

# Transcriptomic and phenotypic convergence of neurodevelopmental disorder risk genes *in vitro* and *in vivo*

Received: 23 August 2024

Accepted: 25 February 2026

Published online: 24 April 2026

 Check for updates

Meilin Fernandez Garcia <sup>1,17</sup>, Kayla Retallick-Townsley <sup>1,2,17</sup>, April Pruitt <sup>3,17</sup>, Elizabeth A. Davidson <sup>4,17</sup>, Novin Balafkan <sup>1,5,17</sup>, Jonathan Warrell <sup>6</sup>, Tzu-Chieh Huang <sup>1,7</sup>, Alfred Kibowen <sup>1</sup>, Zhiyuan Chu <sup>6</sup>, Yi Dai <sup>8</sup>, Sarah E. Fitzpatrick <sup>3,9</sup>, Ran Meng <sup>6</sup>, Annabel Sen <sup>1</sup>, Sophie Cohen<sup>1</sup>, Olivia Livoti<sup>1</sup>, Suha Khan<sup>4</sup>, Charlotte Becker <sup>1</sup>, Andre Luiz Teles e Silva <sup>1,10</sup>, Jenny Liu <sup>1</sup>, Grace Dossou<sup>4</sup>, Jen Cheung<sup>1</sup>, Susanna Liu <sup>6</sup>, Sadaf Ghorbani <sup>1,5</sup>, P. J. Michael Deans <sup>1</sup>, Marisa DeCiucis <sup>7,11,12</sup>, Prashant Emani <sup>6</sup>, Huanyao Gao <sup>1</sup>, Hongying Shen <sup>7,11,12</sup>, Mark Gerstein <sup>6,13</sup>, Zuoheng Wang <sup>8,13</sup>, Laura M. Huckins <sup>1,2,3</sup>, Ellen J. Hoffman <sup>4,14,15</sup> ✉ & Kristen Brennand <sup>1,2,3,7,15,16</sup> ✉

Diverse risk genes have been identified for neurodevelopmental disorders (NDDs), but how these genes converge on similar biological pathways in neurons, and thus give rise to similar phenotypes, is unclear. Here we apply a pooled CRISPR approach to successfully target 23 NDD loss-of-function genes with roles in chromatin biology and examine convergent effects on gene expression across human induced pluripotent stem cell-derived neural progenitor cells, glutamatergic neurons and GABAergic neurons. Points of convergence vary between these cell types, with the greatest number of convergent genes and strongest convergent networks in mature glutamatergic neurons, where they broadly represent synaptic, epigenetic and, unexpectedly, mitochondrial pathways. The most convergent networks were observed between NDD genes with shared biological annotations, clinical associations and co-expression patterns in human post-mortem brain. Drugs that were predicted to reverse convergent transcriptomic signatures and/or arousal and sensory processing behaviors ameliorated behavioral phenotypes in zebrafish NDD gene mutants. These results suggest that convergent effects of NDD risk genes could provide clinically useful insights.

Autism spectrum disorder (ASD) and related developmental delay (DD) are highly heritable<sup>1</sup>. The aggregate impact of common variants of small effect reflects most genetic risk<sup>2</sup>, but in as many as a quarter of cases, potentially damaging rare inherited and *de novo* mutations in risk genes are detected<sup>3</sup>. There is significant overlap between those genes affecting ASD<sup>4</sup> and those more broadly affecting developmental<sup>5</sup> and psychiatric<sup>6,7</sup> disorders. Altogether, neurodevelopmental disorder (NDD) risk genes are typically expressed during cortical development<sup>8</sup>,

particularly the excitatory and inhibitory lineages<sup>4</sup>, and broadly split between two functional classes: neuronal communication (for example, synaptic function) and gene expression regulation (for example, chromatin regulators and transcription factors). Over half of NDD genes have roles in gene expression regulation<sup>4</sup>, sharing substantial overlap in genomic binding sites in the brain<sup>9</sup>, and with targets enriched for NDD risk genes<sup>10</sup>. Yet, evidence to support the parsimonious explanation that regulatory NDD genes preferentially target synaptic NDD genes

A full list of affiliations appears at the end of the paper. ✉ e-mail: [ellen.hoffman@yale.edu](mailto:ellen.hoffman@yale.edu); [kristen.brennand@yale.edu](mailto:kristen.brennand@yale.edu)

is lacking<sup>4</sup>. It remains unclear how disrupting NDD genes with distinct functions yields similar outcomes.

Many have proposed that diverse ASD genes have convergent downstream effects. NDD genes are co-expressed in the brain<sup>11–13</sup>, suggesting that they are regulated together and involved in related biological processes, and result in highly interconnected protein–protein interactomes<sup>14,15</sup>, indicating functional relationships between NDD proteins. Even as the number of NDD genes grows, risk genes continue to converge on a finite number of biological pathways, developmental stages, brain regions and cell types<sup>16</sup>. Disentangling these complex etiologies remains an outstanding challenge.

Given emerging evidence that epigenetic NDD genes have diverse and interconnected roles<sup>17–19</sup>, we tested the hypothesis that the nature of convergence is influenced by developmental and cell-type contexts. We report a pooled CRISPR-knockout (KO) strategy successfully targeting loss-of-function (LoF) mutations to 23 NDD genes, most with roles in chromatin biology, to examine effects on gene expression in induced neural progenitor cells (iNPCs), glutamatergic neurons and GABAergic neurons. We describe convergent networks that were distinct between cell types, strongest in neurons, where they were enriched in synaptic biology, epigenetic regulation and, unexpectedly, mitochondrial function. Machine learning tools allowed us to extend our analyses in silico across all known NDD genes, resolving how the degree of convergence between risk genes was influenced by clinical associations, biological function and co-expression patterns in the human post-mortem brain. Convergent analyses successfully predicted drugs capable of suppressing phenotypes in NDD gene zebrafish mutants, suggesting that analyses of convergent gene expression highlight behaviorally relevant pathways, and may in turn be useful in patient stratification or treatment development.

## Results

### A systematic comparison of NDD gene effects across neuronal cell types

From 102 highly penetrant LoF gene mutations associated with NDD (previously described as 58 gene expression regulation, 24 neuronal communication and 20 other)<sup>4</sup>, we used gene ontology and primary literature to identify 21 epigenetic modifiers specifically involved in chromatin organization, rearrangement and modification (*ASH1L*, *ARID1B*, *ASXL3*, *BCL11A*, *CHD2*, *CHD8*, *CREBBP*, *PPP2R5D*, *KDMSB*, *KDM6B*, *KMT2C*, *KMT5B* (*SUV420H1*), *MBD5*, *MED13L*, *PHF12*, *PHF21A*, *SETD5*, *SIN3A*, *SKI*, *SMARCC2*, *WAC*), as well as two transcription factors with putative roles as chromatin regulators (*FOXP2*, *POGZ*). Three extensively studied synaptic genes (*NRXN1*, *SCN2A*, *SHANK3*) and three under-explored neuronal communication genes (*ANK3*, *DPYSL2*, *SLC6A1*) strongly associated with NDD were added (Supplementary Fig. 1a). Many of these 29 genes differed in relative frequency of LoF gene mutations between ASD ( $n = 16$ ) and DD ( $n = 4$ )<sup>5</sup>, schizophrenia (SCZ)<sup>20</sup> and epilepsy<sup>21,22</sup> (Fig. 1a,b and Supplementary Fig. 1b), as well as general associations with genome-wide association studies (GWASs) for many neuropsychiatric disorders (MAGMA<sup>23</sup>) (Fig. 1c and Supplementary Fig. 1c), indicating a pleiotropic effect consistent with the shared genetic liability across neuropsychiatric disorders<sup>24</sup>. iNPCs, iGLUTs and iGABAs (Supplementary Fig. 2a), as well as their in vivo fetal counterparts (Supplementary Fig. 2b), expressed all genes prioritized herein<sup>25</sup>.

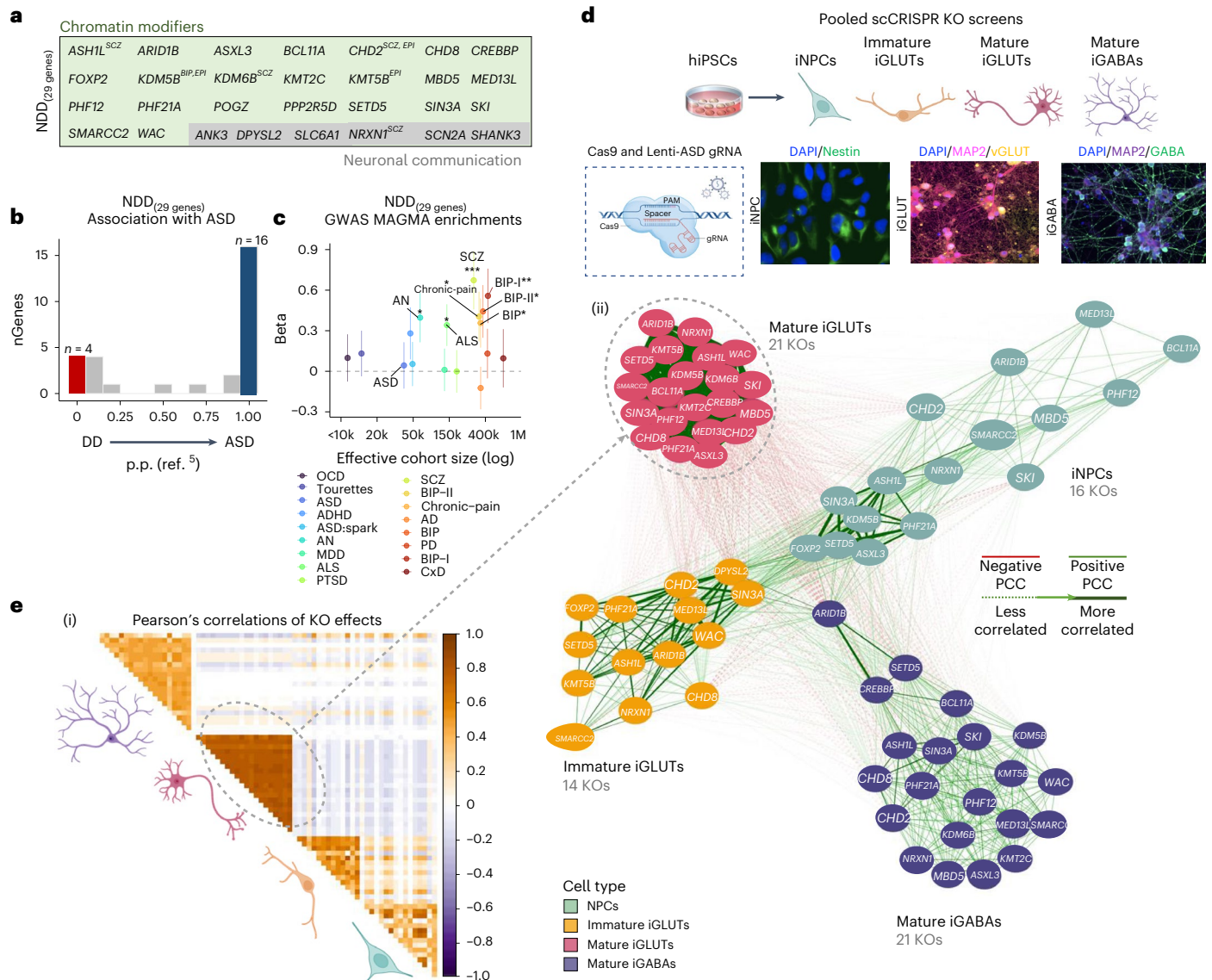
Towards resolving whether regulatory genes confer continuous or distinct periods of susceptibility across neurodevelopment, we knocked out regulatory NDD genes in neural progenitor cells (NPCs) (SNaPs<sup>26</sup>, here termed iNPCs), immature and mature glutamatergic neurons (iGLUTs)<sup>27</sup> and mature GABAergic neurons (iGABAs)<sup>28</sup> (Fig. 1d). A pooled CRISPR approach (expanded CRISPR-compatible cellular indexing of transcriptomes and epitopes by sequencing (ECCITE-seq)<sup>29</sup>) combined direct detection of single guide RNAs and single-cell RNA sequencing (scRNA-seq) to compare LoF effects across 29 NDD genes.

The CRISPR-KO library was generated from pre-validated guide RNAs (gRNAs) (three to four gRNAs per gene; Supplementary Table 1). Sequencing of the gRNA library confirmed the presence of gRNAs targeting 24 genes (*ANK3*, *ARID1B*, *ASH1L*, *ASXL3*, *BCL11A*, *CHD2*, *CHD8*, *DPYSL2*, *FOXP2*, *KMT5B* (*SUV420H1*), *KDMSB*, *KDM6B*, *KMT2C*, *MBD5*, *MED13L*, *NRXN1*, *PHF12*, *PHF21A*, *SCN2A*, *SETD5*, *SIN3A*, *SKI*, *SMARCC2*, *WAC*), but three (*DPYSL2*, *FOXP2*, *SCN2A*) were present at lower frequency (Supplementary Fig. 3b,c).

Control human induced pluripotent stem cells (hiPSCs) were induced to iNPCs, iGLUTs and iGABAs (Supplementary Fig. 3a), transduced first with lentiviral-Cas9v2 (Addgene, cat. no. 98291) and subsequently with the pooled lentiviral gRNA library 3 d before collection, at day 7 (iNPC and immature iGLUT), day 21 (iGLUT) and day 36 (iGABA) (experimental workflow, Supplementary Fig. 4a; computational workflow, Supplementary Fig. 4b; experimental validation of CRISPR editing efficiency, Supplementary Fig. 5). After filtering and quality control (Supplementary Fig. 4c–e), we resolved NDD transcriptomes for 118,436 single cells: 25,402 iNPC, 38,097 immature (day 7) iGLUT, 28,388 mature (day 21) iGLUT and 26,549 mature (day 36) iGABA. Because the original gene-expression-based clustering was driven by cellular heterogeneity, cell quality and sequencing lane effects (Supplementary Fig. 6a), independent of gRNA identity, we removed cells with high expression of subtype markers and adjusted for cellular heterogeneity (Supplementary Fig. 6b,c and Supplementary Tables 2 and 3). ‘Weighted-nearest neighbor’ (WNN) analysis assigned clusters based on both gRNA identity class and gene expression to ensure that cells assigned to a gRNA identity class demonstrated successful perturbation of the targeted NDD gene<sup>30</sup>. For those WNN clusters where most cells were assigned to a single KO target, the transcriptomic signatures were compared with nontargeting scramble control clusters. Altogether, 35,777 cells were used for downstream analyses: 12,107 iNPC, 3,171 immature iGLUT, 11,802 mature iGLUT and 8,697 mature iGABA. An average of 474 cells were assigned to each individual single guide RNA (757 iNPC, 227 immature iGLUT, 562 mature iGLUT and 414 mature iGABA), totaling 33,150 perturbed cells and 2,627 controls (882 iNPC, 90 immature iGLUT, 1,258 mature iGLUT and 397 mature iGABA). The gene expression patterns of nonperturbed iNPCs and iNeurons (>30% of all pooled cells) were significantly correlated with fetal brain cells and cortical adult neurons.

Successful perturbations (scCRISPR-KO) were identified for 23 NDD genes (Supplementary Figs. 6 and 7): 16 in iNPCs, 14 in immature iGLUT neurons and 21 in mature iGLUT and iGABA neurons (Supplementary Fig. 6). Nine NDD genes were perturbed in all four cell types (*ARID1B*, *ASH1L*, *CHD2*, *MED13L*, *NRXN1*, *PHF21A*, *SETD5*, *SIN3A*, *SMARCC2*; Supplementary Fig. 7a,b). For most NDD genes, KO in mature iGLUTs yielded the largest number of differentially expressed genes (DEGs,  $P_{\text{FDR}} < 0.05$  (FDR, false discovery rate)) (Supplementary Fig. 7b), an effect that was not driven by differences in the extent of perturbation of the NDD gene itself between cell types (Supplementary Fig. 7c(i)). The transcriptomic effects of individual NDD genes cluster by cell type: the strongest NDD gene correlations are in mature iGLUTs; that is, all nominally significant ( $P < 0.01$ )  $\log_2$  fold change (FC) DEGs are most highly correlated with each other and least correlated with the other cell types, whether relative to all scramble control cells (Fig. 1e(i),(ii) and Supplementary Fig. 7c(ii)) or to random subsets of scramble control cells (Supplementary Fig. 8a,b). DEGs across individual NDDs shared significant gene ontology enrichments (Supplementary Fig. 8c), with mature iGLUTs frequently enriched for SCZ GWAS genes (12 of 21 NDD genes), and mature iGABAs for migraine GWAS genes (8 of 21) (Supplementary Fig. 9).

Unsurprisingly, given the greater within-cell-type correlations between NDD genes and the unique pathway enrichments across cell types, very few DEGs shared significance and direction of effect for the same NDD gene perturbation across all four cell types (FDR-adjusted

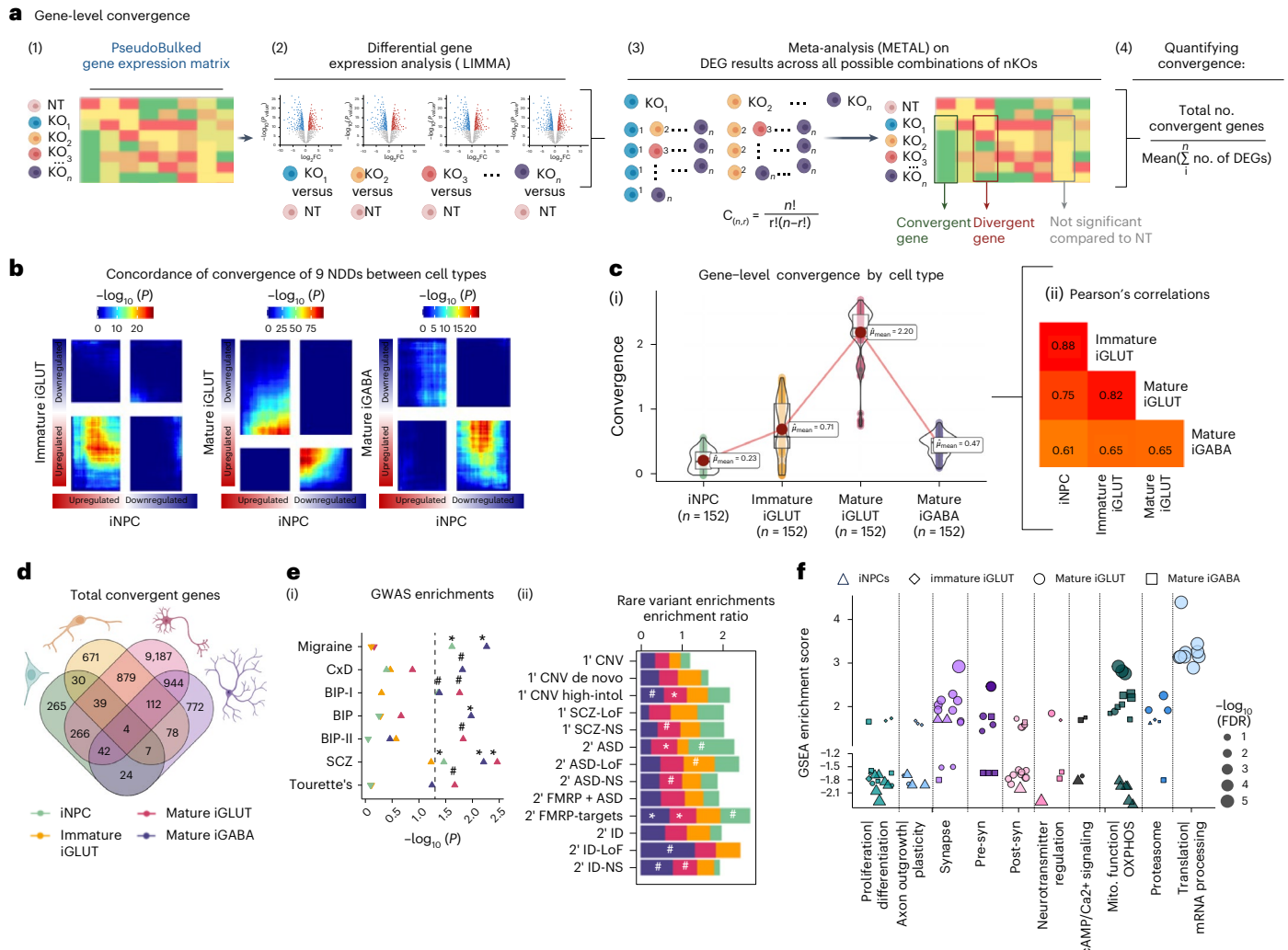


**Fig. 1** KO effects of 21 NDD risk genes are most strongly correlated in mature neurons. **a**, List of rare-variant target risk genes associated with NDDs separated by chromatin modifiers and neuronal communication genes. Bold gene names indicate strong associations with ASD based on ref. 5. Gene targets of rare variants associated with SCZ, epilepsy (EPI) and bipolar disorder (BIP) are annotated. **b**, Strength of association with ASD, as estimated by distribution of posterior probability (p.p.) scores from ref. 5; 4 of 29 NDD genes were more strongly associated with DD (blue; p.p. ≤ 0.1) while 16 of 29 were more strongly associated with ASD (red; p.p. ≥ 0.9). Further annotations of individual risk genes are shown in Supplementary Figs. 1 and 2. **c**, One-sided, positive MAGMA GSEA of targeted genes across GWAS for anorexia nervosa (AN), chronic pain, amyotrophic lateral sclerosis (ALS), SCZ, BIP, BIP-I (bipolar subtype 1) and BIP-II (bipolar subtype 2). FDR multiple testing correction was performed to adjust

for multiple gene set comparisons: #Nominal  $P < 0.05$ , \*FDR  $< 0.05$ , \*\*FDR  $< 0.01$ , \*\*\*FDR  $< 0.001$ . Error bars indicate the standard error of beta (the regression coefficient). **d**, Schematic of hiPSC-derived cell-type-specific scCRISPR-KO screen. Representative immunofluorescence for markers of NPCs (DAPI/Nestin), mature iGLUTs (DAPI/MAP2/vGLUT) and mature iGABAs (DAPI/MAP2/GABA). **e**, Transcriptomic impact of NDD gene KO represented as two-tailed Pearson's correlation across nominally significant ( $P < 0.01$ ) DEGs. (i) Pearson's correlation matrix of  $\log_2FC$  DEGs across all NDDs and cell types. (ii) Cross-cell-type correlation network diagram (based on Pearson's correlations) across NDD perturbations (number of NDD gene KO perturbations resolved indicated in parentheses); the mature iGLUT cluster was most dense, and the iNPC most sparse. Illustrations in **d** and **e** created in BioRender; Townsley, K. <https://biorender.com/rvk1zn2> (2026). PCC, Pearson's correlation coefficient.

$P_{meta} < 0.05$ , Cochran's heterogeneity  $Q$ -test  $P_{Het} > 0.05$ ; computational workflow, Supplementary Fig. 10a); in fact, the only common DEG between cell types was frequently the targeted NDD gene itself. With a more relaxed statistical threshold (nominal  $P < 0.05$ ), modest shared effects of individual NDD genes could be resolved across cell types. These effects rarely resulted in perturbation of the other NDD genes themselves (Supplementary Fig. 10b), showed very little overlap between NDD genes (Supplementary Fig. 10c) and showed no significant enrichments with psychiatric GWAS after multiple testing correction (Supplementary Fig. 10d).

**NDD gene KOs resulted in cell-type-specific convergent genes and networks that were strongest in glutamatergic neurons** 'Convergent genes' (Fig. 2) are those DEGs with significant and shared direction of effect across all NDD gene perturbations (FDR-adjusted  $P_{meta} < 0.05$ , Cochran's heterogeneity  $Q$ -test  $P_{Het} > 0.05$ )<sup>31</sup> (computational workflow, Fig. 2a). Across the nine NDD genes perturbed in all four cell types (ARID1B, ASHL1, CHD2, MED13L, NRXN1, PHF21A, SETD5, SIN3A, SMARCC2), convergence was highly cell-type-specific (Fig. 2, Supplementary Fig. 11a–c and Supplementary Data 2). Although the strength of convergence correlated across cell types (Fig. 2c(ii)), it was



**Fig. 2 | Gene-level convergence is greatest in mature glutamatergic neurons.** In total, nine NDD genes showed evidence of KO across all four cell types: *ARID1B*, *ASH1L*, *CHD2*, *MED13L*, *NRXN1*, *PHF21A*, *SETD5*, *SIN3A*, *SMARCC2*. For these nine, ‘convergent genes’ are defined as those DEGs with significant and shared direction of effect across all NDD gene perturbations. **a**, Schematic explaining cell-type-specific convergence at the individual gene level via differential gene expression meta-analysis (FDR-adjusted  $P_{meta} < 0.05$ , Cochran’s heterogeneity  $Q$ -test  $P_{het} > 0.05$ ). **b**, Convergence across nine NDD genes is unique to each cell type, using rank–rank hypergeometric (RRHO) test to explore correlation of convergent genes shared across nine NDD perturbations (RRHO score =  $-\log_{10}(P \text{ value}) \times \text{sign}(\text{fold change})$ ) between cell types. The top-right quadrant represents downregulated genes (meta-analysis  $Z$ -score  $> 0$ ) for the  $y$ -axis and  $x$ -axis cell types. The bottom-left quadrant represents upregulated convergent genes (meta-analysis  $Z$ -score  $< 0$ ) for the  $y$ -axis and  $x$ -axis cell types. Significance is represented by color, with red regions representing significantly convergent gene expression. **c**, (i) The average strength of convergence, measured as the ratio of convergent genes to the average number of DEGs across all 152 unique combinations of 2–5 genes from the nine NDD genes, was highest in iGLUTs. (ii) The magnitude of convergence between the same NDDs tested in different cell types was highly positively correlated (two-tailed Pearson’s correlation, Holm’s multiple testing correction was performed,  $P_{holm} < 2.2 \times 10^{-16}$ ), with the strongest relationship between immature and mature iGLUTs. In the box plots, the median is represented by the line (center) and the mean as the red point. The lower and

upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The upper and lower whiskers extend up to  $1.5 \times$  interquartile range (IQR). All data points are plotted individually. **d**, Venn diagram representing the absolute overlap of cell-type-specific convergent genes shared across nine NDDs (regardless of whether convergent genes were perturbed in the same direction between cell types). **e**, (i) One-sided, positive MAGMA gene set enrichment  $-\log_{10}(P \text{ value})$  of cell-type-specific convergence and GWAS-risk-associated genes with significance after multiple testing correction indicated as follows: #unadjusted  $P \leq 0.05$ , \*FDR  $\leq 0.05$ , \*\*FDR  $< 0.01$ , \*\*\*FDR  $< 0.001$ . The direction of the triangles indicates a positive (upwards triangle) or negative (downwards triangle) enrichment beta. (ii) Over-representation analysis (ORA) enrichment ratios of cell-type-specific (color of bars) convergence and rare-variant target genes. Significance after multiple testing correction indicated as follows: #unadjusted  $P \leq 0.05$ , \*FDR  $\leq 0.05$ , \*\*FDR  $< 0.01$ , \*\*\*FDR  $< 0.001$ . **f**, Two-sided GSEA identified downstream pathways involved in neural proliferation, neurite outgrowth, synaptic vesicle transport and mitochondrial function as cell-type-specific targets of convergent genes across nine NDDs. FDR multiple testing correction was performed. Results in the figure panel were filtered for pathways with nominal  $P < 0.05$ . Normalized GSEA enrichment scores represent the direction of enrichment based on the meta-analyzed  $Z$ -score for each convergent gene. Cell type is represented by shape and the size of each point represents the  $-\log_{10}(\text{FDR})$ . Illustrations in **a** created in BioRender; Townsley, K. <https://biorender.com/efkzff6> (2026). nKOs, number of KO genes tested for convergence; NT, XXX.

greatest in mature iGLUTs (quantified as the ratio of convergent genes to the average number of DEGs across all 152 unique 2–5-gene combinations of these nine NDD genes) (Fig. 2c(ii)). While the ‘top’ convergent gene was unique for each cell type (Supplementary Table 4), >50% of

convergent genes in NPCs (52%), immature iGLUTs (57%) and mature iGABAs (56%) overlapped with convergent genes in mature iGLUTs; of note, shared convergent genes were not necessarily perturbed in the same direction between cell types (Fig. 2d). Mature iGLUTs had the

greatest total number of convergent genes (11,473); however, immature iGLUTs had the highest ratio of convergent genes after correction for the number of DEGs across perturbations.

Convergent genes were enriched for SCZ GWAS loci (MAGMA<sup>23</sup>, FDR < 0.05) (Fig. 2e(i)), rare ASD and Fragile X Mental Retardation 1 protein (FMRP) target genes (FDR < 0.05) (Fig. 2e(ii)), and pathways involved in neurodevelopment, mitochondrial function and translational regulation (Supplementary Fig. 12). When tested again across the 21 NDD genes perturbed in both iGLUTs and iGABAs, mature iGLUTs again showed the largest absolute number of convergent genes (iGLUTs, 10,557, Supplementary Fig. 13a; iGABAs, 892, Supplementary Fig. 13b). Intriguingly, although convergent genes were highly cell-type-specific, those NDD gene combinations that were highly convergent in one cell type were likely to be convergent in others; in neurons, top convergent sets most frequently included *ARID1B*, *SETD5* and *NRXN1* (Supplementary Fig. 11d).

Given that the biological impact of convergence is likely to be impacted by the strength of shared gene regulatory relationships and functions, we re-examined convergence within the framework of co-expression networks (Bayesian bi-clustering). ‘Convergent networks’ (Fig. 3) are co-expressed genes that share similar expression patterns across NDD gene perturbations<sup>31</sup> (computational workflow, Fig. 3a). The network connectivity score (‘network convergence’) informs the strength and composition across cell types (that is, networks with more interconnectedness and containing genes with greater functional similarity have increased convergence). Convergent networks generated from the nine NDD genes perturbed in all four cell types (Fig. 3b) or across the 21 NDDs genes in both iGLUTs and iGABAs (Supplementary Fig. 13c) revealed the greatest convergent network strength in iGLUTs. Network-level convergence was weakly correlated between cell types (Fig. 3c); the number of convergent unique network nodes was greatest in iGLUTs, distinct across cell types (Fig. 3d, Supplementary Tables 5 and 6 and Supplementary Data 2) and significantly enriched for rare variants linked to SCZ and ASD (Fig. 3e and Supplementary Tables 5 and 6). Convergent networks in iNPCs highlighted pathways associated with neurogenesis (for example, cell cycle, cell division, EPO signaling) (Fig. 3f), while in mature iGLUTs they were enriched for synaptic function (transmembrane transport and receptor signaling, secretory vesicles, SNARE complex) (Fig. 3g).

### Convergent networks are strongest between NDD genes with shared co-expression patterns in the post-mortem brain, biological annotations (synaptic or epigenetic) and clinical outcomes (ASD or DD)

To resolve the extent to which functional similarity and co-expression patterns between NDD genes predicted convergence, we trained a prediction model (random forest linear regression)<sup>32</sup> using 70% of our data, evaluated it using 30% of our data and validated in an external dataset<sup>31</sup> (computational workflow, Fig. 4a; model predictor variables, Fig. 6b; more information, Supplementary Figs. 14 and 15). Cell type, brain co-expression (dorsolateral prefrontal cortex (DLPFC)) and functional similarity (that is, gene ontology) correlate with convergence (Fig. 4c) and well-predicted gene-level convergence (97% variance explained; root mean squared error (RMSE) = 0.021) (Fig. 4d(i)) and moderately predicted network-level convergence (53% variance explained; RMSE = 0.73) (Fig. 4d(ii)). Our trained model accurately predicted gene-level (Pearson’s  $R = 0.998$ ,  $P < 0.001$ , RMSE = 0.15) (Fig. 4e(i) and Supplementary Fig. 15c) and network-level convergence in our testing set ( $R = 0.72$ ,  $P < 2.2 \times 10^{-16}$ , RMSE = 0.85) (Fig. 4e(ii) and Supplementary Fig. 15d), and performed moderately well in predicting network-level convergence ( $R = 0.26$ ,  $P < 0.001$ , RMSE = 0.68) (Fig. 4f(ii) and Supplementary Fig. 15e(ii)) and to a lesser extent gene-level convergence ( $R = 0.14$ ,  $P < 0.001$ , RMSE = 1.75) (Fig. 4f(i) and Supplementary Fig. 15e(i)) in the external dataset.

To query whether convergence reflected clinical associations to ASD or DD, we again quantified convergence as the ratio of convergent genes to the average number of DEGs (Fig. 2e), here across all (2–5-gene) combinations of all NDD genes perturbed in each cell type (for example, 27,824 unique combinations of 21 NDD genes in iGLUTs and iGABAs; Supplementary Fig. 16a). Convergence, both gene-level (Supplementary Fig. 16a,c) and network-level (Supplementary Fig. 16b,d), was greater between genes with stronger associations to ASD compared with DD<sup>5</sup>, particularly in mature neurons (Supplementary Fig. 16e,f). Yet, this analysis was limited by the relatively small numbers of predominantly ASD ( $n = 16$ ) and DD ( $n = 4$ ) included in our dataset (Fig. 1b).

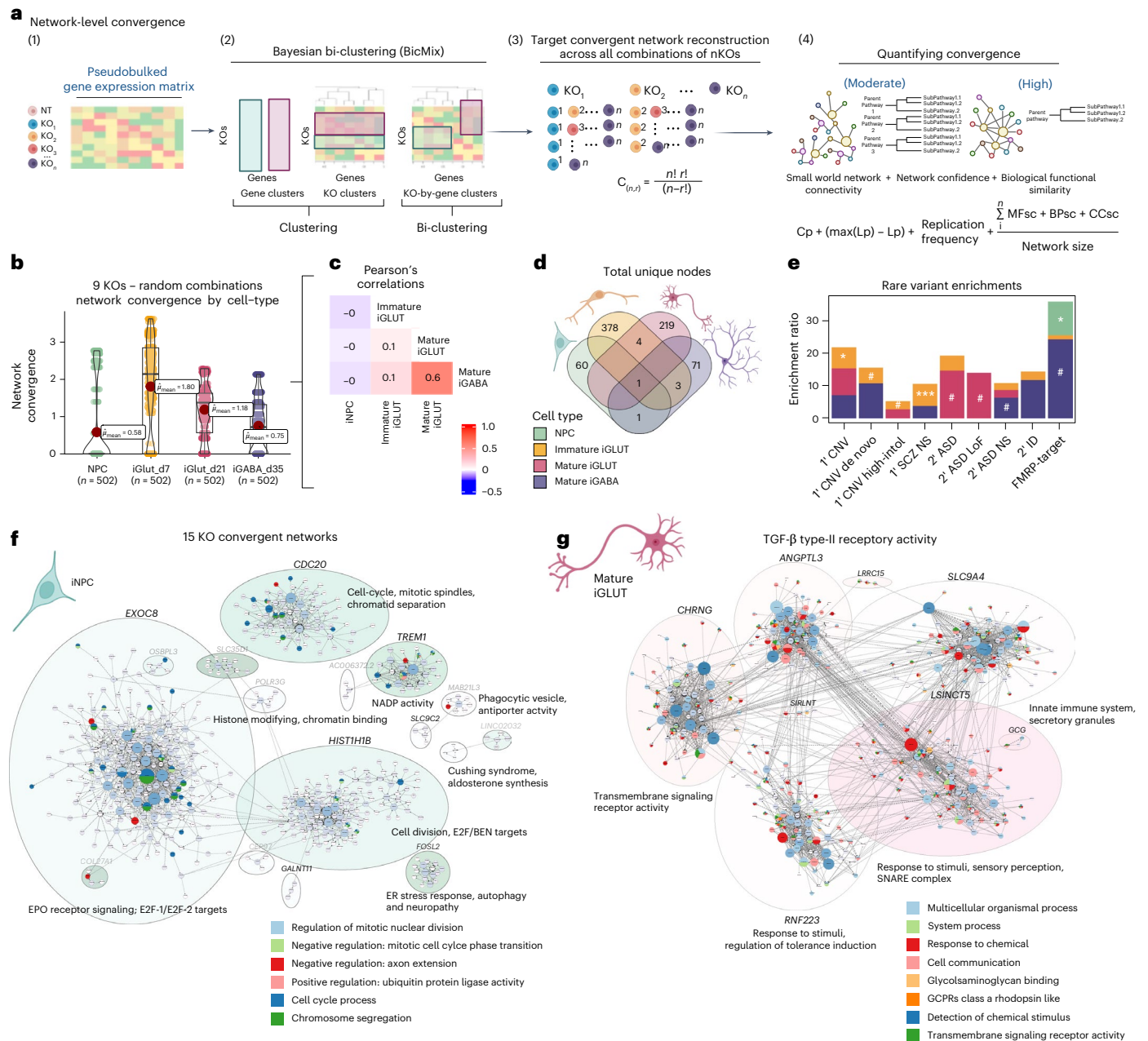
To extend our comparisons of convergence across larger sets of NDD genes, particularly those clinically defined as predominantly ASD or DD genes<sup>5</sup>, or those with biologically annotated synaptic or epigenetic roles<sup>4</sup>, we asked whether it was possible to train a machine learning model to predict cell-type-specific impacts of CRISPR-KO of all 102 NDD genes<sup>4</sup>. An integrative linear network of cell-type phenotypes (LNCTP) model, previously trained on >2.8 million nuclei from the prefrontal cortex (PFC) across 388 individuals, accurately imputed single-cell expression following simulated perturbations<sup>33</sup>. By conditioning the LNCTP model on reduced expression of the 29 NDD genes (Fig. 5a), we resolved downstream gene expression changes within three in silico post-mortem brain network models (bulk PFC tissue, excitatory neurons only and inhibitory neurons only), and compared these changes with convergent genes identified by in vitro scCRISPR-KO; the LNCTP model better replicated experimental iGLUT data (Fig. 5b).

Expanding in silico LNCTP model predictions of downstream gene expression changes to all 102 NDD genes (Fig. 5c–g and Supplementary Fig. 17) demonstrated greatest concordance with in vitro convergence in excitatory neurons, consistent with our findings that convergence was greater in iGLUTs (Figs. 2c and 3d). This concordance was more pronounced for synaptic NDD genes ( $n = 24$ ) relative to regulatory genes ( $n = 58$ ) (Fig. 5c). Predominantly ASD genes ( $n = 50$ ) had greater concordance with convergence in excitatory neurons (Fig. 5c), whereas predominantly DD genes ( $n = 40$ ) were moreso in inhibitory neurons (Fig. 5c). Overall, across functional or clinical categories, despite limited overlap in specific convergent genes (Fig. 5d) and terms (Fig. 5e,f), there was overall enrichment for synaptic, epigenetic and mitochondrial biology (Fig. 5g), consistent with in vitro scCRISPR-KO (Fig. 2f).

### Convergent genes and networks in glutamatergic neurons targeted synaptic, epigenetic and mitochondrial biology

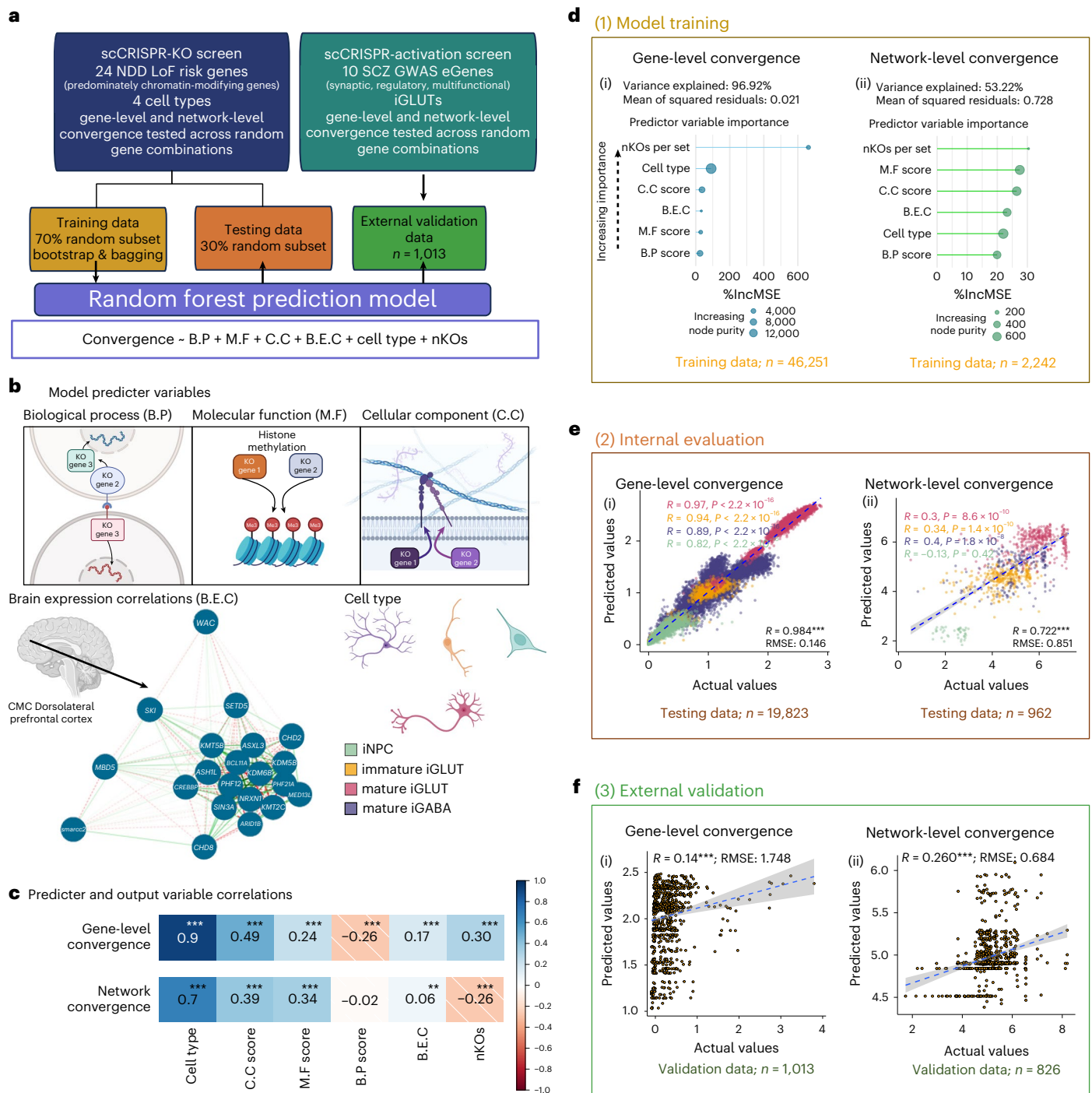
Convergent genes and networks revealed cell-type-specific disease (Fig. 2e) and functional enrichments (Figs. 2f, 5g and 6a,b), many consistent with established NDD etiology in neurogenesis<sup>18,34–41</sup> and synaptic biology<sup>14,15,42,43</sup>. For example, iNPCs were significantly enriched for pathways involved in proliferation and differentiation, whereas mature iGLUTs showed unique enrichments in neuronal communication (for example, pre-synaptic function) and regulation of gene expression (for example, messenger RNA processing and protein translation). Unexpectedly, both mature iGLUT and iGABA neurons were enriched for mitochondrial biology (for example, oxidative phosphorylation: mature iGLUTs: NES = 2.8,  $P < 2.2 \times 10^{-16}$ , FDR < 0.001; mature iGABAs: NES = 1.67,  $P = 0.023$ , FDR < 0.05).

To functionally validate the impact of LoF mutations on convergent pathways, we selected five high-confidence NDD genes (*KMT5B*, *NRXN1*, *CHD8*, *ASH1L*, *ARID1B*) representing distinct functional clusters (for example, chromatin remodelers, synaptic genes) for initial testing. We utilized a well-established doxycycline-inducible Cas9 system (iCas9)<sup>44</sup> that allows for a robust comparison of gRNA-transduced iCas9 ‘induced’ (KO) cells with matched gRNA- and iCas9-transduced cells lacking expression of iCas9 (no doxycycline) to serve as an ‘uninduced’



**Fig. 3 | Network-level convergence resolves cell-type-specific and developmental-specific node genes.** ‘Convergent networks’ are co-expressed genes that share similar expression patterns across NDD gene perturbations, here resolved for the nine NDD KO resolved across all four cell types: *ARID1B*, *ASH1L*, *CHD2*, *MED13L*, *NRXN1*, *PHF21A*, *SETD5*, *SIN3A*, *SMARCC2*. **a**, Schematic explaining cell-type-specific convergence at the network level using Bayesian bi-clustering and unsupervised network reconstruction. **b**, Strength of network convergence across all random combinations of nine NDD KO perturbations by cell type. (i) The mean strength of network convergence is significantly different by cell type, with the highest convergence present in immature iGLUTs. The same KO combinations tested in one cell type may not resolve convergence in another cell type. Each point represents a resolved network, and its calculated convergence strength. Dots that represent the same combinations of KO perturbations, but tested in each cell type, are connected by a line. In the box plots, the median is represented by the line (center) and the mean as the red point. The lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles). The upper and lower whiskers extend from the hinge to the largest or smallest value up to  $1.5 \times$  IQR. All data points are plotted individually. **c**, Convergent network strength was most correlated between mature iGLUTs and iGABAs (two-tailed Pearson’s correlation test with Holm’s multiple testing correction;  $PCC = 0.6$ ,  $P_{Holm} < 2.2 \times 10^{-16}$ ). Convergent network

strength in iNPCs was not correlated with network strength in neurons. **d**, Venn diagrams of the total number of unique node genes within convergent networks for each cell type. The lack of overlapping node genes between cell types (**d**), as well as the weak correlations of convergence strength between immature and mature cell types (**c**), suggest greater cell-type specificity in the magnitude of network-level convergence compared with gene-level convergence. **e**, One-sided enrichment ratios from ORA of cell-type-specific (color of bars) convergent node genes for rare-variant targets. FDR-based multiple testing correction was performed: #unadjusted  $P \leq 0.05$ , \*FDR  $\leq 0.05$ , \*\*FDR  $< 0.01$ , \*\*\*FDR  $< 0.001$ . **f, g**, Representative cell-type-specific network plots for convergence across 15 genes (*ARID1B*, *ASH1L*, *ASXL3*, *BCL11A*, *KDMSB*, *CHD2*, *MBD5*, *MED13L*, *NRXN1*, *PHF12*, *PHF21A*, *SETD5*, *SIN3A*, *SKI*, *SMARCC2*) from iNPCs (**f**) and mature iGLUTs (**g**). Network genes were filtered for protein-coding genes, clustered and annotated based on the primary node gene for each cluster. GSEA of the networks identified unique functions by cell type. Convergent networks in iNPCs were enriched for pathways associated with neurogenesis (for example, cell cycle, cell division, EPO signaling), and in mature iGLUTs for pathways associated with synaptic function (transmembrane transport and receptor signaling, secretory vesicles, SNARE complex). Illustrations in **a** created in BioRender; Townsley, K. <https://biorender.com/efkzzf6> (2026).



baseline control. Five of five NDD genes (*KMT5B*, *NRXN1*, *CHD8*, *ASHIL*, *ARID1B*) tested in iCas9 NPCs (CD184<sup>+</sup>/CD133<sup>+</sup> NPCs) in arrayed format revealed one or more effects on proliferation (Ki-67; Fig. 6c and Supplementary Fig. 18a), neurogenesis (NPCs: CD184<sup>+</sup>/CD133<sup>+</sup>/CD271<sup>+</sup>; neurons: CD184<sup>+</sup>/CD44<sup>+</sup>/CD24<sup>+</sup>; Supplementary Fig. 18b) and/or gliogenesis (astrocytes: CD184<sup>+</sup>/CD44<sup>+</sup>; Supplementary Fig. 18c) that varied between genes. Likewise, a pooled CRISPR analysis in iCas9 cortical organoids confirmed effects on neurogenesis, again with variable effects between NDD genes (Supplementary Fig. 19).

To assess how loss of NDD-associated genes affects mitochondrial function, we performed a pooled CRISPR-KO screen using a nearly identical library (same backbone, guide density and control set) in the HI-iCas9 line. Transduced cells were differentiated into NPCs and iGLUTs by day 21, stained with the mitochondrial inner membrane

potential-sensitive dye JC-1 and sorted by fluorescence-activated cell sorting (FACS) into high- (PE-high) and low- (FITC-high) membrane potential fractions, following amplicon sequencing to quantify gRNA representation in each fraction (Fig. 6d). Of the 15 KOs, ten resulted in elevated mitochondrial membrane potential in both NPCs and iGLUTs, and the remaining five caused cell-type-specific impacts on mitochondrial membrane potential. Pathway enrichment of the ten NDD genes that increased mitochondrial membrane potential revealed a convergence on chromatin remodeling complexes, microRNAs and transcription factors.

Given their strong phenotypes and unexpected enrichment of mitochondrial pathways, we selected three NDD KOs (*ASHIL*, *ARID1B*, *NRXN1*) for further testing of mitochondrial effects in arrayed format, using a platform with the ability to resolve dose-dependent changes

**Fig. 4 | Functional similarity and brain co-expression between NDD genes predict gene-level and network-level convergence, with unique influences by cell type.** **a**, Schematic for training random forest models for gene- and network-level convergence with external validation in an SCZ CRISPRa screen. **b**, Predictor variables included in the model include scores of functional similarity, DLPFC brain co-expression, cell type and the number of KOs (B.P score, semantic similarity of GO: Biological Process membership between KO genes; C.C score, semantic similarity of GO: Cellular Component membership between KO genes; M.F score, semantic similarity of GO: Molecular Functions membership between KO genes; B.E.C, DLPFC expression correlations between KO genes). **c**, Two-tailed Pearson's correlations of predictor variables and gene-level and network-level convergence (\*\* $P_{\text{Bonferroni}} \leq 0.01$ , \*\*\* $P_{\text{Bonferroni}} \leq 0.01$ ). **d**, Functional similarity, brain co-expression, cell type and the number of KOs assayed strongly predicted gene-level convergence (97% variance explained by the model; mean of squared residuals = 0.02) and moderately predicted network-level convergence (53% variance explained; mean of squared residuals = 0.73). (i), (ii) Importance of each of the predictor variables was assessed by two metrics: the percentage mean increase in squared residuals (%IncMSE) and the increase in node purity. In the model, number of KO genes in a set is the most important predictor of convergence based on %IncMSE, but not node purity. However, the impact of nKOs on gene-level convergence is much stronger, likely

an artifact of the method used for measuring convergence. For network-level convergence, each variable has an IncMSE of 20–30%. **e**, Internal evaluation of the model using 30% of the original data resulted in high concordance between convergence predicted by the model and the measured convergence. Predicted gene-level (i) (gene-level convergence:  $n = 19,823$ ; two-tailed Pearson's  $R = 0.984$ ;  $P_{\text{Holm}} < 1 \times 10^{-150}$ ; root mean squared error (RMSE) = 0.15) and network-level (ii) convergence (network-level convergence:  $n = 962$ ;  $\rho = 0.722$ ;  $P_{\text{Holm}} = 2.4 \times 10^{-149}$ ; RMSE = 0.85) by the model strongly correlated with the measured convergence in the testing sets. Correlations of predicted versus accrual convergence values are color-coded by cell type with corresponding color-coded correlations and  $P$  values listed in the upper-right corners of the scatterplots. Lines represent the smooth conditional mean with 95% confidence bands. **f**, External validation of the random forest prediction in an independent scCRISPRa screen of SCZ target genes (10 perturbations) predicted showed moderate, but significant, correlation between convergence by the model and the measured convergence. (i) Gene-level convergence ( $n = 1,013$ , Pearson's  $R = 0.14$ ,  $P = 1.1 \times 10^{-5}$ , RMSE = 1.748). (ii) Network-level convergence ( $n = 826$ , Pearson's  $R = 0.26$ ,  $P = 2.9 \times 10^{-14}$ , RMSE = 0.68). Lines represent the smooth conditional mean with 95% confidence bands. Illustrations in **b** created in BioRender; Townsley, K. <https://biorender.com/5xlpkxm> and <https://biorender.com/rvk1zn2> (2026).

in mitochondrial fragmentation following pharmacological insults (Supplementary Fig. 20). By high content imaging, we analyzed and quantified  $1 \times 10^4$  mitochondria per genotype, with morphological measurements taken for mitochondrial (TOMM20-positive) volume, surface area and sphericity (roundness) as well as total OXPHOS complex, within neuronal dendrites (MAP2-positive) of mature (day 21) iGLUTs. Among the three NDD KOs, *ARID1B* resulted in increased mitochondrial networking (indicated by decreased mitochondrial sphericity and increased branch length; one-way analysis of variance (ANOVA), Šidák's, adjusted  $P = 0.0213$  and  $P = 0.0081$ , respectively) concomitant with increased levels of OXPHOS proteins (one-way ANOVA, Šidák's, adjusted  $P = 0.0024$ ) (Fig. 6e and Supplementary Fig. 21a), overall consistent with increased mitochondrial efficiency. Second, we tested oxygen consumption using the Seahorse Cell Mito Stress test. NRXN1 KO resulted in increased coupled and maximal respiration in iGLUTs (one-way ANOVA,  $P < 0.05$ ; Fig. 6f); increased mitochondrial reliance, in the absence of fused mitochondria, with elevated OXPHOS protein levels points to a possible metabolic overload due to reduced mitochondrial efficiency (Fig. 6e). In contrast, *ARID1B* and *ASH1L* KOs did not show significant changes in these Seahorse parameters (Supplementary Fig. 21b,c). Taken together, both *ARID1B* and *NRXN1* KO neurons show evidence of increased mitochondrial activity, *ARID1B*-KO through enhanced fusion and elevated expression of OXPHOS complexes, and *NRXN1* KO by increasing OXPHOS activity to meet ATP demands. As observed for neurogenesis in iNPCs,

single-gene KO iGLUTs confirmed convergent effects on mitochondrial biology, finding distinct but related phenotypes between NDD genes.

### Pharmacological targeting of convergent genes reversed behavioral phenotypes in mutant zebrafish

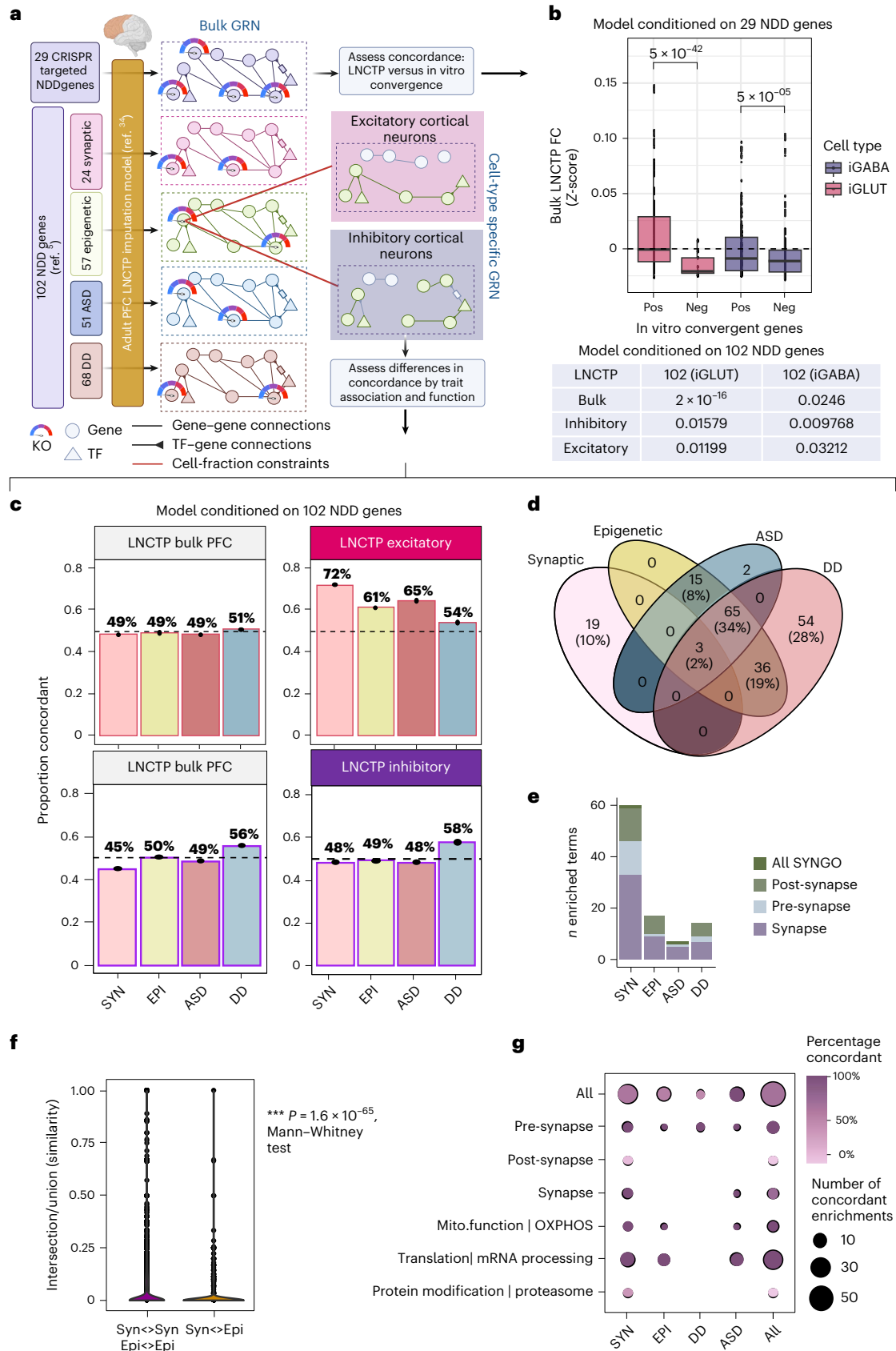
By design, in vitro models substantially limit the complexity of the observed impact of NDD genes, lacking higher circuit-level effects. Towards applying molecular convergence in vitro to explore the mechanisms of phenotypic convergence in vivo, the convergence of sets of NDD genes was next explored on the basis of shared behavioral effects in zebrafish mutants (Fig. 7 and Supplementary Tables 7 and 8). A comprehensive in vivo high-throughput, automated behavioral analysis in larval zebrafish<sup>41</sup> revealed clear stratification of NDD genes based on basic arousal and sensory processing behaviors (Fig. 7a and Supplementary Fig. 22). Given that zebrafish brain gene expression was significantly correlated with in vitro human-derived mature neurons (Fig. 7b and Supplementary Fig. 23), we asked whether behavioral stratification of NDD mutants in larval zebrafish can be attributed to molecular convergence. For 15 NDD genes for which we have matched behavioral and molecular analyses, zebrafish stable mutant lines and CRISPR F0 mutants were clustered based on 24 sleep–wake and visual-startle parameters, yielding four distinct clusters of genes: set 1 (*nrxn1a*, *mbd5*, *kdm5bab*), set 2 (*phf12ab*, *skiab*, *chd2*, *smarcc2*), set 3 (*kdm6bab*, *kmt5b*, *kmt2cab*) and set 4 (*wacab*, *arid1b*, *phf21aab*, *chd8*, *ash1l*) (Fig. 7a and Supplementary Data 3). Gene-level convergence

**Fig. 5 | LNCTP predicts effects of convergent genes in silico.** **a**, LNCTP imputation and perturbation model: an energy-based network model is trained to impute bulk and cell-type-specific expression data in the PFC over a population of post-mortem individuals from PsychENCODE using a panel of 1,325 genes and embedded cell-type-specific GRNs (LNCTP in silico model); a chosen gene is then perturbed by fixing its expression, and the effects on other genes are predicted by the model; in silico category-specific convergent genes are then identified by analyzing the FCs across subjects (LNCTP simulating perturbations). **b**, A perturbation-conditioned version of the LNCTP model applied negative perturbation to mimic the effect of the 29 CRISPR-KOs and predict downstream gene expression changes in the adult PFC. Predicted in silico log FCs for the in vitro positive and negative convergent genes across the 29 CRISPR perturbations, in bulk, excitatory and inhibitory neuron networks evaluated if in vitro convergence was replicated in adult PFC prediction models (LNCTP simulating perturbations, two-tailed  $t$ -test  $P$  values shown). In the box plots, center represents median, hinges represent 1st and 3rd quartiles. The upper and lower whiskers extend up to  $1.5 \times \text{IQR}$ . All data points are plotted

individually. **c**, A perturbation-conditioned version of the LNCTP model applied a negative perturbation to mimic 102 NDD KOs and evaluate whether convergence measured across 29 NDDs in vitro remained convergent in a larger set. Proportion of genes showing same direction FCs in in silico and in vitro perturbations across classes of perturbation and cell type (left), and the intersection of convergent in silico genes across classes of perturbation (LNCTP in silico convergent genes, synaptic–epigenetic genes reduced and ASD–DD genes enriched,  $P < 1 \times 10^{-3}$ , two-tailed hypergeometric test). **d**, Venn diagram of in silico convergent genes across all categories by clinical (ASD versus DD) or functional (synaptic versus epigenetic) annotation. **e**, Number of terms enriched for convergent genes across all categories for 102 in silico perturbations. **f**, Semantic distance of pairs of enriched terms within or between sets determined by synaptic and epigenetic convergent gene rankings (LNCTP semantic distance test, two-tailed Mann–Whitney test). **g**, Percentage of concordant genes in each perturbation and ontology category within the leading-edge enriched genes (LNCTP in silico convergent genes). Illustrations in **a** created in BioRender; Townsley, K. <https://biorender.com/6jv3bge> (2026). TF, transcription factor.

between NDD genes in these sets was distinct (Supplementary Table 9), largely nonoverlapping between cell types and stronger in mature iGLUTs than mature iGABAs (Fig. 7c). Across behavioral sets, rare ASD, SCZ and intellectual disability LoF genes were enriched primarily in

iGLUTs, with all sets converging on FMRP targets, highly intolerant copy number variants (CNVs) and ASD variants (Fig. 7d). Phenotypes related to DD, behavior and motor function showed unique enrichments by set, predominately in the iGLUTs, whereas all sets were enriched for seizure,



hypertonia and abnormal skeletal muscle morphology (Fig. 7e). Candidate drugs predicted to reverse convergent genes (that is, drugs with anti-correlating transcriptomic signatures) in iGLUTs and iGABAs were prioritized from the 520 Connectivity Map (CMap)<sup>45</sup> drugs with matched clinical and experimental zebrafish data. Top enriched drugs included antidepressants, antipsychotics and statins (Supplementary Data 2 and Supplementary Fig. 24a). Whereas some drugs were broadly predicted to reverse convergent signatures in three out of four NDD gene sets (for example, the antipsychotic perphenazine), others uniquely targeted specific sets (for example, naltrexone in set 2 iGLUTs, sirolimus in set 3 iGLUTs and valsartan in set 3 iGABAs). Sets 3 and 4 showed the greatest number of CMap enrichments. By considering existing pharmacological effects of the top drugs on zebrafish behavior<sup>46</sup>, some of the predicted drug reversers were shown to oppose effects on NDD-related phenotypes in zebrafish (Supplementary Fig. 24b). Yet, the direction of effect predicted based on transcriptomic convergence in human neurons did not always align with anti-correlating behavioral effects in zebrafish (for example, moxifloxacin, perphenazine).

The top negatively enriched drugs for iGLUT convergence from CMap and anti-correlating drugs predicted from a pharmaco-behavioral screen of 376 drugs in larval zebrafish<sup>46</sup> were empirically tested in representative mutants from sets 2–4, which showed the strongest CMap enrichments (Fig. 7f). We determined whether the phenotypic impact of mutant × drug combinations led to partial rescue, rescue, over-correction or exacerbation of the mutant phenotype across significant arousal and startle behavioral parameters (Fig. 7g). Ten out of 11 drugs rescued at least one dysregulated behavioral parameter (Fig. 7g and Supplementary Fig. 24c–e). Paclitaxel robustly rescued behavioral parameters in *kdm6bab* FO mutants and pravastatin partially and completely rescued select parameters in *chd2<sup>Δ7/Δ7</sup>* mutants (Fig. 7g(i)), including nighttime sleep bouts in *kdm6bab* FO mutants and responses to lights-ON stimuli in *chd2<sup>Δ7/Δ7</sup>* mutants (Supplementary Fig. 24f(i),(ii)). Interestingly, we also observed over-correction of the *phf21aab* FO mutant phenotype by fluvoxamine (Fig. 7g(ii)), such as increased sleep bouts that were significantly decreased following fluvoxamine treatment (Supplementary Fig. 24f(iii)). Taken together, in vivo behavioral profiling of NDD genes in zebrafish overlaps with in vitro-defined

convergent networks and identifies pharmacological suppressors of specific behavioral phenotypes.

## Discussion

To investigate common pathways affected by loss of NDD risk genes, we targeted 23 NDD genes with roles predominantly in gene regulation using a pooled CRISPR-KO strategy. Transcriptomic convergence across NDD risk genes varied between the cell types of the brain, resolving more convergent targets and stronger convergent networks in mature glutamatergic neurons, where they were enriched not just for pathways with well-established links to ASD etiology (for example, gene regulation, synaptic biology), but also mitochondrial function<sup>47</sup>. Machine learning tools extended the analyses in silico to all known NDD genes, recapitulating observed enrichments for regulatory NDD genes, yet predicting even greater convergence for synaptic NDD genes, and unexpectedly suggesting that predominantly ASD genes converge in excitatory neurons whereas predominantly DD genes converge in inhibitory neurons. Finally, drugs predicted to reverse convergent signatures suppressed behavioral phenotypes in NDD gene mutant zebrafish.

While downstream effects of epigenetic NDD genes unexpectedly targeted mitochondrial genes in neurons, in fact, 5% of NDD cases meet diagnostic criteria for classic mitochondrial disorders<sup>48</sup>. Mitochondrial DNA mutations<sup>49,50</sup>, haplotypes<sup>51</sup> and heteroplasmy<sup>49,52</sup> have all been associated with NDD. Not only do mitochondrial mutations cause synaptic and behavioral phenotypes<sup>53</sup>, but multiple lines of human and animal evidence link NDDs to mitochondrial deficits and oxidative stress<sup>54–59</sup>, with neuronal and/or behavioral phenotypes reversed by antioxidant treatment<sup>55,57–59</sup>.

Perturbations of the same NDD genes resulted in different convergent networks across cell types. For example, KOs of NDD genes in human NPCs<sup>18,34</sup>, cerebral organoids<sup>36–38</sup> and developing mouse<sup>39</sup>, tadpole<sup>40</sup> and zebrafish<sup>41</sup> brains reveal overlapping alterations in neurogenesis and developmental dynamics. Indeed, both regulatory and synaptic genes impact proliferation and patterning of progenitors (for example, *ARID1B*<sup>60</sup>, *CHD8*<sup>36</sup>, *NRXN1*<sup>61</sup>, *SYNGAP1*<sup>62</sup>), excitatory transmission by glutamatergic neurons (for example, *CHD8*<sup>63,64</sup>, *NRXN1*<sup>65</sup>, *SHANK3*<sup>66</sup>, *SYNGAP1*<sup>67</sup>) and inhibitory transmission by GABAergic

**Fig. 6 | NDD KOs converge on mitochondrial function.** **a**, Two-sided GSEA identified downstream pathways involved in neurogenesis, neurite outgrowth, synaptic biology and mitochondrial function as cell-type-specific targets of convergent genes across 15 NDD KOs (*ARID1B*, *ASHIL*, *ASXL3*, *BCL11A*, *KDMSB*, *CHD2*, *MBD5*, *MED13L*, *NRXN1*, *PHF12*, *PHF21A*, *SETD5*, *SIN3A*, *SKI*, *SMARRC2*) in iNPCs and mature iGLUTs. FDR-based multiple testing correction was performed and results were filtered for pathways with nominal  $P < 0.05$  for plotting. Normalized GSEA scores represent the direction of enrichment based on the meta-analyzed Z-score for each convergent gene. Cell type is represented by shape and the size of each point represents the  $-\log_{10}$ (FDR). **b**, Summary of network- and gene-level pathway enrichments (two-sided GSEA, FDR-corrected  $P$  values for enrichment tests are reported in Supplementary Data 1) (from Figs. 2 and 3) for shared effects of nine and 15 NDD KOs in iNPCs and mature iGLUTs. **c**, Proliferation assessment of NPCs using Ki-67 median fluorescence intensity (MFI) measured with flow cytometry, WT (purple, no iCas9 induction) versus KO (green, doxycycline to induce iCas9). Ki-67 MFI was compared between uninduced and induced conditions for each gene using two-sided unpaired  $t$ -tests with Welch's correction, followed by Benjamini–Hochberg FDR correction ( $Q = 5\%$ ). Bars show mean  $\pm$  s.e.m. with individual biological replicates overlaid (*NRXN1*:  $n = 4$  uninduced,  $n = 5$  induced; *ASHIL*:  $n = 5$  uninduced,  $n = 4$  induced; *ARID1B*:  $n = 4$  induced). LoF of target genes significantly decreased Ki-67 for *NRXN1* (mean difference  $-229.8$ ;  $P = 0.0004$ ;  $Q = 0.0008$ ) and increased Ki-67 for *ASHIL* (difference 887.0;  $P = 0.0319$ ;  $Q = 0.0335$ ) and *ARID1B* (difference 1,350;  $P = 0.0535$ ;  $Q = 0.0374$ ). **d**, Scatter plot of gRNA  $\log_2$ FC (high- (PE-high) and low- (FITC-high) mitochondrial inner membrane potential ( $\Delta\psi_m$ )-sensitive dye JC-1 membrane potential fractions) in NPCs ( $x$  axis) and mature iGLUT neurons ( $y$  axis), with points colored by enrichment category (shared NPC and iGLUT in red; distinct between NPC and iGLUT in blue). Right, bar chart of  $-\log_{10}$ (FDR)

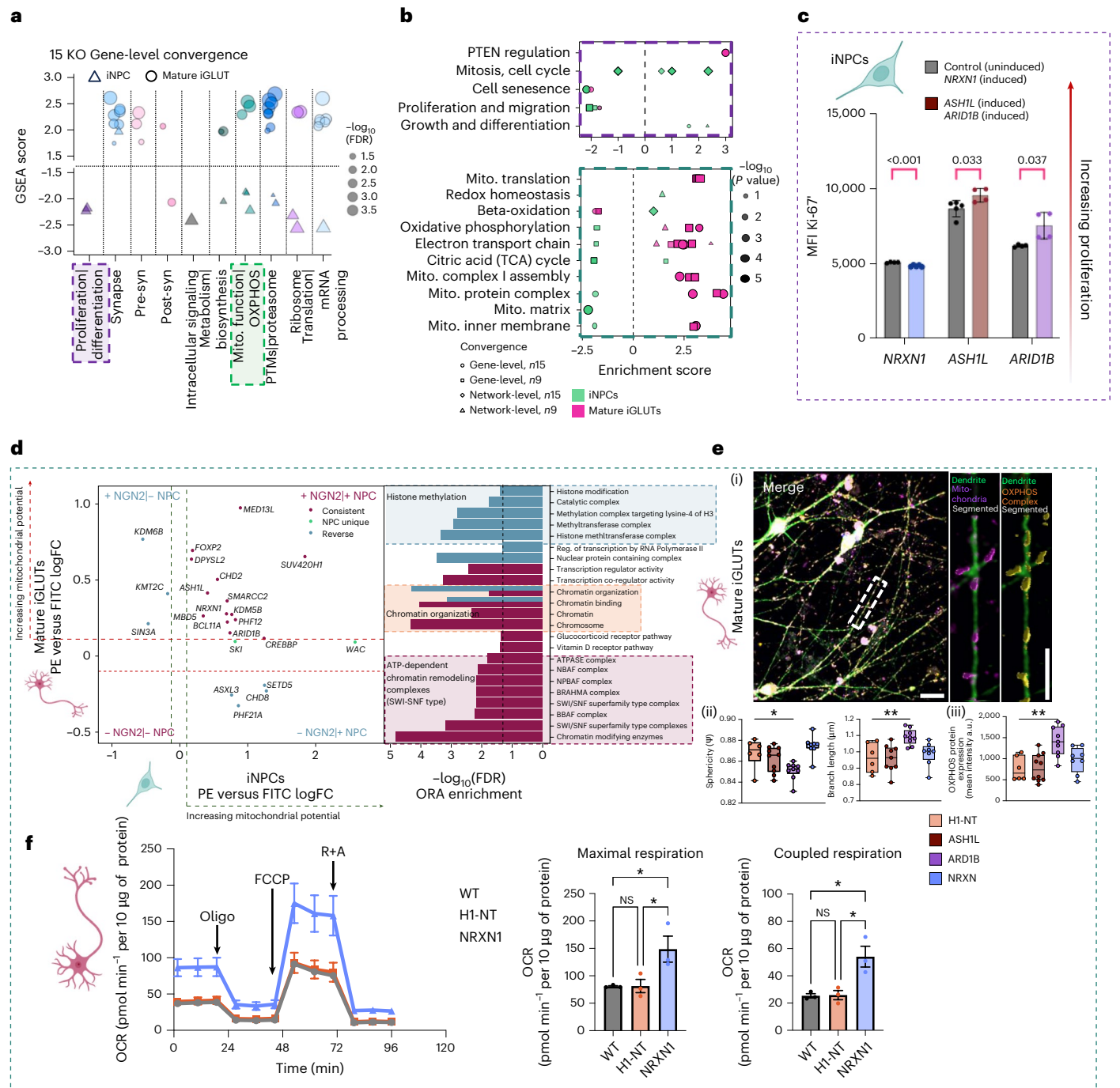
for over-represented gene sets in the gene KOs enriched in both lineages. **e**, (i) High resolution, high-throughput microscopy of mitochondrial morphology (scale bar: 10  $\mu\text{m}$ ): an isolated dendrite labeled with a dendritic marker (MAP2), mitochondrial marker (TOMM20) and marker of the OXPHOS complex (Total OXPHOS) (scale bar: 5  $\mu\text{m}$ ). (ii) Effect of *ASHIL*-KO ( $n = 10, 9$ ), *ARID1B*-KO ( $n = 9$ ) and *NRXN1*-KO ( $n = 8$ ) on mitochondrial sphericity (HI-NT versus *ARID1B*, adjusted  $P = 0.0213$ ) and branch length (HI-NT versus *ARID1B*, adjusted  $P = 0.0081$ ) independent of changes in mitochondrial volume and surface area (Supplementary Figs. 20 and 21) compared with HI-NT (nontargeting control,  $n = 6$ ). (iii) Effect of *ASHIL*-KO, *ARID1B*-KO and *NRXN1*-KO on average fluorescence intensity (HI-NT versus *ARID1B*, adjusted  $P = 0.0024$ ) of OXPHOS proteins compared with HI-NT. Each data point indicates one well of a 96-well plate, representing hundreds of  $\mu\text{m}^2$  of neuronal area and tens of thousands of individual mitochondria (statistical analysis comparing KO with control was performed using a one-way ANOVA comparing HI-NT versus other groups (*ASHIL*, *ARID1B*, *NRXN1*) with Sidak's multiple comparisons, \*adjusted  $P < 0.05$ , \*\*adjusted  $P < 0.01$ ). **f**, Effect of *NRXN1*-KO on maximal respiration (WT versus H-NT:  $P = 0.9995$ ; WT versus *NRXN1*:  $P = 0.0466$ ; H-NT versus *NRXN1*:  $P = 0.0483$ ) and coupled respiration (WT versus H-NT:  $P = 0.9973$ ; WT versus *NRXN1*:  $P = 0.0140$ ; H-NT versus *NRXN1*:  $P = 0.0152$ ) in iGLUTs. Data are presented as mean  $\pm$  s.e.m. Statistical analysis was performed using one-way ANOVA. Each data point in the temporal plot represents the mean and s.e.m. across a 24-well Seahorse assay plate. The experiment was independently replicated twice, with the bar graph showing three wells from one representative replicate. The center of the box plots represents the median, the bounds the 25th and 75th percentiles, and the whiskers the minimum and maximum values of that group. NS, not significant; OCR, oxygen consumption rate; oligo, oligomycin; R+A, rotenone and antimycin A.

neurons (for example, *ARID1B*<sup>68</sup>, *CHD8*<sup>63</sup>, *NRXN1*<sup>69</sup>, *SHANK3*<sup>70</sup>). Many NDD genes seem to have broad roles outside their annotated function; for example, some chromatin regulators (for example, *CHD8*, *CHD2* and *POGZ*) localize to microtubules in the centrosome<sup>17</sup>, mitotic spindle<sup>18</sup> and cilia<sup>19</sup>. This observation connects the pleiotropic nature of many NDD genes and pathophysiological evidence linking multiple cell types and distinct cellular functions to NDD.

What explains phenotypic convergence between NDD genes with distinct annotated functions? The strength of convergence was most highly correlated to common clinical associations, biological annotations and co-expression patterns in the post-mortem brain. Critically, these factors are inter-dependent. ASD and DD risk genes differentially map onto human developmental brain co-expression networks<sup>11</sup>, with those most strongly implicated in DD enriched for expression in progenitor cells and immature neurons, and those in

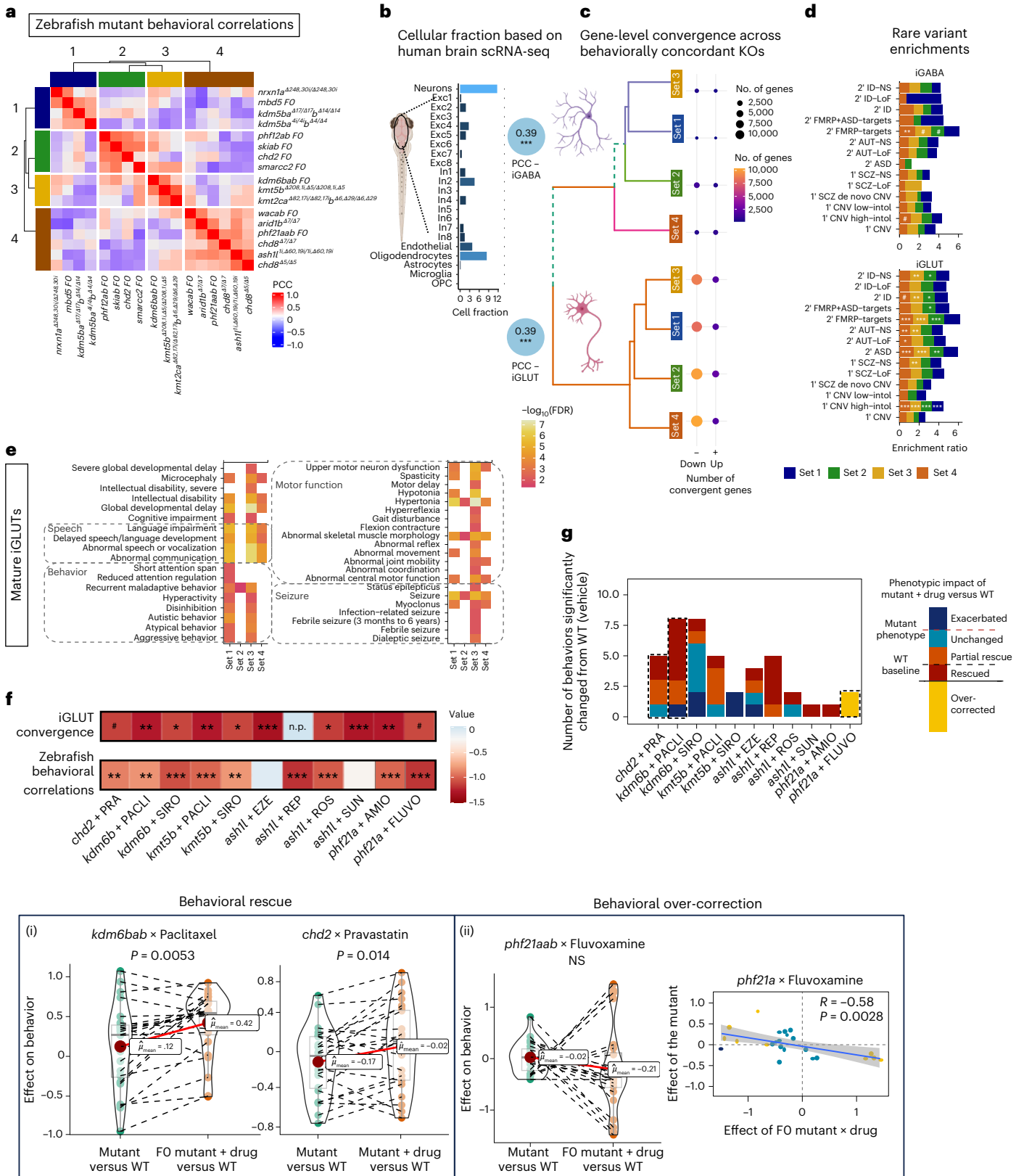
ASD in mature neurons<sup>5</sup>. Indeed, cellular identities and biological pathways are captured by patterns of gene co-expression<sup>71,72</sup>. Transcriptomic and epigenomic analyses of post-mortem brain from NDD cases likewise indicate convergent molecular signatures<sup>73</sup> and subtypes of NDD<sup>74</sup>. Thus, we posit that shared clinical and phenotypic effects of distinct NDD genes in fact reflect the patterns of co-expression in the developing brain.

Personalized medicine seeks to tailor treatments to individual patients<sup>75</sup>; for example, patients with cancer<sup>76</sup> and monogenic disease<sup>77</sup> with specific genetic mutations receive targeted treatments. Previous efforts to classify genes that predict NDD clinical features or treatment response applied gene ontology<sup>4,5</sup> or differential neurodevelopmental KO effects in vitro<sup>34</sup> or in vivo<sup>40</sup>. Here we proposed to stratify risk genes based on convergent molecular impacts in human neurons. Our overarching hypothesis, in doing so, was that by resolving shared



downstream gene targets between multiple NDD genes, we might inform a precision-medicine-based approach that did not necessarily need to target risk genes one-at-a-time. Although convergent networks did not predict behavioral stratification of zebrafish mutants, they did inform drug prediction, with 10 out of 11 drugs tested found

to ameliorate at least one mutant behavioral phenotype in vivo. This ability to reverse, rather than prevent, a behavioral phenotype indicates that targeting convergent networks in post-mitotic neurons may represent a clinically actionable neurodevelopmental window that persists through symptom onset. The extent to which convergent



**Fig. 7 | NDD gene mutants with shared behavioral phenotypes in zebrafish resolve unique and cell-type-specific gene-level convergent signatures and are rescued by predicted medications. a**, NDD risk genes uniquely cluster based on sleep–wake/visual-startle behavioral responses in zebrafish mutants. set 1: *nrxa1a*, *mbd5*, *kdm5bab*; set 2: *phf12ab*, *skiab*, *chd2*, *smarcc2*; set 3: *kdm6bab*, *kmt5b*, *kmt2cab*; set 4: *wacab*, *arid1b*, *phf21aab*, *chd8*, *ash1l*. **b**, Gene expression in human mature iGLUTs and iGABAs correlates with expression in the zebrafish brain. Cellular deconvolution of WT larval zebrafish brain expression based on adult human single-cell brain reference identifying neurons as the largest proportion of cells in the fish brain. Gene expression in WT zebrafish brain significantly positively correlates with gene expression of mature iGLUTs (two-tailed Pearson's correlation with Holm's multiple testing correction;  $R = 0.39$ ,  $P_{\text{Holm}} < 0.001$ ) and iGABAs ( $R = 0.39$ ,  $P_{\text{Holm}} < 0.001$ ). **c**, For each of the four behaviorally defined sets, gene-level convergence (DEGs with significant and shared direction of effect across all NDD genes within each of the four sets ( $P$ value-based DEG meta-analysis (METAL), FDR-adjusted  $P_{\text{meta}} < 0.05$ , Cochran's heterogeneity  $Q$ -test  $P_{\text{Het}} > 0.05$ )) is largely nonoverlapping between mature iGLUTs and iGABAs, with unique enrichments for common psychiatric risk gene targets. Numbers of convergent genes that are upregulated (+) or downregulated (–) for each NDD set are indicated. **d**, In both iGABAs and iGLUTs, all four behavioral sets were enriched for FMRP targets. Gene targets of neurodevelopmental rare variants were significantly enriched for convergent signatures only in mature iGLUTs; behavioral set 4 uniquely significantly enriched for secondary targets of ASD LoF variants and set 3 uniquely enriched for primary targets of SCZ nonsynonymous variants. **e**, In iGLUTs, NDD-related behaviors were enriched only in sets 1 and 3, with enrichments for language, speech and intellectual delays in sets 1, 3 and 4. All sets were enriched for seizure and hypertonia. **f**, Potential 'rescue' drugs for these four phenotypic groups were selected from enrichment scores using CMap and filtered for drugs included in a screen of 376 compounds for behavioral effects in zebrafish. Top candidates that were significantly negatively enriched for iGLUT convergence from CMap and negatively correlated with mutant behavioral features were tested in mutant lines representative of sets 2–4. n.p. indicates

that the drug repaglinide was not present in the CMap dataset. Mutant  $\times$  drug combinations were as follows: *chd2* <sup>$\Delta 7/\Delta 7$</sup>   $\times$  pravastatin; *kdm6bab* FO  $\times$  paclitaxel; *kdm6bab* FO  $\times$  sirolimus; *kmt5b* <sup>$\Delta 208, \text{II}, \Delta 5/\Delta 208, \text{II}, \Delta 5$</sup>   $\times$  paclitaxel; *kmt5b* <sup>$\Delta 208, \text{II}, \Delta 5/\Delta 208, \text{II}, \Delta 5$</sup>   $\times$  sirolimus; *ash1l* <sup>$\text{II}, \Delta 60, 19/\text{II}, \Delta 60, 19$</sup>   $\times$  ezetimibe; *ash1l* <sup>$\text{II}, \Delta 60, 19/\text{II}, \Delta 60, 19$</sup>   $\times$  repaglinide; *ash1l* <sup>$\text{II}, \Delta 60, 19/\text{II}, \Delta 60, 19$</sup>   $\times$  rosuvastatin; *ash1l* <sup>$\text{II}, \Delta 60, 19/\text{II}, \Delta 60, 19$</sup>   $\times$  sunitinib; *phf21aab* FO  $\times$  amiodarone; *phf21aab* FO  $\times$  fluvoxamine. **g**, For behaviors that were significantly different between mutant + DMSO and WT + DMSO ( $P < 0.06$ ), we characterized the effect of the mutant  $\times$  drug on behavior as exacerbated (**a**) (significant effect mutant + drug-versus-WT > significant effect mutant-versus-WT), unchanged (**b**) (significant effect mutant + drug-versus-WT = significant effect mutant-versus-WT), partial rescue (**c**) (significant effect mutant + drug-versus-WT < effect mutant-versus-WT), rescued (**d**) (significant effect mutant-versus-WT, no significant effect mutant + drug-versus-WT) or over-corrected (**e**) (mutant + drug-versus-WT opposite direction of significant effect mutant-versus-WT). All drugs reversed at least one dysregulated behavior except for sirolimus in *kmt5b*. (i) Comparison of the magnitude of effect (beta,  $n = 24$  parameters) on behavior between the mutant + DMSO compared with mutant + drug groups shows rescue of select behavioral features in *kdm6b* and *chd2* mutants by paclitaxel (Shapiro–Wilk's normality:  $W = 0.94301$ ,  $P = 0.02121$ ; two-sided Wilcoxon signed rank test statistic = 52,  $P = 0.0053$ ,  $n = 24$ ) and pravastatin (Shapiro–Wilk's normality:  $W = 0.97587$ ,  $P = 0.4203$ , two-sided paired Welch's  $t$ -statistic =  $-3.533$ ,  $P = 0.01394$ ,  $n = 24$ ), respectively. (ii) The *phf21a* mutant phenotype was strongly opposed by fluvoxamine (left: normality statistic = 0.93744,  $P = 0.01295$ ; two-sided Wilcoxon signed rank test statistic = 196,  $P = 0.19$ ; right: two-tailed Pearson's correlation =  $-0.58$ ,  $P = 0.0028$ ,  $n = 24$ ). In the box plots, the median (black line) and mean (red point) are shown. The lower and upper hinges correspond to 1st and 3rd quartiles. Upper and lower whiskers extend from the hinge to the largest or smallest value up to  $1.5 \times$  IQR. All data points are plotted individually. Illustrations in **b** created in BioRender; Townsley, K. <https://biorender.com/rvk1zn2> (2026). AMIO, amiodarone; EZE, ezetimibe; FLUVO, fluvoxamine; PACLI, paclitaxel; PRA, pravastatin; REP, repaglinide; ROS, rosuvastatin; SIRO, sirolimus; SUN, sunitinib.

downstream targets, whether associated with risk or resilience, can be manipulated to prevent or ameliorate NDD signatures and phenotypes warrants future investigation.

Although rare LoF NDD gene mutations tend to confer large effects in the individuals who carry them, the small effects of common variants account for much of the genetic risk for NDD at the population level<sup>2</sup>. The differences in expressivity and incomplete penetrance of high-effect-size rare variants are frequently attributed to diversity across polygenic backgrounds<sup>78</sup>; in vitro, NDD gene effects are indeed influenced by the individual genomic context<sup>36</sup>. In psychiatry, common genetic variants are more associated with cross-disorder behavioral dimensions<sup>79</sup> and rare variants with co-occurring intellectual disability<sup>80</sup>. Common risk variants interact with rare mutations to determine individual-level liability in ASD<sup>81</sup>, DD<sup>82</sup>, SCZ<sup>83,84</sup> and epilepsy<sup>85</sup>. Our results, highlighting that convergence downstream of NDD gene effects is enriched for cross-disorder GWAS variants and rare LoF genes, inform pleiotropy of genetic risk for psychiatric disorders. Moving forward, we argue that it is critical that empirical functional genomic studies systematically consider the impact of common and rare variants together, including screening the impact of LoF genes in hiPSC lines derived from donors with high and low polygenic risk scores<sup>86</sup>. Intriguingly, even susceptibility to environmental risk factors for NDD (for example, valproic acid) seems to be mediated by genetic background<sup>87</sup>. Deeper phenotypic characterization of NDD effects across donors will be critical in determining how complex genetic (or environmental) interactions shape cellular phenotypes, circuit function and human behavior in the clinic.

In the post-mortem brain, NDD gene signatures are not just associated with downregulation of co-expression modules involving synaptic signaling<sup>88</sup>, but also with upregulation of microglial and astrocyte gene modules<sup>89,90</sup>. The extent to which increased neuroimmune activity in NDD is a response to cellular or environmental sources of inflammation,

or indicative of a role for glia cells in risk, is unclear; evidence supports both possibilities. Consistent with a model of maternal immune activation during neurodevelopment<sup>91</sup>, glucocorticoids and inflammatory cytokines perturb the expression of psychiatric risk genes<sup>92,93</sup>, altering the regulatory activity of psychiatric risk loci<sup>94</sup> and interfering with neuronal maturation in brain organoids<sup>95</sup>. Yet, in vivo analysis of NDD genes in zebrafish revealed global increases in microglia<sup>41</sup> and in vitro screening in human microglia uncovered roles in endocytosis and uptake of synaptic material<sup>96</sup>. Indeed, given the reciprocal relationships between neuronal activity and glial function, epigenetic state and gene expression<sup>97–99</sup>, it seems probable that both cell-autonomous and non-cell-autonomous effects underlie and/or exacerbate NDD gene effects.

In summary, we demonstrate that convergent effects of NDD risk genes vary between cell types. Our analyses suggest that clinical convergence between regulatory and synaptic genes in the etiology of NDD is driven more so by co-expression patterns of risk genes than by direct regulation of epigenetic genes on synaptic targets. If the convergence of multifold risk genes on a smaller number of shared molecular pathways indeed explains how genetically heterogeneous mutations result in similar clinical features, then genetic stratification of cases will inform novel therapeutic targets. We predict that such individualized points of therapeutic intervention may be most effective when targeting mature glutamatergic neurons, which not only harbor the strongest convergent effects but also represent a therapeutic window that is actionable after diagnosis.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41593-026-02247-7>.

## References

1. Sandin, S. et al. The heritability of autism spectrum disorder. *JAMA* **318**, 1182–1184 (2017).
2. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
3. Mahjani, B. et al. Prevalence and phenotypic impact of rare potentially damaging variants in autism spectrum disorder. *Mol. Autism* **12**, 65 (2021).
4. 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 (2020).
5. Fu, J. M. et al. Rare coding variation provides insight into the genetic architecture and phenotypic context of autism. *Nat. Genet.* **54**, 1320–1331 (2022).
6. Singh, T. et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).
7. Palmer, D. S. et al. Exome sequencing in bipolar disorder identifies AKAP11 as a risk gene shared with schizophrenia. *Nat. Genet.* **54**, 541–547 (2022).
8. Willsey, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
9. Fazel Darbandi, S. et al. Five autism-associated transcriptional regulators target shared loci proximal to brain-expressed genes. *Cell Rep.* **43**, 114329 (2024).
10. Markenscoff-Papadimitriou, E. et al. Autism risk gene POGZ promotes chromatin accessibility and expression of clustered synaptic genes. *Cell Rep.* **37**, 110089 (2021).
11. Parikshak, N. N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
12. Voineagu, I. et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
13. Liao, C. et al. Convergent coexpression of autism-associated genes suggests some novel risk genes may not be detectable in large-scale genetic studies. *Cell Genom.* **3**, 100277 (2023).
14. Pintacuda, G. et al. Protein interaction studies in human induced neurons indicate convergent biology underlying autism spectrum disorders. *Cell Genom.* **3**, 100250 (2023).
15. Gao, Y. et al. Proximity analysis of native proteomes reveals phenotypic modifiers in a mouse model of autism and related neurodevelopmental conditions. *Nat. Commun.* **15**, 6801 (2024).
16. Willsey, H. R., Willsey, A. J., Wang, B. & State, M. W. Genomics, convergent neuroscience and progress in understanding autism spectrum disorder. *Nat. Rev. Neurosci.* **23**, 323–341 (2022).
17. O’Neill, A. C. et al. Spatial centrosome proteome of human neural cells uncovers disease-relevant heterogeneity. *Science* **376**, eabf9088 (2022).
18. Sun, N. et al. Autism genes converge on microtubule biology and RNA-binding proteins during excitatory neurogenesis. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.12.22.573108> (2024).
19. Kostyanovskaya, E. et al. Convergence of autism proteins at the cilium. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.12.05.626924> (2025).
20. 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 (2017).
21. Johannesen, K. M. et al. Defining the phenotypic spectrum of SLC6A1 mutations. *Epilepsia* **59**, 389–402 (2018).
22. Heyne, H. O. et al. De novo variants in neurodevelopmental disorders with epilepsy. *Nat. Genet.* **50**, 1048–1053 (2018).
23. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
24. Sullivan, P. F. & Geschwind, D. H. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell* **177**, 162–183 (2019).
25. Schrode, N. et al. Synergistic effects of common schizophrenia risk variants. *Nat. Genet.* **51**, 1475–1485 (2019).
26. Wells, M. F. et al. Natural variation in gene expression and viral susceptibility revealed by neural progenitor cell villages. *Cell Stem Cell* **30**, 312–332 (2023).
27. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
28. Yang, N. et al. Generation of pure GABAergic neurons by transcription factor programming. *Nat. Methods* <https://doi.org/10.1038/nmeth.4291> (2017).
29. Mimitou, E. P. et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).
30. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
31. Deans, P. J. M. et al. Functional implications of polygenic risk for schizophrenia in human neurons. *Nat. Commun.* **17**, 1355 (2026).
32. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
33. Emani, P. S. et al. Single-cell genomics and regulatory networks for 388 human brains. *Science* **384**, eadi5199 (2024).
34. 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 (2020).
35. 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 (2020).
36. Paulsen, B. et al. Autism genes converge on asynchronous development of shared neuron classes. *Nature* **602**, 268–273 (2022).
37. Meng, X. et al. Assembloid CRISPR screens reveal impact of disease genes in human neurodevelopment. *Nature* **622**, 359–366 (2023).
38. Li, C. et al. Single-cell brain organoid screening identifies developmental defects in autism. *Nature* **621**, 373–380 (2023).
39. Jin, X. et al. In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* <https://doi.org/10.1126/science.aaz6063> (2020).
40. 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 (2021).
41. Weinschutz Mendes, H. et al. High-throughput functional analysis of autism genes in zebrafish identifies convergence in dopaminergic and neuroimmune pathways. *Cell Rep.* **42**, 112243 (2023).
42. Wang, B. et al. A foundational atlas of autism protein interactions reveals molecular convergence. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.12.03.569805> (2023).
43. Murtaza, N. et al. Neuron-specific protein network mapping of autism risk genes identifies shared biological mechanisms and disease-relevant pathologies. *Cell Rep.* **41**, 111678 (2022).
44. Gonzalez, F. et al. An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).
45. Subramanian, A. et al. A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell* **171**, 1437–1452 (2017).
46. Jamadagni, P. et al. Pharmacological profiling identifies suppressors of autism gene-associated phenotypes in zebrafish. *Proc. Natl Acad. Sci. USA* **123**, e2518846123 (2026).
47. Manji, H. et al. Impaired mitochondrial function in psychiatric disorders. *Nat. Rev. Neurosci.* **13**, 293–307 (2012).

48. Rossignol, D. A. & Frye, R. E. Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis. *Mol. Psychiatry* **17**, 290–314 (2012).
49. Wang, Y., Picard, M. & Gu, Z. Genetic evidence for elevated pathogenicity of mitochondrial DNA heteroplasmy in autism spectrum disorder. *PLoS Genet.* **12**, e1006391 (2016).
50. Varga, N. A. et al. Mitochondrial dysfunction and autism: comprehensive genetic analyses of children with autism and mtDNA deletion. *Behav. Brain Funct.* **14**, 4 (2018).
51. Chalkia, D. et al. Association between mitochondrial DNA haplogroup variation and autism spectrum disorders. *JAMA Psychiatry* **74**, 1161–1168 (2017).
52. Wang, Y. et al. Association of mitochondrial DNA content, heteroplasmies and inter-generational transmission with autism. *Nat. Commun.* **13**, 3790 (2022).
53. Yardeni, T. et al. An mtDNA mutant mouse demonstrates that mitochondrial deficiency can result in autism endophenotypes. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.2021429118> (2021).
54. Shen, M. et al. Reduced mitochondrial fusion and Huntingtin levels contribute to impaired dendritic maturation and behavioral deficits in Fmr1-mutant mice. *Nat. Neurosci.* **22**, 386–400 (2019).
55. Fernandez, A. et al. Mitochondrial dysfunction leads to cortical under-connectivity and cognitive impairment. *Neuron* **102**, 1127–1142 (2019).
56. Gandal, M. J. et al. Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018).
57. Kanellopoulos, A. K. et al. Aralar sequesters GABA into hyperactive mitochondria, causing social behavior deficits. *Cell* **180**, 1178–1197 (2020).
58. Li, J. et al. Mitochondrial deficits in human iPSC-derived neurons from patients with 22q11.2 deletion syndrome and schizophrenia. *Transl. Psychiatry* **9**, 302 (2019).
59. Li, J. et al. Association of mitochondrial biogenesis with variable penetrance of schizophrenia. *JAMA Psychiatry* **78**, 911–921 (2021).
60. Martins-Costa, C. et al. ARID1B controls transcriptional programs of axon projection in an organoid model of the human corpus callosum. *Cell Stem Cell* **31**, 866–885 (2024).
61. Flaherty, E. et al. Neuronal impact of patient-specific aberrant NRXN1 $\alpha$  splicing. *Nat. Genet.* **51**, 1679–1690 (2019).
62. Birtele, M. et al. Non-synaptic function of the autism spectrum disorder-associated gene SYNGAP1 in cortical neurogenesis. *Nat. Neurosci.* **26**, 2090–2103 (2023).
63. Ellingford, R. A. et al. Cell-type-specific synaptic imbalance and disrupted homeostatic plasticity in cortical circuits of ASD-associated Chd8 haploinsufficient mice. *Mol. Psychiatry* **26**, 3614–3624 (2021).
64. Shi, X. et al. Heterozygous deletion of the autism-associated gene CHD8 impairs synaptic function through widespread changes in gene expression and chromatin compaction. *Am. J. Hum. Genet.* **110**, 1750–1768 (2023).
65. Pak, C. et al. Cross-platform validation of neurotransmitter release impairments in schizophrenia patient-derived NRXN1-mutant neurons. *Proc. Natl Acad. Sci. USA* <https://doi.org/10.1073/pnas.2025598118> (2021).
66. Yi, F. et al. Autism-associated SHANK3 haploinsufficiency causes I<sub>h</sub> channelopathy in human neurons. *Science* **352**, aaf2669 (2016).
67. Vermaercke, B. et al. SYNGAP1 deficiency disrupts synaptic neoteny in xenotransplanted human cortical neurons in vivo. *Neuron* <https://doi.org/10.1016/j.neuron.2024.07.007> (2024).
68. Jung, E. M. et al. Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. *Nat. Neurosci.* **20**, 1694–1707 (2017).
69. Fernando, M. B. et al. Phenotypic complexities of rare heterozygous neurexin-1 deletions. *Nature* **642**, 710–720 (2025).
70. Chen, Q. et al. Dysfunction of cortical GABAergic neurons leads to sensory hyper-reactivity in a Shank3 mouse model of ASD. *Nat. Neurosci.* **23**, 520–532 (2020).
71. Rajarajan, P. et al. Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science* <https://doi.org/10.1126/science.aat4311> (2018).
72. Wen, C. et al. Cross-ancestry atlas of gene, isoform, and splicing regulation in the developing human brain. *Science* **384**, eadh0829 (2024).
73. Wong, C. C. Y. et al. Genome-wide DNA methylation profiling identifies convergent molecular signatures associated with idiopathic and syndromic autism in post-mortem human brain tissue. *Hum. Mol. Genet.* **28**, 2201–2211 (2019).
74. Ramaswami, G. et al. Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism. *Nat. Commun.* **11**, 4873 (2020).
75. Bell, J. Stratified medicines: towards better treatment for disease. *Lancet* [https://doi.org/10.1016/S0140-6736\(14\)60115-X](https://doi.org/10.1016/S0140-6736(14)60115-X) (2014).
76. Tsimberidou, A. M. et al. Molecular tumour boards – current and future considerations for precision oncology. *Nat. Rev. Clin. Oncol.* **20**, 843–863 (2023).
77. Zhang, H., Colclough, K., Gloyn, A. L. & Pollin, T. I. Monogenic diabetes: a gateway to precision medicine in diabetes. *J. Clin. Invest.* <https://doi.org/10.1172/JCI142244> (2021).
78. Schaaf, C. P. et al. A framework for an evidence-based gene list relevant to autism spectrum disorder. *Nat. Rev. Genet.* **21**, 367–376 (2020).
79. Shi, Y. et al. Multi-polygenic scores in psychiatry: from disorder specific to transdiagnostic perspectives. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **195**, e32951 (2024).
80. Andreassen, O. A., Hindley, G. F. L., Frei, O. & Smeland, O. B. New insights from the last decade of research in psychiatric genetics: discoveries, challenges and clinical implications. *World Psychiatry* **22**, 4–24 (2023).
81. 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 (2022).
82. Kingdom, R., Beaumont, R. N., Wood, A. R., Weedon, M. N. & Wright, C. F. Genetic modifiers of rare variants in monogenic developmental disorder loci. *Nat. Genet.* **56**, 861–868 (2024).
83. Bergen, S. E. et al. Joint contributions of rare copy number variants and common SNPs to risk for schizophrenia. *Am. J. Psychiatry* **176**, 29–35 (2019).
84. Akingbuwa, W. A., Hammerschlag, A. R., Bartels, M., Nivard, M. G. & Middeldorp, C. M. Ultra-rare and common genetic variant analysis converge to implicate negative selection and neuronal processes in the aetiology of schizophrenia. *Mol. Psychiatry* **27**, 3699–3707 (2022).
85. Oliver, K. L. et al. Common risk variants for epilepsy are enriched in families previously targeted for rare monogenic variant discovery. *EBioMedicine* **81**, 104079 (2022).
86. Dobrindt, K. et al. Publicly available hiPSC lines with extreme polygenic risk scores for modeling schizophrenia. *Complex Psychiatry* **6**, 68–82 (2021).
87. Anton-Bolanos, N. et al. Brain Chimeroids reveal individual susceptibility to neurotoxic triggers. *Nature* **631**, 142–149 (2024).
88. Gandal, M. J. et al. Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
89. Wamsley, B. et al. Molecular cascades and cell type-specific signatures in ASD revealed by single-cell genomics. *Science* **384**, eadh2602 (2024).

90. Yap, C. X. et al. Brain cell-type shifts in Alzheimer's disease, autism, and schizophrenia interrogated using methylomics and genetics. *Sci. Adv.* **10**, eadn7655 (2024).
91. Han, V. X., Patel, S., Jones, H. F. & Dale, R. C. Maternal immune activation and neuroinflammation in human neurodevelopmental disorders. *Nat. Rev. Neurol.* **17**, 564–579 (2021).
92. 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 (2022).
93. Seah, C. et al. Common genetic variation impacts stress response in the brain. Preprint at *bioRxiv* <https://doi.org/10.1101/2023.12.27.573459> (2023).
94. Retallick-Townsley, K. G. et al. Dynamic stress- and inflammatory-based regulation of psychiatric risk loci in human neurons. Preprint at *bioRxiv* <https://doi.org/10.1101/2024.07.09.602755> (2024).
95. Cruceanu, C. et al. Cell-type-specific impact of glucocorticoid receptor activation on the developing brain: a cerebral organoid study. *Am. J. Psychiatry* <https://doi.org/10.1176/appi.ajp.2021.21010095> (2021).
96. Teter, O. M. et al. CRISPRi-based screen of autism spectrum disorder risk genes in microglia uncovers roles of *ADNP* in microglia endocytosis and uptake of synaptic material. *Mol. Psych.* **30**, 4176–4193 (2025).
97. Ma, Y. et al. Activity-dependent transcriptional program in *NGN2*<sup>+</sup> neurons enriched for genetic risk for brain-related disorders. *Biol. Psychiatry* **95**, 187–198 (2024).
98. Sanchez-Priego, C. et al. Mapping cis-regulatory elements in human neurons links psychiatric disease heritability and activity-regulated transcriptional programs. *Cell Rep.* **39**, 110877 (2022).
99. Boulting, G. L. et al. Activity-dependent regulome of human GABAergic neurons reveals new patterns of gene regulation and neurological disease heritability. *Nat. Neurosci.* **24**, 437–448 (2021).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2026

<sup>1</sup>Division of Molecular Psychiatry, Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. <sup>2</sup>Pamela Sklar Division of Psychiatric Genomics, Department of Genetics and Genomics, Icahn Institute of Genomics and Multiscale Biology, Nash Family Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>3</sup>Interdepartmental Neuroscience Program, Yale School of Medicine, New Haven, CT, USA. <sup>4</sup>Child Study Center, Yale University School of Medicine, New Haven, CT, USA. <sup>5</sup>Mohn Research Centre for Regenerative Medicine, Haukeland University Hospital, Bergen, Norway. <sup>6</sup>Department of Molecular Biophysics & Biochemistry and Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. <sup>7</sup>BD2: Breakthrough Discoveries for Thriving with Bipolar Disorder, Easton, MD, USA. <sup>8</sup>Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA. <sup>9</sup>Medical Scientist Training Program, Yale School of Medicine, New Haven, CT, USA. <sup>10</sup>Hospital Israelita Albert Einstein, São Paulo, Brazil. <sup>11</sup>Department of Cellular & Molecular Physiology, Yale School of Medicine, New Haven, CT, USA. <sup>12</sup>Systems Biology Institute, Yale West Campus, West Haven, CT, USA. <sup>13</sup>Department of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA. <sup>14</sup>Department of Neuroscience, Yale University School of Medicine, New Haven, CT, USA. <sup>15</sup>Wu Tsai Institute, Yale University School of Medicine, New Haven, CT, USA. <sup>16</sup>Department of Genetics, Yale University School of Medicine, New Haven, CT, USA. <sup>17</sup>These authors contributed equally: Meilin Fernandez Garcia, Kayla Retallick-Townsley, April Pruitt, Elizabeth A. Davidson, Novin Balafkan. ✉e-mail: [ellen.hoffman@yale.edu](mailto:ellen.hoffman@yale.edu); [kristen.brennand@yale.edu](mailto:kristen.brennand@yale.edu)

## Methods

### 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 subject research. Post-mortem brain data are similarly lacking identifiable information and are not considered human subject research.

All procedures involving zebrafish were conducted in accordance with Institutional Animal Care and Use Committee (IACUC; protocol no. 2024-20054) regulatory standards at Yale University.

### Generation of neural cells

Informed consent was obtained at the National Institute of Mental Health (NIMH), under the review of their Internal Review Board. hiPSC work was reviewed by the Internal Review Board of the Icahn School of Medicine at Mount Sinai as well as by the Embryonic Stem Cell Research Oversight Committee at the Icahn School of Medicine at Mount Sinai and Yale University. Fibroblasts were genotyped by IlluminaOmni 2.5 bead chip genotyping<sup>100,101</sup>, PsychChip<sup>102</sup> and exome sequencing<sup>102</sup>; hiPSCs<sup>102</sup> were validated by G-banded karyotyping (Wicell Cytogenetics) and genome stability monitored by Infinium Global Screening Array v3.0 (Illumina); single nucleotide polymorphism (SNP) genotype was inferred from all RNA sequencing (RNA-seq) data using the Sequenom SURESelect Clinical Research Exome and Sure Select V5 SNP lists to confirm that neuron identity matched donor. All hiPSCs are available at the Rutgers University Cell and DNA Repository (study 160) NIMH Repository and Genomic Resource is in the process of transferring operations from Rutgers University to Coriell Institute for Medical Research. For urgent requests, please contact NIMH.genomics.resources@mail.nih.gov.

Control hiPSCs were cultured in StemFlex media (Gibco, cat. no. A3349401) supplemented with Antibiotic-Antimycotic (Gibco, cat. no. 15240062) on Geltrex-coated plates (Gibco, cat. no. A1413302). Cells were passaged at 80–90% confluence with 5 mM EDTA (Life Technologies, cat. no. 15575-020) for 3 min at room temperature. EDTA was aspirated and cells dissociated in fresh StemFlex media. Media was replaced every 48–72 h for 4–7 d until the next passage.

Transient transcription factor overexpression from stable clonal hiPSCs was used to induce control hiPSCs to iNPCs (here SNaPs)<sup>26</sup>, iGLUTs<sup>27</sup> and iGABAs<sup>28</sup>. iNPCs are rapidly generated by 48-h induction with *NGN2*<sup>26,103</sup>. iGLUTs are induced via transient overexpression of *NGN2*, and are >95% glutamatergic neurons, robustly express excitatory genes and show spontaneous excitatory synaptic activity by 3–4 weeks in vitro<sup>25,27,61,65,66,104–110</sup>. iGABA neurons are induced via transient overexpression of *ASCL1* and *DLX2*, and are >95% GABAergic neurons, robustly express inhibitory genes and show spontaneous inhibitory synaptic activity by 5–6 weeks<sup>28,69,106,111,112</sup>. iNPCs, iGLUTs and iGABAs express most NDD genes, including all genes prioritized herein<sup>25</sup>.

We transduced hiPSCs from two control donors (553-3, karyotypic XY; 3182-3, karyotypic XX) with lentiviral *pUBIQ-rtTA* (Addgene, cat. no. 20342) and *tetO-NGN2-eGFP-NeoR* (Addgene, cat. no. 99378) for iNPCs and iGLUTs, or *pUBIQ-rtTA* (Addgene, cat. no. 20342), *tetO-ASCL1-PuroR* (Addgene, cat. no. 97329) and *tetO-DLX2-HygroR* (Addgene, cat. no. 97330) for iGABAs. Following transduction by spinfection at 1,000g for 1 h at 37 °C, hiPSCs were subjected to 48-h antibiotic selection (1 mg ml<sup>-1</sup> neomycin G418 (Thermo, cat. no. 10131027), 0.5 µg ml<sup>-1</sup> puromycin (Thermo, cat. no. A1113803) and/or 250 µg ml<sup>-1</sup> hygromycin (Thermo, cat. no. 10687010), and then clonalized by expansion from single colonies. Ultimately, clonal and inducible iNPC/iGLUT 3182-3-clone5 (XX) and iGABA 553-3-clone34 (XY) hiPSCs were validated for lentiviral genome integration by PCR, doxycycline-induced transcription factor expression by quantitative PCR, and robust and consistent neuronal induction confirmed by RNA-seq and immunocytochemistry for relevant cell-type markers. Analyses throughout reflect data from

iGLUT 3182-3-clone5 (iNPC, day 7 iGLUT and day 21 iGLUT) and iGABA 553-3-clone34 (day 36 iGABA).

**iNPCs.** At 0 days in vitro (DIV0), 3182-3-clone5 hiPSCs were dissociated and plated at  $1.5 \times 10^6$  cells per well onto Geltrex-coated six-well plates (1:250 dilution coating) in SNaP Induction Media (DIV0): DMEM/F12 with Glutamax (ThermoFisher, cat. no. 11320082), Glucose (0.3% v/v), N2 Supplement (1:100, ThermoFisher, cat. no. 17502048), doxycycline (2 µg ml<sup>-1</sup>; Sigma-Aldrich, cat. no. D9891), LDN-193189 (200 nM; Stemgent, cat. no. 04-0074), SB431542 (10 µM; Tocris, cat. no. 1614) and XAV939 (2 µM; Stemgent, cat. no. 04-00046) supplemented with 25 ng ml<sup>-1</sup> Chroma I ROCK2 Inhibitor. After 24 h, DIV2, cells were fed with Selection Media: DMEM/F12 with Glutamax, Glucose (0.3% v/v), N2 Supplement (1:100), doxycycline (2 µg ml<sup>-1</sup>), Geneticin (0.5 mg ml<sup>-1</sup>; ThermoFisher, cat. no. 10131035), LDN-193189 (100 nM), SB431542 (5 µM) and XAV939 (1 µM). At 48 h post induction (DIV2), SNaPs were dissociated with Accutase for 10 min at 37 °C, quenched in DMEM, pelleted at 800g for 5 min and replated at  $1.5 \times 10^6$  cells per well onto Geltrex-coated six-well plates in SNaP Selection Media supplemented with Geneticin (0.5 mg ml<sup>-1</sup>). After 16–18 h (DIV3), medium was switched to SNaP maintenance Medium: DMEM/F12 with Glutamax, Penn/Strep (1:100), MEM-NEAA (1:100; Life Technologies, cat. no. 10370088), B27 minus Vitamin A (1:50; Life Technologies, cat. no. 12587010), N2 Supplement (1:100; Life Technologies, cat. no. 17502048), recombinant human EGF (10 ng ml<sup>-1</sup>; R&D Systems, cat. no. 236-EG-200), recombinant human basic FGF (10 ng ml<sup>-1</sup>; Life Technologies, cat. no. 13256029), Geneticin (0.5 mg ml<sup>-1</sup>) and Chroman I (25 ng ml<sup>-1</sup>). Cells were fed every 48 h with SNaP maintenance medium lacking Chroman I and Geneticin. Cells were dissociated and seeded weekly at a density of  $1.25\text{--}1.5 \times 10^6$  cells per well onto Geltrex-coated six-well plates until NPC morphology was observed and persistent. Cells were expanded and cryofrozen.

**DIV7 iGLUTs.** The 3182-3-clone5 iNPCs were thawed and seeded at  $1 \times 10^6$  cells per well onto Geltrex-coated 12-well plates. *NGN2* expression was induced with doxycycline (2 µg ml<sup>-1</sup>) for 24 h (DIV0) with antibiotic selection for 48 h (DIV1-3) in SNaP maintenance medium. At DIV4, SNaPs were dissociated with Accutase, switched into Neuronal Medium (Brainphys (Stemcell, cat. no. 05790), Glutamax (1:100), Sodium Pyruvate (1 mM), Anti-Anti (1:100), N2 (1:100), B27 without vitamin A (1:50), BDNF (20 ng ml<sup>-1</sup>; R&D, cat. no. 248-BD-025), GDNF (20 ng ml<sup>-1</sup>; R&D, cat. no. 212), dibutyryl cAMP (500 µg ml<sup>-1</sup>; Sigma, cat. no. D0627), L-ascorbic acid (200 µM; Sigma, cat. no. A4403), Natural Mouse Laminin (1.2 µg ml<sup>-1</sup>; ThermoFisher, cat. no. 23017015)) and seeded in Geltrex-coated (1:120 dilution coating) 12-well plates. Medium was changed every 24 h until DIV7 collection.

**D21 iGLUTs.** hiPSCs were collected in Accutase (Innovative Cell Technologies, cat. no. AT-104) for 5 min at 37 °C, dissociated into a single-cell suspension, quenched in DMEM, pelleted via centrifugation for 5 min at 1,000g, resuspended in StemFlex containing 25 ng ml<sup>-1</sup> Chroma I ROCK2 Inhibitor and 2.0 µg ml<sup>-1</sup> doxycycline (DIV0), seeded  $1 \times 10^6$  cells per well onto Geltrex-coated six-well plates (1:250 dilution coating) and incubated overnight at 37 °C. The next day, DIV1, hiPSCs were subjected to 48-h antibiotic selection by medium replacement with Induction Media: DMEM/F12 (ThermoFisher, cat. no. 10565018), Glutamax (1:100; ThermoFisher, cat. no. 10565018), N2 (1:100; ThermoFisher, cat. no. 17502048), B27 without vitamin A (1:50; ThermoFisher, cat. no. 12587010), Antibiotic-Antimycotic (1:100) with 1.0 µg ml<sup>-1</sup> doxycycline and 0.5 mg ml<sup>-1</sup> Geneticin. At DIV3, cells were treated with 4.0 µM cytosine-β-D-arabinofuranoside hydrochloride (Ara-C) and 1.0 µg ml<sup>-1</sup> doxycycline to arrest proliferation and eliminate non-neuronal cells in the culture. At DIV4, immature neurons were dissociated with Accutase and 5 units per ml of DNase I at 37 °C for 7–10 min, quenched in DMEM, centrifuged for 5 min at 400g, resuspended in 25 ng ml<sup>-1</sup> Chroma I ROCK2 Inhibitor, 1.0 µg ml<sup>-1</sup> doxycycline and 4.0 µM Ara-C, and

switched to Neuron Medium: Brainphys (Stemcell, cat. no. 05790), Glutamax (1:100), Sodium Pyruvate (1 mM), Anti-Anti (1:100), N2 (1:100), B27 without vitamin A (1:50), BDNF (20 ng ml<sup>-1</sup>; R&D, cat. no. 248-BD-025), GDNF (20 ng ml<sup>-1</sup>; R&D, cat. no. 212), dibutyl cAMP (500 µg ml<sup>-1</sup>; Sigma, cat. no. D0627), L-ascorbic acid (200 µM; Sigma, cat. no. A4403), Natural Mouse Laminin (1.2 µg ml<sup>-1</sup>; ThermoFisher, cat. no. 23017015), and seeded at 7 × 10<sup>5</sup> cells per well onto Geltrex-coated (1:60 dilution coating) 12-well plates and incubated overnight at 37 °C. The next day, DIV6, Chroman I was removed from culture and Ara-C lowered to 2.0 µM with a full Neuronal Medium change. At DIV7, a full Neuronal Medium change was performed to remove doxycycline and Ara-C from culture, to allow for antibiotic-resistant genes silencing. From DIV7 onwards, half Neuronal Medium changes were performed every 72–96 h until mature DIV21 for collection.

**DIV36 iGABAs.** hiPSCs were collected in Accutase (Innovative Cell Technologies, cat. no. AT-104) for 5 min at 37 °C, dissociated into a single-cell suspension, quenched in DMEM, pelleted via centrifugation for 5 min at 1,000g, resuspended in StemFlex containing 25 ng ml<sup>-1</sup> Chroma I ROCK2 Inhibitor and 2.0 µg ml<sup>-1</sup> doxycycline (DIV0), seeded at 1.5–2 × 10<sup>6</sup> cells per well onto Geltrex-coated six-well plates (1:250 dilution coating) and incubated overnight at 37 °C. The next day, DIV1, hiPSCs were subjected to 48-h antibiotic selection by medium replacement with Induction Media: DMEM/F12 (ThermoFisher, cat. no. 10565018), Glutamax (1:100; ThermoFisher, cat. no. 10565018), N2 (1:100; ThermoFisher, cat. no. 17502048), B27 without vitamin A (1:50; ThermoFisher, cat. no. 12587010), Antibiotic-Antimycotic (1:100) with 1.0 µg ml<sup>-1</sup> doxycycline, 1.0 µg ml<sup>-1</sup> puromycin (Sigma, cat. no. P7255) and 250 µg ml<sup>-1</sup> hygromycin (Sigma, cat. no. 10687010). At DIV3, cells were treated with 4.0 µM Ara-C and 1.0 µg ml<sup>-1</sup> doxycycline to arrest proliferation and eliminate non-neuronal cells in the culture. At DIV5, immature neurons were dissociated with Accutase and 5 units per ml of DNase I at 37 °C for 7–10 min, quenched in DMEM, centrifuged for 5 min at 400g, resuspended in 25 ng ml<sup>-1</sup> Chroma I ROCK2 Inhibitor, 1.0 µg ml<sup>-1</sup> doxycycline and 4.0 µM Ara-C, and switched to Neuron Medium: Brainphys (Stemcell, cat. no. 05790), Glutamax (1:100), Sodium Pyruvate (1 mM), Anti-Anti (1:100), N2 (1:100), B27 without vitamin A (1:50), BDNF (20 ng ml<sup>-1</sup>; R&D, cat. no. 248-BD-025), GDNF (20 ng ml<sup>-1</sup>; R&D, cat. no. 212), dibutyl cAMP (500 µg ml<sup>-1</sup>; Sigma, cat. no. D0627), L-ascorbic acid (200 µM; Sigma, cat. no. A4403) and Natural Mouse Laminin (1.2 µg ml<sup>-1</sup>; ThermoFisher, cat. no. 23017015), and seeded at 7 × 10<sup>5</sup> cells per well onto Geltrex-coated (1:60 dilution coating) 12-well plates and incubated overnight at 37 °C. The next day, DIV6, Chroman I was removed from culture and Ara-C lowered to 2.0 µM with a full Neuronal Medium change. At DIV7, a full Neuronal Medium change was performed to remove doxycycline and Ara-C from culture, to allow for antibiotic-resistant genes silencing. From DIV7 onwards, half Neuronal Medium changes were performed every 72–96 h until mature DIV36 for collection.

#### CRISPR-KO gRNA library design (ThermoFisher) and validation

From the 102 highly penetrant LoF gene mutations associated with ASD (58 gene expression regulation, 24 neuronal communication genes, 9 cytoskeletal genes and 11 multifunction genes)<sup>4</sup>, gene ontology and primary literature research identified 26 epigenetic modifiers specifically involved in chromatin organization, rearrangement and modification. ASD gene expression (RNA-seq reads per kilobase million (RPKM) in iGLUTs) was plotted against significance of ASD association (transmission and de novo association analysis (TADA) FDR values), to ensure selection of genes with the highest expression and highest clinical association. Gene expression was confirmed across development in the brain (BrainSpan<sup>113</sup>), and in bulk RNA-seq and scRNA-seq. Twenty-one epigenetic modifiers (*ASH1L*, *ASXL3*, *ARID1B*, *CHD2*, *CHD8*, *CREBBP*, *KDM5B*, *KDM6B*, *KMT2C*, *KMT5B*, *MBD5*, *MED13L*, *PHF12*, *PHF21A*, *POGZ*, *PPP2R5D*, *SETD5*, *SIN3A*, *SKI*, *SMARCC2*,

*WAC*) as well as two transcription factors with putative roles as chromatin regulators (*FOXP2*, *BCL11A*) were selected. Gene regulatory transcription factors, general transcription factors and DNA replication genes were excluded. Three extensively studied synaptic genes (*NRXN1*, *SCN2A*, *SHANK3*) with roles in ASD were included as positive controls and three under-explored genes for ASD roles in neuronal communication genes (*ANK3*, *DPYSL2*, *SLC6A1*) were also included in the library.

Individual DNA samples from glycerol stocks of Invitrogen Lenti-Array Human CRISPR Library gRNAs-PuroR (ThermoFisher, cat. no. A31949) (3–4 individual gRNAs per gene, Supplementary Table 1) were prepared using GeneJET Plasmid Miniprep Kit (K0503) and pooled at an equimolar ratio and a fivefold ratio of scramble control gRNA plasmid. Library quality was confirmed by restriction enzyme digest (10x Cutsart NEB) and agarose gel purification using QIAquick Gel Extraction Kit (cat. no. 28706) to check library purity, followed by MiSeq for gRNA count distribution. Based on the abundance of gRNAs from MiSeq, four NDD gene targets were highly unlikely to be resolved in the final experiments (*POGZ*, *PPP2R5D*, *SHANK3*, *SLC6A1*) and three with low abundance were less likely to be resolved (*SCNA2*, *FOXP2*, *DYPSL2*).

Lentiviral Cas9v2-HygroR (Addgene, cat. no. 98291) and pooled LentiArray-gRNA-PuroR CRISPR-KO Library were packaged as high-titer lentiviruses (Boston Children's Hospital Viral Core) and experimentally titrated in each cell type. The highest viable multiplicity of infection (MOI) was used for Cas9v2 and MOI < 0.5 for the lentivirus gRNA pool library.

**CRISPR and gRNA delivery.** Lentiviral Cas9v2-HygroR (Addgene, cat. no. 98291) transduction of iNPCs, day 4 (iGLUTs) or day 5 (iGABAs), occurred via spinfection (1 h at 1,000g) and was followed by 72 h of hygromycin (250 µg ml<sup>-1</sup>) (except for iGABAs, which express inducible hygromycin resistance at this stage). Pooled Invitrogen LentiArray Human gRNA-PuroR CRISPR-KO Library gRNAs (ThermoFisher, cat. no. A31949) (MOI 0.3–0.5) were transduced via spinfection 3 d before collection (for example, day 4 for day 7 iGLUTs, day 18 for day 21 iGLUTs, day 33 for day 36 iGABAs), with fresh medium containing puromycin (1 µg ml<sup>-1</sup>) added 16–24 h post transduction of gRNAs. For mature iGLUTs and iGABAs, as doxycycline was removed from medium at DIV7, and by DIV18 neurons had lost transcription factor-linked antibiotic resistance, at 24 h post transduction (DIV19 or DIV34) puromycin (1 µg ml<sup>-1</sup>) and hygromycin (250 µg ml<sup>-1</sup>) were added to media for 48-h of antibiotic selection before collection.

#### Dissociation of different neural cell types to single cells for scRNA-seq assays

Cells were dissociated 72 h post gRNA library delivery for single-cell sequencing, as iNPCs, DIV7 and DIV21 iGLUTs, or DIV36 iGABAs, as follows:

iNPCs and DIV7 iGLUTs were dissociated in Accutase for 5 min at 37 °C, washed with DMEM/10%FBS, centrifuged at 1,000g for 5 min, gently resuspended and counted.

DIV21 iGLUTs and DIV36 iGABAs were dissociated with papain. Papain was pre-warmed (39 °C) for 30 min in HBSS (ThermoFisher, cat. no. 14025076), HEPES (10 mM, pH 7.5), EDTA (0.5 mM) and papain (0.84 mg ml<sup>-1</sup>; Worthington-Biochem, cat. no. LS003127). The cells were washed with PBS-EDTA (0.5 mM) and 300 µl of papain solution and 5 units of DNase I were added per well of a 12-well plate and incubated at 37 °C for 10–15 min, 125 rpm. Dissociation was quenched with DMEM-10%FBS. Detached neurons were broken by gentle manual pipetting, pelleted at 600g for 5 min, resuspended in DMEM-10%FBS, filtered through a cell strainer and counted and submitted for 10X sequencing.

Cells were loaded into 10X in four lanes per cell type, targeting 20,000 cells per lane for a total of ~80,000 targeted cells per cell type. scRNA-seq was performed at Yale Genomics Core with the 10X single-cell 5' v2 HT with CRISPR barcode kit.

### Bulk RNA-seq and CRISPR editing efficiency evaluation

The HI hESC line with iCas9 (NIHhESC-10-0043), generously provided by the Huangfu Lab, was used to assess the editing efficiency of the gRNAs<sup>44,114</sup> and to conduct the mitochondrial pooled and arrayed experiments. NPCs were generated using the dual SMAD inhibition approach per the STEMdiff SMADi Neural Induction Kit protocol (STEMCell Technologies, cat. no. 08581).

To validate gene KO, NPCs were transduced with LV particles carrying four gRNAs per target gene. After 48 h of selection with  $1 \mu\text{g ml}^{-1}$  puromycin, Cas9 expression was induced by adding dox at  $2 \mu\text{g ml}^{-1}$  for 72 h. Following induction, cells were collected for bulk RNA-seq. Total RNA was extracted using TRIzol reagent (Invitrogen). PolyA RNA-seq library preparation and sequencing were conducted at the Yale Center for Genomic Analysis (YCGA). Raw fastq files were quality-checked (FastQC; RRID:SCR\_014583), then mapped to human genome reference hg38 (STAR<sup>115</sup>). gRNA-targeted loci for each sample were extracted (SAMtools<sup>116</sup>; RRID:SCR\_002105). Variation/small insertion/deletion at the site of interest and mutation efficiency at corresponding loci were called (CrisprVariants R package<sup>117</sup>), after excluding possible germline variants from Cas9-noninduced samples.

NDD gene effects were resolved either (1) relative to a nontargeting control in the context of iCas9, owing to administration of doxycycline (for example, mitochondrial assays), or (2) across isogenic conditions expressing targeting gRNA, with the comparison being whether iCas9 expression was induced by administration of doxycycline (for example, replication assays).

### Proliferation and neurogenesis analysis

For proliferation analysis using Ki-67, NPCs were seeded into 24-well plates and either treated with doxycycline (induced) to activate Cas9 or left untreated (uninduced). The cells were cultured for 7 d, representing approximately three NPC generations. On day 7, cells were collected, and  $\sim 1 \times 10^6$  cells were stained with Ki-67-FITC (cat. no. 130-117-803, Miltenyi Biotec) using the Foxp3/Transcription Factor Staining Buffer Set (cat. no. 00-5523, Invitrogen), following the manufacturer's protocol.

To evaluate the effects of gene KOs on neurogenesis and gliogenesis, transduced NPC-iCas9 lines were spontaneously differentiated into human cortical neurons and glial cells. Briefly,  $1 \times 10^6$  cells were seeded in GelTrex-coated (1:5) six-well plates and cultured in complete neuronal media containing BrainPhys Neuronal Medium, Glutamax (100X), Sodium Pyruvate (100 mM), B27 (-RA) supplement ( $\times 50$ ), N2 ( $\times 100$ ), Anti-Anti ( $\times 100$ ), Natural Mouse Laminin ( $1 \text{ mg ml}^{-1}$ ), dbcAMP ( $500 \text{ ng ml}^{-1}$ ), L-ascorbic acid ( $200 \mu\text{M}$ ), BDNF ( $20 \mu\text{g ml}^{-1}$ ) and GDNF ( $20 \mu\text{g ml}^{-1}$ ). Medium was refreshed every 3 d. On day 25, cells were collected and stained for FACS analysis using surface markers previously described<sup>118</sup> to differentiate NPCs (CD184<sup>+</sup>/CD44<sup>-</sup>/CD24<sup>+</sup>), neurons (CD184<sup>+</sup>/CD44<sup>-</sup>/CD24<sup>+</sup>) and glia (CD184<sup>+</sup>/CD44<sup>+</sup>). CD271, a marker for mesenchymal stem cells, was excluded from the original panel as NPCs were pre-purified via FACS using CD133<sup>+</sup>/CD184<sup>+</sup>/CD271<sup>-</sup> markers before differentiation. A minimum of 50,000 cells per gate were acquired using a BD LSRFortessa Cell Analyzer at the Yale Flow Cytometry Core. Flow cytometry data were analyzed (FlowJo v10.10 Software; RRID:SCR\_008520; BD Life Sciences).

The full protocol is available from protocols.io via <https://doi.org/10.17504/protocols.io.j8nlkyew1g5r/v1>.

All statistical analyses for flow cytometry assessment were conducted (GraphPad Prism version 9.5.1 (528) for macOS; RRID:SCR\_002798; GraphPad Software). Each well was treated as an independent replicate. Differences between KO (induced) and control (uninduced) groups were assessed by comparing the median fluorescence intensity of the target fluorophore using an unpaired *t*-test with Welch correction to account for individual group variance. Multiple comparisons were corrected using the FDR method with a

two-stage step-up procedure (Benjamini, Krieger and Yekutieli) at an FDR threshold of 5%.

### FACS analysis of mitochondrial membrane potential and CRISPR screen read-out via amplicon sequencing

For our mitochondrial assays, we used a nearly identical library (same backbone, guide density and control set) screened exclusively in the HI inducible Cas9 (HI-iCas9) hPSC line. Mitochondrial inner membrane potential was measured in HI-iCas9, following differentiation to NPCs or iGlut on day 21. Cells were collected, counted and aliquoted at  $1 \times 10^6$  cells per sample. JC-1 dye (MitoProbeJC-1 Assay Kit; Invitrogen, cat. no. M34152) was dissolved in DMSO at a stock concentration of  $200 \mu\text{M}$  and added to each sample to achieve a final concentration of  $2 \mu\text{M}$ , then incubated for 30 min at  $37^\circ\text{C}$  in 5%  $\text{CO}_2$ . A  $50 \mu\text{M}$  CCCP control was included to induce complete mitochondrial depolarization. After staining, cells were washed once in their respective culture medium, resuspended in FACS buffer (Invitrogen eBioscience Staining Buffer, cat. no. 00422226) and analyzed immediately on a ThermoFisher 'Bigfoot' spectral cell sorter using 488-nm excitation with 525/50-nm (FITC) and 585/40-nm (PE) emission filters. Debris and doublets were excluded by forward/side scatter gating, and CCCP-treated samples were used to define FITC and PE gates. Approximately  $1 \times 10^6$  events per sample were recorded. Cells were then pelleted (300g, 5 min) and genomic DNA extracted using the Qiagen DNeasy Blood & Tissue Kit (cat. no. 69504).

Unique molecular identifier (UMI)-tagged amplicon libraries were generated in three PCR steps. In PCR-1, genomic DNA was amplified with Platinum II Hot-Start PCR Master Mix (Invitrogen, cat. no. 14000012) and UMI-containing primers (Forward: 5'-ACACTCTTCCCTACACGACGCTCTCCGATCTACGTGACGTAGAAAGTAATAATTTCTGGGT-3'; Reverse: 5'-GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTN(25)NNNNNNNNACTCGGTGCCACTTTTCAA-3') under the following conditions:  $94^\circ\text{C}$  for 2 min; 4–6 cycles of  $98^\circ\text{C}$  for 5 s,  $60^\circ\text{C}$  for 15 s,  $60^\circ\text{C}$  for 30 s. The resulting  $\sim 180$ -base pair (bp) products were purified and concentrated using the Zymo DNA Clean & Concentrator-5 kit (cat. no. D4013) and eluted in  $10 \mu\text{l}$  of nuclease-free water (ThermoFisher, cat. no. AM9938). In PCR-2, purified product was amplified with adaptor primers (Forward: 5'-ACACTCTTCCCTACACGACGCTCTCCGATCT-3'; Reverse: 5'-GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-3') for 22 cycles under identical cycling conditions in a  $\sim 20$ - $\mu\text{l}$  reaction. A seven-cycle indexing PCR (PCR-3) was performed by the sequencing facility at YCGA before sequencing. Final libraries were sequenced on an Illumina NovaSeq platform (paired-end 150 bp, 5 million reads per sample).

Flanking sequences on both sides of each gRNA were trimmed using BBDuk, and reads were then mapped to gRNA reference sequences and counted (MAGeCK<sup>119</sup>; RRID:SCR\_025016). Raw counts for each gRNA were normalized to counts of scrambled gRNA. Abundance of each target gene was then calculated by summing of all gRNAs targeting that gene. The  $\log_2$ -transformed FCs of gRNA-target abundance were compared between PE-high samples and FITC-high samples.

The full protocol is available from protocols.io via <https://doi.org/10.17504/protocols.io.rm7vz98brgx1/v1>.

### Immunostaining

Cells were fixed with fixative solution (4% sucrose and 4% paraformaldehyde prepared in Dulbecco's PBS (DPBS)) for 10 min at room temperature. Following this, cells were washed twice with DPBS and incubated in blocking solution (2% normal donkey serum prepared in DPBS) supplemented with 0.1% Triton for 2 h at room temperature. After this, cells were incubated overnight at  $4^\circ\text{C}$  in the primary antibody solution prepared in blocking solution. Cells were washed three times with DPBS, incubated at room temperature in secondary antibody prepared in blocking solution, then washed three times with DPBS. In the second wash, cells were incubated in DBPS supplemented with DAPI (Sigma D9542,  $1 \mu\text{g ml}^{-1}$ ) for 2 min at room temperature.

Antibody	Species	Vendor	Cat. no.	RRID	Dilution
Anti-MAP2	Chicken	Invitrogen, Abcam	PA1-10005, ab5392	AB_1076848, AB_2138153	1:1,000
Anti-Nestin	Rabbit	Millipore	ABD69	AB_2744681	1:200
Anti-vGLUT1	Rabbit	Synaptic Systems	135-303	AB_887875	1:200
Anti-GABA	Rabbit	Sigma-Aldrich	A2052	AB_477652	1:200
TOMM20	Mouse	Santa Cruz Biotechnology	sc-17764	AB_628381	1:200
Total OXPHOS	Rabbit	Abcam	AB-317270	not available	1:500
Anti-mouse	Donkey	Jackson ImmunoResearch	715-605-151	AB_2340863	1:500
Anti-rabbit	Donkey	Jackson ImmunoResearch	711-545-152	AB_2313584	1:500
Anti-chicken	Donkey	Jackson ImmunoResearch	715-605-150, 703-545-155	AB_2340862, AB_2340375	1:500

Fixed cultures were acquired using a DragonFly Confocal Dual Spinning Disk confocal microscope, at  $\times 60$  magnification and 1.4 numerical aperture. All images were acquired with a fixed laser intensity and exposure time across experimental conditions. Four images were acquired per well, and 4–10 wells were acquired per experimental condition. Each well represents a biological replicate and statistical data point. Therefore, each replicate represents hundreds of  $\mu\text{m}^2$  of neuronal area and tens of thousands of individual mitochondria.

Mitochondrial morphology features were determined using the Surface module of Imaris 10.2 (RRID:SCR\_007370). Likewise, OXPHOS complex features were determined using the Surface module of Imaris 10.2. The Volume, Area and Sphericity features of the Surface module were selected for analysis. Mitochondrial networking features were determined using published, open-source methods<sup>120</sup>. A one-way ANOVA with a Sidák's multiple comparisons test was performed on data with GraphPad Prism 10.

To validate robustness and sensitivity of the microscopy assay, we treated day 14 iGluts overnight with carbonyl cyanide 4-(trifluoromethoxy) phenylhydrazone (FCCP) (Sigma-Aldrich, cat. no. SML2959) at 5  $\mu\text{M}$ , 10  $\mu\text{M}$  and 50  $\mu\text{M}$  doses. Following this, we conducted the immunostaining, mitochondrial structural analysis and statistical analyses outlined above.

### Seahorse XF Mito Stress Test

Day 5 iGLUTs were plated at  $1.65 \times 10^5$  cells per well in XF24 microplates (Agilent, cat. no. 100777-004) and cultured to day 21. At 1 h before measurement, growth medium was removed, leaving 50  $\mu\text{l}$  per well, and replaced with 1 ml of pre-warmed Seahorse XF DMEM (Agilent, cat. no. 103575-100) supplemented with 25 mM glucose (Agilent, cat. no. 103577-100) and 0.23 mM pyruvate (Agilent, cat. no. 103578-100). Plates were equilibrated for 1 h at 37 °C in a non-CO<sub>2</sub> incubator. Immediately before the assay, the medium was replaced with 500  $\mu\text{l}$  of fresh assay buffer. Oxygen-consumption rate was recorded on a Seahorse XFe24 Analyzer (Agilent) using the standard Mito Stress Test. The program consisted of three sequential injections—1.5  $\mu\text{M}$  oligomycin (Sigma-Aldrich, cat. no. 75351), 1.5  $\mu\text{M}$  FCCP (Sigma-Aldrich, cat. no. C2920) and a mix of 0.5  $\mu\text{M}$  rotenone (Sigma-Aldrich, cat. no. R8875) + 0.5  $\mu\text{M}$  antimycin A (Sigma-Aldrich, cat. no. A8674)—separated by four measurement phases (baseline plus post-injection 1–3). Each phase comprised three cycles of 3 min of mixing, 2 min of waiting and 3 min of measurement. After the assay, cells were lysed using M-PER Mammalian Protein Extraction Reagent (ThermoFisher, cat. no. 78501) supplemented with cOmplete Mini Protease Inhibitor Cocktail (Sigma-Aldrich, cat. no. 11836153001) and PhosSTOP (Sigma-Aldrich, cat. no. 4906845001), according to the manufacturer's instructions. Total protein concentrations were determined using the Pierce Dilution-Free Rapid Gold BCA Protein Assay (ThermoFisher, cat. no. A55860), and oxygen-consumption rate values were normalized to total protein content.

### CRISPR organoid assays

H1-hESC-iCas9 cells were transduced with a pooled gRNA library containing four gRNAs per target gene, with 20% of the library comprising nontargeting gRNAs. Following selection with 1  $\mu\text{g ml}^{-1}$  puromycin, the established cell line was used to generate cortical organoids following a well-established protocol<sup>121</sup> with slight modifications. In brief, embryoid bodies were generated using AggreWell plates (Stemcell Technologies) according to the manufacturer's instructions. Once formed, embryoid bodies were transferred to ultralow-attachment 10-cm plates (Corning) for further culture. Patterning was initiated using StemFlex base media (A3349401, cat. no. Gibco) supplemented with 100 nM LDN-193189 (x) and 10  $\mu\text{M}$  SB431542 (x). The medium was refreshed daily. Organoids were cultured on an orbital shaker at 53 rpm for the duration of the protocol. On day 6, the patterning medium was replaced with growth media: Neurobasal A medium (cat. no. 10888022, Gibco), 1  $\times$  GlutaMAX (cat. no. 35050061, Gibco) and 1  $\times$  B27 (cat. no. 12587010, Gibco), supplemented with 20 ng  $\text{ml}^{-1}$  FGF (PeproTech) and 20 ng  $\text{ml}^{-1}$  EGF (PeproTech). On day 14, Cas9 expression was induced by treating the organoids with 2  $\mu\text{g ml}^{-1}$  doxycycline (Sigma-Aldrich) for 72 h. From day 25, FGF and EGF were replaced with 20 ng  $\text{ml}^{-1}$  BDNF (PeproTech) and 20 ng  $\text{ml}^{-1}$  NT-3 (PeproTech). Media changes were performed every other day. Starting from day 42, organoids were maintained in growth medium without additional supplements. Medium was refreshed 2–3 times per week.

The organoids were maintained in culture for ~80 d, at which point five organoids from three biological replicates were collected for DNA extraction using the DNeasy Blood & Tissue Kit (cat. no. 69504, Qiagen). Extracted DNA was subjected to PCR amplicon sequencing with UMIs using a three-step PCR protocol. In the first step (PCR-1), UMI-containing primers (5'-ACACTCTTCCCTACAC-GACGCTCTCCGATCTACGTGACGTAGAAAGTAATAATTTCTGGGT-3' and 5'-GTGACTGGAGTTCAGACGTGTGCTCTCCGATCTN(25252525)NNNNNNNNACTCGGTGCCACTTTTTCAA-3') were used for four cycles. PCR-2 utilized adaptor primers (5'-ACACTCTTCCCTACACGACGCTCTCCGATCT-3' and 5'-GTGACTGGAGTTCAGACGTGTGCTCTCCGATCT-3') for 22 cycles. PCR-3, performed by the sequencing facility, added sample-specific indexing in seven additional cycles. The prepared libraries were sequenced on a NovaSeq platform with paired-end 150-bp reads, generating 10 million reads per sample, at the YCGA.

Fragments amplified by PCR were sequenced on a NovaSeq 6000 sequencer paired-end at 150 bp with ~10 million reads per sample. Flanking sequences on both side of gRNAs were trimmed using BBDuk, and reads were then mapped to gRNA reference sequences and counted using the MAGeCK package<sup>119</sup>. Raw counts for each gRNA were normalized to counts of scrambled gRNA. Abundances of each gRNA-target gene were then calculated by sum of all gRNAs targeting that gene after excluding gRNAs with low KO efficiency (<5%). Average log<sub>2</sub>-transformed FCs of gRNA-target abundance were

compared between doxycycline-induced versus uninduced samples in day 77 samples.

### Analysis of single-cell CRISPR-KO screens in NPCs, DIV7 and DIV21 iGLUTs, and DIV36 iGABAs

mRNA sequencing reads were mapped to the GRCh38 reference genome using the Cell Ranger Software (RRID:SCR\_017344). To generate count matrices for GDO (gRNA) libraries, the kallisto indexing and tag extraction (kite) workflow was used. Count matrices were used as input into the R/Seurat package<sup>122</sup> (v.5.1.0; RRID:SCR\_007322) to perform downstream analyses, including quality control, normalization, cell clustering, GDO demultiplexing and covariate regression<sup>29,123</sup>. CRISPR-screen experiments in each cell type were processed independently. Within each cell type, ~100–80,000 cells were sequenced across four lanes. gRNA and RNA UMI feature counts were filtered removing the top and bottom deciles of cells based on distribution of counts in each cell type. The percentages of all the counts belonging to the mitochondrial, ribosomal and hemoglobin genes calculated using Seurat::PercentageFeatureSet were filtered with cell-type-specific thresholds, given the relatively high proportion of mitochondrial genes expressed in neurons. Mitochondrial, ribosomal and hemoglobin genes as well as MALAT1 were removed (`^RP[SL][[:digit:]]^RPLP[[:digit:]]^RPSA|^HB[AEGQ][[:digit:]]^HB[ABDMQ]^MT|^MALAT1$`). Lowly expressed genes, those that had at fewer than two read counts in 90% of samples, were also removed. 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. gRNA demultiplexing was performed using the Seurat::MULTIseqDemux function for each lane individually and then counts were merged across lanes (Supplementary Fig. 3b). In NPCs, 94,363 cells were retained after filtering and removal of negatively assigned cells, with 62.7% classified as doublets and 37.3% classified as singlets. In DIV7 and DIV21 iGLUTs, 57,685 and 31,473 cells were retained, with 34% and 9.8% doublets and 66% and 90.2% singlets, respectively. In DIV35 iGABAs, 64,462 cells were retained, with 48.3% doublets and 51.7% singlets. For all downstream analyses, only cells with 'singlet' gRNA classification were used (26,549–38,097 cells per experiment) (Supplementary Fig. 4c–e). The number of singlet cells by gRNA per cell type shown in Supplementary Fig. 6a,b.

### Cell-type-specific population heterogeneity correction

Gene-expression-based clustering was largely driven by cellular heterogeneity, cell quality and sequencing lane effects. gRNA identity was not correlated with these covariates (Supplementary Fig. 7), so we adjusted for transcriptomic variability arising from cellular heterogeneity by applying maturity and cellular subtype scores across both perturbed and nonperturbed cells. First, variation related to cell-cycle phase of individual cells was accounted for by assigning cell-cycle scores using Seurat::CellCycleScoring, which uses a list of cell-cycle markers<sup>124</sup> to segregate by markers of G2/M phase and markers of S phase. Second, to address variance due to cellular heterogeneity within a single experiment, we adapted the method applied by Seurat::CellCycleScoring to calculate a 'Maturity.Score' and 'Subtype.Score' for each cell based on cellular subtype (more variable in mature GABAergic neurons) and developmental time-point-specific markers (more variable in NPCs and immature iGLUTs) (Supplementary Tables 2 and 3). Cells with outlier maturity scores and subtype scores were removed from downstream analyses. RNA UMI count data were then normalized and log-transformed, and the percentages of mitochondrial, hemoglobin and ribosomal genes (markers of cell quality), lane, cell-cycle scores (Phase) and maturity scores regressed out using Seurat::SCTransform. The scaled residuals of this model represent a 'corrected' expression matrix, which was used for all downstream analyses.

Although demultiplexing assigned the correct guide identity to each cell, to remove 'false positives' whereby gRNAs were assigned

but gene expression was unperturbed, the transcriptomes of gRNA clusters were evaluated relative to scramble gRNAs, ensuring that cells assigned to a guide-tag identity class demonstrated successful perturbation of the targeted NDD gene. To remove subsequent 'false negatives', whereby a successful CRISPR-KO may not result in significant downregulation of the targeted gene<sup>29</sup> yet still achieve an overall transcriptomic profile distinct from scramble populations, we performed WNN analysis to assign clusters based on both guide-tag identity class and gene expression<sup>30</sup>. To identify successfully perturbed cells, the transcriptomes of gRNA clusters were compared with scramble-gRNA control clusters by differential gene expression analysis (Wilcoxon rank sum), comparing each cluster with all other clusters. Nontargeting WNN clusters and KO gRNA WNN clusters were filtered by setting a quantile base average expression threshold of target genes based on the distribution of target gene average expression across all other clusters. Clusters were the collapsed by gRNA identity; gRNAs with less than 75 cells were removed from analysis. These cells were then used for downstream differential gene-expression analyses<sup>125</sup>. For each cell type individually, single-cell gene expression matrices were Pseudobulked using scuttle::aggregateAcrossCells function across lanes (four pseudo-bulk samples per perturbation) and lowly expressed genes were removed (leaving 18,000–22,000 genes), followed by edgeR/limma differential gene expression analysis (RRID:SCR\_010943). Concordance between Wilcoxon rank sum differential gene expression analysis using single-cell data and limma:voom using Pseudobulked data was assessed for each gene.

Altogether, Wilcoxon rank sum was applied to measure NDD gene knockdown from single-cell DEG analysis. Given the concordance between the DEG results using single-cell Wilcoxon and pseudo-bulk limma:voom (Supplementary Fig. 6c), all main figures and all Supplementary figures thereafter applied pseudobulked data analyzed with limma.

To validate whether the high correlation within-cell types was due to exactly the same scramble control cells, we re-performed DEG analyses using random selection of a subset of scramble cells for each cell type (Supplementary Fig. 8). Briefly, for each gene, 50% (if number of pseudobulked sample cells >50) or 80% (if number of pseudobulked sample cells <50) of scramble cells were randomly selected using the sample function from R. DEG analyses were then performed as described above using the limma/dreamlet package between KOs and a subset of scrambles different among genes. The process was repeated three times to avoid random selection bias and the median of each gene logFC was used as the final logFC. Average overlap of random scramble cells across different genes is ~50%.

### Meta-analysis of gene expression across perturbations<sup>31</sup>

Across NDD KOs, DEGs were meta-analyzed (METAL<sup>126</sup>), and 'convergent' genes were defined as those with significant and shared direction of effect across all NDD gene perturbations and with nonsignificant heterogeneity (FDR-adjusted  $P_{\text{meta}} < 0.05$ , Cochran's heterogeneity  $Q$ -test  $P_{\text{Het}} > 0.05$ ). To test convergence between NDD KOs, meta-analyses were performed across all possible combinations of 2–5 KO perturbations with and without sub-setting for those shared across cell types (>40,000 combinations across cell types) (Supplementary Data 1).

### Bayesian bi-clustering to identify convergent networks<sup>31</sup>

Across NDD KOs, convergent networks were generated by Bayesian bi-clustering<sup>127</sup> and undirected gene co-expression network reconstruction from the NDD KOs. Not constrained by statistical cut-offs, and able to capture the effect of more lowly expressed genes, convergent networks may be a more sensitive measure of convergence. Networks were built based on bi-clustering (BicMix)<sup>128</sup> using log<sub>2</sub>CPM (counts per million) expression data from all the replicates across each of the NDD gene sets and scramble gRNA jointly. We performed 40 runs of BicMix with the default parameters on these data and the output from

iteration 400 of the variational Expectation-Maximization algorithm was used. Target Specific Network reconstruction<sup>128</sup> was performed to identify convergent networks across all possible combinations of the nine NDD gene KO perturbations shared across cell types ( $n = 502$  combinations per cell type) and randomly sampled combinations of 2–21 KO perturbations without sub-setting for those shared across cell types ( $n = 1,400–2,300$  combinations).

### Influence of functional similarity on convergence degree

To test the influence of functional similarity and brain co-expression between KOs on convergence and compare the degree of convergence between the same KOs in different cell types, we established two methods for defining and measuring convergence. First, gene-level convergence using meta-analysis as described above, with the strength of convergence for each set defined as ratio of convergent genes to the average number of DEGs.

$$\text{Gene level convergence} = \frac{n\text{Convergent genes}}{\text{Mean}(\sum_1^N n\text{DEGs})}$$

Second, network-level convergence based on undirected network reconstruction from Bayesian bi-clustering as described above. Bi-clustering identifies co-expressed genes shared across the downstream transcriptomic impacts of any given set of KO perturbations; thus, the resolved networks are the transcriptomic similarities between distinct perturbations (convergence). We calculated the 'degree of convergence' for each network based on a previously described metric<sup>31</sup>. Briefly, convergence scores are based on: (1) network connectivity, as defined by the sum of the clustering coefficient ( $C_p$ ) and the difference in average length path ( $L_p$ ) from the maximum average length path resolved across all possible sets  $[(\max)L_p - L_p]$ ; (2) similarity of network genes based on biological pathway membership scored by taking the sum of the mean semantic similarity (semsim) scores<sup>129</sup> between all genes in the network; and (3) minimum percentage duplication rate across 40 runs. Duplication thresholds are network-dependent and a metric of confidence in the connections.

$$\begin{aligned} \text{Network level convergence} &= C_p + [(L_p) - L_p] \\ &+ \text{mean} \left( \sum_1^N \text{MF semsim} + \text{BP semsim} + \text{CC semsim} \right) \\ &+ n\text{Duplication runs}/n\text{Total runs} \end{aligned}$$

Functionally, similarity scores across the NDD KO genes represented in each set were calculated using: (1) Gene Ontology Semantic Similarity Scores: the average semantic similarity score based on Gene Ontology pathway membership within Biological Pathway (BP), Cellular Component (CC) and Molecular Function (MF) between NDD genes in a set<sup>129</sup>; and (2) brain expression correlation score: based on the strength of the correlation in NDD gene expression in the Common Mind Consortium (CMC) ( $n = 991$  after quality control) post-mortem DLPFC gene expression data.

We performed Pearson's correlation analysis (Holm's-adjusted  $P$ ) on similarity scores and the degree of network convergence to determine the influence of the similarity of the initial KO genes on downstream convergence. We compared the average strength of convergence across cell types using a parametric Welch's  $F$ -test and pairwise Games–Howell test.

### Enrichment analysis of convergence for risk loci using MAGMA

We intersected cross-cell-type perturbation-specific and cross-perturbation cell-type-specific gene-level convergence with genetic risk of psychiatric and neurological disorders/traits (attention-deficit/hyperactivity disorder<sup>130</sup>, anorexia nervosa<sup>131</sup>, ASD<sup>2</sup>, alcohol dependence<sup>132</sup>, bipolar disorder<sup>133</sup>, cannabis use disorder<sup>134</sup>,

major depressive disorder<sup>135</sup>, obsessive-compulsive disorder<sup>136</sup>, post-traumatic stress disorder<sup>137</sup>, SCZ<sup>138</sup>, Cross disorder<sup>139</sup>, Alzheimer disease<sup>140</sup>, Parkinson disease<sup>141</sup>, amyotrophic lateral sclerosis<sup>142</sup>, Tourette's<sup>143</sup>, migraine<sup>144</sup>, chronic pain<sup>145</sup> and neurotic personality traits<sup>146</sup>, GWAS summary statistics) using multi-marker analysis of genomic annotation (MAGMA)<sup>23</sup>. SNPs were mapped to genes based on the corresponding build files for each GWAS summary dataset using the default method, snp-wise = mean (a test of the mean SNP association). A competitive gene set analysis was then used to test enrichment in genetic risk for a disorder across gene sets with an FDR < 0.05.

To test whether observed effects were due to the differential size of the gene sets for each GWAS or owing to the fact that DEGs are more likely to include neural genes, which are more likely to be associated with brain disorder, GWAS sets were filtered for genes expressed in each cell type before enrichment testing, and enrichment tests were performed after randomly down-sampling GWAS gene sets to 100, 250, 500, 750 and 1,000 genes (Supplementary Fig. 9), and performed ten times within each set size (that is, 50 tests for each GWAS).

### Over-representation analysis, functional enrichment annotation and biological theme comparison of convergence

To identify pathway enrichments unique to individual KOs, convergent genes and convergent networks based on zebrafish behavioral subgroups (see the zebrafish methods below), we performed biological theme comparison and gene set enrichment analysis (GSEA) (ClusterProfiler<sup>147</sup>; RRID:SCR\_016884). Using FUMAGWAS: GENE2FUNC, the 102 ASD genes were functionally annotated and an over-representation gene-set analysis for each convergent gene set was performed<sup>148</sup>. Using WebGestalt (WEB-based GENE SeT Analysis Toolkit<sup>149</sup>; RRID:SCR\_006786), over-representation analysis was performed on all convergent gene sets against publicly available gene set lists GeneOntology, KEGG, DisGenNet and Human Phenotype Ontology, and a curated gene list of rare-variant targets associated with ASD, SCZ and ID<sup>25</sup>.

### Random forest prediction model of convergence strength

To determine how well functional similarity between KOs can predict gene-level and network-level convergence, we trained a random forest model<sup>32</sup> (randomForest package in R; RRID:SCR\_015718) for each type of convergence, evaluated the model in an independent internal dataset and validated the model in an external CRISPRa activation screen<sup>31</sup>. Data from randomly tested gene combinations (2–5 KO sets at the gene level and 2–10 KO sets at the network level) tested across cell types were randomly down-sampled into a training set (70%) and a testing set (30%)—all with comparable proportions of data by cell type. The random forest model was trained with bootstrap aggregation using CC, MF and BP semantic similarity scores, brain expression correlation, number of genes and cell type as predictors. The random forest linear regression model was evaluated in the testing data by comparing actual values with predicted values, estimating the root mean squared error and performing Pearson's correlations. Predictor models were validated using an external dataset of ten CRISPR-activation perturbations of SCZ common variant target genes with multifunctional annotations broadly grouped as signaling/cell communication (*CALN1*, *NAGA*, *FES*, *CLCN3*, *PLCL1*) and epigenetic/regulatory (*SF3B1*, *TMEM219*, *UBE2Q2L*, *ZNF804A*, *ZNF823*)<sup>31</sup>, and assessed based the root mean squared error and Pearson's correlation between actual and predicted convergence strength.

### LNCTP in silico model

To investigate the perturbation of ASD genes in silico, we adapted the LNCTP model<sup>33</sup> to predict the effects of changes in gene expression in the PFC, across neuronal and non-neuronal cell types. The LNCTP is defined as an energy model representing the joint distribution of a collection of phenotypes of interest conditioned on the genotype.

Since we are interested primarily in the effects of gene expression perturbations on the expression of other genes, we use only the imputation segment of the LNCTP model (excluding the prediction of higher-order phenotypes and cell–cell interactions).

The probabilistic model for the imputation-based LNCTP may be expressed as:

$$\begin{aligned}
 P_{\text{LNCTP}}(x_i, |, z_i) &= \exp(-E(x_i, |, z_i))E(x_i, |, z_i) \\
 &= x_{i0}^T J x_{i0} + \sum_g x_{i0g}^T b(z_i, \beta_g) + \sum_c (x_{ic}^T J_c x_{ic} + x_{ic}^T b_c) \\
 &\quad + \lambda \sum_g (x_{i0g} - f(z_i))^T x_{i,1\dots c,g}^2
 \end{aligned}
 \tag{1}$$

Here  $z_i$  represents the genotype of individual  $i$ , and  $x_i$  represents bulk and cell-type-specific gene expression from individual  $i$ . We further index the gene expression by  $C$  cell types (which are here: excitatory neurons, inhibitory neurons, oligodendrocytes, astrocytes, oligodendrocyte precursor cells, endothelial cells and microglia), which will be denoted  $x_1, x_2, \dots, x_C$ , and we will use  $x_0$  to denote the bulk expression. The variables  $f_{1\dots C}$  represent the estimated cell fractions in the bulk observations (predicted from the genotype,  $z$ ). The parameters of the model are  $\theta = \{\beta_{1\dots G}, J_{0\dots C}\}$  and  $\lambda$  acts as a hyperparameter. The parameters  $\beta_{1\dots G}$  and  $J_{0\dots C}$  reflect the gene-specific expression biases and pairwise interactions, respectively, whose nonzero elements are determined by the sparsity structure arising from expression quantitative trait loci (eQTLs) and gene regulatory network (GRN) linkages, respectively; the nonzero elements of  $J_c$  occur only between genes connected in the GRN of cell type  $c$ .

Further details on the training of the model in Eq. (1) can be found in ref. 33; here, we outline the specific differences in the training for the purposes of our analysis. As in ref. 33, we used genetics and expression data from post-mortem PFC samples from the PsychENCODE consortium. However, we group together samples from all higher-order phenotypes during training (control, SCZ, bipolar disorder and ASD), and split the data into three partitions of size 760, 100 and 100 for training, validation and testing, respectively (each including samples from all higher-order phenotypes). Further, we include all 29 CRISPR targeted genes, 102 NDD genes<sup>5</sup>, transcription factors<sup>33</sup> and neuropsychiatric transcriptome-wide association analysis (TWAS)-selected genes<sup>33</sup>, and the top 100 up- and downregulated CRISPR convergent genes in iGLUT and iGABA cells (400 genes in total), in the model, generating 1,325 genes in total. The expression quantitative trait loci (eQTL) and GRN linkages from PsychENCODE are then restricted to this subset of genes.

### LNCTP simulating perturbations

To perform perturbations in this model corresponding to the 29 CRISPR targeted genes, we use the following perturbation-conditioned version of the LNCTP model:

$$\begin{aligned}
 P_{\text{LNCTP}}(x_{i-(c^*, g^*)}, |, z_i, x_{i,c^*, g^*} = \{k, -k\}) \\
 &= \exp(x_{i-(c^*, g^*)}, |, z_i, x_{i,c^*, g^*} = \{k, -k\}) \\
 E(x_{i-(c^*, g^*)}, |, z_i, x_{i,c^*, g^*} = \{k, -k\}) &= x_{i0}^T J x_{i0} \\
 &\quad + \sum_g x_{i0g}^T b(z_i, \beta_g) + \sum_c (x_{ic}^T J_c x_{ic} + x_{ic}^T b_c) \\
 &\quad + \lambda \sum_g (x_{i0g} - f(z_i))^T x_{i,1\dots c,g}^2 + K \delta(x_{i,c^*, g^*} = \{k, -k\})
 \end{aligned}
 \tag{2}$$

where  $(c^*, g^*)$  denotes the perturbed gene and cell type, whose expression is set to  $k$  or  $-k$ ,  $\delta(a)$  is a delta function whose value is 0 if expression  $a$  is true, and 1 otherwise, and  $K$  is an arbitrarily large value. We perturb each of the CRISPR targeted genes in turn in the bulk network, using  $k = 2$ , and applying a negative perturbation to mimic the effect of the CRISPR perturbation. We note that, since the model is trained

on Z-scored log-normalized expression counts, this corresponds to introducing a large negative FC to the selected gene. The in silico-predicted log FCs per individual across all genes (per cell type) are then calculated by comparing the expected values before and after perturbation:

$$\Delta_{i,c,g} = \mathbb{E}_{P_{\text{LNCTP}}(\cdot, |, z_i)}[x_{i,c,g}] - \mathbb{E}_{P_{\text{LNCTP}}(\cdot, |, z_i, x_{i,c^*, g^*} = \{k, -k\})}[x_{i,c,g}]
 \tag{3}$$

and the final predicted log FCs are calculated by taking the expectation across individuals. We use the sampling approach in ref. 33 to evaluate the expectations in Eq. (3).

To perform perturbations across all 102 NDD genes, for efficiency, we learn a reduced model by removing the dependency on  $z_i$  in Eq. (1). We sample cell-type-specific expression values for each individual from the full model, and then fit the reduced model by refitting the model parameters to maximize the likelihood of the full data vectors (consisting of the original bulk and sampled cell-specific expression vectors for each individual). Perturbations are performed in the reduced model as in Eq. (2) and FCs are calculated as in Eq. (3), while removing the dependency on  $z_i$  and the  $i$  subscripts, respectively.

### LNCTP in silico convergent genes

To identify in silico convergent genes for a set of perturbations,  $S = \{(c_1^*, g_1^*), \dots, (c_N^*, g_N^*)\}$ , we calculate  $\Delta_{c,g}$  using Eq. (3) for each perturbation, writing  $\Delta_{c,g}^{c^*, g^*}$  for the log FC to  $(c, g)$  generated by applying perturbation  $(c^*, g^*)$ , and  $\Delta_{c,g}^S$  for the vector of log FCs created by applying all perturbations in  $S$ . Then, the set of in silico convergent genes for  $S$  is found by selecting those for which  $P_{\text{sign}}(\Delta_{c,g}^S [|\Delta_{c,g}^S| \geq \tau]) < 0.1$ , where  $P_{\text{sign}}(\cdot)$  is the  $P$  value from a two-tailed one-sample sign-test, and juxtaposition represents element-wise multiplication of vectors. The threshold  $\tau$  is introduced to reduce noise from perturbations that are estimated to generate small log FCs, and throughout we set  $\tau = 0.3$ .

For the comparison of in silico convergent genes derived from different perturbation sets  $S$ , we apply two-sided hypergeometric tests to the gene sets defined as above (using all 1,325 genes in our model as the background set). For GSEA of convergent genes derived from  $S$ , we apply clusterprofiler<sup>147</sup> to the full set of genes in our model, ranked by  $P_{\text{sign}}(\Delta_{c,g}^S [|\Delta_{c,g}^S| \geq \tau])$  as defined above.

### LNCTP semantic distance test

To test the semantic distance between enriched terms for two sets of perturbations,  $S_1$  and  $S_2$ , we generate the set of enriched terms  $T_1$  and  $T_2$  by applying GSEA to each set as described above (using Benjamini–Höcherberg correction and an FDR threshold of 0.2 to select enriched terms  $T_1$  and  $T_2$ ). We then calculate the similarity between terms  $t_1$  and  $t_2$  by evaluating  $s(t_1, t_2) = |G(t_1) \cap G(t_2)| / |G(t_1) \cup G(t_2)|$ , where  $G(t)$  denotes the set of genes occurring in the leading edge of term  $t$ . We test for a significant semantic distance between  $S_1$  and  $S_2$  by evaluating  $s(t_1, t_2)$  between all pairs  $t_1 \in S_1, t_2 \in S_2$ , versus all pairs  $t_1 \in S_1, t_2 \in S_1$  and  $t_1 \in S_2, t_2 \in S_2$ , and applying a one-sided rank-sum test for the smaller similarity in the former pairs versus the latter.

### Transcriptional correlations between hiPSC-derived neural cells, fetal and adult brain cell types and the zebrafish brain

We compared wild-type (WT) zebrafish brain expression with gene expression in our hiPSC-derived models and with sign-cell expression data for the fetal and adult PFC (PsychENCODE<sup>150,151</sup>; [http://resource.psychencode.org/Datasets/Derived/SC\\_DeComp/DER-20\\_Single\\_cell\\_expression\\_processed\\_TPM.tsv](http://resource.psychencode.org/Datasets/Derived/SC_DeComp/DER-20_Single_cell_expression_processed_TPM.tsv)). We first filtered zebrafish gene names and converted them to the appropriate *Homo sapiens* orthologs using the R package orthogene (v3.2.1<sup>152</sup>); genes without matched orthologs were dropped from both species. Pseudo-bulk expression data from scramble control cells were used as the baseline expression across NPCs, day 7 iGLUTs, day 21 iGLUTs and day 36 iGABAs. Pearson's correlation coefficients between in vitro cells, fetal and adult

post-mortem brain cells and zebrafish brain were calculated and a Bonferroni correction applied.

## Zebrafish

All procedures involving zebrafish were conducted in accordance with IACUC (protocol no. 2024-20054) regulatory standards at Yale University. Zebrafish larvae were raised at 28 °C on a 14-h:10-h light:dark cycle. Larvae were grown in 150-mm Petri dishes in blue water (0.3 g l<sup>-1</sup> Instant Ocean, 1 mg l<sup>-1</sup> methylene blue, pH 7.0) at a density of 60–80 larvae per dish. Behavioral assays were conducted in zebrafish larvae at 5–7 days post fertilization (dpf). At these developmental stages, sex is not yet determined.

## Zebrafish mutant generation

We performed automated, high-throughput, quantitative behavioral profiling of larval zebrafish to measure arousal and sensorimotor processing as a read-out of circuit-level deficits resulting from gene perturbation<sup>41</sup>. We quantified 24 parameters across sleep–wake activity and visual-startle responses in 18 stable homozygous mutant or FO mosaic crispant lines for 15 NDD genes (Supplementary Tables 7 and 8). Stable zebrafish lines were generated by our lab (*arid1b*<sup>Δ7/Δ7</sup>, *chd2*<sup>Δ7/Δ7</sup>, *chd8*<sup>Δ7/Δ7</sup>, *chd8*<sup>Δ5/Δ5</sup>, *kdm5ba*<sup>Δ17/Δ17</sup>, *b*<sup>Δ14/Δ14</sup>, *kdm5ba*<sup>Δ4/4i</sup>, *b*<sup>Δ4/Δ4</sup>)<sup>41</sup> or provided as a generous gift from the Thyme lab (*ash1l*<sup>11i,Δ60,19i/1i,Δ60,19i</sup>, *kmt5b*<sup>Δ208,1i,Δ5/Δ208,1i,Δ5</sup>, *kmt2ca*<sup>Δ82,17i/Δ82,17i</sup>, *b*<sup>Δ6,Δ29/Δ6,Δ29</sup>, *nrxn1a*<sup>Δ218/Δ218</sup>)<sup>153,154</sup>.

*kmt5b* uab320/+; ZDB-ALT-220315-26  
*kmt2ca* uab318/+; ZDB-ALT-220315-24  
*kmt2cb* uab319/+; ZDB-ALT-220315-25  
*ash1l* uab305/+; ZDB-ALT-220315-11  
*chd8* ya509/+; ZDB-ALT-240514-1  
*chd8* ya510/+; ZDB-ALT-240514-2  
*kdm5ba* ya523/+; ZDB-ALT-240514-15  
*kdm5bb* ya525/+; ZDB-ALT-240514-17  
*kdm5ba* ya524/+; ZDB-ALT-240514-16  
*kdm5bb* ya526/+; ZDB-ALT-240515-1  
*nrxn1a* uab403/+; ZDB-ALT-220720-20

FO crispants for the following genes were generated as in ref. 155: *chd2*, *kdm6bab*, *mbd5*, *phf12ab*, *phf21aab*, *skiab*, *smarcc2*, *wacab*. Briefly, we designed two CRISPR crRNAs per allele, prioritizing early exons for targeting. CRISPR RNPs were assembled individually and then combined before injection at the one-cell stage. The number of scrambled guides injected into the control group was matched to the number of CRISPR guides used for the experimental group. Injected embryos were raised to 5 dpf at which point the behavioral assays (described below) were conducted. We identified unique behavioral fingerprints for each NDD gene mutant, revealing convergent and divergent phenotypes across mutants (Supplementary Fig. 22b). To classify convergent behavioral subgroups that may share circuit-level functions, we performed correlation analyses with hierarchical clustering across mutants. We identified four distinct subgroups of NDD genes with highly correlated behavioral features (Fig. 7a).

## Behavioral assays

Larvae were placed into individual wells of a 96-well plate (cat. no. 7701-1651; Whatman) containing 650 μl of standard embryo water (0.3 g l<sup>-1</sup> Instant Ocean, 1 mg l<sup>-1</sup> methylene blue, pH 7.0) per well within a Zebrabox (ViewPoint Life Sciences). Locomotion was quantified with an automated video-tracking system (Zebrabox and ZebraLab software). The visual-startle assay (VSR) was conducted at 5 dpf (ref. 41). Briefly, to assess larval responses to lights-off stimuli (VSR-OFF), larvae were acclimated to white light for 1 h, and baseline activity was tracked for 30 min followed by five 1-s dark flashes with intermittent white light for 29 s. To evaluate larval responsivity to lights-on stimuli (VSR-ON), the assay was reversed, where larvae were acclimated to darkness for 1 h, and baseline activity was tracked for

30 min followed by five 1-s white light flashes with intermittent darkness for 29 s. For VSR-OFF and VSR-ON, six behavioral parameters were quantified using custom MATLAB code<sup>41</sup> (available from GitHub via <https://github.com/ehoffmanlab/Weinschutz-Mendes-et-al-2023-behavior>; <https://doi.org/10.5281/zenodo.7644898> (ref. 156)): (1) average intensity of all startle responses; (2) average post-stimulus activity; (3) average activity after first stimulus; (4) stimulus versus post-stimulus activity; (5) intensity of responses to the first stimulus; (6) intensity of responses to the final stimulus. To assess arousal responses in daytime and nighttime, the sleep–wake paradigm was conducted between 5 and 7 dpf, following the VSR-OFF and VSR-ON assays. During a 14-h:10-h white light:darkness cycle, larval activity and sleep patterns were tracked within the Zebrabox and analyzed with custom MATLAB code<sup>41,157,158</sup> (available from GitHub via (<https://github.com/JRihel/Sleep-Analysis/tree/Sleep-Analysis-Code>; <https://doi.org/10.5281/zenodo.7644073> (ref. 159))). Six behavioral parameters were quantified for daytime and nighttime: (1) total activity; (2) total sleep; (3) waking activity; (4) rest bouts; (5) sleep length; (6) sleep latency. Across VSR-OFF, VSR-ON and sleep–wake assays, we analyzed 24 parameters. For stable mutant lines, behavioral assays were performed blind to genotype using sibling- or cousin-matched larvae from heterozygous in-crosses with genotyping performed after each experiment. Experiments were repeated at least in duplicate for each line. Our analysis focused on homozygous mutants, which typically exhibit more robust phenotypes than heterozygotes. For CRISPR FO mutants, experiments were performed once with experimenters aware of which larvae were injected with scrambled versus gene-targeting RNPs. Larval movement was quantified blind to genotype using commercial tracking software as described above. Larvae were genotyped after each experiment to confirm the presence of on-target mutations. Baseline behavioral screening sample sizes are as follows: *kmt5b* WT *n* = 62, HOM *n* = 71; *mbd5* scrambled *n* = 39, crispant *n* = 56; *kdm5bab* 4i:del4 WT *n* = 42, HOM *n* = 55; *kdm5bab* del17:del14 *n* = 68, HOM *n* = 69; *phf12ab* scrambled *n* = 32, crispant *n* = 15; *skiab* scrambled *n* = 45, crispant *n* = 33; *chd2* scrambled *n* = 43, crispant *n* = 40; *smarcc2* scrambled *n* = 36, crispant *n* = 35; *kdm6bab* scrambled *n* = 40, crispant *n* = 34; *kmt2cab* WT *n* = 38, HOM *n* = 20; *wacab* scrambled *n* = 32, crispant *n* = 32; *arid1b* WT *n* = 81, HOM *n* = 81; *phf21a* scrambled *n* = 52, crispant *n* = 44; *chd8* del7 WT *n* = 180, HOM *n* = 164; *chd8* del5 WT *n* = 143, HOM *n* = 167; *ash1l* WT *n* = 19, HOM *n* = 23; *nrxn1a* WT *n* = 24 HOM *n* = 28. Based on our power analyses for baseline behavioral experiments, we estimated that a sample size of 28–41 fish per group provides 80% power to detect a 20–30% change in behavior (ANOVA,  $\alpha$  set at 0.05).

## Behavioral analysis

Linear mixed models (LMMs) were used to compare phenotypes of each behavioral parameter between homozygous mutant versus WT or crispant versus scramble-injected fish for each gene of interest. Variations of behavioral phenotypes across experiments were accounted for by including the date of the experiment as a random effect in LMM. Hierarchical clustering analysis was performed to cluster mutants and behavioral parameters based on signed  $-\log_{10}$ -transformed *P* values from LMM, where sign indicates direction of the difference in behavioral phenotype when comparing stable mutant with WT or crispant with scrambled-injected. Pearson's correlation analysis was used to assess correlations between mutants based on the difference in the 24 parameters. Difference was evaluated using signed  $-\log_{10}$ -transformed *P* values. For the behavioral data, we assumed the data distribution to be normal but not of equal variance. We did not formally test the normality assumption.

## Drug prioritization based on zebrafish pharmaco-behavioral profiles

NDD gene-associated mutant and crispant behavioral phenotypes were compared with a dataset of 376 US Food and Drug

Administration-approved drugs that were screened for their behavioral effects in larval zebrafish using the visual-startle and sleep–wake assays described above<sup>46</sup>. These drugs have a significant effect on at least two behavioral parameters (LMM,  $P < 0.05/3$ , corrected for three behavioral assays). Pearson's correlation analysis was used to identify drugs that significantly correlate (correlation  $> 0.5$ ,  $P < 0.05$ ,  $t$ -statistic) or anti-correlate (correlation  $< -0.5$ ,  $P < 0.05$ ,  $t$ -statistic) with mutant behavioral signatures (Supplementary Data 2 and 3).

### Drug prioritization based on perturbation signature reversal in LiNCs neuronal cell lines

To identify drugs that could reverse cell-type-specific convergence across different KOs, we used the Query tool from The Broad Institute's CMap Server<sup>45</sup>. 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 WTCS by dividing by the signed mean within each perturbation (NCS) and computes FDR as fraction of 'null signatures' (DMSO) where the absolute NCS exceeds reference signature. We prioritized drugs that were negatively enriched for convergent signatures specifically in neuronal cells (either neurons or NPCs with NCS  $\leq -1.00$ , FDR  $\leq 0.05$ ) and filtered for drugs that had clinical data in humans and paired behavioral phenotyping in zebrafish (Supplementary Data 2).

### Targeted drug rescue of behavioral phenotypes in zebrafish

For mutant  $\times$  drug experiments, larval activity was monitored from 5–7 dpf using the behavioral assays described above. Individual WT zebrafish larvae were added to each well of a 96-well plate containing 650  $\mu$ l of standard embryo water. A 5 mM stock solution of each compound dissolved in DMSO or DMSO alone (control) was pipetted directly into each well after which the visual-startle and sleep–wake assays were performed. Drugs were tested at a final concentration of 10  $\mu$ M (0.2% DMSO final concentration) in background-matched homozygous and WT larvae or crispant and scrambled control-injected larvae. For stable mutant lines, behavioral assays were performed blind to genotype using sibling- or cousin-matched larvae from heterozygous in-crosses with genotyping performed after each experiment. For CRISPR FO mutants, experiments were performed once with experimenters aware of which larvae were injected with scrambled versus gene-targeting RNPs. Larval movement was quantified blind to genotype using commercial tracking software as described above. Larvae were genotyped after each experiment to confirm the presence of on-target mutations. Mutant  $\times$  drug behavioral assays were repeated 1–3 $\times$  for stable mutant lines and 1 $\times$  for crispants. Sample sizes are as follows: *phf21a* + amiodarone scrambled DMSO  $n = 24$ , FO DMSO  $n = 22$ , FO AMIO  $n = 23$ ; *phf21a* + fluvoxamine scrambled DMSO  $n = 24$ , FO DMSO  $n = 22$ , FO FLUVO  $n = 21$ ; *chd2* + pravastatin WT DMSO  $n = 27$ , HOM DMSO  $n = 21$ , HOM PRAVA  $n = 23$ ; *kdm6b* + paclitaxel scrambled DMSO  $n = 24$ , FO DMSO  $n = 24$ , FO SIRO  $n = 23$ ; *kmt5b* + paclitaxel WT DMSO  $n = 20$ , HOM DMSO  $n = 22$ , HOM PACLI  $n = 26$ ; *kmt5b* + sirolimus WT DMSO  $n = 27$ , HOM DMSO  $n = 26$ , HOM SIRO  $n = 20$ ; *ash1l* + sunitinib WT DMSO  $n = 14$ , HOM DMSO  $n = 13$ , HOM SUN  $n = 12$ ; *ash1l* + ezetimibe WT DMSO  $n = 15$ , HOM DMSO  $n = 12$ , HOMEZE  $n = 9$ ; *ash1l* + rosuvastatin WT DMSO  $n = 27$ , HOM DMSO  $n = 22$ , HOM ROSU  $n = 22$ ; *ash1l* + repaglinide WT DMSO  $n = 11$ , HOM DMSO  $n = 16$ , HOM REPAG  $n = 14$ . Based on our power analysis, we estimate that a sample size of 10–12 fish per group will provide 95% power to detect a 40–43% change in behavior (ANOVA,  $\alpha$  set at 0.05), based on our previous pharmaco-behavioral screen<sup>160</sup>.

For behaviors that were nominally significantly different between mutant + DMSO and WT + DMSO ( $P < 0.06$ ), we characterized the effect of the mutant  $\times$  drug on behavior as: (1) 'exacerbated' (significant effect mutant + drug-v-WT  $>$  significant effect mutant-v-WT) if mutant behavior  $P$  value  $\leq 0.06$  and mutant  $\times$  drug behavior  $P$  value  $\leq$  mutant behavior  $P$  value with increased absolute beta values

(that is, stronger  $P$  value with appreciable difference in the magnitude of effect but not direction); (2) 'unchanged' (significant effect mutant + drug-v-WT = significant effect mutant-v-WT); (3) 'partial rescue' (significant effect mutant + drug-v-WT  $<$  significant effect mutant-v-WT) if mutant behavior  $P$  value  $\leq$  mutant  $\times$  drug behavior  $P$  value with reduced effects on the absolute beta value; (4) 'rescued' (significant effect mutant-v-WT, no significant effect mutant + Drug-v-WT) if mutant behavior  $P \leq 0.06$  and mutant  $\times$  drug behavior  $P > 0.06$ ; (5) 'over-corrected' (mutant + drug-v-WT opposite direction of significant effect mutant-v-WT) if mutant behavior  $P \leq 0.06$  and mutant  $\times$  drug behavior  $P \leq 0.06$ , with opposing directions of effect.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Key Resource Table available at <https://doi.org/10.5281/zenodo.19457279>. All hiPSCs are available from the Rutgers University Cell and DNA Repository (study 160; <http://www.nimhstemcells.org/>). Protocols are available from <https://www.protocols.io/workspaces/brennand-laboratory>. scRNA-seq data reported in this paper are publicly available from the Gene Expression Omnibus (GSE319096). Previously published SCZ CRISPRa screen datasets that were used for external validation of random forest models are available from the GEO (GSE200774) and from Synapse (syn27819129). Secondary summary statistics from the main analysis are available from Synapse (syn72039767). Source data are provided with this paper.

### Code availability

The full analysis pipeline (including code and processed data objects) used for analysis of single-cell CRISPR-KO data, evaluation and characterization of gene-level and network-level convergence, and predictive modeling using random forests are publicly available from Synapse (syn72039767). Custom MATLAB software developed by the Hoffman Lab to analyze visual-startle response parameters is available from GitHub via <https://github.com/ehoffmanlab/Weinschutz-Mendes-et-al-2023-behavior>; <https://doi.org/10.5281/zenodo.7644898> (ref. 156). Custom MATLAB software developed by J. Rihel to analyze sleep–wake assays is available from GitHub via <https://github.com/JRihel/Sleep-Analysis/tree/Sleep-Analysis-Code>; <https://doi.org/10.5281/zenodo.7644073> (ref. 159).

### References

- Ahn, K. et al. High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol. Psychiatry* **19**, 568–572 (2014).
- Ahn, K., An, S. S., Shugart, Y. Y. & Rapoport, J. L. Common polygenic variation and risk for childhood-onset schizophrenia. *Mol. Psychiatry* <https://doi.org/10.1038/mp.2014.158> (2014).
- 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 (2017).
- Guss, E. J. et al. Protocol for neurogenin-2-mediated induction of human stem cell-derived neural progenitor cells. *Star Protoc.* **5**, 102878 (2024).
- 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 (2021).
- Ho, S. M. et al. Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods* **101**, 113–124 (2016).
- Marro, S. G. et al. Neuroligin-4 regulates excitatory synaptic transmission in human neurons. *Neuron* **103**, 617–626 (2019).

107. Zhang, Z. et al. The fragile X mutation impairs homeostatic plasticity in human neurons by blocking synaptic retinoic acid signaling. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aar4338> (2018).
108. Meijer, M. et al. A single-cell model for synaptic transmission and plasticity in human iPSC-derived neurons. *Cell Rep.* **27**, 2199–2211 (2019).
109. Zhang, S. et al. Allele-specific open chromatin in human iPSC neurons elucidates functional disease variants. *Science* **369**, 561–565 (2020).
110. Sun, Y. et al. A deleterious Nav1.1 mutation selectively impairs telencephalic inhibitory neurons derived from Dravet syndrome patients. *eLife* <https://doi.org/10.7554/eLife.13073> (2016).
111. Barretto, N. et al. ASCL1- and DLX2-induced GABAergic neurons from hiPSC-derived NPCs. *J. Neurosci. Methods* **334**, 108548 (2020).
112. Powell, S. K. et al. Induction of dopaminergic neurons for neuronal subtype-specific modeling of psychiatric disease risk. *Mol. Psychiatry* **28**, 1970–1982 (2023).
113. Miller, J. A. et al. Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014).
114. Shi, Z. D. et al. Genome editing in hPSCs reveals GATA6 haploinsufficiency and a genetic interaction with GATA4 in human pancreatic development. *Cell Stem Cell* **20**, 675–688 (2017).
115. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
116. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* <https://doi.org/10.1093/gigascience/giab008> (2021).
117. Lindsay, H. et al. CrispRVariants charts the mutation spectrum of genome engineering experiments. *Nat. Biotechnol.* **34**, 701–702 (2016).
118. Yuan, S. H. et al. Cell-surface marker signatures for the isolation of neural stem cells, glia and neurons derived from human pluripotent stem cells. *PLoS ONE* **6**, e17540 (2011).
119. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
120. Chaudhry, A., Shi, R. & Luciani, D. S. A pipeline for multidimensional confocal analysis of mitochondrial morphology, function, and dynamics in pancreatic  $\beta$ -cells. *Am. J. Physiol. Endocrinol. Metab.* **318**, E87–E101 (2020).
121. Birey, F. et al. Assembly of functionally integrated human forebrain spheroids. *Nature* **545**, 54–59 (2017).
122. 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 (2018).
123. Papalexi, E. et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nat. Genet.* **53**, 322–331 (2021).
124. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
125. Tian, R. et al. CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* <https://doi.org/10.1016/j.neuron.2019.07.014> (2019).
126. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
127. Saha, A. et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. *Genome Res.* **27**, 1843–1858 (2017).
128. 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 (2016).
129. Yu, G. Gene ontology semantic similarity analysis using GOSemSim. *Methods Mol. Biol.* **2117**, 207–215 (2020).
130. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
131. Duncan, L. et al. Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *Am. J. Psychiatry* **174**, 850–858 (2017).
132. Walters, R. K. et al. Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
133. Mullins, N. et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* **53**, 817–829 (2021).
134. Johnson, E. C. et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032–1045 (2020).
135. Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
136. International Obsessive Compulsive Disorder Foundation Genetics Collaborative (IOCDF-GC) and OCD Collaborative Genetics Association Studies (OC GAS). Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
137. Nievergelt, C. M. et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat. Commun.* **10**, 4558 (2019).
138. Trubetskoy, V. et al. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022).
139. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell* **179**, 1469–1482 (2019).
140. Marioni, R. E. et al. GWAS on family history of Alzheimer’s disease. *Transl. Psychiatry* **8**, 99 (2018).
141. Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson’s disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
142. Nicolas, A. et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron* **97**, 1268–1283 (2018).
143. Yu, D. et al. Interrogating the genetic determinants of Tourette’s syndrome and other tic disorders through genome-wide association studies. *Am. J. Psychiatry* **176**, 217–227 (2019).
144. Hautakangas, H. et al. Genome-wide analysis of 102,084 migraine cases identifies 123 risk loci and subtype-specific risk alleles. *Nat. Genet.* **54**, 152–160 (2022).
145. Johnston, K. J. A. et al. Genome-wide association study of multisite chronic pain in UK Biobank. *PLoS Genet.* **15**, e1008164 (2019).
146. Lo, M. T. et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* **49**, 152–156 (2017).
147. 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 (2012).
148. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
149. Wang, J. & Liao, Y. WebGestaltR: gene set analysis toolkit WebGestaltR. R package version 0.4.3 <https://CRAN.R-project.org/package=WebGestaltR> (2020).
150. Darmanis, S. et al. A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl Acad. Sci. USA* **112**, 7285–7290 (2015).

151. Lake, B. B. et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
152. Schilder, B. M. & Skene, N. G. orthogene: an R package for easy mapping of orthologous genes across hundreds of species. *Bioconductor* <https://doi.org/10.18129/B9.bioc.orthogene> (2022).
153. Capps, M. E. S. et al. Disrupted diencephalon development and neuropeptidergic pathways in zebrafish with autism-risk mutations. *Proc. Natl Acad. Sci. USA* **122**, e2402557122 (2025).
154. Calhoun, C. C. S. et al. Removal of developmentally regulated microexons has a minimal impact on larval zebrafish brain morphology and function. *eLife* **13**, RP101790 (2025).
155. Kroll, F. et al. A simple and effective FO knockout method for rapid screening of behaviour and other complex phenotypes. *eLife* <https://doi.org/10.7554/eLife.59683> (2021).
156. Hoffman, E. ehoffmanlab/Weinschutz-Mendes-et-al-2023-behavior: Visual Startle. *Zenodo* <https://doi.org/10.5281/zenodo.7644897> (2023).
157. Rihel, J. et al. Zebrafish behavioral profiling links drugs to biological targets and rest/wake regulation. *Science* **327**, 348–351 (2010).
158. Prober, D. A., Rihel, J., Onah, A. A., Sung, R. J. & Schier, A. F. Hypocretin/orexin overexpression induces an insomnia-like phenotype in zebrafish. *J. Neurosci.* **26**, 13400–13410 (2006).
159. Rihel, J. JRihel/Sleep-Analysis: Sleep\_Code\_v1. *Zenodo* <https://doi.org/10.5281/zenodo.7644072> (2023).
160. Hoffman, E. J. et al. Estrogens suppress a behavioral phenotype in zebrafish mutants of the autism risk gene, CNTNAP2. *Neuron* **89**, 725–733 (2016).

## Acknowledgements

This work was supported by grant no. F31MH130122 (K.R.-T.); the HHMI Gilliam Fellowship (A.P.); the Autism Science Foundation (A.P.); grant nos. T32MH014276 (M.F.G.), T32GM136651 (E.D., S.E.F.), R01MH123155 (K.B.), RM1MH132648 (K.B., E.J.H.), R01MH121074 (K.B.), R01MH116002 (E.J.H.), R21MH133245 (E.J.H.), R01ES033630 (L.M.H., K.B.), R01MH124839 (L.M.H.) and R01MH118278 (L.M.H.); the Simons Foundation (grant nos. 1012863KB, 573508EH and 345993EH); the Spector Fund (E.J.H.); the Swebilius Foundation (E.J.H.); the Kavli Foundation (E.J.H.); BD2: Breakthrough Discoveries for Thriving with Bipolar Disorder (grant no. DG230102, H.S., M.D., T.-C.H., K.B.); the European Union's Horizon 2020 research and innovation programme

under the Marie Skłodowska-Curie grant (no. 101065629, N.B.); and the Interdepartmental Neuroscience Program at Yale T32NS041228 (A.P.). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Special thanks to M. Talkowski and D. Ruderfer for countless discussions on convergence and to S. Thyme and the Thyme lab for sharing zebrafish mutant lines. For the purpose of open access, the authors have applied a CC BY public copyright license to all Author Accepted Manuscripts arising from this submission.

## Author contributions

M.F.G. designed and executed the NDD gene ECCITE-seq in iNPCs, iGLUTs and iGABAs, with technical assistance from S.C., O.L. and J.C. and support from P.J.M.D. K.R.-T. conducted all bioinformatic analyses and generated all figures, with technical assistance from A.S. and H.G. Arrayed KO lines and neuronal cultures were generated, validated, phenotyped and analyzed by N.B. N.B., T.-C.H., A.K., M.D., H.G., C.B., A.L.T.S. and J.L. conducted mitochondrial phenotypic analyses, in collaboration with H.S.; N.B. and S.G. conducted neurogenesis analyses. A.P., E.D., Y.D., S.E.F. and S.K. generated, phenotyped and analyzed zebrafish mutants with technical assistance from G.D. and bioinformatic supervision from Z.W. J.W., Z.C., R.M., S.L., P.E. and M.G. conducted LNCTP analyses. Funding and mentorship was provided by L.M.H., E.J.H. and K.B. The manuscript was written by K.B. with extensive feedback from L.M.H., E.J.H. and K.R.-T. as well as contributions from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41593-026-02247-7>.

**Correspondence and requests for materials** should be addressed to Ellen J. Hoffman or Kristen Brennand.

**Peer review information** *Nature Neuroscience* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

#### Data collection

single cell cDNA and gRNA sequencing was collected from cells were loaded into 10X in 4 lanes per cell type, targeting 20,000 cells per lane for a total of ~80,000 targeted cells per cell type. scRNA-seq was performed at Yale Genomics Core with the 10X single cell 5' v2 HT with CRISPR barcode kit

Automated, high-throughput, quantitative behavioral profiling of larval zebrafish measured arousal and sensorimotor processing as a readout of circuit-level deficits resulting from gene perturbation and/or drug exposure. Individual larvae were added to each well of a 96 well plate containing 650 $\mu$ l of standard embryo water and, for drug exposure experiments, a 5mM stock solution of the drug of interest was pipetted directly into each well at a final concentration of 10  $\mu$ M (0.1% DMSO final concentration). We quantified 24 visual-startle and sleep-wake parameters. Visual-startle OFF and Visual-startle ON assays measured 6 parameters each: (i) average intensity of all startle responses; (ii) average post-stimulus activity; (iii) average activity after first stimulus; (iv) stimulus versus post-stimulus activity; (v) intensity of responses to the first stimulus; (vi) intensity of responses to the final stimulus. Sleep-wake assays measured 6 parameters in daytime and nighttime: (i) total activity; (ii) total sleep; (iii) waking activity; (iv) rest bouts; (v) sleep length; (vi) sleep latency.

#### Data analysis

Detailed data analysis is outlined in the methods; full code for the analysis workflow will be uploaded to Synapse (syn72039767) and made publicly available prior to publication.

All RNA-sequencing FASTQ files underwent matching analytical procedures. In brief, 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 to perform downstream analyses, including QC, normalization, cell clustering, and gRNA and HTO/GDO demultiplexing. Data was transformed into pseudobulk samples (by

lane) land transformed into og2CPM using VOOM (Limma), tested for differential expression using EdgeR and Limma, and co-expression networks resolved by Bayesian bi-clustering using BicMix.

The full analysis pipeline (including code and processed data objects) used for analysis of single-cell CRISPR-KO data, evaluation and characterization of gene-level and network level convergence, and predictive modeling using random forest are publicly available through Synppase (syn72039767).

Custom MATLAB software developed by the Hoffman Lab to analyze visual-startle response parameters is available on github at <https://github.com/ehoffmanlab/Weinschutz-Mendes-et-al-2023-behavior>; <https://doi.org/10.5281/zenodo.7644898>. Custom MATLAB software developed by Jason Rihel to analyze sleep-wake assays is available on github at <https://github.com/JRihel/Sleep-Analysis/tree/Sleep-Analysis-Code>; <https://doi.org/10.5281/zenodo.7644073>.

kallisto/0.46.1  
 bustools/0.40.0  
 kb\_python/0.26.4  
 python/ 2.x and 3.x  
 cellRanger/v7.2  
 Metal/version release 2018-08-28  
 MAGMA/v1.10  
 FUMA/v1.5.2  
 WebGestalt/version release 2024  
 cMAP Query Tool and LINCx 2020 updated 11/23/2021

R version 4.1.0-4.2.0

R packages:

Seurat v5.1.0 was used for all analysis of single-cell CRISPR screens prior to pseudo-bulking.

[1] MASS_7.3-60	randomForest_4.7-1.1
[3] GGally_2.1.2	ggExtra_0.10.0
[5] ggstatsplot_0.9.4	WebGestaltR_0.4.6
[7] factoextra_1.0.7	lsr_0.5.2
[9] GOSemSim_2.22.0	DOSE_3.22.0
[11] enrichplot_1.16.1	GeneNet_1.2.16
[13] fdrtool_1.2.17	longitudinal_1.1.13
[15] corpcor_1.6.10	clusterProfiler_4.4.4
[17] org.Hs.eg.db_3.15.0	AnnotationDbi_1.58.0
[19] ggpubr_0.4.0	venn_1.11
[21] ggrepel_0.9.1	scuttle_1.6.2
[23] psych_2.2.5	statmod_1.4.36
[25] Glimma_2.6.0	pamr_1.56.1
[27] survival_3.5-7	cluster_2.1.4
[29] sva_3.44.0	genefilter_1.78.0
[31] mgcv_1.8-40	nlme_3.1-157
[33] RColorBrewer_1.1-3	corrplot_0.92
[35] doParallel_1.0.17	iterators_1.0.14
[37] foreach_1.5.2	variancePartition_1.26.0
[39] BiocParallel_1.32.6	forcats_1.0.0
[41] stringr_1.5.1	purrr_1.0.1
[43] readr_2.1.2	tidyr_1.2.0
[45] tibble_3.2.1	tidyverse_1.3.1
[47] cowplot_1.1.1	presto_1.0.0
[49] data.table_1.14.2	Rcpp_1.0.10
[51] edgeR_3.38.1	limma_3.54.2
[53] ggpointdensity_0.1.0	DropletUtils_1.18.1
[55] SingleCellExperiment_1.18.1	SummarizedExperiment_1.28.0
[57] Biobase_2.56.0	GenomicRanges_1.48.0
[59] GenomeInfoDb_1.34.9	IRanges_2.32.0
[61] S4Vectors_0.36.2	BiocGenerics_0.44.0
[63] MatrixGenerics_1.10.0	matrixStats_0.62.0
[65] glmGamPoi_1.10.2	Matrix_1.6-4
[67] BUSpaRse_1.16.0	reshape2_1.4.4
[69] scales_1.3.0	sctransform_0.4.1
[71] dplyr_1.1.2	viridis_0.6.2
[73] viridisLite_0.4.2	patchwork_1.2.0.9000
[75] ggplot2_3.5.0.9000	devtools_2.4.3
[77] usethis_2.1.6	Seurat_5.0.1
[79] SeuratObject_5.0.1	sp_1.5-0
[81] BicMix	

For zebrafish behavioral assays: Custom MATLAB code is available at <https://github.com/ehoffmanlab/Weinschutz-Mendes-et-al-2023-behavior> and <https://github.com/JRihel/Sleep-Analysis/tree/Sleep-Analysis-Code>; DOI: 10.5281/zenodo.7644073.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All source donor hiPSCs have been deposited at the Rutgers University Cell and DNA Repository (study 160; <http://www.nimhstemcells.org/>). sc-RNA sequencing data reported in this paper is available on Gene Expression Omnibus (GSE319096). Previously published SCZ-CRISPRa screen datasets that were used for external validation of random forest models are available on the GEO (GSE200774) and on Synapse (syn27819129). Raw data for the main figures, as well as secondary summary statistics from main analysis are available through Synapse (syn72039767)

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Two human-induced pluripotent stem cell-lines were used in this study - one XX and one XY karyotypic sex. However, this study is not powered to make any claims related to sex differences.

Reporting on race, ethnicity, or other socially relevant groupings

NA

Population characteristics

hiPSCs (human-induced pluripotent stem cells) were derived from an XX and XY donor, representing neurotypical controls with no history of psychiatric disorders and who are of European ancestry.

Recruitment

NA. Recruitment of donors was not a part of this study. All lines were selected from a previously reported case/control hiPSC cohort of childhood onset SZ (COS) (Hoffman et al, Nat Comm 2017)

Ethics oversight

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.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

For each KO target 3-4 gRNAs were used. The number of cells for each scCRISPR screen in each celltype was 20,000 cells per lane for a total of ~80,000 targeted cells. Following demultiplexing and QC a minimum of 75 cells was necessary for analysis with an average of 474 cells per sgRNA - single-cell data was pseudobulked across Lanes to establish 4 samples per gene KO used in pseudobulk differential gene expression analysis.

For zebrafish baseline behavioral screening, sample sizes are as follows: kmt5b WT n=62, HOM n=71; mbd5 scrambled n=39, crisprant n=56; kdm5b 4i:del4 WT n=42, HOM n=55; kdm5b del17:del14 n=68, HOM n=69; phf12ab scrambled n=32, crisprant n=15; skiab scrambled n=45, crisprant n=33; chd2 scrambled n=43, crisprant n=40; smarcc2 scrambled n=36, crisprant n=35; kdm6bab scrambled n=40, crisprant n=34; kmt2cab WT n=38, HOM n=20; wacab scrambled n=32, crisprant n=32; arid1b WT n=81, HOM n=81; phf21a scrambled n=52, crisprant n=44; chd8 del7 WT n=180, HOM n=164; chd8 del5 WT n=143, HOM n=167; ash1l WT n=19, HOM n=23; nrnx1a WT n=24 HOM n=28.

For zebrafish gene-x-drug experiments, sample sizes are as follows: phf21a + amiodarone scrambled DMSO n=24, F0 DMSO n=22, F0 AMIO n=23; phf21a + fluvoxamine scrambled DMSO n=24, F0 DMSO n=22, F0 FLUVO n=21; chd2 + pravastatin WT DMSO n=27, HOM DMSO n=21, HOM PRAVA n=23; kdm6b + paclitaxel scrambled DMSO n=24, F0 DMSO n=24, F0 PACLI n=24; kdm6b + sirolimus scrambled DMSO n=24, F0 DMSO n=24, F0 SIRO n=23; kmt5b + paclitaxel WT DMSO n=20, HOM DMSO n=22, HOM PACLI n=26; kmt5b + sirolimus WT DMSO n=27, HOM DMSO n=26, HOM SIRO n=20; ash1l + sunitinib WT DMSO n=14, HOM DMSO n=13, HOM SUN n=12; ash1l + ezetimibe WT DMSO n=15, HOM DMSO n=12, HOM EZE n=9; ash1l + rosuvastatin WT DMSO n=27, HOM DMSO n=22, HOM ROSU n=22; ash1l + repaglinide WT DMSO n=11,

HOM DMSO n=16; HOM REPAG n=14.

Data exclusions	<p>Cells were excluded based on poor cell quality based on the overall percent of mitochondrial, hemoglobin, and ribosomal genes expressed in each cell and lack of sufficient read depth. After demultiplexing for gRNA identity, cells with more than one (singlet) gRNA identity assigned were excluded from downstream analysis. False positive (cells with gRNA identity but no impact on the transcriptome) identified based on gRNA an transcriptomic clustering using weighted nearest neighbor analysis, cells that did not cluster with other cells of the same gRNA identity were removed and clusters were evaluated from global KO of the targeted gene.</p> <p>For zebrafish larvae, heterozygous mutants were not included in the analysis.</p>
Replication	<p>For each CRISPR-KO screen in each cell-type, 12 technical replicates were used and pooled together. Then submitted as 20,000 cells per lane across 4 lanes which represented the technical replicates in the downstream analyses.</p> <p>For heterozygote zebrafish in-crosses, behavioral assays were repeated 2-4 times.</p>
Randomization	<p>For both experiments, technical replicate wells of iNPCs, iGLUTs, and iGABAs were pooled into a single library, randomly separated into 4 batches sequenced across 4 lanes (10x Genomics scRNA sequencing).</p> <p>For zebrafish heterozygous in-crosses, larvae were exposed to experimental conditions and genotyped after the conclusion of the experiment.</p>
Blinding	<p>Following stem cell reprogramming, the clinical origin of each sample was blinded to those performing neuronal differentiations, RNA purification, and RNA sequencing in the validation of the cell lines. For the experiment, gRNA identity and proportion is inherently unknown - technical replicates were pooled and randomly divided across 4 lanes.</p> <p>Experimenters were blind to genotype for behavioral heterozygous in-crosses until after the experiment concluded.</p>

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used	<p>Antibody Species Vendor Catalog # Dilution</p> <p>anti-MAP2 chicken Invitrogen, Abcam PA1-10005, ab5392 1:1000</p> <p>anti-Nestin rabbit Millipore ABD69 1:200</p> <p>anti-vGLUT1 rabbit Synaptic systems 135-303 1:200</p> <p>anti-GABA rabbit Sigma-Aldrich A2052 1:200</p> <p>TOMM20 mouse Santa Cruz Biotechnology sc-17764 1:200</p> <p>Total OXPHOS n/a Abcam AB-317270 1:500</p> <p>anti-mouse donkey Jackson ImmunoResearch 715-605-151 1:500</p> <p>anti-rabbit donkey Jackson ImmunoResearch 711-545-152 1:500</p> <p>anti-chicken donkey Jackson ImmunoResearch 715-605-150, 703-545-155 1:500</p>
Validation	<p>All antibodies were previously validated (in Brennand et al 2011, Brennand et al 2015, and Ho et al 2017) for immunocytochemistry in human cells.</p> <p>Validation and references for MAP2 antibody provided at <a href="https://www.abcam.com/map2-antibody-ab5392.html">https://www.abcam.com/map2-antibody-ab5392.html</a></p> <p>Validation and references for Synapsin1 antibody provided at <a href="https://sysy.com/product/106011">https://sysy.com/product/106011</a></p>

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	<p>Cultured fibroblast-reprogrammed hiPSCs</p> <p>Validated control hiPSCs for CRISPRa/RNAi were selected from a previously reported case/control hiPSC cohort of childhood onset SZ (COS) (Hoffman et al, Nat Comm 2017). The following controls were used for CRISPRa/RNAi (hiPSC NPCs NSB553(XY), NSB3182(XX)).</p>
---------------------	--

Authentication	Commercial cell lines: HEK293T cells for virus generation: Verma Lab ( <a href="https://jvi.asm.org/content/73/1/576">https://jvi.asm.org/content/73/1/576</a> )
Authentication	Samples were confirmed to key karyotypically normal using the Illumina Core Exome Genotyping Chip (Illumina, 20030770) and cnvPartition 3.2.0 (Illumina, Genome Studio). No cell lines displayed karyotypic abnormalities (no reported CNVs >2.5 MB in size). All reported CNVs are shown in each certificate of analysis. hiPSC lines were confirmed to be viable post-thaw, achieving a minimum of 50% confluency within 10 days). Sample identity testing was performed using the SNPTrace assay, confirming correct sample association between parental fibroblast and hiPSC line. Gene expression analysis was using a custom Nanostring panel <sup>27</sup> to confirm expression of pluripotency markers such as POU5F1, NANOG, and SOX2, and lack of expression of early differentiation markers such as AFP (Mesoderm), SOX17 (Endoderm), and NR2F2 (Ectoderm). A scorecard panel was used to confirm propensity to differentiate <sup>27</sup> . All hiPSC lines used in this study passed the above QC and have a certificate of analysis.
Mycoplasma contamination	As part of the hiPSC validation process, all samples were tested for the absence of Mycoplasma (Lonza, LT07-710) and confirmed to be sterile (Hardy Diagnostics, K82). NPCs were tested for mycoplasma monthly with all test being negative.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	N/A

## Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	Zebrafish larvae were raised at 28°C on a 14:10 hour light:dark cycle. Larvae were grown in 150 mm Petri dishes in blue water (0.3g/L Instant Ocean, 1 mg/L methylene blue, pH 7.0) at a density of 60-80 larvae per dish. Behavioral assays were conducted in zebrafish larvae at 5-7 dpf.
Wild animals	NA
Reporting on sex	At these developmental stages, sex is not yet determined.
Field-collected samples	NA
Ethics oversight	All procedures involving zebrafish were conducted in accordance with Institutional Animal Care and Use Committee (IACUC; Protocol #2024-20054) regulatory standards at Yale University.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Plants

Seed stocks	NA
Novel plant genotypes	Na
Authentication	NA