A Consensus Network of Gene Regulatory Factors in the Human Frontal Lobe

Stefano Berto [1,2,3]*, Alvaro Perdomo-Sabogal [1], Daniel Gerighausen [1], Jing Qin [4,5] and Katja Nowick [1,2]*

[1] Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University Leipzig, Leipzig, Germany, [2] Paul-Flechsig Institute for Brain Research, University of Leipzig, Leipzig, Germany, [3] Department of Neuroscience, University of Texas Southwestern Medical Center, Dallas, TX, USA, [4] Department of Mathematics and Computer Sciences, University of Southern Denmark, Odense, Denmark, [5] Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria

Cognitive abilities, such as memory, learning, language, problem solving, and planning, involve the frontal lobe and other brain areas. Not much is known yet about the molecular basis of cognitive abilities, but it seems clear that cognitive abilities are determined by the interplay of many genes. One approach for analyzing the genetic networks involved in cognitive functions is to study the coexpression networks of genes with known importance for proper cognitive functions, such as genes that have been associated with cognitive disorders like intellectual disability (ID) or autism spectrum disorders (ASD). Because many of these genes are gene regulatory factors (GRFs) we aimed to provide insights into the gene regulatory networks active in the human frontal lobe. Using genome wide human frontal lobe expression data from 10 independent data sets, we first derived 10 individual coexpression networks for all GRFs including their potential target genes. We observed a high level of variability among these 10 independently derived networks, pointing out that relying on results from a single study can only provide limited biological insights. To instead focus on the most confident information from these 10 networks we developed a method for integrating such independently derived networks into a consensus network. This consensus network revealed robust GRF interactions that are conserved across the frontal lobes of different healthy human individuals. Within this network, we detected a strong central module that is enriched for 166 GRFs known to be involved in brain development and/or cognitive disorders. Interestingly, several hubs of the consensus network encode for GRFs that have not yet been associated with brain functions. Their central role in the network suggests them as excellent new candidates for playing an essential role in the regulatory network of the human frontal lobe, which should be investigated in future studies.

Keywords: transcription factor, coexpression network, weighted topological overlap network, consensus network, cognitive abilities, cognitive disorders, prefrontal cortex (PFC)

INTRODUCTION

Broadly defined, cognition refers to the biological mechanisms through which animals perceive, learn and memorize information from the environment and decide to act upon them (Shettleworth, 2009). In humans, cognitive processes such as use of language, social behavior, and decision-making have been attributed to the frontal lobe (Duncan et al., 1996; Chayer and Freedman, 2001). However, the actual molecular mechanisms that underlie these morphological changes are still not well understood.

Candidate genes that are involved in the molecular mechanisms of cognition can be identified through biomedical studies on cognitive disorders. For example, causative mutations point to the genes that should in their wild-type variants be important for providing for healthy cognitive abilities. Research on cognitive disorders such as Alzheimer's disease (AD; Bullido et al., 1998), intellectual disability (ID; Kaufman et al., 2010), autism spectrum disorder (ASD; Bailey et al., 1996; Voineagu et al., 2011; Berg and Geschwind, 2012; Ecker et al., 2012), schizophrenia (SZ; Andreasen, 1995), circadian rhythm and bipolar disorder (BD; Akula et al., 2014, 2016; Takahashi, 2015), Parkinson's disease (PD; Polymeropoulos, 2000), and several syndromes or disorders associated with ID or cognitive impairment (SY; Greydanus and Pratt, 2005) has thus already identified several candidate genes involved in cognition. Importantly, these studies also revealed that most cognitive disorders are complex and phenotypically and genetically heterogeneous (Sebat et al., 2007; Tsankova et al., 2007; Voineagu et al., 2011; Weyn-Vanhentenryck et al., 2014), thus creating challenges for studying these disorders.

Transcriptome and network analyses bear great potential for overcoming some of these challenges and uncovering the genetic interactions and molecular mechanisms causing such complex disorders. For example, recent studies have used network approaches to identify coexpressed ASD and ID modules implicated in synaptic development, chromatin remodeling and early transcriptional regulation (Parikshak et al., 2013; Willsey et al., 2013; De Rubeis et al., 2014). However, coexpression networks can have many false positive inferences. One way to reduce the effect of false positives is to calculate weighted topological overlap (wTO) networks (Zhang and Horvath, 2005; Nowick et al., 2009). Another drawback is that most network studies so far have only analyzed data from one dataset. However, it is unclear how variable independently derived networks are and depend, for instance, on the technical platform or on the particular samples/individuals that were used to produce the dataset. We thus analyzed and compared here 10 different transcriptome datasets from individual human frontal lobe samples, which have been produced with different platforms (microarrays and RNA-Seq), and developed a method for integrating the coexpression wTO networks calculated from them into one consensus network of high confidence level.

Several reasons prompted us to especially focus on the role of gene regulatory factors (GRFs) in the consensus network of the frontal lobe. First, because GRFs regulate the expression of many genes, they are expected to be among the most important players in these networks and might provide important insights about the molecular mechanisms taking place in this tissue. Second, primate specific zinc finger genes with a *Krüppel*-associated box (KRAB-ZNFs) are also enriched among the genes expressed during frontal lobe development (Zhang et al., 2011), which leads to the hypothesis that at least some GRFs might contribute to human specific cognitive abilities. Third, we show here that GRFs are enriched among the candidate genes for ID and ASD, thus suggesting an important role of GRFs in the gene regulatory processes and circuitry of such cognitive disorders. Taken together, GRFs are thus good candidates for providing essential information about the molecular mechanisms that set the stage for cognition.

To identify and analyze GRF proteins with potential implications in cognition in more detail, we used our in-house list of all 3315 human GRFs (Perdomo-Sabogal et al., under preparation). This catalog includes information from the most relevant studies in the area of human GRF inventories (see Section Materials and Methods), and includes information about proteins involved in different regulatory mechanisms such as DNA-binding proteins, cofactors that associate with transcription factors, histone and chromatin modifiers, among others. We also performed a comprehensive literature survey and compiled a list of 676 GRFs that are known to be important during human brain development or that have been associated with cognitive disorders. We will refer to this set of 676 GRFs as "Brain-GRFs" (Table S1). Using our high-confidence consensus network we identified here several GRFs, including 166 "Brain-GRFs" that are hubs and thus seem to be important for the gene regulatory processes in the human frontal lobe.

MATERIALS AND METHODS

Data Sets

The raw and processed data from microarrays and RNA-Seq were downloaded from ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) and Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/). Microarrays were analyzed using the R programming language and Bioconductor packages (Ihaka and Gentleman, 1996). For the microarrays, we determined gene expression levels (RMA values) and MAS5 detection p-value from the probes using the "affy" and "oligo" package, respectively of the platform used (Gautier et al., 2004; Carvalho and Irizarry, 2010). We considered only the probesets significantly detected in at least one individual ($p < 0.05$). Furthermore, for genes represented by more than one expressed probeset, we calculated the mean of the expression values of all its probesets. For the RNA-Seq data, we used published RPKM values when available (BrainSpan). Otherwise, we processed and analyzed the raw data by mapping of the reads using segemehl (Hoffmann et al., 2009) and calculating RPKM values using R programming language and R libraries such as GenomicRanges, GenomicFeatures, and Rsamtools (Lawrence et al., 2013). All the raw data were mapped to the hg19 genome. All expression values were then filtered for RPKM values > 0.5 for 90% of the samples. All samples were used from the following datasets: FrontalVal

[GSE25219] (Kang et al., 2011), NeoVal [GSE11512] (Somel et al., 2009), KhatVal [SRA028456] (Somel et al., 2011), and GexVal [GSE22521] (Liu et al., 2012). Only the data from the control individuals were selected from the DisVal [GSE53987], BipRval [GSE53239] (Akula et al., 2014), and BipVal [GSE5388] (Ryan et al., 2006) datasets. From the BrainSpan dataset we selected the samples from the frontal lobe regions and subset them such that individuals with same ages (13 total individuals per dataset) were used.

Catalog of Gene Regulatory Factor Proteins

The GRF catalog we used for building our GRFs consensus network of the human frontal lobe was initially built by Perdomo-Sabogal et al. (under preparation). For this catalog the information for 3315 GRF proteins sourced from the most seminal studies in the area of human GRF inventories (Messina et al., 2004; Vaquerizas et al., 2009; Ravasi et al., 2010; Nowick et al., 2011; Corsinotti et al., 2013; Tripathi et al., 2013; Wingender et al., 2013, 2015) that are associated with gene ontology terms for regulation of transcription, DNA-depending transcription, RNA polymerase II transcription cofactor and co-repressor activity, chromatin binding, modification, remodeling, or silencing, among others, were manually curated.

Gene Sets

The ASD gene list was compiled using the SFARI gene database (09/20/2015, 740 genes; Basu et al., 2009; Banerjee-Basu and Packer, 2010). In the analysis, we included all the 740 genes. In addition, we also calculated the overlap between GRFs and ASD genes with strong association with S category (syndromic) and strong evidence (levels 1–4). ASD modules (asdM12 and asdM16) were obtained from an independent genome-wide expression study that compared ASD with healthy post-mortem brain tissues (Voineagu et al., 2011).

GRFs with association with Parkinson's disease, Alzheimer's disease, and Schizophrenia where filtered according to their significant evidence in more than two GWAS studies (Allen et al., 2008; Bertram, 2009; Jia et al., 2010; Lill et al., 2012). Additional schizophrenia GRFs were derived from independent publication with 108 loci implicated in schizophrenia (Consortium SWGotPG, 2014). ID and FMRP targets genes were collected from independent publications (Inlow and Restifo, 2004; Ropers, 2008; Darnell et al., 2011; van Bokhoven, 2011; Lubs et al., 2012; Consortium SWGotPG, 2014).

Other brain related GRFs were manually selected using web sources such as OMIM and independent databases such as SGZR (Hamosh et al., 2005; Jia et al., 2010). We prioritize GRFs that have evidence on brain functions, synaptic transmission, and brain development.

wTO Calculation

Spearman rank correlations were used to correlate the expression values of the GRF genes with the expression values of all genes, separately in each of the 10 datasets. Note that only expressed genes were considered in each dataset and that the number of expressed GRFs and other genes differs between the datasets. We

extracted all significant correlations ($p < 0.05$) for calculating the weighted topological overlap values ($\omega = wTO$) between all pairs of expressed GRF genes for each dataset as previously described (Nowick et al., 2009). The calculation is based on a real symmetric matrix $A = [a_{ij}]$, in which a_{ij} is a real number ranging between -1 to 1 that indicates the correlation coefficient between the i-th and j-th GRF in the dataset. In particular, we have $a_{ii} = 0$. Comparing with the previous method (Zhang and Horvath, 2005), our method incorporates both significant (Spearman rank correlation; $p < 0.05$) positive and negative correlations of two GRFs' correlated gene sets (u) described as follow: $a_{ij} \epsilon [0, 1]$ when $a_{ij} \geq 0 \rightarrow a_{iu}a_{ju} \geq 0$ for all u and $a_{ij} \epsilon [-1, 0]$ when $a_{ij} \leq 0 \rightarrow a_{iu}a_{ju} \leq 0$ for all u. This condition results in a positive wTO value for the GRFs i and j if they are both correlated in the same direction with u, while in a negative wTO value if i and j are correlated with u in the opposite direction.

Inserting the weighted connectivity of a node i as:

$$K_i = \sum_i a_{ij},$$

and the connectivity between i and j as:
$C = A * A^T$, the weighted topological overlap is calculated as:

$$\omega_{ij} = \frac{c_{ij} + a_{ij}}{min\left(K_i, K_j\right) + 1 - |a_{ij}|}$$

To evaluate the reliability of each wTO network, we performed 100 permutations by randomizing the expression values of each individual. This effectively assigned a random expression value to each gene of a particular individual out of all the available gene expression values for that individual. The permutation was done separately for each individual. We then calculated 100 permuted wTO networks for each dataset. We determined the number of links in the empirically derived ("real") network for multiple wTO cutoffs [0.1:0.6] and compared it to the number of links with the same wTO cutoff in the 100 permuted networks. This method allowed us to determine a p-value for how different the empirical networks are from random expectation and to calculate a false positive rate for the links in each network. All empirically derived networks had more links at all tested wTO values compared to the permuted networks, demonstrating that the empirically derived networks are different from random expectation (Table S2D).

Consensus Network Construction

To construct the consensus network, we first analyzed the distributions of the wTO values of all GRF-GRF pairs across all datasets using the boxplot.stats function in R (Williamson et al., 1989) to have an overall view of the data sets. Our results show that the distributions of wTO values of the datasets BipRVal, DisVal, and FrontalVal are different from the other datasets (**Figure 2**). Based on these observations, we chose the Wilcoxon rank sum test for our subsequent analysis, since it is a non-parametric test and hence robust against outliers. Thus, we are able to construct the consensus network by taking all the wTO values from all the datasets into consideration. Furthermore, to identify significant GRF-GRF pairs, we performed another Wilcoxon rank sum test with alternative hypothesis greater than

|wTO| > 0.3. By applying this test, we avoided potential false positive links due to high variation of wTO values across the datasets. If the result was significant ($p < 0.05$), we considered this GRF-GRF pair as a significant pair. For each of these detected significant GRF-GRF pairs, we then calculated its consensus wTO value as the median of all 10 individual wTO values. Note here, we opted for |wTO| > 0.3 as cutoff in the hypothesis, because this was the mean of the cutoffs at which the 10 networks differed from random expectation with $p < 0.01$.

Network Visualization

For network visualization, we used Cytoscape 3.0. Node attributes were used according to our manually curated Brain-GRF list, the Human Proteome map (Kim et al., 2014), and the FMRP targets (Darnell et al., 2011). We included the Cytoscape session (the file is publically available on http://www.nowick-lab.info/?page_id=470) for manual visualization of GRF-GRF interactions as additional file.

Statistics

For gene set enrichments, p-values were calculated with a one-sided Fisher's exact test function in R (alternative = "g," confidence level = 0.99, simulated p-value with 1000 replicates). A one-sided Wilcoxon ranked test was implemented to evaluate the enrichment of the connectivity between species (alternative = "g," confidence level = 0.99, paired = FALSE). P-values for overlaps were calculated with hypergeometric tests using a custom made R script. We retained an independent background (BrainSpan expressed gene = 15585 genes). P-values were subsequently adjusted for multiple comparisons using Benjamini-Hochberg FDR procedure. Two-way permutation test of 1000 was performed to validate the overlaps. First we randomized the external gene sets (e.g., ASD genes) by randomly selecting the same number of genes from an independent brain expressed genes list (e.g., BrainSpan gene set) and subsequently calculating the overlap p-values with the GRF gene set. The second approach randomized the internal gene sets (e.g., GRF gene set) by randomly selecting the same number of genes as GRFs that were expressed and subsequently calculating the overlap p-values. Analysis for RNA-seq, microarray, and correlation filtering were performed using custom made R and SQL scripts. To calculate the correlation and wTO, we developed a Java-based program.

Enrichment for Transcription Factor Binding Sites (TFBS)

For the TFBS enrichment, we focused on the 5421 genes that are expressed in all datasets and correlated with at least one GRF in each of the 10 different datasets. To test whether correlated genes might be target genes of the respective GRF, we performed a ChIP Enrichment Analysis (ChEA) using the ENCODE database and data from Chip-Seq, Chip-Chip, Chip-PET, and DamID experiments (Lachmann et al., 2010). We also performed a TFBS enrichment analysis using the Jolma and JASPAR databases (Jolma et al., 2013; Mathelier et al., 2014). We tested for enrichment of TFBSs included in those databases within the 2 kb upstream region of the 5421 genes using
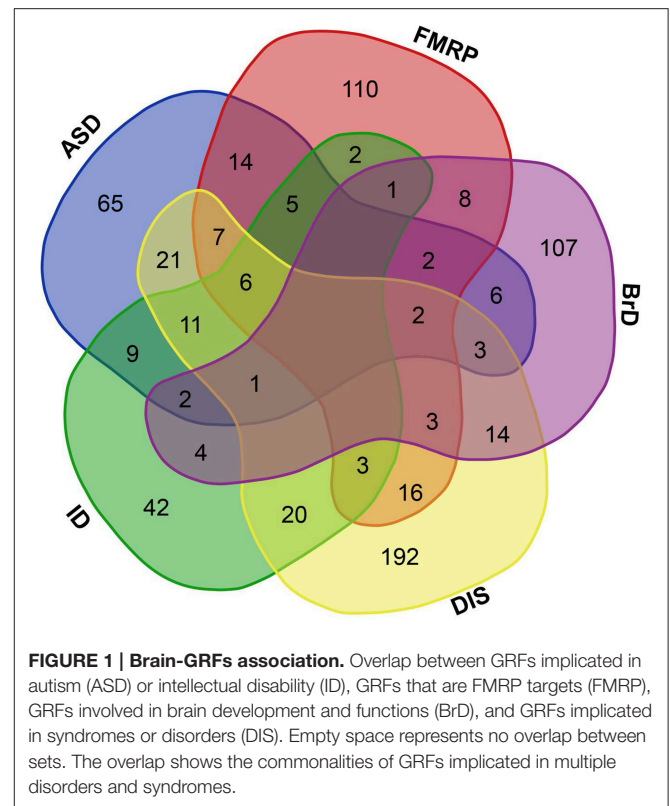

FIGURE 1 | Brain-GRFs association. Overlap between GRFs implicated in autism (ASD) or intellectual disability (ID), GRFs that are FMRP targets (FMRP), GRFs involved in brain development and functions (BrD), and GRFs implicated in syndromes or disorders (DIS). Empty space represents no overlap between sets. The overlap shows the commonalities of GRFs implicated in multiple disorders and syndromes.

CentriMo (default parameters) implemented in the MEME suite (Bailey et al., 2009; Bailey and Machanick, 2012). As background, we used the 2 kb upstream regions of the remaining protein coding genes and CpG islands.

Protein–Protein-Interactions Enrichment

Protein–Protein-Interactions (PPIs) were compiled from BioGRID and InWeb using the method described in Parikshak et al. (2013). We used the set of 5421 genes commonly expressed in all 10 datasets. Then we determined the GRF-gene pairs that were called to interact as proteins according to BioGRID and InWeb (Rossin et al., 2011; Chatr-Aryamontri et al., 2013). GRF-gene pairs that were present in each of the 10 datasets and were indicated to interact as proteins were then combined to a consensus PPI network. Fisher's exact test was used for testing the enrichment of PPI in Brain-GRFs and other GRFs.

GO Enrichment

For the GO enrichment analysis in the consensus network, we first ranked the genes of each dataset according to the number of GRFs they were correlated with. Then we summed up the ranks across the 10 datasets. The ranked list of the sums of the ranks was used as input for the Wilcoxon test implemented in FUNC (Prüfer et al., 2007) for the GO enrichment analysis. This method allowed us to understand the relative importance of a gene in each dataset according to the rank position. We next summarized the ranks across the 10 datasets, thus obtaining a general rank (rank-sum). The GO enrichment test was performed using FUNC

TABLE 1 | Platforms description.

Dataset	Names	Assession number	Sample	Type	Permutation (|wTO|)
BipRVal	*Bipolar RNA-seq Values*	GSE53239	11	Adult	>0.39
BipVal	*Bipolar Microarray Values*	GSE5388	31	Adult	>0.24
DfcVal	*DFC RNA-seq Values*	BrainSpan	13	Developmental	>0.36
DisVal	*Disorder Microarray Values*	GSE53987	19	Adult	>0.30
FrontVal	*Frontal Pole Microarray Values*	GSE25219	348	Developmental	>0.10
GexVal	*Gene Expression Microarray Values*	GSE22521	25	Developmental	>0.25
KhatVal	*RNA-seq Values from a Khaitovich study*	SRA028456	12	Developmental	>0.36
NeoVal	*Neoteny Microarray Values*	GSE11512	44	Developmental	>0.19
OfcVal	*OFC RNA-seq Values*	BrainSpan	13	Developmental	>0.37
VfcVal	*VFC RNA-seq Values*	BrainSpan	13	Developmental	>0.37

We used multiple platforms to uncover the GRF—GRF interactions.
The first column represents the chosen name of each dataset.
The second column showed complete name of the dataset, the used platform, and Values as indication for the wTO calculation.
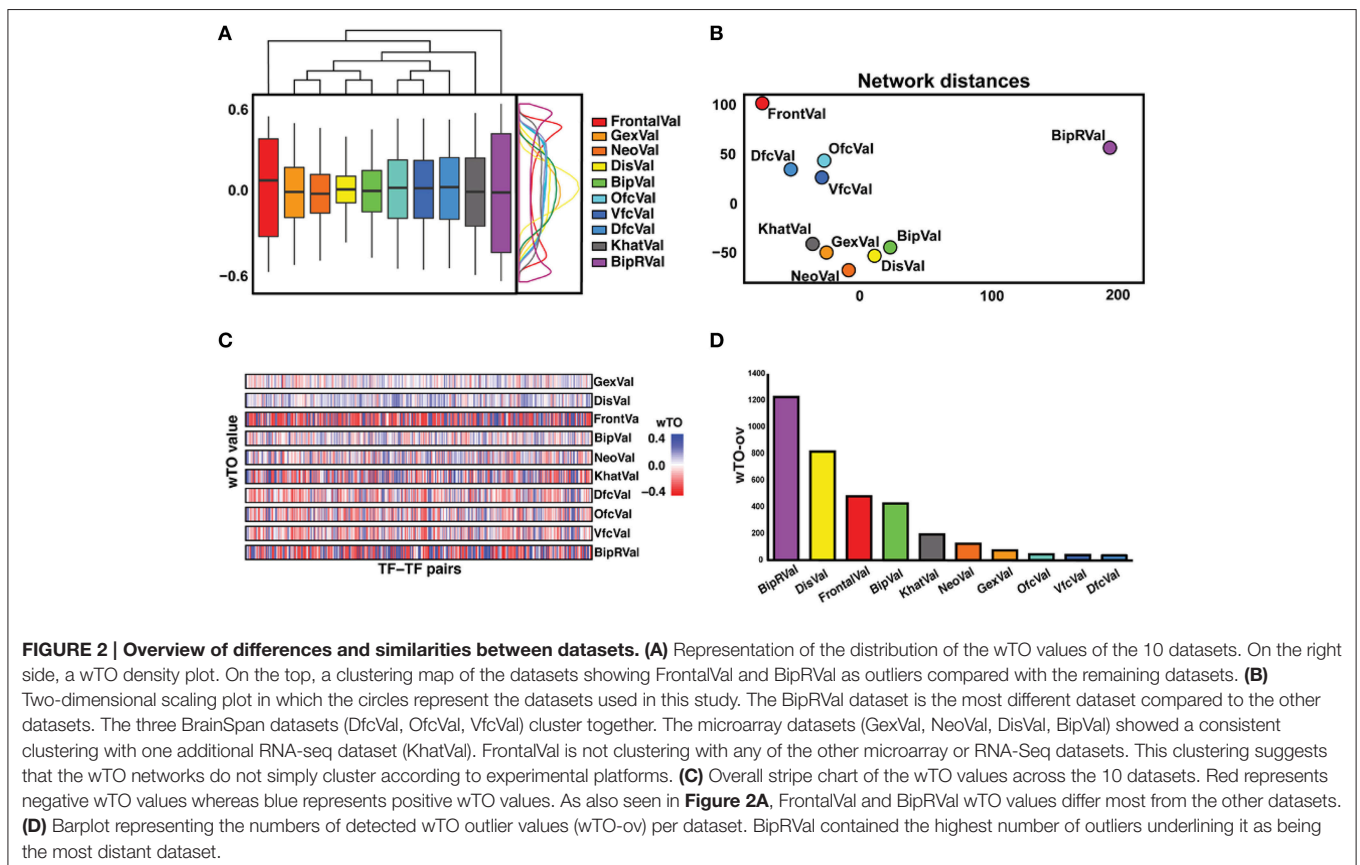*The third column contains the accession numbers of each dataset.*
The fourth column indicates the number of samples used for the analysis.
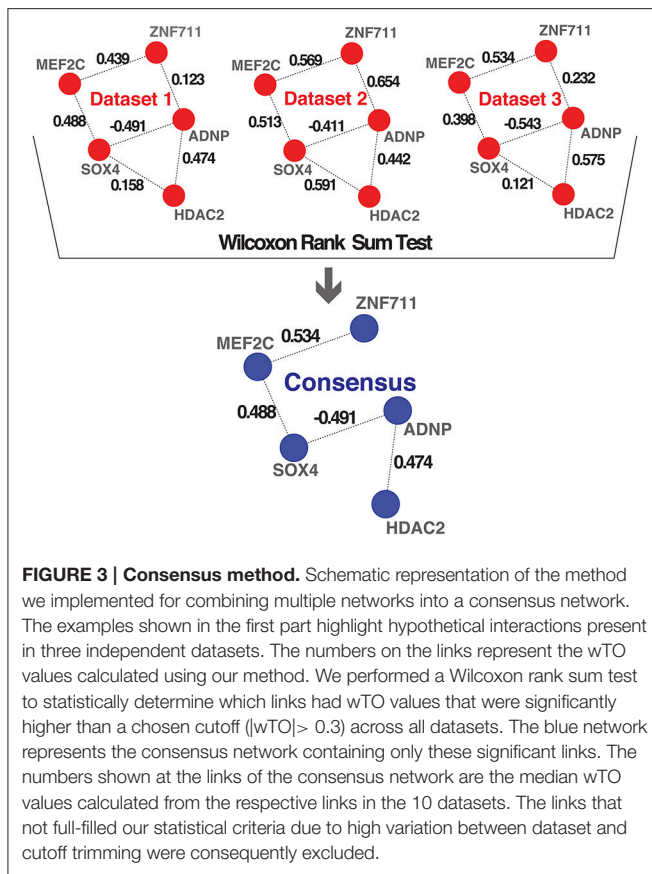*The fifth column indicates the type of dataset.*
The sixth column shows the |wTO| cutoff at which the dataset has significantly more links in the empirical network than in the permuted ones.
For BipRval, BipVal, DisVal, we used only the control samples consisting of healthy individuals (see Materials and methods).

FIGURE 2 | Overview of differences and similarities between datasets. (A) Representation of the distribution of the wTO values of the 10 datasets. On the right side, a wTO density plot. On the top, a clustering map of the datasets showing FrontalVal and BipRVal as outliers compared with the remaining datasets. **(B)** Two-dimensional scaling plot in which the circles represent the datasets used in this study. The BipRVal dataset is the most different dataset compared to the other datasets. The three BrainSpan datasets (DfcVal, OfcVal, VfcVal) cluster together. The microarray datasets (GexVal, NeoVal, DisVal, BipVal) showed a consistent clustering with one additional RNA-seq dataset (KhatVal). FrontalVal is not clustering with any of the other microarray or RNA-Seq datasets. This clustering suggests that the wTO networks do not simply cluster according to experimental platforms. **(C)** Overall stripe chart of the wTO values across the 10 datasets. Red represents negative wTO values whereas blue represents positive wTO values. As also seen in **Figure 2A**, FrontalVal and BipRVal wTO values differ most from the other datasets. **(D)** Barplot representing the numbers of detected wTO outlier values (wTO-ov) per dataset. BipRVal contained the highest number of outliers underlining it as being the most distant dataset.

(Prüfer et al., 2007). We used a Wilcoxon rank-based test for GO enrichment among the genes with highest rank-sums. For the GO analyses we only analyzed GO groups with at least 20 genes per group. We report GO groups with enrichment with $p < 0.01$ before and after refinement.

For the analysis of GO enrichment within each individual network among genes correlated with the selected Brain-GRF hubs we collected for each hub its correlated genes in all the 10 datasets. The remaining set of expressed genes was used as background set. We used the hypergeometric test implemented

FIGURE 3 | Consensus method. Schematic representation of the method we implemented for combining multiple networks into a consensus network. The examples shown in the first part highlight hypothetical interactions present in three independent datasets. The numbers on the links represent the wTO values calculated using our method. We performed a Wilcoxon rank sum test to statistically determine which links had wTO values that were significantly higher than a chosen cutoff (|wTO|> 0.3) across all datasets. The blue network represents the consensus network containing only these significant links. The numbers shown at the links of the consensus network are the median wTO values calculated from the respective links in the 10 datasets. The links that not full-filled our statistical criteria due to high variation between dataset and cutoff trimming were consequently excluded.

in FUNC for the GO enrichment analysis considering only GO groups with at least 20 genes per group. We report GO groups with enrichment with $p < 0.01$ before and after refinement. Finally, we summarized the 10 lists of significant GO categories into one single list, thus removing duplicated GO categories. We also parsed the analyzed GO categories into a list of developmental categories using CateGOrizer (Hu et al., 2008).

RESULTS

Gene Regulatory Factors Involved in Brain Development and Cognitive Disorders

Within this list of human GRFs we identified 676 GRFs that are involved in cognitive functions, brain development, and disorders by using different sources (see Materials and Methods; **Figure 1** and Table S1). A prevalence of genes coding for GRFs among genes associated with some cognitive disorders has been observed before (Hong et al., 2005; West and Greenberg, 2011; Parikshak et al., 2013; De Rubeis et al., 2014; Nord et al., 2015). We here tested if this observation represents a significant overrepresentation of GRF genes among genes implicated in cognitive disorders. Among the 401 genes implicated in ID, we identified 106 genes coding for GRFs, which represents a highly significant enrichment of GRFs among all ID genes (hypergeometric test, $p = 2.03 \times 10^{-07}$). The SFARI database (Basu et al., 2009; Banerjee-Basu and Packer, 2010) currently contains 740 genes implicated in autism. Among those, 297

genes show strong evidence of ASD association. We identified 154 GRFs among the 740 genes (78 among the 297 ASD genes with strong association), which demonstrates that there is also a highly significant overrepresentation of GRFs among genes associated with autism (hypergeometric test, $p = 0.0001$). We further investigated whether GRFs are enriched among the target genes of the Fragile-X Mental Retardation Protein (FMRP). This protein was previously shown to play an important role in ASD-pathways by exerting translational regulation during human brain development (Darnell et al., 2011). Among the set of 842 FMRP target genes predicted by HITs-CLIP, we identified 179 GRF genes revealing a significant overrepresentation of GRF genes (hypergeometric test, $p = 0.0001$). In addition, GRFs are also significantly enriched for genes highly expressed in neurons (hypergeometric test, $p < 0.001$) and astrocytes (hypergeometric test, $p < 0.05$) compared with other brain cell-type expressed genes (Zhang et al., 2014).
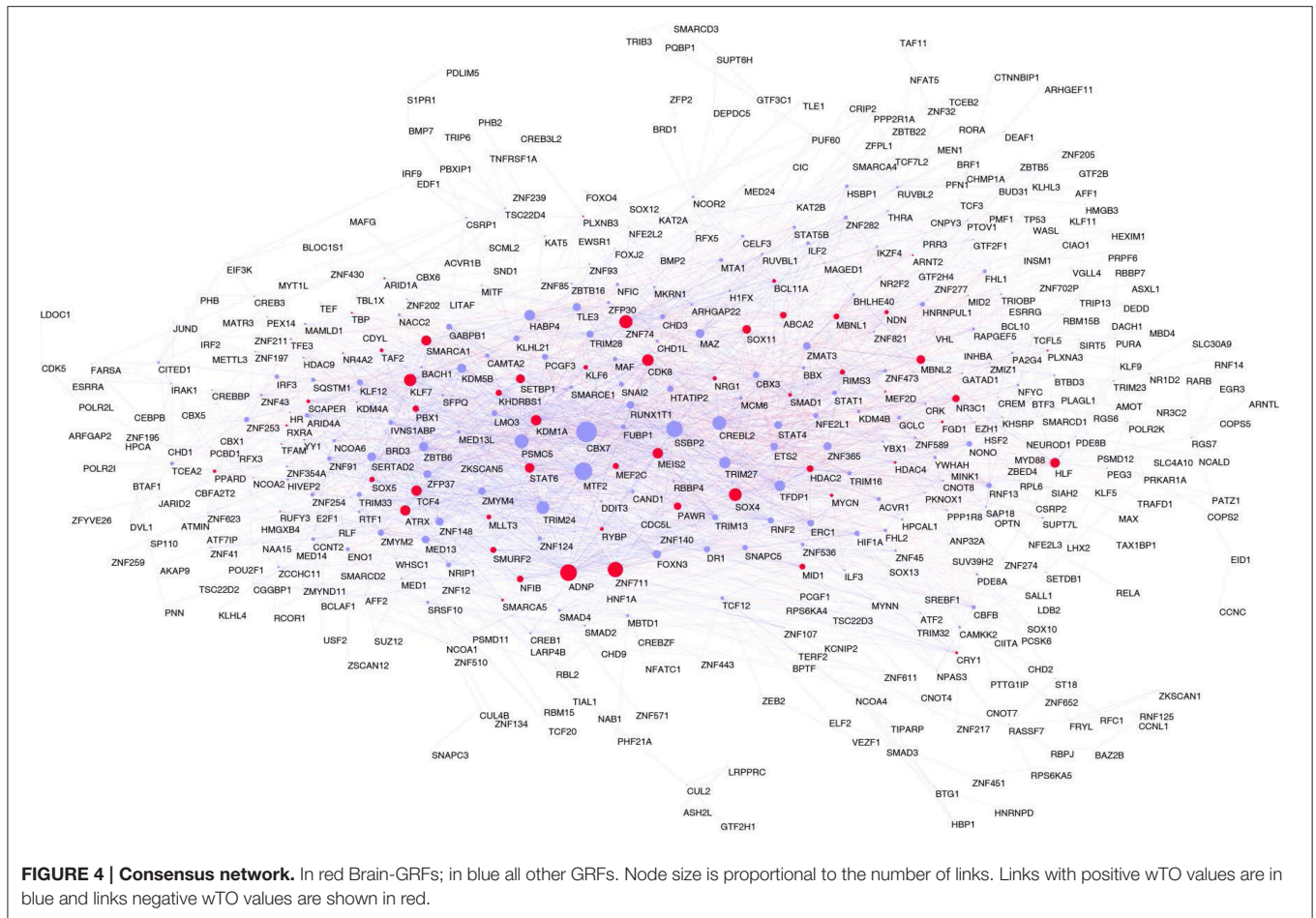
Taken together, these findings show that GRF genes are enriched among candidate genes for cognitive disorders and cell important for brain functions, metabolism, and structure. Therefore, they are likely to be good candidates for providing essential information about the molecular processes involved in the organization and functioning of neural circuits that support healthy cognitive abilities.

A Consensus Network of High Confidence

To investigate the roles of all GRFs in the frontal lobe, we analyzed 10 genome-wide expression datasets comprised of frontal lobe samples from individuals of different ages and obtained with different techniques (**Table 1**). We first analyzed each dataset independently to investigate the consistency of the coexpression networks derived from these independent datasets.

Specifically, from each dataset, we constructed a weighted topological overlap (wTO) network taking into account all expressed GRFs and their coexpressed genes (Nowick et al., 2009). For constructing this wTO network, we first identified all genes that are significantly correlated in expression (i.e., coexpressed) with a particular GRF. These genes include putative target genes and genes coding for interaction partners of that GRF. The wTO of a pair of GRFs then represents the commonality of these two GRFs in their sets of coexpressed genes. Because GRFs can function as activators or repressors of gene expression, we take into account the sign of the correlation when calculating the wTO. Pairs of GRFs with |wTO| values above a certain cutoff are connected by a link in the wTO network visualization (see Materials and Methods).

Even though each network is supported to significantly differ from random expectation, we noted differences between the 10 networks, for instance, in the distribution of the wTO values and when comparing the wTO values for particular links between the datasets (**Figures 2A,B**). The differences between the networks can probably be explained by biological variation between individuals, but also by technical variations such as in RNA extraction methods, RIN values, and RNA library preparation procedures. We observed that the dataset BipRVal differs the most from the other datasets by having the highest number of wTO outliers, followed by datasets DisVal and FrontalVal

FIGURE 4 | Consensus network. In red Brain-GRFs; in blue all other GRFs. Node size is proportional to the number of links. Links with positive wTO values are in blue and links negative wTO values are shown in red.

(**Figures 2B,C**). All in all, we found that merely 19% (287930) of all links between GRFs are present in all 10 wTO networks. Given such variation between the networks, we think it is dangerous to rely on only one dataset when making inferences about biological relationships. Instead, multiple datasets should be combined to alleviate the dependence of the results on a particular set of individuals, developmental time points, different RNA library preparations, and gene expression measurement platforms and to focus on the most consistently observed links.

To combine the 10 independently derived networks into a consensus network with higher confidence, we considered them as biological replicates. We evaluated for each GRF—GRF pair, whether the distribution of strengths of their links across the 0 datasets is significantly higher than a chosen cutoff (Wilcoxon rank sum test, $p < 0.05$; **Figure 3** and see Materials and Methods). If so, the link was included into our consensus network. The resulting consensus network for $|wTO| > 0.3$ consists of 2516 links (**Figure 4** and Tables S2A,B). This method allowed us to pinpoint the links with the strongest consistency across multiple networks. To determine the final weight of the links in the consensus network, we calculated the median of all wTO values for the respective GRF—GRF pair.

Brain-GRF Genes Are Often Hubs and Highly Interconnected in the Frontal Lobe Consensus Network

Focusing on the most consistent links as determined by our consensus network, we next analyzed how the known Brain-GRFs are integrated into this consensus network. Of the total of 676 Brain-GRFs, 166 are present in the consensus network. Interestingly, this represents a significant enrichment of Brain-GRFs among the 498 GRFs of the consensus network (Fisher exact test, $p = 1.79 \times 10^{-11}$, Odd Ratio = 2.2). Remarkably, the group of Brain-GRFs has a higher connectivity (number of links) compared to other GRFs in the consensus network (Wilcoxon rank sum test, $p = 0.015$). Those finding suggests that known Brain-GRFs have stronger and more consistent functional relationships amongst each other than other GRFs in the frontal lobe.

To confirm the transcriptional pathways suggested by our consensus network, we examined whether there is enrichment of the GRF binding sites in the regulatory sequences of the 5421 genes that are correlated with at least one of the 498 GRFs of the consensus network (Table S2C). To this end, we first performed a ChIP enrichment analysis (ChEA) using the updated ENCODE database and a manually curated list of
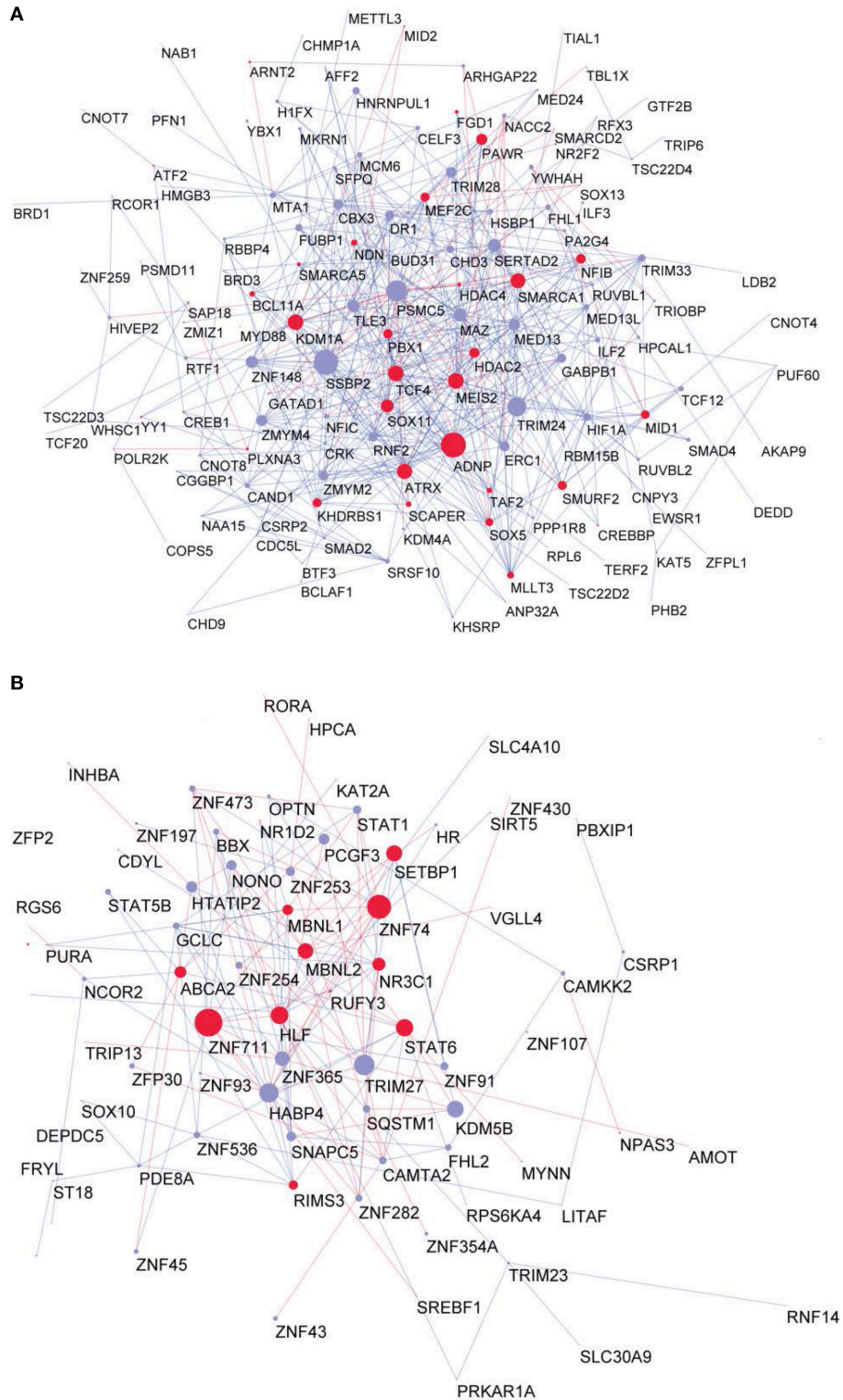
FIGURE 5 | Proteome GRF modules with red nodes representing the Brain-GRFs whereas in blue the other GRFs. Links with positive wTO values are in blue and links negative wTO values are shown in red. **(A)** Fetal module. **(B)** Adult module. Brain-GRFs are significantly enriched in the fetal module showing higher connectivity compared with the other GRFs.
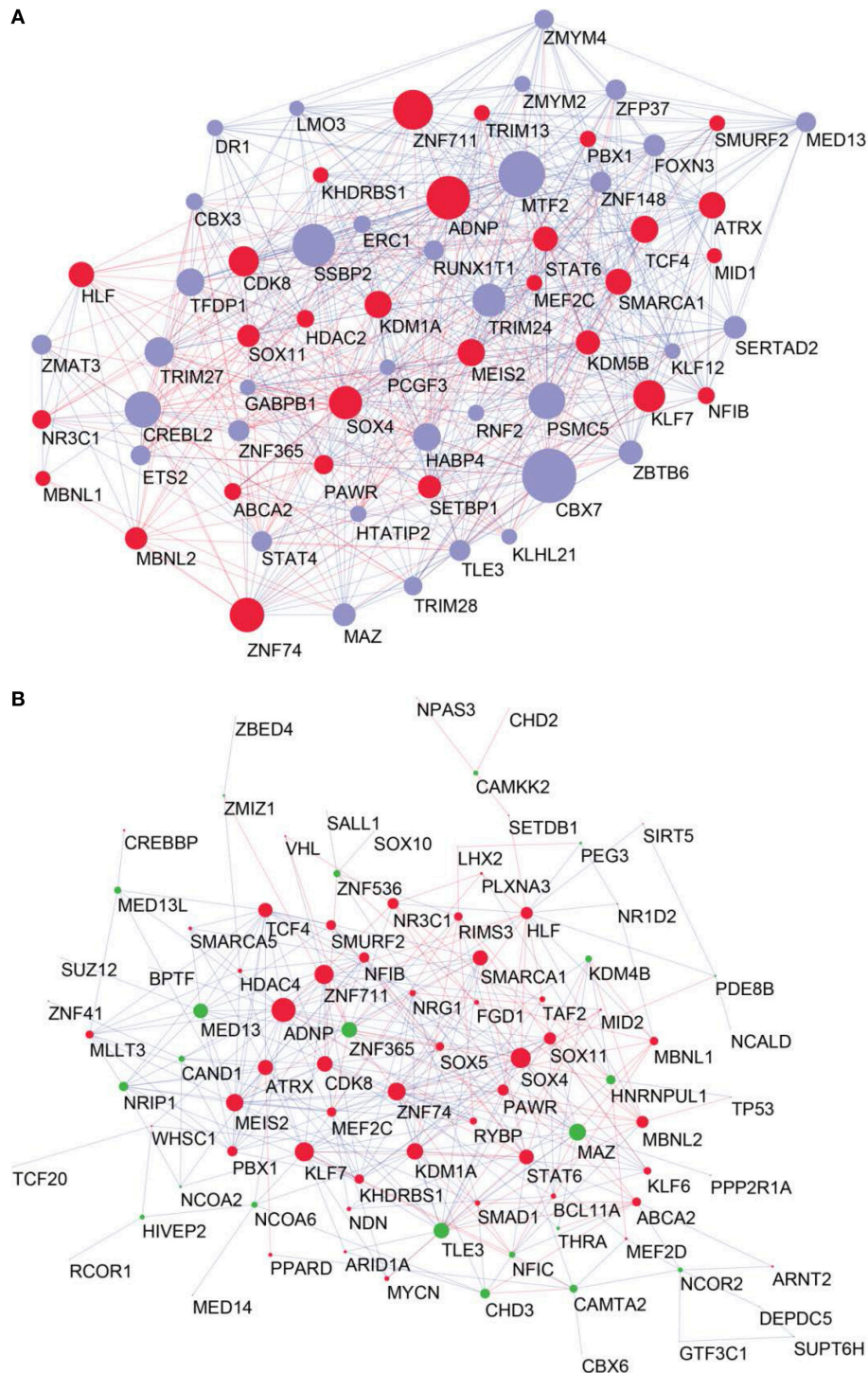
FIGURE 6 | High confident consensus network and proteomics networks. (A) Representation of the frontal lobe consensus network. Shown are the most highly connected hubs (degree > 25). Red nodes highlight Brain-GRFs, while blue nodes represent all other GRFs. The size of a node is proportional to its number of links: bigger nodes represent hubs in the network. Links with positive wTO values are in blue and links with negative wTO values are shown in red. **(B)** Brain-GRFs and FMRP targets module. Red nodes highlight the Brain-GRFs, while the green nodes highlight GRFs that are FMRP targets. The size of the nodes is proportional to their number of links.

target genes uncovered by ChIP-Seq, Chip-chip, ChIP-PET, and DamID from multiple studies (Lachmann et al., 2010). We found that the TFBS of 55 GRFs in the consensus network are significantly enriched among the regulatory sequences of the 5421 genes ($p < 0.05$ after Benjamini-Hochberg correction). Among those 55 GRFs, we found, for instance, *HDAC2* involved in synaptic plasticity and neural circuits (Guan et al., 2009), *ATF2* linked to neuronal apoptosis and cell migration (Yuan et al., 2009), and *CHD2* implicated in ASD and epilepsy (Rauch et al., 2012; Table S3A). Secondly, using the Jaspar and Jolma databases (Jolma et al., 2013; Mathelier et al., 2014), we found an enrichment of binding sites for 34 additional GRFs of the consensus network within the 2 kb region upstream of the transcription start site of the 5421 genes (Fisher exact test, $p < 0.05$ after Benjamini-Hochberg correction; Jolma et al., 2013; Mathelier et al., 2014). Here, we found enrichment for binding sites of *ARNTL*, important for circadian rhythm associated with BD (Nievergelt et al., 2006), *MEF2D*, involved in neuronal differentiation and PD (Yang et al., 2009), and *MEF2C*, involved in ASD, ID, and epilepsy (Novara et al., 2010) among others (Table S3B).

Coexpressed genes can also indicate protein interaction partners. Thus, we next examined protein—protein interactions (PPI) among the 498 GRFs and the 5421 correlated genes utilizing the annotations from BioGRID (Stark et al., 2006) and InWeb. We found that correlated GRF-gene pairs were significantly enriched within the PPI interactions (Fisher exact test, $p = 2.2 \times 10^{-6}$, Odd Ratio > 3), thus providing an additional confirmation of the potential functional interactions between GRFs and their correlated genes (Table S4).

In addition to the Brain-GRF enrichment, we examined the overlap between our consensus network with two coexpression modules, asdM12 and asdM16, that have been implicated in ASD previously (Voineagu et al., 2011). Remarkably, the consensus network overlaps significantly with the asdM12 module that is associated with synaptic development and dysregulated in ASD brains (hypergeometric test, $p = 0.045$). This result suggests that functional relationship of the GRFs in our consensus network plays a role in ASD.

To investigate whether the GRFs are also highly expressed at protein level in a fetal or adult brain, we superimposed our consensus network with a proteome map of the human brain at different stages, which was derived using mass-spectrometry proteomics (Kim et al., 2014). This strategy allowed us to understand the potential roles of the GRFs in the period of brain development and circuitry formation compared with an adult brain. Interestingly, overall the GRFs of our consensus network have higher expression and significantly more links in the fetal module compared to the adult module (Wilcoxon rank sum test, $p = 0.006$). The known Brain-GRFs are specifically enriched in the fetal module (Fisher exact test, $p = 0.03$, OR = 1.5) with generally higher number of links in comparison to other GRFs (Wilcoxon rank sum test, $p = 0.002$; **Figures 5A,B**).

To determine the most important GRFs in the consensus network of the human frontal pole, we determined the GRFs with the highest numbers of links (**Figures 6A,B**). Examples of

such hubs include *ADNP*, *ZFN711*, *ZNF74*, and *SOX4*, which are all Brain-GRFs. Interestingly, those Brain-GRFs are also strongly interconnected with other Brain-GRFs (e.g., *MEF2C*, *PBX1*, *SMARCA1*, an *SOX11*) and GRFs that are FMRP-targets (e.g., *KDM4B*, *MED13*, *NRIP1*, and *ZNF365*), suggesting a high functional interrelationship between various Brain-GRFs (**Figure 7**). Of note, in addition to the Brain-GRFs, the consensus network also contained hubs that yet are not implicated in brain functions or disorders. For example we detected GRFs important for embryogenesis (e.g., *CBX7*, *TFDP1*, and *TLE3*; Dehni et al., 1995; Morey et al., 2013; Laing et al., 2015) and energy metabolism (e.g., *PSMC5* and *SERTAD2*; Hoyle et al., 1997; Liew et al., 2013). Due to their strong connectivity to known Brain-GRFs in the consensus network, it seems likely that also these GRFs play an important role in the human frontal lobe circuitries. Taken together, our results suggest GRFs that are important for shaping the transcriptional circuitry of the human frontal lobe, including novel candidates for experimental validation of their roles at brain level and potential association with cognitive disorders.

To infer more about the functions of the GRFs in the consensus network, we performed a Gene Ontology (GO) enrichment analysis among the genes correlated with the GRFs (see Materials and Methods). We found significant enrichment for genes involved in metabolism, signaling, transport, translation, and RNA splicing (**Figure 8A**). We also specifically tested for GO enrichment of the genes correlated with three Brain-GRFs that are the strongest hubs in the consensus network: *ADNP*, *ZNF711*, and *ZNF74* (see Materials and Methods). Overall, we found similar GO groups enriched for these hubs like we did for the consensus network as a whole. However, there were also hub-specifically enriched GO categories such as brain development, methylation, and regulation of synaptic transmission, which suggests a specific role of these three GRFs in the regulation of genes important for these particular brain functions (**Figures 8B–D**; Table S5).

DISCUSSION

Comprehending the characteristic complexity of cognitive disorders, such as ASD and ID, still represents a challenge in neurosciences. An important step toward understanding this complexity is to elucidate the molecular networks of healthy human brains. In this study, we specifically compiled a set of 676 "Brain-GRF" genes implicated in brain development and cognitive disorders and analyzed their co-expression networks to gain first insights into which gene regulatory pathways these genes may be involved in in the frontal lobe of healthy individuals. Importantly, we discovered that networks derived from independent studies differ considerably from each other, highlighting a potential danger of relying on just one dataset. After combining these independent networks into a consensus network containing the links that are the most conserved across them, we were able to identify robust relationships between GRFs in the coexpression network of the frontal lobe of healthy human individuals. We further discovered that, while
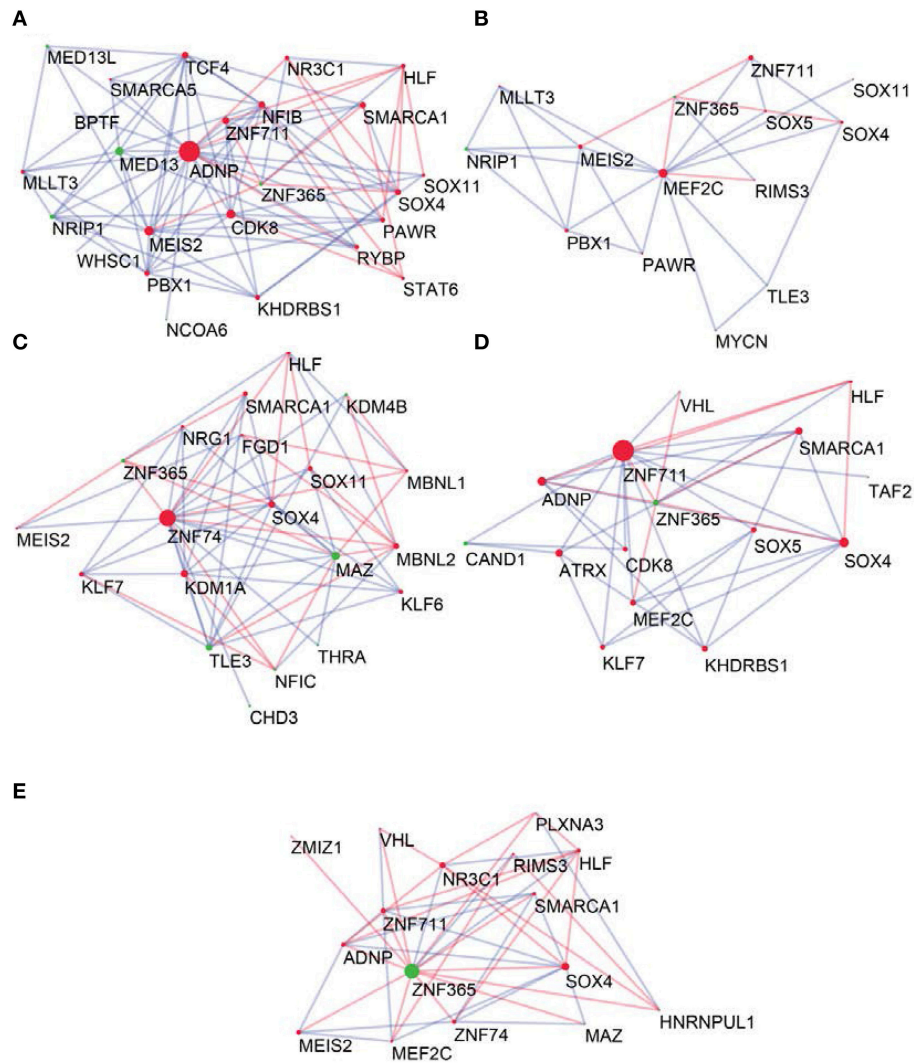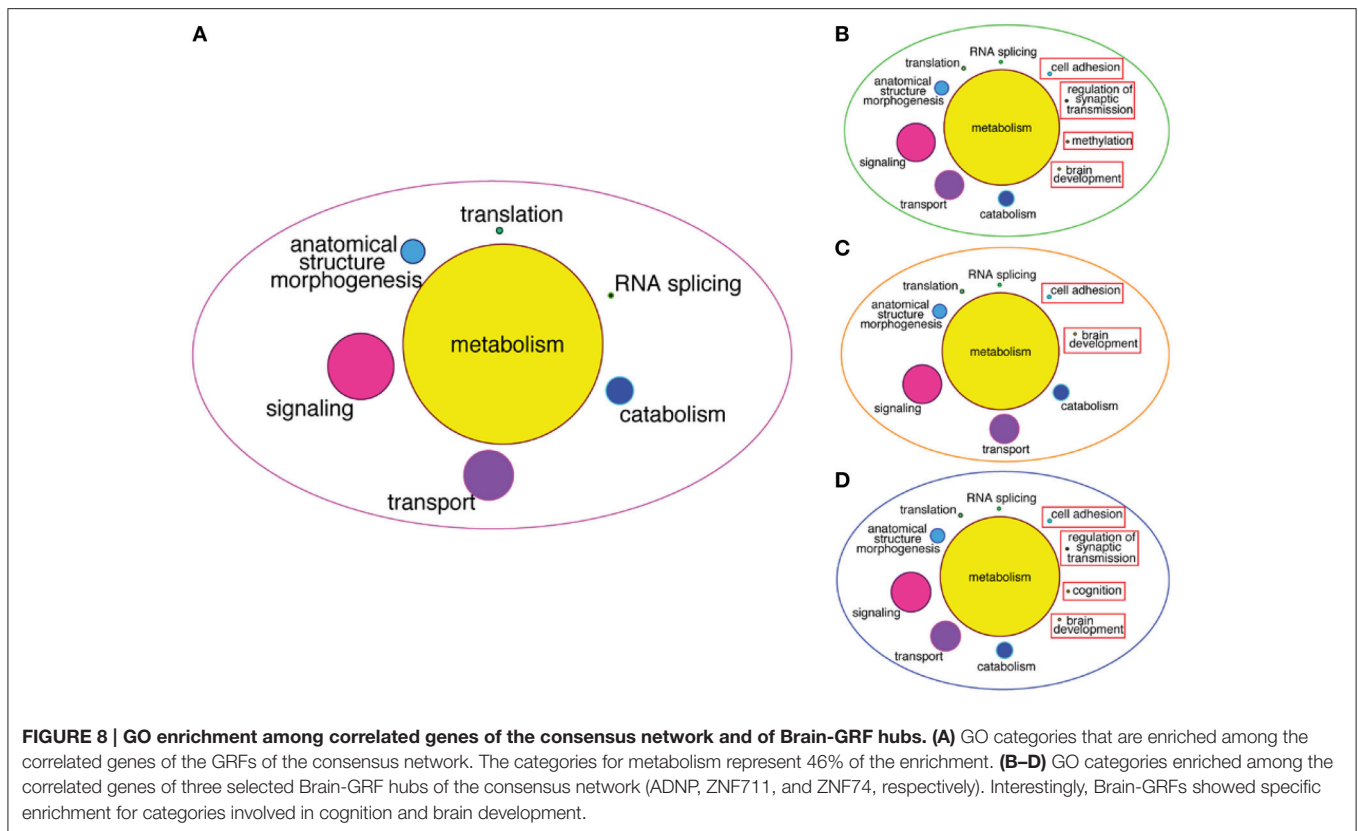
FIGURE 7 | Neighbors of hub Brain-GRFs and their strongly connected partners. (A) ADNP module, **(B)** MEF2C module, **(C)** ZNF74 module, **(D)** ZNF711 module, and **(E)** ZNF365 module. Red nodes highlight Brain-GRFs whereas green nodes represent FMRP targets. Links with positive wTO values are in blue and links negative wTO values are shown in red. Each hub Brain-GRFs is interestingly associated with other known Brain-GRFs highlighting potential interactions and common pathways.

some hubs in the consensus network are known "Brain-GRF" genes, others have not been linked to functions in the brain before.

The function of most GRFs is still only insufficiently characterized. However, insights into the functions and interactions of our human frontal lobe consensus network can be gained from the expression patterns of the GRFs, the GO enrichment of the genes correlated with the GRFs, and disorders the GRFs have been associated with. Many hubs of the consensus network are also expressed in tissues other than brain. However, we observed that a considerable number of them (115 in total), for example *ZNF711, ADNP, MEF2C, SOX11,* and *CBX7*, have higher expression in mouse neurons than in other brain cells, such as glia, astrocytes, oligodendrocytes, myelinating oligodendrocytes, and endothelial cells (Zhang et al.,

2014), suggesting that they have an essential role in neurons. In addition, we also discovered that the GRFs of our network play dominant roles in the fetal proteome module, (Kim et al., 2014) supporting the reasoning that these GRFs might regulate important processes during brain development such as forming the necessary brain structures for proper brain functions, including cognitive functions. Despite being ubiquitously expressed, it is plausible that some GRFs might only be hubs in the frontal lobe, a possibility that needs to be investigated further when data becomes available.

Our GO analysis revealed that the hub GRFs of the frontal lobe consensus network are likely to regulate genes involved in splicing, translation, metabolism, signaling, and synaptic transmission in the frontal lobe. Interestingly, these GO categories seem to be important for several brain functions.

FIGURE 8 | GO enrichment among correlated genes of the consensus network and of Brain-GRF hubs. (A) GO categories that are enriched among the correlated genes of the GRFs of the consensus network. The categories for metabolism represent 46% of the enrichment. **(B–D)** GO categories enriched among the correlated genes of three selected Brain-GRF hubs of the consensus network (ADNP, ZNF711, and ZNF74, respectively). Interestingly, Brain-GRFs showed specific enrichment for categories involved in cognition and brain development.

For instance, translational mechanisms have been shown to play a role in the mechanisms of memory formation and synaptic plasticity (Richter and Klann, 2009) and RNA splicing mechanisms have been implicated in neuronal development (Li et al., 2007; Weyn-Vanhentenryck et al., 2014). Genes involved in metabolism might be important to provide the brain with the necessary energy for its functions. Signaling and synaptic transmission are important for the communication between neurons and relevant to allow for cognitive abilities. We thus suggest the interactions of the GRFs in the frontal lobe network are critically underlying the regulatory processes that allow for these vital brain functions.

We found a significant enrichment of known Brain-GRFs, including GRFs implicated in ASD, ID, or SY in our consensus network, indicating that it forms the basis for setting the stage for healthy cognitive abilities. For instance, the three strongest hubs are *ZNF711*, associated with ID (Tarpey et al., 2009), *ADNP*, involved in ID and ASD (Helsmoortel et al., 2014; Iossifov et al., 2014), and *ZNF74*, involved in ID and SY (Ravassard et al., 1999). Being in these central network positions presumably renders them to risk genes that increase the likelihood for developing brain disorders. We speculate that interaction between *ZNF711* and *ZNF74* reflect biological pathways that might be important for intellectual abilities. In line with this potential, genes correlated with *ZNF711* and *ZNF74* are enriched for functions such as axon development, brain development and regulation of synaptic transmission, which are

likely important for the development and maintenance of healthy cognitive skills. Another hub in our GRF consensus network is *MEF2C*, a GRF that is important for synaptic plasticity and has been implicated in ASD (Ebert and Greenberg, 2013). *MEF2C* is also strongly associated with other Brain-GRFs such as *ZNF711*, *SOX11*, and *SOX5*, defining a strongly interconnected module of GRFs involved in regulatory pathways that might control cognitive functions (Uwanogho et al., 1995; Jankowski et al., 2006; Tarpey et al., 2009; Schanze et al., 2013). Our analysis highlighted also hubs that are targeted by FMRP, pointing to pathways that might be (dys)regulated at the post-transcriptional level. For instance, *CREBBP*, a GRF associated with ASD and ID (Barnby et al., 2005), *HDAC4*, implicated in ID and ASD (Pinto et al., 2014), *ZNF365*, which has been discovered in a module strongly associated with ASD in a brain expression study (Voineagu et al., 2011), and *KDM5B* and *KDM4B*, recently implicated in ASD using another weighted network approach (TADA; De Rubeis et al., 2014; Iossifov et al., 2014). *CREB* transcription factors and *HDAC4* are further known to regulate synaptic plasticity and memory formation (Silva et al., 1998; Hardingham et al., 2001; Vecsey et al., 2007; Thomson et al., 2008; Kim et al., 2012; Sando et al., 2012). These observations lead us to speculate that Brain-GRFs are strongly dependent on each other by sharing functional pathways and target genes. Further experimental studies are needed to identify shared targets of these and other GRFs to confirm their role in human frontal lobe functions and disorders.

Supporting our speculation that Brain-GRFs depend on each other, we found that Brain-GRFs have significantly more links than other GRFs and are strongly interconnected in the human frontal lobe network. Importantly, in addition to 30 known Brain-GRFs that are hubs, we identified further 36 GRF genes that are hubs in the frontal lobe consensus network but were not included in our Brain-GRFs list. Interestingly, one of these hubs, *GABPB1* encodes for a subunit of the hetero-tetrameric GABP consisting of two GABPA and two GABPB subunits (Batchelor et al., 1998). GABPA was recently found to bind human-specific binding sites and regulate gene expression of at least four genes (*ALDOA, HSPA8, TP73,* and *TMBIM6*) that have been associated with cognitive diseases such as autism, AZ, PD and other brain disorders (Perdomo-Sabogal et al., 2016). To 0 explore if more of these hubs might be associated with brain functions, we mined the (non-curated) data from DisGeNET (Piñero et al., 2015). We found that at least 12 of these hub GRFs may be connected with mental diseases and other neurological pathologies such as AZ (*DR1, ETS2, TFDP1,* and *TRIM13*), PD (*RUNX1T1*), SZ (*ZNF365*), developmental verbal dyspraxia (*ERC1*) and central neuroblastoma (*LMO3, PSMC5, TRIM13, TRIM24, ZMAT3*), among others. This suggests that with our method we have potentially identified novel candidates for being associated with important, if not essential, functions in the brain. We speculate that sequence and regulatory changes altering the regulatory activity or expression of these 36 hub GRFs could have medical relevance. It would thus be highly interesting to experimentally investigate their functions at brain level.

The structure and organization of the consensus network we are presenting here provides insights into regulatory circuits of the human frontal lobe. However, a yet unanswered question is how the network that we described for the human frontal lobe differs from the network of other brain regions, tissues or species. We expect that the relevant data for addressing this question will become available soon. We also expect that more GRFs will be discovered to be involved in brain functions. In future studies similar strategies as we presented here can then be implemented to enrich our knowledge about the molecular basis and regulatory networks underlying cognitive abilities.

AUTHOR CONTRIBUTIONS

SB designed and executed research; AP contributed material; DG contributed analysis programs and visualizations; JQ contributed methods and designed research; KN designed research; All authors wrote and discussed the manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fgene.2016.00031

REFERENCES

Akula, N., Barb, J., Jiang, X., Wendland, J. R., Choi, K. H., and Sen, S. K. (2014). RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol. Psychiatry* 19, 1179–1185. doi: 10.1038/mp.2013.170

Akula, N., Wendland, J. R., Choi, K. H., and McMahon, F. J. (2016). An integrative genomic study implicates the postsynaptic density in the pathogenesis of bipolar disorder. *Neuropsychopharmacology* 41, 886–895. doi: 10.1038/npp.2015.218

Allen, N. C., Bagade, S., McQueen, M. B., Ioannidis, J. P., Kavvoura, F. K., Khoury, M. J., et al. (2008). Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat. Genet.* 40, 827–834. doi: 10.1038/ng.171

Andreasen, N. C. (1995). Symptoms, signs, and diagnosis of schizophrenia. *Lancet* 346, 477–481. doi: 10.1016/S0140-6736(95)91325-4

Bailey, A., Phillips, W., and Rutter, M. (1996). Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *J. Child Psychol. Psychiatry* 37, 89–126. doi: 10.1111/j.1469-7610.1996.tb01381.x

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208. doi: 10.1093/nar/gkp335

Bailey, T. L., and Machanick, P. (2012). Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.* 40, e128–e128. doi: 10.1093/nar/gks433

Banerjee-Basu, S., and Packer, A. (2010). SFARI Gene: an evolving database for the autism research community. *Dis. Model. Mech.* 3, 133–135. doi: 10.1242/dmm.005439

Barnby, G., Abbott, A., Sykes, N., Morris, A., Weeks, D. E., Mott, R., et al. (2005). Candidate-gene screening and association analysis at the autism-susceptibility locus on chromosome 16p: evidence of association at GRIN2A and ABAT. *Am. J. Hum. Genet.* 76, 950–966. doi: 10.1086/430454

Basu, S. N., Kollu, R., and Banerjee-Basu, S. (2009). AutDB: a gene reference resource for autism research. *Nucleic Acids Res.* 37, D832–D836. doi: 10.1093/nar/gkn835

Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L., and Wolberger, C. (1998). The structure of GABPα/β: an ETS domain-ankyrin repeat heterodimer bound to DNA. *Science* 279, 1037–1041. doi: 10.1126/science.279.5353.1037

Berg, J. M., and Geschwind, D. H. (2012). Autism genetics: searching for specificity and convergence. *Genome Biol.* 13, 247. doi: 10.1186/gb-2012-13-7-247

Bertram, L. (2009). Alzheimers disease genetics current status and future perspectives. *Int. Rev. Neurobiol.* 84, 167–184. doi: 10.1016/S0074-7742(09)00409-7

Bullido, M. J., Artiga, M. J., Recuero, M., Sastre, I., Garcia, M. A., Aldudo, J., et al. (1998). A polymorphism in the regulatory region of APOE associated

with risk for Alzheimers dementia. *Nat. Genet.* 18, 69–71. doi: 10.1038/ng 0198-69

Carvalho, B. S., and Irizarry, R. A. (2010). A framework for oligonucleotide microarray preprocessing. *Bioinformatics* 26, 2363–2367. doi: 10.1093/bioinformatics/btq431

Chatr-Aryamontri, A., Breitkreutz, B. J., Heinicke, S., Boucher, L., Winter, A., Stark, C., et al. (2013). The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* 41, D816–D823. doi: 10.1093/nar/gks1158

Chayer, C., and Freedman, M. (2001). Frontal lobe functions. *Curr. Neurol. Neurosci. Rep.* 1, 547–552. doi: 10.1007/s11910-001-0060-4

Consortium SWGotPG (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 511, 421–427. doi: 10.1038/nature13595

Corsinotti, A., Kapopoulou, A., Gubelmann, C., Imbeault, M., Santoni de Sio, F. R., Rowe, H. M., et al. (2013). Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS ONE* 8:e56721. doi: 10.1371/journal.pone.0056721

Darnell, J. C., Van Driesche, S. J., Zhang, C., Hung, K. Y., Mele, A., Fraser, C. E., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* 146, 247–261. doi: 10.1016/j.cell.2011.06.013

Dehni, G., Liu, Y., Husain, J., and Stifani, S. (1995). TLE expression correlates with mouse embryonic segmentation, neurogenesis, and epithelial determination. *Mech. Dev.* 53, 369–381. doi: 10.1016/0925-4773(95)00452-1

De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215. doi: 10.1038/nature13772

Duncan, J., Emslie, H., Williams, P., Johnson, R., and Freer, C. (1996). Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cogn. Psychol.* 30, 257–303. doi: 10.1006/cogp.1996.0008

Ebert, D. H., and Greenberg, M. E. (2013). Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* 493, 327–337. doi: 10.1038/nature11860

Ecker, C., Suckling, J., Deoni, S. C., Lombardo, M. V., Bullmore, E. T., Baron-Cohen, S., et al. (2012). Brain anatomy and its relationship to behavior in adults with autism spectrum disorder: a multicenter magnetic resonance imaging study. *Arch. Gen. Psychiatry* 69, 195–209. doi: 10.1001/archgenpsychiatry.2011.1251

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy–analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20, 307–315. doi: 10.1093/bioinformatics/btg405

Greydanus, D. E., and Pratt, H. D. (2005). Syndromes and disorders associated with mental retardation. *Indian J. Pediatr.* 72, 859–864. doi: 10.1007/BF02731116

Guan, J. S., Haggarty, S. J., Giacometti, E., Dannenberg, J. H., Joseph, N., Gao, J., et al. (2009). HDAC2 negatively regulates memory formation and synaptic plasticity. *Nature* 459, 55–60. doi: 10.1038/nature07925

Hamosh, A., Scott, A. F., Amberger, J. S., Bocchini, C. A., and McKusick, V. A. (2005). Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 33, D514–D517. doi: 10.1093/nar/gki033

Hardingham, G. E., Arnold, F. J., and Bading, H. (2001). Nuclear calcium signaling controls CREB-mediated gene expression triggered by synaptic activity. *Nat. Neurosci.* 4, 261–267. doi: 10.1038/85109

Helsmoortel, C., Vulto-van Silfhout, A. T., Coe, B. P., Vandeweyer, G., Rooms, L., van den Ende, J., et al. (2014). A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat. Genet.* 46, 380–384. doi: 10.1038/ng.2899

Hoffmann, S., Otto, C., Kurtz, S., Sharma, C. M., Khaitovich, P., Vogel, J., et al. (2009). Fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput. Biol.* 5:e1000502. doi: 10.1371/journal.pcbi.1000502

Hong, E. J., West, A. E., and Greenberg, M. E. (2005). Transcriptional control of cognitive development. *Curr. Opin. Neurobiol.* 15, 21–28. doi: 10.1016/j.conb.2005.01.002

Hoyle, J., Tan, K. H., and Fisher, E. M. (1997). Localization of genes encoding two human one-domain members of the AAA family: PSMC5 (the thyroid hormone receptor-interacting protein, TRIP1) and PSMC3 (the Tat-binding protein, TBP1). *Hum. Genet.* 99, 285–288. doi: 10.1007/s004390050356

Hu, Z.-L., Bao, J., and Reecy, J. M. (2008). CateGOrizer: a web-based program to batch analyze gene ontology classification categories. *Online J. Bioinformatics* 9, 108–112. Available online at: http://onljvetres.com/geneontologyabs2008.htm; http://www.animalgenome.org/tools/catego/index.html

Ihaka, R., and Gentleman, R. (1996). R: a language for data analysis and graphics. *J. Comput. Graph. Stat.* 5, 299–314.

Inlow, J. K., and Restifo, L. L. (2004). Molecular and comparative genetics of mental retardation. *Genetics* 166, 835–881. doi: 10.1534/genetics.166.2.835

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515, 216–221. doi: 10.1038/nature13908

Jankowski, M. P., Cornuet, P. K., McIlwrath, S., Koerber, H. R., and Albers, K. M. (2006). SRY-box containing gene 11 (Sox11) transcription factor is required for neuron survival and neurite growth. *Neuroscience* 143, 501–514. doi: 10.1016/j.neuroscience.2006.09.010

Jia, P., Sun, J., Guo, A., and Zhao, Z. (2010). SZGR: a comprehensive schizophrenia gene resource. *Mol. Psychiatry* 15, 453–462. doi: 10.1038/mp.2009.93

Jolma, A., Yan, J., Whitington, T., Toivonen, J., Nitta, K. R., Rastas, P., et al. (2013). DNA-binding specificities of human transcription factors. *Cell* 152, 327–339. doi: 10.1016/j.cell.2012.12.009

Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., et al. (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483–489. doi: 10.1038/nature10523

Kaufman, L., Ayub, M., and Vincent, J. B. (2010). The genetic basis of non-syndromic intellectual disability: a review. *J. Neurodev. Disord.* 2, 182–209. doi: 10.1007/s11689-010-9055-2

Kim, M.S., Akhtar, M. W., Adachi, M., Mahgoub, M., Bassel-Duby, R., Kavalali, E. T., et al. (2012). An essential role for histone deacetylase 4 in synaptic plasticity and memory formation. *J. Neurosci.* 32, 10879–10886. doi: 10.1523/JNEUROSCI.2089-12.2012

Kim, M.S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., et al. (2014). A draft map of the human proteome. *Nature* 509, 575–581. doi: 10.1038/nature13302

Lachmann, A., Xu, H., Krishnan, J., Berger, S. I., Mazloom, A. R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics* 26, 2438–2444. doi: 10.1093/bioinformatics/btq466

Laing, A. F., Lowell, S., and Brickman, J. M. (2015). Gro/TLE enables embryonic stem cell differentiation by repressing pluripotent gene expression. *Dev. Biol.* 397, 56–66. doi: 10.1016/j.ydbio.2014.10.007

Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., et al. (2013). Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* 9:e1003118. doi: 10.1371/journal.pcbi.1003118

Li, Q., Lee, J.-A., and Black, D. L. (2007). Neuronal regulation of alternative pre-mRNA splicing. *Nat. Rev. Neurosci.* 8:819–831. doi: 10.1038/nrn2237

Liew, C. W., Boucher, J., Cheong, J. K., Vernochet, C., Koh, H.-J., Mallol, C., et al. (2013). Ablation of TRIP-Br2, a regulator of fat lipolysis, thermogenesis and oxidative metabolism, prevents diet-induced obesity and insulin resistance. *Nat. Med.* 19, 217–226. doi: 10.1038/nm.3056

Lill, C. M., Roehr, J. T., McQueen, M. B., Kavvoura, F. K., Bagade, S., Schjeide, B. M., et al. (2012). Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. *PLoS Genet.* 8:e1002548. doi: 10.1371/journal.pgen.1002548

Liu, X., Somel, M., Tang, L., Yan, Z., Jiang, X., Guo, S., et al. (2012). Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res.* 22, 611–622. doi: 10.1101/gr.127324.111

Lubs, H. A., Stevenson, R. E., and Schwartz, C. E. (2012). Fragile X and X-linked intellectual disability: four decades of discovery. *Am. J. Hum. Genet.* 90, 579–590. doi: 10.1016/j.ajhg.2012.02.018

Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 42, D142–D147. doi: 10.1093/nar/gkt997

Messina, D. N., Glasscock, J., Gish, W., and Lovett, M. (2004). An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res.* 14, 2041–2047. doi: 10.1101/gr.2584104

Morey, L., Aloia, L., Cozzuto, L., Benitah, S. A., and Di Croce, L. (2013). RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell Rep.* 3, 60–69. doi: 10.1016/j.celrep.2012.11.026

Nievergelt, C. M., Kripke, D. F., Barrett, T. B., Burg, E., Remick, R. A., Sadovnick, A. D., et al. (2006). Suggestive evidence for association of the circadian genes

PERIOD3 and ARNTL with bipolar disorder. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* 141B, 234–241. doi: 10.1002/ajmg.b.30252

Nord, A. S., Pattabiraman, K., Visel, A., and Rubenstein, J. L. (2015). Genomic perspectives of transcriptional regulation in forebrain development. *Neuron* 85, 27–47. doi: 10.1016/j.neuron.2014.11.011

Novara, F., Beri, S., Giorda, R., Ortibus, E., Nageshappa, S., Darra, F., et al. (2010). Refining the phenotype associated with MEF2C haploinsufficiency. *Clin. Genet.* 78, 471–477. doi: 10.1111/j.1399-0004.2010.01413.x

Nowick, K., Fields, C., Gernat, T., Caetano-Anolles, D., Kholina, N., and Stubbs, L. (2011). Gain, loss and divergence in primate zinc-finger genes: a rich resource for evolution of gene regulatory differences between species. *PLoS ONE* 6:e21553. doi: 10.1371/journal.pone.0021553

Nowick, K., Gernat, T., Almaas, E., and Stubbs, L. (2009). Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proc. Natl. Acad. Sci. U.S.A.* 106, 22358–22363. doi: 10.1073/pnas.0911376106

Parikshak, N. N., Luo, R., Zhang, A., Won, H., Lowe, J. K., Chandran, V., et al. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* 155, 1008–1021. doi: 10.1016/j.cell.2013.10.031

Perdomo-Sabogal, A., Nowick, K., Piccini, I., Sudbrak, R., Lehrach, H., Yaspo, M. L., et al. (2016). Human lineage-specific transcriptional regulation through GA-binding protein transcription factor alpha (GABPa). *Mol. Biol. Evol.* doi: 10.1093/molbev/msw007. [Epub ahead of print].

Piñero, J., Queralt-Rosinach, N., Bravo, À., Deu-Pons, J., Bauer-Mehren, A., Baron, M., et al. (2015). DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database* 2015:bav028. doi: 10.1093/database/bav028

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., et al. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* 94, 677–694. doi: 10.1016/j.ajhg.2014.03.018

Polymeropoulos, M. H. (2000). Genetics of Parkinson's disease. *Ann. N.Y. Acad. Sci.* 920, 28–32. doi: 10.1111/j.1749-6632.2000.tb06901.x

Prüfer, K., Muetzel, B., Do, H. H., Weiss, G., Khaitovich, P., Rahm, E., et al. (2007). FUNC: a package for detecting significant associations between gene sets and ontological annotations. *BMC Bioinform.* 8:41. doi: 10.1186/1471-2105-8-41

Rauch, A., Wieczorek, D., Graf, E., Wieland, T., Endele, S., Schwarzmayr, T., et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* 380, 1674–1682. doi: 10.1016/S0140-6736(12)61480-9

Ravasi, T., Suzuki, H., Cannistraci, C. V., Katayama, S., Bajic, V. B., Tan, K., et al. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. *Cell* 140, 744–752. doi: 10.1016/j.cell.2010.01.044

Ravassard, P., Côté, F., Grondin, B., Bazinet, M., Mallet, J., and Aubry, M. (1999). ZNF74, a gene deleted in DiGeorge syndrome, is expressed in human neural crest-derived tissues and foregut endoderm epithelia. *Genomics* 62, 82–85. doi: 10.1006/geno.1999.5982

Richter, J. D., and Klann, E. (2009). Making synaptic plasticity and memory last: mechanisms of translational regulation. *Genes Dev.* 23, 1–11. doi: 10.1101/gad.1735809

Ropers, H. H. (2008). Genetics of intellectual disability. *Curr. Opin. Genet. Dev.* 18, 241–250. doi: 10.1016/j.gde.2008.07.008

Rossin, E. J., Lage, K., Raychaudhuri, S., Xavier, R. J., Tatar, D., Benita, Y., et al. (2011). Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* 7:e1001273. doi: 10.1371/journal.pgen.1001273

Ryan, M. M., Lockstone, H. E., Huffaker, S. J., Wayland, M. T., Webster, M. J., and Bahn, S. (2006). Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol. Psychiatry* 11, 965–978. doi: 10.1038/sj.mp.4001875

Sando, R.IIIrd, Gounko, N., Pieraut, S., Liao, L., Yates, J., and Maximov, A. (2012). HDAC4 governs a transcriptional program essential for synaptic plasticity and memory. *Cell* 151, 821–834. doi: 10.1016/j.cell.2012.09.037

Schanze, I., Schanze, D., Bacino, C. A., Douzgou, S., Kerr, B., and Zenker, M. (2013). Haploinsufficiency of SOX5, a member of the SOX (SRY-related HMG-box) family of transcription factors is a cause of intellectual disability. *Eur. J. Med. Genet.* 56, 108–113. doi: 10.1016/j.ejmg.2012.11.001

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., et al. (2007). Strong association of *de novo* copy number mutations with autism. *Science* 316, 445–449. doi: 10.1126/science.1138659

Shettleworth, S. J. (2009). *Cognition, Evolution, and Behavior.* New York, NY: Oxford University Press.

Silva, A. J., Kogan, J. H., Frankland, P. W., and Kida, S. (1998). CREB and memory. *Annu. Rev. Neurosci.* 21, 127–148. doi: 10.1146/annurev.neuro.21.1.127

Somel, M., Franz, H., Yan, Z., Lorenc, A., Guo, S., Giger, T., et al. (2009). Transcriptional neoteny in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5743–5748. doi: 10.1073/pnas.0900544106

Somel, M., Liu, X., Tang, L., Yan, Z., Hu, H., Guo, S., et al. (2011). MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol.* 9:e1001214. doi: 10.1371/journal.pbio.1001214

Stark, C., Breitkreutz, B.-J., Reguly, T., Boucher, L., Breitkreutz, A., and Tyers, M. (2006). BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* 34, D535–D539. doi: 10.1093/nar/gkj109

Takahashi, J. S. (2015). Molecular components of the circadian clock in mammals. *Diabetes Obes. Metab.* 17(Suppl. 1), 6–11.

Tarpey, P. S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., et al. (2009). A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat. Genet.* 41, 535–543. doi: 10.1038/ng.367

Thomson, D. M., Herway, S. T., Fillmore, N., Kim, H., Brown, J. D., Barrow, J. R., et al. (2008). AMP-activated protein kinase phosphorylates transcription factors of the CREB family. *J. Appl. Physiol.* 104, 429–438. doi: 10.1152/japplphysiol.00900.2007

Tripathi, S., Christie, K. R., Balakrishnan, R., Huntley, R., Hill, D. P., Thommesen, L., et al. (2013). Gene Ontology annotation of sequence-specific DNA binding transcription factors: setting the stage for a large-scale curation effort. *Database* 2013:bat062. doi: 10.1093/database/bat062

Tsankova, N., Renthal, W., Kumar, A., and Nestler, E. J. (2007). Epigenetic regulation in psychiatric disorders. *Nat. Rev. Neurosci.* 8, 355–367. doi: 10.1038/nrn2132

Uwanogho, D., Rex, M., Cartwright, E. J., Pearl, G., Healy, C., Scotting, P. J., et al. (1995). Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. *Mech. Dev.* 49, 23–36. doi: 10.1016/0925-4773(94)00299-3

van Bokhoven, H. (2011). Genetic and epigenetic networks in intellectual disabilities. *Annu. Rev. Genet.* 45, 81–104. doi: 10.1146/annurev-genet-110410-132512

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263. doi: 10.1038/nrg2538

Vecsey, C. G., Hawk, J. D., Lattal, K. M., Stein, J. M., Fabian, S. A., Attner, M. A., et al. (2007). Histone deacetylase inhibitors enhance memory and synaptic plasticity via CREB: CBP-dependent transcriptional activation. *J. Neurosci.* 27, 6128–6140. doi: 10.1523/JNEUROSCI.0296-07.2007

Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., et al. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* 474, 380–384. doi: 10.1038/nature10110

West, A. E., and Greenberg, M. E. (2011). Neuronal activity–regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* 3:a005744. doi: 10.1101/cshperspect.a005744

Weyn-Vanhentenryck, S. M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., et al. (2014). HITS-CLIP and integrative modeling define the Rbfox splicing-regulatory network linked to brain development and autism. *Cell Rep.* 6, 1139–1152. doi: 10.1016/j.celrep.2014.02.005

Williamson, D. F., Parker, R. A., and Kendrick, J. S. (1989). The box plot: a simple visual method to interpret data. *Ann. Intern. Med.* 110, 916–921. doi: 10.7326/0003-4819-110-11-916

Willsey, A. J., Sanders, S. J., Li, M., Dong, S., Tebbenkamp, A. T., Muhle, R. A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* 155, 997–1007. doi: 10.1016/j.cell.2013.10.020

Wingender, E., Schoeps, T., and Dönitz, J. (2013). TFClass: an expandable hierarchical classification of human transcription factors. *Nucleic Acids Res.* 41, D165–D170. doi: 10.1093/nar/gks1123

Wingender, E., Schoeps, T., Haubrock, M., and Dönitz, J. (2015). TFClass: a classification of human transcription factors and their rodent orthologs. *Nucleic Acids Res.* 43, D97–D102. doi: 10.1093/nar/gku1064

Yang, Q., She, H., Gearing, M., Colla, E., Lee, M., Shacka, J. J., et al. (2009). Regulation of neuronal survival factor MEF2D by chaperone-mediated autophagy. *Science* 323, 124–127. doi: 10.1126/science.1166088

Yuan, Z., Gong, S., Luo, J., Zheng, Z., Song, B., Ma, S., et al. (2009). Opposing roles for ATF2 and c-Fos in c-Jun-mediated neuronal apoptosis. *Mol. Cell. Biol.* 29, 2431–2442. doi: 10.1128/MCB.01344-08

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:17. doi: 10.2202/1544-6115.1128

Zhang, Y., Chen, K., Sloan, S. A., Bennett, M. L., Scholze, A. R., O'Keeffe, S., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34, 11929–11947. doi: 10.1523/JNEUROSCI.1860-14.2014

Zhang, Y. E., Landback, P., Vibranovski, M. D., and Long, M. (2011). Accelerated recruitment of new brain development genes into the human genome. *PLoS Biol.* 9:e1001179. doi: 10.1371/journal.pbio.1001179