Science Translational Medicine
AAAS

www.sciencetranslationalmedicine.org/cgi/content/full/10/472/eaat8178/DC1

Supplementary Materials for

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, Rujia 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, Cuihua Xia, Yongjun Wang, Yunpeng Wang, Judith A. Badner, Elliot S. Gershon, Kevin P. White, Chunyu Liu*

*Corresponding author. Email: chenchao@sklmg.edu.cn (C. Chen); liuch@upstate.edu (C.L.)

Published 19 December 2018, *Sci. Transl. Med.* **10**, eaat8178 (2018)
DOI: 10.1126/scitranslmed.aat8178

This PDF file includes:

Materials and Methods

SMRI Gene expression (mRNA) profiling and quality control

We extracted total RNA of PC tissue specimens from the SMRI Neuropathology Consortium and Array collections using the RNeasy Mini kit (Qiagen, Hilden, Germany). The concentration and A260/A280 ratio were measured with a NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The 28S:18S rRNA ratio and RNA Integrity Number (RIN) were measured using an RNA LabChip kit on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Only RNA samples with a RIN > 6 were used for expression profiling.

We performed whole-genome transcriptome profiling using an Affymetrix Human Gene 1.0 ST Array (Affymetrix, Santa Clara, CA) at the NIH Neuroscience Microarray Consortium facility at Yale University. Single-nucleotide polymorphisms (SNPs) in probe regions can affect probe hybridization efficiency (*46*). We created the list of probes containing SNPs and removed them before data preprocessing: 39,529 out of 805,481 probes were eliminated from the analysis. We provided customized library files (http://bioinfo.psych.uic.edu/ArrayGenes/SNPsInProbes.jsp) for the Robust Multichip Average (RMA) preprocessing steps: background correction, quantile normalization, and gene level summarization (*47*). Afterward, for the convenience of comparison, only genes with Entrez IDs were retained.

SMRI miRNA expression profiling and quality control

Total RNA was extracted from tissues using the mirVana miRNA Isolation Kit (Ambion, Austin, TX) according to the manufacturer's instructions. RNA quality and the presence of small RNAs were measured on a 2100 Bioanalyzer (Agilent Technologies). After strict RNA quality assurance, 1.5 µg of

total RNA was used for small RNA library creation using Illumina's DGE small RNA sample prep kit per the manufacturer's instructions. Purified cDNA was quantified with the Quant-iT PicoGreen dsDNA Kit (Thermo Fisher Scientific) and diluted to 3 pM for sequencing on the Illumina 1G Genome Analyzer (University of Houston). Each library was sequenced in a single lane.

Sequence reads with 36-nt read length were picked for miRNA mapping. Reads that did not pass the Illumina chastity and no-calls filter were removed (*48*). FastQC (v0.11.2) was used to check for homopolymers, adapters, and distribution of base quality. After trimming for adapters, sequences that read length < 10nt, copy number < 4, or more than 10 consecutive, repetitive nucleotides were discarded. The miRBase database release 19 was used to identify miRNAs, and Bowtie 2 was used for mapping (*49*). Average valid sequence reads were 15M in each sample, and the total count was used for sample-wise normalization (*50*). ComBat, a batch effect adjustment program, was used to remove batch effects from both miRNA and mRNA data sets (*51*).

SMRI Genotyping experiment and quality control

Genomic DNA was extracted from frozen PC tissues provided by SMRI following a modified phenol/chloroform/isoamyl alcohol protocol. The DNA was resuspended in 0.1 mM EDTA TE buffer. Genomic DNA was evaluated by a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific) for concentration and by 1% agarose gel to validate DNA integrity. We used Affymetrix GeneChip Mapping 5.0K Array and Assay Kits (Affymetrix) for genotyping according to the Affymetrix protocol. Genotypes were called with the BRLMM-p algorithm (Affymetrix) with all arrays, and SNPs with call rates < 99% were excluded from the analysis. SNPs showing the departure from Hardy-Weinberg

equilibrium (HWE) were also filtered out ($P < 0.05$). Of the remaining SNPs, only SNPs showing minor allele frequency (MAF) > 10% were used for further analysis.

BrainGVEX data as Replication

The BrainGVEX study (synapse accession doi:10.7303/syn4590909) is part of the PsychENCODE project, which is a public resource of multi-dimensional genomic data using unbiased genome-wide approaches from human post-mortem brains (*16*). Currently, we have approximately 1,000 brains, including SCZ, major depression, BD, and controls without any known history of psychiatric disorders. There are two main sources of the data collection, Stanley Medical Research Institute (SMRI) brain collections, and Banner Sun Health Research Institute (BSHRI) brain collection. Because our discovery data is also from SMRI, we excluded the overlapping samples which have been used as primary data.

BrainGVEX RNA preparation

Total RNA was isolated at University of Illinois at Chicago and University of Chicago using a Qiagen miRNeasy mini kit. Approximately 50mg fresh-frozen brain tissue was homogenized by FastPrep-24 system in QIAzol Lysis Reagent with Lysing Matrix D, then mixed well with chloroform. The separated aqueous layer was recovered, mixed with ethanol and applied to a miRNeasy mini column. Columns were treated with Qiagen RNase-free DNase digestion set, then washed with the appropriate miRNeasy mini kit buffers. Total RNA was eluted with RNase-free water.

Total RNA was quantified by either Qubit 2.0 RNA BR assay kit or Xpose spectroscopy; the quality of total RNA was assayed by Agilent RNA 6000 Nano Kit on the Agilent Bioanalyzer. Total RNA

samples that passed QC for library generation have a concentration of ≥ 100ng/uL assayed by Qubit 2.0 RNA BR Assay or Xpose, and RIN score ≥ 5.5 assayed by Agilent Bioanalyzer RNA 6000 Nano assay kit.

BrainGVEX library preparation and sequencing

All total RNA from BSHRI collections were processed into rRNA-depleted stranded libraries for sequencing on the Illumina HiSeq2000 using the TruSeq Stranded Total RNA sample prep kit with Ribo Zero Gold HMR. For some libraries, 2ul of 1:100 ERCC RNA ExFold Spike In Mix 1 was added to total RNA starting material before ribo-depletion step as an internal way of tracking library prep and sequencing quality. Libraries were PCR amplified for 12 cycles and cleaned with 0.60X Ampure XP beads.

Libraries Quality Control were processed at the University of Chicago HGAC by quantification with the Qubit 2.0 dsDNA HS assay kit and quantification and quality check with the Agilent Bioanalyzer DNA HS assay kit. Libraries were sequenced on Illumina's HiSeq2000 on a high output flow cell for 100bp PE sequencing. Libraries are 3-plexed per lane to reach 40M paired-end reads per library.

BrainGVEX data processing

Fastq files had adapters removed (Phred score < 30 at ends) using cutadapt (v1.12) (*52*), then the resulting adapter-trimmed fastq files were checked for quality using FastQC (v0.11.2). Trimmed reads were then aligned to the GRCH37.p13 (hg19) reference genome via STAR (2.4.2a) (*53*) using comprehensive gene annotations from Gencode (v19). BAM files were produced in both genomic and transcriptome coordinates and sorted using samtools (v1.3) (*54*). Gene-level quantifications were calculated using RSEM (v1.2.29).

Before network construction, we applied log 2 transformation to miRNA sequencing data, ran background correction, quantile normalization and summarization on mRNA data. Another round of quantile normalization was applied on combined miRNA+mRNA data, which is proved very useful on cross platform data combination by Thompson *et al*'s evaluation (*55*).

Characteristics of each module

For each module, we used multiple regression to remove the effects of sex, age, pH, RIN, PMI from the module eigengenes before testing the correlation with the diagnosis. Other covariates such as medication effects, death status or cell type proportion differences were also considered and systemically evaluated (See details in Supplementary Methods). We performed the functional enrichment test using DAVID v6.8 (http://david.abcc.ncifcrf.gov/tools.jsp) (*13*), and plotted the pairwise connection network using VisANT (*56*).

eQTL and miQTL analysis

Quantile normalization was applied for both mRNA and miRNA data. Using an additive genetic model by PLINK, linear regression analysis was performed to test for correlation between the normalized data and the number of minor alleles (*57*). From this analysis, an asymptotic p-value from the Wald statistic was obtained as a measure of association of each SNP with the expression of any given mRNA or miRNA. Permutations were performed to correct for an expression-SNP combination and multiple testing within a *cis*-region or whole genome scan. The detailed procedure can be found in our previous paper (*58*).

Enrichment analysis of common and rare variants

The common variants datasets we used were downloaded from PGC (https://www.med.unc.edu/pgc/results-and-downloads), where we selected the GWAS of SCZ and BD. PGC2 SCZ GWAS includes up to 36,989 cases and 113,075 controls (*3*) and PGC2 BD GWAS includes 11,974 cases and 51,792 controls (*4*).

We conducted a gene-based analysis using MAGMA version 1.06 for the enrichment test (*17*). SCZ-associated SNPs were mapped to genes according to their position in the NCBI 37.3 build map. The reference data files of annotation are from the 1,000 Genomes data (1,000 Genomes European panel), which is used to account for linkage disequilibrium between SNPs. The annotation analysis resulted in 19,246 protein-coding genes containing at least one SNP in the SCZ data. An additional annotation using a 10 kilobase window around each gene was made, yielding 19,350 genes, to determine the effect of using a window on relative performance. We first obtained the intermediate results file, i.e., individual gene analysis on SNP p-values from the SCZ meta-analysis. Next, two gene-set analyses were applied: self-contained gene-set analysis and competitive gene-set analysis. Self-contained gene-set analysis tests whether genes in a gene set showed joint association with SCZ, and competitive gene-set analysis tests whether those genes show a different association with SCZ compared with other genes in the rest of the genome. Bonferroni test was used for the p-values' correction. We also used INRICH to test the overlap between genomic intervals associated with SCZ and daM genes (*18*). We considered linkage disequilibrium, gene lengths, and gene density in the analysis. Permutation method was used for multiple testing corrections. PLINK's clump was used to create LD-based intervals based on samples of European ancestry from the 1000 Genomes Project. We

selected the index SNPs using following criteria: significance thresholds 1×10^{-5}, the top 0.1% (P < 5.67e-9), top 1% (P < 3.20e-4) and top 5% (P < 0.0120) of all significant SNPs (*59*).

For genes implicated in *de novo* or rare variants, we selected three data sources: NPdenovo database, which collected genes with *de novo* mutations from SCZ family trios (*19*), and two large SCZ exome sequencing studies which detected genes with the rare variant burden (*20, 21*). Due to the overlapping genes from different sources, we also used the combined gene set of the three datasets for the enrichment test. We tested whether the daM enriched for genes implicated in rare/*de novo* variants using hypergeometric test. Since Genovese's study (*21*) provided the disruptive and damaging ultra-rare variant counts for each gene, we tested the enrichment using a logistic regression model, which uses exome-wide disruptive and damaging ultra-rare variants (dURV) count as a covariate to correct for average burden (*21*). Gene length was normalized before enrichment test using Uddin's method (*60*).

We also considered the potential enrichment of genes in diseases associated CNV regions. The CNV regions were collected based on the previous investigation on SCZ and BD (*61, 62*). We retrieved the CNV Genes from RefGene (hg19) where at least one bp overlapped with associated CNVs. We tested whether the daM enriched for genes implicated in CNVs using logistic regression model.

We utilized genes associated with type 2 diabetes by GWAS as independent negative control gene sets (*22*). We also used genes from other brain coexpression modules in our study as control to test their enrichments of rare variants.

TFs and their DNA binding targets

TF information including 199 TFs and their putative 199,950 TF-to-target DNA interactions, was obtained from Kheradpour et al.'s study (*25*). We also considered data from another publication of experimental TF binding data from gene-centered eY1H assays, which examined 1,086 full-length human TFs with 390,960 putative TF-DNA interactions (*24*). Using the TFs and targets information above, we extracted the TF-target relationships from genes in the daM and selected the most connected TFs as interesting hubs for further investigation.

miRNA binding targets

To test the validity of the miRNA-mRNA relationships observed in WGCNA, we investigated miRNA-target interaction pairs using miRWalk2.0 (*23*). MiRWalk2.0 compiles the miRNA binding prediction results from 12 existing miRNA-target prediction programs (DIANA-microTv4.0, DIANA-microT-CDS, miRanda, mirBridge, miRDB4.0, miRmap, miRNAMap, doRiNAi.e., PicTar2, PITA, RNA22v2, RNAhybrid2.1, and Targetscan6.2) with experimentally verified miRNA-target interactions (miRTarBase, PhenomiR2.0, miR2Disease and HMDD). Using the miRNAs and targets information above, we extracted the miRNA-targets relationships from miRNAs and mRNAs in the daM and selected the most connected miRNAs as interesting hubs for further investigation.

Induction of human neural progenitor cell (hNPC)

We induced hNPC from normal human induced pluripotent stem cell (hiPSC; Catalog number: DYR0100, Stem Cell Bank, Chinese Academy of Sciences) using STEMdiff neural induction medium

(05835, Stem cell, Canada). Briefly, hiPSCs were seeded onto Matrigel (354277, Corning, USA) coated 12-well plate at a density of 4×10^5 cells per well and fed with mTeSR medium (05850, Stem cell, Canada) with 2 μM Y-27632 (Sigma). The hiPSCs were treated with the induction medium at the following day (Day 0) with medium changed daily. Cell confluence reached approximately 90%-100%. Cells were further cultured in the neural induction medium until passage 3 (Day 14), then cells could be expanded in the STEMdiff Neural Progenitor Medium (05833, Stem cell, Canada) with medium changed daily. At this time point, immunofluorescence characterization confirmed that the expression of NPC markers (Nestin, 1:300, CST, L2117; PAX6, 1:300, CST, F2712) in induced hNPCs (fig. S3), indicating that we have successfully induced hNPCs from hiPSCs.

Lentivirus infection

Lentiviral shRNA targeting human *POU3F2* (seq 5′-GGCGGAUCAAACUGGGAUUTT -3′) was shuttled to the GV248 vector (Genechem, Shanghai). 1×10^6 NPCs were infected with 10 μl packaged lentivirus supernatant. (10^8TU/mL). 72 hours after viral infection, cells were treated with puromycin (1μg/mL) for three days to select NPCs stably express *POU3F2* shRNA or control shRNA. RT-qPCR were used to confirm *POU3F2* knockdown efficiency (fig. S4).

Knockdown and overexpression assay

To experimentally confirm the regulation relationship, we used RNAi and gene overexpression to induce TF/miRNA alteration in SH-SY5Y cell or NPC lines and performed real-time quantitative PCR (RT-qPCR) to detect expression changes of other genes. We also tested expression of genes not predicted as targets of TFs or miRNAs, as negative controls of regulations. For the knockdown experiment, we designed small interfering RNA (siRNA) pairs using the Genscript online tool

(https://www.genscript.com/ssl-bin/app/rnai). SiRNA targeting *POU3F2* (Si-*POU3F2*) and *hsa-miR-320e* inhibitor were synthesized by Genepharma. The SH-SY5Y cells were cultured in DMEM (High Glucose) and supplemented with 10% fetal bovine serum, 1% NEAA, and 1% Glutamax. The NPCs were maintained in Neural Progenitor Medium (Stemcell Technologies) and the plates coated with Matrigel(corning). When cell confluence reached 70-80%, cells were transfected with Lipofectamine RNAiMAX (Thermo Fisher Scientific) according to the manufacturer's instructions. For *POU3F2* overexpression assay, *POU3F2* overexpression vector was transfected into SH-SY5Y using Lipo2000 (Thermo) reagent. Total RNA was extracted using the miRNeasy mini kit (Qiagen) at the time points of 24hr, 36hr, and 48 hr after transfection. RNA quantity and purity were assessed using Nanodrop spectrophotometer (Thermo Fisher Scientific) and agarose gel electrophoresis. cDNA was synthesized using the HiScriptII Q Select RT SuperMix (Vazyme) and further used for RT-qPCR using ChamQ SYBR qPCR Master Mix (Vazyme). β-actin and U6 genes were used as internal references in RT-qPCR. Three biological replicates were used in each knockdown and overexpression experiments, and for each biological replicate we had three technical replicates.

Cell type proportion estimation

Our samples were extracted from parietal cortex, which is a heterogeneous tissue with multiple cell types. Cell type specific expression and proportions may contribute to gene expression variability, and lead to expression differences between cases and controls. To examine this potential artifact by cell types, we applied a supervised method called CellMix (*63*), to estimate the proportion of cell types in cases and controls.

For *a priori* information on cell type expression in brain, Zhang published two matrices of cell-type specific fragments per kilobase million (FPKM) values for genes from mouse cortex (mouse matrix1) and from the human cortex (human matrix2) (*64*). Seven cells types were included in the mouse matrix: neurons, astrocytes, oligodendrocyte precursors, newly-formed oligodendrocytes, myelinating oligodendrocytes, microglia, and endothelial cells, and five cell types were included in the human matrix: neurons, astrocytes, oligodendrocytes, microglia, and endothelial cells. For the mouse matrix, we retained the common genes between the mouse matrix and our dataset. We sorted the common genes by their variance across five cell types in descending order and choose the top 5,000 genes to be the mouse reference profile. For the human matrix, we averaged the FPKM values of the same gene across different age groups in the matrix and sorted the genes by the variance across the five cell types in descending order. We choose the top 5,000 genes and their expression values to be the human reference profile. We performed ref-based deconvolution using the CellMix3 package (lsfit option) and used the two reference profiles separately. For convenience, we separated the cell types into two categories: neuron and non-neuron. The estimated cell fraction by human and mouse reference profiles can be found in fig. S2.

Effects of drug treatment and postmortem process

Drug treatment and postmortem process may affect gene expression profiling. We designed two strategies to examine the drug effects. We used a method described by Fromer *et al.* where they retrieved drug treatment signatures from previous studies using monkeys, rodents, human cell lines, and whole blood tissues (*5*). For genes in modules of interest, they compared the distributions of disease significance with or without those drug signatures. Following their strategy, in our study we did not observe significant distribution differences (Kolmogorov–Smirnov test, $P = 0.21$), indicating

the disease association was not driven by drug effect. A second strategy was to use the lifetime antipsychotics information in BrainGVEX cases. Genes were identified exhibiting association with antipsychotics after controlling other covariates. We observed 65 genes associated with lifetime antipsychotics, which are likely to be affected by drug treatment gradually, and none of them overlapped with genes in the daM in this study. Taking the two strategies' results together, our findings suggest that gene expression changes in this module were not driven by drug treatment.

Death status such as hypoxia, fever and unexpected death may affect the detection of coexpression network. Detailed death information was unavailable from SMRI data. However, in an independent brain project, we systematically evaluated how death status and postmortem process affect genes expression and their networks. We found that the genes easily affected by death status and postmortem process were not part of the daM in this study. In addition, our results were reproducible in another two data sets, FCTX and BrainGVEX, of which the samples' death status and postmortem processes are different from our discovery SMRI data. This evidence suggests that the diseases association is not caused by postmortem processes.

Supplementary Figures

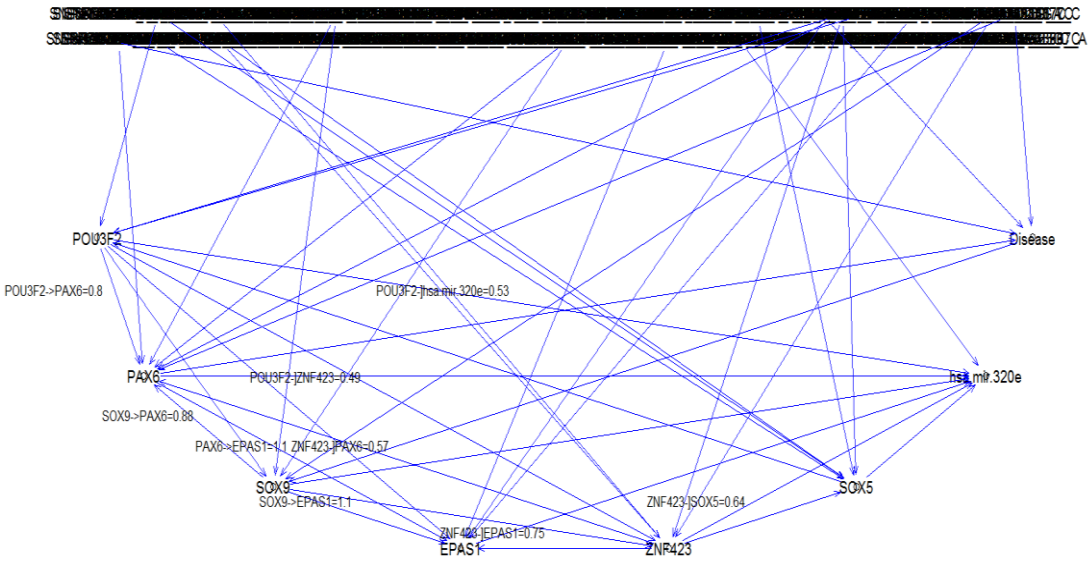

Fig. S1. Causal relationship between transcription factors and miRNAs.

We extracted 263 SNPs as QTL signals of one miRNA and six TFs. The edge orienting scores were calculated based on the correlations between genetic markers and expression values. The candidate pleiotropic anchor (CPA) score was used to test single marker edge orienting. We used a threshold of 0.8, which implies that the model probability of the causal model is $10^{0.8}$ = 6.3 fold higher than that of the next best model. The orthogonal causal anchor (OCA) score was used to test multiple genetic markers, with a threshold of 0.3, implying that the model probability of the causal model is $10^{0.3}$ = 2 fold higher than that of the next best model. Only directions above the threshold were labeled.

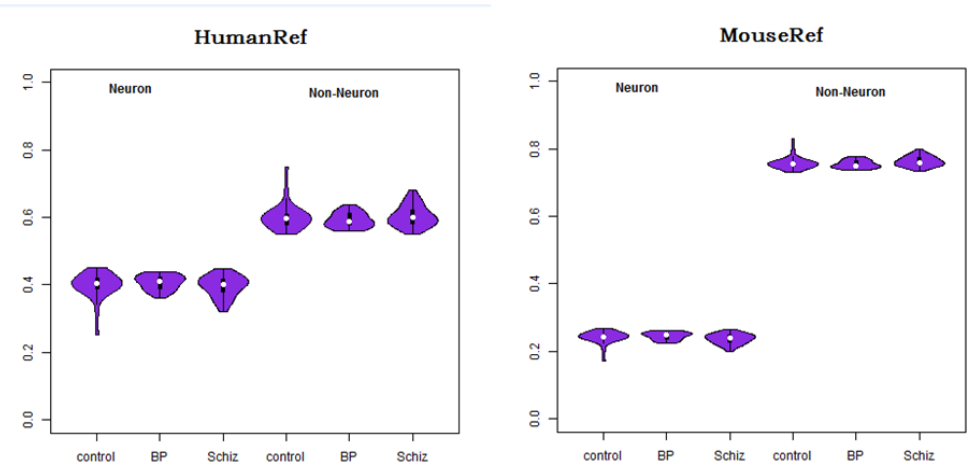

Fig. S2. Differences in cell type proportions between cases (BD and SCZ) and controls.

We used cell type specific matrix from human cortex (HumanRef) and mouse cortex (MouseRef) in Zhang's study (*64*). Totally 14,256 genes in both our data and Zhang matrix were used for the deconvolution test. We estimated cell fractions from SMRI PC tissues using the CellMix package, lsfit option. Using two different types of references, there were no significant differences of estimated fractions among control, BD and SCZ groups.

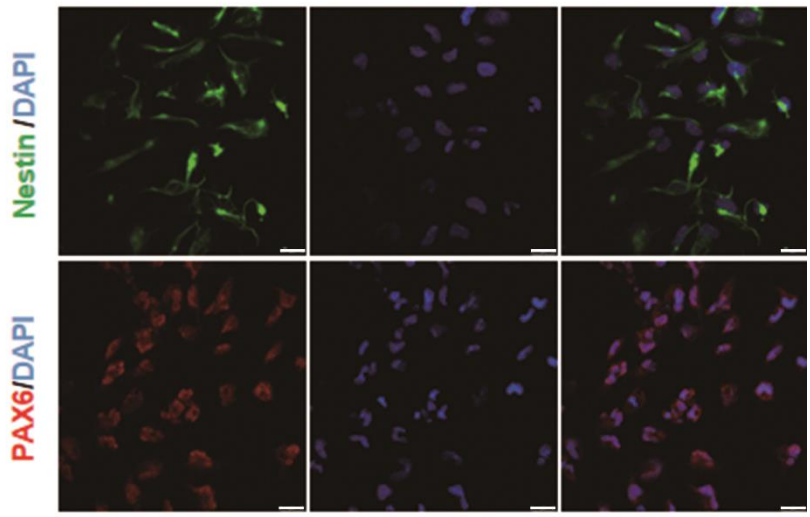

Fig. S3. Immunofluorescence staining of human NPCs.

Typical hNPC markers (Nestin and PAX6) were used for characterization of hNPCs derived from hiPSCs. Nuclei were stained with DAPI (blue). Scale bar, 25 μm.

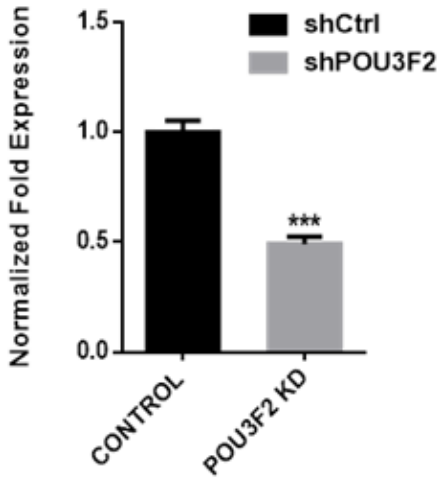

Fig. S4. RT-qPCR analysis of *POU3F2* knockdown efficiency.

Data are shown as means ± SEM of three independent biological replicates. Two-tailed Student's t-test, ***P < 0.001.

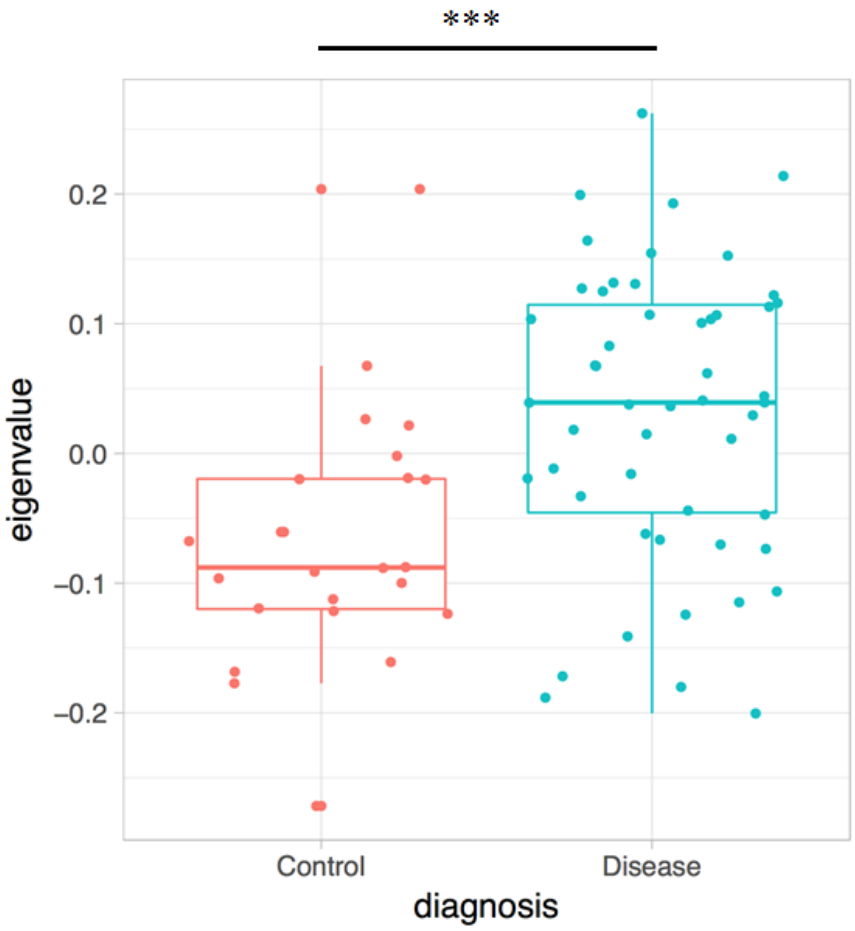

Fig. S5. Distribution of eigengene values in daM.

The module eigengene is one representative gene by summarizing all genes' expression profile in that module. Red dots represent eigengene values of control samples and cyan dots represent values of SCZ and BD patients. ***$P < 0.001$.

Supplementary Tables

Table S1. Demographic and clinical information for discovery SMRI data.

	SCZ	BD	CTRL	P-value
Age	42.0 (20-60)	43.5 (20-65)	46.0 (30-70)	0.27
Sex (F/M)	11/14	13/13	11/13	0.90
Race (Caucasian/Non-Caucasian)	25/0	26/0	24/0	1
PMI	32.1 (13-65)	36.3 (12-84)	25.9 (8-58)	0.34
RIN	8.16 (7.4-8.8)	7.93 (6.7-8.9)	7.95 (6.9-9.0)	0.38
pH	6.42 (5.8-6.93)	6.51 (5.98-6.97)	6.53 (6.0-6.94)	0.55
Left brain (Fixed/Frozen)	13/12	11/15	14/10	0.51
Hemisphere (Left/Right)	12/13	16/10	10/14	0.35
Drug Use (Heavy/Moderate/Social/None)	8/2/2/13	6/10/3/7	0/0/1/23	**0.00002**
Smoking (Yes/No/Unknown)	14/3/8	12/6/8	8/5/11	0.51
Diagnosis	SCZ PARANOID 4 SCZ UNDIFF 21	BD I 21 BD II 5	Control 24	N/A

Table S2. Demographic and clinical information for replicated BrainGVEX data.

	SCZ	BD	CTRL	P-value
Age	42.7 (18-61)	46 (19-64)	50.7 (20-70)	0.09
Sex (F/M)	17/53	20/28	17/46	0.10
Race (Caucasian/Non-Caucasian)	70/0	48/0	63/0	1
PMI	42.7 (9-142)	37.6 (11-115)	23.9 (1.5-90)	0.14
RIN	7.33 (5.5-9.2)	7.34 (5.5-9)	7.70 (5.5-9.3)	0.38
Brain Weight	1430 (1090-1670)	1431 (1080-1690)	1441 (1100-1980)	0.25
Hemisphere (Left/Right)	29/41	22/26	19/44	0.20
Diagnosis	SCZ 70	BD 48	Control 63	N/A
Total Reads	63.3M (36.8-97.7)	63.7M (37.6-144)	59.0M (34-126)	0.46

Table S3. Overlap of genes in the daM and genes with rare variants.

Total	**Genes**
106	*ABI1, ABLIM1, ACO1, ADD3, AHCYL1, AHCYL2, AHNAK, ALDH1L1, ALDH2, ALDH4A1, ALDH6A1, ALDOC, APOE, AQP4, ATP1A2, ATP1B2, AXL, BCAR3, C17orf49, CLTCL1, CLU, CNTFR, CPE, CPT1C, CST3, CTNND2, DCHS2, DDAH1, DTNA, EZR, FADS2, FAM171B, FAT1, FGFR3, FGGY, FIBIN, GJA1, GLI3, GPAM, GPR98, HMGCR, IDH2, ITM2C, ITPR2, KCTD6, KCTD8, KIAA0195, LAMB2, LIMD1, LPAR3, LRP4, LSAMP, LY75-CD302, MAOA, MAP3K5, MYO10, MYOM1, NCAN, NDRG2, NDUFB7, NHSL1, NTRK2, NWD1, OLFM2, PAPSS2, PBXIP1, PCDHGC3, PDGFC, PEX11A, PLCG1, PRDX6, PREX2, PTCH1, PTPRF, PTPRZ1, RPUSD3, RYR3, SFXN5, SLC1A2, SLC25A18, SLC4A4, SLC6A1, SLC7A11, SLC9A3R1, SLC9C2, SNTA1, SPARCL1, SRI, THSD4, TNIK, TNRC6B, TNS3, TRAK1, TRIM32, TRPM3, TRRAP, TTPA, TTYH1, TUBA1B, TUBB2B, WASF3, WBSCR22, WLS, WWC1, ZHX3, ZMIZ1*

Table S4. Transcription factor expression associated with disease in SMRI and BrainGVEX data.

	Differential P in SMRI	Differential P in BrainGVEX
PAX6	0.000152	0.0466
SOX9	0.000738	0.0887
POU3F2	0.00468	0.0899
EPAS1	0.0411	0.199
SOX5	0.0346	0.0312
ZNF423	0.150	0.336

Table S5. Transcription factor expression correlations in BrainGVEX data.

Correlation (r) in BrainGVEX	*SOX9*	*POU3F2*	*PAX6*	*EPAS1*	*SOX5*	*ZNF423*
SOX9		**0.340**	**0.563**	**0.245**	**0.382**	**0.127**
POU3F2			**0.436**	**0.139**	**0.375**	**0.157**
PAX6				**0.264**	**0.440**	**0.108**
EPAS1					**0.237**	0.0288
SOX5						**0.143**
ZNF423						

r is calculated by Pearson's correlation, and bold numbers indicate the correlations are significant ($P < 0.05$).

Table S6. OCA (LEO.NB.OCA) and CPA (LEO.NB.CPA) scores for six transcription factors and one miRNA.

EDGE	LEO.NB.OCA	LEO.NB.CPA
PAX6 -> EPAS1	**1.08**	**2.24**
ZNF423 -> SOX5	**0.639**	**2.04**
POU3F2 -> PAX6	**0.805**	**1.75**
SOX5 -> hsa-miR-320e	**0.396**	**1.08**
POU3F2 -> ZNF423	**0.492**	**1.74**
POU3F2 -> hsa-miR-320e	**0.526**	**1.55**
ZNF423 -> hsa-miR-320e	0.137	0.117
POU3F2 -> SOX9	**0.327**	**1.48**
EPAS1 -> hsa-miR-320e	0.0805	0.654
PAX6 -> hsa-miR-320e	0.066	0.652
SOX9 -> hsa-miR-320e	0.0533	**0.824**
POU3F2 -> EPAS1	0.0637	0.587
ZNF423 -> PAX6	**0.57**	**2.62**
SOX5 -> POU3F2	0.0765	0.452
ZNF423 -> EPAS1	**0.75**	**1.3**
SOX9 -> EPAS1	**1.07**	**2**
SOX9 -> PAX6	**0.879**	**2.57**
SOX9 -> ZNF423	0.0317	**1.9**

ZNF423 -> SOX9	-0.0317	**2.02**
PAX6 -> SOX9	-0.879	**1.14**
EPAS1 -> SOX9	-1.07	0.0412
EPAS1 -> ZNF423	-1.21	-1.11
PAX6 -> ZNF423	-0.57	**1.55**
PAX6 -> SOX5	-0.272	**1.09**
SOX5 -> PAX6	-0.621	0.0911
PAX6 -> POU3F2	-1.42	-1.82
EPAS1 -> POU3F2	-0.0637	0.277
EPAS1 -> PAX6	-1.33	-1.17
hsa-miR-320e -> POU3F2	-1.04	-1.27
POU3F2 -> SOX5	-0.398	-1.09
hsa-miR-320e -> SOX9	-0.0533	0.706
hsa-miR-320e -> PAX6	-0.066	0.411
SOX5 -> EPAS1	-0.103	-0.0746
SOX5 -> SOX9	-0.904	-0.207
SOX9 -> POU3F2	-0.327	0.381
hsa-miR-320e -> EPAS1	-0.0805	0.154
EPAS1 -> SOX5	-0.129	-0.18
ZNF423 -> POU3F2	-0.492	0.788
SOX9 -> SOX5	-0.555	-0.326
SOX5 -> ZNF423	-0.859	-0.625

hsa-miR-320e -> SOX5	-0.534	-0.589
hsa-miR-320e -> ZNF423	-0.283	-0.712

Notes: Numbers in bold indicate scores those reach significance thresholds and support the regulation directions in the first column. They indicate the scores that exceed the recommended threshold of 0.3 (LEO.NB.OCA) or 0.8 (LEO.NB.CPA).

Table S7. Sequences of RT-qPCR primers and siRNA.

qPCR Sequences	
hsa-miR-320e-F	ACACTCCAGCTGGGAAAGCTGGGTTG
hsa-miR-320e-R	TGGTGTCGTGGAGTCG
CHMP2A-F	ACCGCGAGCGACAGAAACTA
CHMP2A-R	GCACCAAGTCTTTTGCCATGA
VPS29-F	CGCTTGCAATTGCTGGAACT
VPS29-R	TAGCTGGCAAACTGTTGCAC
POU3F2-F	GTTGCCGTTTTGGGGGATTT
POU3F2-R	AATGGAGAGTGGCCAAGAGC
U6-F	CTCGCTTCGGCAGCACA
U6-R	AACGCTTCACGAATTTGCGT
ECM7-F	AGAGCACGTCGGTTTCCTTA
ECM7-R	TCCACTCGAACGGGATCAAA
PSMB4-F	TACTACCGAGATGCCCGTTC

PSMB4-R	TGGGCAATATCCCAGTTGGT
VCP-F	TCCTGATGAGAAGTCCCGTG
VCP-R	TGATGGGTTTGTCTGCCTC
actin-F	CATGTACGTTGCTATCCAGGC
actin-R	CTCCTTAATGTCACGCACGAT
siRNA Sequences	
POU3F2-Homo-1009	GGCGGAUCAAACUGGGAUUTT
Negative control	UUCUCCGAACGUGUCACGUTT

Table S8. Top 50 connected genes in the daM, miRNAs, transcription factors, and their WGCNA parameters.

GENE SYMBOL	P.GS.DIS	MM.GREEN	P.MM.GREEN
PAX6	0.000152	0.902663	1.91E-28
SOX9	0.000738	0.771757	5.32E-16
POU3F2	0.004684	0.72384	1.11E-15
EPAS1	0.041072	0.639195	6.75E-10
SOX5	0.034557	0.587694	2.97E-08
ZNF423	0.150469	0.559124	1.85E-07
Hsa-miR-320e	0.308219	-0.5006	4.79E-06
Hsa-miR-320d	0.358915	-0.484	1.09E-05
Hsa-miR-585	0.068325	0.480296	1.30E-05
Hsa-miR-320c	0.356813	-0.48014	1.31E-05
Hsa-miR-320b	0.416662	-0.45737	3.70E-05
NOTCH2	0.000911	0.962765	3.27E-43
RANBP3L	2.70E-05	0.956877	6.26E-41
MLC1	0.002687	0.956702	7.23E-41
SLC1A3	0.000241	0.951326	4.71E-39
ATP1A2	0.001278	0.950048	1.19E-38
SLC4A4	0.000616	0.947237	8.33E-38
METTL7A	0.001749	0.943896	7.38E-37

GJA1	0.001644	0.941599	3.06E-36
ACSS3	0.0001	0.936206	7.00E-35
AHCYL1	0.000375	0.933401	3.20E-34
CLDN10	0.000279	0.932067	6.45E-34
SPON1	0.004754	0.926392	1.09E-32
ATP13A4	0.001691	0.925287	1.84E-32
GPR98	0.000827	0.921537	1.03E-31
SLC39A12	0.000523	0.920538	1.60E-31
PREX2	0.005226	0.919541	2.48E-31
SLC1A2	0.000105	0.919238	2.83E-31
GABRG1	0.000463	0.919109	2.99E-31
HEPH	0.003477	0.918603	3.72E-31
ACBD7	1.18E-05	0.918531	3.84E-31
GRAMD1C	8.08E-06	0.915702	1.27E-30
GPAM	0.004699	0.914997	1.70E-30
EMX2	9.13E-05	0.91472	1.90E-30
SLC15A2	0.000323	0.912317	5.02E-30
PPAP2B	0.005242	0.91019	1.16E-29
AMOT	0.000731	0.908586	2.15E-29
ALDH6A1	0.000286	0.908428	2.28E-29
MSI2	0.001155	0.908291	2.40E-29
DOCK7	2.37E-05	0.906977	3.95E-29

BBOX1	0.002465	0.905452	6.95E-29
P2RY1	0.008546	0.905398	7.09E-29
GOLIM4	0.003802	0.903587	1.37E-28
LRP4	0.029656	0.903307	1.52E-28
TMEM47	0.000374	0.903108	1.63E-28
GLI3	0.000224	0.90244	2.07E-28
SLC25A18	4.00E-05	0.902367	2.12E-28
ETNPPL	0.00164	0.901824	2.57E-28
UG0898H09	0.000842	0.901504	2.88E-28
GPC5	8.96E-05	0.90022	4.51E-28
NTSR2	0.00705	0.899891	5.06E-28
ACSBG1	0.001933	0.898149	9.20E-28
FUT10	0.000127	0.897883	1.01E-27
RYR3	0.014679	0.896606	1.55E-27
BMPR1B	9.30E-05	0.893657	4.10E-27
RAB31	0.000159	0.891939	7.13E-27
PDLIM5	0.002596	0.890748	1.04E-26
GPR125	0.000255	0.889676	1.46E-26
STON2	0.016622	0.889324	1.63E-26
PBXIP1	0.00563	0.888006	2.45E-26
AHCYL2	0.006422	0.887699	2.69E-26

P.GS.DIS is the disease-associated p-value; MM.RED and p.MM.RED indicates the genes Pearson's correlation r and p-value with module eigengene.

Table S9. *POU3F2* and its putative targets in daM.

TFs	**Putative targets**
POU3F2	*ADD3, ASCL1, DDAH1, ECE1, EFCAB5, EPHX1, FGD6, FIBIN, HIST2H2AC, LRP2BP, MID1, NDP, NEBL, NUMBL, PAX6, PPP1R1B, RANBP3L, SELENBP1, SLC1A3, SLC44A3, SOX9, STK33, TBL1X, TNRC6B, TSPAN12, WDFY3-AS2*