The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders

Chao Chen1,2*, Qingtuan Meng1, Yan Xia1,3, Chaodong Ding1,3, Le Wang1,4, Rujia Dai1,3, Lijun Cheng2, Preethi Gunaratne3, Richard A. Gibbs5, Shishi Min1, Cristian Coarfa6, Jeffrey G. Reid6, Chunling Zhang9, Chuan Jiao3, Yi Jiang1,10, Gina Giase11, Amber Thomas5, Dominic Fitzgerald5, Tonya Brunetti5,12, Annie Shieh3, Cuihua Xia1, Yongjun Wang13, Yunpeng Wang14,15, Judith A. Badner16, Elliot S. Gershon17, Kevin P. White5,18, Chunyu Liu1,3,19* 

Schizophrenia and bipolar disorder are complex psychiatric diseases with risks contributed by multiple genes. Disregulation of gene expression has been implicated in these disorders, but little is known about such dysregulation in the human brain. We analyzed three transcriptome datasets from 394 postmortem brain tissue samples from patients with schizophrenia or bipolar disorder or from healthy control individuals without a known history of psychiatric disease. We built genome-wide coexpression networks that included microRNAs (miRNAs). We identified a coexpression network module that was differentially expressed in the brain tissue from patients compared to healthy control individuals. This module contained genes that were principally involved in glial and neural cell genesis and glial cell differentiation, and included schizophrenia risk genes carrying rare variants. This module included five miRNAs and 545 mRNAs, with six transcription factors serving as hub genes in this module. We found that the most connected transcription factor gene POU3F2, also identified on a genome-wide association study for bipolar disorder, could regulate the miRNA hsa-mir-320e and other putative target mRNAs. These regulatory relationships were replicated using PsychENCODE/BrainGVEX datasets and validated by knockdown and overexpression experiments in SH-SY5Y cells and human neural progenitor cells in vitro. Thus, we identified a brain gene expression module that was enriched for rare coding variants in genes associated with schizophrenia and that contained the putative bipolar disorder risk gene POU3F2. The transcription factor POU3F2 may be a key regulator of gene expression in this disease-associated gene coexpression module.

INTRODUCTION

Schizophrenia (SCZ) and bipolar disorder (BD) are severe psychiatric diseases that affect millions of people worldwide (1). Despite a century of evidence establishing their genetic basis, only recently have specific genetic risk factors been identified (2). However, there is not a simple Mendelian model between the risk alleles and these psychiatric disorders. Instead, these psychiatric disorders are very complex and are polygenic in nature involving hundreds of genes with small effect sizes (3, 4). With the polygenic nature of these disorders, many studies have focused on converging individual genes into functional networks to reveal the underlying disease etiology. For example, Fromer et al. (5) have reported a brain coexpression network that captured SCZ association and impact of polygenic risk for SCZ. Gandal et al. (6) identified several disease-shared and disorder-specific coexpression modules in parallel with polygenic overlap among five major psychiatric disorders. However, it is still unclear how disease-associated module (dAM) genes interact and contribute to the pathophysiology of SCZ and BD.

The most studied regulators of gene expression are transcription factors and microRNAs (miRNAs). Increasing evidence implicates connections between transcription factors and miRNAs with SCZ and BD. Transcription factors and miRNAs are known to affect brain development and are differentially expressed in postmortem brain tissue from patients with SCZ and BD, and their target genes are enriched in SCZ and BD risk loci (7–9).

On the basis of the assumption that coexpression implies coregulation (10), and hub genes in coexpression modules are likely to be the regulators of gene coexpression, we integrated genotype, mRNA, and miRNA data from brain tissue samples from patients with SCZ or BD to search for transcription factors and miRNAs at the hub of dAMs and to experimentally validate their putative regulatory relationships. We identified one hub gene, POU class 3 homeobox 2 (POU3F2), as a master regulator in the dAM for SCZ and BD.
RESULTS
Because SCZ and BD share genetic components and similar gene expression patterns (6, 11), and with the goal to explore the biology underlying potential shared disease mechanisms between SCZ and BD, we used postmortem brain transcriptome data from 95 patients with SCZ, 74 patients with BD, and 225 control individuals from multiple datasets. We combined SCZ and BD as major psychiatric disorders in all case-control studies. Both mRNA and miRNA data from microarray and RNA sequencing (RNA-seq) were incorporated. Detailed demographic information and quality control steps for the postmortem brain tissue data are provided in Materials and Methods and tables S1 and S2.

Identification of a disease-associated miRNA and mRNA coexpression network module
To detect disease-associated coexpression modules, we first analyzed the mRNA and miRNA expression in parietal cortex (PC) tissue samples from the Stanley Medical Research Institute (SMRI) using weighted gene coexpression network analysis (WGCNA) (12). To capture expression of a group of genes exhibiting case-control differences, we combined case-control data to construct networks and identified 46 coexpression modules. We found one daM after correcting for the effects of sex, age, brain tissue pH, RNA integrity number, and postmortem interval [control versus SCZ + BD; \( P = 4.3 \times 10^{-5} \), false discovery rate (FDR) \( q = 7.0 \times 10^{-3} \)]. Five miRNAs and 545 genes were included in the daM. Functional enrichment testing by DAVID v6.8 (13) analysis showed that these genes were enriched for gliogenesis \( (P = 1.5 \times 10^{-11}, \text{FDR} \ q = 2.8 \times 10^{-8}) \), glial cell differentiation \( (P = 1.5 \times 10^{-8}, \text{FDR} \ q = 2.8 \times 10^{-5}) \), and neurogenesis \( (P = 3.8 \times 10^{-8}, \text{FDR} \ q = 7.2 \times 10^{-5}) \).

To investigate whether gene networks were conserved among patients with SCZ or BD compared to controls, we also constructed gene coexpression networks in other ways. We separated SCZ, BD, and control samples to run WGCNA analysis independently and compared their module differences. We used \( Z_{\text{summary}} \) to assess whether the connectivity level and pattern of a module in one dataset were preserved in another, where \( Z_{\text{summary}} > 2 \) implied moderate preservation and \( Z_{\text{summary}} > 10 \) indicated high preservation (14). We found that all modules detected in control samples were well preserved in SCZ or BD patients \( (Z_{\text{summary}} > 2; \text{see Materials and Methods}) \). We also applied WGCNA analysis to SCZ + control and BD + control brain tissue samples to determine whether any disease-specific modules existed. We detected one module showing significant association with SCZ in SCZ + control samples, but this did not survive multiple testing corrections \( (P < 0.05, \text{FDR} \ q > 0.05) \). Genes in this module significantly overlapped with those in the daM of the combined dataset \( (P < 0.01) \). We did not detect any daM in BD + control brain tissue samples.

Module preservation in independent datasets
To test preservation of the daM, two independent datasets were used as replicates: One group comprised control samples from Andrew Singleton’s group, which also included miRNA and mRNA microarray expression data from 138 frontal cortex tissue (FCTX) samples (GSE15745) (15); the other group comprised RNA-seq data for prefrontal cortex tissue samples from 63 controls, 70 SCZ patients, and 48 BD patients from the PsychENCODE/BrainGVEX project (Fig. 1A) (16). The daM in the SMRI PC tissue data had \( Z_{\text{summary}} = 36.8 \) in frontal cortex data from GSE15745 (Fig. 1B) and \( Z_{\text{summary}} = 10.9 \) in prefrontal cortex data from the BrainGVEX dataset (Fig. 1C), indicating well-preserved membership and connectivity of the daM. It is worth noting that the corresponding module in the BrainGVEX dataset also exhibited significant disease association in the same direction \( (P < 0.01) \).

Enrichment of genetic variants associated with SCZ or BD in the daM
We tested whether genes in the daM were related to genetic association with SCZ or BD. For the genetic variants, we focused on common or rare single-nucleotide variants and copy number variations (CNVs). For common variants, we applied MAGMA (17) and INRICH (18) to test the enrichment of daM genes with signals from SCZ or BD genome-wide association study (GWAS) but did not detect any significant enrichment (Fig. 2A). For rare variants, we examined the rare variant burden of module genes associated with disease. Genes in the daM were significantly enriched for genes implicated in rare and de novo variants from three selected independent sources (enriched \( P < 0.05; \text{Fig. 2, A and B} \)) (19–21). The enrichment was more significant when we combined the three gene sets (overlapped gene number = 106;
enriched $P = 4.31 \times 10^{-7}$; Fig. 2A and table S3). As a negative control, we used type 2 diabetes–associated genes from GWAS to test the enrichment with genes containing rare variants (22), and found no significant overlap with any of the three sources, nor the combined set. Other brain coexpression modules from our data were also used as controls to test the enrichment. We observed another 9 of 46 modules significantly enriched for genes with rare coding variants, and the daM was the second significant module after multiple testing correction (adjusted $P < 0.0001$). For CNVs, we tested the enrichment of genes implicated in CNV regions but did not observe any significant enrichment.

**Potential key regulators and their roles in the daM**

In the daM, we focused on identifying transcription factors and miRNAs as key regulators and explored their functional roles. Five miRNAs were in the module: hsa-miR-585 and four hsa-miR-320 family members including hsa-miR-320b, hsa-miR-320c, hsa-miR-320d, and hsa-miR-320e. Six transcription factors were also included in this module: Pou3f2, endothelial PAS domain protein 1 (Epas1), paired box 6 (Pax6), zinc finger protein 423 (Znf423), Sry-box 5 (Sox5), and Sry-box 9 (Sox9).

We first explored different aspects of regulatory roles for miRNAs in the daM: module membership, pairwise weighted correlation with miRNA, and the enrichment of target genes with predicted binding. By calculating the module membership, we found that these five miRNAs had significant correlations with the module eigengene ($P < 0.001$; Table 1). By extracting pairwise weighted correlation coefficient ($r_w$) between the miRNA and mRNA, we detected strong correlations between miRNAs and mRNAs ($r_w > 0.05$; original correlation $r > 0.607$). Numbers of correlated miRNAs for each miRNA are provided in Table 1. We observed that expression of most of the mRNAs was inversely correlated with expression of miRNAs (476 of 545, 87.3%), suggesting that miRNAs may play a role in regulating the transcriptome through direct down-regulation of their mRNA targets. Using mirWalk2.0, we examined the predicted binding targets of each miRNA and calculated the overlapped genes in the daM (23). We found that mRNA genes predicted to be targets of the five miRNAs were significantly enriched in this module (hypergeometric probability test, $P < 0.05$; Table 1). This suggested that the predicted regulation relationships from mirWalk2.0 were consistent with the expression correlation results from WGCNA.

We next explored the regulatory roles of transcription factors in the daM. One hundred one putative targets of the six transcription factors were included using transcription factor binding information obtained from Fuxman et al. (24) and Kheradpour and Kellis (25). Among the six transcription factors detected in the daM, Pou3f2 had the most regulated putative targets ($n = 26$) in this module, including transcription factors Pax6 and Sox9. The other transcription factors Epas1, Pax6, Znf423, and Sox9 had 9, 21, 24, 10, and 11 putative targets, respectively. We have provided those transcription factor pairwise correlation and disease association $P$ values in replicated BrainGVEX data (tables S4 and S5).

We visualized the relationships between transcription factors and their miRNA targets in this module, with the six transcription factors and five miRNAs as hub genes (Fig. 3). We found that the transcription factor–miRNA target binding relationship was consistent with the relationship extracted from module correlation testing. Hsa-miR-320e was the most connected node in this module according to the correlation data (68 nodes with $r_w > 0.05$; Table 1).

**Causal relationships among key regulators in the daM**

We tested whether transcription factors were upstream or downstream regulators of miRNAs or were regulated by other transcription

---

**Fig. 2. Enrichment of common and rare genetic variants in the daM for SCZ or BD postmortem brain tissue.** (A) Enrichment of genes in the daM with common and rare variants. For common variants, we applied MAGMA and INRICH to test the enrichment. Self-contained gene set analysis tested whether genes in a gene set showed joint association with SCZ, and competitive gene set analysis tested whether those genes showed differential association with SCZ compared with other genes in the rest of the genome. Thresholds of $1 \times 10^{-5}$, top 0.1%, top 1%, and top 5% of all significant single-nucleotide polymorphisms (SNPs) were used as index SNPs in INRICH, and none of them detected significant enrichment. For rare variants, data from two exome sequencing studies and one database were used to test the enrichment. We applied a hypergeometric method to test the overlap with genes collected in the study of Purcell et al. (20), the NPdenovo database (19), and the combined gene sets from these two studies. We applied logistic regression to test the rare variant burden using disruptive and damaging ultra-rare variant counts for each gene from the study of Genovese et al. (21). (B) Number of genes containing de novo or rare mutations from three sources (19–21) overlapping with genes in the daM.
factors through integration of genetic markers. Causal inference of correlated nodes (miRNAs, transcription factors, and their target protein-coding genes in this study) was inferred by the network edge orienting (NEO) method (26). Because we were interested in the causal relationship between five miRNAs and six transcription factors, only miRNA quantitative trait locus (miQTL) signals associated with the five miRNAs and expression QTL (eQTL) signals associated with the six transcription factors were included in this analysis. Among the five miRNAs, only hsa-miR-320e had significant miQTL signals \( (P < 0.05, \text{FDR} q < 0.05) \), so this miRNA and six transcription factors were selected for regulation direction tests.

We used modified NEO (see Materials and Methods) to build a local structure equation model and to obtain edge-oriented scores. The orthogonal causal anchors (OCAs) \( (\text{LEO.NB.OCA}) \ (A \rightarrow B) > 0.3 \) and candidate pleiotropic anchor (CPA) \( (\text{LEO.NB.CPA}) \ (A \rightarrow B) > 0.8 \) indicated that the regulation direction was A to B. We observed that \( \text{LEO.NB.OCA} \ (\text{POU3F2} \rightarrow \text{hsa-miR-320e}) = 0.526 \) and \( \text{LEO.NB.CPA} \ (\text{POU3F2} \rightarrow \text{hsa-miR-320e}) = 1.55 \), which suggested that POU3F2 may be an upstream regulator that affected hsa-miR-320e’s expression (fig. S1 and table S6).

Meanwhile, NEO results indicated that POU3F2 was the upstream regulator of other transcription factors (PAX6, ZNF423, and SOX9) (fig. S1 and table S6). These results suggested that POU3F2 may be a key regulator in the daM.

Experimental validation of the putative causal regulatory relationship in the daM

With the hub positions of POU3F2 and hsa-miR-320e in the regulation network, we tried to confirm their predicted relationships through in vitro experiments. We used RNA interference (RNAi) and a gene overexpression assay to induce expression alterations of POU3F2 and hsa-miR-320e in SH-SY5Y neuroblastoma cells and examined the expression changes of their predicted targets.

The expression of POU3F2 decreased by 41% after RNAi in SH-SY5Y cells \( (P < 0.001) \). As a result, hsa-miR-320e expression increased by 170% \( (P < 0.001) \); expression of two negative controls (ECM7 and PSMB4) that were not predicted targets of POU3F2 did not change.

### Table 1. Characteristics of miRNAs in daM and their predicted binding targets.

<table>
<thead>
<tr>
<th>miRNA</th>
<th>Chr</th>
<th>Correlated with ME ((P\text{ value}))</th>
<th>Genes predicted as binding targets in daM ((\text{no. of total target genes}))</th>
<th>Significance of binding targets in module ((\text{enrichment} \ P\text{ value}))</th>
<th>No. of correlated genes in module ((r_w &gt; 0.05))</th>
</tr>
</thead>
<tbody>
<tr>
<td>hsa-miR-320b</td>
<td>Chr1</td>
<td>3.70 \times 10^{-5}</td>
<td>333 (10,126)</td>
<td>(&lt;1 \times 10^{-10})</td>
<td>14</td>
</tr>
<tr>
<td>hsa-miR-320c</td>
<td>Chr18</td>
<td>1.31 \times 10^{-5}</td>
<td>303 (9,480)</td>
<td>(&lt;1 \times 10^{-10})</td>
<td>34</td>
</tr>
<tr>
<td>hsa-miR-320d</td>
<td>Chr13</td>
<td>1.09 \times 10^{-5}</td>
<td>297 (8,904)</td>
<td>(&lt;1 \times 10^{-10})</td>
<td>44</td>
</tr>
<tr>
<td>hsa-miR-320e</td>
<td>Chr19</td>
<td>4.79 \times 10^{-6}</td>
<td>263 (7,887)</td>
<td>(&lt;1 \times 10^{-10})</td>
<td>68</td>
</tr>
<tr>
<td>hsa-miR-585</td>
<td>Chr5</td>
<td>1.30 \times 10^{-5}</td>
<td>139 (7,742)</td>
<td>0.046</td>
<td>0</td>
</tr>
</tbody>
</table>

*\(r_w\) is the weighted correlation coefficient from the transformed pairwise correlation matrix, where \(r_w > 0.05\) is equivalent to the original \(r > 0.607\).
in vitro results confirmed experimentally that knockdown and overexpression experiments (Fig. 4, C and D). These potential targets of *hsa-miR-320e* included *CHMP2A*, *VPS29*, Two negative controls (Fig. 4B). The blue bars indicate gene expression in control groups before knocking down or overexpressing *POU3F2* or *hsa-miR-320e*. The three biological replicates were used, and for each biological replicate, we designed three technical replicates.** 

**DISCUSSION**

Previous work has documented abnormalities in coordinated gene expression networks in postmortem brain tissue from patients with SCZ or BD. However, these studies have involved either a single-dimension correlated network that cannot resolve the driver node within the module or have not done a regulatory relationship analysis. Most findings have been presented as correlations instead of causal relationships. In this study, we integrated multidimensional datasets and revealed a role for *POU3F2* as a regulator of a gene coexpression network. *POU3F2* is clearly only one regulator, and there are many pathways that may potentially be involved in the etiology of SCZ or BD. Our study provides a framework to capture such pathways and to tease out their regulatory relationships.

*POU3F2* has been reported to be associated with SCZ using brain activation measurements from functional magnetic resonance imaging as a quantitative phenotype (36). *POU3F2* has been shown to lie close to the BD risk loci in the most recent Psychiatric Genomics Consortium BD GWAS (37). Expression changes in *POU3F2* have been observed in neurons derived from SCZ patient–specific induced pluripotent stem cells (38). *POU3F2* was discovered on the basis of the sequence similarity of the POU domain and is also known as Brain-2 (*Brn-2*), because it is expressed in the central nervous system (39). The function of *POU3F2* was initially studied in melanocytic cells. In our study, we observed that *POU3F2* knockdown could promote NPC proliferation and inhibit neuronal differentiation (Fig. 5, A to D). The aberrant expression of *POU3F2* could lead to alterations in neuron number (40). In overexpression experiments, *POU3F2*’s expression increased by nearly 10-fold (P < 0.001). As a result, *hsa-miR-320e*’s expression significantly decreased by 33% (P < 0.001), and the expression of two negative controls was not significantly changed (Fig. 4B).

In the case of knocking down *hsa-miR-320e*, *hsa-miR-320e*’s expression decreased by 33% (P < 0.001) but had no effect on expression of *POU3F2* (Fig. 4C). Overexpression of *hsa-miR-320e* (increased by 120%; P < 0.001) did not change expression of *POU3F2* (Fig. 4D). Two negative controls (*CHMP2A* and *VPS29*), which were not potential targets of *hsa-miR-320e*, were not significantly changed in both knockdown and overexpression experiments (Fig. 4, C and D). These in vitro results confirmed experimentally that *POU3F2* was the upstream regulator of *hsa-miR-320e*.

*POU3F2*’s effects on the proliferation and differentiation of neural progenitor cells

Accumulating evidence suggests that *POU3F2* is primarily expressed in the central nervous system and plays an important role in brain development and cell differentiation (27, 28). To further characterize its functional roles, we decided to knock down *POU3F2* in human neural progenitor cells (NPCs). NPCs were transfected with small hairpin RNAs (shRNAs) against *POU3F2*, and their proliferation and differentiation were evaluated using an immunofluorescence assay. After knocking down *POU3F2* (expression decreased by 51%; P < 0.001), we found that the proliferation ratio of EdU+ cells (5-ethyl-2′-deoxyuridine; a marker for proliferating cells) to DAPI+ cells (4′,6-diamidino-2-phenylindole; a marker of living cells) was significantly increased compared to control groups (P < 0.001; Fig. 5, A and B). We next analyzed the differentiation of NPCs to neurons and found that the proportions of Tuj1+ cells (a marker of immature neurons) and MAP2+ cells (a marker of mature neurons) were significantly decreased compared to control groups (P < 0.001; Fig. 5, C and D). These results indicated that *POU3F2* knockdown could promote the proliferation ability of NPCs and inhibit the differentiation of NPCs into neurons.

To investigate how *POU3F2* affected cell proliferation and differentiation capabilities, we examined expression changes of six putative targets of *POU3F2* in the daM in NPCs after *POU3F2* knockdown or overexpression. These putative targets included *SOX9*, *PAX6*, *ZNF423*, *NOTCH2*, *CLU*, and *TRIM8*. The three transcription factors (*SOX9*, *PAX6*, and *ZNF423*) and the most connected gene in the daM (*NOTCH2*) have been reported to regulate brain development and neural differentiation in many studies (29–32). *NOTCH2* was also reported to be associated with SCZ (33). *TRIM8* and *CLU* were also hub genes in the daM. These two genes are located in the 108 significant SCZ GWAS loci (3) and are involved in tumor cell proliferation and differentiation (34, 35). Expression of *SOX9*, *ZNF423*, *NOTCH2*, *CLU*, and *TRIM8* significantly decreased by 10, 7.9, 39, 17, and 15% after *POU3F2* knockdown and significantly increased by 57, 8.0, 39, 23, and 40% after *POU3F2* overexpression (P < 0.05; Fig. 5, E and F). Expression of *PAX6* was not significantly increased after *POU3F2* overexpression, suggesting that it may not be regulated by *POU3F2* in our NPC model. Expression of negative controls (*VPS29* and *VCP*) was not significantly changed (Fig. 5, E and F).
and may be one possible explanation for anatomical changes in the brain tissue of patients with BD (40).

The genes in daM significantly overlapped with those from the PsychENCODE Capstone One study, which detected one module associated with disease and was functionally enriched for genes involved in glia differentiation (geneM3/isoM1; overlap $P < 0.01$) (41). Genes in the daM also overlapped with our previous findings using datasets from multiple brain banks (33) and with results from the study of Torkamani et al. (42) (overlapping $P < 0.01$).

In our study, we observed that the daM harbored risk genes carrying rare variants but not the common variants identified by GWAS. However, the genes in daM could still be related to common risk variants. For example, rs11191359, rs4146429, and rs4146428 are eQTL signals of TRIM8 located in the promoter region of TRIM8. They are located in a SCZ-associated region and have linkage disequilibrium with SCZ GWAS SNP rs7907645 (GWAS association; $P = 1.27 \times 10^{-11}$) (3). TRIM8 expression is regulated by the transcription factors POU3F2 and PAX6 (24). Variants in the TRIM8 promoter region may disrupt the binding efficiency of POU3F2 and PAX6 and reduce expression of TRIM8. Integrating genetic risk variants with regulatory networks may provide new insights into the transcriptional regulatory architecture that could underlie certain psychiatric disorders.

Our study does have several limitations. First, although hundreds of brain samples were used, the sample size was still relatively small. Possibly due to the small sample size, analyses of the coexpression modules in the uncombined SCZ and BD samples did not yield robust findings. Second, limited data were generated containing both miRNA and mRNA expression for building networks involving miRNAs. This could lead to missing some important miRNA regulatory relationships. Third, gene expression changes may be affected by altered cellular populations in patients (43). In our study, we estimated the proportions of each cell type using a deconvolution method and found no significant difference (fig. S2). Larger sample sizes and new methods may be required to better estimate cell numbers. Fourth, we only validated a few regulatory pathways in the daM. Large-scale validation should be completed in the future. Last, we observed the daM genes involved in gliogenesis, glial cell differentiation, and neurogenesis,
but we only validated the regulatory loop in neuroblasta and NPC lines. Further investigation is needed to clarify whether these regulatory networks also occur in non-neural cells.

By integrating genotype, miRNA, and mRNA expression data, as well as transcription factor and miRNA binding information, we found cascade regulation relationships from SNP variants to transcription factors to miRNAs or other target genes. Our results suggest that complex diseases such as SCZ and BD require a systems biology approach, with an integration of multidimensional datasets to elucidate a better understanding of disease risk.

**MATERIALS AND METHODS**

**Study design**

This study was designed to investigate the dysfunctional regulatory network and key regulators in postmortem brain tissue from patients with SCZ or BD. We analyzed genome-wide mRNA, miRNA, or genotyping data from postmortem brain tissue from 169 patients with SCZ or BD and 225 healthy control individuals who did not have a known history of psychiatric disorders. We first applied WGCNA and QTL analysis to reveal mRNA-miRNA and genotype-mRNA/miRNA relationships. The daM was detected because it exhibited case-control expression differences after removing confounding factors. We next used web resources to explore transcriptional regulators (transcription factors and miRNA binding information) and identified POU3F2 as a master regulator of other transcription factors and miRNAs in the daM. The function of POU3F2 was examined in cell differentiation and proliferation experiments, and its regulatory activities were validated using RNAi, gene overexpression, and luciferase reporter experiments.

**Samples**

**Discovery data**

PC tissue specimens from the SMRI Neuropathology Consortium and Array collections included SCZ, BD, and control samples (table S1). We removed non-Europeans, duplicates, and samples missing any of the mRNA, miRNA, or genotyping data. After filtering, we retained 75 samples (51 patients and 24 controls), yielding data for 19,984 mRNAs, 470 miRNAs, and 1,452,078 SNPs for subsequent analyses. The detailed demographic, clinical information and their distribution differences among groups are provided in table S1.

**Replication data**

The two replication datasets were microarray data from Andrew Singleton’s group [Gene Expression Omnibus (GEO) accession no. GSE15745] and RNA-seq data of BrainGVE from PsychENCODE. GSE15745 contains frozen FCTX samples from 138 neurologically normal Caucasian subjects after quality control. The BrainGVE project includes samples from SMRI, so we excluded those overlapping samples when using PsychENCODE as replicates. Seventy SCZ, 48 BD, and 63 controls from PsychENCODE were used for validation. The detailed demographic, clinical information and their distribution differences among groups are provided in table S2.

**Module construction and preservation statistics**

We identified mRNA and miRNA with correlated expression patterns using WGCNA (12). We calculated a correlation matrix for all possible pairwise nodes (mRNA and miRNA) and chose the power = 6 for weighting the correlation matrix following an approximate scale-free topology. We set minimum block size as 30 and bivariate midcorrelation (bicor) to build the network. We detected network modules using the dynamic tree cut algorithm, with the mergeCutHeight as 0.05 and deepSplit as 2. WGCNA and the dynamic tree cut algorithm were implemented in R (version 3.1.3) (12). The unsigned network type was used to keep the negative relationships between miRNAs and mRNAs. We plotted the pairwise connection network using Cytoscape v3.6.1 (45).

Because our sample size is relatively small, we used two additional datasets to assess the module preservation. The validation datasets included samples from BrainGVE and GSE15745. We applied Zsummary test to assess module preservation between expression datasets (14). The recommended thresholds are as follows: $Z_{\text{summary}} < 2$ implies no evidence for module preservation, $2 < Z_{\text{summary}} < 10$ implies weak to moderate evidence, and $Z_{\text{summary}} > 10$ implies strong evidence for module preservation.

**NEO analysis of transcription factor–miRNA interactions**

In addition to binding information, we used modified NEO to investigate the causal relationship between transcription factors and miRNAs (26). The imported data were the expression data of transcription factors and miRNAs, as well as genotype data. The outputs are local-structure edge orienting scores, which use the likelihoods of local structural equation models to integrate selected traits and markers to assess the causal relationship between correlated quantitative variables. We selected the genotype data that included eQTLs of transcription factors and miRNAs from our analysis of SMRI samples and from results of the GTEx portal (www.gtexportal.org), CommonMind Consortium (commonmind.org/), and UK Brain Expression Consortium (www.brainec.org/). In total, 901 SNPs were included in NEO analysis. The CPA model was used to test single marker edge orienting, and the OCA model was used to test multiple genetic markers. The likelihood-based CPA score assessed whether the chosen model would yield a higher likelihood than did the alternative models. We used a threshold of 0.8, as the software suggested, which implies that the model likelihood score of the causal model was $10^{0.8} = 6.3$-fold higher than that of the next best model. For the OCA score, we used a threshold of 0.3, as suggested, which implies that the model likelihood score of the causal model was $10^{0.3} = 2$-fold higher than that of the next best model.

**NPC proliferation and differentiation assay**

We investigated the function of POU3F2 in neuronal cell proliferation using the BeyoClick EdU Cell Proliferation Kit with Alexa Fluor 555 (C0075S, Beyotime). Briefly, after POU3F2 knockdown, NPCs were incubated with 10 μM EdU solution for 5 hours to label proliferating cells. Then, cells were fixed with 4% paraformaldehyde and permeabilized with 0.3% Triton X-100/phosphate-buffered saline (PBS). The prepared click additive solution was used to detect the EdU-incorporated cells. Last, cells were labeled with Hoechst 33342 to count the proportion of EdU+ cells. For neuron differentiation assay, we used the STEMdiff Neuron Differentiation Kit (08500, STEMCELL Technologies) and the STEMdiff Neuron Maturation Kit (08510S, STEMCELL Technologies) to generate neurons according to the manufacturer’s instructions. For immunofluorescence staining assay, cells were fixed with 4% paraformaldehyde and permeabilized with 0.3% Triton X-100/PBS. Then, cells were blocked using 5% bovine serum albumin/PBS for 30 min at room temperature, followed by incubation with indicated primary antibodies for 1 hour at room temperature: Tuj1 (1:300; 5568, Cell Signaling Technology) and MAP2
SUPPLEMENTARY MATERIALS
www.sciencetranslationalmedicine.org/cgi/content/full/10/472/eaat8178/DC1
Materials and Methods
Fig. S1. Causal relationship between transcription factors and miRNAs.
Fig. S2. Differences in cell type proportions between cases (BD and SCZ) and controls.
Fig. S3. Causal relationship between transcription factors and miRNAs.
Fig. S4. RT-qPCR analysis of POUSF2 knockdown efficiency.
Fig. S5. Distribution of eigengene values in daM.
Table S1. Demographic and clinical information for discovery SMRI data.
Table S2. Demographic and clinical information for replicated BrainGEX data.
Table S3. Overlap of genes in the daM and genes with rare variants.
Table S4. Transcription factor expression association with disease in SMRI and BrainGEX data.
Table S5. Transcription factor expression correlations in BrainGEX data.
Table S6. OCA (LEO.NB.OCA) and CPA (LEO.NB.CPA) scores for six transcription factors and one miRNA.
Table S7. Sequences of RT-qPCR primers and siRNA.
Table S8. Top 50 connected genes in the daM, transcription factors, and their WGCNA parameters.
Table S9. POUSF2 and its putative targets in daM.
REFERENCES AND NOTES
L.W., and Q.M. performed the NPC proliferation and differentiation assays and RNAi and overexpression experiments. Y.X., C.Z., and C.J. provided gene expression analysis. Y.J., S.M., and C.X. performed the eQTL and miQTL analysis. R.D. performed the cell type deconvolution analysis. C. Chen performed all other data analysis. J.A.B., E.S.G., Yunpeng Wang, and K.P.W. contributed to study design and the writing of this manuscript. Competing interests: K.P.W. is the president and a shareholder of Tempus Labs Inc. The other authors declare that they have no competing interests. Data and materials availability: All data associated with this study are in the paper or in the Supplementary Materials. The Stanley brain expression data have been submitted to GEO with accession no. GSE35977. To apply for PsychENCODE Consortium data, researchers can submit an application form via the Synapse website: www.synapse.org/#!Synapse:syn4921369/wiki/235539.

Submitted 9 April 2018
Resubmitted 26 July 2018
Accepted 7 November 2018
Published 19 December 2018
10.1126/scitranslmed.aat8178

The transcription factor POU3F2 regulates a gene coexpression network in brain tissue from patients with psychiatric disorders

Chao Chen, Qingtuan Meng, Yan Xia, Chaodong Ding, Le Wang, Ruijia Dai, Lijun Cheng, Preethi Gunaratne, Richard A. Gibbs, Shishi Min, Cristian Coarfa, Jeffrey G. Reid, Chunling Zhang, Chuan Jiao, Yi Jiang, Gina Giase, Amber Thomas, Dominic Fitzgerald, Tonya Brunetti, Annie Shieh, Cuilhua Xia, Yongjun Wang, Yunpeng Wang, Judith A. Badner, Elliot S. Gershon, Kevin P. White and Chunyu Liu

Sci Transl Med 10, eaat8178.
First published 13 December 2018
DOI: 10.1126/scitranslmed.aat8178

Illuminating the genomic mysteries of psychiatric diseases

Schizophrenia and bipolar disorder are complex psychiatric diseases with risks contributed by multiple genes. Studies by the PsychENCODE Consortium, including two in this issue (Chen et al. and Meng et al.), seek to elucidate the genomic elements and regulatory pathways that underpin several psychiatric disorders. Chen et al. analyzed transcriptome data from postmortem brain tissue from patients with schizophrenia or bipolar disorder. They report that the transcription factor POU3F2 is a core regulator of a gene coexpression network associated with these disorders. In a genome-wide analysis of control human brain samples from the adult and developing brain, Meng et al. report that the IncRNA DGCR5, which lies within the 22q11.2 deletion associated with schizophrenia risk, regulates expression of several SCZ-associated protein-coding genes.