TITLE:

Convergent and non-additive impact of schizophrenia risk genes in human neurons

AUTHORS:

PJ Michael Deans*[1], Kayla G. Townsley*[2,3,4], Aiqun Li**[2.3], Carina Seah**[2,3,4], Jessica Johnson[2]***, Judit Garcia Gonzalez[2]***, Evan Cao[2], Nadine Schrode[2,3], Alex Yu[2,3], Sam Cartwright[2,3], Georgios Voloudakis[2,3,4,5], Wen Zhang[2,3,4], Minghui Wang[2,3], John F. Fullard[2,3,4,5], Kiran Girdhar[2,3,4,5], Eli Stahl[2,3], Schahram Akbarian[2,3,4], Bin Zhang[2,3], Panos Roussos[2,3,4,5,6,7], Paul O'Reilly[2,4], Laura M. Huckins[1,2,3,8#], Kristen J. Brennand[1,2,3,4#]

AFFILIATIONS

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

[2] 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

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

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

[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

*, ** and ***: These authors contributed equally to this work.

Correspondence: kristen.brennand@yale.edu and laura.huckins@yale.edu

ABSTRACT

Genetic studies of schizophrenia reveal a complex polygenic risk architecture comprised of hundreds of risk variants; most are common in the population at-large, non-coding, and act by genetically regulating the expression of one or more gene targets ("eGenes"). It remains unclear how genetic variants predicted to confer individually small effects combine to yield substantial clinical impacts in aggregate. Here, we demonstrate that eGenes have shared downstream transcriptomic effects ("convergence") that may underlie unexpected interactions ("non-additive effects") observed when eGenes are manipulated in combination. We apply a pooled CRISPR approach to perturb schizophrenia eGenes in human induced pluripotent stem cell-derived glutamatergic neurons. The strength and specificity of convergence increased between functionally similar eGenes. Predicting that convergence might impact additive relationships between risk loci when inherited together, we use an arrayed approach to explore bidirectional combinatorial perturbations of a partially overlapping set of fifteen schizophrenia eGenes. When specifically considering groups of synaptic or epigenetic eGenes, combinatorial eGene perturbations yield changes that are smaller than predicted by summing individual eGene effects ("sub-additive effects"). Moreover, convergent and non-additive downstream transcriptomic effects overlap, suggesting that functional redundancy of eGenes may be a major mechanism underlying non-additivity. Combinatorial perturbations result in outcomes that are not yet well-predicted by single eGene perturbations alone, indicating that the effects of polygenic risk cannot necessarily be extrapolated from experiments testing one risk gene at a time.

KEYWORDS

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

INTRODUCTION

The genetic architecture of schizophrenia is complex and polygenic. Highly penetrant rare mutations underlie only a fraction of cases[1]. Rather, genome wide association studies (GWAS) indicate that schizophrenia is predominantly associated with genetic variation that is common in the population[2]. These risk loci have small effect sizes, are typically found in non-coding regions, and regulate the expression of one or more genes[3-5]. Mapping GWAS loci to their target genes (termed "eGenes", as defined by significant genetic regulation of expression) remains challenging, but is informed by expression quantitative trait loci (eQTL)[6-9], chromatin accessibility[10-12], enhancers[13-16], and 3D chromatin architecture[17-22]. The regulatory activity of risk loci can be empirically evaluated using massively parallel reporter assays[23-25] and pooled CRISPR screens[26], and causal gene targets and functions definitively resolved by genetic engineering in human induced pluripotent stem cells (hiPSCs)[10,11,17,27,28].

Schizophrenia eGenes are particularly expressed during fetal cortical development[29-31] and in glutamatergic neurons (as well as medium spiny neurons, and certain interneurons)[31-33]. They are highly co-expressed in human brain tissue[34] and cultured

neurons[17], show high connectivity in protein-protein interaction networks[17,35-37], and are enriched for roles in synaptic function and gene regulation[2,17,38-42]. Likewise, transcriptomic studies of schizophrenia post-mortem brains also identify aberrant expression of genes associated with synaptic function and chromatin dynamics in neurons[43-45]. The mechanism by which hundreds of distinct eGenes lead to shared molecular pathology is unknown.

We predicted that eGenes linked to schizophrenia would share substantial downstream transcriptomic changes with a common direction of effect (termed "convergence"). Although convergence has already been described in the context of loss-of-function autism spectrum disorder risk genes[46-56], these rare mutations almost never co-occur in the same individual. The convergent impact of common variants – which are frequently inherited together, and the impacts of which are apparent only in aggregate – remain unknown. We targeted twenty schizophrenia eGenes in iGLUTs using a pooled CRISPRa approach, successfully perturbed ten, and resolved their convergent impact (*CALN1, CLCN3, FES, NAGA, PLCL1, TMEM219*, *SF3B1, UBE2Q2L, ZNF823, ZNF804A*). To test if convergence impacts additive relationships between eGenes inherited in combination (i.e. if eGene effects sum linearly according to the additive model[26]), we manipulated eGenes one at a time and together in biological groups defined by known roles at the synapse ("synaptic": *SNAP91, CLCN3, PLCL1, DOC2A, SNCA*), or regulating transcription ("regulatory": *ZNF823, INO80E, SF3B1, THOC7, GATAD2A*), or with un-related non-synaptic, non-regulatory biology ("multi-function": *CALN1, CUL9, TMEM219, PCCB, FURIN),* and random combinations thereof. Combinatorial eGene perturbations within biological functions resulted in transcriptomic changes that were predominantly smaller than predicted by the additive model ("sub-additive"); sub-additive genes overlapped with convergent genes of the same eGenes.

Our work begins to address the long-standing question of how risk variants with individually minuscule effects combine to yield substantial impacts on neuronal biology. Overall, convergent signatures were robust, observable across three partially overlapping lists of schizophrenia eGenes, whether manipulated in arrayed or pooled approaches. Altogether, our work suggests that combinatorial interactions between eGenes[57,58] should be empirically measured rather than predicted from single eGene studies.

RESULTS

Convergence of downstream transcriptomic impacts across schizophrenia eGene perturbations.

We[27,59-61] and others[11,62-68] demonstrated that iGLUTs are >95% glutamatergic neurons, robustly express glutamatergic genes, release neurotransmitters, produce spontaneous synaptic activity, and recapitulate the impact of psychiatric trait associated genes. iGLUTs express most schizophrenia eGenes, including all eGenes prioritized herein[27].

eGenes whose brain expression was predicted to be up-regulated by GWAS loci[2] were prioritized for a pooled CRISPR activation (CRISPRa) experiment, which are currently restricted to one direction of effect. eGenes that were non-coding, located in the MHC

locus, or poorly expressed in iGLUTs were excluded. First, transcriptome and epigenome imputation (EpiXcan[69]) of schizophrenia GWAS[2] risk loci from post-mortem brain[43,70] prioritized seven schizophrenia eGenes (**SCZ1**: *NEK4, PLCL1, UBE2Q2L, NAGA, FES, CALN1,* and *ZNF804A*) (**Table 1A; Fig. 1A**). Second, transcriptomic imputation (prediXcan[71-73], $p<6x10^{-6}$) of SCZ GWAS[2] identified ~250 eGenes (**SI Table 1)**, subsequently narrowed by considering colocalization (COLOC[74,75], PP4 > 0.8) between schizophrenia GWAS[2] and post-mortem brain expression quantitative loci (eQTL) peaks[6], which identified 25 eGenes (**SI Table 1)**. 22 eGenes overlapped between approaches, ten of which were coding genes associated with increased expression in schizophrenia (**SCZ2**: *CALN1*, *CLCN3, CUL9, DOC2A, PLCL1, INO80E, SF3B1, SNAP91, TMEM219, ZNF823)* (**Table 1B; Fig. 1A**). Of note, our eGene selection, derived in bulk post-mortem brain, is largely preserved using an excitatory neuron-specific PrediXcan analysis (ExN-PrediXcan, **Tables 1A and 1B**).

Pooled CRISPR screening combined single-cell RNA sequencing readouts and direct detection of sgRNAs[76]. Two independently designed, constructed, and validated pooled CRISPRa libraries (**SCZ1** and **SCZ2**) were transduced into iGLUTs from two donors in independent experiments at unique developmental time-points (DIV7 or DIV21, **SI Fig 1F**). Non-perturbed cells from both **SCZ1** and **SCZ2** demonstrated gene expression patterns which correlated with expression in the adult postmortem DLPFC in neurotypical controls (**SI Fig. 12A-C**) and cortical neurons. Specifically, these cells were most strongly correlated with fetal cells transitioning to neuronal fate, fetal excitatory neurons, and cortical adult neurons (**SI Fig. 13**). Pooled CRISPR screens may be confounded by non-cell autonomous effects, which is of particular concern in neurons, as maturation is necessarily activity-dependent and therefore requires normal functioning of nearby cells; here, this is partially mitigated by the large number of presumably wildtype neurons in the population expressing either a scramble gRNA or no detectable gRNA at all (>60% of all pooled cells, see **SI Fig. 4A**). There was no significant difference in the degree of variance in maturity of the cell population between experiments and imputed cell fractions were not correlated with perturbation status (**SI Fig. 2 and 3**). An unsupervised framework, Weighted Nearest Neighbor Analysis[77], assigned successful perturbations; in total, we resolved perturbations of six of seven SCZ1 eGenes (ten gRNAs each) and four of ten SCZ2 eGenes (three gRNAs each). For 5401 and 6352 cells, respectively, we identified 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, normalization, and removal of doublets (cells containing more than one sgRNA), an average of 316 cells per sgRNA were successfully perturbated (ranging from 93-552) for a total of 3640 perturbed cells and 210 scramble controls (**SI Fig. 4-7**). Upregulation of eGenes by CRISPRa ranged from 0.2 to 3 log2 fold-change (Log2FC), comparable to the predicted effect sizes [SCZ1 (0.08 to 0.35); SCZ2 (0.2 to 0.77)] and eGene expression changes (Log2FC range 0.3-5.2) in the post-mortem dorsolateral prefrontal cortex (**Fig. 1B, C**; **SI Table 2,3; Fig. 8**). Effects of different gRNAs targeting the same eGenes were highly concordant, even when the degree of perturbation varied (**SI Fig. 15**). Differentially expressed genes (DEGs, $p_{FDR}<0.05$) were enriched for neuroactive ligand-receptor interaction, protein processing in the endoplasmic reticulum, proteasome, and spliceosome Gene Ontology and KEGG Pathways terms (**SI Data 3**), suggesting that

diverse eGenes might impact similar neural processes and pathways.

We define "convergence" as the independent development of transcriptomic changes in the same direction resulting from all eGene perturbations. DEGs were meta-analyzed (using METAL[78], p< 1.92E-06), and "convergent" genes were defined as those with shared direction of effect across all eGene perturbations and with non-significant heterogeneity between eGenes (Cochran's heterogeneity Q-test $p_{Het} > 0.05$) (**Fig. 2**). Across all schizophrenia eGenes, 790 significantly down-regulated genes and 10 significantly up-regulated genes were identified (Bonferroni meta p-value<=0.05) (**SI Data 3**), enriched for brain development, neuronal morphology, signaling, and transcriptional regulation (**SI Data 3**).

To identify groups of genes with similar expression patterns across eGene perturbations we define "convergent networks" as relationships between genes that are co-regulated by shared biological mechanisms. Unsupervised Bayesian bi-clustering[78] and gene co-expression network reconstruction from the pooled CRISPRa single cell RNAseq (n=3850 cells, 16851 genes, donor/batch corrected and normalized to adjust for covariates such as cell heterogeneity) identified high-confidence co-expressed gene networks. Across the pooled single-cell experiments, 1048 protein-coding source node genes (>5 edges) were identified, with a total network membership of 1869 genes that clustered together in at least 20% of the runs (**Fig. 2A**, **SI Data 3**), and significant enrichments for gene targets of schizophrenia GWAS loci as well as transcription factors and miRNAs that regulate schizophrenia GWAS loci, such as *TCF4*[79,80] (**Fig. 2A, i-iii)**. The cross-target convergent network was enriched in biological pathways implicated in schizophrenia etiology (**SI Fig. 9**); over representation analysis revealed schizophrenia, bipolar disorder, intellectual disability, and autism spectrum disorder common and rare risk genes to be significantly over-represented in node genes shared across all eGene perturbations (**Fig. 2A, i; SI Data 3**).

To study the strength and composition of convergent networks, we define "network convergence" as the sum of the network connectivity score (i.e., networks with fewer nodes and more interconnectedness have increased convergence). We endeavored to identify the biological factors (e.g., number of eGenes, functional similarity of eGenes, and eGene co-expression) that influenced network convergence. eGene number tested the number of eGenes used to generate a convergent network. Functional similarity (i.e., the degree of shared biological functions amongst eGenes) was calculated two ways: Gene Ontology semantic similarity scoring (within biological pathway, cellular component, and molecular function)[81], and synaptic/signaling score (proportion of eGenes with annotated function as either "signaling" for pooled or "synaptic" for arrayed). The brain expression correlation was calculated as the strength of the correlation of eGene expression in the post-mortem dorsolateral prefrontal cortex[6] (see **Methods**, **SI Fig. 10**). Bayesian reconstruction[82] was performed across all random combinations of eGene perturbations from the pooled screen (1003 unique eGene-Convergent Network sets) and arrayed experiment described in the following section (32752 sets) and resolved distinct networks (**Fig. 3B,E**). Principal components analysis tested the effect of biological factors on the network convergence scores (**Fig. 3C-D, F-G**; **SI Fig. 10-11**). Only brain expression correlation and the proportion of synaptic/signaling genes were significantly positively correlated with network

convergence across all sets in both the pooled [brain expression correlation: Pearson's r=0.24, adj. p-value<0.001, signaling proportion: Pearson's r=0.14, adj. p-value<0.01, n=826] and arrayed experiments [brain expression correlation: Pearson's r=0.083, Bonferroni adjusted p-value<0.001, signaling proportion: Pearson's r=0.25, adj. p-value<0.001, n=16319] (**Fig. 3D,G**). The average expression of perturbed eGenes was positively correlated with network convergence but is only significantly associated in the arrayed screen (**Fig. 3D, G**). Finally, although **SCZ1** and **SCZ2** pooled CRISPR screens were generated from distinct differentiation timepoints, the proportion of eGene perturbations by experiment did not correlate with the degree of network convergence, indicating that we have adequately controlled for variation in neuronal maturation (**Fig. 3D**; Pearson's r=0.062, bonferroni p-value=1).

Combinatorial perturbation of schizophrenia eGenes results in sub-additive impacts on transcription that reflect convergence.

To test if convergence between eGenes indeed impacts additive relationships between risk variants, we manipulated eGenes in combination to better approximate the polygenic nature of schizophrenia. Given that genes implicated in synaptic biology and epigenetic/transcriptional regulation are enriched for the schizophrenia risk[2,17,38-42], we sought to generate three groups of eGenes, linked to synaptic biology, gene regulation, or neither (**Fig. 1A**, arrayed screen). Unconstrained by the unidirectionality of pooled CRISPR screens, we did not restrict our list to eGenes with a single direction of effect. From the 18 coding genes prioritized by the intersection of transcriptomic imputation and colocalization, eGenes were separated into discrete functional categories based on gene ontology annotations. Our final gene list included five synaptic genes (*SNAP91, CLCN3, PLCL1, DOC2A, SNCA*), five regulatory genes (*ZNF823, INO80E, SF3B1, THOC7, GATAD2A*), and five genes with non-synaptic, non-regulatory functions, termed "multi-function" (*CALN1, CUL9, TMEM219, PCCB, FURIN*) (**Table 1B; Fig. 1A**).

We applied an arrayed design (i.e., distinct conditions in each well) to manipulate schizophrenia eGenes alone and in combination, allowing us to capture cell autonomous and non-cell autonomous effects in a manner not possible in the pooled design (**Fig. 4, SI Table 4, SI Fig. 19**). Endogenous expression was increased and decreased (via CRISPRa and shRNAs, respectively) in the direction associated with schizophrenia risk. Three to five vectors per gene were tested in 7-day-old (D7) iGLUTs, identifying the single vector that best achieved the level of significant perturbation predicted by eQTL analyses as confirmed by qPCR (**SI Fig. 1E**). Each eGene was perturbed in 21-day-old (D21) iGLUTs for 72 hours (**Fig. 1D, SI Fig. 1F-G, SI Fig. 19A**, **SI Fig. 21A**), individually and jointly, including appropriate vector and scrambled controls, from two neurotypical donors with average polygenic risk scores (one experimental batch per donor). Three groups of five random genes, one group of ten random genes, and one group of all fifteen genes were also included. Significant ($p<0.05$) changes in eGene expression in iGLUTs were confirmed by RNAseq in 13/15 eGenes (**SI Fig. 1G**, **SI Fig. 19A**); we validated the magnitude and direction of experimental eGene perturbation relative to the dosage effects of the top predicted causal SNPs (e.g. eQTL effect size) and predicted eGene expression changes (**Fig. 1B, D; SI Tables 2,3,4**). Across donors, donor status did not significantly impact the degree

of eGene perturbation (**SI Fig. 1H,** p=0.75, paired t test). Single perturbation of eGenes by CRISPRa ranged from 0.07 to 0.44 log2 fold change and RNAi ranged from -0.22 to -0.87 log2 fold change, comparable to EpiXcan effect sizes of 0.10 to 0.31 and -0.06 to -0.20 and PrediXcan effect sizes of 0.22 to 0.77 and -0.17 and -0.38 for corresponding eGenes.

Across the majority of the schizophrenia eGenes in our arrayed screen, competitive gene-set enrichment analysis using 698 manually curated neural[60] gene-sets (**SI Fig. 19B, C, SI Fig. 20A)** resulted in DEGs (p_{FDR}<0.05) that were strongly enriched for gene-sets related to rare and common psychiatric disorder risk genes (11/15) (**SI Fig. 20B)**, pre-synaptic biology (10/15) (**SI Fig. 20C**), and glutamatergic neurotransmission (10/15) (**SI Fig. 20D**).

Differential expression following simultaneous perturbation of all five eGenes within each of the synaptic, regulatory, multi-function group ("measured combinatorial"), or random combinations thereof, was compared to the sum of differential expression for each single eGene perturbation ("expected additive") (**Fig. 4; SI Fig. 21-23, Box 1; SI Table 4**). The difference between the measured combinatorial and expected additive model for each group represents the observed non-additive effects[83] (**Fig. 4A**). Most genes were expressed as predicted by the expected additive model (i.e., did not differ significantly from the model), but combinatorial perturbations of synaptic regulatory, and multi-function eGenes resulted in significant non-additive effects (Bayes moderated t-statistics, FDR p < 0.1) across 16.8%, 20.2%, and 0% of the total transcriptome (π1 synergy coefficients[83] of 43.86, 42.74, and 0.00, respectively) (**Fig. 4B,** non-additive p value distributions seen in **SI Fig 21E**). The log2FC of genes classified as "non-additive" was significantly larger than "additive" genes across all joint eGene perturbations (**SI Fig 23C**), and baseline expression of non-additive genes largely overlapped with those of additive genes (Log2(CPM) values within scramble control conditions, **SI Fig 23D**). Testing for overrepresentation in our curated neural gene sets showed that non-additive genes in the synaptic and regulatory joint eGene perturbations were significantly enriched for genes associated with risk for SZ as well as for pre- and postsynaptic gene sets (**Fig. 4C**). This was not explained by the magnitude of eGene perturbation between individual and combinatorial perturbations, which did not significantly differ (**SI Fig. 21B**, p>0.05 Wilcoxon ranked sum test). The degree of perturbation was strongly correlated with non-additivity for synaptic and regulatory eGenes, but not multifunctional eGenes – supporting that the type of eGene perturbed and not the magnitude of the perturbation alone drives non-additive effects (**SI Fig. 16A)**. Additionally, the strength of these correlations points to specific eGenes that may drive non-additive effects. For example, log2FC of *CLCN3* (synaptic) and *INO80E* (regulatory) are the most correlated with synergy coefficients (**SI Fig. 16C**). When evaluated across all eGene sets, the proportion of synaptic (Pearson's r=0.49) and regulatory (r=0.45) eGenes in a set positively correlated with non-additivity, while proportion of multifunctional eGenes was strongly negatively correlated (r=-0.94). Additionally, higher semantic similarity scores based on molecular function (r=0.6), but not cellular component (r=0) or biological function (r=-0.05) positively correlated with non-additivity (**SI Fig. 16B**), overall indicating that the functional role of a given eGene strongly impacts interacting effects on transcription.

In addition to the impact of function on non-additivity, we also examined whether increasing the number of perturbed eGenes in a joint perturbation contributed to non-additive effects. Among randomised eGene sets, the size of the observed non-additive transcriptional effect increased directly with the total number of eGenes simultaneously perturbed, indicating that increasing the number of eGenes perturbed increases the degree of interactive effects on transcription (compare joint perturbations of random subsets of 5, 10 and 15 eGenes**, Fig. 4D, SI Fig 23A**).

Most (>95%) non-additive genes (whether up- or down-regulated, FDR $p < 0.1$) had "sub-additive" effects, defined as having less differential expression than predicted by the additive model (i.e., changes that were "less up" or "less down" than expected) (**SI Fig. 21C, Box 1**). Sub-additive genes were enriched for synaptic and brain disorder-associated gene-sets (**SI Data 1**).

Sub-additive effects do not reflect lack of combinatorial perturbation at the single cell level; observed non-additive effects were similar whether tested from a single multiplexed vector expressing all gRNAs[27], or from independent expression vectors (**SI Fig. 22A-F**). Likewise, the strength and specificity to the non-additive effects resulting from pooled gRNA vectors were accurately recapitulated using a polycistronic gRNA vector (**SI Fig. 22G,H**), with 95% of non-additive effects validated. Notably, non-additive effects were more sensitively resolved using polycistronic vectors. Finally, modified ECCITE-seq confirmed a high number of unique gRNA integrations at the single cell level (**SI Fig. 22I)**.

To test the hypothesis that sub-additive effects arose from shared downstream signatures of individual eGene perturbations, we tested for convergent transcriptomic effects within each functional group (using METAL[78], $p< 1.92E-06$). There was robust within-group convergence for the synaptic (1070 genes) and regulatory (1070 genes) eGene groups, but limited convergence across the multi-function eGenes (71 genes) (**Fig. 5B-E**). Most of the convergent genes overlapped with the non-additive genes (Fisher's exact test, $p<2.2E-16$ for both synaptic and regulatory eGene groups), with 71% (761 of 1070) and 94% (1000 of 1070) of the convergent genes downstream of synaptic and regulatory eGenes, respectively, being including in their respective non-additive gene lists (**Fig. 5C,D**). Across all eight combinations of eGenes studied here (synaptic, regulatory, multi-function and random combinations of five, ten and fifteen eGenes), convergence across individual perturbations is correlated with the degree of non-additive effect seen in the corresponding joint perturbation condition (**Fig. 5A**, Pearson's $r^2 = 0.6569$, $p=0.0147$). Convergent effects of synaptic eGenes were enriched for synaptic function (e.g., mGluR5 interactors, $p=1.64E-03$) and brain disorders (e.g., schizophrenia GWAS, $p=8.41E-05$) gene-sets (**Fig. 5F**). Furthermore, convergent effects of regulatory eGenes were also enriched for brain disorder gene-sets (e.g., bipolar disorder, $p=9.92E-06$) (**Fig. 5G)**. Taken together, these findings highlight converging effects of schizophrenia eGenes on specific downstream genes relating to synaptic function and brain disorders.

Pathway-specific polygenic risk scores (PRS) (PRSet[84]) were calculated from non-additive signatures from synaptic (4,306 genes in PRS; $R^2=0.0431$), regulatory (5,249 genes in PRS; $R^2=0.0419$) and all fifteen eGenes (4,988 genes in PRS, $R^2=0.0425$) (**SI Fig. 24A**). PRSet enrichment results explained the largest proportion of variance in

schizophrenia case-control status, accounted for almost half of the total genome-wide PRS (19,340 genes plus SNPs in regions outside gene annotations in genome-wide PRS, R^2= 0.0925) (**SI Fig. 24B**). Altogether, non-additive genes are enriched in GWAS signal in aggregation and harbor polygenic signal to the same degree as curated synaptic genes (**SI Fig. 24C-E**).

Bayesian network reconstruction[82] from the arrayed experiment (n=63 samples, 4/sgRNA or shRNA, 25487 genes, and normalized to adjust for covariates such as donor) revealed a high-confidence co-expressed gene network that replicated across a maximum number of iterations (**Fig. 2B**). A densely interconnected network of 255 genes clustered together in at least 10% of the runs (**Fig. 2B**). This convergent network was significantly enriched for biological pathways implicated in schizophrenia etiology; over representation analysis revealed that target genes of schizophrenia, intellectual disability, and autism spectrum disorder common and rare variants were significantly over-represented in the network (**Fig. 2B,i**; **SI Data 3**), as well as genes regulated by miRNAs and transcription factors implicated in schizophrenia etiology, such as *hsa-miR-124a*[85] and *NKX2*[86] (**Fig. 2B,ii-iii)**. Separation of schizophrenia eGenes based on either signaling (**SI Fig. 9A**) or regulatory (**SI Fig. 9B**) function resolved unique convergent networks with no overlap in node genes, suggesting that the functional similarity of schizophrenia eGenes affects downstream convergence. Each of these networks included neuropsychiatric risk genes as well as those annotated for synaptic and immune signaling function (**SI Data 3**).

Although only nine node genes overlapped between convergent networks derived from arrayed and pooled experiments (**Fig. 2D**), both networks shared significant enrichments for targets of key miRNAs and transcription factors (**Fig. 2C**) – many with strong evidence of regulating gene expression associated with schizophrenia. The overlap in enrichments for regulatory factors suggests that while these unique groupings of eGene perturbations resolve largely disparate co-expression networks, they share an underlying regulatory structure.

Prediction and empirical validation of drugs predicted to reverse convergent signatures in neurons.

We identified drugs predicted to manipulate top node genes[87]. Across all eGene perturbations, reversers of convergent node signatures were enriched for mechanisms previously associated with psychiatric disorders, including HDAC inhibitors[88] (normalized connectivity score (NCS)=-1.63; FDR adjusted pval<0.08), ATPase inhibitors[89] (NCS=-1.61; FDR<0.08), and sodium channel blockers[90] (NCS=-1.59; FDR<0.08). Mimickers of convergent node signatures were enriched for pathways associated with stress response, including glucocorticoid receptor agonists (NCS=1.66, FDR<0.08) and NFKB pathway inhibitors (NCS=1.60; FDR<0.2) (**SI Data 4**). Finding only nominally significant enrichments in non-neuronal cell lines suggests these may be neuron-specific drug responses.

Three drugs that opposed the transcription signatures of top convergent nodes specifically in neurons or neural progenitor cells (NPCs) were prioritized (see **Methods**, **Box 2**): anandamide (reverser of convergent network signature, NCS=-1.38, FDR=1 as well as *CALN1* signature alone, NCS=-1.23, FDR=0.16), simvastatin (NCS=-1.14,

FDR=1; *TMEM219*, NCS=-0.8823, FDR=0.26), and etomoxir (NCS=-1.35, FDR=1; *CALN1,* NCS=-1.42, FDR=0.04; *TMEM219*, NCS=-1.09, FDR=0.01) (**SI Data 4**). These drugs were tested *in vitro* for their ability to reverse, or oppose, the effects of paired schizophrenia eGene perturbations. iGLUT CRISPRa for eGenes was followed by treatment with matched reverser drugs (*CALN1*: anandamide and etomoxir; *TMEM219*: simvastatin and etomoxir). Downstream transcriptomic and phenotypic assays were assessed to resolve eGene-drug effects on neuronal molecular, morphological, and physiological phenotypes (**Fig. 6**, **SI Fig. 28-30**). All drugs reversed or suppressed the transcriptomic impact of the CRISPRa perturbation alone. Notably, etomoxir not only limited the transcriptomic impact of perturbations of *CALN1* and *TMEM219*, but also impacted cell morphology (**Fig. 6C-D**), altering neurite length (adj. pval=0.0162) in a direction significantly opposed by *CALN1* perturbation (adj. pval=0.008) (**Fig. 6A-B**) and nominally opposed by *TMEM219* perturbation (adj. pval=0.082). Thus, it is possible to pharmacologically target convergent effects, even when schizophrenia eGenes themselves are not druggable.

DISCUSSION

Perturbations of schizophrenia eGenes resulted in convergent and non-additive transcriptional effects between functionally similar genes. First, a pooled CRISPRa study of twenty eGenes revealed a convergent effect concentrated on neurodevelopmental pathways, synaptic transmission, and voltage-gated ion channel activity. Second, across an independent but overlapping set of fifteen eGenes, within-function combinatorial perturbations resulted in sub-additive impacts on transcriptional profiles of hundreds of genes. The overlap of convergent genes and sub-additive genes suggests that redundant effects of eGenes with shared biological functions leads to saturation of directional impacts on gene expression. Overall, convergent signatures were robust, detected in three partially overlapping lists of schizophrenia eGenes, whether manipulated in arrayed or pooled experimental designs, and regardless of whether iGLUTs shared a common donor, cell type of origin, or developmental time point.

Would any large set of genetic perturbations uncover convergence and non-additivity? The issue of “negative control” is an important but challenging scientific question, particularly given the lack of power in some disease GWAS and the challenges in accurately mapping GWAS loci to gene targets highlighted by reviewers, but is one that must be addressed moving forward. Our results argue that future studies should expand beyond studying the effects of GWAS loci and their target eGenes in isolation, testing risk variant effects across multiple independent genetic backgrounds and in combination. We note that combinatorial eGene manipulations resulted in phenotypic changes in neuronal morphology, synaptic density, and neuronal activity that drastically differed from individual eGene perturbations (**SI Fig. 1F, SI Fig. 25-27**), reinforcing that polygenic risk cannot be extrapolated from experiments that test one risk gene at a time. These unexpected combinatorial effects could reflect compensatory homeostatic responses, or merely technical differences in the developmental timing of eGene perturbations and/or the inclusion of human astrocytes between the transcriptomic and phenotypic experimental paradigms. There is a need to investigate additivity,

interactivity, and convergence across more neuronal and glia cell types and developmental stages.

Altogether, the experimental eGene perturbations approximated the magnitude and direction of predicted eGene effect associated with schizophrenia, and generally resulted in downstream gene expression changes related to synaptic biology and psychiatric disorder risk. Nonetheless, further gene set enrichment analysis using 493 inflammation and cell death gene-sets[47] revealed enrichments related to cell stress and neurodegenerative diseases across many perturbations (**SI Data 1**). This enrichment was not seemingly associated with viral burden, being present whether single, combinatorial, or multiplexed vectors were applied. Indeed, if our *in vitro* system, defined by repeated lentiviral transduction, antibiotic selection, eGene perturbation, and single cell dissociation, stresses human neurons more than accounted for by the scramble gRNA controls, this would represent a concern of relevance to all applications of pooled CRISPR single screening experiments in human neurons. However, neither high content imaging nor multi-electrode array analyses indicated decreased cell survival or a cessation of neuronal activity (**SI Fig. 26, 27**). Moreover, inflammation[91] and oxidative stress[92], and particularly fetal exposures to inflammation, stress, and hypoxia[93,94] are indeed associated with schizophrenia risk.

Mapping GWAS associations to eGene targets is challenging and can yield false positives. Our schizophrenia eGenes remain to be validated through newer methods (e.g., activity-by-contact model[95]) that do not rely on tissue-specific eQTL data. How well our three eQTL-based methods prioritized causal eGenes remains a critical question. In fact, there are frequent hotspots of multiple TWAS-associated genes in the same locus[72], with co-regulation known to underlie pleiotropic TWAS associations[96]. We note that three eGenes prioritized in this study were linked to a single SNP (rs3814883) found within disease-associated copy number variant at 16p11.2, a locus that harbors the greatest excess of psychiatric common polygenic influences[97]. While it is possible that one or more eGenes at this locus were misidentified, an alternative possibility is that a causal GWAS SNP may co-regulate multiple adjacent and distal genes at this loci through chromatin contacts. For example, schizophrenia GWAS SNP rs2027349 altered expression of multiple genes: *VPS45*, lncRNA *AC244033*.2 and a distal gene, *C1orf54*; indeed, combinatorial perturbation of these eGenes results in non-additive impacts on transcriptomic and cellular phenotypes[28].

Given the extent of polygenicity associated with schizophrenia, our conclusions are constrained by the small proportion of eGenes successfully manipulated relative to the total eGenes impacted by schizophrenia GWAS loci. Moreover, we selected for empirical evaluation only those schizophrenia eGenes with the very strongest evidence of genetically regulated gene expression. The generalizability of our observations to all schizophrenia eGenes is unclear, particularly if there are non-linear responses to gradual changes in gene dosage[98]. Thus, future investigation to test across larger gene sets, graded changes in expression[98], *in vivo* brain regions[9] and *in vitro* cell types[99], developmental timespans[100], drug/environmental contexts[101] and donor backgrounds[102] will inform the cell-type-specific and context-dependent nature of convergence and non-additivity. Of course, all of this must be considered within the caveat that *in vitro* CRISPR perturbations do not exactly recapitulate the possession of multiple genetic

variants in human cases and controls. Despite this, it is worth noting that the limited number of perturbations used in our combinatorial conditions is still broadly relevant to studies of common variant interactions. When analyzing the full dataset of 105 S-PrediXcan SCZ eGenes in the post-mortem adult DLPFC[6], a median of ten and a maximum of 37 eGenes had outlying expression in the direction of risk association per individual (**SI Fig. 17**). Across the twenty SCZ eGenes targeted in either the pooled or arrayed screens, a median of two and a maximum of eight eGenes had outlying expression in the direction of risk association per individual (**SI Fig. 18)**.

It is very challenging to extrapolate the extent to which our *in vitro* studies of CRISPR perturbations inform the physiological effects of polygenicity occurring in human cases and controls. Non-additive impacts remain difficult to assess using genomic approaches, despite the large number of methods that have been developed to improve detection[103]. Contrary to population-level additive models[104], our findings suggest that at the individual-level some risk variants may have non-additive effects. We attempted to resolve this through eGene imputation in GWAS using two approaches. First, transcriptomic imputation of brain eGene expression (see **Methods, SI Fig. 1A**) revealed a dose-dependent effect: schizophrenia case-control status ($p<0.01774$, $OR>1.10$) was best predicted when three or more eGenes were perturbed ($OR_{3\ eGenes} = 1.47$ vs. $OR_{1\ eGene} = 1.10$). Second, transcriptomic risk scores were calculated (see **Methods**) to test whether the aggregate effect of eGenes was additive: schizophrenia risk was better predicted from larger ($p<2.2 \times 10^{-16}$) (**SI Fig. 1C**) or more biologically diverse ($R=0.19$, $p<2.2\times10^{-16}$) (**SI Fig. 1D**) gene groups. We speculate that population-level schizophrenia risk increases with the number of genes and pathways impacted (rather than within groups of functionally similar eGenes) due to the lack of individuals, either case or control, with strong imputed within-function perturbations. Pathway-specific polygenic risk scores[84] that incorporate biological pathways, co-expression patterns, convergence, and/or non-additivity may improve patient stratification or better predict drug response.

How does our genetic analysis of convergent and combinatorial eGene effects inform precision medicine for patients with psychiatric disorders? How does our genetic analysis of convergent and combinatorial eGene effects inform precision medicine for patients with psychiatric disorders? First, convergence resolved across combinations of different SCZ eGenes may inform molecular subtypes of disease. For example, when we cluster individuals based on shared patterns of schizophrenia eGene up-regulation in the post-mortem DLPFC (**SI Data 5),** clusters were distinguished by the diagnosis of individuals within (Pearson's Chi-squared; $X^2=140$, df=21, p-value=9.51e-20) (e.g., cluster 8, SCZ, $X^2=3.286$; cluster 12, affective disorders (AFF), $X^2=5.57$; cluster 16, control, $X^2=3.014$) (**SI Fig. 31-32**). A diagnosis of affective disorder (AFF) was significantly associated with up-regulation of *FES*, *NAGA*, *CALN1*, *CLCN3*, *SF3B1* and *ZNF804A* (cluster 12). Convergence analysis across these 6 eGenes in iGLUTs identifies the central node gene *ABCG2*—a biomarker associated with increased negative symptoms[105], down-regulated in a neuroimmune molecular subtype (SCZ Type II)[106,107], and associated with SCZ treatment resistance[108]. Second, we not only predicted drugs capable of reversing convergent transcriptomic signatures, but also demonstrated opposing action of drug treatment and schizophrenia eGene perturbations *in vitro*. We highlight statins, particularly simvastatin, which crosses the

blood brain barrier and shows promise as an add-on treatment in schizophrenia[109]. Two double-blind placebo-controlled trials of simvastatin did not reduce overall symptom severity[110,111] but highlighted the possibility that simvastatin may decrease negative symptoms in some patients. Secondary analysis highlights that baseline inflammatory profiles[112] and treatment-induced changes in insulin receptor levels[113] may serve as biomarkers capable of predicting simvastatin response.

That convergent genes were enriched for risk factors and transcriptomic signatures associated with a range of other brain disorders indicated that convergent effects may in part explain shared features of psychiatric disorders and cross-disorder pleiotropy of risk. Consistent with this, common and rare risk variants for schizophrenia[2,17,38-42,114-116], autism spectrum disorder[117-119] and more broadly across the neuropsychiatric disorder spectrum[30,120-122] are all highly enriched for genes involved in synaptic biology and gene regulation, further highlighting the need to explore the impact of functional similarity in the interactions between rare and common variants. Our findings support the hypothesis that common and rare psychiatric risk variants converge on the same biological pathways[27]. 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 as-yet unidentified regulatory nodes that lead to the disruptions in synaptic function believed to underlie psychiatric endophenotypes represent exciting and unexplored therapeutic targets. Our overarching goal is to advance the field towards an era of precision medicine[123], 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

Schizophrenia eGene Prioritization: "eGenes: are defined as genes with significant genetic regulation of gene expression levels. In total, across the pooled and arrayed analyses, 20 unique eGenes were prioritized based on statistical and epigenetic evidence supporting genetic (dys)regulation of expression in schizophrenia (see**, Table 1**), rather than GWAS or eQTL effect size; predicted direction and magnitude of eGene effect available in **SI Table 1**.

i) "**SCZ1** eGenes": EpiXcan[69] was used to impute brain transcriptomes from Psychiatric Genomics Consortium 3 (PGC3)-SCZ GWAS[2] at the level of genes and isoforms from the PsychENCODE post-mortem datasets of genotyped individuals (brain homogenate, n=924)[43,70]; EpiXcan increases power to identify trait-associated genes under a causality model by integrating epigenetic annotation[124] (from REMC[125]); 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[73] method (using a 50kbp window instead of 1Mbp for transcripts) utilizing the recently described ChIPseq datasets[21]; summary-level imputation was performed with S-PrediXcan[71]. Peaks were assigned to genes with the ChIPseeker R package[126]. In addition, PrediXcan[73] imputed H3K27ac (brain homogenate, n=122; neuronal, n=191) and H3K4me3 (neuronal, n=163)[21] to more confidently identify *cis* regulatory elements

associated with risk for SCZ. Overall, SCZ eGenes were prioritized from GWAS based on: i) significant genetic up-regulation of expression (z-score >6 for genes), ii) epigenetic support (imputed epigenetic activity ($p<0.01$) across at least one of the three assays), iii) exclusion of non-coding genes or those located in the MHC locus, iv) robust expression in our hiPSC neuron RNAseq. Genes were ranked based on the association z-score for imputed gene expression. For pooled experiments (day 7 hiPSC-derived iGLUT), six top coding genes and one top pseudo-gene were selected: *NEK4, PLCL1, UBE2Q2L, NAGA, FES, CALN1,* and *ZNF804* (**Table 1A**).

ii) "**SCZ2** eGenes": First, transcriptomic imputation (prediXcan[71-73]) identified ~250 significant genes ($p<6x10^{-6}$) with predicted differential expression between SCZ-cases and controls using SCZ GWAS[2] and post-mortem CommonMind Consortium (CMC)[6] data (623 samples). Second, colocalization (COLOC[74,75]) of fine-mapped PGC3-GWAS[2] loci (65,205 cases and 87,919 controls) with post-mortem brain[6] eQTL (537 EUR samples)[6] identified 25 loci with very strong evidence (high posterior probability that a single shared variant is responsible for both signals, PP4 > 0.8[74]). There was significant overlap between the two analyses (binomial test p-value $3.03x10^{-112}$); of the 25 COLOC genes, 22 were also significant by PrediXcan. For each eGene, the magnitude and direction of perturbation associated with SCZ risk was predicted, and expression confirmed in hiPSC neuron RNAseq[27]. eGenes were further separated into discrete functional categories based on gene ontology annotations (http://geneontology.org/). From these 22, we prioritized the top coding genes across three broad categories: synaptic, regulatory, and multifunction (defined as not synaptic, regulatory, and seemingly unrelated to each other). To complete selection of five genes from each category, three additional top-ranked synaptic genes from the prediXcan analysis were included: *DOC2A*[75], *CLCN3*[75] and *PLCL1*[122]. Overall, 15 SCZ eGenes were prioritized from GWAS based on i) significant genetic regulation by COLOC and/or PrediXcan, ii) exclusion of non-coding genes and those located in the major histocompatibility complex (MHC) locus, iii) robust expression in our hiPSC neuron RNAseq.

For arrayed experiments (day 21 NPC-derived iGLUT), our final gene list for combinatorial perturbations included five synaptic genes (*SNAP91, CLCN3, PLCL1, DOC2A, SNCA*), five regulatory genes (*ZNF823, INO80E, SF3B1, THOC7, GATAD2A*), and five genes with non-synaptic, non-regulatory functions, termed "multi-function" (*CALN1, CUL9, TMEM219, PCCB, FURIN*) (**Table 1B**). For pooled experiments (day 21 NPC-derived iGLUT), the ten coding genes with significant genetic up-regulation were selected: *CALN1*, *CLCN3, CUL9, DOC2A, PLCL1, INO8E0, SF3B1, SNAP91, TMEM219, ZNF823*. This list was combined with our eGene set previously evaluated in hiPSC-neurons [27]; one functionally validated gRNA was included for each of these three genes (*SNAP91*, *TSNARE1*, and *CLCN3)*[27].

gRNA design: CRISPRa gRNA design and cloning were conducted as described previously[127]. For the fifteen eGenes prioritized by a combination of COLOC and PrediXcan, we designed three gRNAs each. For the seven eGenes prioritized by EpiXcan and PrediXcan, we designed ten gRNAs each. For the three previously tested eGenes[27] (intended a positive control), we used one pre-validated gRNA each.

iGLUT induction from hiPSC-derived NPCs[27,59-61] or hiPSCs[66,128].

Validated control hiPSCs for eGene perturbation were selected from a previously reported case/control hiPSC cohort of childhood onset schizophrenia[129]. Informed consent was obtained from all fibroblast donors at the National Institute of Mental Health, under the review of the Internal Review Board of the NIMH. All hiPSC work was reviewed by the Internal Review Board of the Icahn School of Medicine at Mount Sinai. This work was also reviewed by the Embryonic Stem Cell Research Oversight Committee at the Icahn School of Medicine at Mount Sinai and Yale University. The following control hiPSC/NPCs were used: NSB553-S1-1 (male), NSB2607-2/NSB2607-1-4 (male), NSB690-2 (male). All fibroblast samples were genotyped by IlluminaOmni 2.5 bead chip genotyping[130,131], PsychChip[129], and exome sequencing[129]. Parental hiPSCs were validated by G-banded karyotyping (Wicell Cytogenetics), with ongoing genome stability monitored by Infinium Global Screening Array v3.0 (Illumina). Critically, SNP genotype is inferred from all RNAseq data using the Sequenom SURESelect Clinical Research Exome (CRE) and Sure Select V5 SNP lists to confirm that neuron identity matches donor.

i) Validated control hiPSC-derived NPCs for CRISPRa/shRNA were selected from a previously reported case/control hiPSC cohort of childhood onset SCZ (COS)[129]: NSB553-S1-1 (male, average SCZ PRS, European ancestry), NSB2607-1-4 (male, average SCZ PRS, European ancestry). hiPSC-NPCs expressing dCas9-VPR were generated and validated as previously described)[129] and cultured in hNPC media (DMEM/F12 (Life Technologies #10565), 1x N2 (Life Technologies #17502-048), 1x B27-RA (Life Technologies #12587-010), 1x Antibiotic-Antimycotic, 20 ng/ml FGF2 (Life Technologies)) on Matrigel (Corning, #354230). hiPSC-NPCs at full confluence (1-1.5x10^7 cells / well of a 6-well plate) were dissociated with Accutase (Innovative Cell Technologies) for 5 mins, spun down (5 mins X 1000g), resuspended, and seeded onto Matrigel-coated plates at 3-5x10^6 cells / well. Media was replaced every two-to-three days for up to seven days until the next split.

At day -2, dCas9-VPR hiPSC-NPCs were seeded as 1.2x10^^{6} cells / well in a 12-well plate coated with Matrigel. At day -1, cells were transduced with rtTA (Addgene 20342) and *NGN2* (Addgene 99378) lentiviruses. Medium was switched to non-viral medium four hours post infection. At day 0 (D0), 1 µg/ml dox was added to induce *NGN2*-expression. At D1, transduced hiPSC-NPCs were treated with antibiotics to select for lentiviral integration (300 ng/ml puromycin for dCas9-effectors-Puro, 1 mg/ml G-418 for NGN2-Neo). At D3, NPC medium was switched to neuronal medium (Brainphys (Stemcell Technologies, #05790), 1x N2 (Life Technologies #17502-048), 1x B27-RA (Life Technologies #12587-010), 1 µg/ml Natural Mouse Laminin (Life Technologies), 20 ng/ml BDNF (Peprotech #450-02), 20 ng/ml GDNF (Peprotech #450-10), 500 µg/ml Dibutyryl cyclic-AMP (Sigma #D0627), 200 nM L-ascorbic acid (Sigma #A0278)) including 1 µg/ml Dox. 50% of the medium was replaced with fresh neuronal medium once every second day.

For pooled analysis, on day 5, young hiPSC-NPC *NGN2*-neurons were replated onto matrigel-coated plates and cells were dissociated 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 $5{\cdot}10^5$ per 12-well plate. For arrayed analysis, neurons were not replated, owing to the complexity of conditions involved.

At D13, iGLUTs were treated with 200 nM Ara-C to reduce the proliferation of non-neuronal cells in the culture, followed by half medium changes. At D18, Ara-C was completely withdrawn by full medium change while adding media containing individual shRNA/gRNA vectors or pools of mixed shRNA and gRNA vectors (Addgene 99374), either targeting eGenes or scramble controls. Medium was switched to non-viral medium four hours post infection. At D19, transduced iGLUTs were treated with corresponding antibiotics to the gRNA lentiviruses (1 mg/ml HygroB for lentiguide-Hygro/lentiguide-Hygro-mTagBFP2) followed by half medium changes until neurons were harvested at D21.

ii) Clonal hiPSCs from two control donors of European ancestry (NSB690-2 (male, average SCZ PRS, European ancestry) and NSB2607-2 (male, average SCZ PRS, European ancestry)[129] with lenti-EF1a-dCas9-VPR-Puro (Addgene #99373), pLV-TetO-hNGN2-eGFP-Neo (Addgene #99378), 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 $1.2{\cdot}10^6$ 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 $1{\cdot}10^6$ 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 $5{\cdot}10^5$ per 12-well plate. On day 7, neurons were harvested for scRNA sequencing.

Neuronal Pooled CRISPRa screens. Expanded CRISPR-compatible CITE-seq (ECCITE-seq)[76], combines Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq) and Cell Hashing for multiplexing and doublet detection[132] 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[133-135]. CRISPRa hiPSC iGLUT neurons (2607 (male) and 690 (male)) were transduced with the pooled gRNA at day -1. After maturation, 7-day-old iGLUT neurons were dissociated to single cell suspensions with papain, antibody-hashed[132], and bar-coded single cell cDNA generated using 10X Genomics Chromium[136]. NPC-derived iGLUT neurons (2607 (male) and 553 (male)) 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.

Analysis of single-cell CRISPRa screens in DIV 7 and DIV 21 iGLUT Neurons. 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[137] to perform downstream analyses, including QC, normalization, cell clustering, HTO/GDO demultiplexing, and DEG analysis[76,138].

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. 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. 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. In both experiments, 8-10% of retained cells contained multiple gRNAs and were assigned as doublets after de-multiplexing. To remove variation related to cell-cycle phase of individual cells, cell cycle scores were assigned using Seurat::CellCycleScoring which uses a list of cell cycle markers[139] 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, hemoglobulin, and ribosomal genes, batch, donor (HTO-maxID; as a biological replicate), cell cycle scores (Phase) 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, we performed ‘weighted-nearest neighbor’ (WNN) analysis, to assign clusters based on both guide-tag identity class and gene expression[77]. To identify successfully perturbed cells, we calculated a p-value based on the Wilcox rank sum test and Area Under the Curve (AUC) statistic, which reflects the power of each gene (or gRNA) to serve as a marker of a given cluster using Presto. WNN Clusters were then filtered based on two criteria (1) single gRNA-identity with an AUC statistic of >= 0.8 (where 1 means the gRNA is a perfect marker of a given cluster) and (2) a logFC >= 2 standard deviations of the mean or logFC > 0 and p-val > 0.05, of the target gene (but no other target genes) compared to scramble (non-targeting sgRNAs) controls (**SI Fig. 4-8**). These clusters were then used for downstream

analyses[140].

Cell Fraction Imputation and Quantification of Heterogeneity in Composition of iGLUT neurons. Using CiberSortx, we imputed the cell-faction identity of randomly sampled scramble control cells from each experiment (n=100/exp) using the PsychEncode scRNAseq dataset as a reference (100 permutations). To determine if the level of heterogeneity of iGLUT neuron maturity and subtype was similar between DIV7 and DIV21 iGLUT neurons in the given experiments, we performed a non-parametric Levene's Test for Homogeneity of Variance (LT-test) on the imputed cell fraction matrices. Although we observed heterogeneity in relative central and peripheral nervous system marker expression across the cell fractions, this heterogeneity was not due to gRNA identity and the level of variance in our data due to cellular heterogeneity was not significantly different by time-point. We were underpowered to compare gRNAs between cells with higher expression of different cell markers.

Meta-analysis of gene expression across perturbations. We performed a meta-analysis and Cochran's heterogeneity Q-test (METAL[78]) using the p-values and direction of effects (t-statistic), weighted according to sample size across all sets of perturbations in both the arrayed and pooled assays (Target vs. Scramble DEGs). Genes were defined as "convergent" if they (1) had the same direction of effect across all 5, 10, or 15 target combinations, (2) were Bonferroni significant in our meta-analysis (Bonferroni adjusted p-value <= 0.05), and (3) had a heterogeneity p-value = >0.05.

Bayesian Bi-clustering to identify Target-Convergent Networks. eGene-Convergent gene co-expression Networks (eGCN)[82] were built using an unsupervised Bayesian biclustering model, BicMix[141], on the log2CPM expression data from all the replicates across each of the 5-target sets and scramble gRNA jointly or all the cells across 10 targets and scramble gRNA jointly. for the arrayed and pooled assays respectively. To account for neuronal maturity differences in the single-cell screen, expression matrices were batch corrected and normalized and the scramble cells from both experiments (matched scramble gRNA across experiments) used as a single control population. To perform this as a joint analysis across two experiments, (1) Count matrices from each experiment were combined and RNA transcripts, mitochondrial, ribosomal, and hemoglobin genes were removed ('^MT-|^RP[SL][[:digit:]]|^RPLP[[:digit:]]|^RPSA|^HB[ABDEGMQZ][[:digit:]]') as well as genes that had at fewer than 2 read counts in 90% of samples, (2) and limma:voom normalization and transformation was used to compute the log2cpm counts from the effective library sizes of each cell (16851 genes). 40 runs of BicMix were performed on these data and the output from iteration 400 of the variational Expectation-Maximization algorithm was used. The hyperparameters for BicMix were set based on previous extensive simulation studies[142]. Convergent networks were identified across all possible combinations of 2-14 as well as all 15 of the targets (n=32752 combinations) in the arrayed assay, and all possible combinations of 2,3,4,5,6,7 or 8 as well as all 10 of the targets (n=1003 combinations) in the pooled screen. Network connections that did not replicate in more than 10% of the runs were excluded. Nodes with less than 5 edges or non-coding genes were removed from gene set enrichment analysis (GSEA). (The threshold of >5 edges is based on the likelihood of more than 5 edges being present by chance, with 10% being the percentage of runs where the connection was identified,

see[82,141]. Duplication thresholds are network-dependent and a metric of confidence in the connections; including those with especially low duplication rates were not included in downstream analysis.) Of all random sets tested in the pooled screen, 64.8% resolved a convergent network passing at least a 10% duplication threshold; of all random sets tested in the arrayed experiment, ~50% resolved a convergent network with a 5-255 threshold of duplication depending on the node-edge connection. Using FUMAGWAS: GENE2FUNC, the protein-coding genes were functionally annotated and overrepresentation gene-set analysis for each network gene set was performed[143]. Using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit)[144], over-representation analysis (ORA) was performed on all convergent network gene sets against a curated list of common and rare variant target genes across ASD, BIP, SCZ, and ID[27].

Influence of Functional Similarity on Convergence Degree. Functionally similarity scores across the eGenes represented in each set was calculated using three metrics: (1) Gene Ontology Scores: the average semantic similarity score based on Gene Ontology pathway membership (within Biological Pathway (BP), Cellular Component (CC), and Molecular Function (MF) between genes in a set[81], (2) Brain expression correlation (B.E.C.) score: based on the strength of the correlation in gene expression in the CMC (n=991 after QC) post-mortem dorsa-lateral prefrontal cortex (DLPFC) gene expression data[6], and (3) Signaling Score: based on the proportion of eGenes whose basic functional annotation was categorized as "signaling" (*CALN1, CLCN3, FES, NAGA, PLCL1, TMEM219*; with *PLCL1* and *CLCN3* further separated as specific synaptic genes) or four "epigenetic/regulatory" target genes (*SF3B1, UBE2Q2L, ZNF823, ZNF804A*; with *ZNF823, ZNF804A* as specific transcription factors*)* using FUMAGWAS: GENE2FUNC[143] (**SI Fig. 10**).

Bi-clustering identifies co-expressed genes shared across the downstream transcriptomic impacts of any given set of eGene perturbations, thus, the resolved networks are the transcriptomic similarities between distinct perturbations (convergence)**.** While bi-clustering resolves convergent gene co-expression networks, the strength of convergence within a network can vary. We assigned each network a "degree of convergence" based on (1) network connectivity and (2) similarity of network genes based on biological pathway membership. We performed a principal components analysis on the functional similarity scores and the degree of network convergence. PCA loadings determined the effect of the included variables on the variability across all resolvable sets (arrayed=16320, pooled=827, variables=6). To quantify this, we calculated two small world connectivity network coefficients: the cluster connectivity coefficient based on the proportion of edges present out of all possible edges (*Cp*) and the average path length (*Lp*)[145].

Here we define convergence as (1) increased connectivity of the resolved networks, and (2) functional similarity of genes within the network. Network connectivity was defined by the sum of the clustering coefficient (Cp) and the difference in average length path (Lp) from the maximum average length path resolved across all possible sets [(max)Lp-Lp]. Network functional similarity was scored by taking the sum of the mean semantic similarity scores between all genes in the network. Overall, convergence degree represented the sum of the network connectivity score and the network functional similarity score.

$$convergence = Cp + [\max(Lp) - Lp] + \sum_{1}^{N} MFsemsim + BPsemsim + CCsemsim$$

Convergent networks with matched patterns of gene expression in the post-mortem brain. We clarify that this approach asks how often eGenes are up-regulated together in individual post-mortem brains. To do this, we ran target-convergent network reconstruction in our scRNA-seq data, not the CMC bulk tissue data, for sets of eGenes defined by the clustering observed in the CMC bulk tissue data. We found zero individuals in the CMC data with significant upregulation of all ten risk eGenes. Instead of only evaluating convergence on the basis on eGene functional similarity as in the first portion of the manuscript, we define eGene pairings more broadly based on the signatures of these eGenes in the post-mortem DLPFC – increasing the relevancy to risk at the individual level. Target sets based on gene expression patterns in the CMC (n=991 after QC) post-mortem dorsa-lateral pre-frontal cortex (DLPFC). We performed K-means clustering to subset the data into clusters based on the Z-scored gene expression of the 10 target genes. Although initial silhouette analysis identified the optimal number of clusters as two, visualization by a scree plot suggested the optimal number to be between 4-6 clusters. Given that data clustered by case/control status (2 clusters), and sub-diagnosis of BP, SCZ, AFF, and Controls (4 clusters), to assess clustering based on 10 eGenes, we tested the impact of using 10 clusters and 20 clusters (**SI Fig. 32**). Perturbation identities were assigned based on average positive Z-scores of >=0.5 within each cluster. 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 10% duplication rate and calculated convergence scores and performed GSEA using protein-coding network genes. Of the twenty clusters, networks were recovered for the combination of targets represented in cluster 4 (2 targets; 913 cells; 15% duplication; 13 node genes), cluster 5 (3 targets; 1260 cells; 15% duplication; 13 node genes), cluster 6 (6 targets; 2035 cells; 15% duplication; 34 node genes), cluster 9 (6 targets, 1822 cells, 20% duplication, 108 node genes), cluster 11 (5 targets; 1640 cells; 15% duplication; 25 node genes), cluster 12 (6 targets; 2357 cells; 20% duplication, 152 node genes), cluster 13 (5 targets; 1741 cells; 17.5% duplication, 17 node genes), cluster 18 (6 targets; 1884 cells, 15% duplication, 25 nodes), cluster 19 (6 targets, 2327 cells, 20% duplication, 153 nodes), cluster 20 (6 targets, 2015 cells, 20% duplication, 33 nodes), while low confidence convergence was resolved for cluster 1 (5 targets, 1600 cells; 7.5% duplication; 38 node genes), cluster 8 (3 targets, 1233 cells, 7.5% duplication, 38 node genes), cluster 14 (3 targets, 1020 cells, 5% duplication, 23 nodes) and 16 (4 targets, 1177 cells, 2.5% duplication, 16 nodes). To determine if convergent networks were distinct between diagnostic groups, we first 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 AFF, BIP and SCZ within the clusters and then calculated Jaccard Similarity Indices between clusters based on convergent network gene membership.

Drug prioritization based on perturbation signature reversal in LiNCs Neuronal Cell Lines: To identify drugs that could reverse the resolved convergent perturbation signature across all ten targets, and within each target individually, we used the Query

tool from The Broad Institute's Connectivity Map (Cmap) Server. Briefly, the tool computes weighted enrichment scores (WTCS) between the query set and each signature in the Cmap LINCs gene expression data (dose, time, drug, cell-line), normalizes the WRCS by dividing by signed mean w/in each perturbation (NCS), and computes FDR as fraction of "null signatures" (DMSO) where the absolute NCS exceeds reference signature[146]. We prioritized drugs that reversed signatures specifically in neuronal cells (either neurons (NEU) of neural progenitor cells (NPCs) with NCS <= -1.00) and filtered for (i) drugs that cross the blood-brain barriers, (ii) drugs that target genes expressed in iGLUT neurons based on bulk RNA-sequencing data from our lab and (ii) drugs that are currently launched or in clinical trial according the cMAP Drug Repurposing database and without evidence of neurotoxicity (**Box 2**).

CRISPRa/shRNA Validation[27]*.* At day -2, dCas9-VPR hiPSC-NPCs were seeded as 0.6x10^6 cells / well in a 24-well plate coated with Matrigel. At day -1, cells were transduced with rtTA (Addgene 20342) and *NGN2* (Addgene 99378) lentiviruses. Medium was switched to non-viral medium four hours post infection. At D0, 1 µg/ml dox was added to induce *NGN2*-expression. At D1, transduced hiPSC-NPCs were treated with corresponding antibiotics to the lentiviruses (1 mg/ml G-418 for *NGN2*-Neo) in order to increase the purity of transduced hiPSC-NPCs. At D3, NPC medium was switched to neuronal medium (Brainphys (Stemcell Technologies, #05790), 1x N2 (Life Technologies #17502-048), 1x B27-RA (Life Technologies #12587-010), 1 µg/ml Natural Mouse Laminin (Life Technologies), 20 ng/ml BDNF (Peprotech #450-02), 20 ng/ml GDNF (Peptrotech #450-10), 500 µg/ml Dibutyryl cyclic-AMP (Sigma #D0627), 200 nM L-ascorbic acid (Sigma #A0278)) including 1 µg/ml Dox. 50% of the medium was replaced with fresh neuronal medium once every second day. At D4 individual shRNA/gRNA vectors (Addgene 99374), either targeting eGenes or scramble controls. 3-5 vectors were tested per eGene. Medium was switched to non-viral medium four hours post infection. At D5, transduced iGLUTs were treated with corresponding antibiotics to the gRNA lentiviruses (1 mg/ml HygroB for lentiguide-Hygro/lentiguide-Hygro-mTagBFP2) before harvesting at D7 in order to assess eGene perturbation efficacy via qPCR.

Real time-quantitative PCR. Real time qPCR was performed as previously described[127]. Specifically, cell cultures were harvested with Trizol and total RNA extraction was carried out following the manufacturer's instructions. Quantitative transcript analysis was performed using a QuantStudio 7 Flex Real-Time PCR System with the Power SYBR Green RNA-to-Ct Real-Time qPCR Kit (all Thermo Fisher Scientific). Total RNA template (25 ng per reaction) was added to the PCR mix, including primers listed below. qPCR conditions were as follows; 48°C for 15 min, 95°C for 10 min followed by 45 cycles (95°C for 15 s, 60°C for 60 s). All qPCR data is collected from at least three independent biological replicates of one experiment. A one-way ANOVA with posthoc Dunnett's multiple comparisons test was performed on data for the set of targeting vectors for each eGene relative to the scramble control vector. Data analyses were performed using GraphPad PRISM 6 software.

Gene	**Forward primer sequence**	**Reverse primer sequence**
SNAP91	AGGACCCATTAGCGGATCTTAACA	GCTCCCTTTGAAACTCAGCATCAA

CLCN3	AATCATAGGTCAAGCAGAGGGTCC	CCACAGGCATATGGAGCAAATACC
PLCL1	GGAGCCTCTGGGTGCAATTC	CATGGAACTCCATCCCTGCT
DOC2A	CCAAGACCCTGGAAGTCACC	GACACGCCACCAATGAAGTC
SNCA	CTTGCCTTCAAGCCTTCTGC	ACACCACACTGTCGTCGAAT
ZNF823	ACGCCCTGGAAGGTCTGTAT	GCCACTGAGTCCATTTCCCA
INO80E	AAACTCAAGATGGCGGTGGG	AGGGACCGTGCTCAGGAT
SF3B1	ATCCTTTTGCAGATGGAGGGA	CTTTTTCTGCTAGCTGTTGCCT
THOC7	GCCGTGACTGACGACGA	TCCGATCATCTCCAGCACCAT
GATAD2A	ACTTCGGTTGGTCAGCACAT	TCGGTCATTCTGAACTGGCT
CALN1	CGATGCGGCTGCCAGAG	AACGGCATCTTTTCCCAGGT
CUL9	AGAAGGATGAAGGCCGAACC	TCCAGCACCAGACAAACCAG
TMEM219	CGGGAGTCCCACTGGTCTA	CAAGAAGTTGATTCACACCTGAG
PCCB	AAGTTTCCTGGAGACAGCGT	ACTGTAAAATCCTGACTGAAGACAT
FURIN	AGAAGGTCTTCACCAACACG	GAAGTGGTAATAGTCCCCGAAG
GAPDH	AGGGCTGCTTTTAACTCTGGT	CCCCACTTGATTTTGGAGGGA

Immunostaining and high-content imaging microscopy

Neurite analysis: Immature iGLUTs were seeded as $1.5x10^4$ cells/well in a 96-well plate coated with 4x Matrigel at day 3. iGLUTs were plated in media containing individual shRNA/gRNA vectors or pools of mixed shRNA and gRNA vectors (Addgene 99374), either targeting eGenes or scramble controls. Medium was switched to non-viral medium four hours post infection. At day 4, transduced iGLUTs were treated with corresponding antibiotics to the gRNA lentiviruses (1 mg/ml HygroB for lentiguide-Hygro/lentiguide-Hygro-mTagBFP2) followed by half medium changes until the neurons were fixed at day 7. At day 7, cultures were fixed using 4% formaldehyde/sucrose in PBS with Ca^{2+} and Mg^{2+} for 10 minutes at room temperature (RT). Fixed cultures were washed twice in PBS and permeabilized and blocked using 0.1% Triton/2% Normal Donkey Serum (NDS) in PBS for two hours. Cultures were then incubated with primary antibody solution (1:1000 MAP2 anti chicken (Abcam, ab5392) in PBS with 2% NDS) overnight at 4°C. Cultures were then washed 3x with PBS and incubated with secondary antibody solution (1:500 donkey anti chicken Alexa 647 (Life technologies, A10042) in PBS with 2% NDS) for 1 hour at RT. Cultures were washed a further 3x with PBS with the second wash containing 1 µg/ml DAPI. Fixed cultures were then imaged on a CellInsight CX7 HCS Platform with a 20x objective (0.4 NA) and neurite tracing analysis performed using the neurite tracing module in the Thermo Scientific HCS Studio 4.0 Cell Analysis Software. 12-24 wells were imaged per condition across a minimum of 2 independent cell lines, with 9 images acquired per well for neurite tracing analysis; each “N” therefore represents an average of hundreds of neurons per image. A one-way ANOVA with a post hoc Bonferroni multiple comparisons test was performed on data for neurite length per neuron using Graphpad Prism.

Synapse analyses: Commercially available primary human astrocytes (pHAs, Sciencell, #1800; isolated from fetal female brain) were seeded on D3 at $0.85x10^4$ cells per well on a 4x Matrigel-coated 96□W plate in neuronal media supplemented with 2% fetal bovine serum (FBS). iGLUTs were seeded as $1.5x10^5$ cells/well in a 96-well plate coated with 4x Matrigel at day 5. Half changes of neuronal media were performed twice a week until fixation. At day 13, iGLUTs were treated with 200 nM Ara-C to reduce the proliferation of non-neuronal cells in the culture. At day 18, Ara-C was completely withdrawn by full medium change while adding media containing individual shRNA/gRNA vectors or pools of mixed shRNA and gRNA vectors (Addgene 99374), either targeting eGenes or scramble controls. Medium was switched to non-viral medium four hours post infection. At day 19, transduced iGLUTs were treated with corresponding antibiotics to the gRNA lentiviruses (1 mg/ml HygroB for lentiguide-Hygro/lentiguide-Hygro-mTagBFP2) followed by half medium changes until the neurons were fixed at day 21. At day 21, cultures were fixed and immunostained as described previously, with an additional antibody stain for Synapsin1 (primary antibody: 1:500 Synapsin1 anti mouse (Synaptic Systems, 106 011); secondary antibody: donkey anti mouse Alexa 568 (Life technologies A10037)). Stained cultures were imaged and analyzed as above using the synaptogenesis module in the Thermo Scientific HCS Studio 4.0 Cell Analysis Software to determine SYN1+ puncta number, area, and intensity per neurite length in each image. 20 wells were imaged per condition across a minimum of 2 independent cell lines, with 9 images acquired per well for synaptic puncta analysis. A one-way ANOVA with a post hoc Bonferroni multiple comparisons test was performed on data for puncta number per neurite length using Graphpad Prism.

Antibody	**Species**	**Supplier**	**Product number**	**Dilution**
MAP2	Ck	Abcam	ab5392	1:500
SYNAPSIN1	Ms	Synaptic Systems	106 011	1:500
Alexa 568 anti-Mouse	Ms	Life technologies	A10037	1:500
Alexa 647 anti-Chicken	Ck	Life technologies	A10042	1:500

Multiple Electrode array (MEA): Commercially available primary human astrocytes (pHAs, Sciencell, #1800; isolated from fetal female brain) were seeded on D3 at $1.7x10^4$ cells per well on a 4x Matrigel-coated 48□W MEA plate (catalog no. M768-tMEA-48W; Axion Biosystems) in neuronal media supplemented with 2% fetal bovine serum (FBS). At D5, iGLUTs were detached, spun down, and seeded on the pHA cultures at $1.5x10^5$ cells per well. Half changes of neuronal media supplemented with 2% FBS were performed twice a week until day 42. At day 13, co-cultures were treated with 200 nM Ara-C to reduce the proliferation of non-neuronal cells in the culture. At Day 18, Ara-C was completely withdrawn by full medium change. At day 25, a full media change was performed to add media containing individual shRNA/gRNA vectors or pools of mixed shRNA and gRNA vectors (Addgene 99374), either targeting eGenes or scramble controls. Medium was switched to non-viral medium four hours post infection. If drug

treatments were included, D26 neurons were treated for 48hrs with either Anandamide (10µM), Etomoxir (10µM), Simvastatin (10µM), or matched vehicles. Electrical activity of iGLUTs was recorded at 37°C twice every week from day 28 to day 42 using the Axion Maestro MEA reader (Axion Biosystems). Recording was performed via AxiS 2.4. Batch mode/statistic compiler tool was run following the final recording. Quantitative analysis of the recording was exported as Microsoft excel sheet. Data from 6-12 biological replicates were analyzed using GraphPad PRISM 6 software or R.

RNAseq: RNA Sequencing libraries were prepared using the Kapa Total RNA library prep kit. Paired-end sequencing reads (100bp) were generated on a NovaSeq platform. Raw reads were aligned to hg19 using STAR aligner[147] (v2.5.2a) and gene-level expression were quantified by featureCounts[148] (v1.6.3) based on Ensembl GRCh37.70 annotation model. Genes with over 10 counts per million (CPM) in at least four samples were retained. After filtering, the raw read counts were normalized by the voom[149] function in limma and differential expression was computed by the moderated t-test implemented in limma[150]. Differential gene expression analysis was performed between each CRISPRa/shRNA target group and scramble control group. Bayes shrinkage (limma::eBayes) estimated modified t- and p- values and identified differentially expressed genes (DEGs) based on an FDR <= 0.05 (limma::TopTable)[151]. Gene Ontology/pathways were evaluated using Gene-set Enrichment Analysis (GSEA)[152], with genes expressed in iGLUTs as our baseline comparison. In these analyses, the t-test statistics from the differential expression contrast were used to rank genes in the GSEA using the R package ClusterProfiler[153]. Permutations (up to 100,000 times) were used to assess the GSEA enrichment P value. Log2 fold changes in expression were calculated across all RNA-seq samples in our arrayed dataset.

Analysis of additive and non-additive effects. We applied our published approach to resolve distinct additive and non-additive transcriptomic effects after combinatorial manipulation of genetic variants and/or chemical perturbagens, developed[27], applied[59], and described in detail[83]. The expected additive effect was modeled through addition of the individual comparisons; the non-additive effect was modeled by subtraction of the additive effect from the combinatorial perturbation comparison. Fitting of this model for differential expression identifies genes that show a difference in the expected differential expression computed for the additive model compared to the observed combinatorial perturbation. Briefly, the non-additive effect between eGenes was identified using a limma's linear model analysis. The coefficients, standard deviations and correlation matrix were calculated, using *contrasts.fit,* in terms of the comparisons of interest. Empirical Bayes moderation was applied using the *eBayes* function to obtain more precise estimates of gene-wise variability. P-values were adjusted for multiple hypotheses testing using false discovery rate (FDR) estimation, and differentially expressed genes were determined as those with FDR ≤ 10%, unless stated otherwise. Two methods were used to compare the extent of synergy between data sets. First, we calculated the fraction of synergistic genes (FDR<10%) measure the extent of synergy. Second, we calculated a synergy coefficient, π1, as the fraction of non-null synergistic P-values, to inform the existence of a synergistic component, even if the P-values themselves are not significant genome-wide.

However, interpretation of the resulting DEGs depends on several factors, such as the

direction of fold change (FC) in all three models. To identify genes whose magnitude of change is larger in the combinatorial perturbation vs. the additive model, we categorized all genes by the direction of their change in both models and their $\log_2$(FC) in the non-additive model. First, $\log_2$(FC) standard errors (SE) were calculated for all samples. Genes were then grouped into 'positive non-addition' if their FC was larger than SE and 'negative non-addition' if smaller than -SE. If the corresponding additive model $\log_2$(FC) showed the same or no direction, the gene was classified as "more" differentially expressed in the combinatorial perturbation than predicted. GSEA was performed on a curated subset of the MAGMA collection using the limma package camera function, which tests if genes are ranked highly in comparison to other genes in terms of differential expression, while accounting for inter-gene correlation. Due to the small sample size in this study and moderate fold changes in some eGene perturbations, changes in gene expression may be small and distributed across many genes. However, powerful enrichment analyses in the limma package may be used to evaluate enrichment based on genes that are not necessarily genome-wide significant and identify sets of genes for which the distribution of t-statistics differs from expectation. Over-representation analysis (ORA) was performed when subsets of DEGs were of interest; genes of interests were ranked by –log10 (p-value) and enrichment was performed against a background of all expressed genes using the WebGestaltR package.

Dataset for population-level analysis of synergy: Individuals from the Sweden-SCZ Population-Based cohort were obtained from the database of Genotypes and Phenotypes, Study Accession: phs000473.v2.p2 (N_{Cases} = 5,232, $N_{Controls}$ = 6,468)[154].

Pathway polygenic risk scores: Pathway-specific polygenic risk score (PRS) analyses were performed using PRSice-2 (v2.3.5) on individual genotype data for the Sweden-SCZ Population based cohort. A total of 4,834 individuals diagnosed with SCZ and 6,128 controls were included after quality control. To calculate the scores, we used a version of the summary statistics from the PGC SCZ GWAS that excludes the Sweden-SCZ data to prevent inflation of results. SNPs were annotated to genes and pathways based on GTF files obtained from ENSEMBL (GRCh37.75). To include potential gene regulatory elements, gene coordinates were extended 35 kilobases (kb) upstream and 10 kb downstream of each gene. We excluded from analyses the MHC region (chr6:25Mb-34Mb), ambiguous SNPs (A/T and G/C), and SNPs not present in both GWAS summary statistics and genotype data.

To obtain empirical "competitive" *P*-values, that assess GWAS signal enrichment while accounting for pathway size, we performed the following permutation procedure: first, a "background" pathway containing all genic SNPs is constructed, and clumping is performed within this pathway. For each pathway with *m* SNPs, *N*=10,000 null pathways are generated by randomly selecting *m* SNPs from the "background" pathway. The competitive *P*-value can then be calculated as

$$competitive\ P-value = \frac{\sum_{n=1}^{N} I(P_n < P_o) + 1}{N+1}$$

where *I*(.) is an indicator function, taking a value of 1 if the association *P*-value of the observed pathway (P_0) is larger than the one obtained from the nth null pathway (P_n),

and 0 otherwise (see[84] for additional details).

For the analyses testing whether non-additive genes from synaptic/regulatory pathways explain larger R^2 than *the same number* of non-additive genes from random combinations (**SI Fig. 24**), we took 2,799 random genes from the non-additive synaptic and regulatory transcriptome, which corresponds to the number of genes with non-additive effects in one of the random joint perturbations. For the GTF NULL permutation analyses, we selected n=2,799 random genes from the GTF file GRCh37.75. Pathway-specific PRS for each sample of 2,799 genes was calculated using PRSet[84], as described above. This procedure was repeated 1000 times.

Transcriptomic Risk Score (TRS) Analyses: In order to test the impact of non-additive genetic effects *in silico*, we used transcriptomic imputation methods to calculate genetically-regulated gene expression (GREX) for individuals from the Sweden-SCZ Population-Based cohort (**SI Table 3**). Brain GREX was calculated using PrediXcan[73] with CMC dorsolateral prefrontal cortex (CMC-DLPFC) models[6]. Predicted GREX levels were calculated for the fifteen eGenes. An initial test of aberrant gene expression was performed by counting the number of genes with dysregulated GREX (defined as predicted GREX in the top or bottom decile of overall expression of that gene, defined in the direction of effect of that gene's association with SCZ from S-PrediXcan analyses (top decile for positive effect, bottom decile for negative effect) for each of the five-gene groups (synaptic, regulatory, multi-function), and summed the number of aberrant genes present in each individual for each perturbed gene group (Synaptic, Regulatory, and Multi-function). We then looked at the SCZ case/control proportion within each group of individuals with 3+, 1-2, and any genes with aberrant GREX.

Association of Synaptic, Regulatory, and Multi-function gene-sets with SCZ: We tested for association of each of the fifteen eGene GREX individually with SCZ (SCZ ~ GREX), and then calculated composite scores of group GREX (Synaptic, Regulatory, and Multi-function) using a Transcriptomic Risk Score (TRS), calculated as the sum of each GREX weighted by the direction of gene perturbation (1 for activation, -1 for inhibition) from *in vivo* experiments, divided by the total number of genes (N) in the gene-set (Equation X):

$$TRS = \frac{\sum_1^N GREX \times direction\ of\ perturbation}{N}$$

We then tested for association of each TRS (Synaptic, Regulatory, and Multi-function) with SCZ status in the Swedish cohort.

Permutation tests. We performed permutation tests to assess the impact of (1) the number of genes included in our TRS gene group and (2) the number of pathways impacted by those genes on SCZ case status. We used S-PrediXcan to find genes with CMC-DLPFC GREX associated with SCZ in a large SCZ cohort (N_{Cases} = 11,260, $N_{Controls}$ = 24,542)[39]. From this resulting list of genes, we assigned genes to two groups: nominally-significant genes (N=1,963, Bonferroni $p<0.05$), and tissue-specific significant genes (N=144, $p<0.05/N_{\text{Genes in CMC-DLPFC PrediXcan model}}$). We created pathway sets affected by these genes using the overlap with Kyoto encyclopedia of genes and genomes (KEGG)[155] and gene ontology (GO)[156,157]. This gave us a sampling pool of 1,465 genes affecting 8,324 pathway sets for the nominally-significant group, and 110 genes

affecting 2,382 pathway sets for the tissue-specific group. We then performed permutation sampling analyses (for nominally-significant and tissue-specific significant gene-pathway set pools) where we randomly sampled sets of five, ten, or fifteen genes from the sampling pool (adjusted for the size of each pathway set), calculated TRS from the sampled gene-set, and looked at the association of TRS with SCZ. We performed sampling 100,000 times for each gene-set size. For this analysis, TRS was calculated by taking the sum of each gene in the gene-sets GREX weighted by the direction of effect of the gene association with SCZ from our S-PrediXcan analysis (1 or -1) (Equation X):

$$TRS = \frac{\sum_1^N GREX \times direction\ of\ perturbation}{N}$$

We then looked at the overall association the number of pathways hit by each TRS (based on the annotated lists) with SCZ variance explained (**SI Fig. 1A-C**). To determine if the type of pathways hit by our perturbed genes was important to SCZ risk (i.e. is it more important to hit multiple, similar pathways or more diverse pathways to increase SCZ variance explained), we additionally assessed whether the similarity in make-up of pathways affected by the TRS was associated with SCZ. To do this, we used the R "GeneOverlap" package to calculate the average Jaccard Index of pathways for each TRS, and looked at the association of that index with SCZ.

STATEMENT OF ETHICS

Yale University Institutional Review Board waived ethical approval for this work. 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 data are similarly lacking identifiable information and are not considered human subjects research.

CONFLICT OF INTEREST STATEMENT

E.S. is today an employee at Regeneron. K.J.B is a scientific advisor to Rumi Scientific Inc. and Neuro Pharmaka Inc.

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), R01MH106056 (K.J.B and S.A.), U01DA047880 (K.J.B and S.A), R01DA048279 (K.J.B and S.A), T35DK104689-07 (E.C.), K08MH122911 (G.V.). R01MH125246 (P.R.), U01MH116442 (P.R.), R01MH109677 (P.R.), I01BX002395 (P.R.), and by the State of Connecticut, Department of Mental Health and Addiction Services. This publication does not express the views of the Department of Mental Health and Addiction Services or the State of Connecticut.

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. iGLUT transcriptomic and phenotypic studies were conducted by P.J.M.D, with assistance by E.C. Synergy analysis was conceptualized by N.S. and applied by P.J.M.D; convergent analyses were conceptualized by K.G.T. and applied by C.S. Transcriptomic imputation was conducted by J.J. and L.H.; pathway-specific PRS by J.G.G. and P.O. 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 iGLUT neuron 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 P.J.M.D, 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): GSE200774. Processed data and accompanying code can be accessed through Synapse (syn27819129)

REFERENCES

1 Singh, T. *et al.* Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509-516, doi:10.1038/s41586-022-04556-w (2022).

2 Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508, doi:10.1038/s41586-022-04434-5 (2022).

3 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).

4 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).

5 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).

6 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).

7 Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nature neuroscience* **21**, 1117-1125 (2018).
8 Zeng, B. *et al.* Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat Genet* **54**, 161-169, doi:10.1038/s41588-021-00987-9 (2022).
9 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).
10 Forrest, M. P. *et al.* Open Chromatin Profiling in hiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci. *Cell stem cell*, doi:10.1016/j.stem.2017.07.008 (2017).
11 Zhang, S. *et al.* Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science* **369**, 561-565, doi:10.1126/science.aay3983 (2020).
12 Zhu, K. *et al.* Multi-omic profiling of the developing human cerebral cortex at the single-cell level. *Sci Adv* **9**, eadg3754, doi:10.1126/sciadv.adg3754 (2023).
13 Huo, Y., Li, S., Liu, J., Li, X. & Luo, X. J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat Commun* **10**, 670, doi:10.1038/s41467-019-08666-4 (2019).
14 Alver, M., Lykoskoufis, N., Ramisch, A., Dermitzakis, E. T. & Ongen, H. Leveraging interindividual variability of regulatory activity for refining genetic regulation of gene expression in schizophrenia. *Mol Psychiatry* **27**, 5177-5185, doi:10.1038/s41380-022-01768-4 (2022).
15 Casella, A. M., Colantuoni, C. & Ament, S. A. Identifying enhancer properties associated with genetic risk for complex traits using regulome-wide association studies. *PLoS Comput Biol* **18**, e1010430, doi:10.1371/journal.pcbi.1010430 (2022).
16 Dong, P., Voloudakis, G., Fullard, J. F., Hoffman, G. E. & Roussos, P. Convergence of the dysregulated regulome in schizophrenia with polygenic risk and evolutionarily constrained enhancers. *Mol Psychiatry*, doi:10.1038/s41380-023-02370-y (2023).
17 Rajarajan, P. *et al.* Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* **362**, doi:10.1126/science.aat4311 (2018).
18 Song, M. *et al.* Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes. *Nat Genet* **51**, 1252-1262, doi:10.1038/s41588-019-0472-1 (2019).
19 Espeso-Gil, S. *et al.* A chromosomal connectome for psychiatric and metabolic risk variants in adult dopaminergic neurons. *Genome Med* **12**, 19, doi:10.1186/s13073-020-0715-x (2020).
20 Hu, B. *et al.* Neuronal and glial 3D chromatin architecture informs the cellular etiology of brain disorders. *Nat Commun* **12**, 3968, doi:10.1038/s41467-021-24243-0 (2021).

21 Girdhar, K. *et al.* Chromatin domain alterations linked to 3D genome organization in a large cohort of schizophrenia and bipolar disorder brains. *Nat Neurosci* **25**, 474-483, doi:10.1038/s41593-022-01032-6 (2022).
22 Girdhar, K. *et al.* The neuronal chromatin landscape in adult schizophrenia brains is linked to early fetal development. *medRxiv*, doi:10.1101/2023.10.02.23296067 (2023).
23 Guo, M. G. *et al.* Integrative analyses highlight functional regulatory variants associated with neuropsychiatric diseases. *Nat Genet* **55**, 1876-1891, doi:10.1038/s41588-023-01533-5 (2023).
24 Rummel, C. K. *et al.* Massively parallel functional dissection of schizophrenia-associated noncoding genetic variants. *Cell* **186**, 5165-5182 e5133, doi:10.1016/j.cell.2023.09.015 (2023).
25 McAfee, J. C. *et al.* Systematic investigation of allelic regulatory activity of schizophrenia-associated common variants. *Cell Genom* **3**, 100404, doi:10.1016/j.xgen.2023.100404 (2023).
26 Yang, X. *et al.* Functional characterization of gene regulatory elements and neuropsychiatric disease-associated risk loci in iPSCs and iPSC-derived neurons. *bioRxiv*, 2023.2008.2030.555359, doi:10.1101/2023.08.30.555359 (2023).
27 Schrode, N. *et al.* Synergistic effects of common schizophrenia risk variants. *Nat Genet* **51**, 1475-1485, doi:10.1038/s41588-019-0497-5 (2019).
28 Zhang, S. *et al.* Multiple genes in a single GWAS risk locus synergistically mediate aberrant synaptic development and function in human neurons. *Cell Genom* **3**, 100399, doi:10.1016/j.xgen.2023.100399 (2023).
29 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).
30 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).
31 Liu, D. *et al.* Impact of schizophrenia GWAS loci converge onto distinct pathways in cortical interneurons vs glutamatergic neurons during development. *Mol Psychiatry* **27**, 4218-4233, doi:10.1038/s41380-022-01654-z (2022).
32 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).
33 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).
34 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).
35 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).
36 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).
37 Hsu, Y. H. *et al.* Using brain cell-type-specific protein interactomes to interpret neurodevelopmental genetic signals in schizophrenia. *iScience* **26**, 106701, doi:10.1016/j.isci.2023.106701 (2023).
38 Bipolar, D., Schizophrenia Working Group of the Psychiatric Genomics Consortium. Electronic address, d. r. v. e., Bipolar, D. & Schizophrenia Working Group of the Psychiatric Genomics, C. Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705-1715 e1716, doi:10.1016/j.cell.2018.05.046 (2018).
39 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).
40 Hall, L. S. *et al.* A transcriptome-wide association study implicates specific pre- and post-synaptic abnormalities in schizophrenia. *Hum Mol Genet* **29**, 159-167, doi:10.1093/hmg/ddz253 (2020).
41 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).
42 Network, Pathway Analysis Subgroup of the Psychiatric Genomics, C., International Inflammatory Bowel Disease Genetics, C. & International Inflammatory Bowel Disease Genetics Consortium, I. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat Neurosci* **18**, 199-209, doi:10.1038/nn.3922 (2015).
43 Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, doi:10.1126/science.aat8127 (2018).
44 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).
45 Ruzicka, W. B. *et al.* Single-cell multi-cohort dissection of the schizophrenia transcriptome. *medRxiv*, 2022.2008.2031.22279406, doi:10.1101/2022.08.31.22279406 (2022).
46 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).
47 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).
48 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).
49 Lalli, M. A., Avey, D., Dougherty, J. D., Milbrandt, J. & Mitra, R. D. High-throughput single-cell functional elucidation of neurodevelopmental disease-associated genes reveals convergent mechanisms altering neuronal differentiation. *Genome Res* **30**, 1317-1331, doi:10.1101/gr.262295.120 (2020).

50 Willsey, H. R. *et al.* Parallel in vivo analysis of large-effect autism genes implicates cortical neurogenesis and estrogen in risk and resilience. *Neuron* **109**, 1409, doi:10.1016/j.neuron.2021.03.030 (2021).
51 Weinschutz Mendes, H. *et al.* High-throughput functional analysis of autism genes in zebrafish identifies convergence in dopaminergic and neuroimmune pathways. *Cell reports* **42**, 112243, doi:10.1016/j.celrep.2023.112243 (2023).
52 Meng, X. *et al.* Assembloid CRISPR screens reveal impact of disease genes in human neurodevelopment. *Nature* **622**, 359-366, doi:10.1038/s41586-023-06564-w (2023).
53 Li, C. *et al.* Single-cell brain organoid screening identifies developmental defects in autism. *Nature* **621**, 373-380, doi:10.1038/s41586-023-06473-y (2023).
54 Wang, B. *et al.* A foundational atlas of autism protein interactions reveals molecular convergence. *bioRxiv*, 2023.2012.2003.569805, doi:10.1101/2023.12.03.569805 (2023).
55 Pintacuda, G. *et al.* Protein interaction studies in human induced neurons indicate convergent biology underlying autism spectrum disorders. *Cell Genomics* **3**, doi:10.1016/j.xgen.2022.100250 (2023).
56 Murtaza, N. *et al.* Neuron-specific protein network mapping of autism risk genes identifies shared biological mechanisms and disease-relevant pathologies. *Cell reports* **41**, 111678, doi:10.1016/j.celrep.2022.111678 (2022).
57 Wainschtein, P. *et al.* Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet* **54**, 263-273, doi:10.1038/s41588-021-00997-7 (2022).
58 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).
59 Wang, M. *et al.* Transformative Network Modeling of Multi-omics Data Reveals Detailed Circuits, Key Regulators, and Potential Therapeutics for Alzheimer's Disease. *Neuron* **109**, 257-272 e214, doi:10.1016/j.neuron.2020.11.002 (2021).
60 Flaherty, E. *et al.* Neuronal impact of patient-specific aberrant NRXN1alpha splicing. *Nat Genet* **51**, 1679-1690, doi:10.1038/s41588-019-0539-z (2019).
61 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).
62 Pak, C. *et al.* Cross-platform validation of neurotransmitter release impairments in schizophrenia patient-derived NRXN1-mutant neurons. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2025598118 (2021).
63 Marro, S. G. *et al.* Neuroligin-4 Regulates Excitatory Synaptic Transmission in Human Neurons. *Neuron* **103**, 617-626 e616, doi:10.1016/j.neuron.2019.05.043 (2019).
64 Zhang, Z. *et al.* The fragile X mutation impairs homeostatic plasticity in human neurons by blocking synaptic retinoic acid signaling. *Science translational medicine* **10**, doi:10.1126/scitranslmed.aar4338 (2018).
65 Yi, F. *et al.* Autism-associated SHANK3 haploinsufficiency causes Ih channelopathy in human neurons. *Science* **352**, aaf2669, doi:10.1126/science.aaf2669 (2016).

66 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).
67 Meijer, M. *et al.* A Single-Cell Model for Synaptic Transmission and Plasticity in Human iPSC-Derived Neurons. *Cell reports* **27**, 2199-2211 e2196, doi:10.1016/j.celrep.2019.04.058 (2019).
68 Sun, Y. *et al.* A deleterious Nav1.1 mutation selectively impairs telencephalic inhibitory neurons derived from Dravet Syndrome patients. *Elife* **5**, doi:10.7554/eLife.13073 (2016).
69 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).
70 Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).
71 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).
72 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).
73 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).
74 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).
75 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).
76 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).
77 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
78 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).
79 Batiuk, M. Y. *et al.* Upper cortical layer-driven network impairment in schizophrenia. *Sci Adv* **8**, eabn8367, doi:10.1126/sciadv.abn8367 (2022).
80 Teixeira, J. R., Szeto, R. A., Carvalho, V. M. A., Muotri, A. R. & Papes, F. Transcription factor 4 and its association with psychiatric disorders. *Translational psychiatry* **11**, 19, doi:10.1038/s41398-020-01138-0 (2021).
81 Yu, G. Gene Ontology Semantic Similarity Analysis Using GOSemSim. *Methods in molecular biology* **2117**, 207-215, doi:10.1007/978-1-0716-0301-7_11 (2020).

82 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).
83 Schrode, N., Seah, C., Deans, P. J. M., Hoffman, G. & Brennand, K. J. Analysis framework and experimental design for evaluating synergy-driving gene expression. *Nat Protoc* **16**, 812-840, doi:10.1038/s41596-020-00436-7 (2021).
84 Choi, S. W. *et al.* The power of pathway-based polygenic risk scores. *Research Square*, doi:https://doi.org/10.21203/rs.3.rs-643696/v1 (2021).
85 Namkung, H. *et al.* The miR-124-AMPAR pathway connects polygenic risks with behavioral changes shared between schizophrenia and bipolar disorder. *Neuron* **111**, 220-235 e229, doi:10.1016/j.neuron.2022.10.031 (2023).
86 Malt, E. A., Juhasz, K., Malt, U. F. & Naumann, T. A Role for the Transcription Factor Nk2 Homeobox 1 in Schizophrenia: Convergent Evidence from Animal and Human Studies. *Front Behav Neurosci* **10**, 59, doi:10.3389/fnbeh.2016.00059 (2016).
87 Corsello, S. M. *et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med* **23**, 405-408, doi:10.1038/nm.4306 (2017).
88 Kurita, M. *et al.* HDAC2 regulates atypical antipsychotic responses through the modulation of mGlu2 promoter activity. *Nat Neurosci* **15**, 1245-1254, doi:10.1038/nn.3181 (2012).
89 Lichtstein, D. *et al.* Na(+), K(+)-ATPase Signaling and Bipolar Disorder. *Int J Mol Sci* **19**, doi:10.3390/ijms19082314 (2018).
90 Rees, E. *et al.* Targeted Sequencing of 10,198 Samples Confirms Abnormalities in Neuronal Activity and Implicates Voltage-Gated Sodium Channels in Schizophrenia Pathogenesis. *Biol Psychiatry* **85**, 554-562, doi:10.1016/j.biopsych.2018.08.022 (2019).
91 Kim, E., Huh, J. R. & Choi, G. B. Prenatal and postnatal neuroimmune interactions in neurodevelopmental disorders. *Nature Immunology*, doi:10.1038/s41590-024-01797-x (2024).
92 Hardingham, G. E. & Do, K. Q. Linking early-life NMDAR hypofunction and oxidative stress in schizophrenia pathogenesis. *Nat Rev Neurosci* **17**, 125-134, doi:10.1038/nrn.2015.19 (2016).
93 Cannon, M., Jones, P. B. & Murray, R. M. Obstetric complications and schizophrenia: historical and meta-analytic review. *Am J Psychiatry* **159**, 1080-1092, doi:10.1176/appi.ajp.159.7.1080 (2002).
94 Estes, M. L. & McAllister, A. K. Maternal immune activation: Implications for neuropsychiatric disorders. *Science* **353**, 772-777, doi:10.1126/science.aag3194 (2016).
95 Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669, doi:10.1038/s41588-019-0538-0 (2019).
96 Wainberg, M. *et al.* Opportunities and challenges for transcriptome-wide association studies. *Nat Genet* **51**, 592-599, doi:10.1038/s41588-019-0385-z (2019).

97 Weiner, D. J. *et al.* Statistical and functional convergence of common and rare genetic influences on autism at chromosome 16p. *Nat Genet* **54**, 1630-1639, doi:10.1038/s41588-022-01203-y (2022).
98 Domingo, J. *et al.* Non-linear transcriptional responses to gradual modulation of transcription factor dosage. *bioRxiv*, 2024.2003.2001.582837, doi:10.1101/2024.03.01.582837 (2024).
99 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).
100 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).
101 Seah, C. *et al.* Modeling gene x environment interactions in PTSD using human neurons reveals diagnosis-specific glucocorticoid-induced gene expression. *Nat Neurosci* **25**, 1434-1445, doi:10.1038/s41593-022-01161-y (2022).
102 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).
103 Carre, C. *et al.* Next-Gen GWAS: full 2D epistatic interaction maps retrieve part of missing heritability and improve phenotypic prediction. *Genome biology* **25**, 76, doi:10.1186/s13059-024-03202-0 (2024).
104 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).
105 Li, Z. *et al.* Identification of potential biomarkers and their correlation with immune infiltration cells in schizophrenia using combinative bioinformatics strategy. *Psychiatry Res* **314**, 114658, doi:10.1016/j.psychres.2022.114658 (2022).
106 Childers, E., Bowen, E. F. W., Rhodes, C. H. & Granger, R. Immune-Related Genomic Schizophrenic Subtyping Identified in DLPFC Transcriptome. *Genes (Basel)* **13**, doi:10.3390/genes13071200 (2022).
107 Luo, C. *et al.* Subtypes of schizophrenia identified by multi-omic measures associated with dysregulated immune function. *Mol Psychiatry* **26**, 6926-6936, doi:10.1038/s41380-021-01308-6 (2021).
108 Zubiaur, P. *et al.* Variants in COMT, CYP3A5, CYP2B6, and ABCG2 Alter Quetiapine Pharmacokinetics. *Pharmaceutics* **13**, doi:10.3390/pharmaceutics13101573 (2021).
109 Shen, H. *et al.* Adjunctive therapy with statins in schizophrenia patients: A meta-analysis and implications. *Psychiatry Res* **262**, 84-93, doi:10.1016/j.psychres.2018.02.018 (2018).
110 Sommer, I. E. *et al.* Simvastatin Augmentation for Patients With Early-Phase Schizophrenia-Spectrum Disorders: A Double-Blind, Randomized Placebo-Controlled Trial. *Schizophr Bull* **47**, 1108-1115, doi:10.1093/schbul/sbab010 (2021).
111 Weiser, M. *et al.* A randomized controlled trial of add-on naproxen, simvastatin and their combination for the treatment of schizophrenia or schizoaffective disorder. *European neuropsychopharmacology : the journal of the European*

College of Neuropsychopharmacology **73**, 65-74, doi:10.1016/j.euroneuro.2023.04.007 (2023).
112 Aichholzer, M. *et al.* Inflammatory monocyte gene signature predicts beneficial within group effect of simvastatin in patients with schizophrenia spectrum disorders in a secondary analysis of a randomized controlled trial. *Brain Behav Immun Health* **26**, 100551, doi:10.1016/j.bbih.2022.100551 (2022).
113 Zaki, J. K. *et al.* Exploring peripheral biomarkers of response to simvastatin supplementation in schizophrenia. *Schizophr Res* **266**, 66-74, doi:10.1016/j.schres.2024.02.011 (2024).
114 Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185-190, doi:10.1038/nature12975 (2014).
115 Fromer, M. *et al.* De novo mutations in schizophrenia implicate synaptic networks. *Nature* **506**, 179-184, doi:10.1038/nature12929 (2014).
116 Marshall, C. R. *et al.* Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* **49**, 27-35, doi:10.1038/ng.3725 (2017).
117 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).
118 Sanders, S. J. *et al.* Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215-1233, doi:10.1016/j.neuron.2015.09.016 (2015).
119 De Rubeis, S. *et al.* Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209-215, doi:10.1038/nature13772 (2014).
120 Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nature Neuro*, doi:10.1038/s41593-020-0603-0 (2020).
121 Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* **51**, 1339-1348, doi:10.1038/s41588-019-0481-0 (2019).
122 Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482 e1411, doi:10.1016/j.cell.2019.11.020 (2019).
123 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).
124 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).
125 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).
126 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).
127 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).

128 Powell, S. K. *et al.* Induction of dopaminergic neurons for neuronal subtype-specific modeling of psychiatric disease risk. *Mol Psychiatry*, doi:10.1038/s41380-021-01273-0 (2021).
129 Hoffman, G. E. *et al.* Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nat Commun* **8**, 2225, doi:10.1038/s41467-017-02330-5 (2017).
130 Ahn, K. *et al.* High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol Psychiatry* **19**, 568-572, doi:10.1038/mp.2013.59 (2014).
131 Ahn, K., An, S. S., Shugart, Y. Y. & Rapoport, J. L. Common polygenic variation and risk for childhood-onset schizophrenia. *Mol Psychiatry*, doi:10.1038/mp.2014.158 (2014).
132 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).
133 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).
134 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).
135 Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297-301, doi:10.1038/nmeth.4177 (2017).
136 Zheng, G. X. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049, doi:10.1038/ncomms14049 (2017).
137 Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420, doi:10.1038/nbt.4096 (2018).
138 Papalexi, E. *et al.* Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat Genet* **53**, 322-331, doi:10.1038/s41588-021-00778-2 (2021).
139 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).
140 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).
141 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).
142 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).

143 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).
144 Wang, J. & Liao, Y. *WebGestaltR: Gene Set Analysis Toolkit WebGestaltR. R package version 0.4.3.*, <https://CRAN.R-project.org/package=WebGestaltR> (2020).
145 Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442, doi:10.1038/30918 (1998).
146 Subramanian, A. *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417, doi:10.1016/j.cell.2017.10.049 (2017).
147 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
148 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
149 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).
150 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).
151 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).
152 Subramanian, A., Tamayo, P., Mootha, V. K. & others. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* **102**, 15545-15550 (2005).
153 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).
154 Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* **45**, 1150-1159, doi:10.1038/ng.2742 (2013).
155 Du, J. *et al.* KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol Biosyst* **10**, 2441-2447, doi:10.1039/c4mb00287c (2014).
156 Gene Ontology, C. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res* **49**, D325-D334, doi:10.1093/nar/gkaa1113 (2021).
157 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29, doi:10.1038/75556 (2000).

TABLES

Table 1A. Top PGC3 SCZ-GWAS eGenes, prioritized by EpiXcan and epigenetic annotation, as epigenetic/regulatory (blue) and signaling (red) for pathway studies. *NEK4* , *PLCL1* , *UBE2Q2L* , *NAGA* , *FES* , *CALN1* , and *ZNF804A* are present in the pooled screen **SCZ1.**

			EpiXcan		ExN-PrediXcan			Epigenetic Annotation			
Group	**Chr**	**Gene**	**z**	**r2**	**z**	**effect_size**	**P**	**ATACseq**	**H3K27ac**	**H3K4me3**	**H3K27ac (neuron)**
regulatory	3	***NEK4***	8.64	0.26	NA	NA	NA	NA	Promoter	NA	Promoter
signaling	2	***PLCL1***	7.2	0.04	5.214354	0.403086	1.84E-07	NA	Intron	Promoter	Intron
regulatory	15	***UBE2Q2L***	6.89	0.02	NA	NA	NA	NA	Promoter	Promoter	Promoter
signaling	22	***NAGA***	6.67	0.25	6.983019	0.148533	2.89E-12	NA	Promoter	Promoter	NA
signaling	15	***FES***	6.63	0.05	3.917081	0.289907	8.96E-05	NA	Promoter	Promoter	NA
signaling	7	***CALN1***	6.57	0.07	5.356091	0.171432	8.50E-08	NA	Intron	Promoter	Intron
regulatory	2	***ZNF804A***	6.27	0.05	2.867957	5.119801	4.13E-03	NA	Promoter	Intron	NA

Table 1B. Top PGC3 SCZ-GWAS eGenes, prioritized by COLOC and PrediXcan, as synaptic (green), regulatory/epigenetic (blue), and multi-function (purple). *CALN1* , *CLCN3* , *CUL9* , *DOC2A* , *PLCL1* , *INO80E* , *SF3B1* , *SNAP91* , *TMEM219* , *ZNF823* are present in the pooled screen **SCZ2**. All 15 eGenes are present in the subsequent arrayed screen.

	GWAS					COLOC	PrediXcan		ExN-PrediXcan		
Group	**Chr**	**SNP**	**P**	**LD range**	**Gene**	**PP4**	**z**	**P**	**z**	**effect_size**	**P**
synapse	6	rs2022265	3.74E-10	83262613…85419243	***SNAP91***	0.893265	5.185282	2.16E-07	6.209717	0.434707	5.31E-10
synapse	4	rs356183	3.37E-08	89645368…91759130	***SNCA***	0.847582	-4.494898	6.96E-06	-1.209231	-0.067749	2.27E-01
synapse	16	rs3814883	8.82E-15	30016830…30034591	***DOC2A***	0.521296	6.088335	1.14E-09	5.910559	0.739169	3.41E-09
synapse	4	rs61405217	5.39E-11	170533784…170644824	***CLCN3***	0.718835	5.771783	7.84E-09	6.21805	0.659242	5.03E-10
synapse	2	rs1451488	6.72E-17	198669426…199437305	***PLCL1***	0.035058	4.910285	0.000000909	5.214354	0.403086	1.84E-07
regulatory	19	rs72986630	3.07E-10	10832225…12839819	***ZNF823***	0.999253	6.350252	2.15E-10	5.475881	0.542843	4.35E-08
regulatory	16	rs3814883	8.82E-15	29007489…31016970	***INO80E***	0.894347	7.639276	2.18E-14	7.28675	0.28085	3.18E-13
regulatory	3	rs704364	8.41E-10	62819885…64848612	***THOC7***	0.893413	-5.236222	1.64E-07	-5.694926	-0.30103	1.23E-08
regulatory	2	rs2914983	1.10E-14	197256700…199299078	***SF3B1***	0.888092	7.256793	3.96E-13	6.353697	3.288414	2.10E-10
regulatory	19	rs1858999	7.97E-14	18497024…20619093	***GATAD2A***	0.858178	-7.214675	5.41E-13	-7.297041	-0.251158	2.94E-13
proprotein convertase	15	rs4702	2.15E-23	90414642…92426654	***FURIN***	0.999943	-10.06869	7.60E-24	-8.776284	-0.52838	1.69E-18
ubiquitination	6	rs113113059	2.29E-11	42150013…44192158	***CUL9***	0.953677	5.754229	8.70E-09	5.052675	0.198153	4.36E-07
calcium signalling	7	rs2944821	1.90E-09	70244499…72912040	***CALN1***	0.913966	5.547481	2.90E-08	5.356091	0.171432	8.50E-08
metabolism	3	rs7432375	5.32E-15	134969406…137055316	***PCCB***	0.906413	-7.547429	4.44E-14	-7.973074	-0.238932	1.55E-15
signalling	16	rs3814883	8.82E-15	28952638…30984212	***TMEM219***	0.846736	6.291546	3.14E-10	6.781431	0.358578	1.19E-11

FIGURES

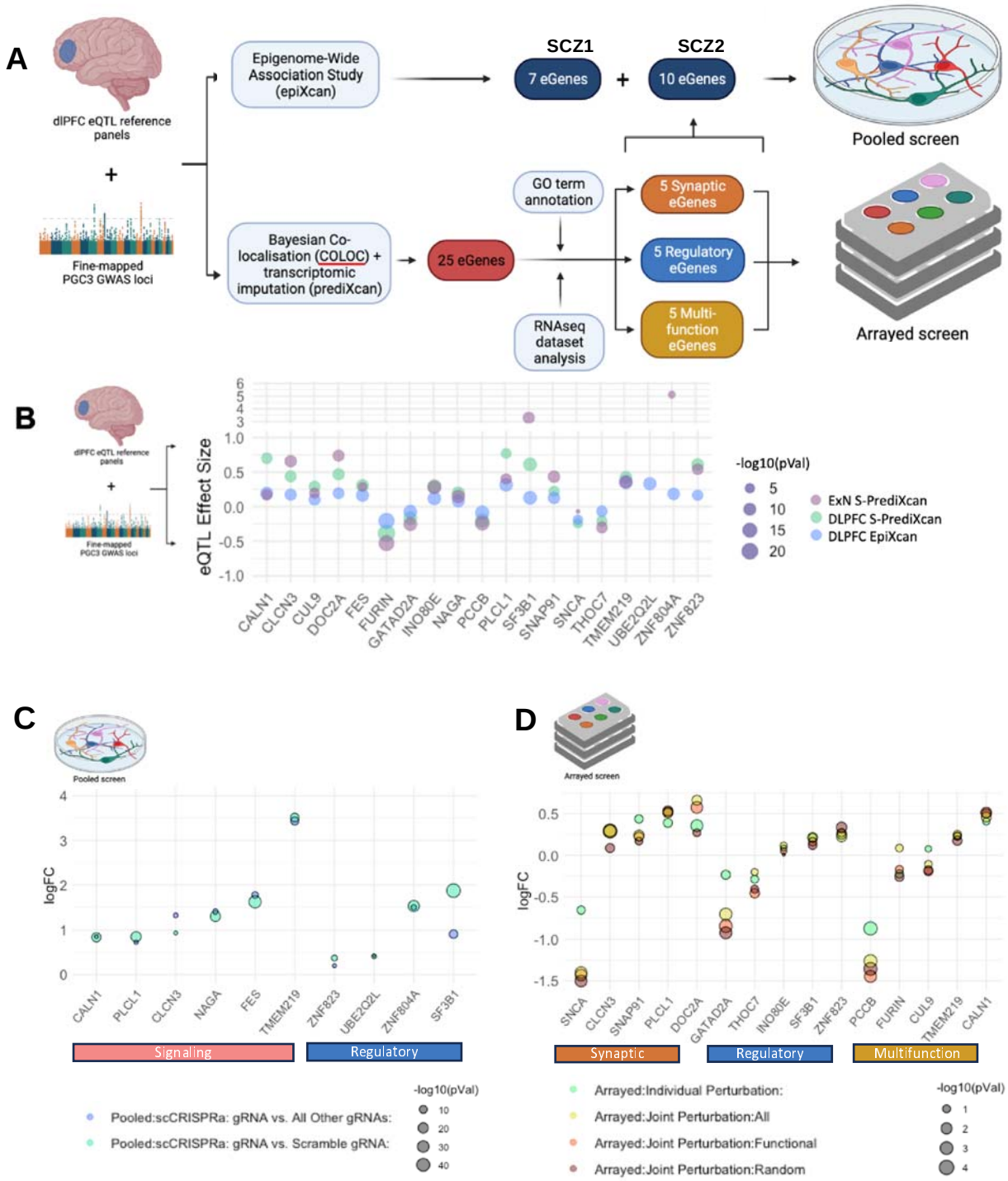
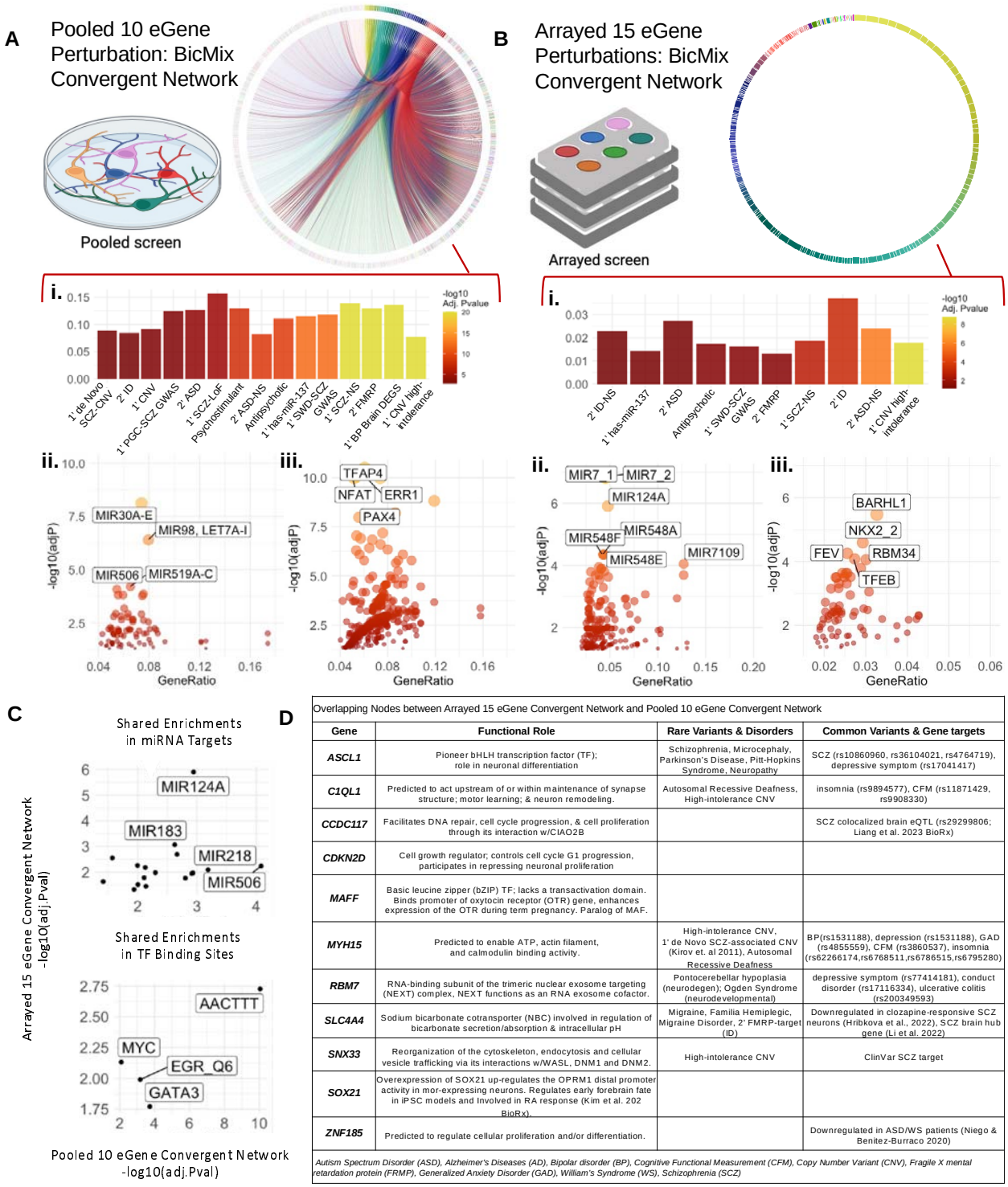


Figure 1. Prioritization and manipulation of synaptic, regulatory, and multi-function brain eGenes regulated by schizophrenia.

A. Schematic of schizophrenia eGene identification and prioritization. Schizophrenia eGenes were prioritized by fine-mapping (COLOC), transcriptomic imputation

(PrediXcan), and/or epigenomic imputation (EpiXcan) schizophrenia GWAS using post-mortem brain expression data. **B.** Effect sizes of significant eGenes from either dorsolateral prefrontal cortex (DLPFC) EpiXcan (blue), DLPFC S-PrediXcan (green) or excitatory neuron (ExN) S-PrediXcan (purple) transcriptomic imputation studies. Size of circles corresponds with the -log10(adjusted p-value) **C.** Log2(fold change) of all eGenes in the pooled experiments **SCZ1** and **SCZ2** comparing all perturbed cells of one target eGene identity to all other cells of different eGene identities (blue) or compared to only Scramble gRNA (teal). Size of circles corresponds with the -log(adjusted p-value) from a one-way pairwise Wilcox Rank Sum. **D.** Log2(fold change) of all eGenes in the arrayed experiment following single (teal) and joint perturbations across all 15 eGenes (yellow) or functional (orange) or random (maroon) sets of 5 eGenes in D21 hiPSC-NPC derived iGLUTs, using individual vectors. Size of circles corresponds with the -log(adjusted p-value) from a one-tailed t-test.



Overlapping Nodes between Arrayed 15 eGene Convergent Network and Pooled 10 eGene Convergent Network

Gene	Functional Role	Rare Variants & Disorders	Common Variants & Gene targets
ASCL1	Pioneer bHLH transcription factor (TF); role in neuronal differentiation	Schizophrenia, Microcephaly, Parkinson's Disease, Pitt-Hopkins Syndrome, Neuropathy	SCZ (rs10860960, rs36104021, rs4764719), depressive symptom (rs17041417)
C1QL1	Predicted to act upstream of or within maintenance of synapse structure; motor learning; & neuron remodeling.	Autosomal Recessive Deafness, High-intolerance CNV	insomnia (rs9894577), CFM (rs11871429, rs9908330)
CCDC117	Facilitates DNA repair, cell cycle progression, & cell proliferation through its interaction w/CIAO2B		SCZ colocalized brain eQTL (rs29299806; Liang et al. 2023 BioRx)
CDKN2D	Cell growth regulator; controls cell cycle G1 progression, participates in repressing neuronal proliferation		
MAFF	Basic leucine zipper (bZIP) TF; lacks a transactivation domain. Binds promoter of oxytocin receptor (OTR) gene, enhances expression of the OTR during term pregnancy. Paralog of MAF.		
MYH15	Predicted to enable ATP, actin filament, and calmodulin binding activity.	High-intolerance CNV, 1' de Novo SCZ-associated CNV (Kirov et. al 2011), Autosomal Recessive Deafness	BP(rs1531188), depression (rs1531188), GAD (rs4855559), CFM (rs3860537), insomnia (rs62266174,rs6768511,rs6786515,rs6795280)
RBM7	RNA-binding subunit of the trimeric nuclear exosome targeting (NEXT) complex, NEXT functions as an RNA exosome cofactor.	Pontocerebellar hypoplasia (neurodegen); Ogden Syndrome (neurodevelopmental)	depressive symptom (rs77414181), conduct disorder (rs17116334), ulcerative colitis (rs200349593)
SLC4A4	Sodium bicarbonate cotransporter (NBC) involved in regulation of bicarbonate secretion/absorption & intracellular pH	Migraine, Familia Hemiplegic, Migraine Disorder, 2' FMRP-target (ID)	Downregulated in clozapine-responsive SCZ neurons (Hribkova et al., 2022), SCZ brain hub gene (Li et al. 2022)
SNX33	Reorganization of the cytoskeleton, endocytosis and cellular vesicle trafficking via its interactions w/WASL, DNM1 and DNM2.	High-intolerance CNV	ClinVar SCZ target
SOX21	Overexpression of SOX21 up-regulates the OPRM1 distal promoter activity in mor-expressing neurons. Regulates early forebrain fate in iPSC models and Involved in RA response (Kim et al. 202 BioRx).		
ZNF185	Predicted to regulate cellular proliferation and/or differentiation.		Downregulated in ASD/WS patients (Niego & Benitez-Burraco 2020)

Autism Spectrum Disorder (ASD), Alzheimer's Diseases (AD), Bipolar disorder (BP), Cognitive Functional Measurement (CFM), Copy Number Variant (CNV), Fragile X mental retardation protein (FRMP), Generalized Anxiety Disorder (GAD), William's Syndrome (WS), Schizophrenia (SCZ)

Figure 2. Downstream target-convergent networks identified by Bayesian bi-clustering resolve distinct networks enriched for schizophrenia common and rare variant target genes and transcription factor binding motifs.

A. Convergent networks resolved across the downstream transcriptomic impacts of all ten target perturbations in the pooled screens **SCZ1** and **SCZ2** identified 1869

convergent genes with enrichments for **(i)** brain-related GWAS genes, **(ii)** transcription factor binding sites of know schizophrenia-associated TFs (TFAP4, NFAT and ERR1), and **(iii)** common and rare variant target genes. **B.** Convergent networks resolved across the downstream transcriptomic impacts of all fifteen target perturbations in the arrayed assay identified 255 convergent genes with enrichments for **(i)** miRNA targets and **(ii)** transcription factor binding sites of know schizophrenia-associated TFs (TFAP4, NFAT and ERR1), and **(iii)** common and rare variant target genes. **C.** While largely distinct, the resolved convergent networks from the arrayed and pooled screen shared 16 significant enrichments for miRNA targets and 4 significant enrichments for TF targets – many of which are thought to play a role in regulation of schizophrenia-associated genes. **D.** Overlapping nodes between the two networks were often involved in neuronal proliferation, and differentiation.

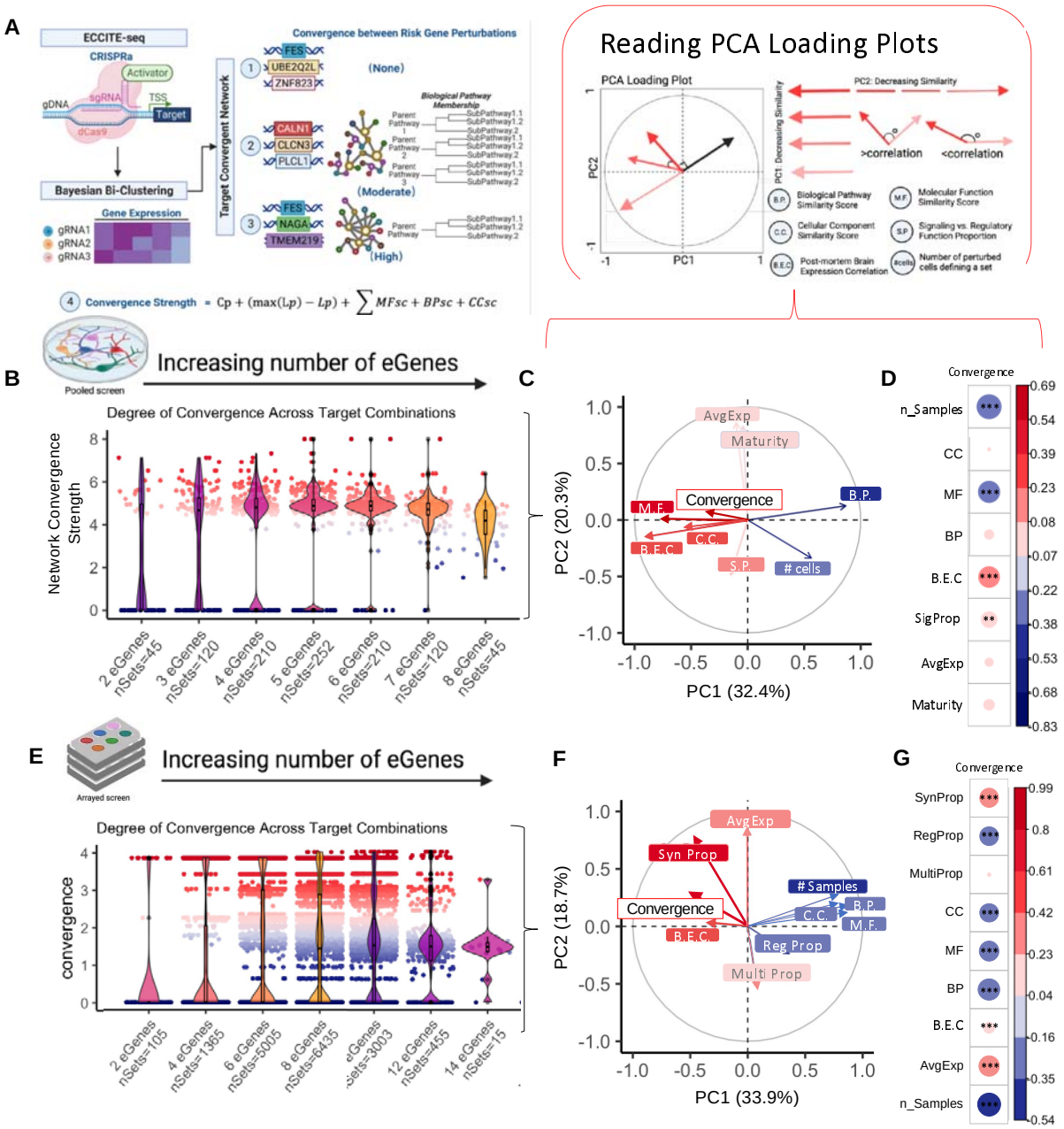


Figure 3. The degree of network convergence is influenced by functional similarity of target perturbations.

A. Defining convergence and calculating convergent network strength. *Here we define convergence as the independent development of transcriptomic similarities between separate "gene perturbations" that move towards union or uniformity of biological function*. Bayesian bi-clustering identifies co-expressed genes that are shared across the downstream transcriptomic impacts of any given eGene perturbations, thus, the resolved networks are the transcriptomic similarities between distinct perturbations**.** While bi-clustering resolves convergent gene co-expression networks, the strength of

convergence within a network can be defined by (i) the degree of network connectivity as define by two small-world network connectivity coefficients (edge density and average path length) and (ii) the degree of functional similarity or unity between genes represented within the network. Given this definition, **(1)** represents perturbations with no convergent downstream effects, **(2)** represents a network with a moderate degree of convergence because it (i) has resolved gene co-expression clusters that can be constructed into a network, (ii) has a moderate degree of network connectivity and (ii) is enriched in biological pathways with some redundancy, while **(3)** represents a highly convergent network because the degree of network connectivity is stronger and there is greater uniformity in biological pathway gene membership. Overall, we quantify the strength or degree of convergence using the function in **(4)**, where *Cp* is the edge density (the proportion of edges present given all possible edges) and *Lp* is the average path length (the mean of the shortest distance between each pair of nodes), *MFsc* is the average semantic similarity score between each pair of nodes in the network based on Molecular Function Gene Ontology, *BPsc* is the average semantic similarity score based on Biological Pathway Gene Ontology and *CCsc* is the average semantic similarity score based on Cellular Component Gene Ontology. Semantic similarity is based on the idea that genes with similar function have similar Gene Ontology annotations. Semantic similarity scores were calculated by aggregating four information content-based methods and one graph structure-based method with the R package GoSemSim. Created with BioRender.com. **B-G.** Principal component analysis (PCA) of the convergence scores, the three Gene Ontology scores (M.F., B.P., C.C.), brain expression correlation (B.E.C), and sample size across all resolved networks in both the pooled and arrayed assays revealed that some functional scores have similar influence on variance as convergence (**SI Fig. 19**). **B & E.** Distribution of the degree of convergence (x-axis) of networks across all possible combinations of 2 to 8 (y-axis; number of sets tested within each set) target perturbations from the single-cell pooled screen (**B**) and **(E)** across all possible combinations of 2 to 14 target perturbations from the arrayed screen show an influence of sample size on the ability to resolve a network. **C & F.** For both the pooled screen (**C**) and the **(F)** arrayed screen, PCs 1 (x-axis) and 2 (y-axis) explain ~62% of the variance between networks. PC loadings demonstrate the influence of each variable on the variance between networks; within the first two PCs the influence of brain expression correlation (B.E.C) and proportion of signaling genes perturbed (S.P) on PCs 1 and 2 on variance explained are more strongly related to convergence degree compared to other functional scores**.** Since degree of convergence is influenced by number of eGene perturbed we ran PCA analysis within networks of the same set size and found that the pattern of influence of signaling proportion and brain expression correlation is maintained when convergence is ranked within set size shown in **SI Fig. 20**. **D & G.** This corresponds to an overall significant positive correlation between network convergence degree, signaling/synaptic proportion of perturbed genes in a set, and brain expression correlation between genes in a set (Bonferroni adjusted p-value of Pearson's correlations: *****<0.05, ******<0.01, *******<0.001.

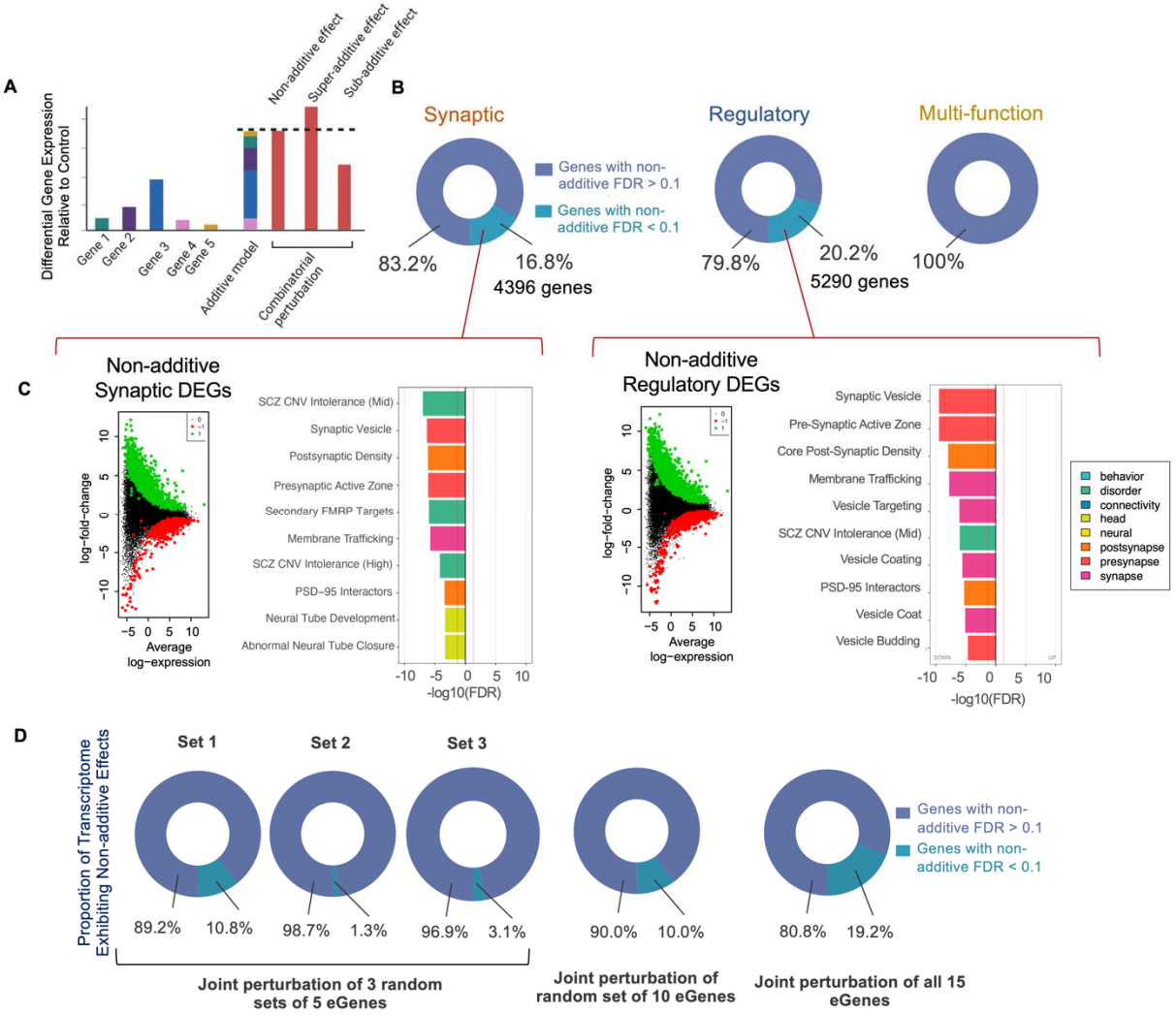


Figure 4. Perturbation of schizophrenia eGenes within functional categories results in non-additive effects on transcription impacting expression of genes linked to brain disorders and synaptic function.

A. Schematic of differential expression analysis. Individual eGene perturbations, the implementation of the expected additive model based on the latter and the measured combinatorial perturbation permitting the detection of interactive effects through comparison with the additive model. **B.** Combinatorial perturbation of synaptic and regulatory eGenes resulted in non-additive effects on expression across 16.8% (synaptic) and 20.2% (regulatory) of the transcriptome. No significant non-additive effects were seen following joint perturbation of non-synaptic, non-transcriptional regulatory eGenes. **C.** GSEA of non-additive genes in the Synaptic eGene set demonstrated significant enrichment for genes relating to brain disorders and synaptic function. GSEA of non-additive genes in the Regulatory eGene set demonstrated significant enrichment for genes relating to brain disorders and synaptic function. SCZ = schizophrenia, CNV = copy number variant, FMRP = Fragile X Mental Retardation

Protein, FDR = false discovery rate**. D.** Non-additive effects following combinatorial perturbation of sets of five, ten, and fifteen eGenes randomly assigned from the synaptic, regulatory, and multi-function eGene groups. The proportion of the transcriptome exhibiting significant non-additive effects increased with increasing numbers of perturbed eGenes (average of 5.1%, 10.0% and 19.2% of the transcriptome with non-additive FDR<0.1 after joint perturbations of five, ten, and fifteen eGenes respectively).

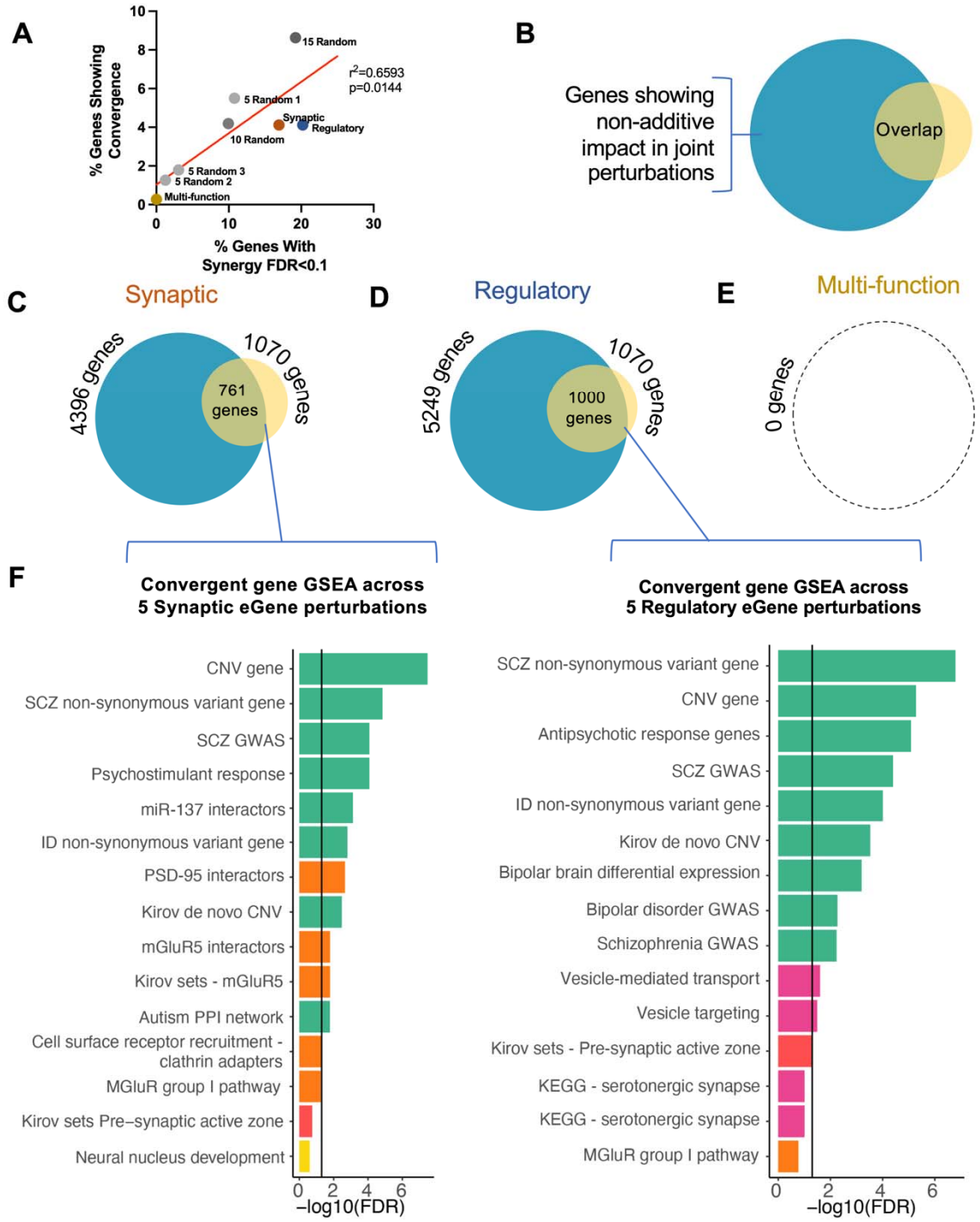


Figure 5. Convergence accounts for non-additive effects within functional pathways.

A-E. Meta-analysis of differentially expressed genes (DEGs) elicited by individual eGene perturbations for each five-gene grouping using METAL to identify DEGs that showed altered expression consistently in the same direction across all five eGene

perturbation conditions for each set of eGenes. **A.** Convergence across individual eGene perturbations is correlated with the degree of non-additive effect seen in the corresponding joint perturbation condition. Pearson's r^2 = 0.6569, p=0.0147. **B.** For each joint eGene perturbation group, non-additive impacts on transcription were compared with genes showing significant convergence across individual perturbations for the same eGene set. **C.** Evidence of convergence was found in 1070 genes across the synaptic eGene perturbations, 761 of which also exhibited non-additive effects in the additive-combinatorial comparison for the same set. **D.** Evidence of convergence was found in 1070 genes across the regulatory eGene perturbations, 1000 of which also exhibited non-additive effects in the additive-combinatorial comparison for the same set. **E.** No significant non-additive effects and only minimal convergence could be seen in eGene perturbations across functional pathways. **F.** GSEA of convergent genes in the synaptic and regulatory eGene groups demonstrated significant enrichment for genes relating to brain disorders and synaptic function**.**

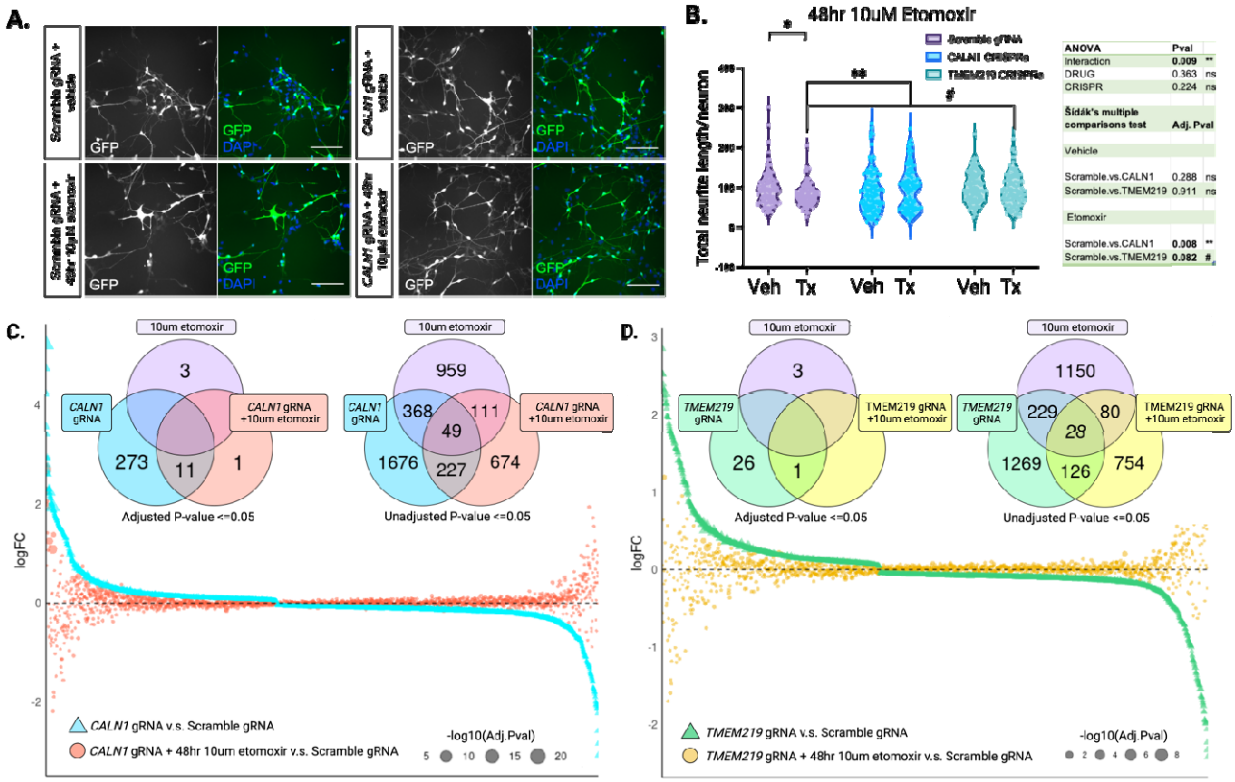


Figure 6. ***In vitro validation identifies opposing effects of in silico drug predictions and top schizophrenia eGenes.***

In vitro validation of drug-eGene phenotypic interactions. **A.** Imaging of Day 21 iGLUT neurons 48hrs after treatment with 10 µM etomoxir or vehicle after CRISPR activation of *CALN1* (compared to scramble gRNA control). Scale bar = 100 µm. **B.** 48-hour treatment with the small-molecule inhibitor of fatty acid oxidation (FAO) etomoxir significantly reduced neurite length in scramble control iGLUT neurons (Tukey's multiple comparisons test; adj. pval=0.0162). Up-regulation of *CALN1* significantly rescued the effect of etomoxir, demonstrating opposing effect of the drug etomoxir and *CALN1* activation (adj. pval=0.008). Additionally, up-regulation of *TMEM219* demonstrated a nominal opposing effect to etomoxir (adj. pval=0.082) Volcano plots of total neurite length in CRISPRa Scramble/*CALN1*/*TMEM219* gRNA cells 48hrs after treatment with 10µM etomoxir or vehicle. ANOVA results table from a two-way Grouped ANOVA and Sidak's multiple comparisons test of the interactive effects of CRISPRa perturbation and treatment with 48hr 10µM etomoxir. **C-D.** Treatment of cells perturbed with either *CALN1* CRISPRa **(C)** or *TMEM219* CRISPRa **(D)** with 10 µM etomoxir reverses or suppresses the transcriptomic impacts of the schizophrenia eGene perturbations alone (**SI Fig. 21-23**).