TITLE

Convergent impact of schizophrenia risk genes

AUTHORS

Kayla G. Townsley*[1,2,3] Aiqun Li*[1,2], PJ Michael Deans[1,2,4], John F. Fullard[1,2,3], Alex Yu[1,2], Sam Cartwright[1,2], Wen Zhang[1,2,3], Minghui Wang[1,2], Georgios Voloudakis[1,2,3,5], Kiran Girdhar[1,2,3,5], Eli Stahl[1,3], Schahram Akbarian[2,3], Bin Zhang[1,2], Panos Roussos[1,2,3,5,6,7], Laura M. Huckins[1,3,8], Kristen J. Brennand[1,2,3,4#]

AFFILIATIONS

[1] Pamela Sklar Division of Psychiatric Genomics, Department of Genetics and Genomics, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029

[2] Nash Family Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029

[3] Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029

[4] Departments of Psychiatry and Genetics, Division of Molecular Psychiatry, Department of Genetics, Wu Tsai Institute, Yale University School of Medicine, New Haven, CT 06511

[5] Center for Disease Neurogenomics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[6] Illness Research Education and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, 10468, NY, USA

[7] Center for Dementia Research, Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

[8] Seaver Autism Centre for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

* These authors contributed equally to this work.

Correspondence:

kristen.brennand@yale.edu,

Number of Tables: 3
Number of Figures: 6

ABSTRACT

Schizophrenia is a highly heritable psychiatric disorder with a complex genetic risk architecture that reflects the additive impact of hundreds of risk variants. While many schizophrenia-associated risk variants are thought to regulate the expression of target genes in a cell-type-specific manner, the mechanisms by which the effect of these myriad variants combine to contribute to risk remain unclear. Here we apply a CRISPR-based approach to evaluate in parallel twelve schizophrenia eGenes (that encompass common variation) in human glutamatergic neurons. Querying the shared neuronal impacts across risk genes uncovers a convergent effect concentrated on pathways of brain development and synaptic signaling. Our analyses reveal shared and divergent downstream effects of these twelve genes, independent of their previously annotated biological roles. General convergence of gene expression increases with increasing polygenicity, while the specificity of convergence increases between functionally similar genes. Convergent networks show brain-region and developmental period-specific enrichments, as well as disorder-specific enrichments for both rare and common variant target genes across schizophrenia, bipolar disorder, autism spectrum disorder, and intellectual disability. These gene targets are drug-able and potentially represent novel points of therapeutic intervention. Convergent signatures are also resolved in the post-mortem brain. Overall, convergence suggests a model to explain how non-additive interactions arise between risk genes and may explain cross-disorder pleiotropy of genetic risk for psychiatric disorders.

KEYWORDS

schizophrenia; psychiatric genomics; human induced pluripotent stem cells; CRISPR-screen; neurons

INTRODUCTION

Genome-wide association studies (GWAS) of single nucleotide polymorphisms (SNPs) continue to identify loci (250 and growing[1-3]) that are significantly associated with risk for schizophrenia (SCZ). These common variant risk loci are enriched for genes expressed in pyramidal excitatory neurons (and a subset of GABAergic interneurons)[4,5], particularly synaptic pathways[6,7]. Risk genes are expressed during fetal cortical development[8,9], highly co-expressed in human brain tissue[10] and cultured neurons[11], and show high connectivity in protein-protein interaction (PPI) networks[11-13]. Transcriptomic studies of SCZ post-mortem brains likewise identify common downregulation of neuronal and synaptic genes as well as upregulation of immune genes in astrocytes[14,15]. An unanswered question is whether individual risk variants sum linearly[16,17] or whether they are amplified[18] or buffered[19] by epistatic interactions. Here we ask how the neuronal impact(s) of risk genes converge in a common genetic background. Identifying the regulatory node(s) downstream of risk genes could reveal novel SCZ biology and point to new therapeutic targets.

CRISPR-mediated genetic perturbation of SCZ risk genes in human induced pluripotent stem cell (hiPSC)-derived neurons can definitively reveal their causal impact on gene expression, cellular function, and synaptic activity[20-23]. Such genetic studies are typically conducted in a one-gene-at-a-time manner and require a substantial time and resource investment, but state-of-the-art CRISPR screens now incorporate a pooled multiplexed design, minimizing experimental variability, increasing the power of the associated analyses, and more closely recapitulating the observed polygenic architecture of disease risk. Recent studies queried an overlapping set of loss-of-function ASD genes *in vitro* in human neural progenitor cells (27 genes)[24] and human brain organoids (3 genes)[25], and *in vivo* in fetal mouse brains (35 genes)[26] and *Xenopus tropicalis* (10 genes)[27], reporting convergence impacting neurogenesis[24-27], WNT signaling[24], and gene expression[25,26]. This approach has not yet been applied to SCZ genes, loss-of-function[28] or otherwise.

We previously integrated hiPSC-based models with CRISPR gene editing, activation (CRISPRa) and interference (CRISPRi) technologies, in order to study a putative causal SNP (*FURIN* rs4702) and top-ranked putative GWAS target genes (*FURIN, SNAP91, TSNARE1, CLCN3*) associated with SCZ, observing genotype-dependent transcriptomic differences and resolving specific pre- and post-synaptic perturbations[29]. A CRISPR screen of putative GWAS target genes, associated with substantially smaller predicted effect sizes than loss-of-function genes, has not yet been reported in the context of any psychiatric disorder.

Here we resolved the convergent impact of twelve eGenes with strong evidence of up-regulation by Psychiatric Genomics Consortium (PGC3)-SCZ GWAS loci: *CALN1, CLCN3, FES, INO80E, NAGA, NEK4, PLCL1, SF3B1, TMEM219, UBE2Q2L, ZNF823, ZNF804A* (**Fig. 1a**). Two independent pooled single-cell RNA-sequencing CRISPRa-based experiments[30] were applied to resolve the genome-wide transcriptomic consequences of activating SCZ eGene expression at two developmental timepoints (day 7 or day 21) in human *NGN2*-induced

glutamatergic neurons (**Fig. 1b**)[31,32]. Our analyses revealed shared and distinct downstream effects of these twelve SCZ risk eGenes, independent of their previously annotated biological roles. We uncovered a convergent effect that increases with increasing polygenicity and within groups of functionally similar genes. Convergent networks showed brain-region and developmental period-specific enrichments that enrich for signaling networks involved in neurological function with known drug targets. Overall, convergent signatures are incredibly robust, observable across two independent lists of SCZ-GWAS target genes manipulated in two distinct ECCITE-seq[30] experiments, and resolvable in post-mortem brain. That convergent genes were enriched for risk factors and transcriptomic signatures associated with a range of other brain disorders suggests that convergent targets may in part explain shared features of psychiatric disorders and cross-disorder pleiotropy of risk and supports the hypothesis that common and rare psychiatric disease-associated variants converge on the same biological pathways[29], providing crucial insight into potential mechanisms of polygenicity in complex traits.

RESULTS

Prediction and prioritization of brain genes activated by SCZ GWAS loci.

SCZ-associated risk SNPs identified by GWAS are highly enriched in *cis* regulatory elements, including promoters and enhancers, and frequently regulate the expression of one or more target genes[33-35] as cis- or trans-acting expression quantitative trait loci (eQTLs)[36]. Transcriptomic imputation studies predict trait-associated genetically regulated gene expression by combining large-scale genetic data (GWAS summary statistics) with tissue-specific eQTL reference panels[37-39]. These approaches infer pathophysiological consequences of GWAS association statistics, conferring magnitude, directionality, and tissue-level precision. eGenes (i.e. genes with an associated eQTL) whose brain expression was predicted to be up-regulated by PGC3-SCZ GWAS SNPs[3] were selected by two complementary approaches. Each list of genes was used to design, construct, and validate a distinct gRNA library.

First, transcriptome and epigenome imputation were performed for SCZ to prioritize: *NEK4, PLCL1, UBE2Q2L, NAGA, FES, CALN1,* and *ZNF804*. The PsychENCODE transcriptome datasets of genotyped individuals (brain homogenate, n=924)[14,40] were leveraged to impute brain transcriptomes at the level of genes and isoforms with EpiXcan[41]; EpiXcan increases power to identify trait-associated genes under a causality model by integrating epigenetic annotation[42]. In addition, we used PrediXcan[39] to perform epigenomic activity imputation for H3K27ac (brain homogenate, n=122; neuronal, n=191) and H3K4me3 (neuronal, n=163)[43] to more confidently identify *cis* regulatory elements associated with risk for SCZ. Overall, SCZ eGenes were prioritized from PGC3-GWAS[3] based on: i) significant genetic up-regulation of expression (z-score >6 for genes; (**Table 1**), ii) epigenetic support (>1 epigenome assay with association within the promoter of the gene) (**Table 1; Supplementary Fig. 1b**), iii) exclusion

of genes located in the major histocompatibility complex (MHC) locus. Six top coding genes and one top pseudo-gene were selected; all seven were subsequently robustly detected by scRNAseq (see below).

Second, transcriptomic imputation was considered together with the strength of co-localization between GWAS risk loci and brain eQTL peaks to identify: *CALN1*, *CLCN3, CUL9, DOC2A, PLCL1, INO8E0, SF3B1, SNAP91, TMEM219, ZNF823*. Post-mortem Common Mind Consortium (CMC)[44] and PGC3-SCZ GWAS[3] data was leveraged to calculate predicted differential expression in brain (using prediXcan[39]; 623 CMC samples), and colocalization of fine-mapped GWAS and eQTL associations (using COLOC[45,46]; 537 EUR-CMC samples; 65,205 cases and 87,919 controls GWAS). Our analysis yielded ~250 genes with significant ($p<6x10^{-6}$)[39] predicted differential expression between SCZ-cases and controls using PrediXcan and 25 loci with very strong evidence (PP4>0.8) of colocalization of PGC3-GWAS loci and brain eQTLs; there was significant overlap between the two analyses (22/25 COLOC genes were PrediXcan significant; binomial test p-value $3.03x10^{-112}$). Overall, SCZ eGenes were prioritized from PGC3-GWAS[3] based on: i) significant genetic up-regulation (**Table 2**), ii) colocalization of GWAS and eQTL associations, iii) exclusion of genes located in the MHC locus, iv) robust expression in our hiPSC neuron RNAseq (**Table 2**). Although ten top coding genes were selected, only five were subsequently robustly detected by scRNAseq (see below).

A pooled CRISPRa approach to resolve the transcriptomic effects of SCZ eGenes at two developmental time-points in human excitatory neurons.

Fusion of a catalytically inactive dCas9 to the tripartite activator VP64-p65-Rta (VPR)[47,48] achieves efficient up-regulation in hiPSC neurons[49]. Coupling CRISPR screening to single-cell RNA sequencing readouts yields rich, high-dimensional phenotypes from pooled screens and permits direct detection of sgRNAs (Expanded CRISPR-compatible CITE-seq, ECCITE-seq)[30]. To evaluate the impact of the twelve SCZ prioritized genes, the two independently designed, constructed, and validated pooled CRISPRa libraries were transduced into *NGN2*-neurons from two donors in independent experiments (three donors across experiments) at unique developmental time-points.

In order to evaluate the impact of the seven SCZ prioritized genes, a pooled library of ten new sgRNAs per SCZ eGene was designed (*NEK4, PLCL1, UBE2Q2L, NAGA, FES, CALN1, ZNF804A*: 10 gRNAs each), and pooled together with 1 scramble gRNA negative control previously functionally validated by CRISPRa[29]. The synthesized, annealed, and pooled gRNAs were cloned into a lentiviral expression vector (lentiGuide-Hygro-mTagBFP2)[49].

To validate a high-quality pooled sgRNA library, an innovative and more accurate method to determine individual sgRNA abundance was developed. Note that the 20-bp gRNA sequences are the only different small parts in the ~8,000 bp expression plasmid. DNA fragmentation by sonication yields redundant and repetitive DNA fragments that are not of interest. To increase the probability of getting sequences of the specific gRNAs, the gRNA library was digested with two restriction enzymes to yield a ~378 bp DNA fragment containing different gRNAs

(**Supplementary Fig. 1c**). Following Illumina's protocol for library preparation, index adapter sequences were added to the ends of a DNA fragment for MiSeq (25 million sequencing reads and 2× 250 bp read lengths). After confirming the equal distribution of individual gRNA within the library (**Supplementary Fig. 1d**), it was packaged into lentivirus and titrated.

Second, three sgRNAs were designed for each of the genes (*CALN1*, *CLCN3, CUL9, DOC2A, PLCL1, INO8E0, SF3B1, SNAP91, TMEM219, ZNF823).* gRNAs were cloned for expression from our LV expression vector[49], individually expressed, and validated by qPCR. The gRNAs with validated expression predicted to mimic the strength of the eQTL effect most closely was selected. Individual gRNAs were packaged as lentiviruses and tittered prior to pooling**.**

An sgRNA library targeting promoters (low MOI, ~300 cells/sgRNA) was used to generate strong-effect up-regulation and maximize detection of trans effects. After maturation, 7-day-old (7-gene) and 21-day-old (10-gene) *NGN2*-neurons were dissociated to single cell suspensions with papain, and bar-coded single cell cDNA generated using 10X Genomics Chromium. Normalization and downstream analysis of RNA-seq data were performed using the Seurat R package (v.2.3.0), which enables the integrated processing of multimodal single-cell datasets. Differential gene expression analysis was performed on demultiplexed cells for each experiment individually, with a total of 12 target genes being compared to their matched scramble control targeted cells, treating each cell as a replicate.

We were able to successfully resolve cell perturbations of all seven genes queried in the first set (each of which had 10 unique gRNAs) and five of the ten genes queried in the second set (each of which had 3 unique gRNAs). The inability to resolve all ten of the perturbations in the day 21 NGN2s is likely due to the difference in gRNA number used to target the eGenes and suggests that either more cells or more gRNAs per target would be necessary to resolve all 10 perturbations. Overall, scRNA-seq on 8000 and 6000 cells respectively were adequate to resolve: the sgRNA in each cell, the *cis* target gene with differential expression, and the downstream *trans* alterations to pathways resulting from initial *cis* up-regulation. Following QC and normalization, twelve successful perturbations across both experiments were identified, with an average of 76 cells per perturbation (ranging from 21-284) for a total of 892 perturbed cells and 116 scramble controls.

Perturbations of individual SCZ eGenes have shared and distinct transcriptomic impacts. Successful CRISPR-activation of seven and five SCZ eGenes impacted the global transcriptome of 7-day and 21-day *NGN2*-excitatory neurons respectively and revealed shared and distinct effects. (**Supplementary Fig. 2a; i-viii, Supplemental Data 1**). Across both experiments, CRISPRa led to an average 1.5 log fold increase in most target genes [*CALN1*=1.54; *CLCN3*=1.75; *FES*=1.38; *NAGA*=1.6; *NEK4*=1.48; *INO80E*=1.5; *PLCL1*=1.50; *SF3B1*=1.38; *UBE2Q2L*=1.44; *ZNF823*=1.35] and roughly 2.0 log fold increase in the expression of the remainder [*ZNF804A* (logFC=1.93) and *TMEM219* (logFC=2.05)] **(Supplementary. Fig. 2b; i-vi**). Gene set enrichment analyses (GSEA) of the 7-Target and 5-Target perturbations were performed individually across Gene

Ontology Terms and KEGG Pathways GSEA terms. Neurodegenerative disorders pathways were significantly enriched across 3 or more perturbations in the 7 Targets in D7 Neurons and the 5 Targets in D21 Neurons. Additionally, perturbations across both experiments were enriched for neuroactive ligand-receptor interaction, protein processing in the endoplasmic reticulum, as well as enrichments in the proteasome and spliceosome. While five perturbations in the 7-target D7 NGN2 set were notably enriched for WNT signaling and three or more perturbations in the 5-Target set were notably enriched for dopaminergic synapse and synaptic vesicle cycling (**Supplemental Data 2**). Overall, these shared enrichments suggest that perturbation of these genes might impact similar neural processes and pathways. Supporting this, meta-analysis of DEGs across all 12 SCZ eGene perturbations in *NGN2*-neurons identified 790 significantly down-regulated genes and 10 significantly up-regulated genes that shared the same direction of effect across all targets (Bonferroni meta p-value<=0.05), suggesting they may be involved in similar neuronal pathways (**Supplemental Data 1**). These overlapping DEGs were enriched for numerous gene ontology terms related to brain development, neuronal morphology, signaling, and transcriptional regulation (**Supplemental Data 2**). Notably, these shared genes were enriched for the Hallmark Signaling by MTORC1 (adj. p=2.72e-2) and Interferon Alpha (adj. p=2.24e-3) and Interferon Alpha Response (adj. p=3.10e-2), and TNF alpha signaling via NFKB (adj. p=2.72e-2)—pathways contributing to biological mechanisms underlying effects of Maternal Immune Activation (MIA) on risk for development of psychiatric disorders.

Convergent downstream impacts change with increasing polygenicity. To further investigate the potential downstream convergent effects of these perturbations, Bayesian bi-clustering and Target-Convergent Network (TCN) reconstruction was performed using the normalized average gene expression across all cells (N=1006). To account for neuronal maturity differences, expression matrices were batch corrected and normalized and the scramble cells from both experiments (matched gRNA) used as a single control population. Network reconstruction of genes clustering across all 12-Target CRISPRa-perturbed cells, representing the highest level of polygenicity modeled in these experiments, only (and not found in Scramble cells; n=892) identified 1048 node genes, with a total membership of 1248 genes that clustered together in at least 50% of the runs (**Fig. 2a**). Network reconstruction of genes clustering together across all 7 CRISPRa-perturbations (mid-level polygenicity) identified 319 node genes with a total membership of 1029 genes that clustered together in at least 50% of the runs (n=502) (**Fig. 2b**). Network reconstruction of genes across all 5 CRISPRa-perturbations (low-level polygenicity) showed less clustering-based convergence, identifying 163 nodes and 166 total genes with only 15% duplication of nodes to target connections across all 40 runs (n=390) (**Fig. 2c**).

To understand the biological pathways within the convergent networks, node sub-networks were explored, and functional gene set annotation was performed (**Supplementary Data 3,4**).. Target-convergent networks identified by Bayesian bi-clustering across high-, mid-, and low-polygenicity sets show shared and distinct pathway enrichments. Both the high- and mid-polygenicity networks are

significantly enriched for brain and embryonic developmental traits, including neuron subtype differentiation, patterning and regionalization, and brain-region specific development (**Fig. 2d**). Additionally, there are across-network enrichments in shared cellular components including channel and receptor complexes and neuron/synapse parts (**Fig 2e-f**). Enrichments unique to each network were found, including enrichments observed only in the high-polygenicity network, but not mid- or low-polygenicity networks. Overall, the high- and mid-polygenicity networks share numerous enrichments for terms related to brain development, signaling, and receptor activity. However, the high-polygenicity network uniquely converges on neurotransmitter transport and signaling (including glutamate, catecholamine, and norepinephrine) while the mid-polygenicity network uniquely converges on membrane depolarization (adj. p=0.03) and action potential propagation (adj. p=0.03) (**Fig. 2f**).

Overall, our network-based characterization of the global gene expression changes following CRISPRa of these 12 SCZ eGenes in hiPSC-derived glutamatergic neurons revealed multiple points of SCZ-relevant convergence. Of note, developmental pathways involved in patterning, regionalization, and growth (WNT and Homeobox signaling), neuro-active ligand receptor interactions, and voltage-gated ion channels. While many of these enrichments were shared when restricting the network to seven SCZ eGene perturbations, modeling mid-level polygenicity in these experiments, unique convergence was identified between the two sets. Additionally, there were dramatically fewer enrichments across the five SCZ eGene perturbations, indicating increases convergence with increasing polygenicity.

Convergent downstream impacts of functionally similar SCZ eGenes. To investigate whether downstream convergent effects of these perturbations would increase in strength or specificity based on shared function, the 12 targets were subset into two functional groupings of six target genes each. Network reconstruction of clustering across 6 “signaling” target genes (*CALN1, CLCN3, FES, NAGA, PLCL1, TMEM219*; n=385) identified 123 nodes and a total network of 177 genes whose connections duplicated in at least 25% of the bi-clustering runs (**Fig. 3a**). Network reconstruction of clustering across 6 “epigenetic/regulatory” target genes (*NEK4, INO80E, SF3B1, UBE2Q2L, ZNF823, ZNF804A*; n=527) identified 99 node genes with a total membership of 105 genes that clustered together in at least 15% of the runs (**Fig. 3d**).

The signaling network converged on embryonic and brain development and voltage-gated ion channel signaling activity. Specifically, it was significantly enriched for numerous terms related to neurogenesis, pattern specification, regionalization, neuron and glial cell differentiation and voltage-gated potassium channel activity and potassium ion transport (**Fig. 3b-c**). The regulatory network converged on secretory pathways, significantly enriched for multiple gene ontology terms related to regulation of hormone and peptide secretion. It was also significantly enriched for positive regulation of pathway restricted SMAD protein phosphorylation (adj. p<0.01), which plays a role in regulating tissue patterning during embryonic development and receptor protein serine threonine kinase

signaling important (adj. p<0.05) for regulating gene expression and cell homeostasis, proliferation, and death (**Fig. 3e**). Overall, perturbations of SCZ eGenes with shared functional pathways show greater specificity in convergence.

Target-convergent networks show brain-region and developmental-period specific enrichments. Enrichment analysis GTeX v8 brain tissues for the three convergent networks of increasing polygenicity and the two functionally convergent networks revealed that convergent genes identified in the high- and mid-polygenicity networks are most strongly upregulated in the cortex (Frontal Cortex BA9: 12-Target adj. p=2.85e-29; 7-Target adj. p=5.76e-37, Cortex: 12-Target adj. p=2.94e-36, 7-Target adj. p=2.21e-31) while those in the low-polygenicity network are most strongly upregulated in the hippocampus (adj. p=9.77e-18), hypothalamus (adj. p= 2.09e-17), and caudate basal ganglia (adj. p=3.90e-16) relative to its other enrichments (**Fig. 4a**). Enrichment analysis across BrainSpan developmental time-points showed that genes shared across the high- and mid-polygenicity networks are strongly up-regulated during adolescence (12-Target adj. p= 1.09e-07; 7-Target adj. p=3.17e-07) and young adulthood (12-Target adj. p=8.76e-06; 7-Target adj. p=1.01e-06). Overall, individual networks have both shared and distinct developmental and tissue-specific directional enrichments based on (1) polygenicity and (2) function (**Fig. 4b**). Given this developmental and regional specificity, over-representation analysis was run on a curated set of rare and common variant target genes across autism spectrum disorder (ASD), bipolar disorder (BIP), SCZ, and intellectual disability (ID) to determine if target-convergent networks distinctly contribute to psychiatric risk. Convergent genes present in the high- and mid-polygenicity networks are enriched for ASD, BIP, SCZ, and ID common and rare variant genes, while low-polygenicity and functionally curated networks are only enriched for rare variant gene targets. Networks modeling higher polygenicity are enriched for rare and common psychiatric risk variants while low-polygenicity and functional networks are only enriched for rare variants (**Fig. 4c**).

SCZ eGene perturbations converge on signaling networks involved in neurological function with known drug targets. Looking across network membership, functional annotation identified 900 annotated drugs and compounds in the DrugBank database targeting 187 genes (**Fig. 4d**). Over representation analysis revealed that targets of anti-psychotics, anesthetic, hypnotics, and sedatives were broadly present across networks. Distinct clustering of drug enrichments included calcium channel blockers in the mid-polygenicity network, hormone signaling and steroids in the low-polygenicity network, anticholinergics in the signaling network, and anti-depressants and amphetamines in the regulatory network (**Fig. 5A**). Although none of these enrichments were significant after multi-testing correction, they show a tendency for distinct drug-type clusters within each network and provide a list of drug-able targets.

Targetable genes shared across the high-/low-polygenicity and 6-signaling networks include numerous neuroactive ligand receptors such as gamma-aminobutyric acid receptors (*GABBR1, GABRD, GABRG2*), glutamate ionotropic kainate receptors (*GRIK2, GRIK3, GRIK4*), cholinergic muscarinic (*CHRM4*), and

nicotinic receptors (*CHRFAM7A*), glutamate ionotropic NMDA receptors (*GRIN2D*), purinergic receptor (*P2R*), neuro-peptide Y receptor (*NPY5R*) and G protein-coupled receptors (*ADORA2A, ADORA2B, ADRB2, GPER, GPRC5A*). They also included drug-targetable ion channels, notably voltage-gated potassium channels (*KCNB1, KCNB2, KCNH6, KCNJ11, KCNJ12, KCNN2, KCNN4, KCNQ1*), voltage-gated calcium channels (*CACNA1H, CACNA2D3, CACNB2*), and voltage-gated (*SCN1B, SCN2B*) and ligand-gated (*SCNN1D*) sodium ion channels. Additionally, phosphodiesterase protein family genes (*PDE10A, PDE1C, PDE1C, PDE3A, PDE4B, PDE5A*), involved in neuro-inflammatory and cAMP signaling in the brain, and the solute carrier family (*SLC16A3, SLC16A6, SLC29A2, SLC32A1, SLC6A3*), involved in GABAergic and dopaminergic transmission, were represented across all networks with numerous known targeting-drugs (**Supplementary Data 4**).

Loxapine, bupropion, and nomifensine are antipsychotics used in the treatment of schizophrenia spectrum disorders, mania in bipolar disorder, and psychotic depression. They target SLC6A3, a dopamine transporter and node gene shared across the high-/mid-polygenicity and regulatory networks, but not the low-polygenicity or signaling networks (**Fig. 5C**). While trifluptomazine and sulpride are antipsychotics that target BCHE, a cholinesterase involved in cholinergic signaling and a node gene shared between the mid-polygenicity and functional signaling network (**Fig. 5D**).

Replication of convergent signatures in the post-mortem dorsolateral prefrontal cortex (DLPFC). K-means clustering of CMC post-mortem DLPFC genes identified individuals with shared patterns of upregulation in the 12 SCZ eGenes; the average Z-scores of each target gene across all clusters reported in **Supplemental Data 5**. Targets within each cluster with Z-score >= +0.250 were assigned as perturbed (**Fig. 6A**). Target-convergent network reconstruction was run in our scRNA-seq data based on the CMC cluster perturbation assignments. Of the eight clusters, networks were recovered for the combination of targets represented in cluster 1 (7 targets, 417 cells; 22.5% duplication; 52 node genes), cluster 3 (6 targets; 526 cells; 20% duplication; 184 node genes), cluster 4 (4 targets; 300 cells; 20% duplication; 142 node genes), cluster 5 (3 targets, 419 cells, 12.5% duplication, 181 node genes); cluster 6 (2 targets, 60 cells, 5% duplication, 64 node genes), cluster 7 (9 targets; 721 cells; 25% duplication; 772 node genes) and cluster 8 (8 targets; 664 cells; 25% duplication, 325 node genes), but not for cluster 2 (4 targets; 199 cells). For network connections that duplicated in at least 20% out of the 40 runs, we calculated the variance in gene expression of their node genes in the CMC DLPFC.

To assess if convergence of the recovered networks replicated in the post-mortem DLPFC, we compared the gene variance of node genes in one cluster to the gene variance of the same genes in all other clusters. The variance in node gene expression of cluster 3 was less than that in all other clusters, except cluster 4, while the variance in node gene expression for cluster 1 was not less than the variance in all other clusters (**Fig, 6B; Table 3**). CMC clusters were significantly different based on the diagnosis status of individuals within a cluster (Pearson's

Chi-squared; X-square=140.84, df=21, p-value=9.51e-20) (**Fig. 6D**). Cluster 3 was strongly positively associated with SCZ (Chi-square Residual = 3.984), while cluster 1 was negatively associated with SCZ (-3.782) and positively associated with control status (2.946) (**Fig. 6C**).

Overall, we were able to recover convergent networks in our scRNA-seq data using matched perturbation status and replicate convergent signatures of network node genes in the post-mortem DLPFC (**Supplementary Figure 3**). CMC clusters associated with SCZ had less variance in node gene expression than those strongly associated with control status. Additionally, clusters with the most overlap in represented perturbation targets were most like one another, having similar variance in their respective node gene expression.

DISCUSSION

A multiplexed CRISPR-based approach resolved the impact of twelve SCZ eGenes (*CALN1, CLCN3, FES, INO80E, NAGA, NEK4, PLCL1, SF3B1, TMEM219, UBE2Q2L, ZNF823, ZNF804A*) in human glutamatergic neurons, uncovering a convergent effect concentrated on neurodevelopmental pathways, neuroactive ligand receptor signaling, and voltage-gated ion channel activity. Because convergence increased with increasing polygenicity and specificity increased within groups of functionally similar genes, our results are overall consistent with the polygenic additive model[50], but do suggest that variants may sum in different patterns and with greater specificity depending upon the extent to which their target genes converge within the same biological pathways[51]. The most highly shared convergent nodes occur where the impact of aggregate genetic risk is strongest, are predicted to be druggable, and so represent novel points of therapeutic intervention.

Overall, convergent signatures are incredibly robust, observable across two independent lists of SCZ-GWAS target genes manipulated in two distinct ECCITE-seq experiments and are resolvable in post-mortem brain. Two independent pooled single-cell RNA-sequencing CRISPRa-based experiments[30] were applied to resolve the genome-wide transcriptomic consequences of activating SCZ eGene expression in two distinct sets of genes, at two developmental timepoints (day 7 or day 21) in human *NGN2*-induced glutamatergic neurons across three donors and generated either directly from hiPSCs or instead from hiPSC-derived neural progenitor cells (NPCs)[31,32]. The clinical relevance of our results was further supported through replication of target convergent networks reconstruction of perturbation groupings based on post-mortem gene signatures of our targets in the dorsa-lateral pre-frontal cortex (DLPFC). In the post-mortem DLPFC, convergence was stronger in clusters enriched for individuals diagnosed with SCZ, and convergent signatures were more similar between individuals with the most overlap in upregulation of the 12 SCZ eGenes.

Coupling CRISPR-based perturbations to single-cell RNA sequencing of isogenic neuronal experiments, we not only increased the number of genes that can be queried, but likewise reduced experimental variation, altogether increasing our

power to resolve convergence. However, there are limitations to this approach. First, it is necessary to test the generalizability of our findings beyond the set of twelve genes chosen here; new CRISPRa and CRISPRi systems use different gRNA scaffold sequences[50,52], making new bi-directional, combinatorial gene perturbations in the same cell theoretically possible. Second, our CRISPRa analyses are from glutamatergic neurons only, and so do not consider the convergence across cell types or within more complex neuronal circuits. Finally, being most like fetal brain cells, hiPSC neurons cannot capture convergence that occurs at the time of symptom onset and thereafter; transcriptional consequences that may affect neurobiology during childhood or adolescence cannot be easily modelled with hiPSC-based platforms. Thus, future investigation to assess how transcriptional convergence differs across gene sets, drug/environmental contexts[52], brain regions[53] and cell types[54], developmental timespans[55], and donor backgrounds[56] will inform the cell-type-specific and context-dependent nature of convergence.

Given that our gene targets were prioritized based on the genetic and genomic evidence supporting their activation by SCZ-GWAS loci, it will be informative to revisit our findings across a larger number of genes and pathways, specifically considering whether manipulations made within or between biological pathways impact the convergence observed. Likewise, it would be interesting to repeat these studies in neural progenitor cells and/or immature neurons. Intriguingly, although extensive evidence links ASD risk genes to both glutamatergic and GABAergic neurons[57], previously reported convergence of ASD genes impacts occurred in neural progenitor cells[24,26,27], impacting neurogenesis[24,26,27], WNT signaling[24] and gene expression[26]. A pressing question is thus the mechanisms by which convergence ultimately impacts synaptic activity, circuit function, and behavior. Notably, here we included one pseudo-gene; given the unexpected phenotypic impact of its manipulation, and because multiple GWAS variants map to pseudo-genes, which may play critical roles in gene silencing and transcriptional regulation, it would be particularly interesting to explore the functional impact of these less well-explored genomic elements.

Our convergent genes are interesting to consider in the context of additive[16,50], epistatic[58], and omnigenic[18,51] models of inheritance. Do they represent “core genes” as conceptualized in the omnigenic model? Given our previous findings that there is emergent biology that results from combinatorial perturbation of SCZ risk genes beyond what would be expected from the additive effect of individually perturbed genes[29], and that these synergistic effects converges on synaptic function and psychiatric risk genes, an important next step will be to manipulate these risk genes alone and together. Our hope is that high-throughput combinatorial CRISPR-based perturbation methods will better resolve the complex genetic risk architecture of psychiatric disorders, including the convergent gene regulatory networks and/or the identity of peripheral master genes directing genetic risk.

There is great value in continuing to explore as many rare and common risk loci as possible, agnostic to previously defined function, in a cell-type-specific and

context-dependent manner. The translational impact of our work includes potential improvements to additive polygenic risk scores, prioritization of convergent genes for mechanistic follow-up, and identification of pathways that might serve as potential therapeutic targets. Overall, the as-yet unidentified regulatory nodes that lead to disruptions in synaptic function related to psychiatric endophenotypes represent exciting and unexplored therapeutic targets. Our overarching goal is to advance the field towards an era of precision medicine[59], whereby not just each patient's genetic variants, but also the expected interactions between them, can be used to predict symptom development, disorder trajectory, and potential therapeutic interventions.

MATERIALS AND METHODS

Prioritization of Seven SCZ eGene List: Day 7 hiPSC-derived NGN2-neuron library. eGenes are defined as genes with significant genetic regulation of gene expression levels. For prioritization, we used the Psychiatric Genomics Consortium (PGC3)-SCZ GWAS[1,2]. Transcriptomic imputation was performed with the EpiXcan[41] method utilizing the PsychENCODE transcriptome datasets of genotyped individuals[14] and epigenetic annotations from REMC[60] as previously described[61]; transcriptomes were imputed at the gene and isoform levels and features with training cross-validation $R^2 \geq 0.01$ were retained. The epigenetic imputation models were built with the PrediXcan[39] method (using a 50kbp window instead of 1Mbp for transcripts) utilizing the recently described ChIPseq datasets[43]; summary-level imputation was performed with S-PrediXcan[37]. Peaks were assigned to genes with the ChIPseeker R package[62]. Genes were ranked based on the association z-score for imputed gene expression. Genes that had no nominally associated imputed epigenetic activity ($p<0.01$) across the three different assays were not considered. Genes within the MHC were not considered. Six top coding genes and one top pseudo-gene were selected: *NEK4, PLCL1, UBE2Q2L, NAGA, FES, CALN1,* and *ZNF804.*

SCZ eGene Prioritization: Day 21 NPC-derived NGN2-neuron library.

Transcriptomic imputation calculated using PrediXcan[39] from post-mortem CMC[44] and PGC3-SCZ GWAS[3] data identified genes with significant ($p<6x10^{-6}$)[39] predicted differential expression between SCZ-cases and controls. Colocalization (using COLOC[45,46]) identified fine-mapped GWAS and eQTL associations with very strong evidence (PP4>0.8) of colocalization. The top significant SCZ eGenes identified by both analyses were prioritized. Any genes located in the major histocompatibility complex (MHC) locus were subsequently removed from the list. Ten top coding genes were selected: *CALN1*, *CLCN3, CUL9, DOC2A, PLCL1, INO8E0, SF3B1, SNAP91, TMEM219, ZNF823*. This list was intersected with the eGene set previously prioritized and functionally evaluated in hiPSC-neurons[29]; of the 108 SCZ-GWAS loci identified by the Psychiatric Genomics Consortium (PGC2)-SCZ GWAS[1,2], 19 were shown to harbor co-localized SCZ-GWAS and *cis*-eQTL (SNP–gene pairs within 1 Mb of a gene; COLOC PP4>0.8) signals in the CommonMind Consortium (CMC) post-mortem brain RNA-Seq analysis of the

Dorsa-Lateral Pre-Frontal Cortex (DLPFC)[44], five of which were predicted to involve only a single protein-coding gene, four were well-expressed in *NGN2*-neurons, and three were up-regulated in SCZ and could be robustly manipulated by CRISPRa: *SNAP91*, *TSNARE1*, and *CLCN3*. One functionally validated gRNA was included for each of these three genes[29].

Neuronal Differentiation of clonal hiPSCs: Clonal hiPSCs from two donors (#690 and #2607) with lenti-EF1a-dCas9-VPR-Puro (Addgene #99373), pLV-TetO-hNGN2-eGFP-Neo, and lentiviral FUW-M2rtTA (Addgene #20342) were maintained in StemFlex™ Medium (ThermoFisher #A3349401) and passaged with EDTA (Life Technologies #15575-020).

On day 1, induction media (DMEM/F12 (ThermoFisher #10565,), 1× N2 (ThermoFisher #17502-048), 1× B27-RA (ThermoFisher #12587-010), 1× Antibiotic-Antimycotic (ThermoFisher #15240096), and 1 µg/mL doxycycline) was prepared and dispensed 2 mL of suspension at 1.2×10^6 cells/well in induction media onto a 6-well plate coated with matrigel (Corning #354230). On day 3, media is replaced with induction medium containing 1 µg/mL puromycin and 1 mg/mLG418. On day 5, split neurons were replated onto matrigel-coated plates and cells were dissociate with Accutase (Innovative Cell Technologies) for 5-10 min, washed with DMEM/10%FBS, gently resuspended, counted and centrifuged at 1,000×g for 5 min. The pellet was resuspended at a concentration of 1×10^6 cells/mL in neuron media [Brainphys (StemCell Technologies #05790), 1×N2 (ThermoFisher #17502-048), 1×B27-RA (ThermoFisher #12587-010), 1 mg/ml Natural Mouse Laminin (ThermoFisher #23017015), 20 ng/mL BDNF (Peprotech #450-02), 20 ng/mL GDNF (Peptrotech #450-10), 500 mg/mL Dibutyryl cyclic-AMP (Sigma #D0627), 200 nM L-ascorbic acid (Sigma #A0278)] with doxycycline, puromycin, G418 [4µM Ara-C (Sigma #C6645)] and 1×Thiazovivin (Sigma #420220). Cells were seeded 5×10^5 per 12-well plate. On day 7, neurons were harvested for scRNA sequencing.

NGN2-glutamatergic neuron induction of hiPSC-derived NPCs[31,32]: hiPSCs-derived NPCs were dissociated with Accutase Cell Detachment Solution (Innovative Cell Technologies, # AT-104), counted and transduced with rtTA (Addgene 20342) and *NGN2* (Addgene 99378) lentiviruses in StemFlex media containing 10 µM Thiazovivin (Millipore, #S1459). They were subsequently seeded at $1x10^6$ cells/well in the prepared 6-well plate. On day 1, medium was switched to non-viral induction medium (DMEM/F12 (Thermofisher, #10565018), 1% N-2 (Thermofisher, #17502048), 2% B-27-RA (Thermofisher, #12587010)) and doxycycline (dox) was added to each well at a final concentration of 1 µg/mL. At day 2, transduced hiPSCs were treated with 500 µg/mL G418 (Thermofisher, #10131035). At day 4, medium was replaced including 1 µg/mL dox and 4 µM cytosine arabinoside (Ara-C) to reduce the proliferation of non-neuronal cells. On day 5, young neurons were dissociated with Accutase Cell Detachment Solution (Innovative Cell Technologies, # AT-104), counted and seeded at a density of $1x10^6$ per well of a Matrigel-coated 12-well plate. Medium was switched to Brainphys neuron medium (Brainphys (STEMCELL, # 05790), 1% N-2, 2% B--27-RA, 1 µg/mL Natural Mouse Laminin (Thermofisher, # 23017015), 10 ng/mL BDNF

(R&D, #248), 10 ng/mL GDNF (R&D, #212), 500 µg/mL Dibutyryl cyclic-AMP (Sigma, #D0627), 200 nM L-ascorbic acid (Sigma, # A4403)). For seeding, 10 µM Thiazovivin (Millipore, #S1459), 500 µg/mL G418 and 4 µM Ara-C and 1 µg/mLdox were added. At day 6, medium was replaced with Brainphys neuron medium with 4 µM Ara-C and 1 µg/mL dox. Subsequently, 50% of the medium was replaced with fresh neuronal medium (lacking dox and Ara-C) once every other day until the neurons were harvested at day 21.

Neuronal ECCITE-seq. CRISPRa hiPSC *NGN2*-neurons (mid-PRS control donors 2607 (XY) and 690 (XY)) were transduced with the pooled gRNA at day -1. After maturation, 7-day-old *NGN2*-neurons were dissociated to single cell suspensions with papain, antibody-hashed[63], and bar-coded single cell cDNA generated using 10X Genomics Chromium[64]. NPC-derived NGN2-neurons (mid-PRS control donors 2607 (XY) and 553 (XY)) were transduced with the mixed-pooled gRNA vectors (Addgene 99374) at day 17. At day 21, media was replaced by 0.5ml/well accutase containing 10 µm Rock inhibitor, THX (catalog no. 420220; Millipore) for 1 hour to dissociate neurons. Neurons were spun down (3 mins X 300g) and resuspended in DMEM/F12 + THX before proceeding to single cell sequencing.

ECCITE-seq library generation. Expanded CRISPR-compatible CITE-seq (ECCITE-seq)[30], combines Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq) and Cell Hashing for multiplexing and doublet detection[63] with direct detection of sgRNAs to enable single cell CRISPR screens with multi-modal single cell readout. By capturing pol III-expressed guide RNAs directly, this approach overcomes limitations of other single-cell CRISPR methods, which detect guide sequences by a proxy transcript, resulting in barcode switching and lower capture rates[65-67].

Bioinformatics. mRNA sequencing reads were mapped to the GRCh38 reference genome using the *Cellranger* Software. To generate count matrices for HTO and GDO libraries, the kallisto indexing and tag extraction (kite) workflow were used. Count matrices were used as input into the R/*Seurat* package[68] to perform downstream analyses, including QC, normalization, cell clustering, HTO/GDO demultiplexing, and DEG analysis[30,69]

Normalization and downstream analysis of RNA data were performed using the Seurat R package (v.2.3.0), which enables the integrated processing of multimodal single-cell datasets. Each ECCITE-seq experiment was initially processed separately. Hashtag and guide-tag raw counts were normalized using centered log ratio transformation, where counts were divided by the geometric mean of the corresponding tag across cells and log-transformed. For demultiplexing based on hashtag, Seurat::HTODemux function was used; and for guide-tag counts Seurat::MULTIseqDemux function within the Seurat package was performed with additional MULTIseq semi-supervised negative-cell reclassification. Cells with RNA UMI feature counts were filtered (200 < nFeature_RNA < 8000) and the percentage of all the counts belonging to the mitochondrial, ribosomal, and hemoglobin genes calculated using Seurat::PercentageFeatureSet. To remove variation related to cell-cycle phase of individual cells, cell cycle scores were assigned using Seurat::CellCycleScoring which using a list of cell cycle markers[70]

to segregate by markers of G2/M phase and markers of S phase. RNA UMI count data was then normalized, log-transformed and the percent mitochondrial, ribosomal, and hemoglobin genes, cell cycle scores, and HTO-maxID regressed out using Seurat::SCTransform. The scaled residuals of this model represent a 'corrected' expression matrix, that was used for all downstream analyses. To ensure that cells assigned to a guide-tag identity class demonstrated successful perturbation of the target gene, the percentage of the feature counts belonging to the corresponding target gene for each cell was calculated using Seurat::PercentageFeatureSet and cells removed with target gene count percentages equal to 0. For the non-targeting identity class (scramble gRNA), cells were filtered based on having count percentages across all target genes equal to 0. These cells were then used for differential gene expression analysis [71].

Analysis of the transcriptomic impact of activating SCZ risk eGenes. Differential gene expression analysis was performed between each CRISPRa target group (cells assigned by targeting sgRNAs of that gene) and control group (cells assigned by non-targeting control sgRNAs) using the R packages EdgeR[72] and Limma[73] treating each cell as one replicate as in[71]. Mitochondrial, ribosomal, and hemoglobin genes from all downstream analyses were removed (['^MT-|^RP[SL][[:digit:]]|^RPLP[[:digit:]]|^RPSA|^HB[ABDEGMQZ][[:digit:]]') as well as genes that had zero read counts in more than 90% of samples leaving 17,462 genes for analysis in set 1 and 17,471 genes for analysis in set 2. Limma:voom[74] was used to transform data into log2cpm, and limma::lmFit[75] to fit a linear regression model and test the effect of each CRISPRa perturbations against the Scramble-gRNA Control individually. Bayes shrinkage (limma::eBayes) estimated modified t- and p- values and identified differentially expressed genes (DEGs) based on an FDR <= 0.05 (limma::TopTable)[75]. Gene-set Enrichment Analysis (GSEA) for individual perturbations was performed using the R package ClusterProfiler[76] (**Supplemental Data 1**).

Meta-analysis of gene expression across perturbations. We performed a meta-analysis and heterogeneity test (METAL[77]) using the p-values and direction of effects (t-statistic), weighted according to sample size across all twelve perturbations (Target vs. Scramble DEGs). Genes were assigned as convergent if they (1) had the same direction of effect across all 12, 7, and 5 target combinations, (2) had a META Bonferroni adjusted p-value <= 0.05, and (3) had a heterogeneity Bonferroni adjusted p-value = 1 (**Supplemental Table 1**).

Bayesian Bi-clustering to identify Target-Convergent Networks. Target-Convergent gene co-expression Networks (TCN) were built using an unsupervised Bayesian biclustering model, BicMix on the log2cpm expression data from all the cells across 12 targets and Scramble gRNA jointly[78]. To perform this as a joint analysis across two experiments, (1) gene expression matrices retaining genes only present in both experiments (N=16650) were combined, (2) and limma:voom normalization and transformation were used to correct for batch and to compute the log2cpm counts from the effective pooled library sizes. 40 runs of BicMix was conducted on these data and the output from iteration 300 of the variational Expectation-Maximization algorithm was used. The hyperparameters for BicMix

were set based on previous extensive simulation studies[79]. Target-Convergent Network reconstruction[80] was performed to identify convergent networks across all 12 targets, the 7 EpiXcan prioritized targets, the 5 PrediXcan prioritized targets, and specific clustering of 6 targets to generate functionally similar convergent networks for regulatory (*NEK4, INO80E, SF3B1, UBE2Q2L, ZNF804A, ZNF823*) and signaling targets (*CALN1, CLCN3, FES, NAGA, PLCL1, TMEM219*). Network connections that did not replicate in more than 15% of the runs were excluded and nodes with less than 5 edges were removed from gene set enrichment analyses. Using FUMAGWAS: GENE2FUNC, the protein-coding genes were functionally annotated and overrepresentation gene-set analysis for each network gene set was performed[81].

Functional enrichment for drug and disorder specific variant targets. Functional enrichment analysis was performed using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit)[82]. Over-representation analysis (ORA) was performed on all convergent network gene sets against known drug targets from DrugBank and against a curated list of common and rare variant target genes across ASD, BIP, SCZ, and ID[29].

Replication of Bayesian Bi-clustering to identify Target-Convergent Networks in Post-mortem Dorsa-lateral Pre-frontal Cortex. In vivo target convergent network reconstruction was replicated in the Common Mind Consortium (CMC; n=991 after QC) post-mortem dorsa-lateral pre-frontal cortex (DLPFC) gene expression data. We performed K-means clustering to subset the data into 8 clusters based the Z-scored gene expression of the 12 target genes. Perturbation identities were assigned based on average positive Z-scores of >=0.025 within each cluster. Perturbation assignments are as follows: Cluster 1: *CALN1, CLCN3, NAGA, INO80E, PLCL1, TMEM219, ZNF804A*; Cluster2: *INO80E, PLCL1, UBE2Q2L, ZNF804A;* Cluster3: *CLCN3, NEK4, PLCL1, SF3B1, ZNF804A, ZNF823*; Cluster4: *FES, NAGA, NEK4, ZNF823*; Cluster5: *FES, SF3B1, UBE2A2L*; Cluster6: *NEK3, UBE2Q2L;* Cluster7: *CALN1, CLCN3, FES, NAGA, NEK4, PLCL1, SF3B1, TMEM219, ZNF823;* Cluster8: *CLCN3, FES, NAGA, NEK4, INO80E, SF3B1, TMEM29, ZNF823*. We then assigned our single-cell data to clusters based on the overlap of perturbations and performed network reconstruction to replicate our convergent analysis using groups based on CMC post-mortem data. We retained clusters that resolved networks with at least 20% duplication rate and calculated the variance of expression in the CMC DLPFC for the corresponding network node genes and performed a Shapiro-Wilk test for normality. Gene variances were not normally distributed. To assess the specificity of convergence of the node genes of a given cluster, we performed a nonparametric Kolmogorov–Smirnov tests to determine if the central distribution frequency of the gene variance was less than that in all other clusters. We performed a Pearson's chi-squared test to determine whether there was a significant difference between the expected frequencies and the observed frequencies in diagnosis of Affective Disorders (AFF), Bipolar Disorder (BP), and Schizophrenia (SCZ) within the clusters.

STATEMENT OF ETHICS

Ethical approval was not required because the hiPSC lines, lacking association with any identifying information and widely accessible from a public repository, are thus not considered to be human subjects research. Post-mortem DLPFC data are similarly lacking identifiable information, and are not considered human subjects research.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

FUNDING SOURCES

This work was supported by F31MH130122 (K.G.T), R01MH109897 (K.J.B., P.R.), R56MH101454 (K.J.B., E.S., L.H.), R01MH123155 (K.J.B.) and R01ES033630 (L.H., K.J.B.), R01MH124839 (LMH), R01MH118278 (LMH). K08MH122911 (G.V.). R01MH125246 (PR), U01MH116442 (PR), R01MH109677 (PR), I01BX002395 (PR).

AUTHOR CONTRIBUTIONS

SCZ eGene lists were prioritized by E.S., L.H., W.Z., G.V. and P.R, together with K.J.B. Epigenome data provided by K.G., S.A. and P.R. The ECCITE-seq pipeline was adapted to hiPSC-neurons by A.L., supported by P.J.M.D., J.F., A.Y. and S.C. ECCITE-seq *NGN2*-neuronal studies were conducted by A.L. and P.J.M.D. 10x analyses were conducted by J.F, and preliminary ECCITE-seq quality control conducted by MW. Convergent analyses were conceptualized by K.J.B. and L.H., developed and conducted entirely by K.G.T. The paper was written by K.G.T., L.H. and K.J.B., with input from all authors.

Special thanks to Michael Talkowski and Douglas Ruderfer for countless discussions on convergence.

DATA AND CODE AVAILABILITY

All source donor hiPSCs have been deposited at the Rutgers University Cell and DNA Repository (study 160; http://www.nimhstemcells.org/); dCas9-VPR hiPSCs are in the process of being submitted in advance of publication.

The accession number for sc-RNA sequencing data reported in this paper is Gene expression omnibus (GEO): GSE000000. Processed data and accompanying code can be accessed through Synapse (syn27819129)

REFERENCES

1 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

2 Pardinas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*, doi:10.1038/s41588-018-0059-2 (2018).
3 Ripke, S., Walters, J. T. & O'Donovan, M. C. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *medRxiv*, 2020.2009.2012.20192922, doi:10.1101/2020.09.12.20192922 (2020).
4 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet* **50**, 621-629, doi:10.1038/s41588-018-0081-4 (2018).
5 Skene, N. G. *et al.* Genetic identification of brain cell types underlying schizophrenia. *Nat Genet* **50**, 825-833, doi:10.1038/s41588-018-0129-5 (2018).
6 Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci* **21**, 1117-1125, doi:10.1038/s41593-018-0197-y (2018).
7 Hall, L. S. *et al.* A Transcriptome Wide Association Study implicates specific pre- and post-synaptic abnormalities in Schizophrenia. 384560, doi:10.1101/384560 %J bioRxiv (2018).
8 Gulsuner, S. *et al.* Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518-529, doi:10.1016/j.cell.2013.06.049 (2013).
9 Schork, A. J. *et al.* A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. *Nat Neurosci* **22**, 353-361, doi:10.1038/s41593-018-0320-0 (2019).
10 Radulescu, E. *et al.* Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. *Mol Psychiatry*, doi:10.1038/s41380-018-0304-1 (2018).
11 Rajarajan, P. *et al.* Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* **362**, doi:10.1126/science.aat4311 (2018).
12 Li, J. *et al.* Spatiotemporal profile of postsynaptic interactomes integrates components of complex brain disorders. *Nat Neurosci* **20**, 1150-1161, doi:10.1038/nn.4594 (2017).
13 Jia, P., Chen, X., Fanous, A. H. & Zhao, Z. Convergent roles of de novo mutations and common variants in schizophrenia in tissue-specific and spatiotemporal co-expression network. *Translational psychiatry* **8**, 105, doi:10.1038/s41398-018-0154-2 (2018).
14 Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, doi:10.1126/science.aat8127 (2018).
15 Gandal, M. J. *et al.* Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693-697, doi:10.1126/science.aad6469 (2018).

16 Wainschtein, P. *et al.* Recovery of trait heritability from whole genome sequence data. *bioRxiv*, 588020, doi:10.1101/588020 (2019).
17 Nag, A., McCarthy, M. I. & Mahajan, A. Large-Scale Analyses Provide No Evidence For Gene-Gene Interactions Influencing Type 2 Diabetes Risk. *Diabetes*, doi:10.2337/db20-0224 (2020).
18 Liu, X., Li, Y. I. & Pritchard, J. K. Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. *Cell* **177**, 1022-1034 e1026, doi:10.1016/j.cell.2019.04.014 (2019).
19 Hess, J. L. *et al.* A polygenic resilience score moderates the genetic risk for schizophrenia. *Mol Psychiatry*, doi:10.1038/s41380-019-0463-8 (2019).
20 Yi, F. *et al.* Autism-associated SHANK3 haploinsufficiency causes Ih channelopathy in human neurons. *Science* **352**, aaf2669, doi:10.1126/science.aaf2669 (2016).
21 Pak, C. *et al.* Human Neuropsychiatric Disease Modeling using Conditional Deletion Reveals Synaptic Transmission Defects Caused by Heterozygous Mutations in NRXN1. *Cell stem cell* **17**, 316-328, doi:10.1016/j.stem.2015.07.017 (2015).
22 Deneault, E. *et al.* Complete Disruption of Autism-Susceptibility Genes by Gene Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. *Stem Cell Reports* **11**, 1211-1225, doi:10.1016/j.stemcr.2018.10.003 (2018).
23 Srikanth, P. *et al.* Genomic DISC1 Disruption in hiPSCs Alters Wnt Signaling and Neural Cell Fate. *Cell reports* **12**, 1414-1429, doi:10.1016/j.celrep.2015.07.061 (2015).
24 Cederquist, G. Y. *et al.* A Multiplex Human Pluripotent Stem Cell Platform Defines Molecular and Functional Subclasses of Autism-Related Genes. *Cell stem cell* **27**, 35-49 e36, doi:10.1016/j.stem.2020.06.004 (2020).
25 Paulsen, B. *et al.* Autism genes converge on asynchronous development of shared neuron classes. *Nature* **602**, 268-273, doi:10.1038/s41586-021-04358-6 (2022).
26 Jin, X. *et al.* In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, doi:10.1126/science.aaz6063 (2020).
27 Willsey, H. R. *et al.* Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilence. *Neuron* (2021).
28 Singh, T., Neale, B. M., Daly, M. J. & Consortium, o. b. o. t. S. E. M.-A. Exome sequencing identifies rare coding variants in 10 genes which confer substantial risk for schizophrenia. *medRxiv*, 2020.2009.2018.20192815, doi:10.1101/2020.09.18.20192815 (2020).
29 Schrode, N. *et al.* Synergistic effects of common schizophrenia risk variants. *Nat Genet* **51**, 1475-1485, doi:10.1038/s41588-019-0497-5 (2019).
30 Mimitou, E. P. *et al.* Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat Methods* **16**, 409-412, doi:10.1038/s41592-019-0392-0 (2019).

31 Zhang, Y. *et al.* Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785-798, doi:10.1016/j.neuron.2013.05.029 (2013).
32 Ho, S. M. *et al.* Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods* **101**, 113-124, doi:10.1016/j.ymeth.2015.11.019 (2016).
33 Roussos, P. *et al.* A role for noncoding variation in schizophrenia. *Cell reports* **9**, 1417-1429, doi:10.1016/j.celrep.2014.10.015 (2014).
34 Hauberg, M. E. *et al.* Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression. *Am J Hum Genet* **101**, 157, doi:10.1016/j.ajhg.2017.06.003 (2017).
35 Hauberg, M. E. *et al.* Differential activity of transcribed enhancers in the prefrontal cortex of 537 cases with schizophrenia and controls. *Mol Psychiatry*, doi:10.1038/s41380-018-0059-8 (2018).
36 Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am J Hum Genet* **102**, 1169-1184, doi:10.1016/j.ajhg.2018.04.011 (2018).
37 Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825, doi:10.1038/s41467-018-03621-1 (2018).
38 Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet* **50**, 538-548, doi:10.1038/s41588-018-0092-1 (2018).
39 Huckins, L. M. *et al.* Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. *Nat Genet* **51**, 659-674, doi:10.1038/s41588-019-0364-4 (2019).
40 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).
41 Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat Commun* **10**, 3834, doi:10.1038/s41467-019-11874-7 (2019).
42 Cao, C. *et al.* Power analysis of transcriptome-wide association study: Implications for practical protocol choice. *PLoS genetics* **17**, e1009405, doi:10.1371/journal.pgen.1009405 (2021).
43 Girdhar, K. *et al.* Acetylated Chromatin Domains Link Chromosomal Organization to Cell- and Circuit-level Dysfunction in Schizophrenia and Bipolar Disorder. *bioRxiv*, 2021.2006.2002.446728, doi:10.1101/2021.06.02.446728 (2021).
44 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).
45 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10**, e1004383, doi:10.1371/journal.pgen.1004383 (2014).

46 Dobbyn, A. *et al.* Co-localization of Conditional eQTL and GWAS Signatures in Schizophrenia. *bioRxiv*, doi:10.1101/129429 (2017).
47 Chavez, A. *et al.* Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326-328, doi:10.1038/nmeth.3312 (2015).
48 Chavez, A. *et al.* Comparison of Cas9 activators in multiple species. *Nat Methods* **13**, 563-567, doi:10.1038/nmeth.3871 (2016).
49 Ho, S. M. *et al.* Evaluating Synthetic Activation and Repression of Neuropsychiatric-Related Genes in hiPSC-Derived NPCs, Neurons, and Astrocytes. *Stem Cell Reports*, doi:10.1016/j.stemcr.2017.06.012 (2017).
50 Wray, N. R., Wijmenga, C., Sullivan, P. F., Yang, J. & Visscher, P. M. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* **173**, 1573-1580, doi:10.1016/j.cell.2018.05.051 (2018).
51 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186, doi:10.1016/j.cell.2017.05.038 (2017).
52 Breen, M. S. *et al.* Modeling gene x environment interactions in PTSD using glucocorticoid-induced transcriptomics in human neurons. *bioRxiv*, 2021.2003.2001.433391, doi:10.1101/2021.03.01.433391 (2021).
53 Sieberts, S. K. *et al.* Large eQTL meta-analysis reveals differing patterns between cerebral cortical and cerebellar brain regions. *Sci Data* **7**, 340, doi:10.1038/s41597-020-00642-8 (2020).
54 Bryois, J. *et al.* Cell-type specific cis-eQTLs in eight brain cell-types identifies novel risk genes for human brain disorders. *medRxiv*, 2021.2010.2009.21264604, doi:10.1101/2021.10.09.21264604 (2021).
55 Walker, R. L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771 e722, doi:10.1016/j.cell.2019.09.021 (2019).
56 Dobrindt, K. *et al.* Publicly Available hiPSC Lines with Extreme Polygenic Risk Scores for Modeling Schizophrenia. *Complex Psychiatry* **6**, 68–82, doi:10.1159/000512716 (2020).
57 Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* **180**, 568-584 e523, doi:10.1016/j.cell.2019.12.036 (2020).
58 Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-1198, doi:10.1073/pnas.1119675109 (2012).
59 McMahon, F. J. & Insel, T. R. Pharmacogenomics and personalized medicine in neuropsychiatry. *Neuron* **74**, 773-776, doi:10.1016/j.neuron.2012.05.004 (2012).
60 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

61 Fullard, J. F. *et al.* Single-nucleus transcriptome analysis of human brain immune response in patients with severe COVID-19. *Genome Med* **13**, 118, doi:10.1186/s13073-021-00933-8 (2021).
62 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
63 Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome biology* **19**, 224, doi:10.1186/s13059-018-1603-1 (2018).
64 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).
65 Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853-1866 e1817, doi:10.1016/j.cell.2016.11.038 (2016).
66 Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821, doi:10.1016/j.cell.2016.11.048 (2016).
67 Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297-301, doi:10.1038/nmeth.4177 (2017).
68 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411, doi:10.1038/nbt.4096 (2018).
69 Papalexi, E. *et al.* Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nature Genetics* **53**, 322-331, doi:10.1038/s41588-021-00778-2 (2021).
70 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196, doi:10.1126/science.aad0501 (2016).
71 Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron*, doi:10.1016/j.neuron.2019.07.014 (2019).
72 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
73 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
74 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).
75 Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust Hyperparameter Estimation Protects against Hypervariable Genes and Improves Power to Detect Differential Expression. *Ann Appl Stat* **10**, 946-963, doi:10.1214/16-AOAS920 (2016).

76 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
77 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
78 Gao, C., McDowell, I. C., Zhao, S., Brown, C. D. & Engelhardt, B. E. Context Specific and Differential Gene Co-expression Networks via Bayesian Biclustering. *PLoS Comput Biol* **12**, e1004791, doi:10.1371/journal.pcbi.1004791 (2016).
79 Gao C, B. C., Engelhardt BE. A latent factor model with a mixture of sparse and dense factors to model gene expression data with confounding effects. *arXiv* (2013).
80 Saha, A. *et al.* Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res* **27**, 1843-1858, doi:10.1101/gr.216721.116 (2017).
81 Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* **8**, 1826, doi:10.1038/s41467-017-01261-5 (2017).
82 Wang, J. & Liao, Y. *WebGestaltR: Gene Set Analysis Toolkit WebGestaltR. R package version 0.4.3.*, <https://CRAN.R-project.org/package=WebGestaltR> (2020).

Table 1. EpiXcan prioritization of PGC3 SCZ-GWAS genes. Genes annotated as epigenetic/regulatory (blue) and signaling (red) for pathway studies.

GENE	Annotated Function	Rank	z-score	ATACseq annotation	H3K27ac annotation	H3K4me3 annotation	H3K27ac annotation (neuron)	gRNAs
NEK4	serine/threonine protein kinase	1	8.64	NA	Promoter	NA	Promoter	10
PLCL1	inositol phospholipid signaling	2	7.20	NA	Distal Intergenic	Promoter	Intron	10
UBE2Q2L	pseudogene	3	6.89	NA	Promoter	Promoter	Promoter	10
NAGA	lysosomal enzyme	4	6.67	NA	Promoter	Promoter	NA	10
FES	tyrosine-protein kinase	5	6.63	NA	Promoter	Promoter	NA	10
CALN1	calcium signaling	6	6.57	NA	Intron	Promoter	Intron	10
CSPG4P11	pseudogene	pseudogene	6.29	NA	Promoter	NA	3' UTR	
ZNF804A	transcription factor	7	6.27	NA	Promoter	Intron	NA	10

Table 2. SCZ-GWAS loci with strong evidence for eQTL mechanisms. PGC3 SCZ-GWAS genes prioritized as epigenetic/regulatory (blue) and signalling (red) for pathway studies.

GWAS						COLOC	PrediXcan		hiPSC-neuron expression		
Chr	**SNP**	**P**	**LD range**	**Gene**	**Annotated Function**	**PP4**	**z**	**P**	**iGLU (RPKM)**	**iGABA (FPKM)**	**iDOPA (FPKM)**
19	rs72986630	3.07E-10	10832225…12839819	***ZNF823***	epigenetic	0.999253	6.350252	2.15E-10	4.0	2.2	5.4
6	rs113113059	2.29E-11	42150013…44192158	***CUL9***	ubiquitination	0.953677	5.754229	8.70E-09	4.0	9.5	23.2
7	rs2944821	1.90E-09	70244499…72912040	***CALN1***	calcium signaling	0.913966	5.547481	2.90E-08	0.9	3.1	0.9
16	rs3814883	8.82E-15	29007489…31016970	***INO80E***	epigenetic	0.894347	7.639276	2.18E-14	6.2	55.0	41.0
6	rs2022265	3.74E-10	83262613…85419243	***SNAP91***	synapse	0.893265	5.185282	2.16E-07	12.1	82.9	62.9
2	rs2914983	1.10E-14	197256700…199299078	***SF3B1***	epigenetic	0.888092	7.256793	3.96E-13	33.8	123.1	130.4
3	rs75968099	5.16E-11	36034966…38107017	*MLH1*	epigenetic	0.877109	6.376857	1.81E-10	11.9	22.7	24.9
16	rs3814883	8.82E-15	28952638…30984212	***TMEM219***	signaling	0.846736	6.291546	3.14E-10	13.5	62.0	60.4
16	rs3814883	8.82E-15	30016830…30034591	***DOC2A***	synapse	0.521296	6.088335	1.14E-09	2.7	20.9	9.3
4	rs61405217	5.39E-11	170533784…170644824	***CLCN3***	synapse	0.718835	5.771783	7.84E-09	18.6	33.7	33.4
2	rs1451488	6.72E-17	198669426…199437305	***PLCL1***	synapse	0.035058	4.910285	9.09E-07	4.4	3.5	6.2

Table 3: *Two-sample Kolmogorov-Smirnov test of Variance between CMC Cluster Node Genes*

		ClusterV1	**ClusterV2**	**ClusterV3**	**ClusterV4**	**ClusterV5**	**ClusterV6**	**ClusterV7**	**ClusterV8**
Cluster 1	*statistic*		0.067	0.000	0.067	0.067	0.067	0.000	0.133
	p.value		0.936	1.000	0.936	0.936	0.936	1.000	0.766
Cluster 3	*statistic*	0.204	0.184		0.082	0.224	0.133	0.173	0.224
	p.value	**0.017**	**0.037**		0.520	**0.007**	0.178	**0.052**	**0.007**
Cluster 4	*statistic*	0.288	0.188	0.025		0.138	0.100	0.075	0.125
	p.value	**0.001**	0.060	0.951		0.220	0.449	0.638	0.287
Cluster 7	*statistic*	0.216	0.140	0.002	0.002	0.108	0.039		0.099
	p.value	**4.37E-10**	**0.000**	0.998	0.998	**0.005**	0.497		**0.010**
Cluster8	*statistic*	0.154	0.089	0.006	0.000	0.071	0.006	0.006	
	p.value	**0.018**	**0.264**	0.994	1.000	0.427	0.994	0.994	

Alternative hypothesis: the Central Distribution Frequency of x lies below that of y

FIGURES

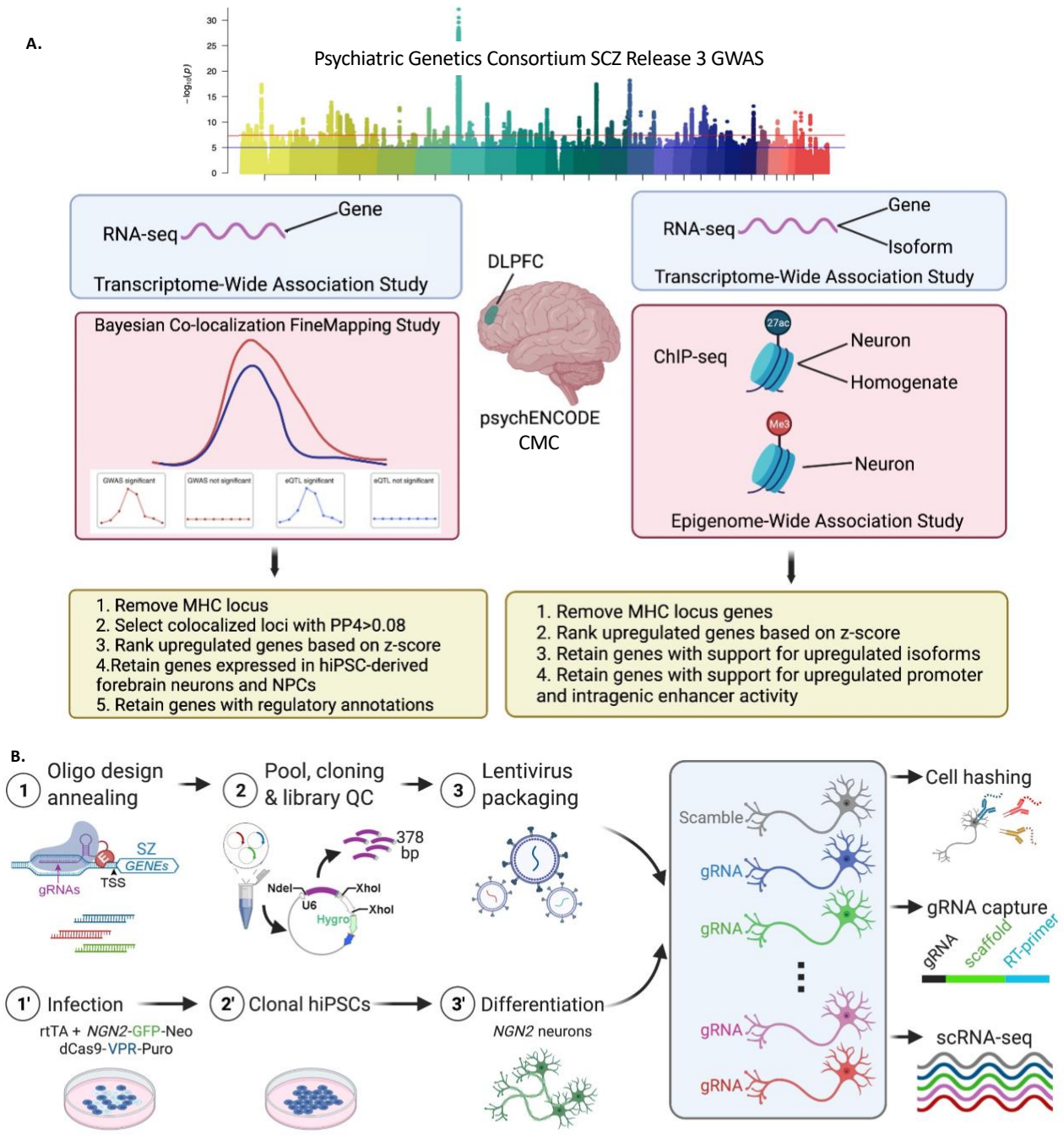


Figure 1. *ECCITE-seq (Expanded CRISPR-compatible Cellular Indexing of Transcriptomes and Epitopes by sequencing) allows us to perturb multiple genes in parallel and examine the transcriptional impact of each.* **(a)** Workflow for prioritization of SCZ eGenes from PGC3-GWAS by transcriptomic imputation, colocalization, and epigenome-wide associations. **(b)** Design of SCZ-GWAS gene CRISPRa gRNA library. First, designed oligos are anneals, pooled, and cloned into a plasmid backbone (1,2). Following cloning, plasmids are digested and ~ 378bp fragment contain gRNA is sequenced to determine the distribution of gRNAs in the pool (Supplementary Figure 1c-d). (3) The final library is packaged into a lenti-viral vector allowing for integration into cultured glutamatergic neurons. (1', 2') hiPSCs are first clonalized with a dox-inducible NGN2 transgene and the puromycin resistant CRISPR dCas9-VPR system. (3') Application of doxycycline induces glutamatergic neuronal development and antibody selection results in relatively pure populations of dCas9-expressing

glutamatergic neurons ready for the gRNA library. Created with BioRender.com. Partial eQTL image modified from Barbeira et. al 2018

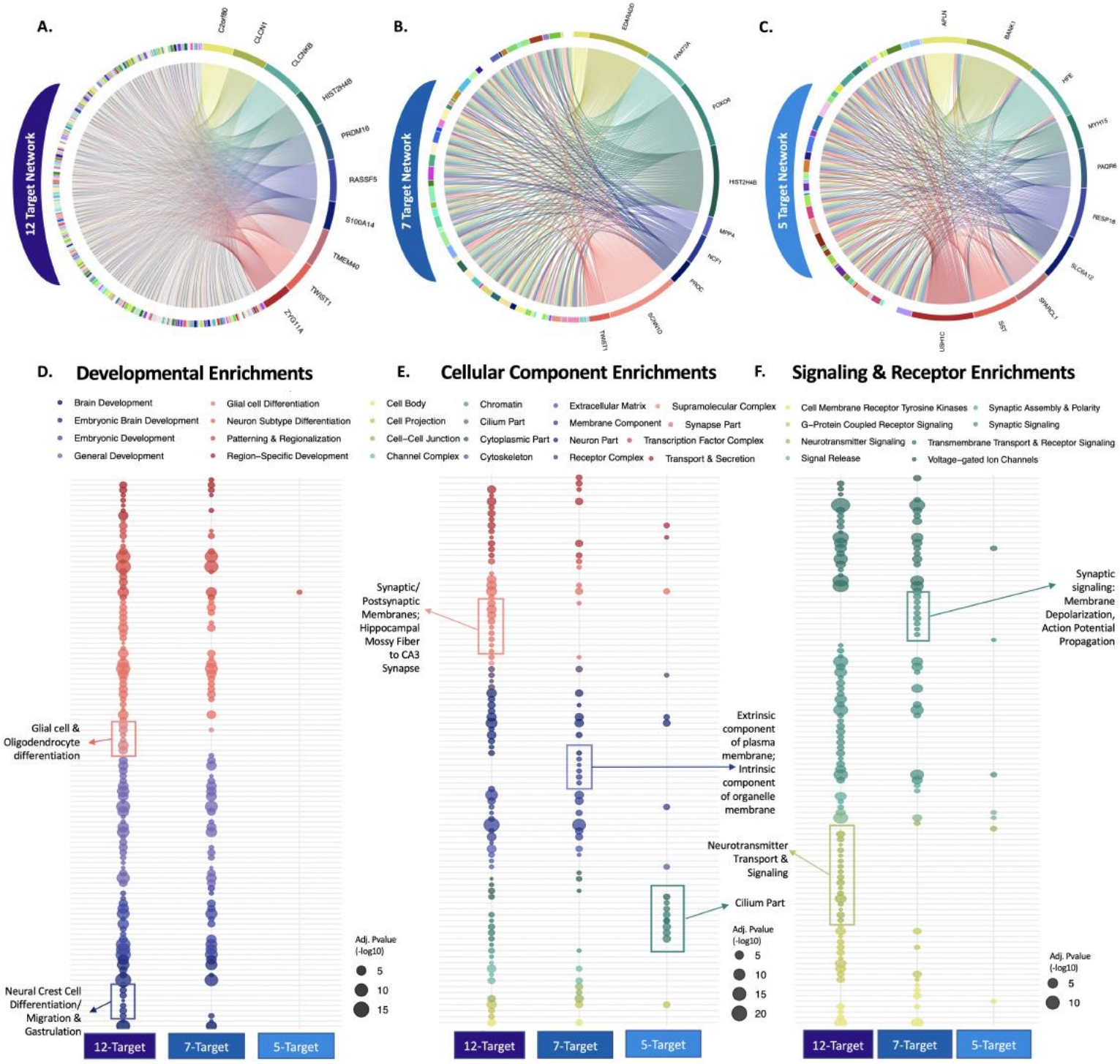


Figure 2. *Target-convergent networks identified by Bayesian Bi-clustering across 12, 7, and 5 SCZ risk gene perturbations show unique and shared pathway enrichments.* **(a-c)** Bayesian Bi-clustering across all samples identified biologically relevant clusters and recovered target-convergent high-polygenicity (12-Target; n=892), mid-polygenicity (7-Target; n=502), low-polygenicity (5-Target; n=390) networks. The top 10 node genes (based on the total number of edges) are represented to the right side of each chord diagram and their connections with all other node genes are represented as strings connecting on color to another. **(d)** The high-polygenicity (12-Target; n=892) and mid-polygenicity (7-Target; n=502) networks are significantly enriched for numerous brain and embryonic developmental traits, including

neuron subtype differentiation, patterning and regionalization, and brain-region specific development. Across all three networks, there are shared and unique enrichments in cellular component **(e)** and signaling and receptor activity **(f)** related gene ontology terms. G.O terms are color-coded by similarity in function, region, or structure and dot size represents -log10 of the adjusted p-value for each term.

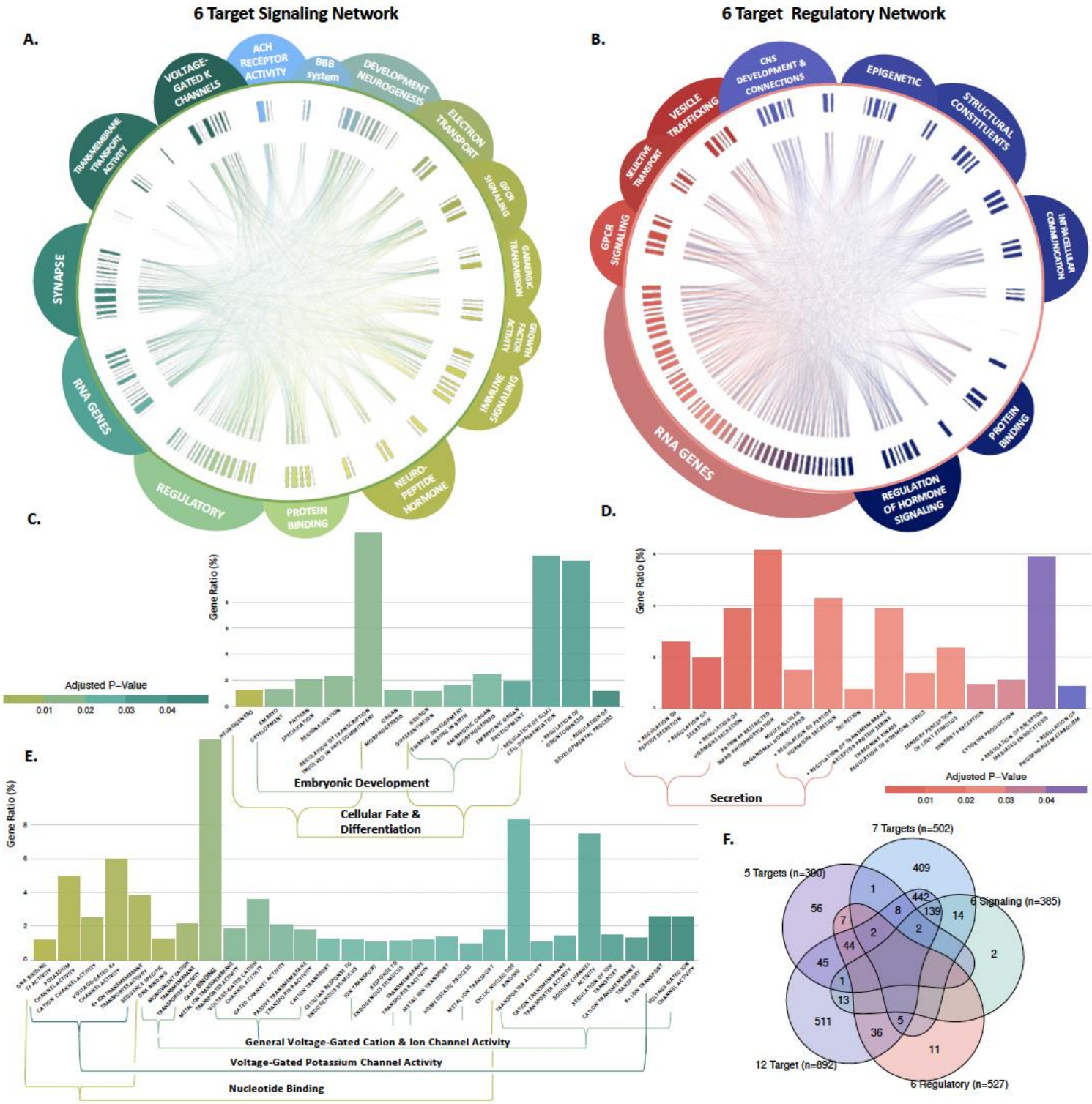


Figure 3. *Functional grouping of targets uncovers pathway-specific convergent perturbation signatures.* Tissue-convergent networks reconstruction of two subsets of 6 targets sharing similar functional annotations resolved unique convergence. The 6-target signaling network (n=385) genes are annotated for synaptic and signaling functions **(a)** and are enriched for neurogenesis and differentiation **(c)** and voltage-gated potassium channel activity **(d).** While the 6-target regulatory/epigenetic network (n=527)

genes largely annotated RNA genes potentially involved in transcriptional regulation (miRNA, lncRNA, lincRNAs) **(b)** and were significantly enriched for pathways of secretions **(e).** Overall, grouping perturbations by functional annotation had shared and unique gene membership and revealed greater specificity of convergence **(f).**

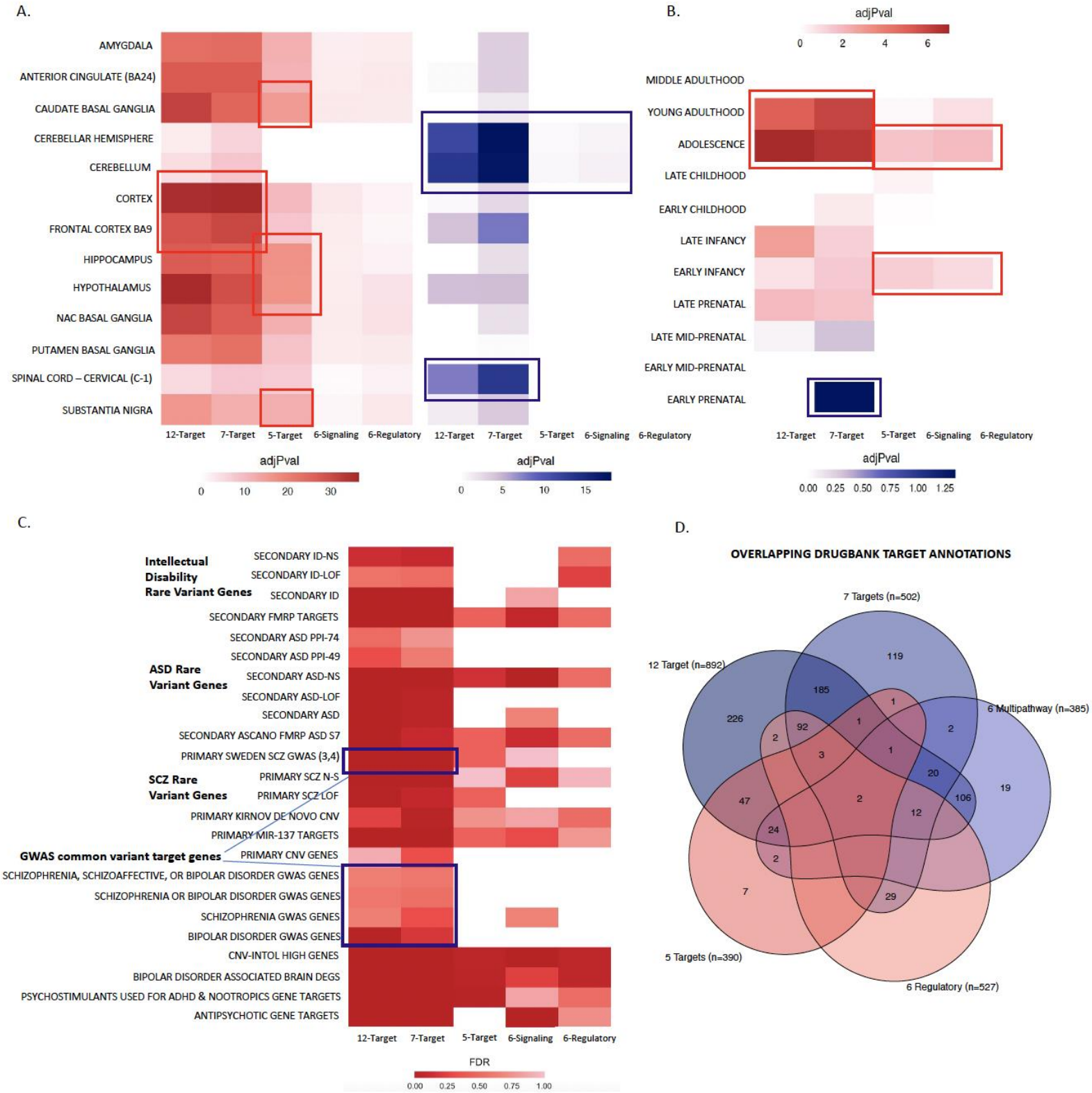


Figure 4. *Convergent networks have shared and unique enrichments regional & developmental Gene Expression and common and rare variant target genes across multiple disorders***. (a)** FUMA Enrichment Analysis across GTeX v8 Brain Region-specific Tissue expression revealed that genes shared across the 12, 7 and 5-Target networks were most strongly up-regulated in cortex. **(b)** FUMA Enrichment

Analysis across networks showed significant downregulation of genes in the 7-Target network during the “Early-Prenatal Period” and significant upregulation of genes in the 12 and 7-Target networks during “Adolescence” and “Young adulthood” based on BrainSpan average gene expression data across developmental time-points. **(c)** Over-representation analysis of convergent genes for a curated list of rare and common variant target genes across Autism Spectrum Disorder (ASD), Bipolar Disorder (BP), and Schizophrenia Spectrum Disorders (i.e schizoafffective/schizophrenia) from Schrode et al. 2019 reveal that high-polygenicity networks are enriched for SCZ, ASD, ID, and BIP common and rare variant genes, while low-polygenicity and functional networks are only enriched for rare variant genes. **(d)** Convergent networks have unique and shared DrugBank annotations for drugs targets.

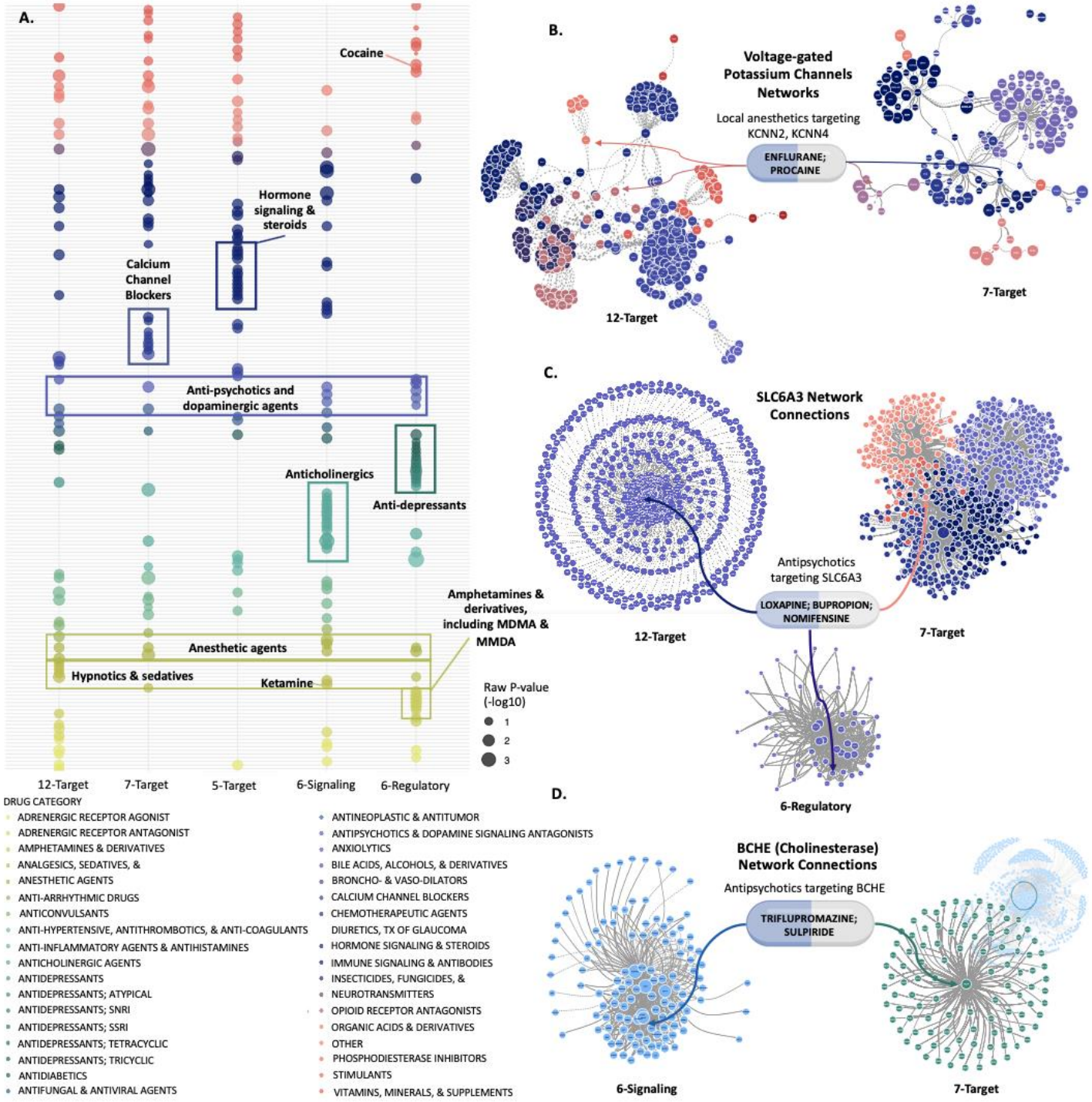
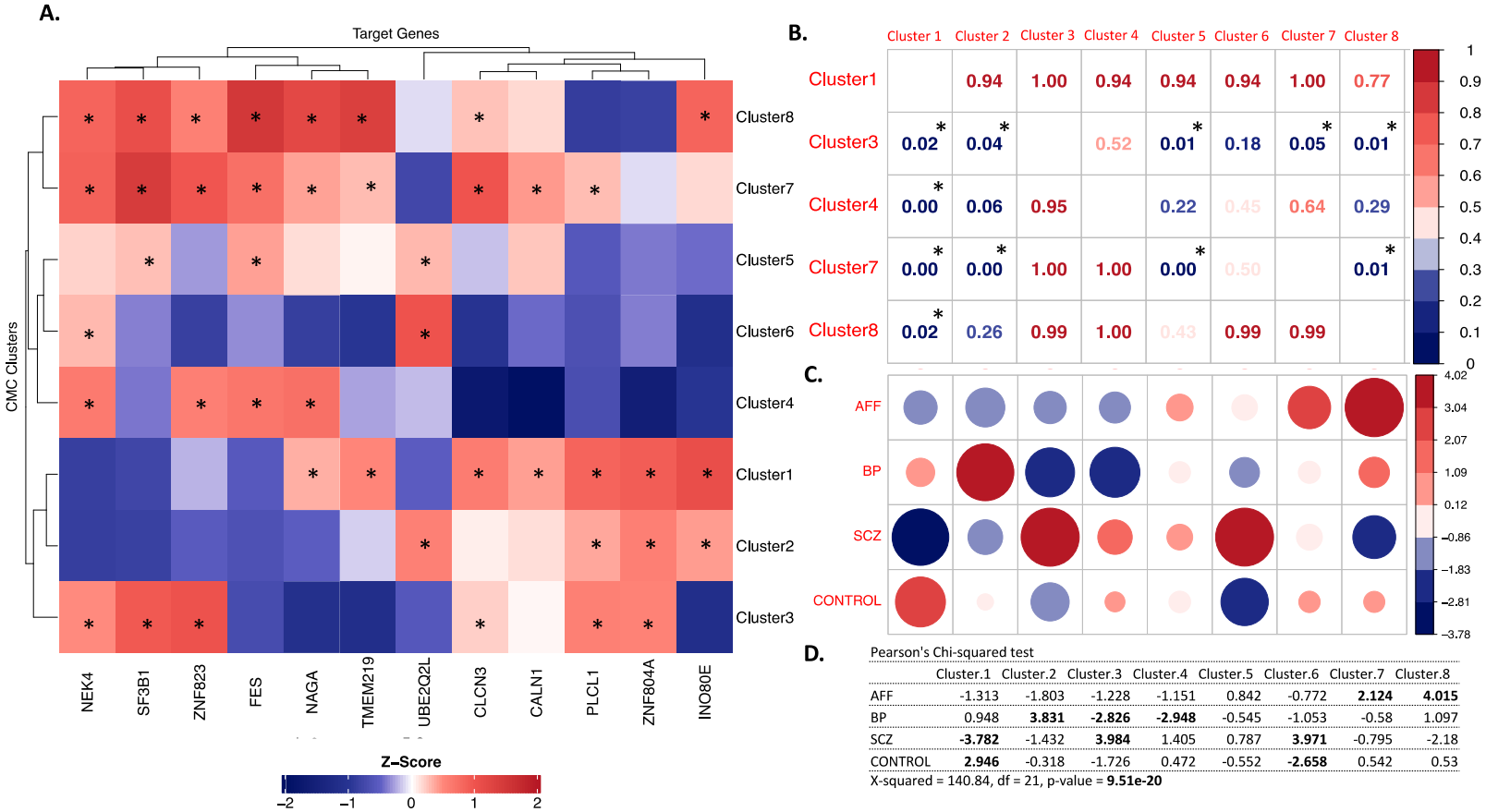


Figure 5. *Target-convergent sub-networks identified by Bayesian bi-clustering include signaling networks associated with neurological function with known drug targets including antipsychotics and antidepressants.* Across network membership, functional annotation identified 900 unique drugs and compounds targeting 168 genes. **(a)** Over representation analysis for known drug targets from the DrugBank dataset (y-axis) across convergent networks (x-axis) identified nominally significant enrichments in multiple medications prescribed to treat psychiatric and neurological symptoms. **(b)** Local anesthetics targeting potassium channels (KCN) share targets across the high- and mid-polygenicity KCN sub-networks. Specific antipsychotic target SLC6A3, a shared node between the high-/mid- polygenicity and regulatory networks **(c)**, while others target BCHE, a shared node in the mid-polygenicity and signaling networks **(d).**



	Cluster.1	Cluster.2	Cluster.3	Cluster.4	Cluster.5	Cluster.6	Cluster.7	Cluster.8
AFF	-1.313	-1.803	-1.228	-1.151	0.842	-0.772	**2.124**	**4.015**
BP	0.948	**3.831**	**-2.826**	**-2.948**	-0.545	-1.053	-0.58	1.097
SCZ	**-3.782**	-1.432	**3.984**	1.405	0.787	**3.971**	-0.795	-2.18
CONTROL	**2.946**	-0.318	-1.726	0.472	-0.552	**-2.658**	0.542	0.53

X-squared = 140.84, df = 21, p-value = **9.51e-20**

Figure 6: *Replication of convergent signatures in the post-mortem dorsa-lateral pre-frontal cortex. In vivo* target network reconstruction matched to perturbation assignments of individuals in the Common Mind Consortium (CMC; n=991) identified node genes with convergent patterns in post-mortem dorsa-lateral pre-frontal cortex (DLPFC). **(a)** Heatmap of average Z-scored gene expression of the 12 ECCITE-seq target genes (columns) across 8 CMC clusters identified by K-means clustering (rows) [***** represents assigned perturbation identities in each cluster (Z-scores of >=0.025)]. **(b)** Variance of node gene expression for CMC Clusters whose perturbation grouping resolved networks with at least 20% duplication rate (rows) in the scRNA-seq data were compared to the variance of the same genes expressed in all other clusters (columns). P-values resulting from Kolmogrov-Smirnov tests of variance are present at each intersection and coded by color from most significant (blue) to least significant (red) (***** indicates p-values <= 0.05). **(c)** Pearson's chi-squared test revealed significant differences in diagnosis status between clusters. The size (larger = more positive) and color (blue = negative, red=positive) of circles in the matrix represent the chi-squared residuals for diagnoses of Affective Disorders (AFF), Bipolar Disorder (BP), and Schizophrenia (SCZ) (rows) for each cluster (columns). **(d)** Table of Chi-squared residuals plotted in c.