DISEASES AND DISORDERS

Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia

Abolfazl Doostparast Torshizi[1,2], Chris Armoskus[3,4], Hanwen Zhang[5], Marc P. Forrest[6,7], Siwei Zhang[5,8], Tade Souaiaia[3,4], Oleg V. Evgrafov[3,4], James A. Knowles[3,4], Jubao Duan[5,8]*, Kai Wang[1,2,4]*

Applying tissue-specific deconvolution of transcriptional networks to identify their master regulators (MRs) in neuropsychiatric disorders has been largely unexplored. Here, using two schizophrenia (SCZ) case-control RNA-seq datasets, one on postmortem dorsolateral prefrontal cortex (DLPFC) and another on cultured olfactory neuroepithelium, we deconvolved the transcriptional networks and identified *TCF4* as a top candidate MR that may be dysregulated in SCZ. We validated *TCF4* as a MR through enrichment analysis of *TCF4*-binding sites in induced pluripotent stem cell (hiPSC)–derived neurons and in neuroblastoma cells. We further validated the predicted *TCF4* targets by knocking down *TCF4* in hiPSC-derived neural progenitor cells (NPCs) and glutamatergic neurons (Glut_Ns). The perturbed *TCF4* gene network in NPCs was more enriched for pathways involved in neuronal activity and SCZ-associated risk genes, compared to Glut_Ns. Our results suggest that *TCF4* may serve as a MR of a gene network dysregulated in SCZ at early stages of neurodevelopment.

INTRODUCTION

Schizophrenia (SCZ) is a debilitating neuropsychiatric disease that affects about 1% of adults (*1*, *2*), with the heritability ranging from 73 to 83% in twin studies (*3*); however, its underlying genetic architecture remains incompletely understood (*4*). Recent genome-wide approaches have identified a plethora of common disease risk variants (*5*), rare copy-number variants (CNVs) of high penetrance (*6*, *7*), and rare protein-altering mutations (*8*) that confer susceptibility to SCZ. Although exciting, the biology underlying these genetic findings remains poorly understood, prohibiting the development of targeted therapeutic strategies. A major challenge is the polygenic nature of SCZ, where risk alleles in many genes, as well as noncoding regions of the genome, across the full allelic frequency spectrum are involved and likely act in interacting networks (*4*). Therefore, given the large number of genes potentially involved, it has been challenging to identify which are the core set of genes that play major roles in the pathways or networks. It becomes even more challenging if the recent hypothesis of omnigenic model is true for SCZ, in which case, almost all of the genes expressed in a disease-relevant cell type may confer risk for the disease through widespread network interactions with a core set of genes (*9*). Thus, it is imperative to identify disease-relevant core gene networks, and possibly, the network master regulators (MRs), which are more likely to be targeted for therapeutic interventions (*10*, *11*), if they exist. We define the term MR for transcription factors (TFs) and genes inflicting regulatory effects on their targets. We have collected a comprehensive set of 2198 genes curated from three sources (*12*–*14*), which are used in the following sections.

Transcriptional networks, as a harmonized orchestration of genomic and regulatory interactions, play a central role in mediating cellular processes through regulating gene expression. One of the most commonly used network-based modeling approaches of cellular processes are scale-free coexpression networks (*15*). However, coexpression networks and other similar networks are not comprehensive enough to fully recapitulate the entire underlying molecular interactions driving the disease phenotype (*16*). Despite the wide adoption of coexpression network analysis approaches, these methods have several limitations (*17*), such as the inability to infer or incorporate causal regulatory relationships, the difficulty to handle mammalian genome-wide networks with many genes, and the presence of high false-positive rates due to indirect connections. In contrast, information-theoretic deconvolution techniques (*16*) aim to infer causal relationships between TFs and their downstream targets and have been recently successfully applied in a wide range of complex diseases such as cancers (*18*) and neurodegenerative diseases (*19*).

Inferring disease-relevant transcriptional gene networks requires well-powered transcriptomic case-control datasets from disease-relevant tissues or cell types. For SCZ, although abnormal gene expression in multiple brain regions (*20*) and different types of neuronal cells (*21*) may be involved in disease pathogenesis, abnormalities in cellular and neurochemical functions of the dorsolateral prefrontal cortex (DLPFC) have been demonstrated in patients with SCZ (*22*). Furthermore, DLPFC controls high-level cognitive functions, many of which are disturbed in SCZ. The CommonMind Consortium (CMC) RNA sequencing (RNA-seq) data on DLPFC, which includes 307 SCZ cases and 245 controls, is currently the largest genomic dataset on postmortem brain samples for neuropsychiatric disorders (*23*). The CMC study validates ~20% of SCZ loci having variants that can potentially contribute to the altered gene expression, where five of which involve only a single gene such as *FURIN*, *TSNARE1*, and *CNTN4* (*23*). A gene coexpression analysis in the CMC study implicated a gene network relevant to SCZ but does not address the question of whether an MR potentially orchestrates the transcriptional architecture

[1]Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA. [2]Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [3]College of Medicine, SUNY Downstate Medical Center, Brooklyn, NY 11203, USA. [4]Zilkhe Neurogenetic Institute, University of Southern California, Los Angeles, CA 90089, USA. [5]Center for Psychiatric Genetics, North Shore University Health System, Evanston, IL 60201, USA. [6]Department of Physiology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [7]Center for Autism and Neurodevelopment, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA. [8]Department of Psychiatry and Behavioral Neurosciences, University of Chicago, Chicago, IL 60015, USA.
*Corresponding author. Email: jduan@uchicago.edu (J.D.); wangk@email.chop.edu (K.W.)

of SCZ. In the current study, using CMC RNA-seq data, we deconvolved the regulatory processes mediating SCZ to identify critical MRs and to infer their role in orchestrating cellular transcriptional networks. Using another set of independent SCZ case-control RNA-seq data on cultured neuronal cells derived from olfactory neuroepithelium (CNON) (*24*), we observed five MRs shared between both datasets. Among the top candidate MRs identified from both datasets, we selected *TCF4*, a leading SCZ risk gene from a genome-wide association study (GWAS) (*5*), for empirical validation of predicted network activity using human-induced pluripotent stem cell (hiPSC)–derived neurons. A general overview of both the computational and experimental stages of this study is provided in Fig. 1. On the basis of the results, we identified *TCF4* as an MR that likely contributes to SCZ susceptibility at the early stages of neurodevelopment.

RESULTS

Deconvolution of transcriptional network of the CMC data uncovers MRs

The CMC transcriptomic study implicated 693 differentially expressed (DE) genes in SCZ postmortem brains (*23*). However, the magnitude of case-control expression differences was small, posing a challenge to inferring their biological relevance. We thus deconvolved the transcriptional regulatory networks to infer the MRs and their transcriptional targets (regulons), from which we constructed their corresponding subnetworks. To achieve this, we used Algorithm for Reconstruction of Accurate Cellular Networks (ARACNe) to reconstruct cellular networks (see Materials and Methods) (*17*). This approach first identifies gene-gene coregulatory patterns by the information theoretic measure mutual information (MI), followed by network pruning through removing indirect relationships in which two genes are coregulated through one or more intermediate entities. This enabled us to observe network relationships that are more likely to represent direct transcriptional interactions (e.g., TF binding to a target gene) or posttranscriptional interactions (e.g., microRNA-mediated gene inhibition). We identified 1466 genes as hubs by computational analysis of the CMC RNA-seq data (table S1), where the corresponding subnetworks for these hub genes contain 24,548 transcriptional interactions (table S2).

Using the reconstructed network, we next performed a virtual protein activity analysis to investigate the activity of the identified hub genes by taking into account the expression patterns of their downstream regulons through a dedicated probabilistic algorithm called VIPER (Virtual Inference of Protein-activity by Enriched Regulon analysis) (*18*). This method exploits the regulator mode of action, the confidence of the regulator-target gene interactions, and the pleiotropic features of the target regulation. We fed the generated network into VIPER to evaluate whether any of the identified hub

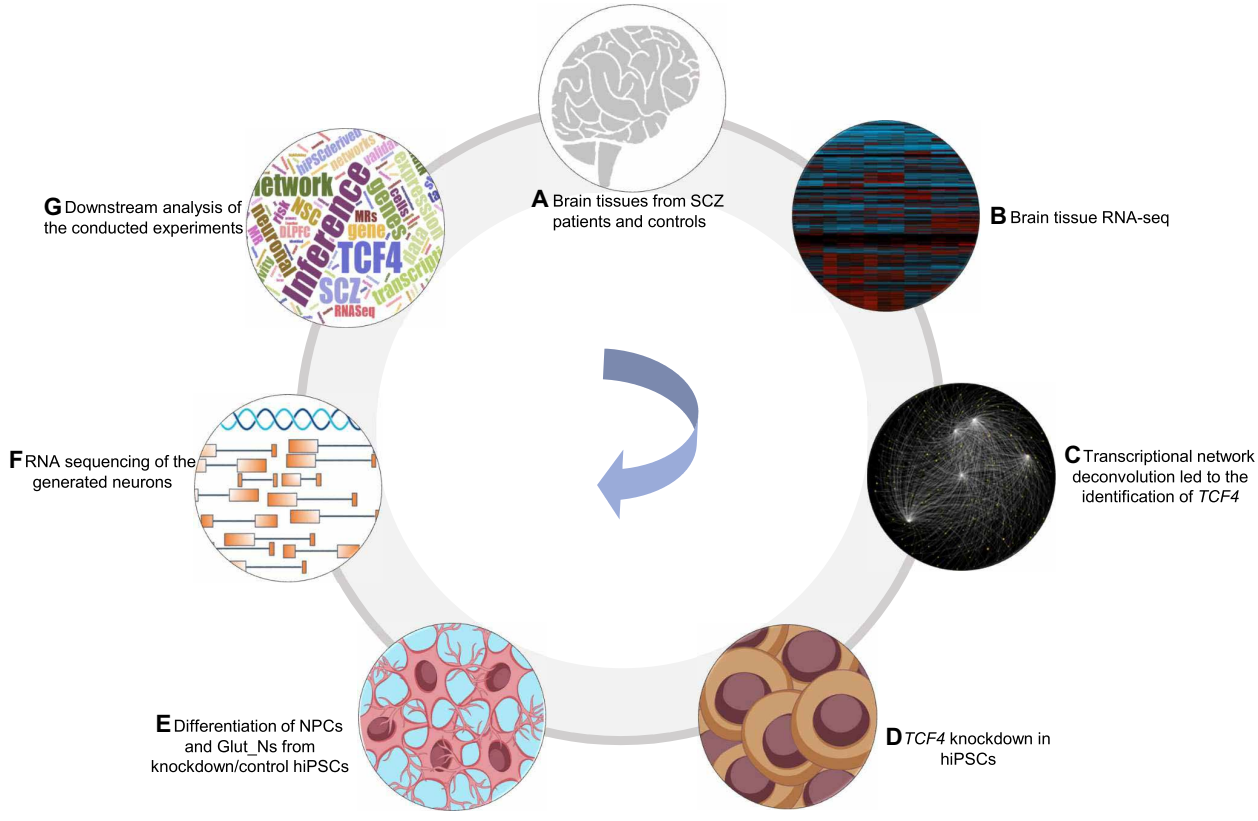

Fig. 1. An overview of the current study. (**A** and **B**) Obtaining gene expression data generated from DLPFC (CMC data) and later CNON data as an independent validation. (**C**) Transcriptional network reconstruction based on the RNA-seq data using algorithm for reconstruction of accurate cellular networks (ARACNe) and identification of *TCF4* as a disease-relevant MR. (**D**) Knockdown of *TCF4* in hiPSCs derived from a patient with SCZ and two unaffected individuals. (**E**) Differentiation of hiPSC into neural progenitor cells (NPCs) and glutamatergic neurons (Glut_Ns). (**F**) RNA-seq of the generated neuronal cells with and without *TCF4* knockdown. (**G**) Downstream inferences and analysis of the conducted experiments.

genes in the network has a significant regulatory role in the expression degrees of its downstream target genes. We then ranked the hub genes by VIPER-adjusted activity *P* values (representing the statistical significance of being an MR; see Materials and Methods). We further defined a short list of genes potentially regulating a large set of targets ($n = 93$). VIPER outputs a list of genes whose activity levels are highly correlated with the expression of their regulons in the input network. However, as recommended by the VIPER developers, many of these genes are not regulators of their corresponding targets in practice. Therefore, they recommended focusing on the highly active genes in the VIPER output, which are TFs or histone modification genes. As a result, among genes with highly significant VIPER *P* values, we only took the genes from the curated list of TFs as potential MRs. Moreover, as suggested by the VIPER developers, we set the activity *P* values at 0.005 to minimize the number of false-positive MRs and to avoid obtaining candidate MRs with very few numbers of regulons despite high-activity *P* values.

To reduce false-positive findings and to focus on MRs known to exert regulatory effects on their targets, we intersected the short list of potential MRs with our curated set of genes (fig. S1). This process resulted in five potential MRs: *TCF4*, *NR1H2*, *HDAC9*, *ZNF436*, and *ZNF10*. On the basis of previous studies (*25*), identification of the activated MRs in a disease state may be confounded when several of its regulons are all targets of a different transcriptional factor, that is, the "shadow effect" (*25*). Because of the highly pleiotropic nature of transcriptional regulations, we evaluated the shadow effects in the constructed networks. We should note that, during the process of network deconvolution, most of the potential co-regulatory edges have been trimmed to preserve the scale-free topology of the network, so we expect to observe no shadow effects. As expected, no shadow connections were observed in this analysis, indicating a high confidence in the constructed transcriptional network to reflect a likely real regulatory connection between an MR and its targets.

Network deconvolution of CNON data independently identifies *TCF4* as an MR

We next attempted a replication of the identified MRs in an independent RNA-seq dataset (with 23,920 transcripts) from CNON of 143 SCZ cases with 112 controls (*24*). The rationale is that if the *TCF4* network identified in the CMC dataset is more of neuronal origin, then confirmation of the *TCF4* MR in the CNON model would add supporting evidence for the predicted MR, given that olfactory cells are promising models in biological psychiatry and neuroscience (*26*). We acknowledge that the CNON culture likely contains non-neuronal cells such as epithelial cells. Because the CMC data are generated from postmortem brain tissues comprising both neuronal and non-neuronal cells, we first estimated the percentage of neuronal cells in postmortem brain DLPFC (used by CMC). On the basis of a recent study of cellular composition of the same brain region (DLPFC) by Lake *et al.* (*27*), in which 36,166 single cells were assayed for gene expression, we found that 26,064 cells (~72%) were neuronal cells, indicating that neuronal cells constitute a significant proportion of the brain cell types in postmortem DLPFC. Furthermore, of 101 *TCF4* target genes predicted in CMC, 31 were found to have significantly higher expression in neuronal cells (enrichment $P = 2.8 \times 10^{-10}$ by Fisher's exact test, fold enrichment = ~3.5), while the rest were all expressed in neuronal cells. We thus conclude that the constructed networks from the CMC dataset stem from neuronal cells, despite possible influence from the small fraction of non-neuronal cells.

Using setup criteria similar to those used on the CMC data, for CNON data, we observed 1836 TFs or expression regulatory nodes in the constructed network, including 34,757 predicted interactions (tables S3 and S4). To further analyze the activity of the identified hub genes, we ran VIPER on the constructed network and identified the top MRs. The five identified MRs in the CMC network (*TCF4*, *NR1H2*, *HDAC9*, *ZNF436*, and *ZNF10*) were observed to be MRs within the network created from CNON data. All of these five MRs were significant in both CMC and CNON data [false discovery rate (FDR) values are provided in table S11]. On the basis of the deconvolution analysis, these five MRs were predicted to regulate the expression of 101, 68, 36, 66, and 43 regulons combined, respectively (table S5). Among these genes, *TCF4* is a well-established SCZ-associated risk factor identified by GWAS (*5*), which has been validated in several additional studies (*28*, *29*). Rare variants in *TCF4* have also been implicated in Pitt-Hopkins syndrome and intellectual disability (*30*). However, its regulatory mechanisms at a systems level are poorly understood. Although the other identified MR may also potentially contribute to SCZ pathogenesis, given the prior biological knowledge on *TCF4*, we have focused on *TCF4* for additional analysis.

To further evaluate the connection between *TCF4* dysregulation and SCZ, we conducted pathway enrichment analysis using WebGestalt (*31*). The predicted *TCF4*-associated regulons were enriched for several pathways such as Notch signaling pathway (FDR = 1×10^{-4}) and long-term depression (FDR = 8×10^{-4}). In addition, to check whether the promoter region of the predicted *TCF4* targets is enriched for *TCF4* binding motif sequence, we conducted TF binding enrichment analysis (TFBEA) using JASPAR (*32*). The *TCF4* motif sequence is extracted from JASPAR and is used for binding enrichment analysis (Fig. 2A). We extracted the sequence of the target genes 2 kb upstream and 1 kb downstream of their transcription start sites (TSSs) and used a significance threshold of 0.80 to test the relative enrichment scores of its target genes. We found that the observed binding enrichment score is significantly higher than expected by chance (two-sided *t* test $P = 2.2 \times 10^{-16}$) based on a random list of genes outside the *TCF4* subnetwork (Fig. 2B), suggesting that the predicted *TCF4* regulons can be directly targeted by *TCF4* binding events.

Predicted *TCF4* regulons are significantly enriched for *TCF4* targets from ATAC-seq and ChIP-seq experiments

Our observed enrichments of TF binding motif of *TCF4* in its regulons were purely a DNA sequence–based prediction, so we next tested whether *TCF4* can physically interact with its regulons. To address this question, we first identified the target genes of *TCF4* by examining the empirical TF occupancy (i.e., TF binding footprints) inferred by the Protein Interaction Quantification (PIQ) tool (*33*, *34*) from our previous assay for transposase-accessible chromatin (ATAC-seq) data in hiPSC and hiPSC-derived glutamatergic neurons (Glut_Ns) (*34*, *35*). Twenty-nine target genes of 101 regulons of *TCF4* were found to be bound by *TCF4* in ATAC-seq from Glut_Ns (total *TCF4* targets in ATAC-seq data in Glut_Ns = 3276) (enrichment $P = 3.3 \times 10^{-5}$ by Fisher's exact test, fold enrichment = ~1.93), whereas only one *TCF4* target was bound by *TCF4* in hiPSCs (total *TCF4* targets in ATAC-seq data in hiPSCs = 172) (no enrichment $P = 0.417$ by Fisher's exact test).

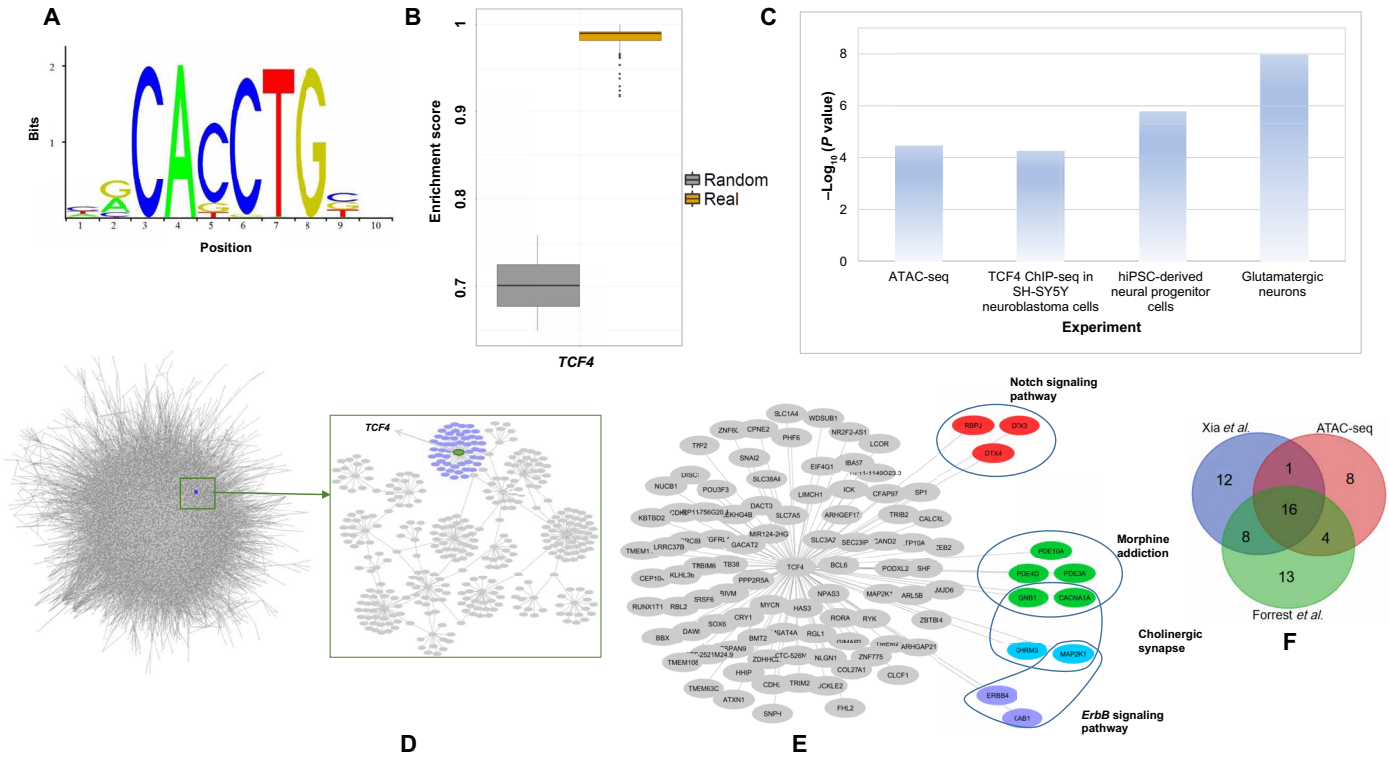

Fig. 2. Summary of *TCF4* regulons in the CMC and CNON data. (**A**) *TCF4* motif from the JASPAR database used in the analysis. (**B**) *TCF4* binding site enrichment scores (based on *TCF4* motifs from the JASPAR database) among the *TCF4* regulons compared to that of a set of random genes. (**C**) Enrichment *P* values of the *TCF4* regulons, compared with several gene sets, including the predicted *TCF4* sites from ATAC-seq, the differentially expressed genes from our *TCF4* knockdown experiments, and the predicted *TCF4* binding sites in neuroblastoma cells by ChIP-seq (*P* value obtained from Fisher's exact test). (**D**) A schematic of the network created from CMC data as well as the *TCF4* targets. (**E**) *TCF4* targets from CMC and CNON data combined with some of the associated pathways. (**F**) List of overlapping predicted *TCF4* targets with an ATAC-seq on hiPSC–Glut_Ns and two ChIP-seq datasets.

Furthermore, with the list of empirically determined *TCF4* targets in a recent *TCF4* chromatin immunoprecipitation sequencing (ChIP-seq) study on human SH-SY5Y neuroblastoma cells (*36*) (a total of 10,604 peaks for 5436 unique candidate target genes), we observed an overlap of 41 *TCF4* targets between *TCF4* ChIP-seq data and our inferred *TCF4* regulons, representing a significant enrichment ($P = 5.28 \times 10^{-5}$ by Fisher's exact test, fold change = ~1.65). Recently, Xia *et al.* (*37*) have examined the genes regulated by *TCF4* in neuron-derived SH-SY5Y cells using ChIP-seq. They have reported a total of 11,322 peaks for 6528 unique candidate target genes. We observed 37 genes overlapping with the *TCF4*-predicted target genes ($n = 101$, $P = 3 \times 10^{-3}$ by Fisher's exact test, fold change = ~1.5). Together, both ATAC-seq and ChIP-seq analyses demonstrated that our list of predicted regulons is significantly enriched for *TCF4* transcriptional regulatory targets, reflecting the differences in different cell types (Fig. 2C). To further probe these overlapping genes, we compiled the predicted targets of *TCF4* with the above ChIP-seq and ATAC-seq datasets on Glut_Ns (Fig. 2F and table S10). We observed that a large fraction of *TCF4* targets are shared between two or all of the compared datasets, which is a strong indication of the reliability of the inferred *TCF4* targets during network construction.

Network perturbation by *TCF4* knockdown in hiPSC-derived neurons supports *TCF4* as an MR and a regulator of neurodevelopment

To further investigate the functional relevance of *TCF4* dysregulation in SCZ, we examined how the predicted *TCF4* transcriptional subnetwork changes with decreased *TCF4* expression in hiPSC-derived neuronal cells. We acknowledge that similar experiments have been performed on the SH-SY5Y neuroblastoma cell line (*36*, *38*), which is a cell type that may be less relevant to SCZ. We first derived relatively pure NPCs ($NESTIN^+/PAX6^+/SOX2^+$) from hiPSCs ($TRA\text{-}1\text{-}60^+/SSEA4^+/NANOG^+/OCT4^+$) (Fig. 3, A to D), which were further differentiated into glutamatergic neurons (Glut_Ns) ($VGLUT1^+/MAP2^+$) (Fig. 3E). hiPSCs were captured from a patient with SCZ (see Materials and Methods). To capture possible differential effects of *TCF4* on gene expression during different developmental stages, we carried out *TCF4* knockdown by lentiviral short hairpin RNA (shRNA) in NPCs of day 3 after initiating neural differentiation (Fig. 3, C and D) and in the early stage of Glut_Ns of day 14 after initiating differentiation (Fig. 3E). We achieved significant *TCF4* knockdown in both cell stages: *TCF4* expression in the knockdown group was reduced to 32.2 and 48.6% of that in the control shRNA groups [as measured by quantitative polymerase chain reaction (qPCR), $P < 0.001$; Fig. 3, F and G]. We also performed RNA-seq analysis on the same set of cells with three biological replicates at both time points in both cell types and found that the reduction of *TCF4* transcript level was consistent with the qPCR results (Fig. 3J).

Transcriptomic analysis of the RNA-seq data of *TCF4* knockdown revealed 4891 DE genes in the day 3 group (NPCs data; Fig. 3H), including 2330 up-regulated and 2561 down-regulated genes (FDR < 0.05) (see Materials and Methods). The day 14 group (Glut_Ns data; Fig. 3I) showed 3152 DE genes, of which 1862 were

up-regulated and 1290 were down-regulated (FDR < 0.05) (table S7). To examine whether the gene network perturbed by *TCF4* knockdown in hiPSC-derived neurons is concordant with the predicted *TCF4* regulons, we compared the list of genes affected by *TCF4* knockdown and the list of 101 *TCF4* regulons from the CMC/CNON data. We found that 39 predicted *TCF4* regulons showed altered expression by *TCF4* knockdown in NPCs (enrichment $P = 1.55 \times 10^{-6}$ by Fisher's exact test, fold change = ~1.75), while 33 predicted *TCF4* regulons were dysregulated after *TCF4* knockdown in Glut_Ns (enrichment $P = 1.05 \times 10^{-8}$ by Fisher's exact test, fold change = ~2.28). This difference between the enrichment *P* values of the conducted experiments compared to the neuroblastoma cell line confirmed the importance of tissue-specific investigation of transcriptional regulation (although the set of regulons may be more conserved across tissues) and further supported the role of *TCF4* as an MR in a transcriptional network (Fig. 2, D and E) relevant to neurodevelopment (Fig. 2E).

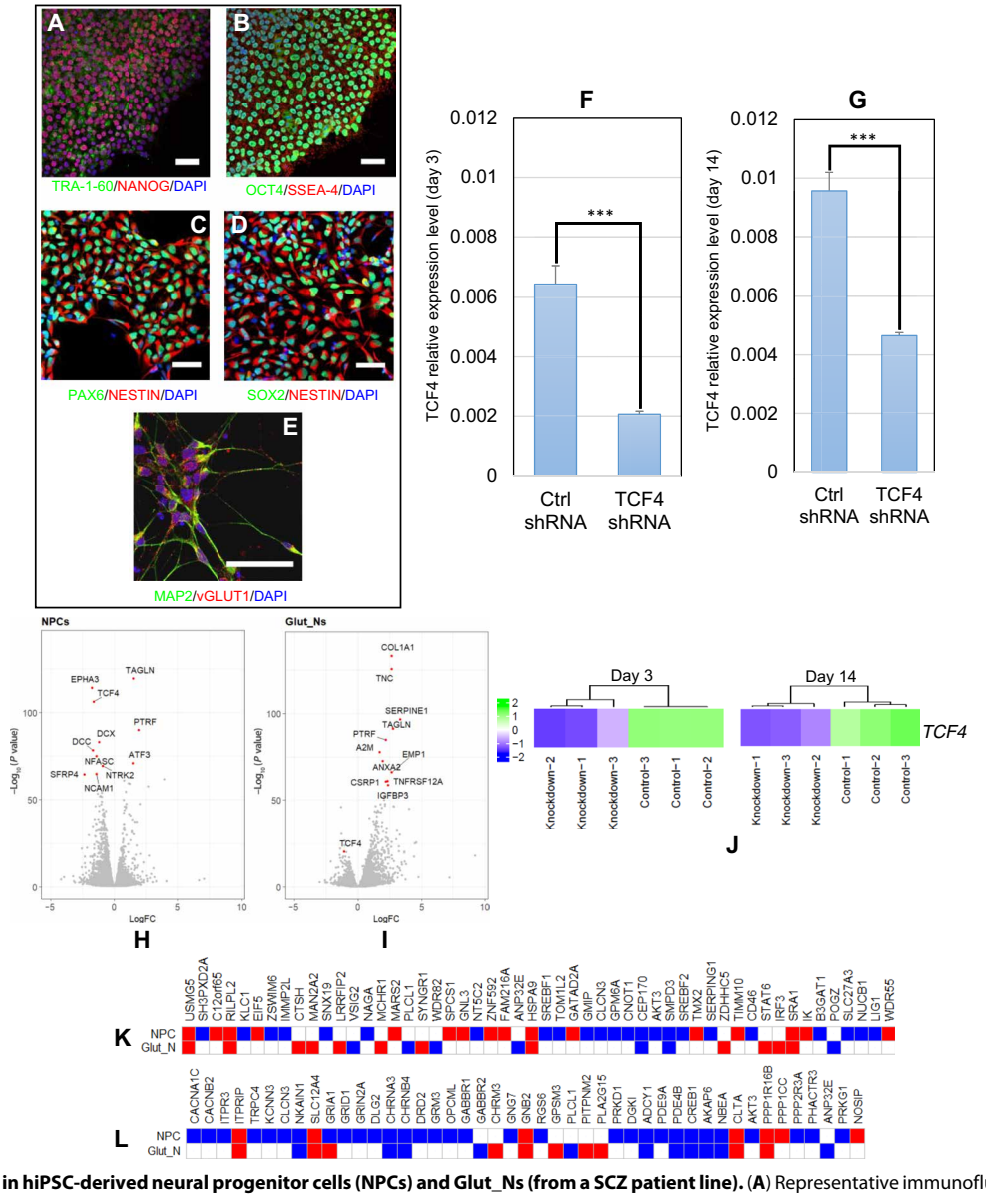

Fig. 3. *TCF4* knockdown in hiPSC-derived neural progenitor cells (NPCs) and Glut_Ns (from a SCZ patient line). (**A**) Representative immunofluorescence (IF)–staining images of hiPSCs. hiPSCs are stained positive for pluripotency markers TRA-1-60 (green) and NANOG (red). (**B**) Representative IF-staining images of hiPSCs. hiPSCs are stained positive for OCT4 (green) and SSEA-4 (red). (**C**) Representative IF-staining images of NPCs (3 days after plating cells): NPCs are stained positive for PAX6 (green) and NESTIN (red). (**D**) Representative IF-staining images of NPCs (3 days after plating cells): NPCs are stained positive for NESTIN (red) and SOX2 (green). (**E**) Representative image of day 14 neurons stained positive for MAP2 (green) and vGLUT1 (red). (**F**) *TCF4* knockdown efficiency measured by qPCR in NPCs (3 days after plating). (**G**) *TCF4* knockdown efficiency measured by qPCR in Glut_Ns (14 days after neuronal induction). (**H**) Volcano plot of DE genes in NPCs. (**I**) Volcano plot of DE genes in Glut_Ns. (**J**) *TCF4* expression levels at days 3 and 14 in RNA-seq data (**K**) Overlap of the DE genes upon *TCF4* knockdown with a list of GWAS-implicated SCZ risk genes (up- and down-regulated genes are shown in red and blue, respectively; white cells indicate that the gene is not DE). (**L**) Overlap of the DE genes upon *TCF4* knockdown with a list of credible GWAS risk loci (up- and down-regulated genes are shown in red and blue, respectively; white cells indicate that the gene is not DE). Scale bars, 50 μm in all images. Cell nuclei are stained with 4′,6-diamidino-2-phenylindole (DAPI) (blue) (A to E), and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is used as the endogenous control to normalize the *TCF4* expression for qPCR. Error bars, mean ± SD ($n = 4$). ***$P < 0.001$, Student's *t* test, two-tailed, heteroscedastic.

As our observed transcriptomic changes were from an SCZ patient line, we further carried out *TCF4* knockdown in cells derived from hiPSC lines of two unaffected individuals (cell lines CD07 and CD09; Fig. 4, A to D) to account for variations due to individual genetic backgrounds. Because *TCF4* regulons were initially found more enriched in the list of genes perturbed by *TCF4* knockdown in Glut_Ns (Fig. 2C), we have focused on examining the robustness of our findings on Glut_Ns with three independent cultures for each condition. Using qPCR, we examined the knockdown efficiency in both cell lines (~40%) (Fig. 4, E and F). We also confirmed the relatively high homogeneity of the differentiated neurons (80 to 85%) (Fig. 4G). Further transcriptomic analysis from bulk RNA-seq data also confirmed the successful reduction in *TCF4* expression in cell line CD09 (negative binomial test, FDR = 1.55×10^{-14}), as well as cell line CD07 (FDR = 2.35×10^{-12}) (Fig. 4H). During the quality assessment procedure, we removed one sequenced sample as an outlier from downstream analysis (fig. S10, see Materials and Methods). In total, we identified 2304 DE genes in CD09 and 1744 DE genes in CD07 (FDR < 0.05). Next, we examined the enrichment of *TCF4* network regulons in the list of DE genes from both cell lines. We observed 22 predicted *TCF4* regulons to be dysregulated after *TCF4* knockdown in Glut_Ns of CD09 (enrichment $P = 6.85 \times 10^{-4}$ by Fisher's exact test, fold enrichment = ~2.1) and 17 predicted *TCF4* regulons in CD07 (enrichment $P = 2 \times 10^{-3}$ by Fisher's exact test, fold enrichment = ~2). Moreover, we compared the log fold change of the genes in the two generated cell lines (CD07 and CD09) and compared them with the initial data on Glut_Ns (Fig. 4, K and L). To check the consistency of cellular responses to the *TCF4* knockdown, we compared the log fold changes of the genes in the CD07 and CD09 lines with the initial SCZ Glut_Ns data. We observed a correlation coefficient of 0.54 and 0.50 for CD07 and CD09, respectively, between the log fold changes of gene expression ($P = 10^{-15}$ and 10^{-14} for CD07 and CD09, respectively), which further supports our findings, implying that *TCF4* knockdown had a consistent effect on the transcriptome change in the several cell lines used in our study. Together, these observations reconfirmed the role of *TCF4* as an MR in the created regulatory networks across individuals.

Although our deconvolution analysis identified 101 genes in the *TCF4* subnetwork, we reasoned that it is likely part of a much larger transcriptional network that would contain neurobiological pathways relevant to SCZ. Previous studies in mice (*39*) and neuroblastoma cells (*38*) have demonstrated that *TCF4* dysregulation affects a large number of genes and pathways; however, *TCF4*-related pathways have never been explored in human NPCs or Glut_Ns, a much stronger model for studying psychiatric disorders. To uncover the cellular pathways regulated by *TCF4* in these cell types, we performed gene ontology (GO) and pathway analyses on the genes perturbed by *TCF4* knockdown.

By using WebGestalt (*31*), we identified a number of shared gene pathways, including focal adhesion, axon guidance, mitogen-activated protein kinase signaling pathway, and apoptosis (table S6 and figs. S2 to S6). In table S6, we showed the altered pathways at both time points as a result of *TCF4* knockdown and the detailed list of up- and down-regulated genes in each pathway. We found almost similar enrichment for "axon guidance" (FDR = 1.521×10^{-6} and fold enrichment = 1.77 in NPCs versus FDR = 5.941×10^{-5} and fold enrichment = 1.81 in Glut_Ns) while observing some unique pathways (long-term potentiation, neurotrophin signaling, and mTOR signaling) relevant to neuronal activity. An independent pathway analysis using MetaCore (see Materials and Methods and figs. S7 and S8) also showed consistent results. The top gene expression changes in NPCs converged on early neurodevelopmental processes such as axon guidance, neurogenesis, and attractive and repulsive receptors, which guide neuronal growth and axon targeting during development. In contrast, the most significant categories in Glut_Ns involved cell adhesion and cell-matrix interactions, suggesting that *TCF4* regulates different subsets of genes during development. In parallel, the MetaCore analysis revealed that the up-regulated genes had distinct functions in NPCs and Glut_Ns, involving mRNA translation and cell-matrix interactions, respectively. However, down-regulated genes shared a number of categories, including axon guidance being the most significant pathway.

For both up-regulated and down-regulated genes by *TCF4* knockdown, NPCs showed more enriched neuronal activity–related gene pathways than in Glut_Ns and with greater significance (tables S8 and S9). It is also noteworthy that most altered neuronal activity–related pathways were enriched in the down-regulated set of genes upon *TCF4* knockdown. These results suggest that *TCF4* gene targets act at the early stage of neurogenesis and may be more relevant to SCZ biology. We conducted a pathway enrichment analysis (Fig. 4, I and J) on our independently sequenced cell lines from unaffected individuals, which yielded very similar outcomes compared to the altered pathways upon *TCF4* knockdown in the Glut_Ns in the SCZ cell line. These similarly altered pathways such as focal adhesion and p53 signaling pathway indicate that, despite cross-individual genetic variations, *TCF4* plays a critical role in rewiring the transcriptional circuitry of SCZ.

Higher *TCF4* gene network expression activity in NPCs may be more reflective of SCZ case-control difference in the prefrontal cortex

Although analyses of *TCF4* knockdown in NPCs and Glut_Ns supported the role of *TCF4* as an MR, the *TCF4* gene network (interactome) expression activity, i.e., the correlation between *TCF4* expression and its regulon's expression patterns, may be different in NPCs and Glut_Ns cellular stages. This cell type–specific expression of the network activity may inform when the temporal expression of *TCF4* is most relevant to SCZ. Toward this end, we carried out gene set enrichment analyses (GSEAs) in the CMC/CNON data, as well as the *TCF4* knockdown RNA-seq data in NPCs and in Glut_Ns. We found that the SCZ case-control expression differences of *TCF4* regulons in CMC data (i.e., prefrontal cortex) appeared to be more correlated with the *TCF4* regulon expression changes upon *TCF4* knockdown in NPCs than in Glut_Ns. To further demonstrate the closer similarity of the *TCF4* interactome expression activity in the prefrontal cortex and in NPCs, we depicted the GSEA enrichment scores in three datasets (CMC/CNON, NPCs, and Glut_Ns) in a Circos plot (fig. S9A). We found that the pattern of SCZ case-control expression differences of *TCF4* and its regulon in the prefrontal cortex appeared to be more similar to the pattern of higher *TCF4* network expression activity in NPCs than in Glut_Ns. Similarly, the GSEA rank metric score of the *TCF4* regulons also showed that the SCZ case-control differential *TCF4* interactome expression pattern in the prefrontal cortex was more comparable to that in NPCs than in Glut_Ns (fig. S9B). We acknowledge that the prefrontal cortex, from which CMC data is generated, largely consists of differentiated cells. In this regard, our observation of higher *TCF4* expression and its positively correlated regulon expression activity

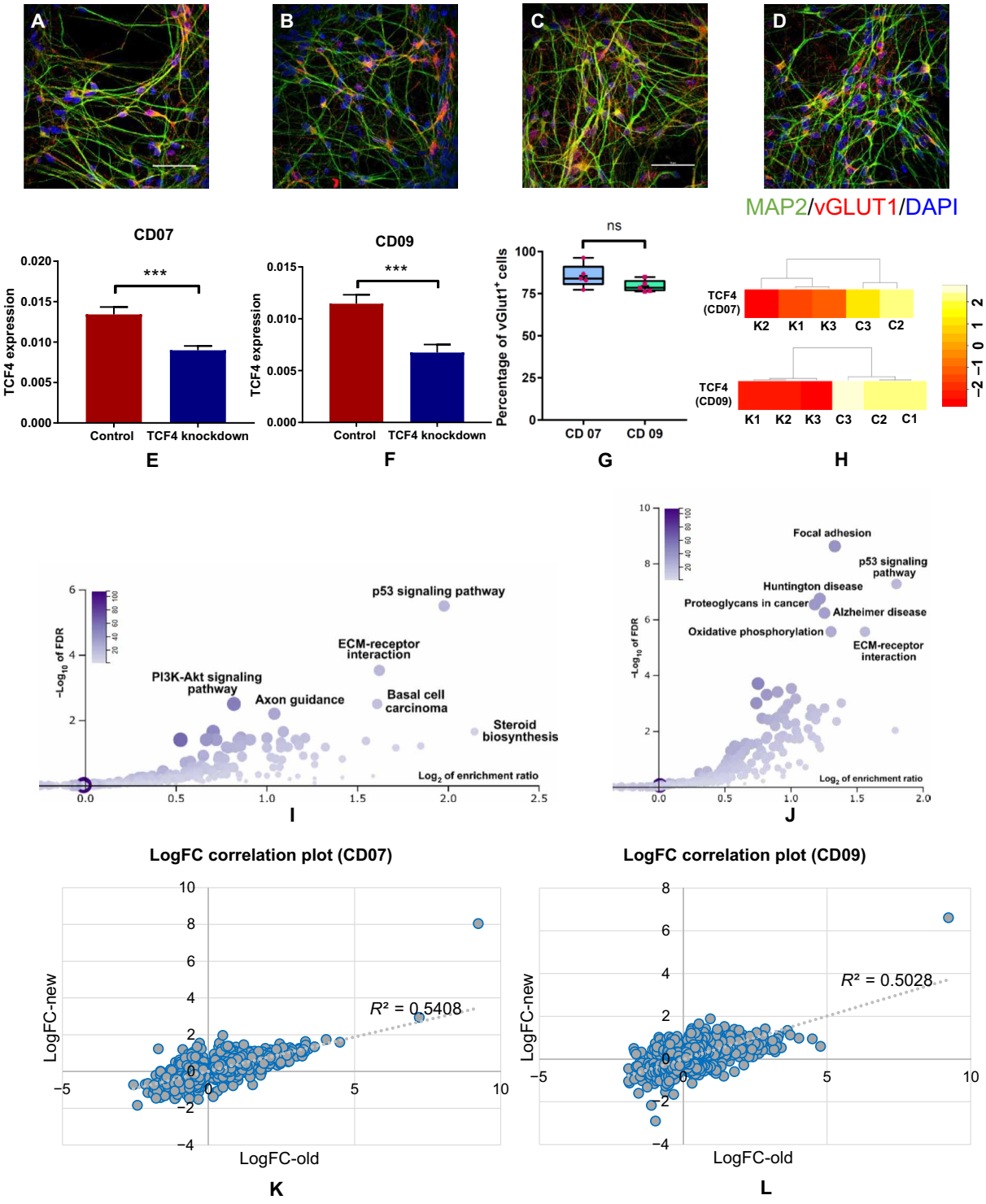

Fig. 4. *TCF4* knockdown in hiPSC-derived Glut_Ns in two independent cell lines CD07 and CD09 from unaffected control subjects. (**A** and **B**) IF staining of hiPSC-derived Glut_Ns from the CD07 control line. (**C** and **D**) IF staining of hiPSC-derived Glut_Ns from the CD09 control line; MAP2, green; vGlut1, red; DAPI, blue. Scale bar, 50 μm. (**E** and **F**) Real-time qPCR (qRT-PCR) result of *TCF4* expression level in two different lines, demonstrating knockdown efficiency; GAPDH is used as the endogenous control to normalize the *TCF4* expression for qPCR; error bars, mean ± SD ($n = 6$ to 8); ***$P < 0.001$, Student's t test, two-tailed, heteroscedastic. (**G**) Percentage of vGlut1-positive cells derived from two different hiPSC lines; cell counts were from five images in each line. (**H**) *TCF4* expression levels in knockdown (K1, K2, and K3) and control samples (C1, C2, and C3) in the generated RNA-seq data on CD07 and CD09. (**I**) Pathway enrichment analysis results on the identified DE genes in CD07 (color legend indicates the number of overlapped DE genes with the corresponding pathway). (**J**) Pathway enrichment analysis results on the identified DE genes in CD09 (color legend indicates the number of overlapped DE genes with the corresponding pathway). (**K**) Correlation plot of the log fold changes (logFC) of the Glut_Ns from the SCZ cell line ("old") and CD07. (**L**) Correlation plot of the log fold changes of the Glut_Ns from the SCZ cell line ("old") and CD09. ns, not significant.

in NPCs is somewhat unexpected. One explanation could be that *TCF4* expression peaks during early cortical development (*40*). One of the limitations of our observation is that both Glut_N and NPC represent early neurodevelopmental stages, which would not recapitulate the actual gene network in prefrontal cortex in adult brain. Nonetheless, our observation is intriguing and may arguably imply that *TCF4* may be an SCZ risk factor at the early stages of neurodevelopment and neurogenesis, which warrants further investigation.

The *TCF4* gene network in NPCs may be more relevant to SCZ disease biology

Given that the expression pattern of *TCF4* and its regulons in the prefrontal cortex is more correlated to NPCs than Glut_Ns (fig. S9A), we reasoned that the *TCF4* gene network in NPCs may be more relevant to SCZ disease biology. To test this hypothesis, we have examined the enrichments of genes regulated by *TCF4* in NPCs, as compared to Glut_Ns for disease-relevant gene pathways, SCZ GWAS risk genes, and de novo SCZ mutations. We conducted a series of comparative tests. To test whether the *TCF4* gene network acting in NPCs is more relevant to the genetic etiology of SCZ, we compared a list of known SCZ susceptibility genes from genome-wide significant risk loci (*5*) that showed gene expression change upon *TCF4* knockdown in NPCs and in Glut_Ns. We found that 52 of those genes [out of the 148 genes reported in (*5*)] were altered by *TCF4* across at least one of two time points in either NPCs or Glut_Ns. Of these 52 SCZ credible genes that showed *TCF4*-altered expression, 63.5% were uniquely found in NPCs (n = 33, Fisher's exact test = 0.019, enrichment factor = 1.5), while 25% (n = 13, Fisher's exact test = 0.950, enrichment factor = 0.6) were uniquely found in Glut_Ns (Fig. 3K). Furthermore, for the 92 SCZ-associated single-nucleotide variants (SNVs) that were indexed for genes related to neuronal electrical excitability and neurotransmission (*41*), 45 had altered expression by *TCF4* knockdown in either NPCs or Glut_Ns. Among them, 24 genes (53.34%, Fisher's exact test = 0.021, enrichment factor = 1.17) were uniquely found in NPCs, while only seven genes (15.6%, Fisher's exact test = 0.90, enrichment factor = 0.54) were uniquely found in Glut_Ns (Fig. 3L). Similarly, we compiled 176 candidate SCZ risk genes from the CLOZUK GWAS study (*42*), and 54 genes were altered by *TCF4* in either NPCs or Glut_Ns, where 72% (n = 39, Fisher's exact test = 0.035, enrichment factor = 1.1) of the genes were uniquely found in NPCs compared to ~17% (n = 9, Fisher's exact test = 0.98, enrichment factor = 0.36) of the genes uniquely found in Glut_Ns. Last, we conducted SCZ de novo SNV (dnSNV) enrichment analysis to probe the overlap of *TCF4*-associated gene expression changes and dnSNV previously reported in SCZ (Fig. 5A) using a list of de novo SNVs from patients with SCZ and control subjects (see Materials and Methods). We found that, while *TCF4*-affected genes in NPCs showed significant enrichment for both loss-of-function (LoF) and missense SCZ dnSNVs, and coding dnSNVs in general (Fig. 5, B and D), only LoF dnSNVs were enriched in Glut_Ns (Fig. 5, C and D). The analyses of both common variants and dnSNVs associated with SCZ in *TCF4*-affected genes nominated NPC as a more disease-relevant cell type and stage of development for the *TCF4* gene network. Together, our results suggest that *TCF4* and its regulons in the early stage of neurodevelopment and neurogenesis may be more relevant for SCZ pathogenesis. However, additional studies should be conducted to further validate this hypothesis.

DISCUSSION

Like other complex disorders, SCZ is polygenic or even possibly omnigenic (*9*), with hundreds or even thousands of susceptibility genes each with a small effect size. A major challenge for understanding the biological implications of genetic findings is to identify the core gene networks in disease-relevant tissues or cell types (*9*). In this study, with the aim of uncovering the gene expression–mediating drivers that may contribute to SCZ pathogenesis, we have used an approach to assembling tissue-specific transcriptional regulatory networks on two independent RNA-seq datasets from DLPFC (CMC data) and cultured cells derived from olfactory neuroepithelium (CNON data). CNON was used as a validation dataset to evaluate whether findings from brain samples can be seen within in vitro cell culture that contains neuronal cells. We identified *TCF4* as an MR based on the deconvolved transcriptional networks from the two independent datasets. For the predicted gene network of *TCF4*, a known leading SCZ GWAS locus, we have empirically validated the predicted *TCF4* gene targets by analyzing both TF binding footprints in neuronal cells and by transcriptomic profiling of *TCF4*-associated gene expression changes in both hiPSC-derived NPCs and Glut_Ns upon *TCF4* knockdown.

Compared to the commonly used weighted gene coexpression network analysis (*43*), the method used in our study aims to find MRs rather than to identify coexpression patterns, bearing a number of unique features, including the ability to elucidate potentially causal interactions, to illustrate the hierarchical structure of the identified transcriptional networks, and to identify potential feed-forward loops between MRs that may cause synergistic effects in the network. The confidence of the identified MRs and their subnetworks was further strengthened by their presence in both transcriptomic datasets and experimentally validated in hiPSC-derived neuronal cells using ATAC-seq data, ChIP-seq data, and transcriptomic profiling.

TCF4 is one of the leading SCZ risk genes identified by GWAS (*5*). However, the molecular mechanism underlying the genetic association remains elusive. Here, we have shown computationally that *TCF4* is one of the top MRs of SCZ, and this was empirically validated in the hiPSC model of mental disorders. Although the downstream gene targets of *TCF4* have been previously profiled in a neuroblastoma cell line (*38*), the hiPSC cellular model used here enabled us to examine the neurodevelopmental relevance of the *TCF4* gene network. By examining the *TCF4*-pertubred gene networks in hiPSC-derived NPCs (early neuronal development/neurogenesis stage) and in differentiated Glut_Ns, we found that *TCF4*-altered genes in NPC are more enriched for gene pathways that are related to neuronal activities. Additional experiments on two independent cell lines from unaffected individuals further validated our findings as to how *TCF4* targets a large body of genes in Glut_Ns. Furthermore, we have shown that *TCF4*-altered genes in NPC are more enriched for credible SCZ GWAS risk genes and for SCZ-associated dnSNVs. Thus, multiple lines of evidence from our study suggest that the *TCF4* gene network expression activity in the early stages of neurodevelopment and neurogenesis may be more important for SCZ pathogenesis. In an attempt to further substantiate our findings on elucidating the transcriptional effects of *TCF4*, we combined the data from all three hiPSC lines (one SCZ case and two control lines) and identified differentially expressed genes upon *TCF4* knockdown in Glut_Ns. This allowed us to boost the power of our analysis where we observed 5584 DE genes (FDR < 0.05) with 3017 up-regulated and 2567 down-regulated genes. Notably, we observed 47 DE genes overlapping

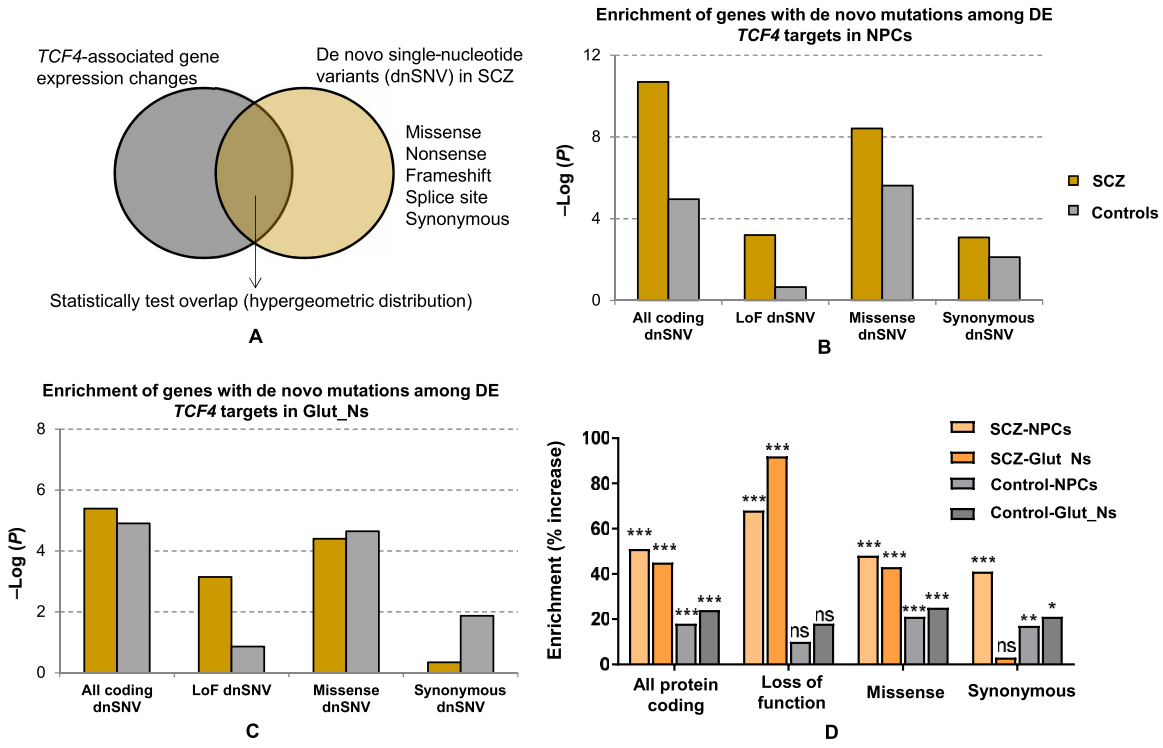

Fig. 5. Enrichment of SCZ de novo SNVs in *TCF4*-associated gene expression changes from NPCs and Glut_Ns. (**A**) Analysis framework. (**B**) Enrichment *P* values ($-\log_{10}$-transformed) of all DE genes at day 3 (NPCs). The *P* value on the *y* axis obtained from a hypergeometric test was used to test the statistical significance of each overlap (number of shared genes between lists) and using all protein coding genes as a background set. (**C**) Enrichment *P* values ($-\log_{10}$-transformed) of all DE genes at day 14 (Glut_Ns). (**D**) Enrichment of SCZ dnSNVs in *TCF4*-associated gene expression changes.

with the 101 predicted *TCF4* regulons (Fisher's exact test, $P = 1.3 \times 10^{-5}$, fold enrichment = 1.83). This analysis further demonstrates the significance of our findings on the regulatory role of *TCF4* on its regulons.

Although our empirical validation of MR has focused on *TCF4*, other MRs that we identified here suggest that additional gene networks are relevant to SCZ. The identified MRs contain many of the TFs previously reported in the literature that may potentially be involved in the pathogenesis of SCZ, such as *JUND* (*44*), *NRG1* and *GSK3B* (*45*), and *NFE2* and *MZF1* (*46*). Among those newly identified MRs, methylation patterns of *TMEM9* (*47*) have been reported to be associated with Parkinson's disease, yet *CRH* is a protein-coding gene associated with Alzheimer's disease, depression, and SCZ (*48*), and *ARPP19* has been identified to be associated with nerve terminal function (*49*). A particularly noteworthy previously unidentified MR candidate is *HDAC9*, a histone deacetylase inhibitor that is important for chromatin remodeling. It has been shown to regulate mouse cortical neuronal dendritic complexities (*50*), as well as hippocampus-dependent memory and synaptic plasticity under different neuropsychiatric conditions (*51*). In our deconvoluted gene network of *HDAC9*, its regulons were moderately enriched for genes related to focal adhesion ($P = 0.0115$), a biological function that has been shown to be altered in SCZ (*52*). Of note, *TCF4*-altered genes in both NPCs and Glut_Ns also showed enrichment of focal adhesion pathway. The MRs we identified and their subnetworks may provide a rich resource of SCZ-relevant gene networks that can be perturbed to increase our understanding of SCZ disease pathophysiology. Moreover, we showed that, despite cross-individual genetic variations that may drive different expression patterns, *TCF4* appears to be a strong SCZ MR in three different cell lines.

The selection of *TCF4* as top MR candidate for validation was based on the consistency between CMC data and CNON datasets. However, we noted that *TCF4* regulons predicted by CMC data and CNON data were different. This difference may have largely risen from the distinct cellular composition of the prefrontal cortex in CMC data and the primary olfactory neuroepithelium in CNON data. CMC data were generated from postmortem human prefrontal cortex, in which tens of different cell types might have been involved. On the other hand, CNON data, which were used as complementary data aimed at further probing the in silico effects of MRs on their targets, are less heterogeneous but with different cellular composition from the prefrontal cortex in CMC. We acknowledge that CNON could be a mixture of dividing neuronal cells and even epithelial cells. Nonetheless, our transcriptomic comparative analysis (fig. S11) indicated that the transcriptomic profiles of CNON, CMC samples, and our iPSC-derived NPCs were all moderately to highly correlated with each other and with GTEx brain frontal cortex and hippocampus (*53*), and CNON resembles CMC data more than NPCs. Alternatively, the different *TCF4* regulons from the two different datasets may be due to the stochastic nature of network deconvolution approaches in general when sample size is moderate. The original developers recommended >100 samples in any analysis to reduce variation, and we had more than this recommended number in both datasets. Last, it is noteworthy that previous work also observed similar differences of predicted regulons when analyzing different cancer datasets from tissues and from cell lines (*54*). Thus, future network deconvolution in single cells within each cell type may minimize these variations.

Despite the possible effects of different cellular compositions on *TCF4* regulons as discussed above, we have focused on our empirical validation in hiPSC-derived Glut_Ns. This is largely based on recent findings that show Glut_Ns to be one of the most relevant cell types to SCZ (*55*, *56*). However, other cell types such as interneurons and microglia have also been implicated in pathogenesis of SCZ. Given cell type–specific expression profiles, the transcriptomic effects of *TCF4* knockdown in different cell types are likely different to some extent, which would be an interesting future research subject. Furthermore, although our relatively homogeneous two-dimensional neuronal cultures (80 to 95%) have their advantages, they may not reflect the in vivo brain circuit where different cell types interact with each other. Therefore, it would be interesting to interrogate possible cell type–specific effects of *TCF4* knockdown in a mixed cell culture or using brain organoids, followed by single-cell RNA-seq (scRNA-seq) analysis.

In summary, by using both a computational approach and a hiPSC neuronal developmental model, we have identified *TCF4* as an MR that likely contributes to SCZ susceptibility at the early stages of neurodevelopment. Although powerful, we acknowledge the limitation of our approach in identifying top MRs, because selecting a top MR is not purely "data-driven"; for instance, *TCF4* was identified as an MR among many other MRs and our selection of *TCF4* as the top MR for validation was not only data-driven but also based on prior knowledge about the *TCF4* association with SCZ. Nonetheless, our study suggests that MRs in SCZ can be identified by transcriptional network deconvolution and that MRs such as *TCF4* as well as other MRs may constitute convergent gene networks that confer disease risk in a spatial and temporal manner. Therefore, *TCF4* and other identified MRs may be a building block of SCZ polygenic/omnigenic architecture, which collectively drives SCZ onset and progression. Transcriptional network deconvolution in larger and more SCZ-relevant cell types/stages, combined with empirical network perturbation, will further deepen our understanding of the genetic contribution to SCZ disease biology.

MATERIALS AND METHODS

Preprocessing of the RNA-seq datasets

RNA-seq data of DLPFC release 1.2 was downloaded from the "normalized SVA-corrected" directory of the CMC Knowledge Portal (www.synapse.org/#!Synapse:syn5609499) using Synapse Python client (http://docs.synapse.org/python/). The data being used in this study contained 307 SCZ cases and 245 controls where the paired-end RNA-seq data had been generated on an Illumina HiSeq 2500. Brain tissues in this dataset were obtained from several brain bank collections, including the Mount Sinai National Institutes of Health (NIH) Brain and Tissue Repository, the University of Pittsburgh NeuroBioBank and Brain and Tissue repositories, the University of Pennsylvania Alzheimer's Disease Core Center, and the National Institute of Mental Health (NIMH) Human Brain Collection Core. Detailed procedure of tissue collection, sample preparation, data generation, and processing can be found in the CMC Knowledge Portal Wiki page (www.synapse.org/#!Synapse:syn2759792/wiki/69613).

The normalized RNA-seq read counts from 143 SCZ cases and 112 controls extracted from cultured primary neuronal cells derived from neuroepithelium were generated as previously described (*24*). Similar to the CMC data, the raw CNON data consisted of 100–base pair (bp) paired-end reads. The data had been preprocessed to exclude ribosomal RNA (rRNA) and mitochondrial genes. Later, normalized counts were calculated by the DESeq2 software.

Transcriptional network reconstruction

ARACNe, an information-theoretic algorithm for inferring transcriptional interactions, was used to identify candidate transcriptional regulators of the transcripts annotated to genes in both CMC and CNON data. First, mutual interaction between a candidate TF(x) and its potential target (y) was computed by pairwise MI, MI(x, y), using a Gaussian kernel estimator. A threshold was applied on MI based on the null hypothesis of statistical independence ($P < 0.05$, Bonferroni-corrected for the number of tested pairs). Second, the constructed network was trimmed by removing indirect interactions using the data processing inequality (DPI), a property of the MI. Therefore, for each (x, y) pair, a path through another TF(z) was considered, and every path pertaining to the following constraint was removed: MI(x, y) < min(MI(x, z),MI(z, y)). A P value threshold of 1×10^{-8} using DPI = 0.1 [as recommended (*17*)] was used when running ARACNe.

Virtual protein activity analysis

The regulon enrichment on gene expression signatures was tested by the VIPER algorithm. First, the gene expression signature was obtained by comparing two groups of samples representing distinctive phenotypes or treatments. In the next step, regulon enrichment on the gene expression signature can be computed using analytic rank-based enrichment analysis (*56*). At the end, significance values (P value and normalized enrichment score) were computed by comparing each regulon enrichment score to a null model generated by randomly and uniformly permuting the samples 1000 times. The output of VIPER is a list of highly active MRs as well as their activity scores and their enrichment P values. Further information about VIPER can be accessed in (*18*).

TF binding site enrichment analysis (TFBEA) and footprint analysis

Human reference genome (hg19) was used to extract the DNA sequence around TSSs for TFBEA. We obtained the gene coordinates from the Ensembl BioMart tool (*57*) and scanned 2000 bp upstream and 1000 bp downstream of the TSS. The motifs of the TFs were obtained from JASPAR, and the extracted sequences of each target were then fed into JASPAR and analyzed versus their corresponding TFs. Then, using the position weight matrix (PWM) of the TF, JASPAR used the modified Needleman-Wunsch algorithm to align the motif sequence with the target sequence in that the input sequence is scanned to check whether or not the motif is enriched. The output is the enrichment score of the input TF in the designated target genes.

We used PIQ to assess the local TF occupancy footprint from ATAC-seq data (*33*, *34*). We extracted the corresponding BED files for TF footprint analysis using the PIQ R package. All the footprints were annotated using the TF matrix with the names of different TFs annotated in the BED files. For each sample, footprints were generated using three different PIQ purity scores (0.7, 0.8, or 0.9; equivalent to an FDR of 0.3, 0.2, or 0.1, respectively). The corresponding files were then extracted using the MR list, and the peak names/coordinates containing *TCF4* gene were collected as a subset of the original BED file. These subsets of genomic coordination were then annotated using the findPeaks.pl included in the HOMER package with the hg19 reference genome.

hiPSC generation and cell culture

The hiPSC lines used for deriving neurons were generated using the Sendai virus method at the Rutgers University Cell and DNA Repository–NIMH Stem Cell Center from cryopreserved lymphocytes of the Molecular Genetics of Schizophrenia (MGS) collection (*58*). The patient line was derived from a 29-year-old male patient with SCZ (cell line ID: 07C65853), and the other two lines were derived from unaffected controls (cell line IDs: 05C39664 and 05C43758, designated as CD07 and CD09, respectively). All three lines do not have large CNVs associated with SCZ. The NorthShore University HealthSystem Institutional Review Board approved the study. hiPSCs were cultured using the feeder-free method in Geltrex (Thermo Fisher Scientific)–coated plates in mTeSR1 medium (StemCell). The media were changed daily, and cells were passaged as clumps every 5 days using ReLeSR (StemCell) in the presence of 5 μm ROCK inhibitor (R&D Systems). hiPSCs were characterized by positive IF staining of pluripotent stem cell markers OCT4, TRA-1-60, NANOG, and SSEA4 (Fig. 3, A and B).

hiPSC-derived NPCs

NPCs were differentiated from hiPSCs using the PSC Neural Induction Medium (Thermo Fisher Scientific) with modifications. In brief, hiPSCs were replated as clumps in Geltrex (Thermo Fisher Scientific)–coated six-well plates (Thermo Fisher Scientific) in mTeSR1 (StemCell) on day 0. From day 2, the medium was switched to PSC Neural Induction Medium and changed daily. On day 11, NPCs were harvested using Neural Rosette Selection Reagent (StemCell). NPCs were maintained in Neural Expansion Medium (Thermo Fisher Scientific) and were passaged every 4 to 6 days in the presence of 5 μM ROCK inhibitor (R&D Systems) until the fourth passage (P4). P4 NPCs were characterized by positive IF staining for Nestin and Pax6 (Fig. 3, C and D).

hiPSC differentiation and shRNA-mediated *TCF4* (ITF-2) knockdown

hiPSC-derived NPCs were plated at a density of 3×10^5 per well in a 12-well plate precoated with Geltrex for 2 hours. The differentiation media used for induced differentiation of NPCs contain neural basal medium, B-27 supplement, L-glutamine, brain-derived neurotrophic factor (10 ng/ml), glial cell line–derived neurotrophic factor (10 ng/ml), and NT-3 (10 ng/ml) (*34*). NPCs were allowed to grow and differentiate for 3 or 14 days after induction. At the corresponding day, 20 μl (1.0×10^5 infectious units of virus) of either *TCF4* or control shRNA lentiviral particles (Santa Cruz sc-61657-V for *TCF4*, which contains a pool of three target-specific, propriety constructs targeting human *TCF4* gene locus at 18q21.2, and sc-108080 for control, which contains a set of scrambled nonspecific shRNA sequence) was added into each well (three replicates for each group), and the transduced cells were further cultured for 48 hours before collection. Total RNAs were extracted from homogenized cell lysate using the RNeasy Plus mini kit (Qiagen), and the quality of extracted total RNA was examined using NanoDrop spectrum analysis and agarose gel electrophoresis. The day 14 excitatory neurons were characterized by positive IF staining for VGLUT1 (Fig. 3E).

qPCR validation of *TCF4* shRNA knockdown efficiency

We used qPCR to verify the knockdown efficiency of shRNA before applying RNA-seq. Briefly, cDNAs were reverse-transcribed from total RNAs using the High-Capacity Reverse Transcription Kit with RNase Inhibitor (Thermo Fisher Scientific). Fifty nanograms of total RNA was used for each reverse-transcription (RT) reaction (20 μl), according to the manufacturer's instructions. cDNA was diluted 15-fold before applying it to qPCR. Subsequent qPCR analysis was performed on a Roche LightCycler 480 II real-time PCR machine using TaqMan probes against *TCF4* and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) (as internal reference). Briefly, 5 μl of diluted RT product was applied to each 10 μl of qPCR reaction using the TaqMan Universal Amplification kit (Thermo Fisher Scientific/Applied Biosystems) with customized TaqMan probes (Hs.PT.58.21450367 for *TCF4*, Integrated DNA Technologies, and Hs03929097_g1 for GAPDH, Applied Biosystems). Cycle parameters were as follows: 95°C, 10 min; 95°C, 15 s; 60°C, 1 min, 45 cycles. Data analysis was performed using the build-in analysis software of Roche 480 II with relative quantification (ΔΔCt method applied).

RNA-seq analysis on hiPSC-derived NPCs and Glut_Ns

RNA-seq libraries were prepared from total RNAs from the collected neuronal cells of 12 samples (three biological replicates, i.e., independent cell cultures, of four distinct groups). These sample groups are for *TCF4* expression knockdown and control at hiPSC-derived NPC stage (day 3 after differentiation) and Glut_N stage (day 14 after differentiation). Total RNAs were extracted as described above. Three main methods for quality control (QC) of RNA samples were conducted, including (i) preliminary quantification, (ii) testing RNA degradation and potential contamination (Agarose Gel Electrophoresis), and (iii) checking RNA integrity and quantification (Agilent 2100) (*10*). After the QC procedures, a library was constructed and library QC was conducted consisting of three steps, including (i) testing the library concentration (Qubit 2.0), (ii) testing the insert size (Agilent 2100), and (iii) precise quantification of library effective concentration (qPCR) (*10*). The quantified libraries were then sequenced using Illumina HiSeq sequencers with 150-bp paired-end reads, after pooling according to its effective concentration and expected data volume.

The paired-end sequenced reads for each sample were obtained as FASTQ files. Initial quality check of the raw data was performed using FastQC. All FASTQ files passed the QC stage, including removing low-quality bases, adapters, short sequences, and checking for rRNA contamination. The average number of preprocessed reads per sample was ~22,740,000. High-quality reads were mapped to the reference genome GRCh38 using STAR (*59*). Sorting and counting were also performed using STAR based on the procedure recommended by the developers. To detect DE genes, the output read counts were fed to edgeR software package (*60*). Normalization, batch correction, and differential expression analysis were conducted using the recommended settings.

Pathway enrichment and GO analysis

Pathway enrichment and GO analysis were conducted using WebGestalt (*31*) v. 2017. Kyoto Encyclopedia of Genes and Genomes (KEGG) was used as the functional database where the list of expressed genes, with fragments per kilobase of transcript per million (FPKM) >1, was used as the background. The maximum and minimum number of genes for each category were set to 2000 and 5, respectively, based on the default setting. Bonferroni-Hochberg multiple test adjustment was applied to the enrichment output. FDR significance threshold was set to 0.05.

MetaCore pathway analysis

Ensembl transcript IDs of differentially expressed genes were imported in MetaCore (v6.32) and processed using the "enrichment analysis" workflow. For each time point (day 3 or day 14), all the up-regulated (FDR < 0.05), all the down-regulated (FDR < 0.05), and the top 2000 differentially expressed genes (FDR < 0.05) were analyzed. Each list was compared to a defined background set of genes consisting of all the expressed transcripts (FPKM > 1) at the relevant time point (day 3 or day 14). The "pathway maps" or "process networks" that were enriched in each analysis were considered significant when FDR < 0.05.

Enrichment analysis on de novo mutations in SCZ

Data on de novo variants in SCZ and controls were obtained from http://denovo-db.gs.washington.edu/denovo-db (accessed 11 April 2017) (*61*). Variants identified in patients with SCZ and control subjects were stratified into four categories (all protein coding, missense, loss of function, and synonymous) based on annotations from the denovo-db. To generate a list of variants affecting protein coding ("all protein coding"), de novo variants outside exonic coding regions were excluded [including those in 5′ untranslated regions (5′UTRs), 3′UTRs, intronic and intergenic regions, and noncoding exons], as well as synonymous variants. The "loss-of-function" list included frameshift, stop gained, stop lost, start lost, splice acceptor, and donor variants. We then compiled a total of six lists containing different sets of differentially expressed genes from the *TCF4* knockdown (KD) experiments at day 3 and day 14, as follows: day3_all_sig (all differentially expressed genes FDR < 0.05; 4898 genes), day3_up_sig (all up-regulated genes FDR < 0.05; 2332 genes), day3_down_sig (all down-regulated genes FDR < 0.05; 2566 genes), day14_all_sig (all differentially expressed genes FDR < 0.05; 3156 genes), day14_up_sig (all up-regulated genes FDR < 0.05; 1864 genes), day14_down_sig (all down-regulated genes FDR < 0.05; 1292 genes). Each of the de novo lists for SCZ and controls was intersected with the relevant list of differentially expressed genes from the *TCF4* KD experiments to determine the shared number of genes in each comparison. A hypergeometric test was used to test the statistical significance of each overlap (number of shared genes between lists) and using all protein coding genes as a background set (20,338, human genome build GRCh38.p10). The data were displayed as the $-\log_{10}$ (P value) for each comparison.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/5/9/eaau4139/DC1

Fig. S1. Representation of the top active MRs identified in the CMC data.
Fig. S2. GOSlim summary of the up-regulated genes on day 3.
Fig. S3. GOSlim summary of the down-regulated genes on day 3.
Fig. S4. GOSlim summary of the up-regulated genes on day 14.
Fig. S5. GOSlim summary of the down-regulated genes on day 14.
Fig. S6. Biological pathways altered upon *TCF4* knockdown in hiPSC-derived neuronal cells.
Fig. S7. MetaCore analysis of top 2000 dysregulated genes upon *TCF4* knockdown in hiPSC-derived neurons.
Fig. S8. MetaCore analysis of all gene sets dysregulated upon TCF4 knockdown in hiPSC-derived neurons.
Fig. S9. The identified TCF4 interactome and their contributions to expression of TCF4.
Fig. S10. Batch correction of CD07 cell line data.
Fig. S11. Correlations between the expression profile (summarized as median gene expression of each gene across samples) of the CMC, CNON, and NPC data with the 53 tissues from the GTEx consortium.
Table S1. Targets of the identified hub genes in the CMC network.
Table S2. Network structure created from the CMC data.
Table S3. Network structure created from the CNON data.
Table S4. Targets of the identified hub genes in the CNON network.
Table S5. Targets of five MRs in the CMC and CNON data.
Table S6. List of the altered pathway upon TCF4 knockdown in NPCs and Glut_Ns in the SCZ cell line.
Table S7. List of differentially expressed genes in NPCs and Glut_Ns in the SCZ cell line.
Table S8. MetaCore enrichment analysis on NPCs in the SCZ cell line.
Table S9. MetaCore enrichment analysis on Glut_Ns in the SCZ cell line.
Table S10. List of TCF4 targets in other studies.
Table S11. VIPER enrichment scores of the identified MRs in the CMC and CNON data.

REFERENCES AND NOTES

1. J. McGrath, S. Saha, D. Chant, J. Welham, Schizophrenia: A concise overview of incidence, prevalence, and mortality. *Epidemiol. Rev.* **30**, 67–76 (2008).
2. R. C. Kessler, H. Birnbaum, O. Demler, I. R. Falloon, E. Gagnon, M. Guyer, M. J. Howes, K. S. Kendler, L. Shi, E. Walters, E. Q. Wu, The prevalence and correlates of nonaffective psychosis in the National Comorbidity Survey Replication (NCS-R). *Biol. Psychiatry* **58**, 668–676 (2005).
3. R. Hilker, D. Helenius, B. Fagerlund, A. Skytthe, K. Christensen, T. M. Werge, M. Nordentoft, B. Glenthoj, Heritability of schizophrenia and schizophrenia spectrum based on the nationwide Danish Twin register. *Biol. Psychiatry* **83**, 492–498 (2018).
4. E. Rees, M. C. O'Donovan, M. J. Owen, Genetics of schizophrenia. *Curr. Opin. Behav. Sci.* **2**, 8–14 (2015).
5. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
6. C. R. Marshall, D. P. Howrigan, D. Merico, B. Thiruvahindrapuram, W. Wu, D. S. Greer, D. Antaki, A. Shetty, P. A. Holmans, D. Pinto, M. Gujral, W. M. Brandler, D. Malhotra, Z. Wang, K. V. F. Fajarado, M. S. Maile, S. Ripke, I. Agartz, M. Albus, M. Alexander, F. Amin, J. Atkins, S. A. Bacanu, R. A. Belliveau Jr., S. E. Bergen, M. Bertalan, E. Bevilacqua, T. B. Bigdeli, D. W. Black, R. Bruggeman, N. G. Buccola, R. L. Buckner, B. Bulik-Sullivan, W. Byerley, W. Cahn, G. Cai, M. J. Cairns, D. Campion, R. M. Cantor, V. J. Carr, N. Carrera, S. V. Catts, K. D. Chambert, W. Cheng, C. R. Cloninger, D. Cohen, P. Cormican, N. Craddock, B. Crespo-Facorro, J. J. Crowley, D. Curtis, M. Davidson, K. L. Davis, F. Degenhardt, J. Del Favero, L. E. DeLisi, D. Dikeos, T. Dinan, S. Djurovic, G. Donohoe, E. Drapeau, J. Duan, F. Dudbridge, P. Eichhammer, J. Eriksson, V. Escott-Price, L. Essioux, A. H. Fanous, K.-H. Farh, M. S. Farrell, J. Frank, L. Franke, R. Freedman, N. B. Freimer, J. I. Friedman, A. J. Forstner, M. Fromer, G. Genovese, L. Georgieva, E. S. Gershon, I. Giegling, P. Giusti-Rodríguez, S. Godard, J. I. Goldstein, J. Gratten, L. de Haan, M. L. Hamshere, M. Hansen, T. Hansen, V. Haroutunian, A. M. Hartmann, F. A. Henskens, S. Herms, J. N. Hirschhorn, P. Hoffmann, A. Hofman, H. Huang, M. Ikeda, I. Joa, A. K. Kähler, R. S. Kahn, L. Kalaydjieva, J. Karjalainen, D. Kavanagh, M. C. Keller, B. J. Kelly, J. L. Kennedy, Y. Kim, J. A. Knowles, B. Konte, C. Laurent, P. Lee, S. H. Lee, S. E. Legge, B. Lerer, D. L. Levy, K.-Y. Liang, J. Lieberman, J. Lönnqvist, C. M. Loughland, P. K. E. Magnusson, B. S. Maher, W. Maier, J. Mallet, M. Mattheisen, M. Mattingsdal, R. W. McCarley, C. McDonald, A. M. McIntosh, S. Meier, C. J. Meijer, I. Melle, R. I. Mesholam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, Y. Mokrab, D. W. Morris, B. Müller-Myhsok, K. C. Murphy, R. M. Murray, I. Myin-Germeys, I. Nenadic, D. A. Nertney, G. Nestadt, K. K. Nicodemus, L. Nisenbaum, A. Nordin, E. O'Callaghan, C. O'Dushlaine, S.-Y. Oh, A. Olincy, L. Olsen, F. A. O'Neill, J. Van Os, C. Pantelis, G. N. Papadimitriou, E. Parkhomenko, M. T. Pato, T. Paunio; Psychosis Endophenotypes International Consortium, D. O. Perkins, T. H. Pers, O. Pietilainen, J. Pimm, A. J. Pocklington, J. Powell, A. Price, A. E. Pulver, S. M. Purcell, D. Quested, H. B. Rasmussen, A. Reichenberg, M. A. Reimers, A. L. Richards, J. L. Roffman, P. Roussos, D. M. Ruderfer, V. Salomaa, A. R. Sanders, A. Savitz, U. Schall, T. G. Schulze, S. G. Schwab, E. M. Scolnick, R. J. Scott, L. J. Seidman, J. Shi, J. M. Silverman, J. W. Smoller, E. Söderman, C. C. A. Spencer, E. A. Stahl, E. Strengman, J. Strohmaier, T. S. Stroup, J. Suvisaari, D. M. Svrakic, J. P. Szatkiewicz, S. Thirumalai, P. A. Tooney, J. Veijola, P. M. Visscher, J. Waddington, D. Walsh, B. T. Webb, M. Weiser, D. B. Wildenauer, N. M. Williams, S. Williams, S. H. Witt, A. R. Wolen, B. K. Wormley, N. R. Wray, J. Q. Wu, C. C. Zai, R. Adolfsson, O. A. Andreassen, D. H. R. Blackwood, E. Bramon, J. D. Buxbaum, S. Cichon, D. A. Collier, A. Corvin, M. J. Daly, A. Darvasi, E. Domenici, T. Esko, P. V. Gejman, M. Gill, H. Gurling, C. M. Hultman, N. Iwata, A. V. Jablensky, E. G. Jonsson, K. S. Kendler, G. Kirov, J. Knight, D. F. Levinson, Q. S. Li, S. A. McCarroll, A. McQuillin, J. L. Moran, B. J. Mowry, M. M. Nothen, R. A. Ophoff, M. J. Owen, A. Palotie, C. N. Pato, T. L. Petryshen, D. Posthuma, M. Rietschel, B. P. Riley, D. Rujescu, P. Sklar, D. St Clair, J. T. R. Walters, T. Werge, P. F. Sullivan, M. C. O'Donovan, S. W. Scherer, B. M. Neale, J. Sebat; CNV and Schizophrenia Working Groups of the Psychiatric Genomics, Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet.* **49**, 27–35 (2017).
7. G. Genovese, M. Fromer, E. A. Stahl, D. M. Ruderfer, K. Chambert, M. Landén, J. L. Moran, S. M. Purcell, P. Sklar, P. F. Sullivan, C. M. Hultman, S. A. McCarroll, Increased burden

of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
8. T. Singh, M. I. Kurki, D. Curtis, S. M. Purcell, L. Crooks, J. McRae, J. Suvisaari, H. Chheda, D. Blackwood, G. Breen, O. Pietiläinen, S. S. Gerety, M. Ayub, M. Blyth, T. Cole, D. Collier, E. L. Coomber, N. Craddock, M. J. Daly, J. Danesh, M. DiForti, A. Foster, N. B. Freimer, D. Geschwind, M. Johnstone, S. Joss, G. Kirov, J. Körkkö, O. Kuismin, P. Holmans, C. M. Hultman, C. Iyegbe, J. Lönnqvist, M. Männikkö, S. A. McCarroll, P. McGuffin, A. M. McIntosh, A. McQuillin, J. S. Moilanen, C. Moore, R. M. Murray, R. Newbury-Ecob, W. Ouwehand, T. Paunio, E. Prigmore, E. Rees, D. Roberts, J. Sambrook, P. Sklar, D. St Clair, J. Veijola, J. T. Walters, H. Williams, S. S. Schizophrenia; INTERVAL Study; DDD Study; UK10 K Consortium, P. F. Sullivan, M. E. Hurles, M. C. O'Donovan, A. Palotie, M. J. Owen, J. C. Barrett, Rare loss-of-function variants in *SETD1A* are associated with schizophrenia and developmental disorders. *Nat Neurosci.* **19**, 571–577 (2016).
9. E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
10. E. Mondragón, L. J. Maher III, Anti-transcription factor RNA aptamers as potential therapeutics. *Nucleic Acid Ther* **26**, 29–43 (2016).
11. B. F. Corbett, J. C. You, X. Zhang, M. S. Pyfer, U. Tosi, D. M. Iascone, I. Petrof, A. Hazra, C. H. Fu, G. S. Stephens, A. A. Ashok, S. Aschmies, L. Zhao, E. J. Nestler, J. Chin, ΔFosB regulates gene expression and cognitive dysfunction in a mouse model of Alzheimer's disease. *Cell Rep.* **20**, 344–355 (2017).
12. M. Lizio, J. Harshbarger, H. Shimoji, J. Severin, T. Kasukawa, S. Sahin, I. Abugessaisa, S. Fukuda, F. Hori, S. Ishikawa-Kato, C. J. Mungall, E. Arner, J. K. Baillie, N. Bertin, H. Bono, M. de Hoon, A. D. Diehl, E. Dimont, T. C. Freeman, K. Fujieda, W. Hide, R. Kaliyaperumal, T. Katayama, T. Lassmann, T. F. Meehan, K. Nishikata, H. Ono, M. Rehli, A. Sandelin, E. A. Schultes, P. A. 't Hoen, Z. Tatum, M. Thompson, T. Toyoda, D. W. Wright, C. O. Daub, M. Itoh, P. Carninci, Y. Hayashizaki, A. R. Forrest, H. Kawaji; FANTOM Consortium, Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol.* **16**, 22 (2015).
13. J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann, N. M. Luscombe, A census of human transcription factors: Function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).
14. H. Han, H. Shim, D. Shin, J. E. Shim, Y. Ko, J. Shin, H. Kim, A. Cho, E. Kim, T. Lee, H. Kim, K. Kim, S. Yang, D. Bae, A. Yun, S. Kim, C. Y. Kim, H. J. Cho, B. Kang, S. Shin, I. Lee, TRRUST: A reference database of human transcriptional regulatory interactions. *Sci. Rep.* **5**, 11432 (2015).
15. I. K. Jordan, L. Mariño-Ramírez, Y. I. Wolf, E. V. Koonin, Conservation and coevolution in the scale-free human gene coexpression network. *Mol. Biol. Evol.* **21**, 2058–2070 (2004).
16. K. Basso, A. A. Margolin, G. Stolovitzky, U. Klein, R. Dalla-Favera, A. Califano, Reverse engineering of regulatory networks in human B cells. *Nat. Genet.* **37**, 382–390 (2005).
17. A. A. Margolin, K. Wang, W. K. Lim, M. Kustagi, I. Nemenman, A. Califano, Reverse engineering cellular networks. *Nat. Protoc.* **1**, 662–671 (2006).
18. M. J. Alvarez, Y. Shen, F. M. Giorgi, A. Lachmann, B. B. Ding, B. H. Ye, A. Califano, Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
19. L. Brichta, W. Shin, V. Jackson-Lewis, J. Blesa, E. L. Yap, Z. Walker, J. Zhang, J. P. Roussarie, M. J. Alvarez, A. Califano, S. Przedborski, P. Greengard, Identification of neurodegenerative factors using translatome–regulatory network analysis. *Nat Neurosci.* **18**, 1325–1333 (2015).
20. V. Haroutunian, P. Katsel, S. Dracheva, D. G. Stewart, K. L. Davis, Variations in oligodendrocyte-related gene expression across multiple cortical regions: Implications for the pathophysiology of schizophrenia. *Int. J. Neuropsychopharmacol.* **10**, 565–573 (2007).
21. D. Arion, S. Horváth, D. A. Lewis, K. Mirnics, Infragranular gene expression disturbances in the prefrontal cortex in schizophrenia: Signature of altered neural development? *Neurobiol. Dis.* **37**, 738–746 (2010).
22. A. L. Guillozet-Bongaarts, T. M. Hyde, R. A. Dalley, M. J. Hawrylycz, A. Henry, P. R. Hof, J. Hohmann, A. R. Jones, C. L. Kuan, J. Royall, E. Shen, B. Swanson, H. Zeng, J. E. Kleinman, Altered gene expression in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol. Psychiatry* **19**, 478–485 (2014).
23. M. Fromer, P. Roussos, S. K. Sieberts, J. S. Johnson, D. H. Kavanagh, T. M. Perumal, D. M. Ruderfer, E. C. Oh, A. Topol, H. R. Shah, L. L. Klei, R. Kramer, D. Pinto, Z. H. Gümüs, A. E. Cicek, K. K. Dang, A. Browne, C. Lu, L. Xie, B. Readhead, E. A. Stahl, J. Xiao, M. Parvizi, T. Hamamsy, J. F. Fullard, Y. C. Wang, M. C. Mahajan, J. M. Derry, J. T. Dudley, S. E. Hemby, B. A. Logsdon, K. Talbot, T. Raj, D. A. Bennett, P. L. De Jager, J. Zhu, B. Zhang, P. F. Sullivan, A. Chess, S. M. Purcell, L. A. Shinobu, L. M. Mangravite, H. Toyoshiba, R. E. Gur, C. G. Hahn, D. A. Lewis, V. Haroutunian, M. A. Peters, B. K. Lipska, J. D. Buxbaum, E. E. Schadt, K. Hirai, K. Roeder, K. J. Brennand, N. Katsanis, E. Domenici, B. Devlin, P. Sklar, Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci.* **19**, 1442–1453 (2016).
24. O. V. Evgrafov, C. Armoskus, B. B. Wrobel, V. N. Spitsyna, T. Souaiaia, J. S. Herstein, C. P. Walker, J. D. Nguyen, A. Camarena, J. R. Weitz, J. M. H. Kim, E. Lopez Duarte, K. Wang, G. M. Simpson, J. L. Sobell, H. Medeiros, M. T. Pato, C. N. Pato, J. A. Knowles, Gene expression in patient-derived neural progenitors provide insights into neurodevelopmental aspects of schizophrenia. bioRxiv 209197 [**Preprint**] 26 October 2017. https://doi.org/10.1101/209197.
25. C. Lefebvre, P. Rajbhandari, M. J. Alvarez, P. Bandaru, W. K. Lim, M. Sato, K. Wang, P. Sumazin, M. Kustagi, B. C. Bisikirska, K. Basso, P. Beltrao, N. Krogan, J. Gautier, R. Dalla-Favera, A. Califano, A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. *Mol. Syst. Biol.* **6**, 377 (2010).
26. K. Yang, A. Sawa, Open sesame: Open chromatin regions shed light onto non-coding risk variants. *Cell Stem Cell* **21**, 285–287 (2017).
27. B. B. Lake, S. Chen, B. C. Sos, J. Fan, G. E. Kaeser, Y. C. Yung, T. E. Duong, D. Gao, J. Chun, P. V. Kharchenko, K. Zhang, Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
28. S. Steinberg, S. de Jong; Irish Schizophrenia Genomics Consortium, O. A. Andreassen, T. Werge, A. D. Børglum, O. Mors, P. B. Mortensen, O. Gustafsson, J. Costas, O. P. Pietiläinen, D. Demontis, S. Papiol, J. Huttenlocher, M. Mattheisen, R. Breuer, E. Vassos, I. Giegling, G. Fraser, N. Walker, A. Tuulio-Henriksson, J. Suvisaari, J. Lönnqvist, T. Paunio, I. Agartz, I. Melle, S. Djurovic, E. Strengman; GROUP, G. Jürgens, B. Glenthøj, L. Terenius, D. M. Hougaard, T. Ørntoft, C. Wiuf, M. Didriksen, M. V. Hollegaard, M. Nordentoft, R. van Winkel, G. Kenis, L. Abramova, V. Kaleda, M. Arrojo, J. Sanjuán, C. Arango, S. Sperling, M. Rossner, M. Ribolsi, V. Magni, A. Siracusano, C. Christiansen, L. A. Kiemeney, J. Veldink, L. van den Berg, A. Ingason, P. Muglia, R. Murray, M. M. Nöthen, E. Sigurdsson, H. Petursson, U. Thorsteinsdottir, A. Kong, I. A. Rubino, M. De Hert, J. M. Réthelyi, I. Bitter, E. G. Jönsson, V. Golimbet, A. Carracedo, H. Ehrenreich, N. Craddock, M. J. Owen, M. C. O'Donovan; Wellcome Trust Case Control Costortium 2, M. Ruggeri, S. Tosato, L. Peltonen, R. A. Ophoff, D. A. Collier, D. St Clair, M. Rietschel, S. Cichon, H. Stefansson, D. Rujescu, K. Stefansson, Common variants at *VRK2* and *TCF4* conferring risk of schizophrenia. *Hum Mol Genet* **20**, 4076–4081 (2011).
29. M. Kharbanda, K. Kannike, A. Lampe, J. Berg, T. Timmusk, M. Sepp, Partial deletion of *TCF4* in three generation family with non-syndromic intellectual disability, without features of Pitt-Hopkins syndrome. *Eur. J. Med. Genet.* **59**, 310–314 (2016).
30. M. P. Forrest, M. J. Hill, A. J. Quantock, E. Martin-Rendon, D. J. Blake, The emerging roles of TCF4 in disease and development. *Trends Mol. Med.* **20**, 322–331 (2014).
31. J. Wang, D. Duncan, Z. Shi, B. Zhang, WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
32. A. Mathelier, O. Fornes, D. J. Arenillas, C. Y. Chen, G. Denay, J. Lee, W. Shi, C. Shyr, G. Tan, R. Worsley-Hunt, A. W. Zhang, F. Parcy, B. Lenhard, A. Sandelin, W. W. Wasserman, JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44**, D110–D115 (2016).
33. R. I. Sherwood, T. Hashimoto, C. W. O'Donnell, S. Lewis, A. A. Barkal, J. P. van Hoff, V. Karun, T. Jaakkola, D. K. Gifford, Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
34. M. P. Forrest, H. Zhang, W. Moy, H. McGowan, C. Leites, L. E. Dionisio, Z. Xu, J. Shi, A. R. Sanders, W. J. Greenleaf, C. A. Cowan, Z. P. Pang, P. V. Gejman, P. Penzes, J. Duan, Open chromatin profiling in hipsc-derived neurons prioritizes functional noncoding psychiatric risk variants and highlights neurodevelopmental loci. *Cell Stem Cell* **21**, 305–318 e308 (2017).
35. S. Zhang, W. Moy, H. Zhang, C. Leites, H. McGowan, J. Shi, A. R. Sanders, Z. P. Pang, P. V. Gejman, J. Duan, Open chromatin dynamics reveals stage-specific transcriptional networks in hiPSC-based neurodevelopmental model. *Stem Cell Res.* **29**, 88–98 (2018).
36. M. P. Forrest, M. J. Hill, D. H. Kavanagh, K. E. Tansey, A. J. Waite, D. J. Blake, The psychiatric risk gene transcription factor 4 (*TCF4*) regulates neurodevelopmental pathways associated with schizophrenia, autism, and intellectual disability. *Schizophr. Bull.* **44**, 1100–1110 (2017).
37. H. Xia, F. M. Jahr, N.-K. Kim, L. Xie, A. A. Shabalin, J. Bryois, D. H. Sweet, M. M. Kronfol, P. Palasuberniam, M. McRae, B. P. Riley, P. F. Sullivan, E. J. van den Oord, J. L. McClay, Building a schizophrenia genetic network: Transcription factor 4 regulates genes involved in neuronal development and schizophrenia risk. *Hum. Mol. Genet.* **27**, 3246–3256 (2018).
38. M. P. Forrest, A. J. Waite, E. Martin-Rendon, D. J. Blake, Knockdown of human *TCF4* affects multiple signaling pathways involved in cell survival, epithelial to mesenchymal transition and neuronal differentiation. *PLOS ONE* **8**, e73169 (2013).
39. A. J. Kennedy, E. J. Rahn, B. S. Paulukaitis, K. E. Savell, H. B. Kordasiewicz, J. Wang, J. W. Lewis, J. Posey, S. K. Strange, M. C. Guzman-Karlsson, S. E. Phillips, K. Decker, S. T. Motley, E. E. Swayze, D. J. Ecker, T. P. Michael, J. J. Day, J. D. Sweatt, *Tcf4* regulates synaptic plasticity, DNA methylation, and memory function. *Cell Rep.* **16**, 2666–2685 (2016).
40. M. D. Rannals, G. R. Hamersky, S. C. Page, M. N. Campbell, A. Briley, R. A. Gallo, B. N. Phan, T. M. Hyde, J. E. Kleinman, J. H. Shin, A. E. Jaffe, D. R. Weinberger, B. J. Maher, Psychiatric risk gene transcription factor 4 regulates intrinsic excitability of prefrontal neurons via repression of SCN10a and KCNQ1. *Neuron* **90**, 43–55 (2016).

41. A. Devor, O. A. Andreassen, Y. Wang, T. Mäki-Marttunen, O. B. Smeland, C.-C. Fan, A. J. Schorks, D. Holland, W. K. Thompson, A. Witoelar, C.-H. Chen, R. S. Desikan, L. K. McEvoy, S. Djurovic, P. Greengard, P. Svenningsson, G. T. Einevoll, A. M. Dale, Genetic evidence for role of integration of fast and slow neurotransmission in schizophrenia. *Mol. Psychiatry* **22**, 792–801 (2017).
42. A. F. Pardiñas, P. Holmans, A. J. Pocklington, V. Escott-Price, S. Ripke, N. Carrera, S. E. Legge, S. Bishop, D. Cameron, M. L. Hamshere, J. Han, L. Hubbard, A. Lynham, K. Mantripragada, E. Rees, J. H. MacCabe, S. A. McCarroll, B. T. Baune, G. Breen, E. M. Byrne, U. Dannlowski, T. C. Eley, C. Hayward, N. G. Martin, A. M. McIntosh, R. Plomin, D. J. Porteous, N. R. Wray, A. Caballero, D. H. Geschwind, L. M. Huckins, D. M. Ruderfer, E. Santiago, P. Sklar, E. A. Stahl, H. Won, E. Agerbo, T. D. Als, O. A. Andreassen, M. Bækvad-Hansen, P. B. Mortensen, C. B. Pedersen, A. D. Børglum, J. Bybjerg-Grauholm, S. Djurovic, N. Durmishi, M. G. Pedersen, V. Golimbet, J. Grove, D. M. Hougaard, M. Mattheisen, E. Molden, O. Mors, M. Nordentoft, M. Pejovic-Milovancevic, E. Sigurdsson, T. Silagadze, C. S. Hansen, K. Stefansson, H. Stefansson, S. Steinberg, S. Tosato, T. Werge; GERAD1 Consortium; CRESTAR Consortium, D. A. Collier, D. Rujescu, G. Kirov, M. J. Owen, M. C. O'Donovan, J. T. R. Walters, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
43. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
44. A. S. Boyajyan, S. A. Atshemyan, R. V. Zakharyan, Association of schizophrenia with variants of genes that encode transcription factors. *Mol. Biol.* **49**, 875–880 (2015).
45. E. S. Emamian, AKT/GSK3 signaling pathway and schizophrenia. *Front. Mol. Neurosci.* **5**, 33 (2012).
46. A. Y. Guo, J. Sun, P. Jia, Z. Zhao, A novel microRNA and transcription factor mediated regulatory network in schizophrenia. *BMC Syst. Biol.* **4**, 10 (2010).
47. E. Masliah, W. Dumaop, D. Galasko, P. Desplats, Distinctive patterns of DNA methylation associated with Parkinson disease: Identification of concordant epigenetic changes in brain and peripheral blood leukocytes. *Epigenetics* **8**, 1030–1038 (2013).
48. M. R. Bennett AO, Stress and anxiety in schizophrenia and depression: Glucocorticoids, corticotropin-releasing hormone and synapse regression. *Aust. N. Z. J. Psychiatry* **42**, 995–1002 (2008).
49. P. R. Maycox, F. Kelly, A. Taylor, S. Bates, J. Reid, R. Logendra, M. R. Barnes, C. Larminie, N. Jones, M. Lennon, C. Davies, J. J. Hagan, C. A. Scorer, C. Angelinetta, M. T. Akbar, S. Hirsch, A. M. Mortimer, T. R. Barnes, J. de Belleroche, Analysis of gene expression in two large schizophrenia cohorts identifies multiple changes associated with nerve terminal function. *Mol. Psychiatry* **14**, 1083–1094 (2009).
50. N. Sugo, H. Oshiro, M. Takemura, T. Kobayashi, Y. Kohno, N. Uesaka, W.-J. Song, N. Yamamoto, Nucleocytoplasmic translocation of HDAC9 regulates gene expression and dendritic growth in developing cortical neurons. *Eur. J. Neurosci.* **31**, 1521–1532 (2010).
51. J. P. Lopez-Atalaya, S. Ito, L. M. Valor, E. Benito, A. Barco, Genomic targets, and histone acetylation and gene expression profiling of neural HDAC inhibition. *Nucleic Acids Res.* **41**, 8072–8084 (2013).
52. Y. Fan, G. Abrahamsen, R. Mills, C. C. Calderón, J. Y. Tee, L. Leyton, W. Murrell, J. Cooper-White, J. J. McGrath, A. Mackay-Sim, Focal adhesion dynamics are altered in schizophrenia. *Biol. Psychiatry* **74**, 418–426 (2013).
53. L. J. Carithers, K. Ardlie, M. Barcus, P. A. Branton, A. Britton, S. A. Buia, C. C. Compton, D. S. DeLuca, J. Peter-Demchok, E. T. Gelfand, P. Guan, G. E. Korzeniewski, N. C. Lockhart, C. A. Rabiner, A. K. Rao, K. L. Robinson, N. V. Roche, S. J. Sawyer, A. V. Segrè, C. E. Shive, A. M. Smith, L. H. Sobin, A. H. Undale, K. M. Valentino, J. Vaught, T. R. Young, H. M. Moore; GTEx Consortium, A novel approach to high-quality postmortem tissue procurement: The GTEx project. *Biopreserv. Biobank.* **13**, 311–319 (2015).
54. H. Ding, E. F. Douglass Jr., A. M. Sonabend, A. Mela, S. Bose, C. Gonzalez, P. D. Canoll, P. A. Sims, M. J. Alvarez, A. Califano, Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. *Nat. Commun.* **9**, 1471 (2018).
55. D. Wang, S. Liu, J. Warrell, H. Won, X. Shi, F. C. P. Navarro, D. Clarke, M. Gu, P. Emani, Y. T. Yang, M. Xu, M. J. Gandal, S. Lou, J. Zhang, J. J. Park, C. Yan, S. K. Rhie, K. Manakongtreecheep, H. Zhou, A. Nathan, M. Peters, E. Mattei, D. Fitzgerald, T. Brunetti, J. Moore, Y. Jiang, K. Girdhar, G. E. Hoffman, S. Kalayci, Z. H. Gümüs, G. E. Crawford; PsychENCODE Cosortium, P. Roussos, S. Akbarian, A. E. Jaffe, K. P. White, Z. Weng, N. Sestan, D. H. Geschwind, J. A. Knowles, M. B. Gerstein, Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
56. H. K. Finucane, Y. A. Reshef, V. Anttila, K. Slowikowski, A. Gusev, A. Byrnes, S. Gazal, P. R. Loh, C. Lareau, N. Shoresh, G. Genovese, A. Saunders, E. Macosko, S. Pollack; Brainstorm Consortium, J. R. B. Perry, J. D. Buenrostro, B. E. Bernstein, S. Raychaudhuri, S. McCarroll, B. M. Neale, A. L. Price, Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
57. A. Yates, W. Akanni, M. R. Amode, D. Barrell, K. Billis, D. Carvalho-Silva, C. Cummins, P. Clapham, S. Fitzgerald, L. Gil, C. G. Girón, L. Gordon, T. Hourlier, S. E. Hunt, S. H. Janacek, N. Johnson, T. Juettemann, S. Keenan, I. Lavidas, F. J. Martin, T. Maurel, W. McLaren, D. N. Murphy, R. Nag, M. Nuhn, A. Parker, M. Patricio, M. Pignatelli, M. Rahtz, H. S. Riat, D. Sheppard, K. Taylor, A. Thormann, A. Vullo, S. P. Wilder, A. Zadissa, E. Birney, J. Harrow, M. Muffato, E. Perry, M. Ruffier, G. Spudich, S. J. Trevanion, F. Cunningham, B. L. Aken, D. R. Zerbino, P. Flicek, Ensembl 2016. *Nucleic Acids Res.* **44**, D710–D716 (2016).
58. J. Shi, D. F. Levinson, J. Duan, A. R. Sanders, Y. Zheng, I. Pe'er, F. Dudbridge, P. A. Holmans, A. S. Whittemore, B. J. Mowry, A. Olincy, F. Amin, C. R. Cloninger, J. M. Silverman, N. G. Buccola, W. F. Byerley, D. W. Black, R. R. Crowe, J. R. Oksenberg, D. B. Mirel, K. S. Kendler, R. Freedman, P. V. Gejman, Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).
59. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
60. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
61. T. N. Turner, Q. Yi, N. Krumm, J. Huddleston, K. Hoekzema, F. S. HA, A. L. Doebley, R. A. Bernier, D. A. Nickerson, E. E. Eichler, denovo-db: A compendium of human de novo variants. *Nucleic Acids Res.* **45**, D804–D811 (2017).

Acknowledgments: We thank the CMC for generating the RNA-seq data on DLPFC from patients and control subjects (the data were generated as part of the CMC supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd., and NIH grants R01MH085542, R01MH093725, P50MH080405, R01MH097276, R01MH075916, P50MH096891, P50MH084053S1, R37MH057881, R37MH057881S1, HHSN271201300031C, AG02219, AG05138, and MH06692). We thank the developers of the ARACNe software for comments on the software usage, and we thank the Wang lab members for helpful suggestions and discussions. **Funding:** This study was supported by NIH grant MH108728 (to K.W.), the Alavi-Dabiri Postdoctoral Fellowship Award (to A.D.T.), HG006465 (to K.W.), the CHOP Research Institute (to K.W.), MH086874 (O.V.E. and J.A.K.), MH102685, MH106575, and MH116281 (to J.D.). **Competing interests:** The authors declare that they have no competing interests. **Author contributions:** A.D.T. conceived the study, designed the study, performed the computational experiments, analyzed the data, and wrote the manuscript. C.A., O.V.E., T.S., and J.A.K. generated and preprocessed the CNON data and edited the manuscript. S.Z. and H.Z. generated hiPSCs, conducted *TCF4* knockdown and other experiments, and wrote the relevant Materials and Methods sections. M.P.F. advised on the study and performed additional enrichment analysis on *TCF4*. K.W. conceived the study, wrote and edited the manuscript, and supervised the project. J.D. supervised the experimental sections of the study and wrote and edited the manuscript. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data generated in this study are deposited in Gene Expression Omnibus under accession number GSE128333. Additional data related to this paper may be requested from the authors.

Submitted 7 June 2018
Accepted 8 August 2019
Published 11 September 2019
10.1126/sciadv.aau4139

Citation: A. Doostparast Torshizi, C. Armoskus, H. Zhang, M. P. Forrest, S. Zhang, T. Souaiaia, O. V. Evgrafov, J. A. Knowles, J. Duan, K. Wang, Deconvolution of transcriptional networks identifies TCF4 as a master regulator in schizophrenia. *Sci. Adv.* **5**, eaau4139 (2019).