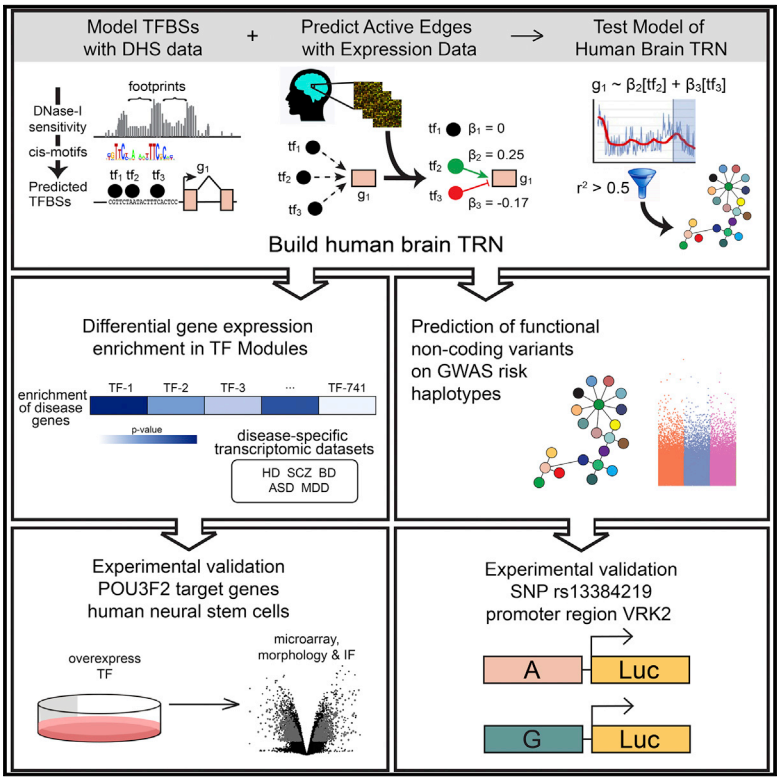
Cell Systems

Article

Genome-Scale Transcriptional Regulatory Network Models of Psychiatric and Neurodegenerative Disorders

Graphical Abstract



Highlights

- Network model predicts binding sites and target genes of 741 TFs in the human brain
- Target genes of key regulator TFs were dysregulated in brain diseases
- *POU3F2* regulates a schizophrenia- and bipolar-disorder-associated gene network

Authors

Jocelynn R. Pearl, Carlo Colantuoni, Dani E. Bergey, ..., Leroy Hood, Nathan D. Price, Seth A. Ament

Correspondence

nprice@systemsbiology.org (N.D.P.), sament@som.umaryland.edu (S.A.A.)

In Brief

Transcriptional regulatory changes in the developing and adult brain are prominent features of brain diseases, but roles for specific transcription factors (TFs) are poorly understood. We reconstructed a model for the binding sites and target genes of 741 TFs in the human brain and predicted key regulators of genetic and genomic changes in psychiatric and neurodegenerative disorders. In neural stem cells, the TF *POU3F2* regulated hundreds of genes enriched for neocortical gene expression changes in schizophrenia and bipolar disorder.

Pearl et al., 2019, Cell Systems *8*, 122–135
February 27, 2019 © 2019 Elsevier Inc.
https://doi.org/10.1016/j.cels.2019.01.002

CellPress

CellPress

Cell Systems
Article

Genome-Scale Transcriptional Regulatory Network Models of Psychiatric and Neurodegenerative Disorders

Jocelynn R. Pearl,[1,2,3] Carlo Colantuoni,[4,5] Dani E. Bergey,[1] Cory C. Funk,[1] Paul Shannon,[1] Bijoya Basu,[1] Alex M. Casella,[4,6] Rediet T. Oshone,[4] Leroy Hood,[1] Nathan D. Price,[1,*] and Seth A. Ament[1,4,7,*]

[1]Institute for Systems Biology, Seattle, WA, USA
[2]Molecular & Cellular Biology Graduate Program, University of Washington, Seattle, WA, USA
[3]Altius Institute of Biomedical Sciences, Seattle, WA, USA
[4]Institute of Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA
[5]Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, USA
[6]Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, USA
[7]Lead Contact
*Correspondence: nprice@systemsbiology.org (N.D.P.), sament@som.umaryland.edu (S.A.A.)
https://doi.org/10.1016/j.cels.2019.01.002

SUMMARY

Transcriptional regulatory changes in the developing and adult brain are prominent features of brain diseases, but the involvement of specific transcription factors (TFs) remains poorly understood. We integrated brain-specific DNase footprinting and TF-gene co-expression to reconstruct a transcriptional regulatory network (TRN) model for the human brain. We identified key regulator TFs whose predicted target genes were enriched for differentially expressed genes in the prefrontal cortex of individuals with psychiatric and neurodegenerative diseases. Many of these TFs were further implicated in the same diseases through disruption of their binding sites by disease-associated SNPs and associations of TF loci with disease risk. Using primary human neural stem cells, we validated network predictions that link the TF *POU3F2* to schizophrenia and bipolar disorder via both *cis*- and *trans*-acting mechanisms. Our models of brain-specific TF binding sites and target genes provide a resource for network analysis of brain diseases.

INTRODUCTION

Convergent evidence suggests that altered transcriptional regulation is a prominent mechanism of common human diseases, including psychiatric and neurodegenerative disorders. Many disease states are accompanied by characteristic tissue-specific changes in gene expression. Neurodegenerative diseases such as Alzheimer's disease (AD) and Huntington's disease involve progressive changes in the expression of thousands of genes in vulnerable brain regions (Hodges et al., 2006; Zhang et al., 2013). Psychiatric disorders such as schizophrenia (SCZ), bipolar disorder (BD), major depression disorder (MDD), and autism (ASD) also involve brain gene expression changes, particularly in neocortical regions involved in cognitive and emotional control (Akula et al., 2014; Fromer et al., 2016; Gandal et al., 2018; Seifuddin et al., 2013; Torkamani et al., 2010; Voineagu et al., 2011).

While non-genetic factors could contribute to brain gene expression changes in psychiatric and neurodegenerative disorders—including effects of medications, lifestyle factors, and differences in cell type distributions—multiple lines of evidence point to a substantial genetic contribution. Hundreds of genetic haplotypes associated with risk for psychiatric and neurodegenerative disorders co-localize with gene expression quantitative trait loci (eQTLs) (Fromer et al., 2016; Gusev et al., 2016; Hauberg et al., 2017). Genetic variants associated with risk for common diseases are enriched in promoters and enhancers, including regions of open chromatin marked by DNase I hypersensitivity (Gusev et al., 2014; Maurano et al., 2012). Thus, it has been proposed that the causal variants at many risk loci alter the expression of nearby target genes via mechanisms such as changes in transcription factor (TF) binding.

Genetically encoded changes in TFs and other transcriptional regulatory proteins may also contribute to disease risk. Both common and rare genetic variation associated with psychiatric disorders are enriched for genes involved in transcriptional regulation and chromatin-remodeling pathways (De Rubeis et al., 2014; PGC, 2015). 29 TFs are located at genome-wide significant-risk loci identified in large-scale genome-wide association studies (GWASs) of SCZ or BD (Hou et al., 2016; Mühleisen et al., 2014; PGC, 2014; Psychiatric GWAS Consortium, 2011). Additional TFs and chromatin remodeling genes have been implicated in risk for neurodevelopmental and psychiatric disorders because of their disruption by rare variants (Kamnasaran et al., 2003; Sanders et al., 2015; Singh et al., 2016).

We hypothesized that psychiatric and neurodegenerative disorders involve dysregulation within networks of TFs, TF binding sites, and TF target genes in the brain. Disease-related changes in transcriptional networks could occur either at the genetic level (i.e., disease-associated variation in the DNA sequence) or at the transcriptional level (i.e., changes in gene expression), or by both

mechanisms. To date, several TFs and cofactors have been implicated in SCZ (Whitton et al., 2018; Wright et al., 2016; Xia et al., 2018), ASD (Darnell et al., 2011; Sugathan et al., 2014), MDD (Wray et al., 2018), or AD (Huang et al., 2017), based on convergent evidence that the gene encoding a TF or cofactor is associated with risk for a disease, and its binding sites or target genes are also associated with risk for that disease or are differentially expressed in the brains of disease cases versus controls. Previous studies have also used gene co-expression to identify disease-perturbed networks in the brain (Gandal et al., 2018; Torkamani et al., 2010; Voineagu et al., 2011; Zhang et al., 2013) and have observed that a subset of differentially expressed networks are enriched for genes that are also associated with genetic risk for the same disease (Voineagu et al., 2011; Zhang et al., 2013). However, specific TFs regulating these gene co-expression networks are generally poorly defined, since methods to reconstruct eukaryotic gene regulatory networks solely from gene co-expression data only rarely predict direct regulatory interactions (Berto et al., 2016; Brichta et al., 2015; Khosravi et al., 2015; Marbach et al., 2012). Other studies have utilized epigenomic profiling and eQTLs to annotate non-coding genetic variation associated with risk for psychiatric and neurodegenerative disorders (Fromer et al., 2016; Hauberg et al., 2017) and to fine-map transcriptional regulatory mechanisms at specific disease risk loci (Kichaev et al., 2014; Li and Kellis, 2016; Won et al., 2016). These studies support the hypothesis that networks of TFs, TF binding sites, and TF target genes influence risk for brain diseases. However, genome-wide studies have not previously integrated all of these levels of transcriptional regulatory network (TRN) organization to understand the transcriptional regulatory architecture of brain diseases.

Here, we reconstructed a TRN model that predicts the genomic binding sites and target genes of 741 TFs in the human brain by integrating tissue-specific DNase-seq footprinting with TF-gene co-expression. We used our TRN model to predict key TFs that regulate transcriptomic changes in psychiatric and neurodegenerative diseases, as well as disease-associated SNPs that disrupt TF binding sites. We integrate these results to characterize *cis*- and *trans*-acting mechanisms linking TRNs to disease and validate our model's predictions for the SCZ and BD-associated TF *POU3F2*.

RESULTS

Reconstruction of a Transcriptional Regulatory Network Model for the Human Brain

To predict the genome-wide binding sites for TFs in the human brain, we performed digital genomic footprinting analysis with 15 DNase-seq experiments from human brain regions and cell types generated by the ENCODE project (Table S1). DNase I cleavage patterns predict occupied binding sites for TFs and other DNA-binding proteins (Hesselberth et al., 2009; Neph et al., 2012). We downloaded raw sequencing reads from each DNase-seq experiment (www.encodeproject.org), aligned the reads to the human genome with SNAP (Zaharia et al., 2011), identified DNase I hypersensitive regions with F-seq (Boyle et al., 2008), and located footprints of DNA-binding proteins with Wellington (Piper et al., 2013). We intersected DNase I footprints with DNA sequence motifs from JASPAR (Mathelier et al., 2014), UniProbe (Hume et al., 2015), and SwissRegulon (Pachkov et al., 2013) to predict binding sites for specific TFs, focusing on 741 TFs— approximately half of all human TFs—which are expressed in the human brain (Hawrylycz et al., 2012) and have known sequence specificity. We used the union of TF binding sites predicted across all 15 DNase-seq experiments to identify 2,637,487 DNase hypersensitive sites (DHSs), containing 1,121,670 footprints.

Next, we generated an integrative model for TF-target gene interactions in the brain by combining our model of brain TFBSs with evidence of co-expression between TF-gene pairs, using an approach similar to our previous work on TRNs in the mouse brain (Ament et al., 2018a). Briefly, we considered TF-gene pairs where the predicted binding sites for that TF were enriched +/−10 kb from a gene's transcription start site (a window shown to maximize overlap of target gene predictions from DNase footprinting with target gene predictions from chromatin immunoprecipitation sequencing [ChIP-seq]; Plaisier et al., 2016). We applied two algorithms to identify a subset of these TF-gene pairs that are also supported by TF-gene co-expression: (1) Pearson correlation between TF-gene pairs and (2) LASSO regression, in which combinations of TFs are used as predictors of the target gene's expression (Ament et al., 2018a; Bonneau et al., 2006; Greenfield et al., 2013; Haury et al., 2012; Marbach et al., 2012). In the LASSO regression model, we retained only those genes for which the LASSO regression model explained >50% of the variance in gene expression (Ament et al., 2018a).

We trained our model using 2,748 gene expression profiles of microdissected brain tissue from the Allen Human Brain Atlas (Hawrylycz et al., 2012). Each of these microdissected tissue samples represents the expression levels of transcripts in $\sim 10^4$ cells, mapped onto a reference atlas with 862 brain structures (Ding et al., 2016). Thus, the Allen Human Brain Atlas defines a cellular resolution map of gene expression in the human brain, where TF-gene co-expression likely reflects co-variation across brain regions and cell types. We generated separate models from five independent training sets (subsets of the Allen Brain Atlas with brains of different individuals). Then, we merged the models to retain only those TF-gene edges supported by both LASSO and Pearson correlation in more than one training set with consistent direction of interaction (activation or repression). Given the sample sizes in our training sets, these thresholds for correlation strength and reproducibility correspond to a Bonferroni-adjusted p value < 0.02, correcting for 8,219,913 possible TF-gene pairs (see STAR Methods). The final TRN model included 741 TFs and 11,093 target genes. 201,218 TF-target gene interactions were supported by co-expression at this threshold in at least two training sets (Table S1). Our model predicted a median of 18 TFs regulating each target gene (interquartile range: 14–22 TFs). The out-degree of TFs was heavy tailed, with a median of 228 targets per TF and 127 apparent "hub" TFs predicted to regulate >500 genes (Figure S1; Table S1).

We systematically evaluated the accuracy of our model's predictions through comparison to independent datasets. First, we asked whether the predicted TF binding sites from our model are enriched in gene regulatory regions identified in independent datasets. The ROADMAP Epigenomics Consortium generated

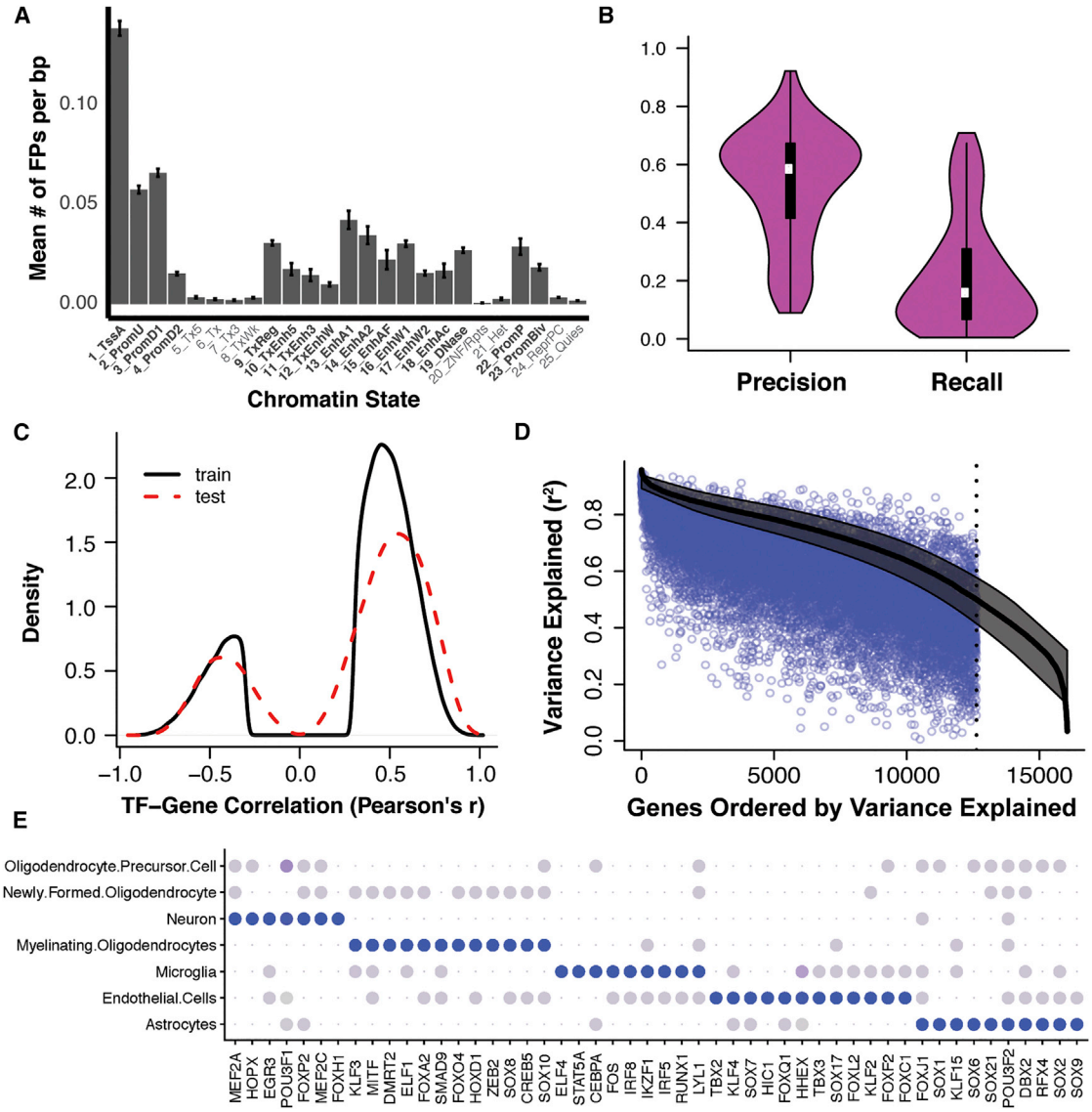


Figure 1. Bioinformatic Validation of DNase I Footprints and Genome-Scale Predictions of TF—Target Gene Interactions in the Human Brain

(A) Density of DNase I footprints within chromatin states predicted with ChromHMM (Kundaje et al., 2015) in independent human brain samples (mean +/− standard error).

(B) Precision and recall of TF target genes predicted from DNase I footprints versus TF target genes predicted from ChIP-seq (target genes defined as those having at least one footprint or at least one ChIP-seq peak +/− 10 kb of the transcription start site).

(C) Distribution of TF-gene correlation coefficients in training data versus independent test sets.

(D) Proportion of gene expression variance explained by the LASSO regression model in five training sets (black line, median; gray shading, range across five training sets) versus holdout samples from the Allen Human Brain Atlas.

(E). Top 50 TFs whose target genes are most strongly enriched in a major brain cell type. Color indicates the scaled $-\log_{10}$ (p value) for the strength of enrichment (Fisher's exact test).

reference maps for chromatin states in the human brain by integrating evidence from genome-wide profiling of 12 chromatin marks in each of 10 human brain samples (Ernst and Kellis, 2012; Kundaje et al., 2015). They identified 25 chromatin states, encompassing regulatory regions (e.g., promoters and enhancers) as well as non-regulatory regions (e.g., quiescent regions and regions that are being actively transcribed). Footprints from our TRN model were >2-fold enriched in chromatin states associated with active gene regulatory elements and depleted in all other chromatin states (Figure 1A; Table S1). Overall, regulatory regions annotated by ROADMAP span 12% of the mappable genome and contain 64% of the footprints from our model (5.3-fold enriched). These results provide independent support for the regulatory activity in the brain for many footprinted regions.

Second, we compared the predicted TF binding sites in our model to TF binding sites identified in TF ChIP-seq experiments. We downloaded uniformly processed ChIP-seq data from the

Gene Transcription Regulation Database (GTRD) for 242 of the 741 TFs in our TRN model (Yevshin et al., 2017). We compared the predicted target genes for each TF supported by ChIP-seq to the predicted target genes from DNase I footprinting. Here, TF-gene pairs are defined as those supported by at least one footprint or at least one ChIP-seq peak +/−10 kb of a gene's TSS; Table S1). None of these ChIP-seq data were from brain tissue, and TFs are known to occupy different binding sites in different tissues, so these analyses are likely to underestimate the accuracy of footprinting. Nonetheless, TF-gene pairs predicted by our footprint model had a median 58.4% precision and 15.9% recall for TF-gene pairs supported by ChIP-seq (Figure 1B). For 198 of the 242 TFs in this analysis, we found significant overlap (Fisher's exact test, false discovery rate [FDR] < 0.05) between the target genes predicted for a TF by DNase I footprints versus ChIP-seq. These results demonstrate the strong accuracy of footprinting, especially considering that the ChIP-seq data are not from brain tissue, but also suggest that additional TF binding sites were missed by our model.

Third, we repeated the ChIP-seq overlap analyses using only those TF-gene interactions that were also supported by TF-gene co-expression. By design, the integrative TRN model has many fewer interactions than the footprint model because we aim to select only those interactions with support from both footprints and TF-gene co-expression. TF-gene pairs in the integrative TRN model had 64.1% precision, similar to the precision of the footprint model, but recall only 3.3% of TF-gene pairs from ChIP-seq. This result clarifies two points. First, our integrative TRN model retains but does not dramatically improve overlap with ChIP-seq compared to footprinting alone. Second, many TF-gene pairs supported by ChIP-seq are not co-expressed in the brain, suggesting that footprinting and TF-gene co-expression provide complementary information about regulatory networks.

Fourth, we asked whether expression correlations between TFs and their predicted target genes hold up in independent transcriptomic datasets. We considered three independent transcriptomic test sets: 893 samples from the Allen Human Brain Atlas held out from TRN reconstruction, 122 samples from eight adult donors in BrainSpan (Li et al., 2018), and 1,231 samples from 134 donors collected by the UK Brain Expression Consortium (Ramasamy et al., 2014). 179,746 of the 201,218 TF-gene edges in our TRN model (89.3%) were supported by a TF-gene correlation of the same magnitude ($|r| > 0.25$) and direction (positive or negative) in at least one of the test sets (Figure 1C; Table S3). As an additional test, we asked whether LASSO regression models learned in training sets could predict gene expression in the holdout samples from the Allen Brain Atlas. This is a more stringent test, since an accurate prediction requires multiple TF-gene pairs to be robust across datasets. The prediction accuracy of LASSO regression models learned in training data was strongly correlated to performance in holdout samples from the Allen Human Brain Atlas (training versus test set accuracy: $r = 0.75$) (Figure 1D; Table S1). The LASSO regression models from training sets explained >50% of the expression variance in holdout samples for 8,369 of the 11,093 genes in the TRN model. Thus, the TF-gene correlation structure is highly reproducible in independent samples.

Fifth, we considered whether reproducibility of TF-gene co-expression could be improved by selecting more stringent thresholds. Of the 201,218 TF-target gene interactions in our TRN model, 112,144, 61,891, and 28,338 interactions were supported by TF-gene co-expression in at least three, four, or five (out of five) training sets, respectively. Of these filtered TF-gene pairs, 112,144 (94.0%), 57,874 (96.7%), and 27,936 (98.6%) were also co-expressed ($|r| > 0.25$, same direction) in at least one of the three independent transcriptomic test sets, compared to 179,746 (89.3%) replicated TF-gene pairs in the full TRN model. 75,372 (67.2%), 45,936 (74.2%), and 22,813 (80.5%) TF-gene pairs, respectively, replicated in at least two test sets, compared to 115,588 (57.4%) replicated TF-gene pairs in the full TRN model. Thus, using more stringent thresholds for TF-gene co-expression provides some improvement in the percent of reproducible interactions but misses many of the reproducible interactions from the full TRN model. Based on these results we chose to use the original (more lenient) thresholds in our primary TRN model. We performed supplementary analyses with the more stringently filtered interactions to confirm the robustness of key results.

Sixth, since many TFs are known to regulate cell-type specific processes in the brain (Rubenstein and Rakic, 2013), we asked whether the predicted target genes for each TF were enriched in specific brain cell types and whether these cell-type enrichments were concordant with the expression patterns of the TFs themselves. We downloaded gene expression profiles of neurons, astrocytes, oligodendrocytes, microglia, or endothelial cells purified from the mouse brain (Zhang et al., 2014). Analyzing these data with the pSI R package, we identified 135 TFs that were expressed specifically in one of the major cell types, as well as 147 TFs whose predicted target genes were enriched for specifically expressed genes in one cell type (Figure 1E; Table S1). 47 TFs could be assigned to a specific cell-type by both methods. Among these 47 TFs, we observed 91% concordance in cell-type assignments (inter-rater reliability: Cohen's kappa = 0.89, $p << 2e{-}16$). Our model correctly assigned known cell-type markers such as *OLIG2* in oligodendrocytes (Zhou et al., 2001) and *NEUROD2* and *NEUROD6* in neurons (Olson et al., 2001; Wu et al., 2005). These results suggest that a subset of TRN modules reflect cell-type-specific processes. TFs whose predicted target genes do not appear to be enriched in a specific cell type may regulate a wide variety of biological functions that are active in multiple cell types. In addition, since these cell-type-specific expression data are from the mouse brain, we may have missed some cell-type-specific TF networks with human-specific expression patterns. TF-gene interactions in our models may also reflect differences in expression across brain regions, which may or may not be cell-type specific.

Key Regulators of Disease-Related Transcriptional Changes in Prefrontal Cortex

Next, we sought to identify TFs that are key regulators of gene expression changes in brain diseases. We focused on gene expression changes in the prefrontal cortex (PFC), a neocortical region that is involved in both emotional control and cognition (Ray and Zald, 2012). Altered PFC structure and function are implicated in a wide range of neurodevelopmental, psychiatric, and neurodegenerative disorders (Bicks et al., 2015; Glausier

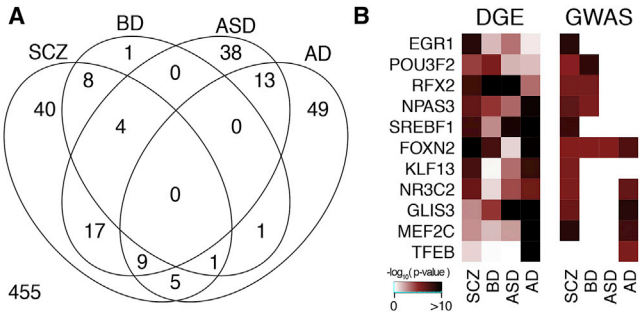


Figure 2. Key Regulators of Disease-Related Transcriptional Changes in PFC

(A) Venn diagram describing overlap of key regulators in schizophrenia (SCZ), bipolar disorder (BD), autism spectrum disorder (ASD), and Alzheimer's disease (AD).

(B) $-\text{Log}_{10}$ (p values) for genetic associations of TF loci and enrichment of each TF's target genes among disease-specific differentially expressed genes for 11 TFs with both a genetic association and a target gene enrichment in one of the four diseases. GWAS p values are based on studies in the GWAS Catalog (MacArthur et al., 2017) (Table S8).

et al., 2014; Hercher et al., 2014; Narayanan et al., 2014; Salat et al., 2001; Weinberger et al., 1986).

We studied post-mortem PFC gene expression profiles from 1,372 human subjects, using data from two to four independent cohorts from each of five common brain diseases: SCZ, BD, MDD, ASD, and AD, as well as non-diseased controls (Akula et al., 2014; Chang et al., 2014; Fromer et al., 2016; Gandal et al., 2018; Parikshak et al., 2016; Reinhart et al., 2015; Torkamani et al., 2010; Voineagu et al., 2011; Wang et al., 2016; Zhang et al., 2013) (Table S2). We identified differentially expressed genes between cases and controls separately in each dataset, using published summary statistics if available or calculating gene expression changes from individual-level data, controlling for covariates of sex, age, post-mortem interval, pH, and RNA integrity. We then tested for over-representation of each TF's target genes among these differentially expressed genes. We considered a TF to be a key regulator in a disease if its predicted target genes were over-represented among the differentially expressed genes in that disease, with a meta-analytic q value < 0.05 across all cohorts and a p value < 0.05 in at least two independent cohorts. We identified 84 key regulators for SCZ, 15 for BD, none for MDD, 81 for ASD, and 78 for AD (Figure 2A; Table S2). Repeating these analyses while restricting our analysis to 112,144 TF-gene interactions with support in at least three (as opposed to two) training sets, we found strongly overlapping results: of 162 key regulator TF-disease associations identified in the restricted model, 143 overlapped TF-disease associations from the full model (inter-rater reliability: Cohen's kappa = 0.75, p value << 2e−16 (Figure S2). Thus, our TRN model revealed associations between key regulator TFs and disease states in the PFC, and these associations are robust to reasonable perturbations in TRN reconstruction methods.

Next, we asked whether key regulators were shared by diseases with overlapping symptoms. We found strong overlap between regulators of disorders with psychotic features (SCZ versus BD, 13 shared TFs, 2 expected by chance, p = 2.5e−11; Figure 2A) and between disorders with a strong neurodevelopmental component (SCZ versus ASD, 30 shared TFs, 9 expected by chance, p = 2.2e−9). By contrast, there was relatively little overlap between regulators of adult-onset psychiatric disorders compared to AD (BD versus AD, 2 shared regulators, 2 expected by chance, p = 0.52; SCZ versus AD, 15 shared regulators, 9 expected by chance, p = 0.04). Taken together, these results suggest that shared key regulator TFs connect diseases with overlapping clinical features.

We sought independent support for key regulator TFs by asking whether these TFs are associated with genetic risk for the same disease. We searched the GWAS catalog (MacArthur et al., 2017) for genetic associations at loci containing TF genes, and we identified 15 instances involving 11 TFs in which a key regulator TF from our TRN model has been reported to be associated with genetic risk for the same disease (Figure 2B; Table S2). This is twice as many instances of overlap between GWAS and TRN regulators as expected by chance (odds ratio = 2.03; p value = 1.6e−2). TFs associated with both genetic and transcriptional changes in the same disease include key regulators for SCZ (*EGR1*, *SREBF1, NPAS3*, *POU3F2*, *RFX2*, *KLF13*, *FOXN2*, and *NR3C2*), BD (*POU3F2*, *NPAS3*, and *FOXN2*), and AD (*MEF2C*, *GLIS3*, *TFEB*, and *NR3C2*). These results highlight the convergence of genetic risk and transcriptional changes on shared TRNs in the brain.

Disease-Associated Genetic Variation Influencing TF Binding Sites

The results above describe putative *trans*-acting effects in which genetic changes at a small number of TF loci are linked to changes in the expression of many downstream target genes. We next sought to elucidate *cis*-acting network perturbations: genetic variants that alter a TF binding site (TFBS), perturbing a single edge in the network. eQTL analyses suggest that such effects are present at most genes, explaining hundreds of GWAS risk loci (Aguet et al., 2017; Fromer et al., 2016; Hauberg et al., 2017; Ward and Kellis, 2016).

To identify TFBS-disrupting SNPs, we intersected our model of human brain TFBSs with genetic variants from Kaviar (Glusman et al., 2011), focusing on SNPs that overlap TFBSs +/− 10 kb of the transcription start site for one of that TF's predicted target genes. We identified 52,705 putative TFBS-disrupting SNPs (Table S3). These SNPs are predicted to modify 67,152 of the 201,218 TF-target gene interactions in our TRN model (the number of modulated TF-target gene interactions is greater than the number of SNPs, since some SNPs overlap the binding sites for more than one TF, and some TFs regulate more than one adjacent gene). These results support the idea that there are widespread effects of non-coding SNPs on gene regulation (Maurano et al., 2012).

To explore the potential impact of these TFBS-disrupting variants on disease risk, we used our model to predict functional variants and target genes at risk loci from published GWAS, focusing on SCZ (PGC., 2014) and AD (Lambert et al., 2013), since these diseases have more genome-wide significant findings than the GWAS of ASD or BD. We identified 13 TFBS-disrupting SNPs in linkage disequilibrium (LD, $r^2 > 0.6$) with lead SNPs (lead SNPs, p < 5e−8) located at 7 of the 108 genome-wide significant SCZ loci reported by the Psychiatric Genomics Consortium (PGC) (PGC, 2014) (Table S3). We also identified 4

TFBS-disrupting SNPs in LD with lead SNPs at 3 of the 19 AD risk loci reported by the International Genomics of Alzheimer's Project (IGAP) (Lambert et al., 2013) (Table S3). 16 of these 17 disease-associated TFBS-disrupting SNPs are also eQTLs associated with changes in the expression of the same gene. Notably, 11 of the 13 SCZ-associated TFBS-disrupting SNPs and 2 of the 4 AD-associated TFBS-disrupting SNPs are predicted to disrupt a binding site for one of the key regulator TFs for the same disease. At SCZ risk loci, these TFs overlap putative binding sites for 18 key regulator TFs. At AD risk loci, these included disrupted binding sites for the key regulator TFs KLF3, KLF11, and PATZ1. Significantly more TFBS-disrupting SNPs at these loci involve key regulator TFs than expected by chance (Fisher's exact test: odds ratio = 1.73, p = 0.03). The enrichment for binding sites of key regulator TFs again highlights the convergence of genetics and gene expression on shared transcriptional networks.

Validation of TF-Disease Associations in Independent Datasets

The analyses above led to numerous predictions about the roles of TF networks in psychiatric and neurodegenerative diseases. We performed additional analyses and experiments to validate key predictions from our model. Here, we describe bioinformatics validation for two of these associations: a role for the MEF2C network in AD and a role for the SREBF1 network in SCZ. In the subsequent section, we describe experimental characterization of a role for a third TF, POU3F2, in BD and SCZ.

Target genes of the myocyte-specific enhancer factor 2C (MEF2C) in our TRN model were enriched for neuron-specific genes (89 of 535 predicted MEF2C target genes; odds ratio = 2.41; p value = 2.5e−11) and for genes that were downregulated in the dorsolateral PFC of individuals with AD (meta-analytic p value = 1.1e−13). Confirming these predictions, we found that genes differentially expressed in the forebrains of mice following conditional *Mef2c* knockout in neurons (Harrington et al., 2016) were enriched for neuron-specific genes (95 of 939 differentially expressed genes; odds ratio = 1.39; p = 4.4e−3) and for genes that were downregulated in AD cases versus controls (meta-analytic p value = 3.7e−4). In addition, *MEF2C* is located at a genome-wide significant risk locus for AD (Lambert et al., 2013). These results suggest that MEF2C has a causal role in the downregulation of neuronal genes in AD.

Target genes of the sterol regulatory element binding transcription factor 1 (*SREBF1*) in our TRN model were enriched for genes expressed specifically in myelinating oligodendrocytes (43 of 440 predicted target genes; odds ratio = 2.8; p value = 3.6e−8) and for genes upregulated in the PFC of SCZ cases versus controls (meta-analytic p value = 2.8e−8). Confirming these predictions, we found that SREBF1 binding sites from ChIP-seq (Yevshin et al., 2017) were enriched +/−10 kb from the transcription start sites of genes expressed specifically in myelinating oligodendrocytes (82 of 1,836 ChIP-seq target genes; odds ratio = 1.39, p value = 6.0e−3) and of genes upregulated in PFC tissue of SCZ cases versus controls (Fromer et al., 2016; Reinhart et al., 2015; Torkamani et al., 2010) (meta-analytic p value = 8.0e−5). Consistent with the known roles of *SREBF1* in lipid biosynthesis, its target genes were enriched for the Gene Ontology terms "lipid binding" (p value = 6.6e−5) and "phospholipid metabolism" (p value = 2.1e−3). Equivalent to *MEF2C*, above, *SREBF1* is located at a genome-wide significant risk locus for SCZ (PGC, 2014). We speculate that genetic perturbations of *SREBF1* contribute to hypomyelination (Kochunov and Hong, 2014) or other lipid biosynthesis-related deficits in SCZ via changes in the expression of SREBF1 target genes.

A Neural Stem Cell- and Astrocyte-Enriched POU3F2-Regulated Network Associated with Bipolar Disorder and Schizophrenia

Among the many TF-disease associations predicted by our model, the association of POU3F2 (POU Class 3 Homeobox 2, also known as *BRN2*) with BD and SCZ stood out because of the convergent evidence from multiple levels of analysis: (1) *POU3F2* was a predicted key regulator of SCZ (p value = 9.7e−6) and BD (p value = 5.2e−6) with predicted target genes most strongly over-represented among upregulated genes in PFC tissue from affected individuals with both diseases; (2) the *POU3F2* genomic locus is associated with risk for BD (Hou et al., 2016; Mühleisen et al., 2014); and (3) a predicted binding site for *POU3F2* was disrupted by an SNP in LD with a lead SNP at a genome-wide significant risk locus for SCZ. Thus, our TRN model implicates POU3F2 in SCZ and BD both through the *cis*-acting effect of a disease-associated variant that disrupts a POU3F2 binding site and through *trans*-acting effects on PFC gene expression. In addition, *POU3F2* has well known roles in neural progenitors and neuronal differentiation (Dominguez et al., 2013), suggesting a tractable, disease-relevant cell type in which to characterize its functions.

To validate a *cis*-acting effect of POU3F2, we characterized the regulatory impact of rs13384219 (hg38, chr2:57907323 A/G). This variant, which has an allele frequency of 0.04 in non-Finnish Europeans (Lek et al., 2016), is in LD (r^2 = 0.65; D′ = 1) with a genome-wide significant lead SNP associated with risk for both SCZ (PGC, 2014) (Figure 3A; lead SNP, rs80256351, p = 3.2e−8; rs13384219, p = 2.1e−5). rs13384219 is located 336 bp upstream of the transcription start site for *VRK2*. Our TRN model predicted a functional effect of this SNP on POU3F2 binding for two reasons. First, the SNP overlaps a DNase I footprint spanning a sequence motif recognized by POU-domain TFs (Figure 3B). Second, *POU3F2* expression was correlated with *VRK2* expression in the Allen Human Brain Atlas (r = 0.67; Figure 3C), suggesting that *POU3F2* and *VRK2* are expressed in overlapping cell types and brain regions.

We used luciferase reporter assays to test the activity of a 436 bp fragment of the *VRK2* promoter containing either the A or G allele of rs13384219. The risk-associated G allele decreased the activity of this *VRK2* promoter fragment in HEK293 cells (p = 4e−24; Figure 3D). We repeated this experiment in combination with POU3F2 overexpression to test the effect of POU3F2 on the activity of the *VRK2* promoter fragment. POU3F2 overexpression caused a dose-dependent decrease in the activity of the *VRK2* promoter fragment with the rs13384219 A allele. By contrast, POU3F2 overexpression caused a dose-dependent increase in the activity of the *VRK2* promoter fragment with the rs13384219 G allele ($P_{SNP \times POU3F2}$ = 8e−4). These results validate our model's predictions that POU3F2 regulates the *VRK2* promoter and that rs13384219 modifies this effect. We note that *POU3F2* and *VRK2* expression are positively correlated in

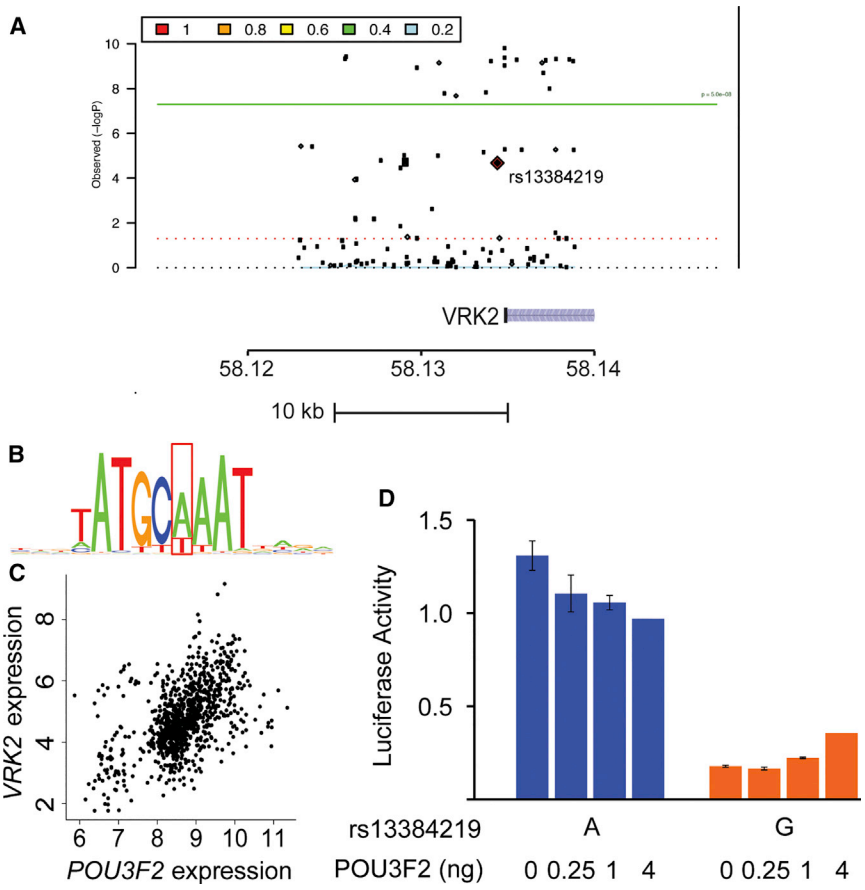


Figure 3. Modulation of a POU3F2 Binding Site by a Schizophrenia-Associated SNP in the *VRK2* Promoter

(A) Region plot of the *VRK2* locus from GWAS of SCZ (PGC, 2014). TRN analysis revealed one SNP at this locus, rs13384219, which overlaps a binding site for a TF that is a predicted regulator of VRK2 and in LD ($r^2 > 0.6$) with a genome-wide significant SNP at this locus.

(B) rs13384219 disrupts a key residue in a sequence motif recognized by POU-domain TFs (the Pou2f2 motif is shown).

(C) *POU3F2* expression is positively correlated with *VRK2* expression in the Allen Human Brain Atlas (r = 0.67).

(D) Dual luciferase reporter assay comparing the activity of a *VRK2* promoter fragment containing either the A or G allele of rs13384219 in HEK293 cells. POU3F2 was overexpressed in combination with transfection of each luciferase construct to assess dose-dependent effects of POU3F2 and interactions with rs13384219 genotype (mean +/− standard error).

the Allen Brain Atlas, whereas our reporter assay suggests a more nuanced relationship. One possibility is that the regulatory context in HEK293 differs from that in the brain. In addition, it is likely that the positive correlation of *POU3F2* and *VRK2* in the Allen Brain Atlas is due to co-expression of these genes in the same cell types and does not directly predict the modulatory effect of *POU3F2* on *VRK2* expression.

To validate *trans*-acting effects of POU3F2 and gain insight into its potential role in psychiatric disease, we overexpressed it in primary human neural stem cells (NSCs). NSCs have low POU3F2 expression, and *in vivo* induction of POU3F2 naturally occurs in early neural precursors, just subsequent to the stage of neurodevelopment at which our NSCs were derived (Dominguez et al., 2013). We transduced NSCs with a POU3F2 lentiviral expression vector and generated microarray gene expression profiles from successfully transfected cells 3 and 10 days following infection. We observed robust overexpression of POU3F2 mRNA and protein at both time points (Figure S3).

First, we asked whether gene expression changes induced by POU3F2 validate predicted target genes from our TRN model. Linear modeling of $\log_2$-transformed, quantile-normalized expression values revealed 2,526 upregulated genes and 1,654 downregulated genes following POU3F2 overexpression (False Discovery Rate (FDR) < 10%; Table S4). 168 of the 592 predicted POU3F2 target genes in our TRN model were differentially expressed, significantly more than expected by chance (odds ratio = 1.41, p value = 2.2e−4). In addition, 265 of the 1,265 genes with a POU3F2 footprint within 10 kb of its TSS were differentially expressed (odds ratio = 1.42; p value = 3.8e−6). Upregulated genes and downregulated genes were independently enriched for target genes in our TRN model and for genes with predicted POU3F2 binding sites ($p < 0.05$). In addition, we observed a trend toward downregulation of *VRK2* after POU3F2 overexpression, consistent with results from luciferase assays, but this was not statistically significant ($p > 0.05$; Figure S3). Therefore, POU3F2 overexpression validated a significantly enriched subset of the POU3F2 target genes predicted by footprinting and TRN modeling. Note that the differentially expressed genes in our POU3F2 overexpression study include both direct and indirect targets of POU3F2, so we do not expect perfect overlap with TRN modeling predictions. The extent of overlap is similar to published reports (with different TFs) comparing target gene predictions from ChIP-seq to TF knockdown experiments (Cusanovich et al., 2014), and the overlap is stronger than has been observed using methods based solely on TF-gene co-expression (Marbach et al., 2012).

Next, we sought to validate the prediction that POU3F2 target genes are enriched for genes that are differentially expressed in the PFC of individuals with SCZ or BD, meta-analyzing the same PFC datasets used previously. The strongest and most consistent finding was that genes downregulated by POU3F2 overexpression in NSCs were enriched for genes that were upregulated in SCZ (Figure 4A; meta-analytic p value = 1.3e−13) and BD (meta-analytic p value = 1.2e−3), validating the predictions from our TRN model. Unexpectedly, we also found that genes that were downregulated in each of these diseases were enriched for genes downregulated after POU3F2 overexpression (SCZ: meta-analytic p value = 3.6e−8; BD: meta-analytic p value = 5.0e−5). We also found an enrichment for genes

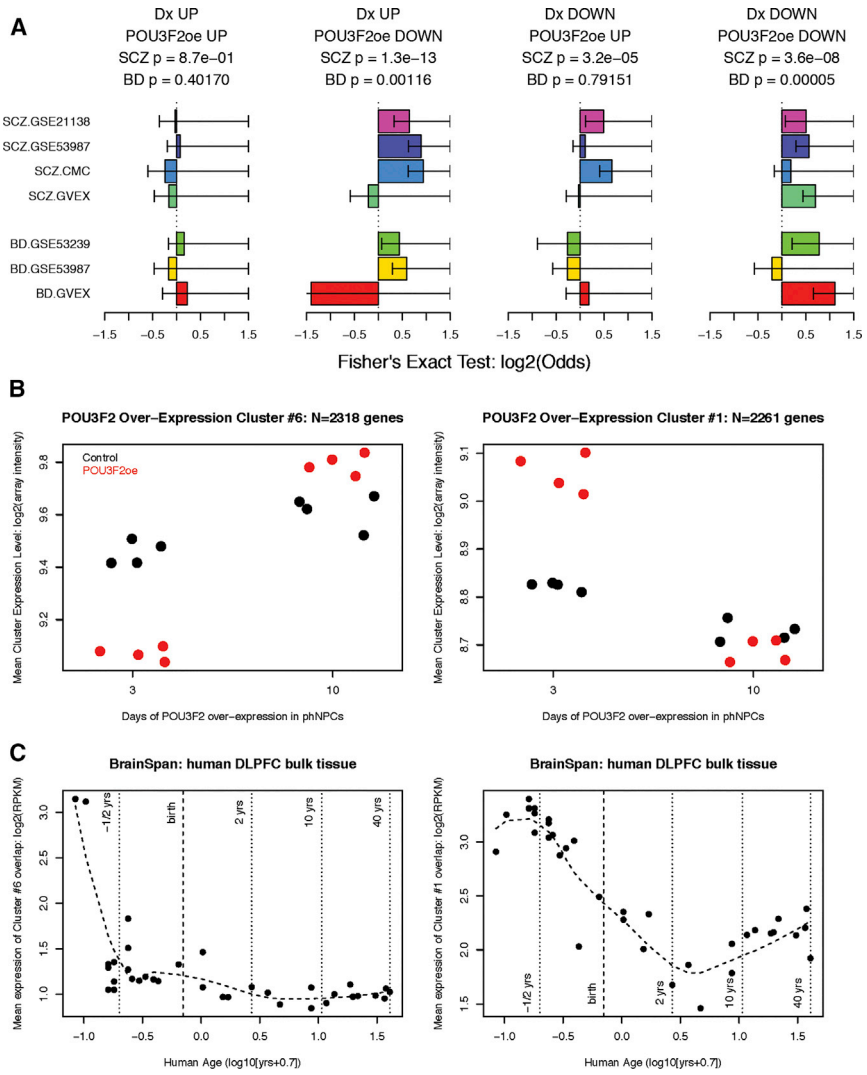


Figure 4. POU3F2 Regulates a Schizophrenia-Related and Bipolar Disorder-Related Gene Network in NSCs

(A) Differentially expressed genes in NSCs overexpressing POU3F2 were enriched for upregulated and downregulated genes in the PFC of individuals with SCZ and BD. Error bars indicate 95% confidence interval for the odds ratio from Fisher's exact test comparing differentially expressed genes in POU3F2 overexpression to differentially expressed geens in cases versus controls from each disease cohort.

(B and C) Each plot shows the mean expression pattern of a gene cluster enriched for differentially expressed genes in NSCs overexpressing POU3F2 or of overlapping clusters in reference data from cortical development. (B) Mean cluster expression level (log_2 array intensity) from genes in clusters C6 and C1 in POU3F2-overexpressing NSCs and controls. (C) Mean expression (log_2 RPKM) of genes in overlapping clusters from RNA-seq of developing dorsolateral PFC from BrainSpan (Li et al., 2018).

upregulated by POU3F2 overexpression among genes downregulated in SCZ (meta-analytics p value = 3.2e−5), but there was no association between these upregulated POU3F2 targets and gene expression changes in BD. Therefore, POU3F2 overexpression in NSCs caused dysregulation of numerous genes that are differentially expressed in the PFC of individuals with SCZ and BD, validating our network prediction that POU3F2 is a key regulator of gene expression changes in these diseases.

We used gene co-expression clustering across multiple datasets to gain perspective on the dynamics of POU3F2 target genes *in vitro* and *in vivo*. We performed k-means clustering (k = 9) to identify clusters of genes with similar expression patterns in our POU3F2 overexpression dataset (Table S4). Four clusters were strongly enriched for genes differentially expressed upon POU3F2 overexpression (Table S4). Clusters C1 and C3 were enriched for upregulated genes (C1: odds ratio = 3.6, p value = 8.0e−74; C3: odds ratio = 10.3, p value = 8.4e−198), with C1 primarily upregulated on day 3 of POU3F2 overexpression and C3 primarily upregulated on day 10. Clusters C6 and C7 were enriched for downregulated genes (C6: odds ratio = 2.3, p value = 8.3e−24; C7: odds ratio = 27.9, p value < 1e−308), with C6 primarily downregulated on day 3, and C7 primarily downregulated on day 10 (Figure 4B). We focused on the clusters associated with gene expression changes at day 3, C1 and C6, because differentially expressed genes at the earlier time point are more likely to represent direct effects of POU3F2.

We explored the expression patterns of these POU3F2-responsive gene clusters in four neurodevelopmental gene expression datasets: (1) RNA sequencing (RNA-seq) of dorsolateral PFC development from BrainSpan (Li et al., 2018); (2) gene expression microarrays of laser-captured regions of the developing macaque cortex (Bakken et al., 2016); (3) single-cell RNA-seq of the developing human cortex (Darmanis et al., 2015); and (4) RNA-seq of *in vitro* differentiation of human-induced pluripotent stem cells toward neurons (van de Leemput et al., 2014).

Genes in C6, downregulated upon POU3F2 overexpression, were enriched for genes involved in the cell cycle (Gene Ontology, p = 5.8e−75). Genes in this cluster overlapped significantly with clusters obtained from *in vivo* neocortical data, showing high expression during the first trimester of *in utero* development and low expression thereafter, in both human and non-human primates (Figures 4C and S4). TRN analysis of the genes in both cluster C6 and the overlapping human DFC cluster indicated the involvement of several TFs known to control cell cycle, including the top hit E2F1 (p = 1.3e−32), which shows high correlation with the DFC cluster itself (r = 0.65, p = 2.5e−5). Single-cell data from the developing human neocortex indicate that these genes are expressed specifically in dividing cells of the prenatal neocortex (Figure S4). Remarkably, a subset of the genes in the C6 cluster, including

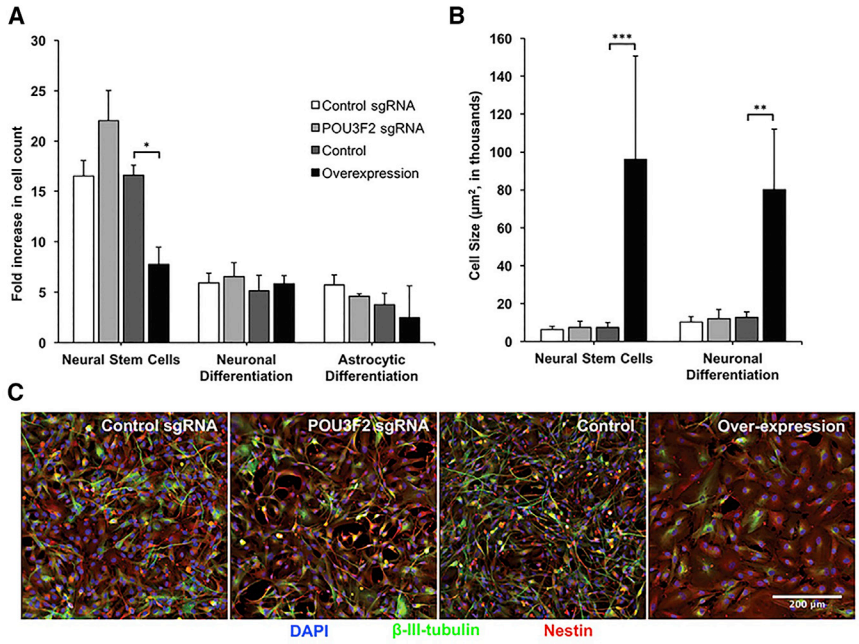


Figure 5. POU3F2 Represses Cellular Proliferation in Primary Human NSCs

We perturbed POU3F2 expression in primary human NSCs by knocking it out with a CRISPR/Cas9 sgRNA or with transgenic overexpression, and then assessed effects on cellular proliferation.

(A) POU3F2 overexpression decreased cellular proliferation in NSCs to levels comparable to proliferation in cells differentiated toward neurons or astrocytes. POU3F2 sgRNA did not significantly influence proliferation (mean +/− stardard error).

(B) POU3F2 overexpression increases cell size, both in NSCs and in cells differentiated toward neurons (mean +/− standard error).

(C) Immunofluorescence images of NSCs and of differentiated neurons, showing effects of POU3F2 overexpression on cell size and morphology.

GLI3, is also highly expressed in astrocytes of the adult cortex (Figure S5).

Genes in C1, upregulated upon POU3F2 overexpression, were enriched for genes involved in transcription (Gene Ontology, p = 4.1e−13). These genes overlapped a cluster of genes whose neocortical expression peaks in the middle of the second trimester and which correlates with POU3F2 expression itself (Figure 4C, r = 0.87). Single-cell data suggest that these genes are expressed in cells of the developing neocortex that are not dividing and that these putative targets of POU3F2 transcriptional activation increase further as neurons mature (Figure S4).

These observations suggest that POU3F2 overexpression in NSCs *in vitro* recapitulates an *in vivo* cellular transition out of an early highly proliferative state and into a subsequent POU3F2-driven state during the second trimester of human neocortical development. Similar cluster overlapping with data from the CORTECON project[60] indicates that the dynamics of these gene clusters are robust in distinct *in vitro* neural differentiation protocols (Figure S4).

To directly test the effects of POU3F2 on the transition from proliferative to non-proliferative cell states, we quantified cell proliferation, cell size, and nucleus size in NSCs and in cells differentiated for 2 weeks toward neurons or astrocytes. We overexpressed POU3F2 as in the gene expression profiling experiments above, and we knocked it out via lentiviral delivery of a CRISPR/Cas9 sgRNA (Figure S3; Table S5). We observed a significant decrease in proliferation in POU3F2 overexpressing NSCs compared to control NSCs (~2-fold decrease, p = 4.3e−4; Figure 5A). In addition, we observed a significant increase in cell size in POU3F2-overexpressing cells compared to controls, both in NSCs (13-fold increase, p < 1e−4) and in cells differentiated toward cortical neurons for 2 weeks (~6-fold increase, p < 1e−3) (Figures 5B and 5C), but the relevance of cell size to neuronal proliferation is unclear. Findings in POU3F2 knockout NSCs were less conclusive, most likely because of the very low baseline levels of POU3F2 expression in NSCs (Figure S3). We observed a trend toward increased proliferation in POU3F2 knockout NSCs, but this was not significant (p = 0.15). Our findings suggest that POU3F2 overexpression inhibits cellular proliferation, possibly via its repression of cell cycle genes and of genes expressed in proliferative NSCs.

DISCUSSION

In this study, we have presented a genome-scale TRN model for the human brain. We identified key regulator TFs that are implicated in gene regulatory changes underlying psychiatric and neurodegenerative disorders, involving changes in gene expression as well as putative *cis*- and *trans*-acting genetic variation. These predictions were systematically validated through ChIP-seq and TF perturbation experiments. Our results describe a convergence of genetic and transcriptional network support for roles of key regulator TFs in multiple diseases, including *cis*- and *trans*-acting effects of the POU3F2-regulated network in SCZ and BD.

Genetic associations and transcriptional changes in a disease often converged on the same key regulator TFs. Convergence on shared networks provides independent support for the associations of these TFs with disease. In addition, these results support the view that transcriptional network changes are a core feature underlying many human diseases. Recently, Boyle et al., (2017) invoked the small-world property of biological networks— nearly all genes in a cell form a single interconnected network of direct and indirect connections— to explain how virtually any gene expressed in a disease-relevant cell type could contribute to genetic risk for a disease (an "omnigenic" model of disease risk). Our results build on this insight and emphasize that an understanding of network structure can reveal a relatively small number of core genes.

Our models identify key regulators of psychiatric and neurodegenerative diseases acting in multiple brain cell types. In AD, neuronally enriched TF networks were often downregulated, whereas microglia-enriched TF networks were upregulated. Most research on SCZ and BD has focused on neuronal

mechanisms. Our results suggest that neuronal gene expression changes in these diseases are mediated by TFs such as *MEF2A*, a key regulator TF in our model whose target genes were enriched in neurons (Table S3) and which has previously been shown to influence synaptic function (Shalizi et al., 2006). Many disease-perturbed networks in psychiatric disorders were enriched in non-neuronal brain cell types. For instance, the oligodendrocyte-enriched *SREBF1* network was implicated by both genetic variation and gene expression changes.

Several key regulator TFs for SCZ and BD are well known for their roles in NSCs, including *HES1*, *MEIS2*, *NKX2-1*, *NPAS3*, *RFX2*, *RFX4*, *SOX2*, and *SOX9*, among others (Rubenstein and Rakic, 2013). Of these, *POU3F2* (Hou et al., 2016; Mühleisen et al., 2014), *SOX2* (PGC, 2014; Won et al., 2016), *NPAS3* (Kamnasaran et al., 2003; Pickard et al., 2009), and *RFX4* (Glaser et al., 2005) are also associated with genetic risk for SCZ or BD, suggesting an etiological role. Interestingly, we identified these NSC-related key regulator TFs based on differential expression in the adult brain. Most of these TFs are also highly expressed in adult astrocytes, and their target genes in our TRN model were enriched for astrocyte-specific genes (Table S1). These pleiotropic effects make it difficult to discern the time points and cell types in which these TFs influence psychiatric disorders. A neurodevelopmental hypothesis is compelling, since adverse events during fetal development are among the strongest non-genetic risk factors for SCZ and mood disorders, and many SCZ risk genes are most highly enriched during fetal brain development (Marenco and Weinberger, 2000; Owen et al., 2011). Changes in adult astrocytes likely also contribute to psychiatric disorders. Several of these NSC and astrocyte TFs have been implicated in cancer stem cells within astrocytic tumors (Suvà et al., 2014) and in de-differentiation of astrocytes and adult neurogenesis (Niu et al., 2015), suggesting functions in proliferative astrocyte states. Understanding how disease risk emerges from changes that occur across all of these cell types and time points is an exciting future challenge.

Convergent lines of evidence led us to focus on POU3F2 as a central regulator of gene expression changes in SCZ and BD. We validated key network predictions that POU3F2 target genes are over-represented among differentially expressed genes in PFC from SCZ and BD cases versus controls and that a risk-associated SNP near *VRK2* influences gene expression through an interaction with POU3F2. We show that POU3F2 represses cell cycle genes and regulates the proliferation of NSCs. These anti-proliferative effects are consistent with two recent reports linking *POU3F2* to neural proliferation phenotypes in stem cell models of ASD (Belinson et al., 2016; Marchetto et al., 2017). Identifying genes involved in BD has been particularly difficult: 30 genome-wide significant risk loci were reported in the largest GWAS to date (Stahl et al., 2017), but biological follow-up experiments have been undertaken for only a few positional candidate genes at these loci (Dao et al., 2010; Leussis et al., 2013). The convergent evidence presented here adds POU3F2 to a very short list of well-supported BD risk genes.

Our network biology approach is broadly applicable to future genetic and genomic studies of human diseases. Our models of binding sites and target genes for 741 TFs, as well as predictions of TFBS-disrupting SNPs, are available online (http://amentlab.igs.umaryland.edu/psych-trn-pfc2016/). These models of tissue-specific gene regulatory networks are likely to improve over time as open chromatin data (DNase-seq, ATAC-seq, and single-nucleus ATAC-seq) become more widely available. These models can be used to predict key regulators of gene expression changes and functional non-coding genetic variation in a growing number of tissues, cell types, and conditions. As such, this study presents a roadmap for understanding brain gene regulation in human health and disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell Lines
- METHOD DETAILS
  - Luciferase Reporter Assay
  - Neural Stem Cell Culture
  - Lentivirus Infections
  - Genome Editing
  - Western Blots
  - qPCR
  - Neural Stem Cell Differentiation
  - Immunofluorescence
  - Microarrays
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Digital Genomic Footprinting
  - FIMO Database
- TRANSCRIPTIONAL REGULATORY NETWORK MODEL
- BIOINFORMATIC VALIDATION OF THE TRN MODEL
- IDENTIFYING KEY REGULATORS OF DISEASE-RELATED GENE EXPRESSION CHANGES
- CELL-TYPE SPECIFIC EXPRESSION ANALYSIS
  - Annotation of TFBS-Disrupting Variants
  - Differentially Expressed Genes in POU3F2 Over-Expression Study
  - Gene Co-Expression Networks
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and five tables and can be found with this article online at https://doi.org/10.1016/j.cels.2019.01.002.

ACKNOWLEDGMENTS

The authors thank Patrick Paddison, Yu Ding, and Chad Toledo (Fred Hutchinson Cancer Research Center) for graciously providing human neural stem cells; Elizabeth Gray, Daniel Stetson, and Kathleen Pestal (University of Washington) for providing LentiCRISPR plasmids and protocols; John Kelsoe and Tatyana Shekhtman (University of California, San Diego) for providing genomic DNA for promoter cloning of *VRK2*; and Nathaniel Peters (W.M. Keck Microscopy Center, University of Washington) for providing immunofluorescence imaging and advice. The author also thank Victor Felix (who maintained resources available at amentlab.igs.umaryland.edu) and Gene Robinson for helpful discussion and providing comments on an early version of this manuscript. This work was supported by a National Science Foundation (United States) Graduate Research Fellowship (J.R.P.), an NARSAD Young Investigator Award from the Brain and Behavior Research Foundation (S.A.A.), the

NIGMS Center for Systems Biology at the Institute for Systems Biology (P50 GM076547; L.H. and N.D.P.), the Big Data for Discovery Science Center of the NIH Big Data to Knowledge program (U54 EB020406; N.D.P. and L.H.), a contract from the NIA (United States; U01 AG046139; N.D.P.), and seed funding from the University of Maryland School of Medicine (United States; S.A.A.).

AUTHOR CONTRIBUTIONS

S.A.A. and J.R.P. designed the experiments. J.R.P., B.B., and D.E.B. performed the experiments. S.A.A., J.R.P., D.E.B., C.C.F., A.M.C., R.T.O., and C.C. analyzed the data. S.A.A., J.R.P., and C.C. wrote the paper. All authors contributed to editing of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 12, 2018
Revised: October 19, 2018
Accepted: January 14, 2019
Published: February 13, 2019

WEB RESOURCES

Allen Brain Atlas, http://human.brain-map.org/static/download

Brain cell type gene expression data, https://web.stanford.edu/group/barres_lab/barreslab_rnaseq.xlsx.

BrainSpan, www.brainspan.org

ENCODE Project, https://www.encodeproject.org

International Genomics of Alzheimer's Project, http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php

JASPAR CORE Vertebrate 2016 collection, http://jaspar.genereg.net/html/DOWNLOAD/

Models of binding sites and target genes, http://amentlab.igs.umaryland.edu/psych-trn-pfc2016/

NIH Blueprint Non-Human Primate Atlas, http://www.blueprintnhpatlas.org

Psychiatric Genomics Consortium GWAS of schizophrenia, https://www.med.unc.edu/pgc/results-and-downloads

R projectR package, https://github.com/genesofeve/projectR

TReNA R package, www.trena.org

REFERENCES

Aguet, F., Brown, A.A., Castel, S.E., Davis, J.R., He, Y., Jo, B., Mohammadi, P., Park, Y., Parsana, P., Segrè, A.V., et al. (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213.

Akula, N., Barb, J., Jiang, X., Wendland, J.R., Choi, K.H., Sen, S.K., Hou, L., Chen, D.T.W., Laje, G., Johnson, K., et al. (2014). RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. Mol. Psychiatry *19*, 1179–1185.

Ament, S.A., Pearl, J.R., Cantle, J.P., Bragg, R.M., Skene, P.J., Coffey, S.R., Bergey, D.E., Wheeler, V.C., MacDonald, M.E., Baliga, N.S., et al. (2018a). Transcriptional regulatory networks underlying gene expression changes in Huntington's disease. Mol. Syst. Biol. *14*, e7435.

Ament, S., Shannon, P., and Richards, M. (2018b). TReNA: Fit transcriptional regulatory networks using gene expression, priors, machine learning.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME suite. Nucleic Acids Res. *43*, W39–W49.

Bakken, T.E., Miller, J.A., Ding, S.-L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Dalley, R.A., Royall, J.J., Lemon, T., et al. (2016). A comprehensive transcriptional map of primate brain development. Nature *535*, 367–375.

Belinson, H., Nakatani, J., Babineau, B.A., Birnbaum, R.Y., Ellegood, J., Bershteyn, M., McEvilly, R.J., Long, J.M., Willert, K., Klein, O.D., et al. (2016). Prenatal β-catenin/Brn2/Tbr2 transcriptional cascade regulates adult social and stereotypic behaviors. Mol. Psychiatry *21*, 1417–1433.

Berto, S., Perdomo-Sabogal, A., Gerighausen, D., Qin, J., and Nowick, K. (2016). A consensus network of gene regulatory factors in the human frontal lobe. Front. Genet. *7*, 31.

Bicks, L.K., Koike, H., Akbarian, S., and Morishita, H. (2015). Prefrontal cortex and social cognition in mouse and man. Front. Psychol. *6*, 1805.

Bonneau, R., Reiss, D.J., Shannon, P., Facciotti, M., Hood, L., Baliga, N.S., and Thorsson, V. (2006). The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol. *7*, R36.

Boyle, A.P., Guinney, J., Crawford, G.E., and Furey, T.S. (2008). F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics *24*, 2537–2538.

Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An expanded view of complex traits: From polygenic to omnigenic. Cell *169*, 1177–1186.

Brichta, L., Shin, W., Jackson-Lewis, V., Blesa, J., Yap, E.-L., Walker, Z., Zhang, J., Roussarie, J.-P., Alvarez, M.J., Califano, A., et al. (2015). Identification of neurodegenerative factors using translatome–regulatory network analysis. Nat. Neurosci. *18*, 1325–1333.

Chang, L.C., Jamain, S., Lin, C.W., Rujescu, D., Tseng, G.C., and Sibille, E. (2014). A conserved BDNF, glutamate- and GABA-enriched gene module related to human depression identified by coexpression meta-analysis and DNA variant genome-wide association studies. PLoS One *9*, e90980.

Cusanovich, D.A., Pavlovic, B., Pritchard, J.K., and Gilad, Y. (2014). The functional consequences of variation in transcription factor binding. PLoS Genet. *10*, e1004226.

Dao, D.T., Mahon, P.B., Cai, X., Kovacsics, C.E., Blackwell, R.A., Arad, M., Shi, J., Zandi, P.P., O'Donnell, P., Bipolar Genome Study (BiGS) Consortium, et al. (2010). Mood disorder susceptibility gene CACNA1C modifies mood-related behaviors in mice and interacts with sex to influence behavior in mice and diagnosis in humans. Biol. Psychiatry *68*, 801–810.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. USA *112*, 7285–7290.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., et al. (2011). FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell *146*, 247–261.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

Ding, S.L., Royall, J.J., Sunkin, S.M., Ng, L., Facer, B.A.C., Lesnar, P., Guillozet-Bongaarts, A., McMurray, B., Szafer, A., Dolbeare, T.A., et al. (2016). Comprehensive cellular-resolution atlas of the adult human brain. J. Comp. Neurol. *524*, 3127–3481.

Dominguez, M.H., Ayoub, A.E., and Rakic, P. (2013). POU-III transcription factors (Brn1, Brn2, and Oct6) influence neurogenesis, molecular identity, and migratory destination of upper-layer cells of the cerebral cortex. Cereb. Cortex *23*, 2632–2643.

Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res. *38*, 4218–4230.

Elkabetz, Y., and Studer, L. (2008). Human ESC-derived Neural Rosettes and Neural Stem Cell Progression. Cold Spring Harb. Symp. Quant. Biol. *73*, 377–387.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216.

Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453.

Gandal, M.J., Haney, J.R., Parikshak, N.N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A.J., Appadurai, V., Buil, A., Werge, T.M., et al. (2018). Shared

molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science *359*, 693–697.

Glaser, B., Kirov, G., Bray, N.J., Green, E., O'Donovan, M.C., Craddock, N., and Owen, M.J. (2005). Identification of a potential bipolar risk haplotype in the gene encoding the winged-helix transcription factor RFX4. Mol. Psychiatry *10*, 920–927.

Glausier, J.R., Fish, K.N., and Lewis, D.A. (2014). Altered parvalbumin basket cell inputs in the dorsolateral prefrontal cortex of schizophrenia subjects. Mol. Psychiatry *19*, 30–36.

Glusman, G., Caballero, J., Mauldin, D.E., Hood, L., and Roach, J.C. (2011). Kaviar: an accessible system for testing SNV novelty. Bioinformatics *27*, 3216–3217.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Greenfield, A., Hafemeister, C., and Bonneau, R. (2013). Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. Bioinformatics *29*, 1060–1067.

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., et al. (2016). Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. *48*, 245–252.

Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhjálmsson, B.J., Xu, H., Zang, C., Ripke, S., Bulik-Sullivan, B., Stahl, E., et al. (2014). Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. Am. J. Hum. Genet. *95*, 535–552.

Harrington, A.J., Raissi, A., Rajkovich, K., Berto, S., Kumar, J., Molinaro, G., Raduazzo, J., Guo, Y., Loerwald, K., Konopka, G., et al. (2016). MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. Elife *5*.

Hauberg, M.E., Zhang, W., Giambartolomei, C., Franzén, O., Morris, D.L., Vyse, T.J., Ruusalepp, A., CommonMind Consortium, Sklar, P., Schadt, E.E., et al. (2017). Large-scale identification of common trait and disease variants affecting gene expression. Am. J. Hum. Genet. *100*, 885–894.

Haury, A.-C., Mordelet, F., Vera-Licona, P., and Vert, J.-P. (2012). TIGRESS: trustful inference of gene REgulation using stability selection. BMC Syst. Biol. *6*, 145.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391–399.

Hercher, C., Chopra, V., and Beasley, C.L. (2014). Evidence for morphological alterations in prefrontal white matter glia in schizophrenia and bipolar disorder. J. Psychiatry Neurosci *39*, 376–385.

Hesselberth, J.R., Chen, X., Zhang, Z., Sabo, P.J., Sandstrom, R., Reynolds, A.P., Thurman, R.E., Neph, S., Kuehn, M.S., Noble, W.S., et al. (2009). Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat. Methods *6*, 283–289.

Hodges, A., Strand, A.D., Aragaki, A.K., Kuhn, A., Sengstag, T., Hughes, G., Elliston, L.A., Hartog, C., Goldstein, D.R., Thu, D., et al. (2006). Regional and cellular gene expression changes in human Huntington's disease brain. Hum. Mol. Genet. *15*, 965–977.

Hou, L., Bergen, S.E., Akula, N., Song, J., Hultman, C.M., Landén, M., Adli, M., Alda, M., Ardau, R., Arias, B., et al. (2016). Genome-wide association study of 40,000 individuals identifies two novel loci associated with bipolar disorder. Hum. Mol. Genet. *25*, 3383–3394.

Huang, K.L., Marcora, E., Pimenova, A.A., Di Narzo, A.F., Kapoor, M., Jin, S.C., Harari, O., Bertelsen, S., Fairfax, B.P., Czajkowski, J., et al. (2017). A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. Nat. Neurosci. *20*, 1052–1061.

Hume, M.A., Barrera, L.A., Gisselbrecht, S.S., and Bulyk, M.L. (2015). UniPROBE, update 2015: new tools and content for the online database of protein-binding microarray data on protein-DNA interactions. Nucleic Acids Res. *43*, D117–D122.

Kamnasaran, D., Muir, W.J., Ferguson-Smith, M.A., and Cox, D.W. (2003). Disruption of the neuronal PAS3 gene in a family affected with schizophrenia. J. Med. Genet. *40*, 325–332.

Khosravi, P., Gazestani, V.H., Pirhaji, L., Law, B., Sadeghi, M., Goliaei, B., and Bader, G.D. (2015). Inferring interaction type in gene regulatory networks using co-expression data. Algorithms Mol. Biol. *10*, 23.

Kichaev, G., Yang, W.Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B. (2014). Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. *10*, e1004722.

Kochunov, P., and Hong, L.E. (2014). Neurodevelopmental and neurodegenerative models of schizophrenia: white matter at the center stage. Schizophr. Bull. *40*, 721–728.

Koohy, H., Down, T.A., Spivakov, M., and Hubbard, T. (2014). A comparison of peak callers used for DNase-seq data. PLoS One *9*, e96303.

Köster, J., and Rahmann, S. (2012). Snakemake—a scalable bioinformatics workflow engine. Bioinformatics *28*, 2520–2522.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. *45*, 1452–1458.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Leussis, M.P., Berry-Scott, E.M., Saito, M., Jhuang, H., de Haan, G., Alkan, O., Luce, C.J., Madison, J.M., Sklar, P., Serre, T., et al. (2013). The ANK3 bipolar disorder gene regulates psychiatric-related behaviors that are modulated by lithium and stress. Biol. Psychiatry *73*, 683–690.

Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., et al. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science *362*, eaat7615.

Li, Y., and Kellis, M. (2016). Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases. Nucleic Acids Res. *44*, e144.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., et al. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res. *45*, D896–D901.

Marbach, D., Costello, J.C., Küffner, R., Vega, N.M., Prill, R.J., Camacho, D.M., Allison, K.R., DREAM5 Consortium, Kellis, M., Collins, J.J., et al. (2012). Wisdom of crowds for robust gene network inference. Nat. Methods *9*, 796–804.

Marchetto, M.C., Belinson, H., Tian, Y., Freitas, B.C., Fu, C., Vadodaria, K.C., Beltrao-Braga, P.C., Trujillo, C.A., Mendes, A.P.D., Padmanabhan, K., et al. (2017). Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. Mol. Psychiatry *22*, 820–835.

Marenco, S., and Weinberger, D.R. (2000). The neurodevelopmental hypothesis of schizophrenia: following a trail of evidence from cradle to grave. Dev. Psychopathol *12*, 501–527.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. *42*, D142–D147.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

Mühleisen, T.W., Leber, M., Schulze, T.G., Strohmaier, J., Degenhardt, F., Treutlein, J., Mattheisen, M., Forstner, A.J., Schumacher, J., Breuer, R.,

et al. (2014). Genome-wide association study reveals two new risk loci for bipolar disorder. Nat. Commun. *5*, 3339.

Narayanan, M., Huynh, J.L., Wang, K., Yang, X., Yoo, S., McElwee, J., Zhang, B., Zhang, C., Lamb, J.R., Xie, T., et al. (2014). Common dysregulation network in the human prefrontal cortex underlies two neurodegenerative diseases. Mol. Syst. Biol. *10*, 743.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Niu, W., Zang, T., Smith, D.K., Vue, T.Y., Zou, Y., Bachoo, R., Johnson, J.E., and Zhang, C.L. (2015). SOX2 reprograms resident astrocytes into neural progenitors in the adult brain. Stem Cell Rep. *4*, 780–794.

Olson, J.M., Asakura, A., Snider, L., Hawkes, R., Strand, A., Stoeck, J., Hallahan, A., Pritchard, J., and Tapscott, S.J. (2001). NeuroD2 is necessary for development and survival of central nervous system neurons. Dev. Biol. *234*, 174–187.

Owen, M.J., O'Donovan, M.C., Thapar, A., and Craddock, N. (2011). Neurodevelopmental hypothesis of schizophrenia. Br. J. Psychiatry *198*, 173–175.

Pachkov, M., Balwierz, P.J., Arnold, P., Ozonov, E., and van Nimwegen, E. (2013). SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. Nucleic Acids Res. *41*, D214–D220.

Parikshak, N.N., Swarup, V., Belgard, T.G., Irimia, M., Ramaswami, G., Gandal, M.J., Hartl, C., Leppa, V., Ubieta, L.T., Huang, J., et al. (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. Nature *540*, 423–427.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC) (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.

Network and Pathway Analysis Subgroup of the Psychiatric Genomic Consortium (PGC) (2015). Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat. Neurosci. *18*, 199–209.

Pickard, B.S., Christoforou, A., Thomson, P.A., Fawkes, A., Evans, K.L., Morris, S.W., Porteous, D.J., Blackwood, D.H., and Muir, W.J. (2009). Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. Mol. Psychiatry *14*, 874–884.

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res. *41*, e201.

Plaisier, C.L., O'Brien, S., Bernard, B., Reynolds, S., Simon, Z., Toledo, C.M., Ding, Y., Reiss, D.J., Paddison, P.J., and Baliga, N.S. (2016). Causal mechanistic regulatory network for glioblastoma deciphered using systems genetics network analysis. Cell Syst. *3*, 172–186.

Psychiatric GWAS Consortium Bipolar Disorder Working Group. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat. Genet. *43*, 977–983.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Ramasamy, A., Trabzuni, D., Guelfi, S., Varghese, V., Smith, C., Walker, R., De, T.; UK Brain Expression Consortium, North American Brain Expression Consortium, and Coin, L., et al. (2014). Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat. Neurosci. *17*, 1418–1428.

Ray, R.D., and Zald, D.H. (2012). Anatomical insights into the interaction of emotion and cognition in the prefrontal cortex. Neurosci. Biobehav. Rev. *36*, 479–501.

Reinhart, V., Bove, S.E., Volfson, D., Lewis, D.A., Kleiman, R.J., and Lanz, T.A. (2015). Evaluation of TrkB and BDNF transcripts in prefrontal cortex, hippocampus, and striatum from subjects with schizophrenia, bipolar disorder, and major depressive disorder. Neurobiol. Dis. *77*, 220–227.

Rubenstein, J.L., and Rakic, P. (2013). Patterning and Cell Type Specification in the Developing CNS and PNS: Comprehensive Developmental Neuroscience (Academic Press).

Salat, D.H., Kaye, J.A., and Janowsky, J.S. (2001). Selective preservation and degeneration within the prefrontal cortex in aging and Alzheimer disease. Arch. Neurol. *58*, 1403–1408.

Sanders, S.J., He, X., Willsey, A.J., Ercan-Sencicek, A.G., Samocha, K.E., Cicek, A.E., Murtha, M.T., Bal, V.H., Bishop, S.L., Dong, S., et al. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron *87*, 1215–1233.

Seifuddin, F., Pirooznia, M., Judy, J.T., Goes, F.S., Potash, J.B., and Zandi, P.P. (2013). Systematic review of genome-wide gene expression studies of bipolar disorder. BMC Psychiatry *13*, 213.

Shalizi, A., Gaudillière, B., Yuan, Z., Stegmüller, J., Shirogane, T., Ge, Q., Tan, Y., Schulman, B., Harper, J.W., and Bonni, A. (2006). A calcium-regulated MEF2 SUMOylation switch controls postsynaptic differentiation. Science *311*, 1012–1017.

Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G., et al. (2016). Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat. Neurosci. *19*, 571–577.

Stahl, E., Forstner, A., McQuillin, A., Ripke, S., Ophoff, R., Scott, L., Cichon, S., Andreassen, O.A., Sklar, P., Kelsoe, J., et al. (2017). Genomewide association study identifies 30 loci associated with bipolar disorder. bioRxiv, 173062.

Sugathan, A., Biagioli, M., Golzio, C., Erdin, S., Blumenthal, I., Manavalan, P., Ragavendran, A., Brand, H., Lucente, D., Miles, J., et al. (2014). *CHD8* regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. Proc. Natl. Acad. Sci. USA *111*, E4468–E4477.

Sun, Y., Pollard, S., Conti, L., Toselli, M., Biella, G., Parkin, G., Willatt, L., Falk, A., Cattaneo, E., and Smith, A. (2008). Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. Mol. Cell. Neurosci *38*, 245–258.

Suvà, M.L., Rheinbay, E., Gillespie, S.M., Patel, A.P., Wakimoto, H., Rabkin, S.D., Riggi, N., Chi, A.S., Cahill, D.P., Nahed, B.V., et al. (2014). Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. Cell *157*, 580–594.

Torkamani, A., Dean, B., Schork, N.J., and Thomas, E.A. (2010). Coexpression network analysis of neural tissue reveals perturbations in developmental processes in schizophrenia. Genome Res. *20*, 403–412.

van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L., et al. (2014). CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron *83*, 51–68.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., and Geschwind, D.H. (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature *474*, 380–384.

Wang, M., Roussos, P., McKenzie, A., Zhou, X., Kajiwara, Y., Brennand, K.J., De Luca, G.C., Crary, J.F., Casaccia, P., Buxbaum, J.D., et al. (2016). Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. Genome Med *8*, 104.

Ward, L.D., and Kellis, M. (2016). HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res. *44*, D877–D881.

Weinberger, D.R., Berman, K.F., and Zec, R.F. (1986). Physiologic dysfunction of dorsolateral prefrontal cortex in schizophrenia: I. Regional cerebral blood flow evidence. Arch. Gen. Psychiatry *43*, 114–124.

Whitton, L., Apostolova, G., Rieder, D., Dechant, G., Rea, S., Donohoe, G., and Morris, D.W. (2018). Genes regulated by SATB2 during neurodevelopment contribute to schizophrenia and educational attainment. PLoS Genet. *14*, e1007515.

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

Won, H., de la Torre-Ubieta, L., Stein, J.L., Parikshak, N.N., Huang, J., Opland, C.K., Gandal, M.J., Sutton, G.J., Hormozdiari, F., Lu, D., et al. (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. Nature *538*, 523–527.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. *50*, 668–681.

Wright, C., Gupta, C.N., Chen, J., Patel, V., Calhoun, V.D., Ehrlich, S., Wang, L., Bustillo, J.R., Perrone-Bizzozero, N.I., and Turner, J.A. (2016). Polymorphisms in MIR137HG and microRNA-137-regulated genes influence gray matter structure in schizophrenia. Transl. Psychiatry *6*, e724.

Wu, S.X., Goebbels, S., Nakamura, K., Nakamura, K., Kometani, K., Minato, N., Kaneko, T., Nave, K.A., and Tamamaki, N. (2005). Pyramidal neurons of upper cortical layers generated by NEX-positive progenitor cells in the subventricular zone. Proc. Natl. Acad. Sci. USA *102*, 17172–17177.

Xia, H., Jahr, F.M., Kim, N.K., Xie, L., Shabalin, A.A., Bryois, J., Sweet, D.H., Kronfol, M.M., Palasuberniam, P., McRae, M., et al. (2018). Building a schizophrenia genetic network: transcription factor 4 regulates genes involved in neuronal development and schizophrenia risk. Hum. Mol. Genet. *27*, 3246–3256.

Yevshin, I., Sharipov, R., Valeev, T., Kel, A., and Kolpakov, F. (2017). GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. Nucleic Acids Res. *45*, D61–D67.

Zaharia, M., Bolosky, W.J., Curtis, K., Fox, A., Patterson, D., Shenker, S., Stoica, I., Karp, R.M., and Sittler, T. (2011). Faster and more accurate sequence alignment with SNAP. https://arxiv.org/abs/1111.5572.

Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell *153*, 707–720.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article17.

Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keeffe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., 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.

Zhou, Q., Choi, G., and Anderson, D.J. (2001). The bHLH transcription factor Olig2 promotes oligodendrocyte differentiation in collaboration with Nkx2.2. Neuron *31*, 791–807.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-POU3F2	Abcam	ab137469
Mouse monoclonal anti-FLAG	Sigma	#F1804 RRID:AB_262044
Rabbit polyclonal anti-GAPDH	Abcam	ab37168
Mouse monoclonal anti-hNestin	R&D	MAB1259 RRID:AB_2251304
Rabbit monoclonal anti-beta-Tubulin	Cell Signaling	5568 RRID:AB_10692510
Mouse monoclonal anti-GFAP	Cell Signaling	3670 RRID:AB_561049
Bacterial and Virus Strains		
TOP10 Competent Cells	ThermoFisher	C404006
pRRL-*POU3F2*-gRNA-Cas9-T2A-puro	This paper	NA
pRRL-Cas9-T2A-puro	This paper	NA
pRRL-SFFV-*POU3F2*-Myc-DDK	This paper	NA
Chemicals, Peptides, and Recombinant Proteins		
EGF	Peprotech	AF-100-15
FGF-2	Peprotech	100-18B
Laminin	Sigma	#L2020
B-27 supplement	ThermoFisher	17504-044
N-2 supplement	ThermoFisher	A1370701
Puromycin	ThermoFisher	A1113803
dibutyryl cAMP	Sigma	D0260
SuperSignal West Dura substrate	ThermoFisher	#34075
Critical Commercial Assays		
Lipofectamine 2000	ThermoFisher	#1168019
Dual-Luciferase® Reporter Assay	Promega	#E1910
Surveyor® Assay	IDT	#706020
Qiagen miRNeasy kit	Qiagen	#217004
SurePrint G3 Human Gene Expression 8x60K v2 Microarray	Agilent	#G4851B
Deposited Data		
Raw and normalized microarray data	This paper	GSE102122
Supplementary datasets	This paper	http://amentlab.igs.umaryland.edu/psych-trn-pfc2016/
DNase-seq data	ENCODE project	www.encodeproject.org
Human reference genome GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/
JASPAR CORE Vertebrate 2016 collection	JASPAR	http://jaspar.genereg.net/html/DOWNLOAD/
Allen Human Brain Atlas	Ding et al., 2016	http://human.brain-map.org/static/download
Brain cell type gene expression data	Zhang et al., 2014	https://web.stanford.edu/group/barres_lab/barreslab_rnaseq.xlsx
BrainSpan adult donor expression data	Brainspan	www.brainspan.org
UK Brain Gene Expression Atlas	Ramasamy et al., 2014	http://www.braineac.org/
CommonMind Consortium differentially expressed genes in prefrontal cortex from SCZ cases vs. controls	Fromer et al., 2016	https://www.nimhgenetics.org/available_data/commonmind/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Prefrontal cortex gene expression from SCZ cases vs. controls	Torkamani et al., 2010	GEO: GSE21138
Prefrontal cortex gene expression from SCZ, bipolar disorder, and major depression cases vs. controls	Reinhart et al., 2015	GEO: GSE53987
Prefrontal cortex gene expression from bipolar disorder cases vs. controls	Akula et al., 2014	GEO: GSE53239
Prefrontal cortex gene expression from major depression cases vs. controls	Chang et al., 2014	GEO: GSE54567/GEO: GSE54568
Prefrontal cortex gene expression from autism cases vs. controls	Voineagu et al., 2011	GEO: GSE28521
Prefrontal cortex gene expression from autism cases vs. controls	Parikshak et al., 2016	GEO: GSE89057
Prefrontal cortex gene expression from Alzheimer's disease cases vs. controls	Narayanan et al., 2014	GEO: GSE33000
Prefrontal cortex gene expression from Alzheimer's disease cases vs. controls	Wang et al., 2016	GEO: GSE84422
CORTECON RNA-seq data	van de Leemput et al., 2014	cortecon.neuralsci.org
Single-cell RNA-seq of human cortex	Darmanis et al., 2015	GEO: GSE67835
NIH Blueprint Atlas of Gene Expression in the Non-Human Primate brain	Bakken et al., 2016	http://www.blueprintnhpatlas.org
Uniformly processed ChIP-seq of human transcription factors	Yevshin et al., 2017	http://gtrd.biouml.org/
Experimental Models: Cell Lines		
HEK293T (Female, transformed cell line)	ATCC	#CRL-3216 RRID:CVCL_0063
HEK293 (Female, transformed cell line)	ATCC	#CRL-1573 RRID:CVCL_0045
CB660 human neural stem cells (Female, primary cell line from embryonic day 50-55 fetal cortical dissection)	Sun et al., 2008	NA
Lenti-transduced POU3F2 overexpressing hNSCs	This paper	NA
Oligonucleotides		
Guide RNA sequence for CRISPR/Cas9 (see supplemental section on genome editing)	This paper	NA
Oligos for PCR amplification of gene edit (see supplemental section on genome editing)	This paper	NA
Recombinant DNA		
pGL4.10[*luc2*]	Promega	#E6651
pRL-CMV Renilla	Promega	#E2261
pVSV-G	Addgene	#36399
psPAX-2	Addgene	#12260
Software and Algorithms		
TReNA R package	Ament et al., 2018b	www.trena.org
projectR R package	NA	https://github.com/genesofeve/projectR
Image J software	FIJI	https://fiji.sc/
SNAP aligner	Zaharia et al., 2011	http://snap.cs.berkeley.edu/
FIMO	Grant et al., 2011	http://meme-suite.org/tools/fimo
pSI R package	Dougherty et al., 2010	https://cran.r-project.org/web/packages/pSI/index.html
bedtools	Quinlan, and Hall, 2010	http://bedtools.readthedocs.io/en/latest/
F-seq	Boyle et al., 2008	http://fureylab.web.unc.edu/software/fseq/
Wellington	Piper et al., 2013	https://github.com/cyversewarwick/wellington-footprint

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Seth A. Ament (sament@som.umaryland.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

We used the 'CB660' primary human neural stem cell line, originally derived in the laboratory of Steven Pollard (Sun et al., 2008). This is a primary cell line with a normal XX (female) karyotype, derived from embryonic day 50-55 fetal cortical dissection. Cells were obtained at passage 35 as a gift from Patrick Paddison (Fred Hutchinson Cancer Research Center). Expression of the neural stem cell markers *SOX2* and *NESTIN* was confirmed by qPCR and immunofluorescence. In addition, we used HEK293T cells, obtained from ATCC (CRL-11268™).

METHOD DETAILS

Luciferase Reporter Assay

VRK2 promoter sequences were amplified from the genomic DNA of an individual whose genome was heterozygous at this position based on whole-genome sequencing. Amplified DNA was cloned into the pGL4.10[*luc2*] reporter plasmid (Promega #E6651) upstream of the *luc* transcription start site. Reporter constructs were co-transfected with a pRL-CMV Renilla vector (ratio of 100 ng luciferase: 5 ng renilla control plasmid) into HEK293 cells using Lipofectamine 2000. After 48 hours, cells were harvested and luciferase and renilla activity was assayed on a Synergy H4 Plate Reader using the Dual-Luciferase® Reporter Assay (Promega #E1910). All reported values were normalized to renilla co-transfection controls. All experiments were performed with 3 biological replicates per condition in a 24-well plate format. Results are representative of at least 2 independent experiments. Barplots are presented as mean +/- s.d.

Neural Stem Cell Culture

Neural stem cell (NSC) lines were grown in neural expansion media (NEM) supplemented with EGF and FGF-2 (20 ng/ml) (PeproTech) on laminin (Sigma L2020) -coated polystyrene plates and passaged as previously described (Sun et al., 2008). Neural expansion media (NEM): NeuroCult NS-A Basal Medium (Stem Cell Technologies), DMEM/F12, Antibiotic-Antimycotic, GlutaMax, B-27 supplement, N-2 supplement, 1:1000 EGF, and 1:1000 FGF.

Lentivirus Infections

VSV-G pseudotyped, self-inactivating lentivirus was prepared by transfecting 293T cells with 1.5 μg pVSV-G, 3 μg psPAX-2, and 6 μg pRRL lentiviral vectors - either lentiCRISPR or lenti-SFFV-*POU3F2*-Myc-DDK. Lentiviral supernatants were collected 48 hours later and transferred to human neural stem cells dishes (MOI = ~ 1). Positively transduced cells were selected with 0.6 ug/mL puromycin for 3 days. Gene editing was evaluated using the Surveyor® Assay (Integrated DNA Technologies). Overexpression was evaluated using qRT-PCR and Western blotting.

Genome Editing

POU3F2 was edited in neural stem cells using a custom guide site targeting the following sequence: ATCGTGCACGCCGA GCCGCCCGG. Genomic DNA was extracted from transduced and puromycin-selected cells (*POU3F2* targeted and non-targeting guide site control) using Thermo genomic DNA purification kit (K0512) and amplified using AccuPrime *Taq* DNA polymerase. The following primers were used to amplify a 431 bp region framing the target site for analysis of editing using the Surveyor® Mutation Detection Kit (IDT #706025): POU3F2_Forward, AGAGCGAGAAGGAGGGAGAG; POU3F2_Reverse, GTGATCCACTGGTGAGC GTG. Four microliters of PCR product were used for TOPO cloning for sequencing (Thermo #K457501) and transformed into TOP10 cells. 34 colonies were picked, grown up, and plasmid DNA extracted (Qiagen Miniprep Kit). Inserts were sequenced at GeneWiz, and percentage gene editing in cell population was determined from alignment to reference sequence. Results from these validation experiments are shown in Figure S3 and Table S5, indicating ~50% editing efficiency.

Western Blots

We used Western blots and qPCR to evaluate POU3F2 overexpression. Cells were harvested following lentiviral infection and selection, washed with PBS, and protein extracted using RIPA buffer on ice for 20 minutes. 30 ul of Nupage LDS buffer was added to 100 ul of sample (in RIPA) and boiled at 80 degrees C for 10 min. 12 microliters was loaded onto 4-12% Bis-Tris gel and run for 1 hour at 180 volts. iBlot transfer system was used to transfer to PVDF membrane, according to manufacturer's instructions. Membrane was cut between 51 and 39 kDa bands indicated by See Blue Plus 2 pre-stained ladder. The following commercial antibodies were used: POU3F2 (Abcam ab137469, 1:1000), FLAG (Sigma #F1804, 1:5000), and GAPDH (Abcam ab37168, 1:5000). A FluorChem was used to image the blots using SuperSignal West Dura chemiluminescent substrate (Thermo #34075). Results are shown in Figure S3.

qPCR

Total RNA was extracted using Qiagen miRNeasy kit. cDNA was reverse transcribed using the VILO mastermix kit (Thermo #11755050). A multiplex qPCR assay was run in triplicate for each sample, from duplicate RNA extractions of each modified or control NSC cell line. Multiplex reactions contained SsoFast Universal Probes mix, POU3F2 assay (FAM; IDT custom primer/probe assay) and ACTB assay (HEX; IDT #Hs.PT.56a.40703009). Samples were run on BioRad CFX96 Real-Time system and quantified using the CFX software. Results are shown in Figure S3.

Neural Stem Cell Differentiation

Protocols for NSC differentiation toward astrocytes and neurons were modified from previously reported methods (Elkabetz and Studer, 2008). Following cell count, 15,000 cells were seeded into each well on 4-well chamber slides (Thermo #154526PK). Cells were treated with either neuronal differentiation media (NeuroCult NS-A Basal Medium Stem Cell Technologies, 2% B-27 Serum-Free Supplement, 1% GlutaMax, 1% Antibiotic-Antimycotic) or astrocyte differentiation media (DMEM L-glutamine, high glucose, 1% N-2 supplement, 1% GlutaMax, 1% FBS). Control cells were maintained in NEM. Media was changed every 3-4 days over an 18-day differentiation period. After one week, 0.05 mM dibutyryl cAMP was added to neuronal differentiation media for the remaining differentiation.

Immunofluorescence

Cells were fixed for 10 minutes in warm 4% paraformaldehyde. Immediately following fixation, standard IF protocol (Cell Signaling Technologies) was conducted with citrate buffer antigen retrieval. Both primary and secondary antibodies were allowed to incubate at 4 degrees overnight. To detect NSC, neuronal, and glial markers we used Nestin (R&D anti-hNestin purified mouse monoclonal IgG #MAB1259), b-III-tubulin (CST beta3-Tubulin Rabbit mAb #5568), and GFAP antibodies (CST GFAP (GA5) mouse mAb #3670S), respectively. Secondary antibodies were purchased from Cell Signaling Technologies (Anti-Mouse IgG - Alexa Fluor 647 conjugate #4410, Anti-Rabbit IgG - Alexa Fluor 488 conjugate #4412, and Anti-Rabbit IgG - Alexa Fluor 555 conjugate #4413). In addition, we imaged cell nuclei with DAPI (CST 4063S).

Image J software (FIJI) was used for all analyses. Images for quantifying cell size were taken on a Leica DM IRBE with a 40X NA 1.25 Oil Immersion objective at University of Washington's W.M. Keck Microscopy Center. A total area of 0.553 mm^2 per cell line per condition was visualized by combining 18 individual images within the imaging program MetaMorph. Cell size was quantified by tracing and recording the area of a randomly selected subset of cells after immunofluorescence staining using the marker that best captured the entire cell (Nestin for NSC state, b-III-tubulin for neuronal differentiation). Cell size was not quantified for astrocytic differentiation due to poor resolution of cell boundaries. Cell proliferation was quantified by using DAPI to count the number of nuclei within an image of each well that was 2.5% of the total well (4.6 mm^2 sampled of 1.7 cm^2 total well area). Images for quantifying cell size were taken on a Leica DM IRBE with a 10X objective at University of Washington's W.M. Keck Microscopy Center. We calculated the fold change in number of cells by extrapolating the number of nuclei in the imaged area to the number of cells total in each well. Finally, given the seed density, the fold increase was quantified.

Microarrays

Total RNA was extracted using the miRNeasy extraction kit (Qiagen) at three days and ten days post-puromycin selection of lentiviral transduced hNSCs. Two individual RNA extractions were performed from each plate at each time point, and each viral construct was transduced into two plates of cells. Gene expression was quantified using SurePrint G3 Human Gene Expression 8x60K v2 Microarray (#G4851B). Samples overexpressing POU3F2 were compared to samples that were transduced with a control vector that did not contain the *POU3F2* coding sequence.

QUANTIFICATION AND STATISTICAL ANALYSIS

Digital Genomic Footprinting

We used publicly available data from 15 DNase-seq experiments with human brain tissue generated by the ENCODE project (Table S1). FASTQ files were downloaded from www.encodeproject.org in January, 2016. FASTQ files were aligned to the human genome (GRCh38) using the SNAP algorithm (Zaharia et al., 2011) with a seed size of 16. As the length of reads was variable, the number of mismatches allowed was adjusted accordingly as a parameter of SNAP. We combined the aligned reads from all FASTQ files associated with each experiment into a combined BAM file. To identify regions of open chromatin from the aligned BAM files, we used F-seq (Boyle et al., 2008) with parameters optimized for DNase-seq following Koohy et al. (2014). Using the output from F-seq, we identified putative footprints using Wellington (Piper et al., 2013). Wellington was run with standard settings and -fdrlimit set to -1.

FIMO Database

Genomic instances of sequence motifs recognized by mammalian TFs were identified in the human genome (GRCh38) using FIMO (Grant et al., 2011). We used JASPAR CORE Vertebrate 2016 collection (http://jaspar.genereg.net/html/DOWNLOAD/), and additional motifs from the HOMOCOMO and Swiss Regulon databases that match TF families not included in JASPAR CORE. Motif point-weight matrices were downloaded with the MEME suite (Bailey et al., 2015). The footprints identified from Wellington were

intersected with these motif instances from FIMO. To maximize coverage, and because of the potential imprecise nature of footprints, if any part of a known motif overlapped with a single base of the footprint, an entry was created. We assigned motifs to their cognate TFs, as well as to additional TFs in the same DNA-binding domain family from the TFClass database (Wingender et al., 2013).

TRANSCRIPTIONAL REGULATORY NETWORK MODEL

We predicted TF-target gene interactions by integrating brain-specific DNase-seq genomic footprints with human brain gene expression profiles from the Allen Brain Atlas (http://human.brain-map.org/static/download). The Allen Brain Atlas consists of 2,748 microarray gene expression profiles of microdissected tissue from six human brains. We applied two gene expression-based methods to predict active TF-target gene interactions among these candidate regulators: (1) Pearson correlation and (2) LASSO regression (see below for details). We constructed separate models using the gene expression data from each of five brains in the ABA, leaving the data from the last of the six Allen Brain Atlas brains as a test set. Finally, we created a consensus model, retaining TF-target gene interactions that were selected by both the Pearson and LASSO methods in at least two of the five brains.

We selected candidate regulators for each gene based on genomic footprinting, as follows. We counted the number of footprints for each TF in proximal and distal regions around each gene's canonical transcription start site (TSS; ENSEMBL v83 gene models), with proximal regions defined as +/- 10kb from the TSS and distal regions defined as +/- 1Mb from the target gene's TSS and >10kb from any gene's TSS. TFs with low-complexity sequence motifs generally have large numbers of motif instances. To reduce the bias of our models toward TFs with low-complexity motifs, we quantile normalized the footprint counts for each TF across all genes. We selected as candidate regulators those TFs that had normalized footprint counts in the upper quartile, in either the proximal or distal region around each gene's TSS.

We applied two methods to predict active TF-target gene interactions and used their intersect. Method 1: Pearson correlation. We calculated the Pearson correlation (r) between each TF-gene pair, across the samples from each of the six ABA brains, separately. TF-gene pairs with $|r| > 0.25$ were considered correlated, as explained in the next paragraph. Pearson correlation is widely used as a measures of gene-gene co-expression and has been shown to yield biologically meaningful interactions (Zhang and Horvath, 2005). Method 2: LASSO regression. Using the glmnet R package, we fit a LASSO regression model to predict the expression of each target gene based on a linear combination of the expression patterns of its candidate regulators. LASSO regression models are among the best performing algorithms for TRN reconstruction from gene expression data (Marbach et al., 2012). We considered candidate regulators with proximal binding sites and with distal binding sites, sequentially. First, we constructed a model considering only TFs whose binding sites were enriched in proximal regions (+/-10kb from the TSS). We then used the predictions from this model as an offset in a second model in which we considered TFs whose binding sites were enriched only in distal regions (>10kb and <1Mb from the TSS).

Next, we defined a consensus model, which retained only high-quality, reproducible TF-gene interactions. We retained genes for which the LASSO regression model explained an average of at least 50% of the gene expression variance across all five training sets ($r^2 > 0.5$). We retained TFs that had at least one target gene in the model. We retained TF-gene interactions that had a non-zero beta-coefficient in the LASSO regression model and a Pearson correlation coefficient $|r| > 0.25$ in at least two of the five Allen Brain Atlas training sets. We determined this threshold for correlation strength and reproducibility based on the following calculations. In the least significant case, a TF-gene pair would have a correlation coefficient, r=0.25, in two training sets and no correlation in the remaining training sets, r=0. We calculated a meta-analytic p-value for this hypothetical TF-gene pair, assuming that the non-zero correlations were observed in the two smallest training sets, which have sample sizes n1=364 and n2=471. For these sample sizes, r=0.25 has a two-tailed p-value of 1.4e-6 and 3.8e-8, respectively. We used Fisher's method to combine these hypothetical p-values from all five training sets (p-values = 1.4e-6, 3.8e-8, 1, 1, 1) and derived a meta-analytic p-value = 2.2e-9. Accounting for 741 TFs x 11,093 genes = 8,219,913 possible TF-gene interactions, this meta-analytic p-value corresponds to a Type I error rate of approximately 0.02.

BIOINFORMATIC VALIDATION OF THE TRN MODEL

We validated our network model by evaluating preservation of TF-gene Pearson correlations in holdout samples from the Allen Brain Atlas and in post-mortem adult brain samples from BrainSpan (www.brainspan.org) and from the UK Brain Expression Consortium (Ramasamy et al., 2014) (GSE60862). We used all available samples from each dataset to define TF-gene correlations. We defined a TF-gene pair to be replicated in each test set if it had a correlation coefficient $|r| > 0.25$ in the same direction as observed in the training sets. In addition, we applied our LASSO regression model to the holdout Allen Brain Atlas samples and evaluated test set prediction accuracy. The latter confirmed that the amount of variance explained by the model in training sets is strongly correlated with the amount of variance explained in the test set ($r = 0.75$). This suggests that the model is not overfit to the training data and validates our strategy to select genes in the consensus TRN model based on the amount of variance explained by the LASSO regression models in training data.

We compared the footprint locations to chromatin states derived from independent datasets. The pooled footprints were compared to 25-state ChromHMM chromatin state models from each of then ten human brain samples from the ROADMAP Epigenomics Consortium (Ernst and Kellis, 2012; Kundaje et al., 2015). We used bedtools to count overlaps between footprints

and chromatin states(Quinlan and Hall, 2010). We summarized these data to counts of footprints per kilobase and plotted the mean and standard error across the ten ChromHMM models. This software pipeline was constructed using Snakemake (Köster and Rahmann, 2012).

We compared the TF-target gene predictions from our model to TF-target gene predictions derived from ChIP-seq data in the GTRD database (Yevshin et al., 2017). The GTRD database provides uniformly processed peak calls from each ChIP-seq experiment. Human ChIP-seq data were available from 400 of the 741 TFs in our TRN model. We counted peaks for each TF that were located +/- 10 kb from transcription start sites of ENSEMBL v83 gene models and considered a gene to be a target of a TF if there was at least one peak in this range. For validation of footprint and integrative TRN models, we used data from 242 TFs for which the ChIP-seq data identified at least 1,000 target genes with a peak +/- 10 kb of the transcription start site.

IDENTIFYING KEY REGULATORS OF DISEASE-RELATED GENE EXPRESSION CHANGES

We used prefrontal cortex gene expression profiles from case-control cohorts with schizophrenia, bipolar disorder, major depression, autism spectrum disorder, and Alzheimer's disease (Table S2). For most datasets, we downloaded normalized read counts (RNA-seq) or intensity levels (microarrays). We evaluated batch effects and outliers in each dataset using hierarchical clustering and principal components analysis. In one dataset – MDD cases and controls from GSE54567 and GSE54568 – we identified batch effects, and we performed batch effects correction with the ComBat function in the sva R package. In the other datasets we did not identify batch effects or outliers. We identified differentially expressed genes in these datasets using the edgeR and limma R packages, adjusting for the following covariates as provided by the authors in the Gene Expression Omnibus: post-mortem interval, RNA integrity, pH, sex, and age. For the CommonMind (Fromer et al., 2016), Parikshak et al. (Parikshak et al., 2016), and Gandal et al. (Gandal et al., 2018) datasets, we instead used published tables of differentially expressed genes, as shown in Table S2. We tested for enrichment of the targets of each TF with up- and down-regulated genes in each dataset, separately, using one-tailed hypergeometric distributions. We considered the intersection of the 11.093 brain-expressed genes in our TRN model with genes quantified in each microarray experiment as the background set for enrichment analyses. We calculated meta-analytic p-values for each disorder using Fisher's method. We calculated a False Discovery Rate for each test using the Benjamini-Hochberg method.

To evaluate the robustness of TF-disease associations, we repeated our network enrichment analyses using subsets of TF-target gene interactions supported at more stringent TF-gene co-expression thresholds. The gene set enrichment analyses described above, utilized the full TRN model with 201,218 TF-gene interactions reproducible in at least two Allen Brain Atlas training sets. We repeated this analysis while restricting our analysis to TF-gene interactions with support in at least three training sets, at least four training sets, or in all five training sets. We evaluated overlap between the key regulator TFs identified in each of these analyses by two metrics. First, we compared the lists of key regulator TFs using inter-rater reliability (Cohen's kappa, k). These analyses revealed strong overlap ($p << 1e-16$) across all TF-gene co-expression inclusion thresholds: full model vs. TRN3+: $k = 0.75$; full model vs. TRN4+, $k = 58$; full model vs. TRN5: $k = 0.41$). We utilized key regulators identified in the full model for subsequent analyses.

CELL-TYPE SPECIFIC EXPRESSION ANALYSIS

We studied over-representation of TF target genes among cell-type specific genes. We downloaded FPKM values from RNA-seq of major brain cell types purified from the mouse brain (Zhang et al., 2014): https://web.stanford.edu/group/barres_lab/barreslab_rnaseq.xlsx. We used the pSI R package (Dougherty et al., 2010) to calculate each gene's cell type specificity and define lists of genes with specific expression in one of the major brain cell types ($pSI < 0.05$). Enrichment of cell-type specific genes among the targets of each TF was calculated using one-tailed hypergeometric distributions.

We utilized the full TRN model for all subsequent analyses.

Annotation of TFBS-Disrupting Variants

Using bedtools(Quinlan and Hall, 2010), we intersected TFBSs with genetic variants from Kaviar (Glusman et al., 2011), restricting our analysis to variants in dbSNP. We included all variants that overlapped a predicted TFBS by at least 1 bp. Next we intersected this table of putative TFBS-disrupting variants with summary statistics Next, we defined LD-independent lead SNPs and correlated SNPs in the GWAS summary statistics from the Psychiatric Genomics Consortium GWAS of schizophrenia (PGC, 2014) (https://www.med.unc.edu/pgc/results-and-downloads) and from the International Genomics of Alzheimer's Project (Lambert et al., 2013) (http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php). Each LD independent lead SNP had an association p-value < 5e-8 and $r^2 < 0.6$ compared to other genome-wide significant SNPs. Additional SNPs were in LD (r^2) with a lead SNP. We intersected these disease-related SNPs with TFBS-overlapping SNPs by merging on rsIDs.

Differentially Expressed Genes in POU3F2 Over-Expression Study

Agilent SurePrint microarrays were scanned with a Agilent DNA microarray scanner. Agilent Feature Extraction software was used for spot identification and intensity measurements. Probe-level intensity measurements were log-transformed and quantile normalized. We fit these data to a linear model using the limma R package. We used the lmFit function to model the effects of a group factor with four levels (two vectors: POU3F2 vs. control x two time points: day 3 vs. day 10), as well as a technical batch effect due to the physical microarray on which samples were run. We then used contrasts.fit() to fit a contrast for the effect of POU3F2 vs. control jointly across

the two time points. We used the eBayes() function to estimate significance and calculated a False Discovery Rate using the Benjamini-Hochberg method. We set a significance threshold FDR < 0.1 to produce lists of differentially expressed genes.

Gene Co-Expression Networks

We analyzed our POU3F2 over-expression experiment together with publicly available references datasets related to brain development: BrainSpan (www.brainspan.org), NIH Blueprint Non-Human Primate Atlas (Bakken et al., 2016) (http://www.blueprintnhpatlas.org); single-cell RNA-seq of the developing and adult human cortex (Darmanis et al., 2015) (GSE67835); and CORTECON (van de Leemput et al., 2014) (cortecon.neuralsci.org). We calculated pairwise gene-gene correlations in each dataset separately using quantile-normalized intensity values (microarrays) or normalized read counts (RNAseq). We then identified clusters of co-expressed genes in each dataset using the k-means algorithm, empirically selecting an optimal k based on the size of each dataset and the observed cluster overlap across datasets. We used the intersectoR() function in the R projectR package (https://github.com/genesofeve/projectR) to explore gene membership overlaps between the gene clusters defined in the POU3F2 over-expression experiment and each comparison dataset. This uses a hypergeometric test, phyper(), to determine the statistical significance of the number of genes shared across clusters in the different experiments. Correlation between individual genes' expression and cluster averages were determined using the cor.test() function. Enrichment of disease gene DEG lists and genes annotated to particular GO terms was calculated using the geneSetTest() function in the limma package within Bioconductor.

DATA AND SOFTWARE AVAILABILITY

Files containing DNase-accessible regions and DNase-I footprints that underlie network reconstruction are presented as a Supplementary Dataset at http://amentlab.igs.umaryland.edu/psych-trn-pfc2016/.

Microarray data have been deposited in the Gene Expression Omnibus: GSE102122.

Digital genomic footprinting and transcriptional regulatory network reconstruction were implemented using a pre-release version of the TReNA R package (www.trena.org). Gene co-expression cluster overlap analysis was performed using the project R R package (https://github.com/genesofeve/projectoR).