* 76, 727–733. doi: 10.1212/WNL.0b013e31820d62d9

Collins, D. L., and Pruessner, J. C. (2010). Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. *Neuroimage* 52, 1355–1366. doi: 10.1016/j.neuroimage.2010.04.193

Collins, D. L., Holmes, C. J., Peters, T. M., and Evans, A. C. (1995). Automatic 3-D model-based neuroanatomical segmentation. *Hum. Brain Mapp.* 3, 190–208. doi: 10.1002/hbm.460030304

Collins, D. L., Neelin, P., Peters, T. M., and Evans, A. C. (1994). Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. *J. Comput. Assist. Tomogr.* 18, 192–205. doi: 10.1097/00004728-199403000-00005

Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., and Collins, D. L. (2011). Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. *Neuroimage* 54, 940–954. doi: 10.1016/j.neuroimage.2010.09.018

Dolz, J., Desrosiers, C., and Ben Ayed, I. (2018). 3D fully convolutional networks for subcortical segmentation in MRI: a large-scale study. *Neuroimage* 170, 456–470. doi: 10.1016/j.neuroimage.2017.04.039

Eskildsen, S. F., Coupé, P., Fonov, V., Manjón, J. V., Leung, K. K., Guizard, N., et al. (2012). BEaST: brain extraction based on nonlocal segmentation technique. *Neuroimage* 59, 2362–2373. doi: 10.1016/j.neuroimage.2011.09.012

Goncharova, I. I., Dickerson, B. C., Stoub, T. R., and deToledo-Morrell, L. (2001). MRI of human entorhinal cortex: a reliable protocol for volumetric measurement. *Neurobiol. Aging* 22, 737–745. doi: 10.1016/S0197-4580(01)00270-6

Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D., and Hammers, A. (2006). Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *Neuroimage* 33, 115–126. doi: 10.1016/j.neuroimage.2006.05.061

Hu, S., Coupé, P., Pruessner, J. C., and Collins, D. L. (2014). Nonlocal regularization for active appearance model: application to medial temporal lobe segmentation. *Hum. Brain Mapp.* 35, 377–395. doi: 10.1002/hbm.22183

Joo, E. Y., Kim, H., Suh, S., and Hong, S. B. (2014). Hippocampal substructural vulnerability to sleep disturbance and cognitive impairment in patients with chronic primary insomnia: magnetic resonance imaging morphometry. *Sleep* 37, 1189–1198. doi: 10.5665/sleep.3836

Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., et al. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 36, 61–78. doi: 10.1016/j.media.2016.10.004

Keihaninejad, S., Heckemann, R. A., Gousias, I. S., Hajnal, J. V., Duncan, J. S., Aljabar, P., et al. (2012). Classification and lateralization of temporal lobe epilepsies with and without hippocampal atrophy based on whole-brain automatic MRI segmentation. *PLoS ONE* 7:e33096. doi: 10.1371/journal.pone.0033096

Khan, A. R., Cherbuin, N., Wen, W., Anstey, K. J., Sachdev, P., and Beg, M. F. (2011). Optimal weights for local multi-atlas fusion using supervised learning and dynamic information (SuperDyn): validation on hippocampus segmentation. *Neuroimage* 56, 126–139. doi: 10.1016/j.neuroimage.2011.01.078

Kim, H., Chupin, M., Colliot, O., Bernhardt, B. C., Bernasconi, N., and Bernasconi, A. (2012a). Automatic hippocampal segmentation in temporal lobe epilepsy: impact of developmental abnormalities. *Neuroimage* 59, 3178–3186. doi: 10.1016/j.neuroimage.2011.11.040

Kim, H., Mansi, T., Bernasconi, N., and Bernasconi, A. (2012b). Surface-based multi-template automated hippocampal segmentation: application to temporal lobe epilepsy. *Med. Image Anal.* 16, 1445–1455. doi: 10.1016/j.media.2012.04.008

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *Siam J. Optimiz.* 9, 112–147. doi: 10.1137/S1052623496303470

Landman, B., and Warfield, S. (2012). "MICCAI 2012 workshop on multi-atlas labeling," in *Proc. Med. Image Comput. Comput. Assisted Intervent. Conf. Grand Challenge Workshop Multi-Atlas Labeling Challenge Result* (Nice).

Lyttelton, O., Boucher, M., Robbins, S., and Evans, A. (2007). An unbiased iterative group registration template for cortical surface analysis. *Neuroimage* 34, 1535–1544. doi: 10.1016/j.neuroimage.2006.10.041

Maccotta, L., Moseley, E. D., Benzinger, T. L., and Hogan, R. E. (2015). Beyond the CA1 subfield: local hippocampal shape changes in MRI-negative temporal lobe epilepsy. *Epilepsia* 56, 780–788. doi: 10.1111/epi.12955

Pruessner, J. C., Kohler, S., Crane, J., Pruessner, M., Lord, C., Byrne, A., et al. (2002). Volumetry of temporopolar, perirhinal, entorhinal and parahippocampal cortex from high-resolution MR images: considering the variability of the collateral sulcus. *Cereb. Cortex* 12, 1342–1353. doi: 10.1093/cercor/12.12.1342

Schramm, J., and Clusmann, H. (2008). The surgery of epilepsy. *Neurosurgery* 62(Suppl. 2), 463–481; discussion 481. doi: 10.1227/01.neu.0000316250.69898.23

Shi, J., Thompson, P. M., Gutman, B., Wang, Y., and Alzheimer's Disease Neuroimaging, I. (2013). Surface fluid registration of conformal representation: application to detect disease burden and genetic influence on hippocampus. *Neuroimage* 78, 111–134. doi: 10.1016/j.neuroimage.2013.04.018

Shoemake, K. (1985). Animating rotation with quaternion curves. *SIGGRAPH Comput. Graph.* 19, 245–254. doi: 10.1145/325165.325242

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97. doi: 10.1109/42.668698

Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognit.* 32, 71–86. doi: 10.1016/S0031-3203(98)00091-0

Styner, M., Lieberman, J. A., Pantazis, D., and Gerig, G. (2004). Boundary and medial shape analysis of the hippocampus in schizophrenia. *Med. Image Anal.* 8, 197–203. doi: 10.1016/j.media.2004.06.004

Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., and Gerig, G. (2006a). "Statistical shape analysis of brain structures using SPHARM-PDM," in *MICCAI 2006 Opensource Workshop* (Copenhagen).

Styner, M., Oguz, I., Xu, S., Brechbuhler, C., Pantazis, D., Levitt, J. J., et al. (2006b). Framework for the statistical shape analysis of brain structures using SPHARM-PDM. *Insight J.* 242–250.

Wang, H. Z., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., and Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 611–623. doi: 10.1109/TPAMI.2012.143

Wang, Z., Neylan, T. C., Mueller, S. G., Lenoci, M., Truran, D., Marmar, C. R., et al. (2010). Magnetic resonance imaging of hippocampal subfields in posttraumatic stress disorder. *Arch. Gen. Psychiatry* 67, 296–303. doi: 10.1001/archgenpsychiatry.2009.205

Check for
updates

Analytic Tools for Post-traumatic Epileptogenesis Biomarker Search in Multimodal Dataset of an Animal Model and Human Patients

Dominique Duncan[1]*, Giuseppe Barisano[1], Ryan Cabeen[1], Farshid Sepehrband[1], Rachael Garner[1], Adebayo Braimah[1], Paul Vespa[2], Asla Pitkänen[3], Meng Law[1] and Arthur W. Toga[1]

[1] Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, United States, [2] Division of Neurosurgery, Department of Neurology, University of California at Los Angeles School of Medicine, Los Angeles, CA, United States, [3] A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland

Epilepsy is among the most common serious disabling disorders of the brain, and the global burden of epilepsy exerts a tremendous cost to society. Most people with epilepsy have acquired forms of the disorder, and the development of antiepileptogenic interventions could potentially prevent or cure epilepsy in many of them. However, the discovery of potential antiepileptogenic treatments and clinical validation would require a means to identify populations of patients at very high risk for epilepsy after a potential epileptogenic insult, to know when to treat and to document prevention or cure. A fundamental challenge in discovering biomarkers of epileptogenesis is that this process is likely multifactorial and crosses multiple modalities. Investigators must have access to a large number of high quality, well-curated data points and study subjects for biomarker signals to be detectable above the noise inherent in complex phenomena, such as epileptogenesis, traumatic brain injury (TBI), and conditions of data collection. Additionally, data generating and collecting sites are spread worldwide among different laboratories, clinical sites, heterogeneous data types, formats, and across multi-center preclinical trials. Before the data can even be analyzed, these data must be standardized. The Epilepsy Bioinformatics Study for Antiepileptogenic Therapy (EpiBioS4Rx) is a multi-center project with the overarching goal that epileptogenesis after TBI can be prevented with specific treatments. The identification of relevant biomarkers and performance of rigorous preclinical trials will permit the future design and performance of economically feasible full-scale clinical trials of antiepileptogenic therapies. We have been analyzing human data collected from UCLA and rat data collected from the University of Eastern Finland, both centers collecting data for EpiBioS4Rx, to identify biomarkers of epileptogenesis. Big data techniques and rigorous analysis are brought to longitudinal data collected from humans and an animal model of TBI, epilepsy, and their interaction.

The prolonged continuous data streams of intracranial, cortical surface, and scalp EEG from humans and an animal model of epilepsy span months. By applying our innovative mathematical tools via supervised and unsupervised learning methods, we are able to subject a robust dataset to recently pioneered data analysis tools and visualize multivariable interactions with novel graphical methods.

Keywords: MRI, EEG, epilepsy, epileptogenesis, informatics, neuroimaging, TBI, biomarker

INTRODUCTION

The goal of the Epilepsy Bioinformatics Study for Antiepileptogenic Therapy (EpiBioS4Rx) is to identify relevant biomarkers of epileptogenesis after traumatic brain injury (TBI) and perform rigorous preclinical trials that permit the future design and performance of economically feasible full-scale clinical trials of antiepileptogenic therapies. Discovering these biomarkers of epileptogenesis is challenging, because this process is multifactorial and involves multiple modalities. We have been collecting and analyzing multimodal data, including neuroimaging, electrophysiology, and molecular/serological/tissue. An informatics infrastructure has been created to facilitate analysis and collaboration among scientists from various centers around the world (Duncan et al., 2018b). We have been developing innovative analytic tools to be shared with the broader epilepsy research community so that others may use our tools in addition to their own tools to advance research in this field. By working on this difficult problem collaboratively among researchers who possess different areas of expertise, we expect to identify several biomarkers of post-traumatic epileptogenesis from the multimodal data collected as part of EpiBioS4Rx and validate those biomarkers.

Substantial research has been devoted to investigate imaging biomarkers of epileptogenesis following TBI in an effort to better understand, prevent, and potentially treat post-traumatic epilepsy (PTE). Although incidence of PTE has been correlated with various factors, these results have been gathered and interpreted independently and are often drawn from models of human temporal lobe epilepsy, animal models of induced TBI via fluid percussion injury (FPI), and pilocarpine or kainic acid-induced status epilepticus. There has been limited investigation directly comparing these models to human cohort studies of epileptogenesis following trauma, which is one area in which our work extends on existing research on PTE. Also, few multimodality studies have been conducted to investigate interrelations among identified potential biomarkers, which could assist in establishing a panel of non-invasive epileptogenic biomarkers that consistently precedes and predicts the development of PTE. EpiBioS4Rx is collecting large-scale imaging data on TBI patients with subsequent seizure activity as well as imaging data on a rodent model of TBI, allowing for a multimodality and multi-species investigation.

Several reviews have summarized electrophysiological (Worrell, 2011; Staba et al., 2014) and imaging (Mishra et al., 2011; van Vliet et al., 2017; Pitkänen et al., 2018) biomarkers identified in rat models and human patients in recent years.

Notably, high frequency oscillations (HFOs), standard frequency between 80 and 600 Hz (Staba et al., 2014), are consistently produced by epileptic neural tissues (Bragin Engel et al., 1999; Jacobs et al., 2012; Zijlmans et al., 2012) and have also been reported in rats after administration of lateral FPI within or adjacent to the injured tissue (Reid et al., 2016). In the same FPI model, pathologic HFOs and repetitive HFOs and spikes (rHFOSs) occurred within 2 weeks of insult only in rats that would later develop seizures (Reid et al., 2016). However, currently there are no validated electrophysiological biomarkers of post-traumatic epileptogenesis (Perucca et al., 2018), so one of our goals is to identify electrophysiological biomarkers that can be validated. As many models of PTE involve continuous EEG recordings, automated seizure detection programs have been investigated to ease data analysis. Approximate entropy (ApEn), in conjunction with neural networks, has been introduced as an analytic tool to discriminate normal and ictal or pre-ictal EEG from epileptic patients and healthy controls (Liang et al., 2010), refining and enhancing seizure detection, which can ultimately expedite the EEG analysis workflow.

Magnetic Resonance Imaging (MRI) and Diffusion Tensor Imaging (DTI) have allowed for non-invasive analysis of molecular and structural alterations of white matter and other neural structures at high spatial resolution. MRI may be used to identify specific abnormalities associated with increased susceptibility to epileptogenesis, including focal lesions (D'Alessandro et al., 1982; Dalessandro et al., 1988), intracerebral hemorrhage (D'Alessandro et al., 1982), biparietal contusions (Englander et al., 2003) and dural penetration from bone or metal fragments (Englander et al., 2003). In a lateral FPI model, diffusion tensor trace alterations in the hippocampus acquired 3 hours after injury were found to predict seizure susceptibility and number of spikes 12 months later (Kharatishvili et al., 2007). A follow up study confirmed that Dav (one third of the trace of the diffusion tensor that is an orientation-independent measure of water diffusion) at 23 days and 2 months and T1p (a longitudinal relaxation in the rotating frame, which can be assumed to be similar to T1 relaxation in the very low magnetic field, thus probing interaction between water and macromolecules in the tissue) at 9 days post insult could predict increased seizure susceptibility following lateral FPI (Immonen et al., 2013).

Axonal damage, visualized with DTI, is seen across all severities of TBI, although irreversible myelin damage, which is correlated with worse cognitive prognoses, is more typically caused by moderate and severe TBI (Kraus et al., 2007). Decreased fractional anisotropy (FA) has been repeatedly found in TBI patients compared with healthy controls (Bendlin et al.,

2008; Sidaros et al., 2008; Irimia et al., 2014), which is especially relevant considering FA ratios have been found to be significantly reduced in TBI patients who developed late post-traumatic seizures compared with non-epileptic TBI patients (Gupta et al., 2005), and along temporal lobe white matter in benign mesial TLE (Labate et al., 2015). Additionally, connectomic studies and tract-based spatial statistics may assist in the understanding of how white matter degeneration patterns lead to neural and cognitive impairment (Irimia et al., 2014), so they may also support a greater understanding in how degeneration patterns specifically lead to PTE. We plan to use our pipelines for connectomics to understand the development of PTE as well as relate these imaging data to the electrophysiological data.

MRI also serves as a useful tool for morphometric analysis. TBI varies significantly in the severity of insult and subsequent lesion(s), so precise lesion quantification is necessary to compare outcomes following stratified severity of injury. Voxel-based morphometry analysis has indicated reduced hippocampal and thalamic volumes in TLE patients (Labate et al., 2008). In a lateral FPI model, Shultz et al. found that hippocampal surface shape analysis (conducted via MRI-based large-deformation high-dimensional mapping) at 1 week post-injury could be predictive of PTE. Rats that later developed PTE showed increased lateral regions while non-epileptic rats showed decreased medial and ventral regions (Shultz et al., 2013). We have developed analysis pipelines to analyze both animal and human imaging data to relate these and explore the translational components of the animal data.

Several supervised and unsupervised models of lesion identification and quantification from T1, T2, and FLAIR images acquired from MRI have been introduced in an effort to automate analysis of multiple sclerosis (Wetter et al., 2016), tumor (Guo et al., 2015), chronic stroke (Pustina et al., 2016; Guo et al., 2018), and TBI (Irimia et al., 2011). Automated quantification of TBI lesions by normalizing and standardizing against standard templates is challenging given that brain morphology is often distorted due to insult (Kim et al., 2008), so our work aims to quantify TBI lesions automatically while maintaining accuracy.

Transforming Research and Clinical Knowledge in TBI (TRACK-TBI) was a study performed at the University of California, San Francisco (main site) that proved the feasibility of large-scale, multi-site analysis of imaging, blood, and clinical data on nearly 3,000 TBI patients. Patient data gathered through TRACK-TBI have been used to examine the relationship between CT and MRI findings that are commonly assessed in emergency trauma facilities and DTI, both of which have been reported as potential biomarkers of epileptogenesis following TBI. In mild TBI cases, FA is significantly reduced in CT/MRI-positive (acute intracranial lesion, including epidural or subdural hematoma, subarachnoid hemorrhage, contusion, axonal injury, or skull fracture) and not reduced in CT/MRI-negative patients (Yuh et al., 2014). DTI can detect alterations in microstructural white matter with greater subtlety than MRI, and FA ratios have been found to be significantly reduced in TBI patients who developed late PTS compared with non-epileptic TBI patients (Gupta et al., 2005). In another study, mild TBI patients with CT/MRI-positive (defined as having any evidence of lesion) and CT/MRI-negative

(no lesions) showed distinct alterations of functional connectivity in resting state fMRI analysis within days of injury that were predictive of cognitive outcomes 6 months later (Palacios et al., 2017).

The EpiBioS4Rx informatics infrastructure contains a thorough and harmonized multimodal database, including imaging and EEG data, which enables researchers to correlate results from imaging analysis to longitudinal epileptiform activity (Duncan et al., 2018b) from both humans and an animal model. Recently, analysis of EpiBioS4Rx data found that early post-traumatic seizures and subsequent development of PTE following severe TBI are strongly correlated with lesions localized to the temporal lobe (i.e., hemorrhagic temporal lobe injury) but not general lesion severity (as measured by the Glasgow Coma Scale) (Tubi et al., 2018).

DATA

The total amount of data that has been and will be collected in the ongoing EpiBioS4Rx includes EEG and video-EEG (video tape recording during EEG monitoring) from cohorts of animals after TBI (using FPI) recorded continuously for 6 months, in addition to prolonged continuous intensive care unit (ICU) EEG recordings from 300 humans, including depth EEG from 100 patients, and intermittent sampling of brain images, blood, and tissue data over 2 years. The collected rat MRI consist of structural and diffusion weighted measures. Sprague-Dawley control rats and TBI rats (left lateral fluid percussion injury) were used with data collected using a Bruker BioSpin MRI GmbH using a dtiEpiT SpinEcho sequence (Duncan et al., 2018b). Patients admitted into the ICU after an acute moderate-severe TBI involving a frontal and/or temporal lobe hemorrhagic contusion are screened for the study. Although a number of sites are collecting data for EpiBioS4Rx, we focus our preliminary analysis on human data from the University of California, Los Angeles (UCLA) and animal data from the University of Eastern Finland, Kuopio.

ANALYSIS METHODS

We present a collection of analytic tools for this multimodal dataset and present examples of some preliminary work on sample data from EpiBioS4Rx as well as future directions for this analysis.

Imaging Methods

We have developed a multimodal image analysis workflow that includes lesion mapping and tractography reconstruction of white matter pathways. Additionally, we have analyzed paravascular spaces (PVS) in the MRI data to aid in our search for post-traumatic epileptogenesis biomarkers.

Lesion Mapping

Lesions were mapped from fluid-attenuated inversion recovery (FLAIR) images with an automated segmentation pipeline using FMRIB Software Library (FSL) tools (Woolrich et al., 2009; Jenkinson et al., 2012; Wetter et al., 2016). FLAIR suppresses

the signal produced by cerebrospinal fluid (CSF) and is sensitive to contrast for mapping lesions in TBI (Gentry et al., 1988; Bigler, 2001; Narayana, 2017). The pipeline begins with skull stripping, smoothing, and intensity normalization. Then lesions are separated from brain tissue and CSF using a histogram-based thresholding algorithm. Finally, lesions not overlapping with white matter (WM) are discarded by registering a WM mask from a standard space into the subject space (FSL-FNIRT is used for the registration). An example is shown in **Figure 1**.

In order to separate periventricular WM hyperintensities from the rest of the WM lesions, we performed a secondary analysis on the T1-weighted (T1w) images. Structural T1w images are less sensitive to periventricular lesions due to CSF partial volume effect, yet they can visualize WM lesions across the brain. The T1w images were analyzed through a similar pipeline as the FLAIR images, and the lesions were mapped accordingly.

Tractography

We have developed diffusion MR image analysis pipelines for quantitative analysis of WM microstructure and connectivity across both rodent and human datasets. Tractography models were created from the diffusion-weighted MRI (dMRI) data using FSL (Jenkinson et al., 2012) and the Quantitative Imaging Toolkit (QIT) (Cabeen et al., 2018). The dMRIs were first skull stripped using FSL Brain Extraction Tool (BET) and then corrected for motion and eddy current artifacts using FSL FMRIB's Linear Image Registration Tool (FLIRT). For this, each diffusion scan was affinely registered to the baseline scan using the mutual information metric, and the associated gradient orientations were rotated to account for the registration. Diffusion tensor models were then estimated from the dMRI using QIT, and the following tensor parameters were extracted: fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD). A study specific template was created using Diffusion Tensor Imaging ToolKit (DTI-TK) (Zhang et al., 2006), and the deformation field for each scan was used to register the data to the Illinois Institute of Technology (IIT) brain template (Zhang et al., 2011) to subject native space. Tractography models of our bundles of interest, including uncinate fasciculus, anterior thalamic radiation, corticospinal tract, inferior longitudinal fasciculus, superior longitudinal fasciculus, fornix, arcuate fasciculus, and five subdivisions of

the corpus callosum were created using a framework for deterministic streamline integration (Cabeen et al., 2016). For each bundle, seed, inclusion, and exclusion masks were manually drawn in the IIT template (Wakana et al., 2007) in reference to a white matter atlas (Catani and Thiebaut de Schotten, 2008). The template masks were then resampled in each subject's native space image to constrain tractography. Other tractography parameters included a step size of 1.0 mm, a maximum angle of 45°C, and a minimum FA of 0.15–25,000 seeds per bundle. Bundle-specific metrics were then computed, including bundle volume, track density, track length, and averages of DTI metrics listed above. In addition to tractography analysis, the human data were also analyzed using voxel-based analysis to obtain diffusion MRI metrics in anatomical regions derived from the Johns Hopkins white matter atlas (Mori et al., 2008; Cabeen et al., 2017). This method applies to human data (**Figure 2**) as well as rodent data (**Figures 3**, **4**). We found that the data allowed multi-fiber modeling to resolve partial volume effects and crossing fiber configurations.

Paravascular Spaces

Many studies have shown that paravascular spaces (PVSs) may play an important role in neuroinflammation: a strong post-traumatic inflammatory reaction was documented in PVSs of contused human brain tissue, suggesting that PVSs' impairment could explain the altered macrophage activity resulting in seizure onset (Holmin et al., 1998; Bechmann et al., 2001; Corraliza, 2014; Abiega et al., 2016). Also structural changes in PVSs may affect their surrounding white matter networks (Taoka et al., 2017). We investigate the role of paravascular spaces in TBI as a potential biomarker for post-traumatic epilepsy.

Study population

We present some analysis performed on human data, focusing on PVSs' role as a potential biomarker of epileptogenesis after TBI; we analyzed clinical data and MRI scans in a sample of 15 patients (12 males, 3 females, age range: 7–68 years old). MRI scans were performed 14 days after trauma using a 3T MRI scanner. PVSs were analyzed on 3D T2 Turbo Spin Echo (TSE) sequences. Six healthy subjects (3 males, 3 females, age range: 12–62 years old) were used as controls. Demographic characteristics of TBI patients and healthy subjects are summarized in **Table 1**.
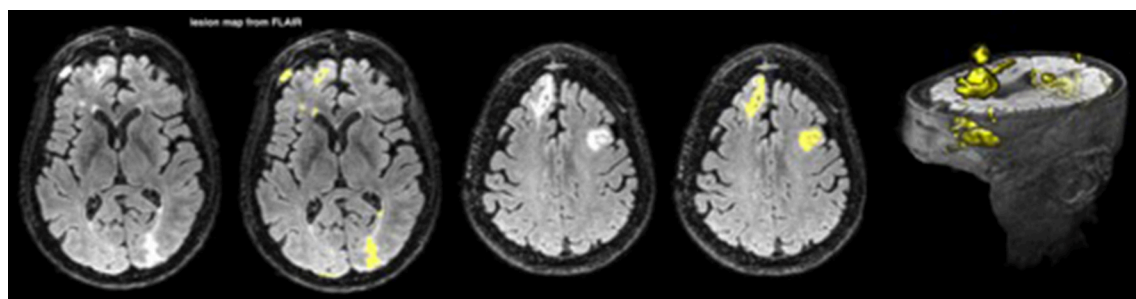

FIGURE 1 | A lesion map from FLAIR for one patient; the lesions are depicted in yellow.

PVS analysis

PVSs were defined as tubular-linear or round-ovoid structures with a CSF-like signal intensity (hyperintense on T2-weighted images) and a diameter of <3 mm. PVSs surround perforating vessels in the brain, and the largest number of PVSs is usually found in the basal ganglia and centrum semiovale. The typical shape, dimensions, and location were used to exclude other possible differential diagnoses (e.g., lacunar infarcts). In this study, we omitted PVS with a diameter of <0.5 mm, because their identification and measurement were not sufficiently reliable.

Image processing on the 3D T2 TSE images was performed in OsiriX Image Viewing Software (Ratib and Rosset, 2006) by a reader blinded to subjects' clinical data. In each subject, we manually marked and counted all PVSs with a diameter between 0.5 and 3 mm. The caliber of PVS was measured with the Ruler Tool in OsiriX. Both the total number of PVSs and the caliber of each PVS were systematically recorded. We categorized PVS by location in the cerebral hemisphere (right and left) to assess the distribution of PVS in the brain. Because of the possible inter-individual variability in the total number of PVSs, we calculated 2 ratios (HR_{right} and HR_{left}) between each hemisphere's amount of PVS (PVS_{right} and PVS_{left}, respectively) and the sum of PVS in the whole brain (PVS_{tot}) for each subject:

Two possible outcomes resulted from these ratios:

- Two equivalent values ($HR_{right} = HR_{left} = 0.5$) if no difference in the number of PVSs was found between the right and left hemispheres;
- Two different values ($HR_{right} \neq HR_{left}$) if the number of PVSs was not the same in the 2 hemispheres. In this situation, we defined HR_{minor} and HR_{major} to be the ratios obtained from the hemisphere with less and more PVS compared to the contralateral hemisphere, respectively. Consequently:

$$0.0 < HR_{minor} < 0.5 \quad and \quad 0.5 > HR_{major} > 1 \qquad (1)$$

Then we calculated the difference between HR_{minor} and HR_{major} as an asymmetry index (AI):
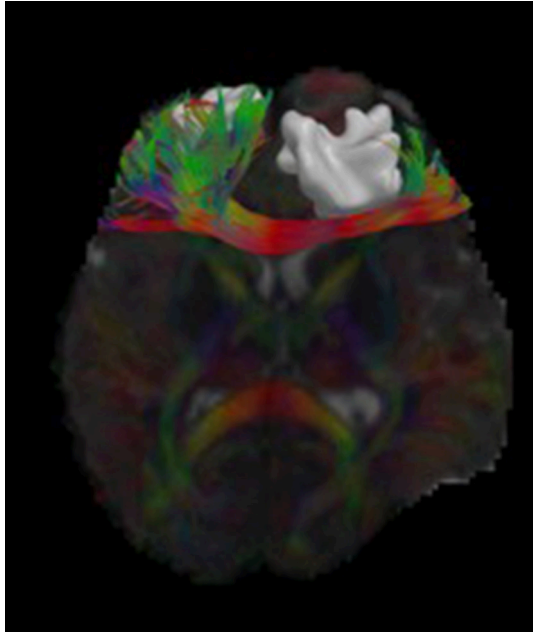

FIGURE 2 | Visualizations of diffusion MRI data from a single human subject. The image shows an axial brain slice rendered with glyphs depicting the underlying multi-compartment diffusion models. A tractography reconstruction of the forceps minor is shown alongside a brain lesion. Through 3D modeling and visualization, we are able to show the impact of the brain trauma on structural connectivity of the frontal lobe.
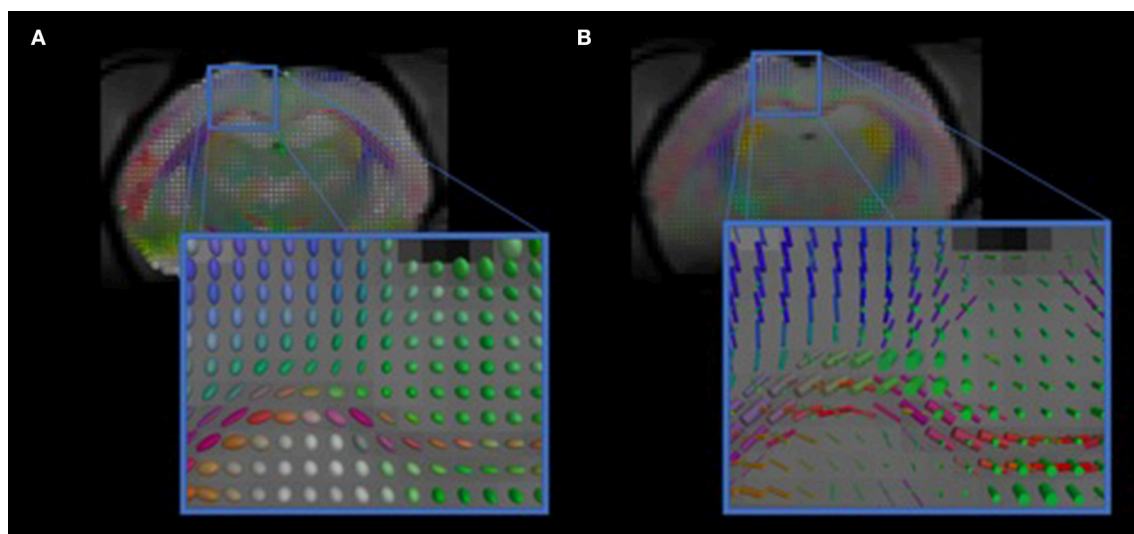

FIGURE 3 | Visualizations of diffusion MRI from the rodent data. The images show diffusion models estimated in each voxel. (A) shows standard diffusion tensor modeling, and (B) shows multi-compartment modeling that resolves complex anatomical features, such as crossing fibers.
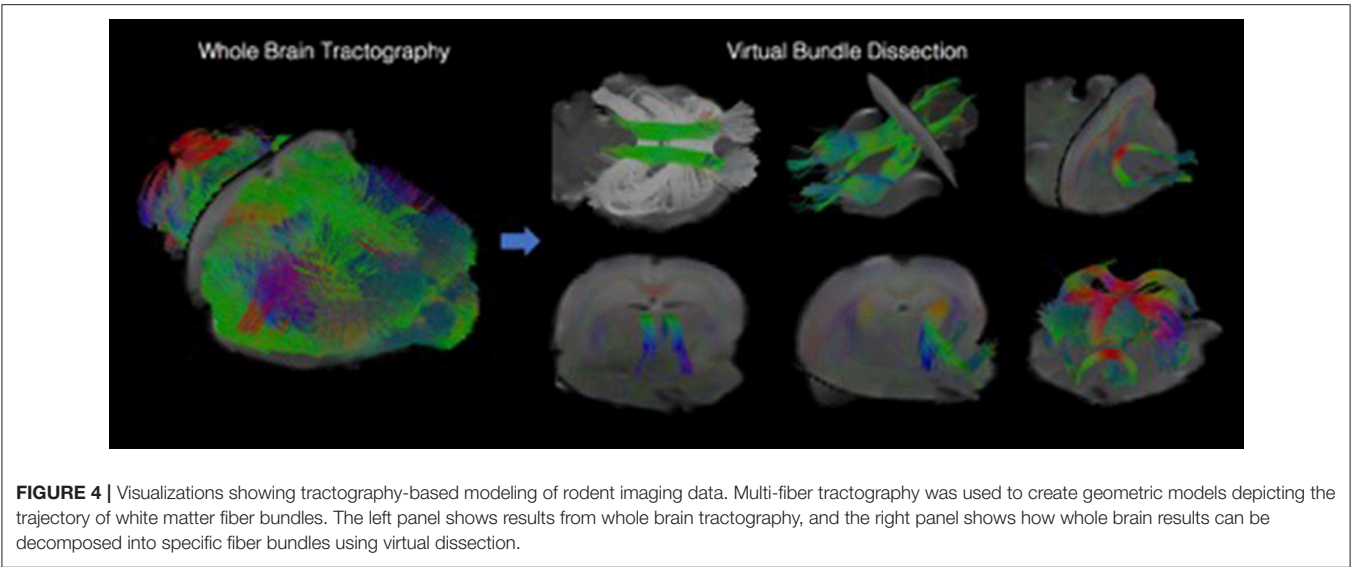
FIGURE 4 | Visualizations showing tractography-based modeling of rodent imaging data. Multi-fiber tractography was used to create geometric models depicting the trajectory of white matter fiber bundles. The left panel shows results from whole brain tractography, and the right panel shows how whole brain results can be decomposed into specific fiber bundles using virtual dissection.

TABLE 1 | Demographic characteristics of TBI patients and healthy subjects.

Characteristics		TBI patients	Healthy controls
Subjects (#)	Total	15	6
	Male	12	3
	Female	3	3
Age at scan (years) [Mean ± standard deviation]	Total	34 ± 23	30 ± 17
	Male	35 ± 22	23 ± 12
	Female	30 ± 29	37 ± 22

$$AI = HRmajor - HRminor \qquad (2)$$

with $0 \leq AI \leq 1$

The higher the AI value was, the more asymmetric the distribution of PVS in the brain was. As a physiological right-left asymmetry in the brain has been reported in previous studies (Asgari et al., 2016; Feldman et al., 2018), and an unbalanced distribution of PVS may be considered normal, we used a threshold of $AI \geq 0.2$ to define a significantly high asymmetry in PVS distribution. This value means that one hemisphere has more than 60% of the total number of PVSs.

We measured the caliber of each marked PVS, and the average of PVS caliber in the right and left hemispheres (C_{right} and C_{left}, respectively) in all subjects. Then we calculated the difference ($|C_{diff}|$) between the mean PVS caliber in the two hemispheres:

$$|Cdiff| = Cright - Cleft \qquad (3)$$

Statistical analysis

A Student's *t*-test was used to determine if there was a difference in the total number and the mean distribution of PVSs between the two cerebral hemispheres in the healthy controls and TBI group. A difference of $p < 0.05$ was considered statistically significant.

RESULTS

Total Number of PVSs

We evaluated the total number of PVSs in TBI patients and healthy controls: the average was 77 ± 48 in the first group, and 80 ± 15 in the latter. No significant difference was found between the two groups ($p = 0.40$).

In our population, we found a weak positive correlation between age and the number of PVSs (Pearson's $\rho = 0.28$, $p = 0.11$), as shown in **Figure 5**.

Asymmetry Analysis

Both TBI patients and healthy controls presented a different number of PVSs in the two cerebral hemispheres. The HR range was 0.29–0.71 in TBI patients and 0.43–0.54 in healthy controls; in the patient group, the mean HR_{minor} and HR_{major} were 0.42 and 0.58, respectively, while in the control group, the values were 0.48 and 0.52, respectively (**Figure 6**). The degree of asymmetry was significantly different in the two groups ($p = 0.001$): the average AI was 0.17 in TBI patients and 0.04 in control subjects.

In the TBI group, we found six patients with a highly asymmetric distribution of PVS (**Figure 7**) in the two cerebral hemispheres (AI $\geq$ 0.2). Five of these patients (83%) experienced at least one seizure within the first six months after TBI (in four cases, the seizure happened within the first month); in three cases, Lateralized Periodic Discharges (LPDs) were detected in the EEG, and in all cases, the affected hemisphere matched with the hemisphere where less PVSs were identified. Furthermore, in all nine TBI patients with intermediate- or high-grade PVS asymmetry, the cerebral hemisphere that suffered the trauma showed a minor number of PVSs compared with the contralateral side.

PVS Caliber Analysis

The mean PVS caliber in TBI patients and healthy controls were 1.37 ± 0.23 mm and 1.31 ± 0.26 mm, respectively: the difference in the two groups was not statistically significant ($p = 0.39$). We

found a significant positive correlation between *AI* and $|C_{diff}|$, as illustrated in the scatter plot in **Figure 8** (Pearson's $\rho = 0.41$, $p = 0.03$).

Patients with a more asymmetric distribution of PVS in the brain had a greater difference in the mean PVS caliber between right and left hemispheres. In patients who had a post-traumatic seizure, smaller PVSs were measured on the side ipsilateral to LPDs and/or affected by the trauma, compared with the contralateral hemisphere. In four patients, the difference in the PVS caliber between the two hemispheres was statistically significant (*p*-values were 0.031, 0.036, 0.034, and 0.049). Thus, the evaluation of PVS distribution and quantification may represent another potential non-invasive neuroimaging biomarker to predict the development of epilepsy after TBI.
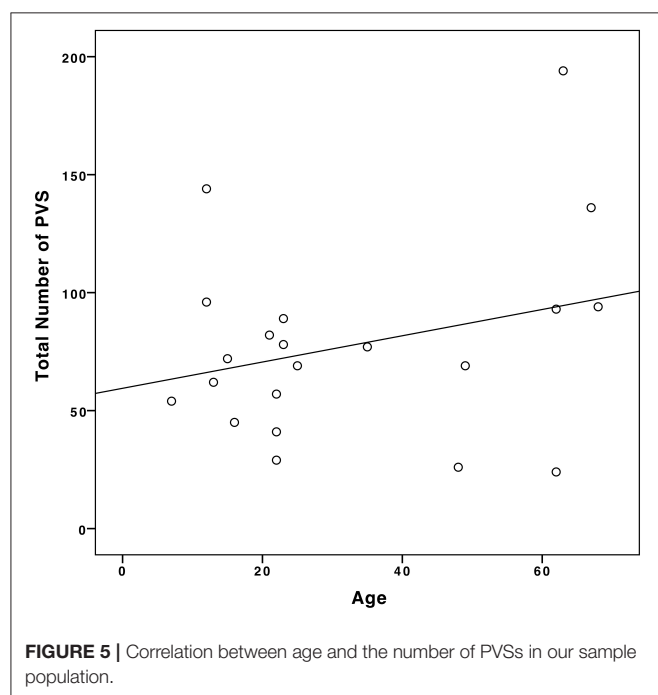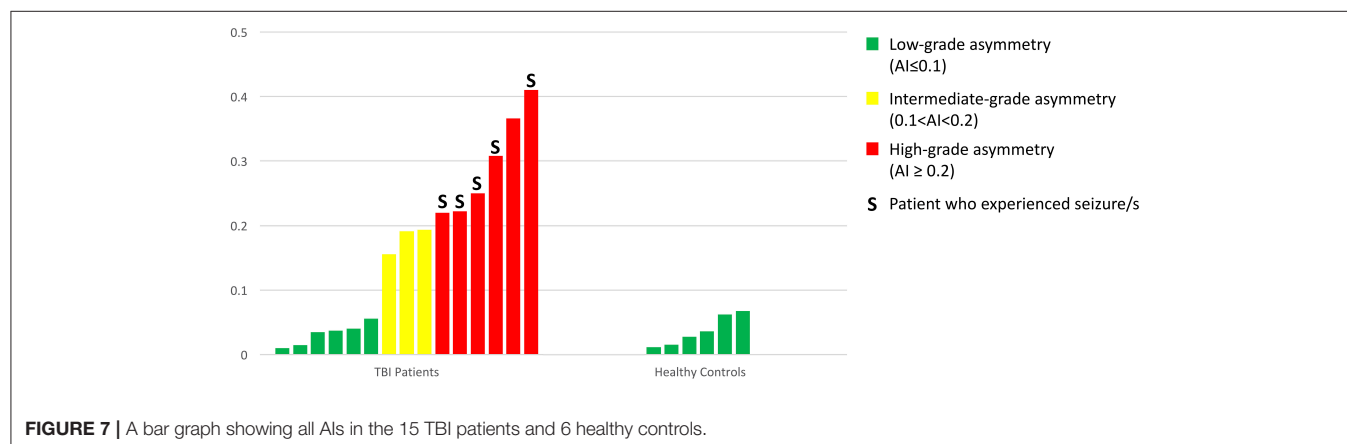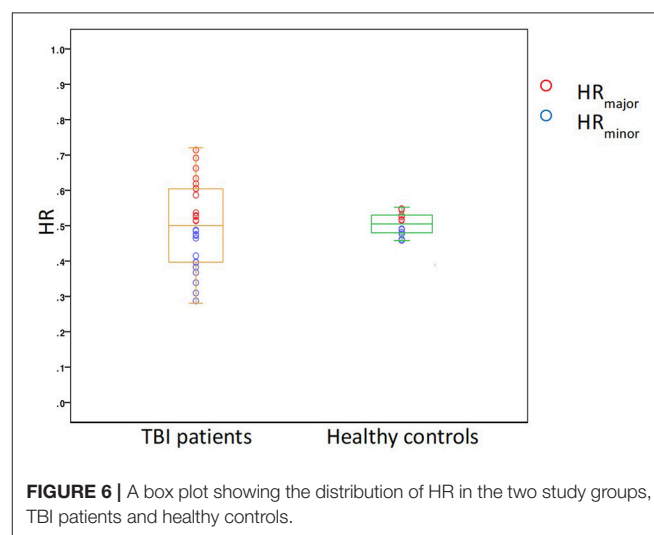
EEG Methods

Various analytic tools were used to analyze both human and rodent EEG. Notably, dimensionality reduction techniques, including diffusion maps and Unsupervised Diffusion Component Analysis (UDCA), were used to elucidate patterns or abnormal activity within large data matrices that may be used to potentially identify biomarkers of epileptogenesis after TBI. Spectral analysis and measures of relationship, such as mutual information, were also conducted. We present an overview of a few analytic tools for EEG with some figures of examples of preliminary results using EpiBioS4Rx data.

Spectral Analysis

As a first step, raw EEG data were imported via EEGLAB in MATLAB (Delorme and Makeig, 2004). The Short Time Fourier Transform (STFT) was applied to the raw, unfiltered EEG data, seen in **Figure 9**, and spectrograms were formed to visualize frequency changes over time. 3D spectrograms, such as **Figure 10**, show the relationship among time, amplitude, and power in addition to the power spectral density (PSD). These

FIGURE 5 | Correlation between age and the number of PVSs in our sample population.

FIGURE 6 | A box plot showing the distribution of HR in the two study groups, TBI patients and healthy controls.

FIGURE 7 | A bar graph showing all AIs in the 15 TBI patients and 6 healthy controls.

plots can be used for visualization purposes or for setting a threshold to focus on a specific frequency range, for example, and then quantifying changes over time.

Persyst Software Tools

We also use Persyst software (Sierra-Marcos et al., 2015) as a tool for visualization of the EEG and for artifact removal, spike detection, and epileptiform activity identification.
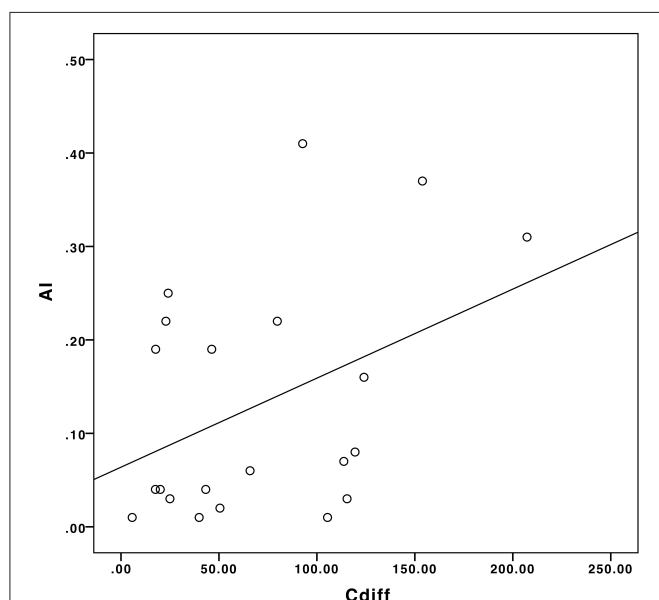
Mutual Information

Another type of analysis that we perform considers measures of relationship, such as mutual information (Duncan et al., 2013a), to study how electrical activity from different areas of the brain relate to each other and how those relationships change over time. We plan to relate these measures of relationship in the EEG to the resting state fMRI to determine if electrode contacts from areas within resting state networks have higher values of mutual information and if these networks differ between patients who develop PTE and those who do not.
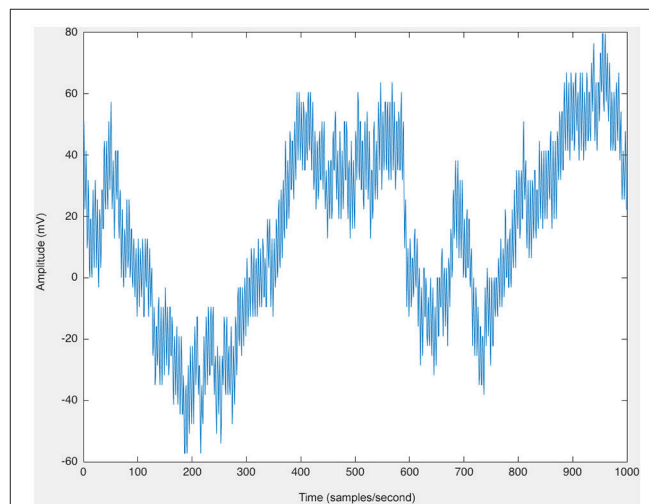

FIGURE 8 | The correlation between C_{diff} and AI in our sample population.

FIGURE 9 | The raw EEG from one channel of human scalp EEG data (200 samples/second).
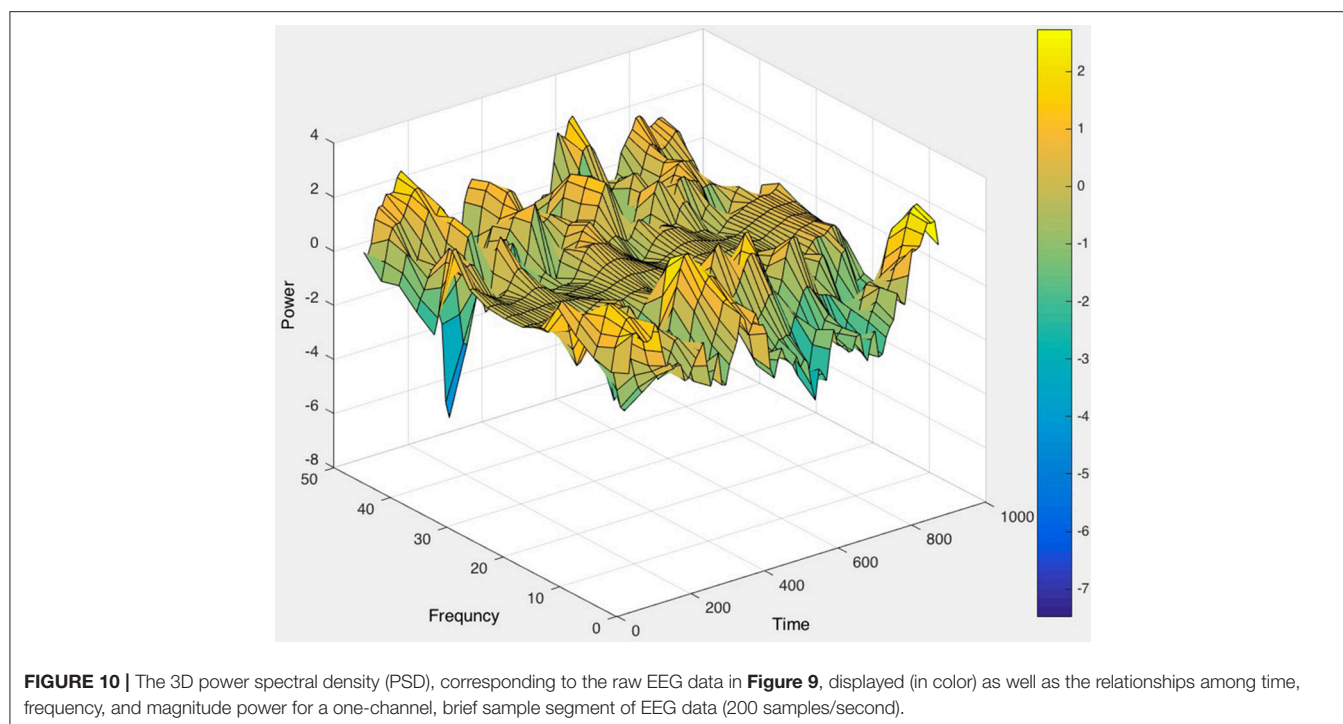

FIGURE 10 | The 3D power spectral density (PSD), corresponding to the raw EEG data in **Figure 9**, displayed (in color) as well as the relationships among time, frequency, and magnitude power for a one-channel, brief sample segment of EEG data (200 samples/second).
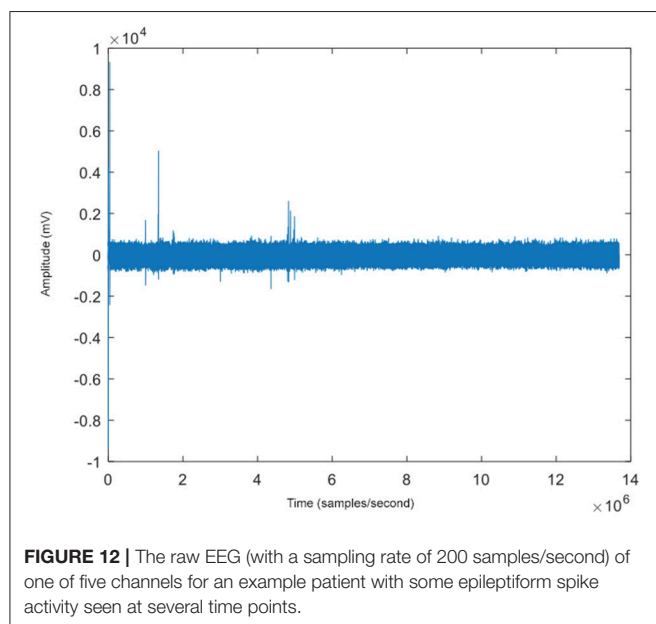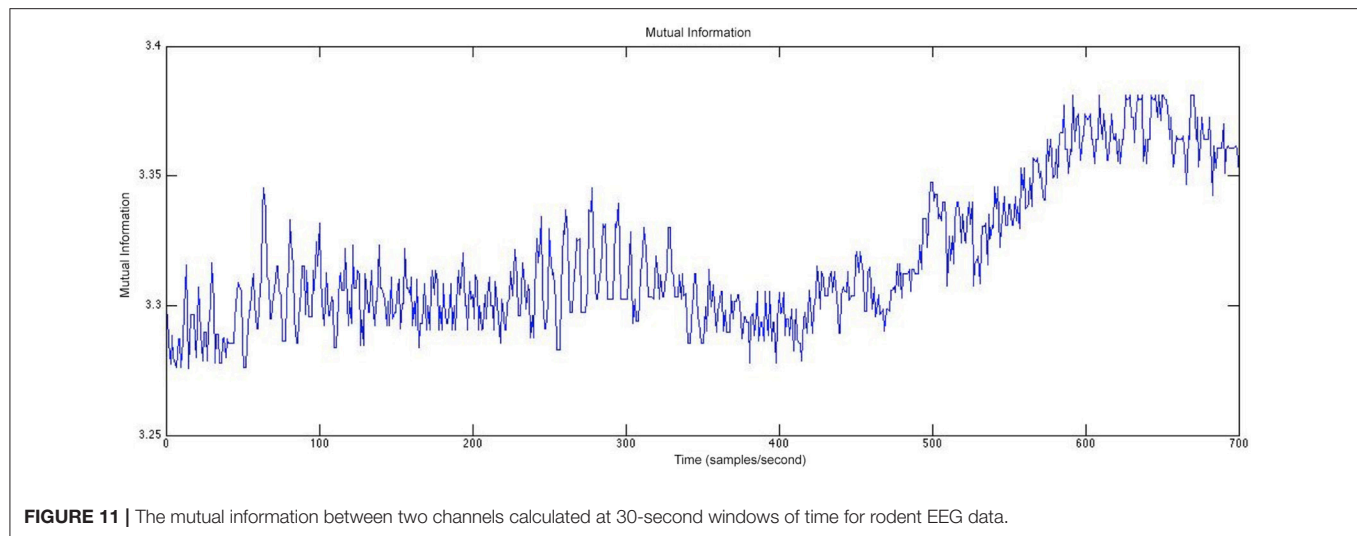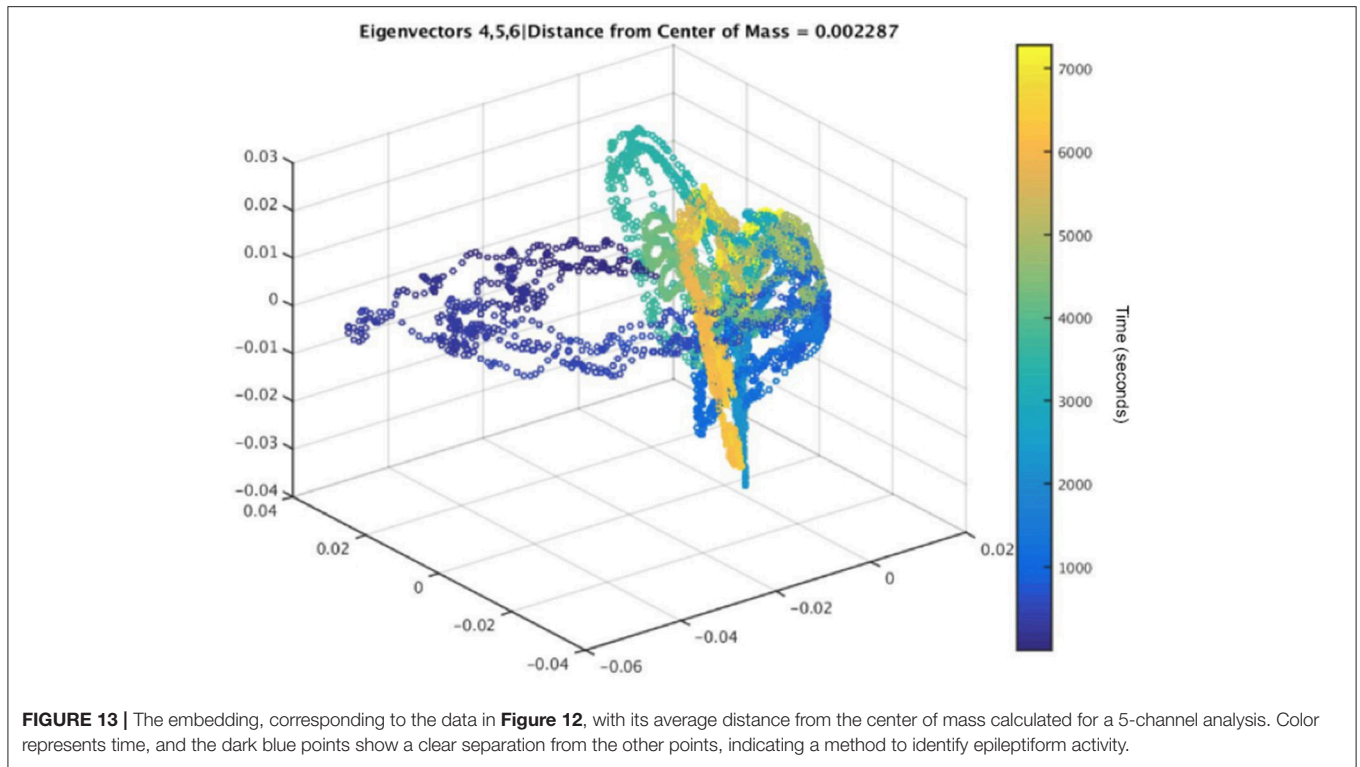
FIGURE 11 | The mutual information between two channels calculated at 30-second windows of time for rodent EEG data.

FIGURE 12 | The raw EEG (with a sampling rate of 200 samples/second) of one of five channels for an example patient with some epileptiform spike activity seen at several time points.

Figure 11 shows an example of the mutual information between two channels of rodent EEG. The mutual information between the two channels was calculated for each consecutive 30-second window and plotted to visualize the relationship between the two channels located in different parts of the brain. This analysis allows us to study how this relationship changes both over time and closer to the occurrence of a seizure, which enables the study of networks in the brain and if those play a role in post-traumatic epileptogenesis. In **Figure 11**, we see a greater relationship between the two electrode contacts chosen for the analysis over time and closer to the seizure onset. Furthermore, we can compare these networks in rats and humans to determine the extent of their similarities.

Dimensionality Reduction

Besides analyzing EEG using spectral analysis, spike detection, and measures of relationship, we can also use dimensionality reduction techniques to analyse the data more extensively and classify epileptiform activity. The EEG amounts to a very large dataset due to the continuous long-term recordings over many electrode contacts. All 300 patients receive 24 h continuous EEG (cEEG) for 72 h minimum during the first 7 days after TBI. Scalp cEEG monitoring is performed using a 16–21 channel bipolar and referential composite montage implemented at each study center based on their established ICU EEG protocols. A subset of 100 patients receive additional depth EEG monitoring during the first 7 days after TBI for higher resolution and pathologic HFOs or repetitive HFOs and spikes detection. Furthermore, we have continuous EEG recordings over 6 months from many cohorts of animals (Duncan et al., 2018b).

An algorithm that we have developed, UDCA (Duncan and Strohmer, 2016; Duncan et al., 2018a), is an extension of diffusion maps (Coifman and Lafon, 2006) and used to reduce the dimensionality of this large amount of data as well as identify patterns in the data that may predict post-traumatic epileptogenesis.

The steps of this algorithm, UDCA, have been previously described (Duncan et al., 2018a); here we briefly explain the steps. The original, raw EEG data matrix (of any number of electrode contacts and any length of time), for example, **Figure 12**, is divided into smaller submatrices that are overlapped by 50% for smoothing purposes. First, the cross-correlation between segments is calculated to ensure minimal variance to ensure similar behavior between the channels that were being analyzed. Channels showing similar waveforms would be expected to have decreased covariance. This is applied to all channels used in the analysis (five channels in the example shown in **Figure 13**), after being split into submatrices. The limit is defined as the difference between the window size, the number of data points in the predefined submatrices, and the window length, the number of data points used to define the lag of the cross-correlation.

FIGURE 13 | The embedding, corresponding to the data in **Figure 12**, with its average distance from the center of mass calculated for a 5-channel analysis. Color represents time, and the dark blue points show a clear separation from the other points, indicating a method to identify epileptiform activity.

Then the time-based covariance matrix is calculated from the covariance of the segment vectors. Singular value decomposition (SVD) is then performed on the covariance matrices. The Mahalanobis distance is applied to inverse covariance matrices that are computed using the SVD to identify outliers; the combination of the Mahalanobis distance and inverse covariance matrices has previously been shown to be a successful tool for denoising data (Talmon et al., 2012). The resulting matrices are constructed from the outputs of the SVD by taking the complex conjugate transpose of the product of the unitary matrix, the inverse of the diagonal matrix, and the other unitary matrix.

The next steps of the algorithm involve constructing the kernel, shown in Equation (3)

$$A = \exp\left(\frac{-d}{4 \bullet k_e}\right) \qquad (4)$$

where d is the Mahalanobis distance (Equation 5), and k_e is the Gaussian kernel (value set to 10, based on the spread of the original data points in the raw EEG data matrix) (Duncan and Strohmer, 2016).

$$d = \left[\text{data}_M - \text{data}_m\right] \cdot \sum_{EEG}^{-1} \cdot \left[\text{data}_M - \text{data}_m\right] \qquad (5)$$

in which data_M is the length of the ith row from the metric data matrix, data_m is the i+1 row, and $\sum_{EEG}^{-1}\left[data_M - data_m\right]$ is the inverse covariance matrix (Duncan and Strohmer, 2016; Duncan et al., 2018a).

Construction of the reference kernel is shown below in Equation (5) using the inverse covariance and the natural extension of AA' (Duncan et al., 2013b, 2018a; Duncan and Strohmer, 2016):

$$W_1 = A_1^* A_1 \qquad (6)$$

in which A_1 is the quotient of A divided element-wise by a repeat matrix of the square root of j_1 with dimensions equal to that of the length of data_M. W = A*A, in which A* is the conjugate transpose, and W is the product of A and its conjugate transpose. Lastly, $j_1 = \sum_i W_{1,i}$ (sum of the elements of W along its columns for row vector) (Duncan et al., 2013b; Duncan and Strohmer, 2016).

$$W_2 = A_2^* A_2 \qquad (7)$$

Additionally, Equation (6) is computed in the same manner as Equation (2), in which A_2 (computed similarly to A_1) with element-wise division by a repeat matrix of the square root of j_2.

The computation of the eigenvectors Equation (7) is performed on W_2, extracting the eigenvalues in a diagonal matrix V and the eigenvectors in a matrix E, corresponding to the eigenvalues, such that:

$$EV = W_2 V \qquad (8)$$

The corresponding eigenvectors are then sorted in a descending order (E_{srt}, V_{srt}). Corresponding point clouds are calculated from Equation (8):

$$V_{clds} = DV_{srt} \qquad (9)$$

in which D is a sparse n x n matrix with the dimensions equal to the length of data$_M$, with values consisting of the square root of one divided-by j$_2$.

Extraction of the two largest eigenvectors was performed according to Equations (9, 10):

$$\varphi_1 = V_{clds_{i,1}} \quad \varphi_2 = V_{clds_{i,2}} \tag{10}$$

Computation of the extension utilized (Equation 11):

$$\omega = \sum_i A_{2_i} \tag{11}$$

in which the column vector ω is the column-wise sum of A$_2$.

Additionally, A$_2$-norm ($\|A_2\|$) is calculated by element-wise division of A$_2$ by a repeat matrix consisting of values from ω, with dimensions equal to that of data$_m$.

$$\hat{\psi} = \frac{\|A_2\| \, V_{srt_i}}{\sqrt{E_{srt_{i+1}}}} \tag{12}$$

Furthermore, $\hat{\psi}$ (Equation 12), is calculated to be the product of $\|A_2\|$ and V_{srt_i} divided element-wise by the square root of the i-th + 1 value of E_{srt}.

Additionally, ψ (initialized as an empty array) is:

$$\psi_i = \psi \hat{\psi} \tag{13}$$

Extended eigenvector extraction corresponding to the two largest eigenvalues (Equations 14, 15):

$$\psi_1 = \psi_{i,1} \tag{14}$$
$$\psi_2 = \psi_{i,2} \tag{15}$$

in which ψ_1 and ψ_2 are tabulated using all values from the rows and columns one and two, respectively.

Preliminary Results Using UDCA

All possible combinations of 3 eigenvectors are used to create the 3D embeddings. Three dimensions were chosen due to this number of dimensions being optimal for visualization, but any number can be chosen and then determined which number of dimensions results in the most important information about the underlying brain activity being extracted, depending on the original data. Embeddings that contained the first eigenvector were excluded due to the normalization that occurs as a result of the SVD analysis (Duncan and Strohmer, 2016). Furthermore, some preliminary results indicated that the embeddings that showed a more diffused spread of points with outliers could be used to indicate preseizure activity in the subject. The determination of the spread for each embedding was calculated by finding each embedded point's Euclidean distance from the center of mass of the embedded points. Embeddings with the largest mean Euclidean distance for each subject were used for preseizure activity evaluation. This method of determining the

optimal embedding allows the algorithm to be automatic and unsupervised, but the algorithm can also be used in a semi-supervised manner as well.

The dark blue points in **Figure 13** represent the time farthest from the seizure in the selected epoch, while the yellow points represent windows of time that are closest to the occurrence of the seizure. **Figure 12** shows an example subject with EEG data from channel 4, in a 5-channel analysis, in which epileptiform spike activity is apparent at several initial time points. The outliers in the embedding shown could be used to correspond with several of the epileptiform spikes in the raw EEG data.

UDCA is a promising method that can be used to detect epileptiform activity that may be a predictor of post-traumatic epileptogenesis. Quantitatively, the evaluation of each embedding can be performed through a variety of methods, such as evaluating the diffusivity in the embedding by calculating the Euclidean distance of each point in the embedding to either the origin or the center of mass of all embedded points or by setting a threshold for the outlier points.

DISCUSSION

We have described some of our analytic tools, including lesion mapping, tractography, PVS analysis, and various types of EEG analysis, including spectral analysis, spike detection, mutual information, and Unsupervised Diffusion Component Analysis, that we are developing and using to analyze the rich, multimodal data from different sites that are collecting data for EpiBioS4Rx. Furthermore, the tools applied to imaging and EEG data are used for both human and animal data so that we can first analyze them separately and then compare the animal model to the human data to determine what translational components exist.

With tractography, we plan to explore the use of a study-specific template that may improve registration performance. We also plan to use the lesion mapping obtained from FLAIR to add lesion statistics to the array of obtained fiber bundle metrics. Based on our analysis of PVS, our results show that PVS may be a potential non-invasive neuroimaging biomarker of post-traumatic epileptogenesis. Moreover, PVS structural analysis combined with DTI analysis can help define the suspected seizure onset area. Ultimately, these results may be of benefit for the design of future clinical trials and for the evaluation of new possible therapeutic targets.

We plan to analyze the EEG using mutual information and compare those results with the resting state fMRI data to study networks in the brain, how they change over time, and how they differ between PTE and non-PTE. With UDCA, our goal is to apply advanced statistical tools to the results of the embeddings to reliably identify epileptiform and preseizure activity in the EEG of humans and rodents.

CONCLUSIONS

As more data are collected in EpiBioS4Rx, we will continue to extract features from neuroimaging and electrophysiologic data as well as molecular, clinical, cognitive, and behavioral measures

to identify candidate diagnostic biomarkers of epileptogenesis. When we apply these methods to new data, we will be able to modify and improve them so that they can be even more effective in our search for biomarkers of epileptogenesis after TBI. Our methods will be used to reveal processes, regions, and stages in epileptogenesis correlated with specific anatomical changes in imaging and changes in the electrical activity in the brain. Furthermore, our tools will allow us and other researchers to easily compare human and animal data to identify their similarities and differences. Innovative statistical techniques will be used to build models of epileptogenesis to predict the probability of developing epilepsy based on biomarker inputs.

AUTHOR CONTRIBUTIONS

DD took the lead in writing the manuscript with input from all authors, performed the EEG analysis, and developed one of the methods, UDCA. GB performed the PVS calculations, analysis,

and interpretation. RC performed the tractography and diffusion MRI analysis as well as the interpretation. FS performed the lesion mapping analysis and interpretation. RG completed the literature review. AB assisted with the EEG analysis. PV collected the human data and assisted with questions relating to the human data. AP collected the rodent data and assisted with questions relating to the rodent data. ML supervised the PVS analysis. AT assisted with data storage issues, supervising all analysis, and directing the project with DD. All authors discussed the results, provided critical feedback, and contributed to the final manuscript.

ACKNOWLEDGMENTS

REFERENCES

Abiega, O., Beccari, S., Diaz-Aparicio, I., Nadjar, A., Layé, S., Leyrolle, Q., et al. (2016). Neuronal hyperactivity disturbs ATP microgradients, impairs microglial motility, and reduces phagocytic receptor expression triggering apoptosis/microglial phagocytosis uncoupling. *PLoS Biol.* 14, 1–48. doi: 10.1371/journal.pbio.1002466

Asgari, M., De Zélicourt, D., and Kurtcuoglu, V. (2016). Glymphatic solute transport does not require bulk flow. *Sci. Rep.* 6:38635. doi: 10.1038/srep38635

Bechmann, I., Priller, J., Kovac, A., Böntert, M., Wehner, T., Klett, F. F., et al. (2001). Immune surveillance of mouse brain perivascular spaces by blood-borne macrophages. *Eur. J. Neurosci.* 14, 1651–1658. doi: 10.1046/j.0953-816X.2001.01793.x

Bendlin, B. B., Ries, M. L., Lazar, M., Alexander, A. L., Dempsey, R. J., Rowley, H. A., et al. (2008). Longitudinal changes in patients with traumatic brain injury assessed with diffusion-tensor and volumetric imaging. *Neuroimage* 42, 503–514. doi: 10.1016/j.neuroimage.2008.04.254

Bigler, E. (2001). The lesion(s) in traumatic brain injury: implications for clinical neuropsychology. *Arch. Clin. Neuropsychol.* 16, 95–131. doi: 10.1016/S0887-6177(00)00095-0

Bragin Engel J., Wilson, C. L., Fried, I., and Buzsaki, G., A. (1999). High-frequency oscillations in human brain. *Hippocampus* 9, 137–142.

Cabeen, R. P., Bastin, M. E., and Laidlaw, D. H. (2016). Kernel regression estimation of fiber orientation mixtures in diffusion MRI. *Neuroimage* 127, 158–172. doi: 10.1016/j.neuroimage.2015.11.061

Cabeen, R. P., Bastin, M. E., and Laidlaw, D. H. (2017). A Comparative evaluation of voxel-based spatial mapping in diffusion tensor imaging. *Neuroimage* 146, 100–112. doi: 10.1016/j.neuroimage.2016.11.020

Cabeen, R. P., Laidlaw, D. H., and Toga, A. W. (2018). "Quantitative imaging toolkit: software for interactive 3D visualization, data exploration, and computational analysis of neuroimaging datasets," in *Proceedings* of the *International Society* for *Magnetic Resonance* in *Medicine* (Paris), 8882.

Catani, M., and Thiebaut de Schotten, M. (2008). A diffusion tensor imaging tractography atlas for virtual *in vivo* dissections. *Cortex* 44, 1105–1132. doi: 10.1016/j.cortex.2008.05.004

Coifman, R. R., and Lafon, S. (2006). Diffusion maps. *Appl. Comput. Harmon. Anal.* 21, 5–30. doi: 10.1016/j.acha.2006.04.006

Corraliza, I. (2014). Recruiting specialized macrophages across the borders to restore brain functions. *Front. Cell. Neurosci.* 8:262. doi: 10.3389/fncel.2014.00262

Dalessandro, R., Ferrara, R., Benassi, G., Lenzi, P. L., and Sabattini, L. (1988). Computed tomographic scans in posttraumatic epilepsy. *Arch. Neurol.* 45, 42–43. doi: 10.1001/archneur.1988.00520250048019

D'Alessandro, R., Tinuper, P., Ferrara, R., Cortelli, P., Pazzaglia, P., Sabattini, L., et al. (1982). CT scan prediction of late post-traumatic epilepsy. *J. Neurol. Neurosurg. Psychiatry* 45, 1153–1155. doi: 10.1136/jnnp.45.12.1153

Delorme, A., and Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Duncan, D., Duckrow, R. B., Pincus, S. M., Goncharova, I., Hirsch, L. J., Spencer, D. D., et al. (2013a). Intracranial EEG evaluation of relationship within a resting state network. *Clin. Neurophysiol.* 124, 1943–1951. doi: 10.1016/j.clinph.2013.03.028

Duncan, D., and Strohmer, T. (2016). Classification of alzheimer's disease using nonlinear independent component analysis. *Math Biosci Eng.* 13, 1119–1130. doi: 10.3934/mbe.2016033

Duncan, D., Talmon, R., Zaveri, H. P., and Coifman, R. R. (2013b). Identifying preseizure state in intracranial EEG data using diffusion kernels. *Math. Biosci. Eng.* 10, 579–590. doi: 10.3934/mbe.2013.10.579

Duncan, D., Toga, A. W., and Vespa, P. M. (2018a). Detecting features of epileptogenesis in EEG after TBI using unsupervised diffusion component analysis. *Discrete Continuous Dyn Syst Ser B.* 23, 161–172. doi: 10.3934/dcdsb.2018010

Duncan, D., Vespa, P., Pitkanen, A., Braimah, A., Lapinlampi, N., and Toga, A. W. (2018b). Big data sharing and analysis to advance research in post-traumatic epilepsy. *Neurobiol. Dis.* doi: 10.1016/j.nbd.2018.05.026. [Epub ahead of print].

Englander, J., Bushnik, T., Duong, T. T., Cifu, D. X., Zafonte, R., Wright, J., et al. (2003). Analyzing risk factors for late posttraumatic seizures: A prospective, multicenter investigation. *Arch. Phys. Med. Rehabil.* 84, 365–373. doi: 10.1053/apmr.2003.50022

Feldman, R. E., Rutland, J. W., Fields, M. C., Marcuse, L. V., Pawha, P. S., Delman, B. N., et al. (2018). Quantification of perivascular spaces at 7 T: A potential MRI biomarker for epilepsy. *Seizure* 54, 11–18. doi: 10.1016/j.seizure.2017.11.004

Gentry, L., Godersky, J., and Thompson, B. (1988). MR imaging of head trauma: review of the distribution and radiopathologic features of traumatic lesions. *Am. J. Roentgenol.* 150, 663–672. doi: 10.2214/ajr.150.3.663

Guo, D., Fridriksson, J., Fillmore, P., Rorden, C., Yu, H., Zheng, K., et al. (2015). Automated lesion detection on MRI scans using combined unsupervised and supervised methods. *BMC Med. Imaging* 15:50. doi: 10.1186/s12880-015-0092-x

Guo, D., Zheng, K., and Wang, S. (2018). "Lesion detection using T1-weighted MRI: a new approach based on functional cortical ROIs," in *Proceedings– International Conference on Image Processing, ICIP* (Athens), 4427–4431.

Gupta, R. K., Saksena, S., Agarwal, A., Hasan, K. M., Husain, M., Gupta, V., et al. (2005). Diffusion tensor imaging in late posttraumatic epilepsy. *Epilepsia* 46, 1465–1471. doi: 10.1111/j.1528-1167.2005.01205.x

Holmin, S., Söderlund, J., Biberfeld, P., and Mathiesen, T. (1998). Intracerebral inflammation after human brain contusion. *Neurosurgery* 42, 291–299. doi: 10.1097/00006123-199802000-00047

Immonen, R., Kharatishvili, I., Gröhn, O., and Pitkänen, A. (2013). MRI Biomarkers for Post-traumatic epileptogenesis. *J. Neurotrauma* 30, 1305–1309. doi: 10.1089/neu.2012.2815

Irimia, A., Chambers, M., and Alger, J. (2011). *Comparison of Acute and Chronic Traumatic Brain Injury Using Semi-Automatic Multimodal Segmentation of MR Volumes. J. 28 VN-r.* Available online at: http://online.liebertpub.com/doi/abs/10.1089/neu.2011.1920.

Irimia, A., Goh, S. Y. M., Torgerson, C. M., Vespa, P. M., and Van Horn, J. D. (2014). Structural and connectomic neuroimaging for the personalized study of longitudinal alterations in cortical shape, thickness and connectivity after traumatic brain injury. *J. Neurosurg. Sci.* 58, 129–144.

Jacobs, J., Staba, R., Asano, E., Otsubo, H., Wu, J. Y., Zijlmans, M., et al. (2012). High-frequency oscillations (HFOs) in clinical epilepsy. *Prog. Neurobiol.* 98, 302–315. doi: 10.1016/j.pneurobio.2012.03.001

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Kharatishvili, I., Immonen, R., Gröhn, O., and Pitkänen, A. (2007). Quantitative diffusion MRI of hippocampus as a surrogate marker for post-traumatic epileptogenesis. *Brain* 130, 3155–3168. doi: 10.1093/brain/awm268

Kim, J., Avants, B., Patel, S., Whyte, J., Coslett, B. H., Pluta, J., et al. (2008). Structural consequences of diffuse traumatic brain injury: A large deformation tensor-based morphometry study. *Neuroimage* 39, 1014–1026. doi: 10.1016/j.neuroimage.2007.10.005

Kraus, MF, Susmaras, T., BP, C., CJ, W., JA, S., and DM, L., (2007). White matter integrity and cognition in chronic traumatic brain injury. a Diffus. *Tensor Imaging Study* 130, 2508–2519. doi: 10.1093/brain/awm216

Labate, A., Cerasa, A., Gambardella, A., Aguglia, U., and Quattrone, A. (2008). Hippocampal and thalamic atrophy in mild temporal lobe epilepsy: a VBM study. *Neurology* 71, 1094–1101. doi: 10.1212/01.wnl.0000326898.05099.04

Labate, A., Cherubini, A., Tripepi, G., Mumoli, L., Ferlazzo, E., Aguglia, U., et al. (2015). White matter abnormalities differentiate severe from benign temporal lobe epilepsy. *Epilepsia* 56, 1109–1116. doi: 10.1111/epi.13027

Liang, S. F., Wang, H. C., and Chang, W. L. (2010). Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection. *EURASIP J. Adv. Signal Process* 2010: 853434. doi: 10.1155/2010/853434

Mishra, A. M., Bai, H., Gribizis, A., and Blumenfeld, H. (2011). Neuroimaging biomarkers of epileptogenesis. *Neurosci. Lett.* 497, 194–204. doi: 10.1016/j.neulet.2011.01.076

Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., et al. (2008). Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. *Neuroimage* 40, 570–582. doi: 10.1016/j.neuroimage.2007.12.035

Narayana, P. A. (2017). White matter changes in patients with mild traumatic brain injury: MRI perspective. *Concussion* 2:CNC35. doi: 10.2217/cnc-2016-0028

Palacios, E. M., Yuh, E. L., Chang, Y.-S., Yue, J. K., Schnyer, D. M., Okonkwo, D. O., et al. (2017). Resting-state functional connectivity alterations associated with six-month outcomes in mild traumatic brain injury. *J. Neurotrauma* 34, 1546–1557. doi: 10.1089/neu.2016.4752

Perucca, P., Smith, G., Santana-Gomez, C., Bragin, A., and Staba, R. (2018). Electrophysiological biomarkers of epileptogenicity after traumatic brain injury. *Neurobiol. Dis.* doi: 10.1016/j.nbd.2018.06.002. [Epub ahead of print].

Pitkänen, A., Ekolle Ndode-Ekane, X., Lapinlampi, N., and Puhakka, N. (2018). Epilepsy biomarkers – toward etiology and pathology specificity. *Neurobiol. Dis.* S0969-9961(18)30145-1. doi: 10.1016/j.nbd.2018.05.007. [Epub ahead of print].

Pustina, D., Coslett, H. B., Turkeltaub, P. E., Tustison, N., Schwartz, M. F., and Avants, B. (2016). Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis. *Hum. Brain Mapp.* 37, 1405–1421. doi: 10.1002/hbm.23110

Ratib, O., and Rosset, A. (2006). Open-source software in medical imaging: development of osiriX. *Int. J. Comput. Assist. Radiol. Surg.* 1, 187–196. doi: 10.1007/s11548-006-0056-2

Reid, A. Y., Bragin, A., Giza, C. C., Staba, R. J., and Engel, J. (2016). The progression of electrophysiologic abnormalities during epileptogenesis after experimental traumatic brain injury. *Epilepsia* 57, 1558–1567. doi: 10.1111/epi.13486

Shultz, S. R., Cardamone, L., Liu, Y. R., Edward Hogan, R., MacCotta, L., Wright, D. K., et al. (2013). Can structural or functional changes following traumatic brain injury in the rat predict epileptic outcome? *Epilepsia* 54, 1240–1250. doi: 10.1111/epi.12223

Sidaros, A., Engberg, A. W., Sidaros, K., Liptrot, M. G., Herning, M., Petersen, P., et al. (2008). Diffusion tensor imaging during recovery from severe traumatic brain injury and relation to clinical outcome: a longitudinal study. *Brain* 131, 559–572. doi: 10.1093/brain/awm294

Sierra-Marcos, A., Scheuer, M. L., and Rossetti, A. O. (2015). Seizure detection with automated EEG analysis: a validation study focusing on periodic patterns. *Clin. Neurophysiol.* 126, 456–462. doi: 10.1016/j.clinph.2014.06.025

Staba, R. J., Stead, M., and Worrell, G. A. (2014). Electrophysiological biomarkers of epilepsy. *Neurotherapeutics* 11, 334–346. doi: 10.1007/s13311-014-0259-0

Talmon, R., Kushnir, D., Coifman, R. R., Cohen, I., and Gannot, S. (2012). Parametrization of linear systems using diffusion kernels. *IEEE Trans. Signal Process.* 60, 1159–1173. doi: 10.1109/TSP.2011.2177973

Taoka, T., Masutani, Y., Kawai, H., Nakane, T., Matsuoka, K., Yasuno, F., et al. (2017). Evaluation of glymphatic system activity with the diffusion MR technique: diffusion tensor image analysis along the perivascular space (DTI-ALPS) in Alzheimer's disease cases. *Jpn. J. Radiol.* 35, 172–178. doi: 10.1007/s11604-017-0617-z

Tubi, M., Lutkenhoff, E., Blanco, M. B., McArthur, D., Villablanca, P., Ellingson, B., et al. (2018). Early seizures and temporal lobe trauma predict post-traumatic epilepsy: A longitudinal study. *Neurobiol. Dis.* S0969-9961(18)30152-9. doi: 10.1016/j.nbd.2018.05.014. [Epub ahead of print].

van Vliet, E. A., Dedeurwaerdere, S., Cole, A. J., Friedman, A., Koepp, M. J., Potschka, H., et al. (2017). WONOEP appraisal: imaging biomarkers in epilepsy. *Epilepsia* 58, 315–330. doi: 10.1111/epi.13621

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., et al. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 36, 630–644. doi: 10.1016/j.neuroimage.2007.02.049

Wetter, N. C., Hubbard, E. A., Motl, R. W., and Sutton, B. P. (2016). Fully automated open-source lesion mapping of T2-FLAIR images with FSL correlates with clinical disability in MS. *Brain Behav.* 6, 1–9. doi: 10.1002/brb3.440

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., et al. (2009). Bayesian analysis of neuroimaging data in FSL. *Neuroimage* 45, (Suppl. 1):S173–86. doi: 10.1016/j.neuroimage.2008.10.055

Worrell, G. A. (2011). Sensing the body electric: biomarkers of epileptic brain. *Epilepsy Curr.* 11, 118–119. doi: 10.5698/1535-7511-11.4.118

Yuh, E., Cooper, S., Mukherjee, P., Yue, J., Lingsma, H., Gordon, W., et al. (2014). Diffusion tensor imaging for outcome prediction in mild traumatic brain injury: a TRACK-TBI study. *J. Neurotrauma* 31, 1457–1477. doi: 10.1089/neu.2013.3171

Zhang, H., Yushkevich, P. A., Alexander, D. C., and Gee, J. C. (2006). Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Med. Image Anal.* 10, 764–785. doi: 10.1016/j.media.2006.06.004

Zhang, S., Peng, H., Dawe, R. J., and Arfanakis, K. (2011). Enhanced ICBM diffusion tensor template of the human brain. *Neuroimage* 54, 974–984. doi: 10.1016/j.neuroimage.2010.09.008

Zijlmans, M., Jiruska, P., Zelmann, R., Leijten, F. S. S., Jefferys, J. G. R., and Gotman, J. (2012). High-frequency oscillations as a new biomarker in epilepsy. *Ann. Neurol.* 71, 169–178. doi: 10.1002/ana.22548

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership