Transcriptomic Convergence and the Female Protective Effect in Autism

Rebecca E. Andersen[1,2,3,4,5,*], Maya Talukdar[1,2,3,4,6,7,*], Tyler Sakamoto[1,2,3,8], David Exposito-Alonso[1,2,3,5], Janet H.T. Song[1,2,3,5], Xuyu Qian[1,2,3,5], Seungil Lee[1,2,3,8], Nila Murugan[1,2,3,9], Ryan N. Delgado[10], Sijing Zhao[1,2,3,11], Gwenyth Eichfeld[1,2,3,12], Julia Harms[1,2,3,13], David C. Page[7,14,15], and Christopher A. Walsh[1,2,3,4,5]

Affiliations:
[1]Division of Genetics and Genomics and Manton Center for Orphan Diseases, Boston Children's Hospital, Boston, MA, USA; [2]Departments of Pediatrics and Neurology, Harvard Medical School, Boston, MA, USA; [3]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA, USA; [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA; [5]Allen Discovery Center for Human Brain Evolution, Boston, MA, USA; [6]Harvard-MIT MD/PhD Program, Program in Biomedical Informatics, Boston, MA, USA; [7]Whitehead Institute, Cambridge, MA, USA; [8]Harvard College, Cambridge, MA, USA; [9]University of Massachusetts Boston, Boston, MA, USA; [10]Department of Genetics, Blavatnik Institute, and Howard Hughes Medical Institute, Harvard Medical School, Boston, MA, USA; [11]Harvard BBS PhD Program, Boston, MA, USA; [12]Colgate University, Hamilton, NY, USA; [13]University of California Berkeley, Berkeley, CA, USA; [14]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA; [15]Howard Hughes Medical Institute, Whitehead Institute, Cambridge, MA, USA.

* These authors contributed equally to this work.

Corresponding author and lead contact:

Christopher A. Walsh, christopher.walsh@childrens.harvard.edu

ABSTRACT

Autism spectrum disorder (ASD) is a common neurodevelopmental condition characterized by deficits in social communication as well as restricted and/or repetitive behaviors. ASD is highly heritable[1], with a complex genetic architecture: large-scale studies have identified dosage-altering copy number variants (CNV) and single nucleotide variants (SNV) that implicate hundreds of genes as individually rare causes of ASD (ASD genes)[2–4], with common variation at multiple loci also contributing substantially to risk[5]. Understanding how disruptions to these functionally diverse genes lead to the shared core features of ASD remains a major challenge[6]. Moreover, ASD is three- to four-fold more common in males than females[7,8], and autistic females tend to carry more autosomal risk alleles for ASD compared to autistic males[9–13], but the biological basis of this "female protective effect" (FPE) is unknown[14,15]. Here we show that individual perturbations of 18 ASD genes converge on shared effects on gene expression, including widespread downregulation of other ASD genes. *De novo* reconstruction of a gene regulatory network (GRN) enabled the identification of central transcriptional regulators, including the prominent ASD gene *CHD8* as well as novel candidates such as *REST*, that drive this transcriptomic convergence in ASD. Furthermore, the X-linked transcription factor *ZFX*, which is expressed from both the active and the "inactive" X chromosomes in females[16], emerged as a key activator of many ASD genes: we propose that the higher *ZFX* expression level observed in female brain can buffer damaging mutations in diverse ASD genes, contributing to the FPE. Together, these results reveal how key GRNs can become broadly and similarly dysregulated upon disruption of individual ASD genes and provide molecular insight into the female protective effect in ASD.

MAIN

Large-scale genetic studies have implicated hundreds of genes in autism spectrum disorder (ASD), with many identified through rare mutations that result in loss of function of one allele[2,17]. This suggests that gene dosage plays an important role in ASD[18], making it critical to determine how appropriate expression levels of ASD genes are maintained. Characterization of gene regulatory networks (GRNs) can provide important insights into mechanisms mediating the proper transcriptional dosage of these ASD genes. Additionally, ASD is significantly male-biased in prevalence, and although some ASD genes have been identified on the X chromosome[19], X-linked genes alone cannot fully explain the male-biased prevalence of ASD (M.T. and D.C.P., under review)[14,20,21]. Moreover, it has been consistently demonstrated that females require greater autosomal genetic risk to manifest ASD compared to males[9–13]. The genetic mechanisms of this female protective effect (FPE) in ASD are unknown

Here, gene regulatory network (GRN) reconstruction identifies key drivers of the expression of ASD genes, including *CHD8* and *REST*. These analyses also uncover the female-biased transcription factor *ZFX*, which is expressed from both the active and "inactive" X chromosome in females, as a key regulator of ASD genes, suggesting a direct transcriptional mechanism that could contribute to the FPE. This study provides a crucial framework for uncovering how variants in diverse genes can cause convergent pathophysiological effects that ultimately result in ASD.

RESULTS

Depletion of ASD Genes Reveals Widespread Transcriptomic Convergence

Functional interrogation of ASD genes revealed strongly overlapping effects on downstream targets. Knockdown (KD) studies were performed for 20 ASD genes, including 18 genes from the Simons Foundation Autism Research Initiative (SFARI) Gene database[22,23] and two long non-coding RNAs (lncRNAs) previously implicated in ASD (*NR2F1-AS1*[24] and *SOX2-OT*[25]). The RNA-degrading CRISPR enzyme CasRx[26,27] and an array of three guide RNAs (gRNAs) per target (**Supplementary Table 1**) were used to induce individual KD of each target gene in control human male (XY) neural progenitor cell (NPC) cultures (**Fig. 1a-b** and **Extended Data Fig. 1a**). RNA-sequencing (RNA-seq)

analysis after 24 hours of KD demonstrated that nearly all perturbations led to a substantial number of significantly differentially expressed genes (DEGs; adj. $p < 0.05$), with a median of 920 DEGs per perturbation (25th percentile: 644, 75th percentile: 1485) (**Supplementary Table 2**). To ensure sufficient statistical power for downstream analyses, we focused on the 18 perturbations (18/20, 90%) that resulted in greater than 20 DEGs.

Transcriptome-wide analyses revealed that ASD gene KDs caused remarkably widespread disruption of other ASD genes. Most sets of DEGs exhibited significant enrichment for known ASD genes in the SFARI gene database (15/18, 83.3%) (**Fig. 1c** and **Supplementary Table 3**). In contrast, there was no significant enrichment for genes implicated in intellectual disability (ID)[28] but not ASD (only ID genes), despite their comparable expression in NPCs (only ID genes: 1062/1242, 85.5%

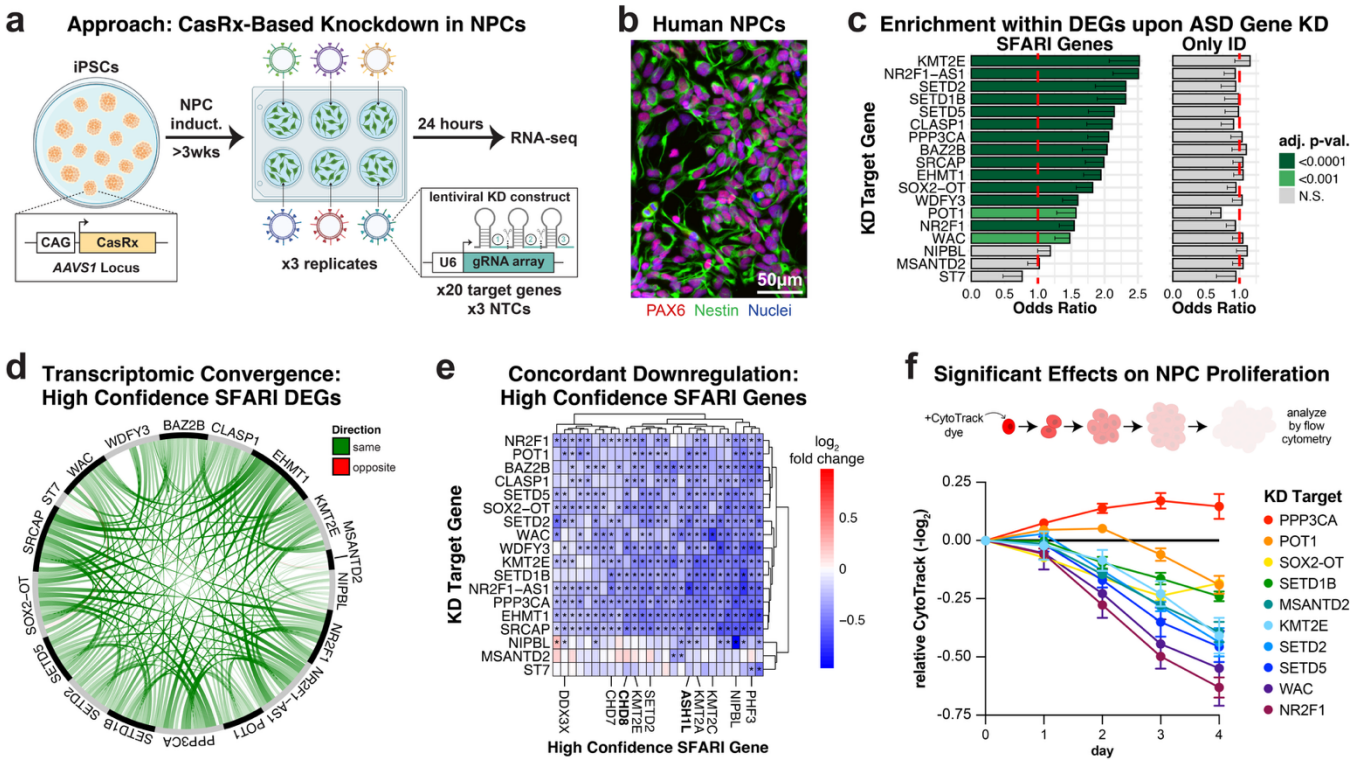


Fig. 1 | Depletion of ASD genes reveals widespread transcriptomic convergence.
a, Experimental design schematic. **b,** Immunocytochemistry of iPSC-derived NPC cultures. **c,** Bar plot depicting enrichment of SFARI genes or only intellectual disability (ID) genes within the DEGs of each KD condition, excluding the targeted gene itself. Error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown). **d,** Circos plot depicting high confidence SFARI DEGs shared between different KD conditions. **e,** Heatmap showing the relative expression of high confidence SFARI genes that are DEGs in at least 10 KD conditions (see also **Extended Data Fig. 1c**). Statistically significant DEGs are indicated with *. Top ASD risk genes from [2] are bolded. **f,** Relative CytoTrack dye intensity ($-\log_2$) over time. Only KD conditions that exhibited statistically significant effects that were consistent over time are displayed. Plot depicts mean +/- standard error of the mean (SEM). Statistical significance was determined using ordinary two-way ANOVA with Dunnett's multiple comparisons test. See **Supplementary Table 6** for full statistical results.

83 expressed; SFARI genes: 797/1230, 64.8% expressed). This suggests that perturbation of ASD genes
84 particularly affects other ASD genes, as opposed to neurodevelopmental genes more broadly.
85 Notably, we discovered a striking pattern of transcriptomic convergence across perturbations, which
86 we define as statistically significant overlap of the same differentially expressed genes (DEGs)
87 (**Extended Data Fig. 1b**). This was particularly evident for high confidence SFARI (HC-SFARI) genes,
88 the SFARI genes in the highest evidentiary tier for association with ASD (**Fig. 1d**). Moreover, HC-SFARI
89 genes tended to shift in a concordant direction as a result of different perturbations even when they were
90 not significant DEGs within a particular KD sample (**Extended Data Fig. 1c**). This included the prominent
91 ASD risk gene *CHD8*[29], which was significantly downregulated by most (11/18, 61.1%) of the KDs.
92 Remarkably, a total of 27 HC-SFARI genes were DEGs in at least 10 of the KDs (27/183, 14.8%, p =
93 4.1e-9), and all were consistently downregulated (**Fig. 1e**), demonstrating strong convergence on many
94 of the strongest risk genes for ASD.
95 This marked transcriptomic convergence suggested that candidate ASD risk genes could be
96 nominated by the following features: 1) consistent downregulation following ASD gene KDs, and 2)
97 dosage sensitivity, as indicated by high genic constraint scores[30] (**Supplementary Table 4** and **Materials**
98 **and Methods**). This approach yielded 397 genes (**Supplementary Table 5**), including 113 SFARI
99 (113/397, 28.5%, adj. p = 2.5e-39) and 49 HC-SFARI (49/397, 12.3%, adj. p = 1.5e-29) genes,
100 demonstrating that these criteria effectively prioritize known ASD genes. Thus, the remaining 284 non-
101 SFARI genes merit consideration as candidate ASD risk genes.
102
103 **ASD Genes Regulate NPC Proliferation**
104 Many ASD gene perturbations also caused similar phenotypic effects in NPCs. Proliferation
105 analysis using heritable dye (CytoTrack) revealed that 9/18 (50%) of the KDs led to a statistically
106 significant decrease in proliferation (**Fig. 1f** and **Supplementary Table 6**), including several genes
107 implicated in microcephaly (*SETD1B*[31], *SETD2*[32], *SETD5*[33], *KMT2E*[34], and *WAC*[35]). In contrast, only
108 1/18 (5.6%) of the KDs led to a significantly increased NPC proliferation rate. This suggests that

reduced NPC proliferation is a common phenotype that results from disruption of several different ASD genes, consistent with previous findings of ASD-related effects on NPC proliferation[36].

Disruption of ASD Genes Alters Cerebral Organoid Development

Inducible KD in cerebral organoids further revealed shared neurodevelopmental phenotypes. To enable these organoid studies, we created the FLEx-based Inducible CRISPR Knockdown (FLICK) construct. FLICK enables control over the timing of KD through doxycycline-inducible genetic recombination that results in constitutive expression of CasRx and an EGFP reporter (**Fig. 2a**),

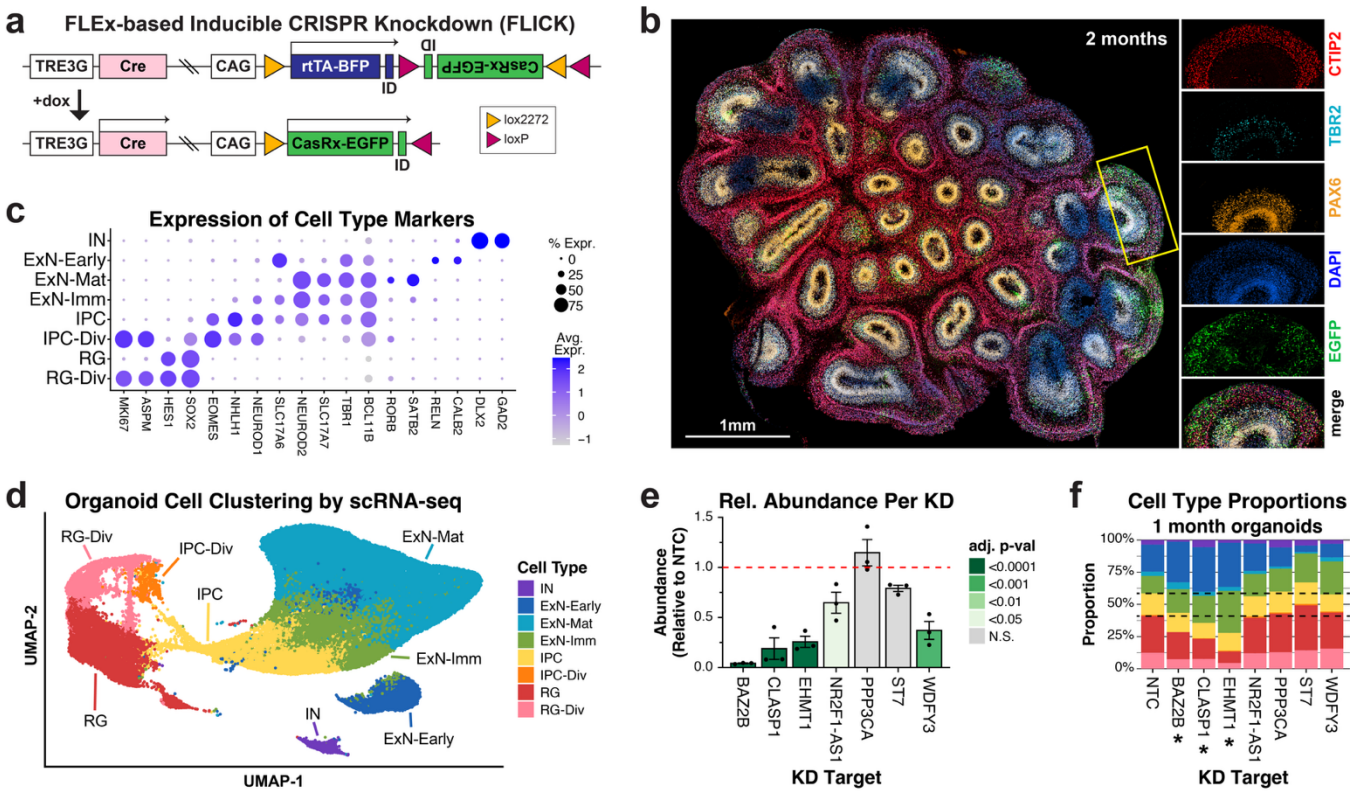


Fig. 2 | Disruption of ASD genes alters cerebral organoid development.
a, Design of FLICK constructs for doxycycline-inducible CasRx KD. **b,** FLICK organoid after 2 months of development. Immunohistochemistry was performed to visualize expression of the following markers: EGFP = FLICK construct; DAPI = nuclei, PAX6 = NPCs; TBR2 = intermediate progenitor cells; CTIP2 = excitatory neurons. **c,** Expression of canonical brain marker genes used for identifying cell types by scRNA-seq. **d,** UMAP of integrated long-term organoid scRNA-seq data collected at one month and two months. **e,** Relative abundance of cells for each KD target from mosaic organoids at one or two months. Results are normalized to the starting proportion for each sample at day 0 and displayed relative to NTCs from the same timepoint. Plot depicts mean +/- SEM. The FLICK constructs had been induced by treatment with doxycycline on d14-d15. Statistical significance was assessed using repeated measures one-way ANOVA with Dunnett's multiple comparisons test. **f,** Cell type composition per KD target in one-month organoids. Dotted lines indicate the NTC proportions of major cell types (radial glia, intermediate progenitors, and neurons). Samples for which the ExN/RG proportion was significantly different than NTC are indicated with * (adj. $p < 0.05$). Statistical analysis was performed using ordinary one-way ANOVA with Dunnett's multiple comparisons test, considering all long-term samples from one month (shown here) and two months (see **Extended Data Fig. 2e**).

117 providing robust KD (**Extended Data Fig. 2a**). The FLICK constructs also contain identifying
118 sequences that can be captured during single-cell RNA-seq (scRNA-seq) and used to determine the
119 KD target in a given cell, enabling the analysis of mosaic cerebral organoids.

120 For long-term studies, we targeted ASD genes whose KDs had not been found to impair NPC
121 proliferation in our CytoTrack experiments (**Supplementary Table 6**). FLICK constructs targeting these
122 genes were separately transposed into iPSCs, which were later pooled to generate mosaic organoids
123 (day 0). KD was induced by doxycycline treatment from day 14-15, when the organoids predominantly
124 consisted of PAX6+ NPCs (**Extended Data Fig. 2b**). These organoids were subsequently grown until
125 either one or two months of development (**Fig. 2b**) and analyzed by scRNA-seq, which identified all
126 expected cell types (**Fig. 2c-d** and **Extended Data Fig. 2c-d**) including radial glial progenitors (RGs),
127 intermediate progenitor cells (IPCs), inhibitory neurons (INs), and excitatory neurons (ExNs).

128 Long-term organoid KDs revealed shared phenotypes across multiple perturbations. Several KDs,
129 particularly those targeting *BAZ2B*, *CLASP1*, and *EHMT1*, were depleted from the organoids compared
130 to NTCs (**Fig. 2e**), even though these KDs did not impair proliferation in NPC cultures (**Supplementary**
131 **Table 6**). Interestingly, pathogenic variants in *CLASP1*[37] and *EHMT1*[38] have been implicated in
132 microcephaly, consistent with this phenotype. Furthermore, these KDs also altered cell type proportions
133 in organoids, producing relatively greater numbers of ExNs and fewer RGs (adj. p = 0.021 for *BAZ2B*,
134 0.018 for *CLASP1*, and 0.012 for *EHMT1*) (**Fig. 2f** and **Extended Data Fig. 2e**). Together, this
135 suggests that disruption of these genes may result in increased ExN differentiation at the expense of
136 RG self-renewal. Most of the other KDs also trended toward an increased proportion of ExNs by the 2-
137 month timepoint, suggesting this is a convergent neurodevelopmental phenotype, in line with previous
138 observations[39,40].

139 Depletion of ASD genes resulted in prominent transcriptomic convergence in cerebral organoids
140 (**Fig. 3a** and **Extended Data Fig. 3a**). RGs with different KDs exhibited highly overlapping DEGs,
141 further supporting our findings from NPC cultures. Indeed, the DEGs identified in organoid RGs strongly
142 overlapped with the DEGs identified from KD of the same target in NPC cultures (**Extended Data Fig.**
143 **3b**), suggesting that these effects on gene expression are not specific to a particular model system.

144 Transcriptomic convergence was also observed within additional cell types including IPCs as well as
145 ExNs (**Fig. 3a** and **Extended Data Fig. 3a**), demonstrating that the convergence of ASD genes on
146 shared downstream effects is not unique to neural progenitors.

147 Although shared ASD DEGs have been identified from postmortem brain tissue from individuals
148 with distinct causes of ASD[41–47], evidence of transcriptomic convergence in previous experimental
149 studies has been limited[39,48,49]. Indeed, based on studies in cerebral organoids, it has been proposed
150 that different ASD mutations largely impact distinct downstream genes[39]. We find that detection of
151 overlapping DEGs is heavily dependent upon the number of cells sequenced for each perturbation, as
152 demonstrated by down-sampling analysis (**Fig. 3b**). This sample-size dependence likely explains the
153 lack of observable convergence in previous studies with smaller sample sizes.

154 Finally, while some convergently disrupted genes were cell type-specific, many genes were
155 consistently disrupted across multiple cell types, including SFARI DEGs (**Fig. 3c**). This suggests that
156 despite cell-type-specific features impacting the transcriptomic consequences of ASD gene

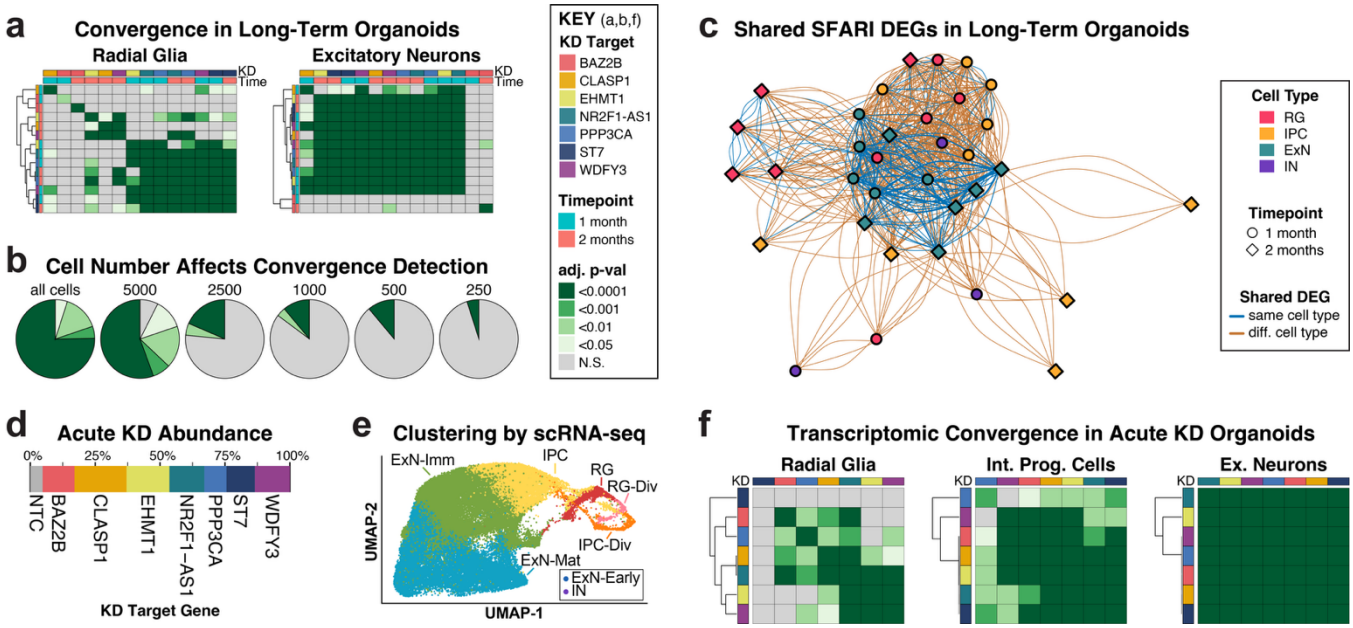


Fig. 3 | Transcriptomic convergence upon ASD gene perturbations in cerebral organoids.
a) Heatmap of cross-enrichment of DEGs for radial glia and excitatory neurons in long-term organoids. **b)** Down-sampling the number of cells per condition to the number indicated above each pie chart decreases the significant cross-enrichment (transcriptomic convergence) observed between sets of DEGs. **c)** Network visualization of DEGs shared amongst different KDs conditions and cell types. The thickness of edges represents the number of shared DEGs. Nodes plotted based on clustering with the Fruchterman-Reingold algorithm (*igraph*). Nodes that are closer together have more shared DEGs on average. **d)** Proportion of cells with each KD target from scRNA-seq of mosaic organoids with acute KD. **e)** UMAP of acute KD mosaic organoids from scRNA-seq. **f)** Heatmap of cross-enrichment between different KD samples in acute KD organoids.

157 perturbations, certain effects could be broadly shared by diverse cell types. This is consistent with
158 widespread observations of convergent ExN phenotypes such as disordered synaptic structure or
159 activity upon disruption of distinct ASD genes[50,51], and suggests that transcriptomic convergence does
160 not reflect merely a shared alteration of cell type or fate.
161

162 **Transcriptomic Convergence Following Acute Perturbations in Cerebral Organoids**

163 Acute disruption of ASD genes in 2-month organoids revealed even greater convergence on shared
164 DEGs. An additional set of mosaic FLICK organoids was grown for two months prior to inducing KD
165 with doxycycline treatment from day 61-62 followed by scRNA-seq two days later, allowing us to
166 determine the acute effects of disrupting these genes within organoids that had developed normally.
167 This approach resulted in similar numbers of cells per KD (**Fig. 3d** and **Extended Data Fig. 3c**),
168 enabling deeper analysis of the KDs that had been depleted from the organoids with long-term KD (**Fig.
169 2e**). As expected, these organoids contained a mixture of cell types (**Fig. 3e** and **Extended Data Fig.
170 3d**), with ExNs by far the most abundant cell type at this developmental timepoint (**Extended Data Fig.
171 3e**). Interestingly, many KDs resulted in ExN DEGs that were enriched for the DEGs identified from
172 NPC cultures (**Extended Data Fig. 3f**), again suggesting shared changes in gene expression across
173 cell types upon KD of ASD genes. Furthermore, acute KD of the target genes that had been depleted
174 from organoids over the course of long-term KD (**Fig. 2e**) allowed us to obtain much greater cell
175 numbers for these perturbations, revealing that these KDs also result in strong transcriptomic
176 convergence (**Fig. 3f**).
177

178 **Key Transcription Factors, Including *REST*, Target Recurrently Dysregulated Genes**

179 Transcriptional regulators that could mediate transcriptomic convergence in ASD were identified
180 through MAGIC analysis[52], which uses chromatin immunoprecipitation sequencing (ChIP-seq) data
181 from ENCODE to identify transcription factors and cofactors whose binding targets are enriched within
182 a gene set of interest. This uncovered transcriptional regulators whose target genes were recurrently
183 differentially expressed across many of the ASD gene KDs in NPCs (**Fig. 4a** and **Supplementary**

184 **Table 7**), including HC-SFARI genes such as *CHD2* and *SIN3A*. Of the recurrent MAGIC regulators *not*
185 previously implicated in ASD, several were both highly constrained (**Supplementary Table 4)** and co-
186 expressed with known ASD genes (**Fig. 4a**), features demonstrated to be predictive of novel ASD risk
187 genes[6].

188 MAGIC analysis also revealed consistent drivers of convergence in organoids (**Extended Data Fig.**
189 **4** and **Supplementary Table 7**). Of the top recurrent transcriptional regulators identified in NPCs, half
190 (13/26, 50%) were also identified in organoid RGs (**Fig. 4b**). MAGIC also identified recurrent ExN
191 regulators shared across developmental stages and across long-term vs acute KD timescales,
192 including many SFARI genes. Interestingly, nine of these genes were also identified as top
193 transcriptional regulators in neural progenitors, suggesting that they are key drivers of convergence in
194 multiple cellular contexts.

195 Of particular interest, the RE1 silencing transcription factor (REST) was significant by MAGIC in
196 17/18 (94.4%) NPC perturbations (**Fig. 4a**) and consistently emerged as a top MAGIC regulator across
197 all tested conditions (**Fig. 4b**). While REST has not previously been directly associated with ASD, it has
198 critical neurodevelopmental roles, functioning as a master negative regulator of neurogenesis[53,54]. Re-
199 analysis of human and mouse REST ChIP-seq binding sites[55] identified strong enrichment of SFARI

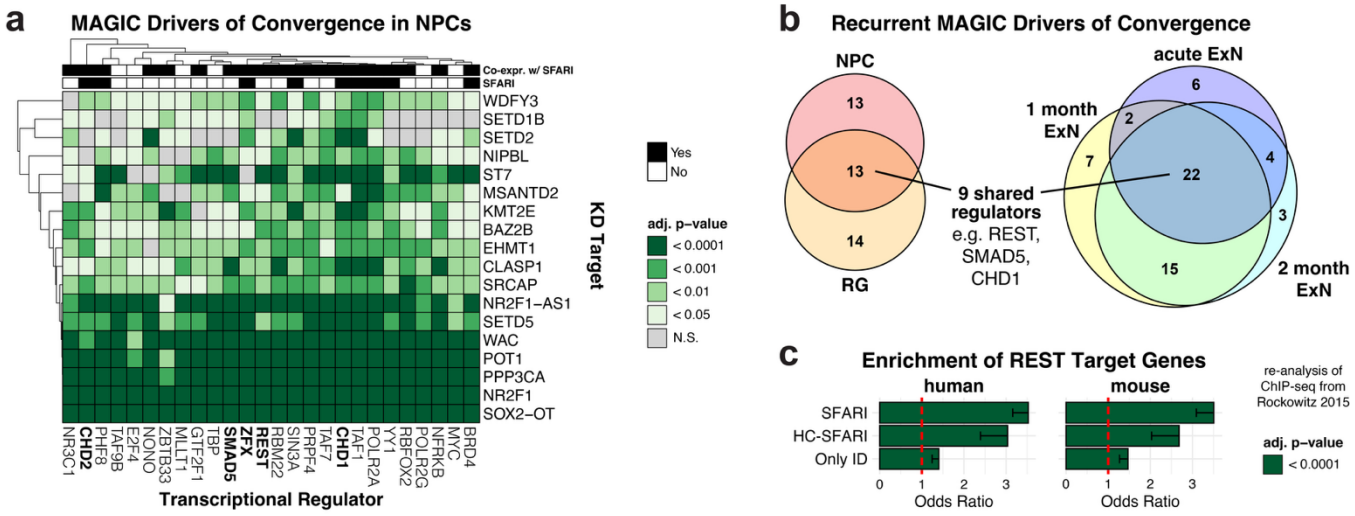


Fig. 4 | Key transcription factors, including REST, target recurrently dysregulated genes.
a, MAGIC analysis identifying transcriptional regulators whose targets were enriched within the DEGs from at least 16 NPC KDs. Transcriptional regulators of particular interest are bolded. **b,** Quantitative Euler diagrams depicting the overlapping MAGIC transcriptional regulators identified in different conditions. **c,** Bar plot depicting enrichment of SFARI or only intellectual disability (ID) genes within REST targets from human and mouse embryonic stem cells. REST target data from Rockowitz and Zheng 2015. Error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown).

genes (adj. p = 3.6e-67 and 7.7e-50, respectively), including HC-SFARI genes (adj. p = 2.2e-13 and 3.7e-8, respectively), within REST targets (**Fig. 4c**), supporting a key role for REST in regulating ASD-relevant gene expression.

Reconstruction of a Gene Regulatory Network Reveals Highly Interconnected ASD Genes

De novo gene regulatory network (GRN) reconstruction from our NPC RNA-seq data using ARACNe[56,57] enabled associating additional transcription factors[58] and chromatin modifiers[59] – together referred to as "regulators" – with their downstream targets (regulons). We validated this approach by separately constructing "held-out" GRNs that excluded data from samples in which the KD target was a potential regulator, which revealed strong cross-enrichment between the GRN-predicted targets from these held-out networks and observed DEGs (adj. p for each < 3.98e-116, combined p = 4.08e-937) (**Fig. 5a** and **Extended Data Fig. 5a**). Thus, ARACNe effectively captures *bona fide* regulatory relationships. The full NPC GRN revealed a high degree of interconnection between ASD genes and their upstream regulators (**Fig. 5b**), with ASD genes strongly enriched within central regulators of the NPC GRN (adj. p = 7.6e-6 for SFARI genes, 8.6e-5 for HC-SFARI genes, centrality assessed by *PageRank* algorithm[60]) (**Fig. 5c**). This suggests that genes associated with ASD are key drivers of NPC gene expression in general. Central regulators were also enriched for pathways[61] associated with the G2M checkpoint (adj. p = 3.3e-4) and mitotic spindle (adj. p = 1.1e-5), consistent with the NPC proliferation phenotypes we observed (**Fig. 1g** and **Supplementary Table 6**).

GRN Analysis Identifies *CHD8* as a Top Driver of Convergence in ASD

We next identified GRN regulators that act as key drivers of ASD-associated transcriptomic convergence. This analysis revealed 70 top drivers of convergence whose activity was consistently altered by ASD gene KD (*VIPER*[62]), and whose targets are both enriched for SFARI genes and central to the GRN (**Supplementary Table 8**). Nearly half of these top drivers of convergence are SFARI genes (31/70, 44.3%, adj. p = 3.9e-11), and nearly one third are HC-SFARI genes (22/70, 31.4%, adj. p = 6.7e-12). Moreover, these top drivers of convergence were consistently downregulated across nearly

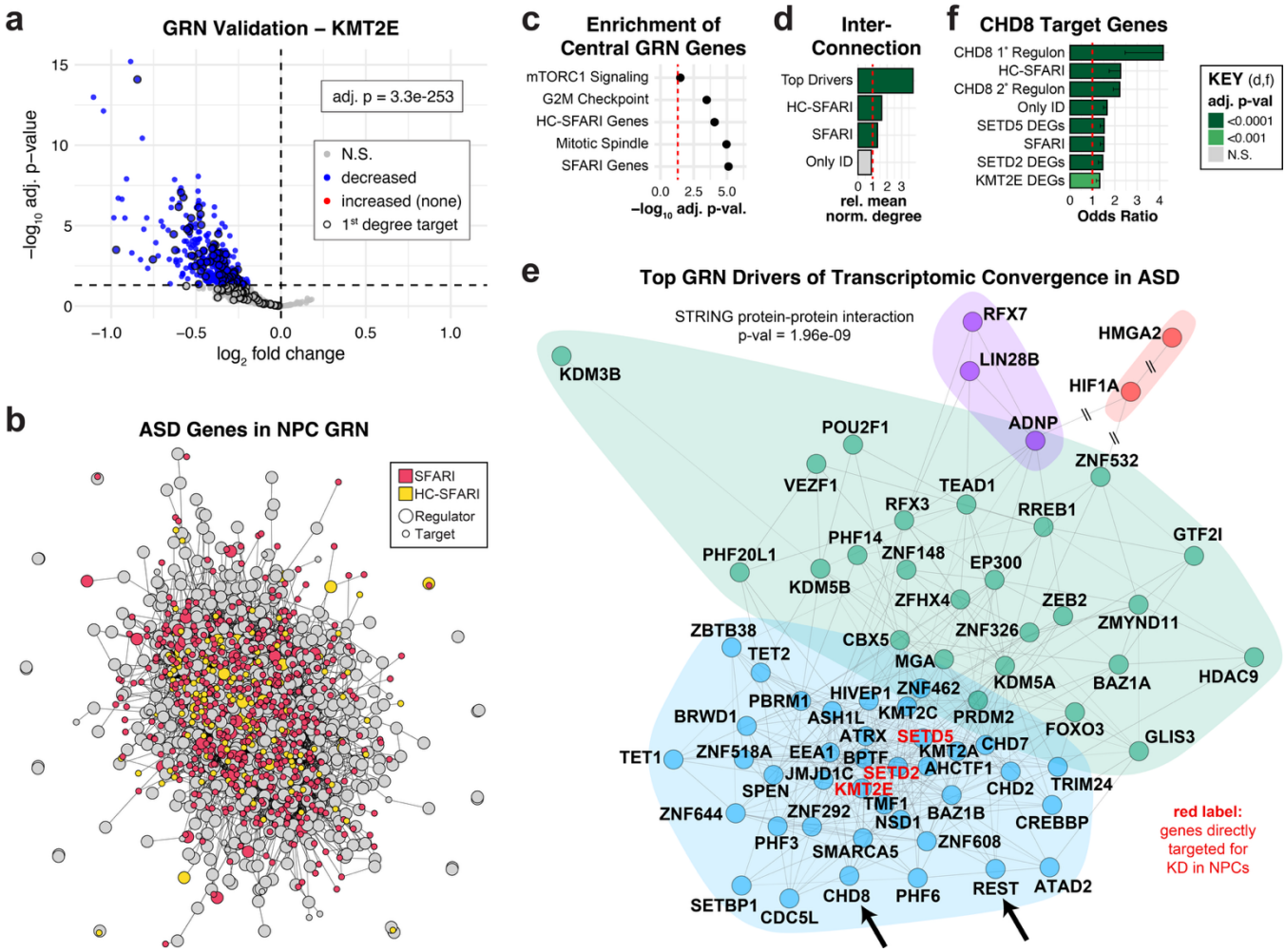


Fig. 5 | Gene regulatory network analysis identifies *CHD8* as a top driver of convergence in ASD. **a,** Volcano plot of genes predicted to be positively regulated by KMT2E, depicting their differential expression upon KD of KMT2E. Statistically significant DEGs with decreased expression are shown in blue, while DEGs with increased expression are shown in red (none). Statistically significant enrichment of 2nd degree regulon genes within the downregulated DEGs is indicated with the adjusted p-value. **b,** Depiction of a subset of the NPC GRN constructed using ARACNE. All regulators are shown along with first degree target genes that are SFARI genes. **c,** Gene set enrichment analysis (GSEA) demonstrating enrichment of genes with high PageRank scores within the gene sets shown. **d,** Bar graph depicting mean normalized degree, a measure of interconnectedness, across regulators from different gene sets. **e,** Depiction of modules of top regulators identified through iterative walktrap clustering. Hashed lines indicate that those lines are not to scale. Genes of particular interest are highlighted with large arrows. KD targets that are further analyzed in panel E are highlighted in red font. **f,** Bar plot depicting enrichment of CHD8 targets within several gene sets. Data from CHD8 ChIP-seq in NPCs from Sugathan et al. 2014. Error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown).

all ASD gene KDs (**Extended Data Fig. 5b**). They are also highly interconnected in the GRN ($p < 1e-6$) (**Fig. 5d**), as well as the STRING database of protein-protein interactions[63] ($p = 2.0e-9$). Taken together, this suggests that these genes not only act as central regulators of many ASD genes but also cross-regulate each other, driving transcriptomic convergence in ASD.

231 Iterative walktrap clustering identified four gene modules within the 70 top drivers of convergence
232 (**Fig. 5e**), one of which (blue) was especially strongly interconnected. This module exhibited very high
233 cross-interaction through STRING analysis ($p < 1e-16$) and included *REST* (**Fig. 4** and **Extended Data**
234 **Fig. 4**) and several notable ASD genes including *CHD8*, one of the strongest known risk genes for
235 ASD[29]. Importantly, the inferred *CHD8* regulon from the GRN was strongly enriched (1st degree regulon
236 adj. p = 4.6e-7, 2nd degree regulon adj. p = 3.9e-20) for direct CHD8 target genes previously identified
237 through ChIP-seq from NPCs[64] (**Fig. 5f**), further supporting the validity of ARACNe-predicted regulatory
238 relationships. Moreover, consistent with previous reports[64,65], CHD8 ChIP-seq targets were also highly
239 enriched for NPC-expressed SFARI genes (adj. p = 2.6e-8) and HC-SFARI genes (adj. p = 8.5e-8). The
240 *CHD8*-containing module also included three transcriptional regulators (*SETD2*, *SETD5*, and *KMT2E*)
241 that we had directly perturbed, and DEGs from these KDs exhibited significant enrichment
242 for CHD8 ChIP-seq targets (adj. p = 1.2e-5, 7.8e-8, and 2.6e-4, respectively) (**Fig. 5f**), further
243 demonstrating convergent regulation of shared targets by critical ASD genes. This analysis
244 highlights *CHD8*, a high confidence ASD risk gene, as a key driver of transcriptomic convergence in
245 ASD.

246 **The X-Linked Transcription Factor ZFX is a Female-Biased Regulator of ASD Genes**

247 Prior studies have found that genes downregulated in ASD tend to be more highly expressed in
248 neurotypical female brain versus male[66,67], raising the possibility that this baseline female-biased
249 expression may contribute to the FPE[66,67]. We identified 68 genes that were consistently downregulated
250 by ASD gene KD and exhibited female-biased expression in the brain[68–70], 22 of which were highly
251 constrained (**Extended Data Fig. 6a**). This included nine SFARI genes (OR = 9.10, adj. p = 2.94e-5)
252 (**Fig. 6a**), two of which – *ZFX* and *ZNF711* – were also GRN regulators of targets enriched for SFARI
253 genes (adj. p = 6.0e-4 for *ZFX*, 0.044 for *ZNF711*) (**Fig. 6b**).

254 Among these, *ZFX* is of particular interest because it is expressed from both the active and
255 "inactive" X chromosome in females[16], making it one of the most highly and consistently female-biased
256 genes across diverse tissues[69–74], including the developing cerebral cortex[68,75] (**Extended Data Fig.**
257 **6b**). *ZFX* encodes a potent transcriptional activator of genome-wide targets[76–78], and we identified many

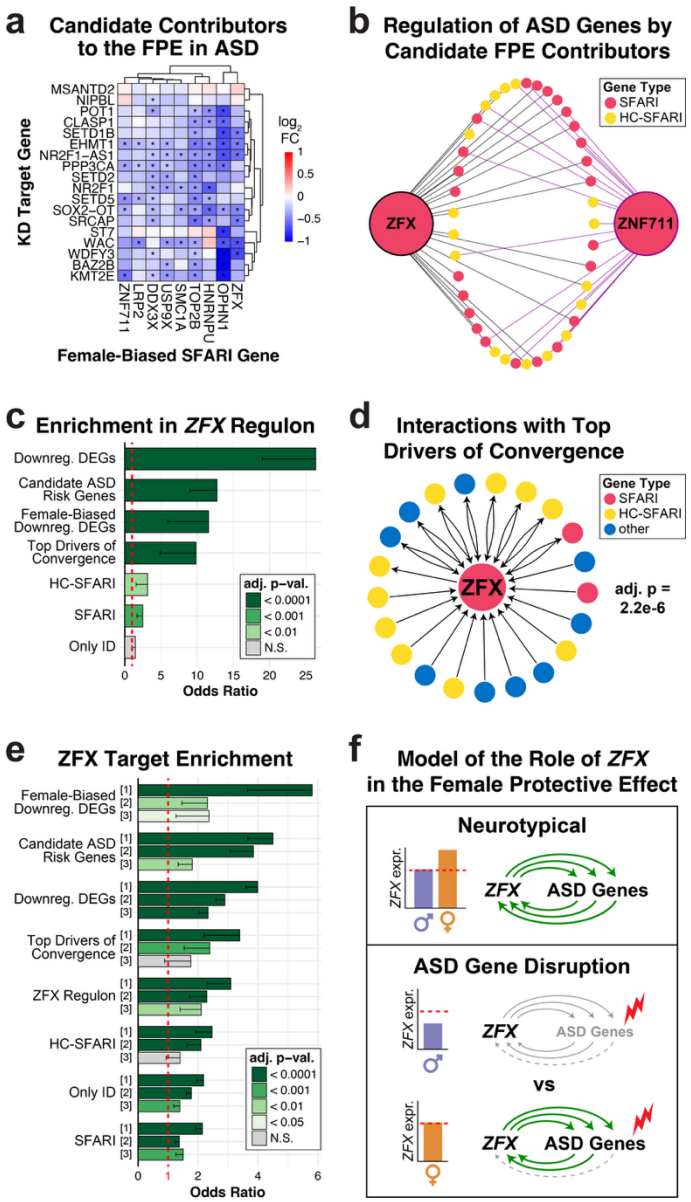


Fig. 6 | The X-linked transcription factor ZFX is a female-biased regulator of ASD genes. **a,** Heatmap showing the relative expression of SFARI genes that are candidate contributors to the FPE in ASD across KD conditions. Statistically significant DEGs are indicated with *. **b,** Depiction of regulatory relationships between the candidate FPE contributors *ZFX* and *ZNF711* and their target SFARI genes, as identified by GRN analysis. Connections between *ZFX* and its targets are shown in black, while connections between *ZNF711* and its targets are shown in purple. **c,** Enrichment of ASD-relevant gene sets within the *ZFX* regulon, as identified by GRN analysis. Error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown). **d,** Depiction of regulatory relationships between *ZFX* and the top drivers of convergence in ASD (previously identified in **Fig. 5e**), as identified by GRN analysis. **e,** Enrichment of ASD-relevant gene sets within the following *ZFX* target sets: [1] targets in C4-2B cells, identified in San Roman et al. 2024; [2] common targets identified in Rhie et al. 2018; and [3] targets in MCF-7 cells, identified in San Roman et al. 2024. Error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown). **f,** Schematic depicting a model of the contribution of *ZFX* to the FPE in ASD.

regulatory relationships between *ZFX* and ASD-relevant gene sets in our GRN (**Fig. 6c**). This included our previously identified top drivers of transcriptomic convergence in ASD (OR = 9.8, adj. p = 2.2e-6). We also found that ZFX is a downstream target of many of these top drivers, with several reciprocal regulatory relationships (**Fig. 6d**). The strong enrichment of ZFX targets for ASD-relevant genes was further confirmed by re-analysis of ZFX ChIP-seq data[76,79] (**Fig. 6e**). As *ZFX* is recurrently downregulated upon disruption of ASD genes, its female-biased expression in combination with its role as a regulator of ASD-relevant genes could thus result in cascading and broad sex-differential impacts on ASD gene expression, potentially contributing to the FPE in ASD (**Fig. 6f**).

266 **DISCUSSION**

267 Here, we show widespread transcriptomic convergence following depletion of ASD genes and
268 highlight key chromatin modifiers and transcription factors, including *CHD8* and *REST*, as critical
269 drivers of this convergence. These drivers were highly interconnected with each other, often via mutual
270 regulatory relationships, revealing how disruption of a single regulator can lead to changes that
271 propagate throughout a GRN. We also identified the transcription factor *ZFX*, which is expressed from
272 both the active and "inactive" X chromosome in females, as a key transcriptional regulator of ASD
273 genes; the female-biased expression of *ZFX* could thus buffer the effects of ASD gene disruptions in
274 females, contributing to the FPE in ASD. Taken together, this work provides a model for how diverse
275 genes can converge on core transcriptomic changes and ultimately generate shared ASD phenotypes,
276 as well as how sex-differential activity of these genes can lead to male-biased risk for ASD.

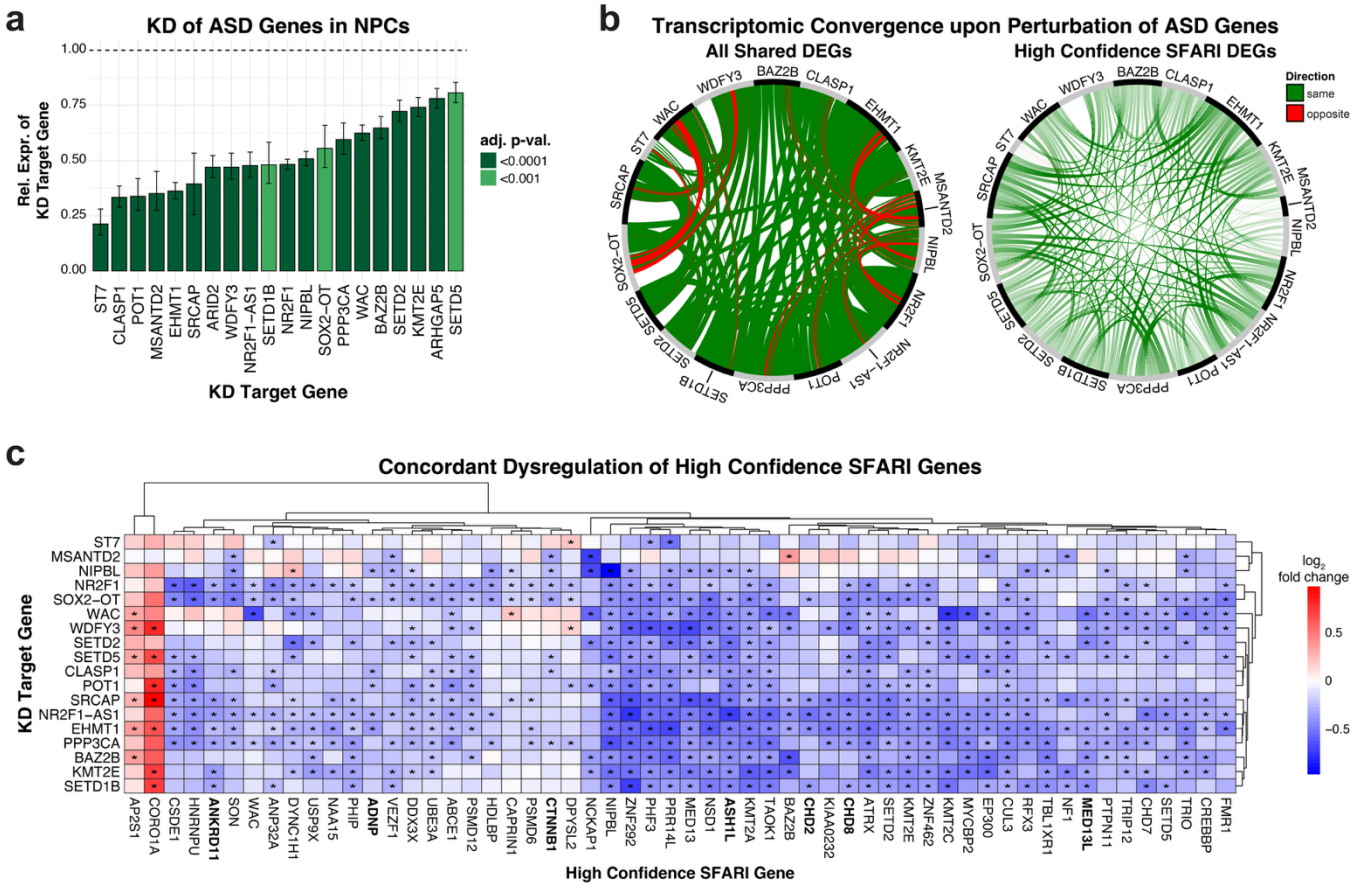
277 Building an understanding of the complex relationships within GRNs is critical for identifying key
278 "rate-limiting" genes that affect the GRN as a whole and ultimately lead to a particular biological
279 outcome[80]. Prior studies of Hirschsprung disease have identified an analogous GRN centered on the
280 rate-limiting gene *RET*, which can be impacted by direct mutations in *RET* itself or disruptions of
281 upstream regulators of *RET*[81]. Here, we have similarly identified critical genes such as *CHD8*, *REST*,
282 and *ZFX*, whose effects can propagate throughout the GRN and drive transcriptomic convergence in
283 ASD.

284 Our functional studies in cerebral organoids uncovered transcriptomic convergence of ASD genes
285 in multiple cell types including ExNs. Future work to reconstruct GRNs from heterogeneous samples
286 through scRNA-seq will be valuable to understand how regulatory relationships are reorganized in
287 different cellular contexts[82–85], though these approaches are currently hampered by limited RNA capture
288 inherent to scRNA-seq techniques. As methodologies advance, it will be possible to reconstruct robust
289 GRNs from cerebral organoids and patient tissue samples, which will further refine our understanding
290 of gene regulation in complex biological systems.

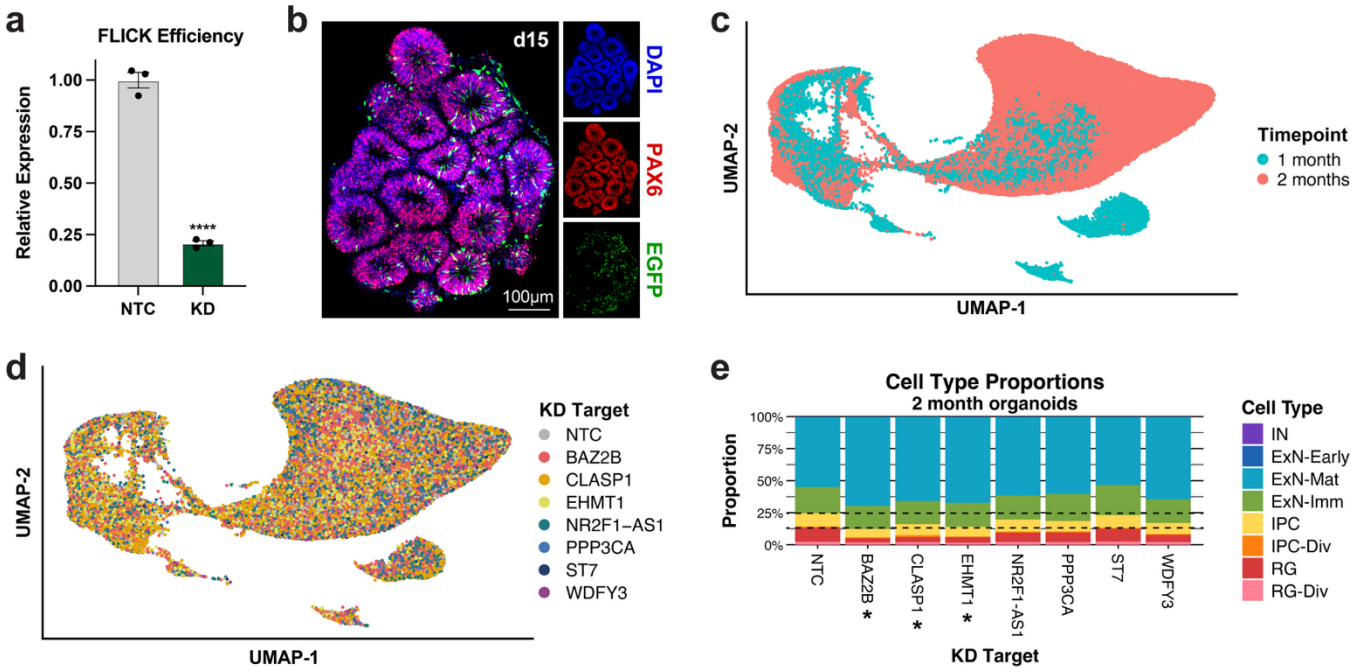
291 Our results have wide-ranging implications for understanding neurodevelopment and disorders
292 such as ASD. Our identification of *ZFX* as a key transcriptional regulator of ASD genes is consistent

293 with a recent report implicating pathogenic *ZFX* variants in ASD[86]. Moreover, downregulation of *ZFX* by
294 KD of many different ASD genes suggests that decreased *ZFX* expression may be a recurrent
295 transcriptional feature of ASD; *ZFX*'s higher baseline level in the female brain due to its expression
296 from both X chromosomes could thus buffer females against this ASD-associated downregulation.
297 While males have a Y-linked homolog of *ZFX* called *ZFY*, it encodes a weaker transcriptional activator
298 than *ZFX*[76], and *ZFY*'s GRN regulon showed no enrichment for SFARI genes (OR = 1.1, adj. p = 0.36),
299 suggesting divergent roles with respect to ASD. In future studies, it will be important to determine
300 whether other genes expressed from both the active and "inactive" X chromosome in females also
301 contribute to the FPE in ASD (M.T. and D.C.P., under review).
302 While our perturbations focused on genes associated with ASD through rare single-gene mutations,
303 it is likely that other classes of ASD-associated variants also disrupt this same convergent network.
304 Supporting this, studies of postmortem brain tissue have demonstrated that idiopathic ASD cases share
305 core transcriptomic changes[42,47] and overlapping DNA methylation signatures[87] with cases of ASD
306 caused by rare mutations of large effect size, such as chromosome 15q11-13 duplication. Moreover,
307 reports suggest that common alleles of smaller effect size[88] implicated in ASD can also drive changes
308 that converge with those caused by rare ASD variants[89]. This raises the intriguing possibility that
309 distinct causes of ASD – including rare single-gene mutations, large copy number variants (CNVs), and
310 common genetic variation – could result in similar phenotypes by each converging on disruption of
311 shared ASD-relevant GRNs. Thus, work such as this that characterizes these GRNs has broad
312 potential to clarify our understanding of the genetic architecture of ASD across a wide variety of genetic
313 variant classes, uncovering potential therapeutic targets that could benefit individuals with diverse
314 causes of ASD. More broadly, while this work focused on ASD, the approaches used here are widely
315 applicable to different cell types and conditions. Thus, this study serves as a framework for uncovering
316 critical transcriptional regulators across a wide diversity of contexts, from development through disease.

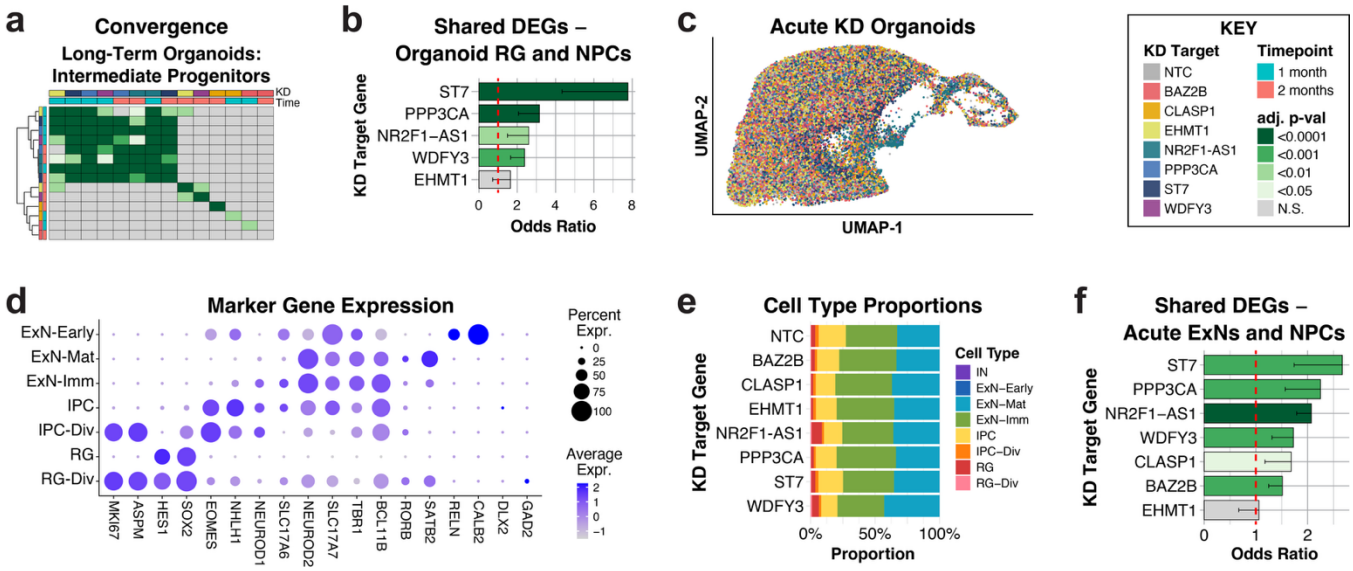
EXTENDED DATA FIGURES



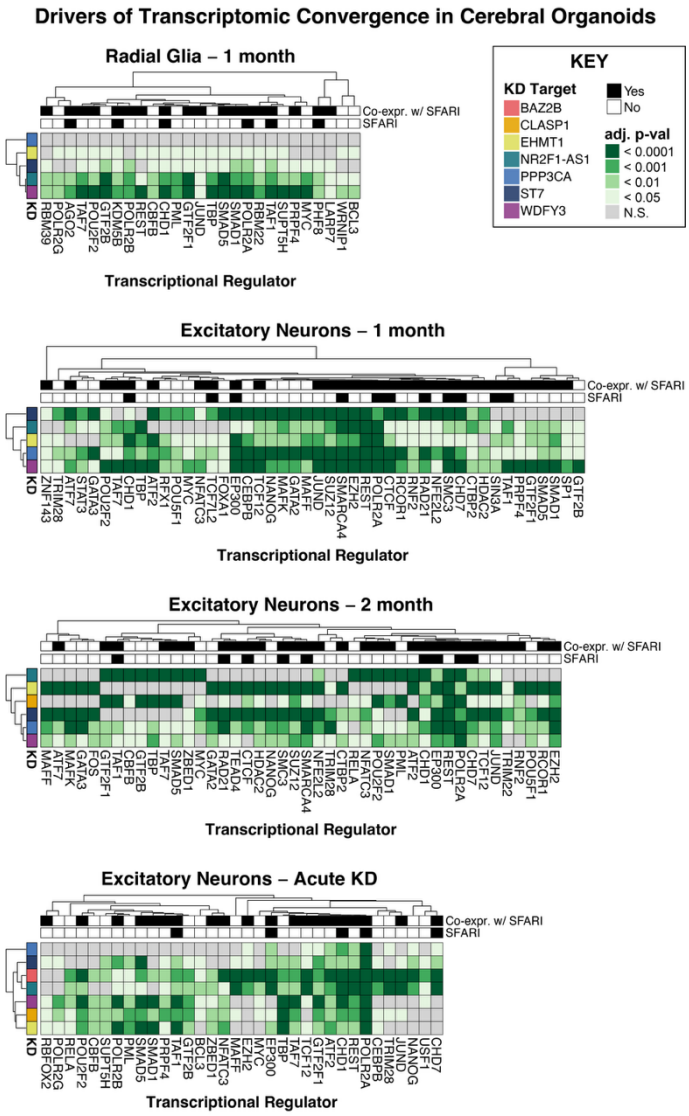
Extended Data Fig. 1 | Transcriptomic convergence upon depletion of ASD genes in NPCs.
a, Relative expression of the targeted gene within each KD sample compared to non-targeting control (NTC) samples. Error bars: 95% confidence interval. **b,** Circos plots depicting DEGs shared between different KD conditions. For direct comparison, the circos plot for high confidence SFARI genes (shown in **Fig. 1d**) is also shown here. **c,** Heatmap showing the relative expression of high confidence SFARI genes that are DEGs in at least 5 KD conditions. Statistically significant DEGs are indicated with *. Genes that were identified amongst the 26 strongest ASD risk genes in [2] are bolded.



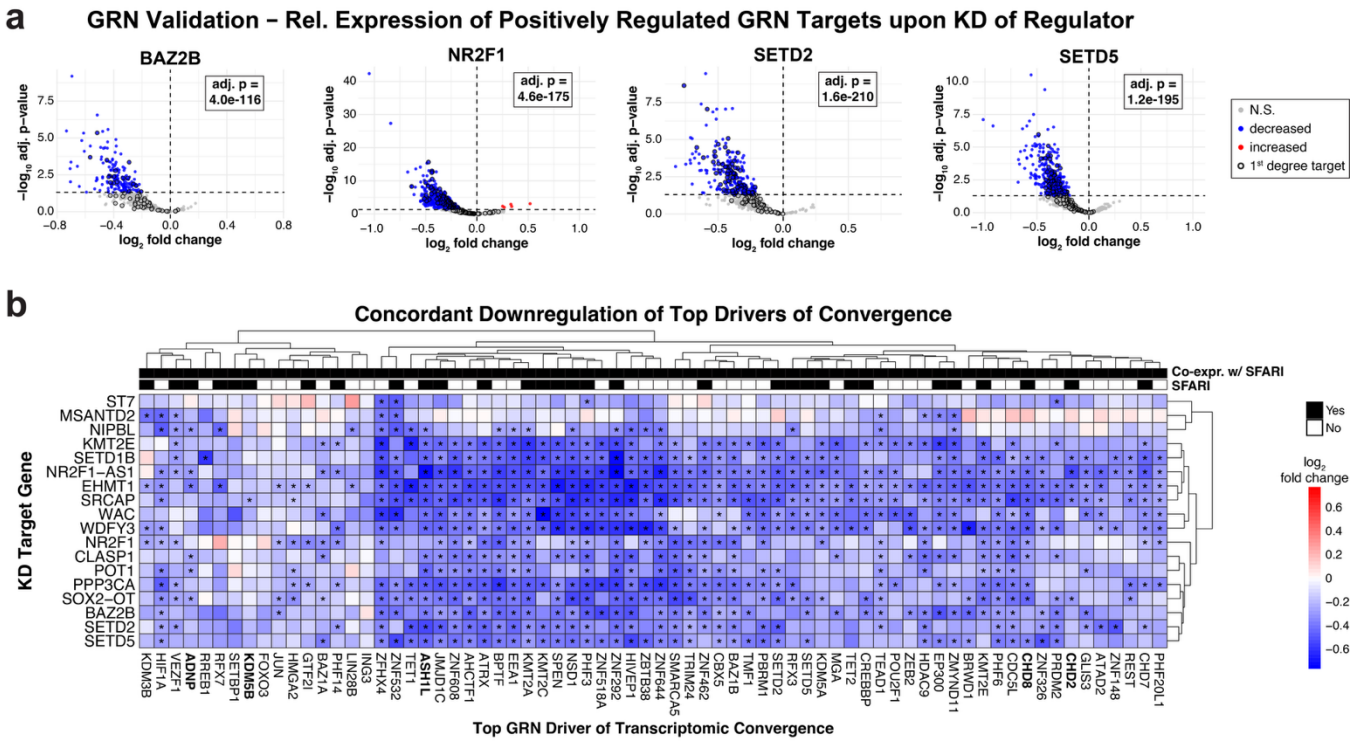
Extended Data Fig. 2 | Mosaic cerebral organoids with KD of ASD genes via the FLICK construct.
a, KD efficiency of FLICK constructs. Bar plot depicts the relative expression of the gene *ENSG00000261840* in iPSCs expressing FLICK constructs with non-targeting controls (NTC) or gRNAs that target the gene for degradation (KD), displayed as mean +/- SEM. **** $p < 0.0001$, two-tailed t test. **b,** Immunohistochemistry of a FLICK organoid at day 15. Nuclei are labeled with DAPI (blue), cells with active FLICK constructs are labeled with the reporter EGFP (green), and NPCs are identified by expression of PAX6 (red). **c,** UMAP of organoid scRNA-seq data colored by timepoint. **d,** UMAP of organoid scRNA-seq data colored by targeted gene. **e,** Cell type composition per KD target in two-month organoids. Dotted lines indicate the NTC proportions of major cell types (radial glia, intermediate progenitors, and neurons). Samples for which the ExN/RG proportion was significantly different than NTC are indicated with * (adj. $p < 0.05$). Statistical analysis was performed using ordinary one-way ANOVA with Dunnett's multiple comparisons test, considering all long-term samples from one month (see **Fig. 2f**) and two months (shown here).



Extended Data Fig. 3 | KD of ASD genes in cerebral organoids reveals transcriptomic convergence. **a,** Heatmap of cross-enrichment of differentially expressed genes (DEGs) from intermediate progenitor cells. **b,** Bar plot of cross-enrichment of DEGs from radial glia (RG) from one-month organoids versus DEGs from previous NPC cultures upon KD of the same target gene. Only perturbations with at least 20 DEGs were tested for cross-enrichment. Error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown). **c,** UMAP of scRNA-seq data from acute KD organoids, colored by KD target gene. **d,** Expression of canonical brain marker genes used for identifying cell types by scRNA-seq. **e,** Cell type composition across the different perturbations in acute KD organoids. **f,** Bar plot of cross-enrichment of DEGs from excitatory neurons (ExNs) from acute KD mosaic organoids versus DEGs from previous NPC cultures upon KD of the same target gene. Error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown).

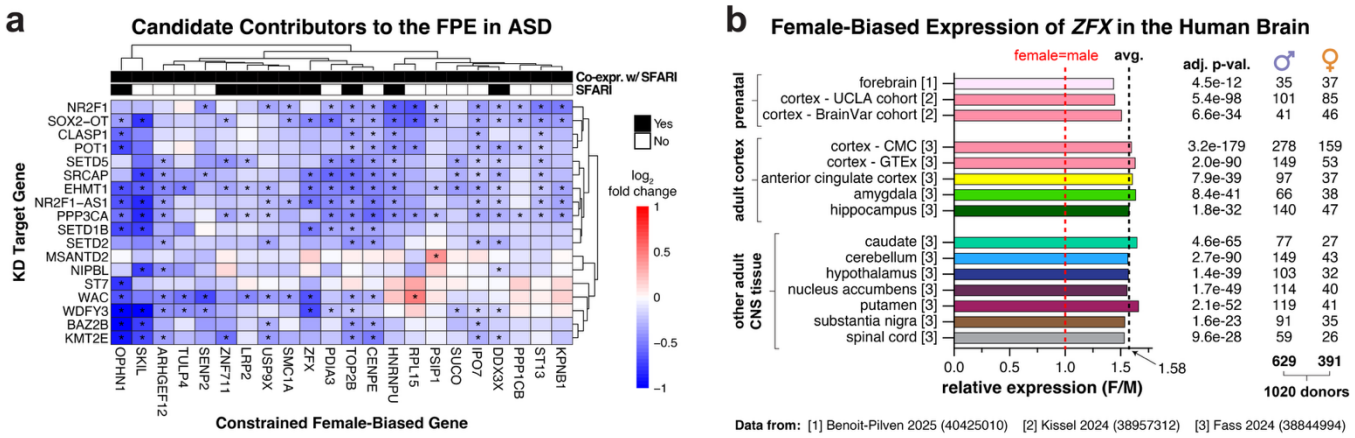


Extended Data Fig. 4 | Genes that exhibit recurrent dysregulation upon ASD gene KD in cerebral organoids are targets of key transcriptional regulators.
MAGIC analysis identifying transcriptional regulators whose targets are enriched within the DEGs from different cell types and stages of organoid development following KD of ASD genes.



Extended Data Fig. 5 | Top drivers of convergence identified through reconstruction of an NPC gene regulatory network.

a, Volcano plot of genes predicted to be positively regulated by the regulator shown above the plot, depicting their differential expression upon KD of the regulator. Statistically significant DEGs with decreased expression are shown in blue, while DEGs with increased expression are shown in red. Statistically significant enrichment of 2nd degree regulon genes within the downregulated DEGs is indicated with the adjusted p-value. **b,** Heatmap showing the relative expression of the top GRN drivers of convergence across the different ASD gene KD samples. Statistically significant DEGs are indicated with *. Genes that were identified amongst the 26 strongest ASD risk genes in [2] are bolded.



Extended Data Fig. 6 | *ZFX* exhibits female-biased expression in the brain and is recurrently downregulated upon ASD gene KD.

a, Heatmap showing the relative expression of candidate contributors to the FPE in ASD across KD conditions. Statistically significant DEGs are indicated with *. **b,** Bar plot depicting the significantly female-biased expression of *ZFX* across several tissues of the human central nervous system (CNS), including prenatal and adult cortex.

REFERENCES

1. Bai, D. *et al.* Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. *JAMA Psychiatry* **76**, 1035 (2019).
2. Satterstrom, F. K. *et al.* Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism Article Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell* 1–17 (2020) doi:10.1016/j.cell.2019.12.036.
3. Trost, B. *et al.* Genomic architecture of autism from comprehensive whole-genome sequence annotation. *Cell* **185**, 4409-4427.e18 (2022).
4. Ruzzo, E. K. *et al.* Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* **178**, 850-866.e26 (2019).
5. Gaugler, T. *et al.* Most genetic risk for autism resides with common variation. *Nat Genet* **46**, 881–5 (2014).
6. 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 Genomics* **3**, 100277 (2023).
7. Loomes, R., Hull, L. & Mandy, W. P. L. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry* **56**, 466–474 (2017).
8. Maenner, M. J. *et al.* Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2020. *MMWR Surveill Summ* **72**, 1–14 (2023).
9. Iossifov, I. *et al.* The contribution of de novo coding mutations to autism spectrum disorder. *Nature* **515**, 216–21 (2014).
10. Doan, R. N. *et al.* Recessive gene disruptions in autism spectrum disorder. *Nat Genet* **51**, 1092–1098 (2019).
11. Jacquemont, S. *et al.* A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *Am J Hum Genet* **94**, 415–25 (2014).
12. Shin, T. *et al.* Rare variation in non-coding regions with evolutionary signatures contributes to autism spectrum disorder risk. *Cell genomics* **4**, 100609 (2024).
13. Antaki, D. *et al.* A phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. *Nat Genet* **54**, 1284–1292 (2022).
14. Werling, D. M. The role of sex-differential biology in risk for autism spectrum disorder. *Biol Sex Differ* **7**, 58 (2016).
15. Dougherty, J. D. *et al.* Can the "female protective effect" liability threshold model explain sex differences in autism spectrum disorder? *Neuron* **110**, 3243–3262 (2022).
16. Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. & Page, D. C. ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation. *Cell* **57**, 1247–58 (1989).
17. Litman, A. *et al.* Decomposition of phenotypic heterogeneity in autism reveals underlying genetic programs. *Nat Genet* **57**, 1611–1619 (2025).
18. Iakoucheva, L. M., Muotri, A. R. & Sebat, J. Getting to the Cores of Autism. *Cell* **178**, 1287–1298 (2019).
19. Mendes, M. *et al.* Chromosome X-wide common variant association study in autism spectrum disorder. *Am J Hum Genet* **0**, (2024).
20. Ropers, H.-H. & Hamel, B. C. J. X-linked mental retardation. *Nat Rev Genet* **6**, 46–57 (2005).
21. Martin, H. C. *et al.* The contribution of X-linked coding variation to severe developmental disorders. *Nat Commun* **12**, 627 (2021).
22. Abrahams, B. S. *et al.* SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).
23. Banerjee-Basu, S. & Packer, A. SFARI Gene: an evolving database for the autism research community. *Dis Model Mech* **3**, 133–5 (2010).
24. Ang, C. E. *et al.* The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *Elife* **8**, (2019).

25. Andersen, R. E. *et al.* Chromosomal structural rearrangements implicate long non-coding RNAs in rare germline disorders. *Hum Genet* **143**, 921–938 (2024).
26. Konermann, S. *et al.* Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell* **173**, 665-676.e14 (2018).
27. Wei, J. *et al.* Deep learning and CRISPR-Cas13d ortholog discovery for optimized RNA targeting. *Cell Syst* **14**, 1087-1102.e13 (2023).
28. Kochinke, K. *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am J Hum Genet* **98**, 149–64 (2016).
29. Haddad Derafshi, B. *et al.* The autism risk factor CHD8 is a chromatin activator in human neurons and functionally dependent on the ERK-MAPK pathway effector ELK1. *Sci Rep* **12**, 22425 (2022).
30. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
31. Weerts, M. J. A. *et al.* Delineating the molecular and phenotypic spectrum of the SETD1B-related syndrome. *Genetics in Medicine* **23**, 2122–2137 (2021).
32. Rabin, R. *et al.* Genotype-phenotype correlation at codon 1740 of SETD2. *Am J Med Genet A* **182**, 2037–2048 (2020).
33. Crippa, M. *et al.* SETD5 Gene Haploinsufficiency in Three Patients With Suspected KBG Syndrome. *Front Neurol* **11**, 631 (2020).
34. O'Donnell-Luria, A. H. *et al.* Heterozygous Variants in KMT2E Cause a Spectrum of Neurodevelopmental Disorders and Epilepsy. *Am J Hum Genet* **104**, 1210–1222 (2019).
35. Quental, R. *et al.* Congenital heart defects associated with pathogenic variants in WAC gene: Expanding the phenotypic and genotypic spectrum of DeSanto-Shinawi syndrome. *Am J Med Genet A* **188**, 1311–1316 (2022).
36. Connacher, R. *et al.* Autism NPCs from both idiopathic and CNV 16p11.2 deletion patients exhibit dysregulation of proliferation and mitogenic responses. *Stem Cell Reports* **17**, 1380–1394 (2022).
37. Alsafh, R. *et al.* Multiplex Consanguineous Family Highlights CLASP1 as a Candidate Gene for Lissencephaly. *Neurol Genet* **10**, e200172 (2024).
38. Rots, D. *et al.* Comprehensive EHMT1 variants analysis broadens genotype-phenotype associations and molecular mechanisms in Kleefstra syndrome. *The American Journal of Human Genetics* **111**, 1605–1625 (2024).
39. Paulsen, B. *et al.* Autism genes converge on asynchronous development of shared neuron classes. *Nature* **602**, 268–273 (2022).
40. Jourdon, A. *et al.* Modeling idiopathic autism in forebrain organoids reveals an imbalance of excitatory cortical neuron subtypes during early neurogenesis. *Nat Neurosci* **26**, 1505–1515 (2023).
41. Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–4 (2011).
42. Parikshak, N. N. *et al.* Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
43. Ramaswami, G. *et al.* Integrative genomics identifies a convergent molecular subtype that links epigenomic with transcriptomic differences in autism. *Nat Commun* **11**, 4873 (2020).
44. Schwede, M. *et al.* Strong correlation of downregulated genes related to synaptic transmission and mitochondria in post-mortem autism cerebral cortex. *J Neurodev Disord* **10**, 18 (2018).
45. Velmeshev, D. *et al.* Single-cell genomics identifies cell type–specific molecular changes in autism. *Science (1979)* **364**, 685–689 (2019).
46. Gandal, M. J. *et al.* Broad transcriptomic dysregulation occurs across the cerebral cortex in ASD. *Nature* **611**, 532–539 (2022).
47. Wamsley, B. *et al.* Molecular cascades and cell type-specific signatures in ASD revealed by single-cell genomics. *Science* **384**, eadh2602 (2024).
48. 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).

49. Li, C. *et al.* Single-cell brain organoid screening identifies developmental defects in autism. *Nature* **621**, 373–380 (2023).
50. Deneault, E. *et al.* Complete Disruption of Autism-Susceptibility Genes by Gene Editing Predominantly Reduces Functional Connectivity of Isogenic Human Neurons. *Stem Cell Reports* **11**, 1211–1225 (2018).
51. DeRosa, B. A. *et al.* Convergent Pathways in Idiopathic Autism Revealed by Time Course Transcriptomic Analysis of Patient-Derived Neurons. *Sci Rep* **8**, 8423 (2018).
52. Roopra, A. MAGIC: A tool for predicting transcription factors and cofactors driving gene sets using ENCODE data. *PLoS Comput Biol* **16**, e1007800 (2020).
53. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–3 (1995).
54. Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–57 (1995).
55. Rockowitz, S. & Zheng, D. Significant expansion of the REST/NRSF cistrome in human versus mouse embryonic stem cells: potential implications for neural development. *Nucleic Acids Res* **43**, 5730–43 (2015).
56. Margolin, A. A. *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7 Suppl 1**, S7 (2006).
57. Lachmann, A., Giorgi, F. M., Lopez, G. & Califano, A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. *Bioinformatics* **32**, 2233–2235 (2016).
58. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
59. Boukas, L. *et al.* Coexpression patterns define epigenetic regulators associated with neurological dysfunction. *Genome Res* **29**, 532–542 (2019).
60. Page, L., Brin, S., Motwani, R. & Winograd, T. The PageRank Citation Ranking : Bringing Order to the Web. *The Web Conference* (1999).
61. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).
62. Alvarez, M. J. *et al.* Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* **48**, 838–47 (2016).
63. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638–D646 (2023).
64. Sugathan, A. *et al.* CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc Natl Acad Sci U S A* **111**, E4468-77 (2014).
65. Cotney, J. *et al.* The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat Commun* **6**, 6404 (2015).
66. Werling, D. M., Parikshak, N. N. & Geschwind, D. H. Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. *Nat Commun* **7**, 10717 (2016).
67. Velmeshev, D. *et al.* Single-cell analysis of prenatal and postnatal human cortical development. *Science (1979)* **382**, (2023).
68. Kissel, L. T. *et al.* Sex-Differential Gene Expression in Developing Human Cortex and Its Intersection With Autism Risk Pathways. *Biological psychiatry global open science* **4**, 100321 (2024).
69. Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369**, (2020).
70. Naqvi, S. *et al.* Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, (2019).
71. Blanton, L. V *et al.* Stable and robust Xi and Y transcriptomes drive cell-type-specific autosomal and Xa responses in vivo and in vitro in four human cell types. *Cell genomics* **4**, 100628 (2024).
72. San Roman, A. K. *et al.* The human inactive X chromosome modulates expression of the active X chromosome. *Cell genomics* **3**, 100259 (2023).
73. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).

74. DeCasien, A. R. *et al.* Sex effects on gene expression across the human cerebral cortex at single cell resolution. *bioRxiv* 2025.06.30.661781 Preprint at https://doi.org/10.1101/2025.06.30.661781 (2025).
75. Benoit-Pilven, C. *et al.* Early establishment and life course stability of sex biases in the human brain transcriptome. *Cell genomics* **5**, 100890 (2025).
76. San Roman, A. K. *et al.* The human Y and inactive X chromosomes similarly modulate autosomal gene expression. *Cell genomics* **4**, 100462 (2024).
77. Raznahan, A. *et al.* Sex-chromosome dosage effects on gene expression in humans. *Proc Natl Acad Sci U S A* **115**, 7398–7403 (2018).
78. Zhang, X. *et al.* Integrated functional genomic analyses of Klinefelter and Turner syndromes reveal global network effects of altered X chromosome dosage. *Proc Natl Acad Sci U S A* **117**, 4864–4873 (2020).
79. Rhie, S. K. *et al.* ZFX acts as a transcriptional activator in multiple types of human tumors by binding downstream from transcription start sites at the majority of CpG island promoters. *Genome Res* **28**, 310–320 (2018).
80. Chakravarti, A. & Turner, T. N. Revealing rate-limiting steps in complex disease biology: The crucial importance of studying rare, extreme-phenotype families. *Bioessays* **38**, 578–86 (2016).
81. Chatterjee, S. *et al.* Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease. *Cell* **167**, 355-368.e10 (2016).
82. Obradovic, A. *et al.* Single-cell protein activity analysis identifies recurrence-associated renal tumor macrophages. *Cell* **184**, 2988-3005.e16 (2021).
83. Vlahos, L. *et al.* Systematic, Protein Activity-based Characterization of Single Cell State. *bioRxiv* 2021.05.20.445002 Preprint at https://doi.org/10.1101/2021.05.20.445002 (2021).
84. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).
85. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* **20**, 1355–1367 (2023).
86. Shepherdson, J. L. *et al.* Variants in ZFX are associated with an X-linked neurodevelopmental disorder with recurrent facial gestalt. *Am J Hum Genet* **111**, 487–508 (2024).
87. 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).
88. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet* **51**, 431–444 (2019).
89. 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).
90. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
91. Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc Natl Acad Sci U S A* **101**, 10380–5 (2004).
92. Yusa, K., Zhou, L., Li, M. A., Bradley, A. & Craig, N. L. A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci U S A* **108**, 1531–6 (2011).
93. Kumamoto, T. *et al.* Direct Readout of Neural Stem Cell Transgenesis with an Integration-Coupled Gene Expression Switch. *Neuron* **107**, 617-630.e6 (2020).
94. Schnütgen, F. *et al.* A directional strategy for monitoring Cre-mediated recombination at the cellular level in the mouse. *Nat Biotechnol* **21**, 562–5 (2003).
95. Kaczmarczyk, S. J. & Green, J. E. A single vector containing modified cre recombinase and LOX recombination sequences for inducible tissue-specific amplification of gene expression. *Nucleic Acids Res* **29**, E56-6 (2001).
96. Qian, K. *et al.* A simple and efficient system for regulating gene expression in human pluripotent stem cells and derivatives. *Stem Cells* **32**, 1230–8 (2014).
97. Schlaeger, T. M. *et al.* A comparison of non-integrating reprogramming methods. *Nat Biotechnol* **33**, 58–63 (2015).

98. de Leeuw, V. C. *et al.* An efficient neuron-astrocyte differentiation protocol from human embryonic stem cell-derived neural progenitors to assess chemical-induced developmental neurotoxicity. *Reprod Toxicol* **98**, 107–116 (2020).
99. Giandomenico, S. L., Sutcliffe, M. & Lancaster, M. A. Generation and long-term culture of advanced cerebral organoids for studying later stages of neural development. *Nat Protoc* **16**, 579–602 (2021).
100. Han, X. *et al.* Millisecond-timescale optical control of neural dynamics in the nonhuman primate brain. *Neuron* **62**, 191–8 (2009).
101. Lohman, B. K., Weber, J. N. & Bolnick, D. I. Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Mol Ecol Resour* **16**, 1315–1321 (2016).
102. Rabe, B. A. Investigation of the Multiple Roles of Notch Signaling in Embryonic Retinal Development Using Novel High-Throughput Techniques. (Harvard University Graduate School of Arts and Sciences, 2020).
103. Ye, J. *et al.* Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134 (2012).
104. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
105. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545–15550 (2005).

METHODS

Knockdown with CasRx

CasRx[26], specifically NLS-RfxCas13d-NLS (Addgene #109049), was used in combination with target-specific guide RNAs (gRNAs) to knock down (KD) targeted transcripts. A deep-learning-based tool[27] was used to design gRNA spacer sequences targeting the genes of interest. These spacer sequences were filtered to exclude homopolymers (greater than 4 As, greater than 3 Cs, Gs, or Ts). Spacer sequences with potential off-target interactions were identified using BLAST[90] and excluded. For each target gene, 3 separate spacer sequences were selected and combined into a tandem pre-gRNA array, with each spacer sequence preceded by the CasRx direct repeat (DR36) sequence[27]: CAAGTAAACCCCTACCAACTGGTCGGGGTTTGAAAC. Tandem pre-gRNA arrays are processed by CasRx into individual gRNAs for KD[26]. See **Supplementary Table 1** for full sequences.

Molecular Cloning

To create constructs for the initial NPC KD experiments, the lentiviral pSicoR plasmid[91] (Addgene #11579) was used as the backbone. The mouse U6 promoter was replaced with the human U6 promoter, which was used to drive expression of the CasRx gRNA arrays. These constructs were used in combination with a CasRx knock-in cell line (described below).

For NPC proliferation assays, an all-in-one construct was created that contained CasRx and EGFP expressed from the CAG promoter element and gRNA arrays expressed from the human U6 promoter. The backbone also contained inverted terminal repeats (ITRs), allowing the construct to be integrated into the genome by piggyBac transposase[92]. These constructs were based on the LiOn design[93] (Addgene #154019), such that expression of CasRx and EGFP was dependent upon recombination of the construct by piggyBac transposase, preventing expression from the episomal plasmid.

For organoid experiments, we created the FLEx-based Inducible CRISPR Knockdown (FLICK) construct by engineering several features into a backbone incorporating the FLEx design[94]. In particular, we included a modified Cre recombinase that includes an intron, which prevents recombination of the plasmid during cloning[95]. We also included CasRx[26] as well as all necessary components of the Tet-On 3G system[96] to enable doxycycline-inducible recombination. Furthermore, we included identifying barcode sequences shortly upstream of the polyA sites for the fluorescent reporters, which enabled determination of which KD construct was incorporated in each cell through

587 single-cell RNA sequencing (scRNA-seq). Finally, we incorporated ITRs, allowing the construct to be
588 integrated into the genome by piggyBac transposase[92].
589
590 **Cell Culture**
591 All human cell culture work was performed under protocols approved by Boston Children's Hospital
592 IRB #S07-02-0087. The control male human iPSC line "280" was previously generated and fully
593 characterized[97]. Human iPSCs and iPSC-derived NPCs were grown on tissue culture plates pre-coated
594 with Matrigel (Corning #354277) or Geltrex (ThermoFisher #A1413302). The cells were passaged as
595 needed using Accutase (Millipore Sigma #SCR005) or ReleSR (StemCell Technologies #100-0483).
596 Human iPSCs were cultured in mTeSR™ Plus medium (StemCell Technologies #100-0276). Human
597 NPCs[98] were generated from iPSCs using the STEMdiff™ SMADi Neural Induction Kit (StemCell
598 Technologies #08582) by following the Monolayer Culture Protocol for at least three weeks. NPCs were
599 continuously grown in Neural Induction Medium + SMADi and were not switched to Neural Progenitor
600 Medium. The medium was supplemented with 10uM Y-27632 (StemCell Technologies #72304) for
601 approximately 1 day following passaging of iPSCs or NPCs.
602 Cerebral organoids were generated from 280-iPSCs using the STEMdiff™ Cerebral Organoid Kit
603 (StemCell Technologies # 08570) with some recommendations from Giandomenico et al., 2021[99] as
604 well as additional modifications. In particular, the organoids were embedded in Matrigel (Corning
605 #354277) at day 7 and grown in Expansion Medium until day 14, at which point they were removed
606 from Matrigel and switched to Maturation Medium on an orbital shaker. Please see full details on
607 protocols.io.
608
609 **Derivation of CasRx Knock-in iPSCs**
610 CasRx[26] was knocked into the *AAVS1* safe harbor locus of 280-iPSCs using targeting constructs
611 that were previously described[96]. Individual iPSC colonies were picked and expanded. Southern blot
612 analysis was performed by Celplor to confirm correct targeting of the *AAVS1* locus. Karyotype analysis
613 was performed using KaryoStat (ThermoFisher), which revealed a 2.4Mb gain on chromosome 1, which
614 is described in microarray nomenclature as arr[GRCh37] 1q32.1(204108163_206539638)x3. This
615 region did not contain any of the target genes.
616
617 **Lentivirus**
618 Lentivirus was generated using standard protocols that have previously been described[100].
619 Lentiviruses were applied at a multiplicity of infection (MOI) of 4.0, as this MOI yields >98% transduced
620 cells. RNA was extracted 24 hours later.
621
622 **Library Preparation for RNA Sequencing (RNA-Seq)**
623 Bulk RNA was extracted using the Direct-zol RNA Microprep Kit (Zymo Research #R2062), and 3'
624 RNA-seq libraries were prepared following a modified version of the TagSeq protocol[101], incorporating
625 recommendations and pooling strategies from the CheapSeq protocol[102]. Briefly, 200ng of total RNA
626 from each sample was combined with an indexed oligo(dT) primer and buffer (from NEB #M0466L) as
627 described for CheapSeq. RNA was fragmented in a pre-heated thermocycler at 94C for 2.5 min without
628 the lid, after which the samples were moved immediately to ice. Each sample was then mixed with a
629 template-switching oligo (TSO) that contained an index and a unique molecular identifier (UMI) along
630 with additional reagents required for reverse transcription, as described for CheapSeq. The reverse
631 transcription reaction was performed at 42C for 90 minutes, without heat inactivation. Remaining single-
632 stranded RNA and oligos were then digested by adding 0.5uL of Buffer EB (Qiagen #19086), 0.5uL of
633 thermolabile ExoI (NEB #M0568S), and 0.25uL of RNase A (ThermoFisher #EN0531), and incubating
634 the samples at 37C for 30min, followed by heat inactivation at 85C for 5min and a hold step at 4C.
635 Up to 8 samples with different index sequences were then pooled, and cDNA was purified by
636 performing a 1.4x left-sided SPRI selection (Beckman Coulter #B23317). Q5 UltraII Mastermix (NEB
637 #M0544L) was then used along with partial Read1 and partial Read2 primers to amplify the cDNA for 7
638 cycles using a C1000 Touch Thermal Cycler (Bio-Rad #1851197), following the "CheapSeq PCR"
639 parameters described below. Amplified cDNA was purified through a 0.7x left-sided SPRI selection
640 (Beckman Coulter #B23317). Final library construction was performed by using Q5 UltraII Mastermix

(NEB #M0544L) along with standard TruSeq indexing primers (NEB #E6440S) and amplifying for 8 cycles following the "CheapSeq PCR" parameters described below. Libraries were purified by performing a 0.4x-0.65x double-sided SPRI selection (Beckman Coulter #B23317). Purified libraries were assessed using Tapestation HSD1000 (Agilent #5067-5584 and #5067-5585). For oligo sequences, see **Supplementary Table 9**.

CheapSeq PCR:

1. 98C for 30s
2. X cycles of 98C for 10s, 65C for 75s
3. 65C for 5min
4. 4C forever

Differentially Expressed Gene (DEG) Analysis of Bulk RNA-seq from NPCs

Three distinct NPC samples were analyzed per condition. Individual KD conditions were compared to three non-targeting control (NTC) conditions, which were considered together as a set of 9 control samples.

Gene expression was quantified using alevin-fry with UMI deduplication (salmon-v1.9). A salmon index and transcript-to-gene map were built using Gencode v45. Counts from distinct transcripts that mapped to the same gene were summed. Only genes with a baseMean expression of >10 in NTC samples were considered for further analyses. Differential expression analysis was performed using DESeq2 (v1.42.0), with the model ~Batch + sgRNA.

For KD confirmation, p-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing for the set of 20 directly targeted genes. The raw p-value for SRCAP was derived from Droplet Digital PCR (see below), as SRCAP was filtered from RNA-seq analysis due to low expression. See **Supplementary Table 2** for details.

For transcriptome-wide analyses, p-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing for the set of 11,168 analyzed genes (baseMean >10 in NTCs).

Droplet Digital PCR (ddPCR)

To confirm KD of SRCAP, which was filtered from RNA-seq analysis due to low expression, the same RNA samples that had been used for RNA-seq (above) were assessed by ddPCR. RNA was first converted to cDNA using the SuperScript™ IV VILO™ Master Mix (ThermoFisher Scientific # 11756050). Next, ddPCR was performed by combining cDNA at an RNA-equivalent concentration of 0.25ng/uL (5ng per 20uL reaction), primers at a final concentration of 0.1uM, and QX200™ ddPCR™ EvaGreen Supermix (Bio-Rad # 1864033). The ddPCR reactions were performed using the QX200 AutoDG Droplet Digital PCR System (Bio-Rad). Expression levels were normalized using the expression of the housekeeping gene *HPRT1*. The results were analyzed using the model Expr ~ Batch + type, where type was either NTC or the targeted gene (SRCAP). The resulting raw p-value was adjusted along with those for the other KD targets (from RNA-seq, above) using Benjamini-Hochberg correction for multiple hypothesis testing for the set of 20 directly targeted genes. See **Supplementary Table 2** for details.

Primer Design

Primers (**Supplementary Table 10**) were designed using Primer-BLAST[103], which assesses potential off-target interactions using BLAST[90]. Oligos were ordered through Azenta or IDT. Primer specificity was confirmed by Sanger sequencing of PCR products (Azenta).

Gene Sets

1. SFARI genes: all genes included in the SFARI gene database[22,23] (4/3/25 release).
2. High confidence SFARI (HC-SFARI) genes: SFARI genes with a score of 1.
3. Intellectual disability (ID) genes: genes from the SysNDD database[28] (accessed 8/23/25) that were:
    a. associated with an ID term (HP:0001249, HP:0001256, HP:0002342, HP:0010864, HP:0002187)
    b. in the "definitive" category

c. not SFARI genes

Enrichment Analyses

Enrichment analyses were performed using hypergeometric tests, and p-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing. When displayed as bar plots, main bars indicate odds ratio and error bars indicate lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown).

Analysis of Expression Correlation with SFARI Genes

To determine whether a gene of interest was significantly co-expressed with SFARI genes in our NPC RNA-seq data, we calculated the median Pearson correlation between the gene of interest and each SFARI gene. To assess whether this degree of correlation was significant, we created an empirical null distribution by calculating the median correlation between the gene of interest and 1000 randomly selected gene sets of the same size as the SFARI gene set, using the universe of NPC-expressed genes (baseMean > 10 in NTCs). A one-sided empirical p-value was then calculated as the proportion of null values greater than or equal to the observed median correlation. Regulators with p-values < 0.05 were considered significantly correlated with SFARI genes.

Identification of Consistently Downregulated DEGs

Consistently downregulated DEGs were defined as genes that exhibited: 1) statistically significant downregulation by at least three of the KDs, and 2) at least 80% concordance in the direction of significant differential expression across all KDs.

Constraint Analysis

LOEUF scores (oe_lof_upper) were obtained from gnomad v2[30] to evaluate genic constraint. For each gene, we calculated a per-gene LOEUF score by calculating the median of the LOEUF values of each gene's associated transcripts. While LOEUF is a continuous metric, we utilized LOEUF < 0.35 as recommended to identify highly constrained genes.

Nucleofection

Nucleofections were performed using the Neon Transfection System (ThermoFisher Scientific). For NPCs, 2M cells were nucleofected with 10ug of KD plasmid and 2ug of a plasmid expressing piggyBac transposase[92] using the 100uL tips. For iPSCs, 1M cells were nucleofected with 5ug of KD plasmid and 1ug of a plasmid expressing piggyBac transposase[92] using the 100uL tips. For small-scale nucleofections, 10uL tips were used and the cell number and plasmid amounts were reduced proportionally. Nucleofection of NPCs was performed using the following settings: voltage = 1600V, width = 10, pulses = 3. Nucleofection of iPSCs was performed using the following settings: voltage = 950V, width = 30, pulses = 2.

NPC Proliferation Assays

LiOn-based KD constructs (see "Molecular Cloning") were nucleofected into 280-NPCs along with a piggyBac-transposase-expressing plasmid (see "Nucleofection" for details). Four days later, the cells were treated for 20 minutes with dye from the CYTOTRACK™ Red 628/643 Cell Proliferation Assay Kit (Bio-Rad #1351205) diluted 1:500 in media. The cells were then rinsed with PBS (ThermoFisher Scientific # 10010031) and passaged, with 1/5 of each sample plated into a well of 4 separate plates to be used for the subsequent 4 timepoints. The remaining cells were used for flow cytometry to assess the initial CytoTrack intensity on day 0 (d0). For the next 4 days, 1 well of each sample was dissociated each day and analyzed by flow cytometry. The CytoTrack intensity values (negative log2) from the EGFP+ cells expressing the KD construct were normalized to the EGFP- cells from the same well to control for well-to-well variability in proliferation rates. They were further normalized to their initial CytoTrack intensity from d0 to control for variability in dye incorporation. The results for each KD were displayed relative to the mean value from the NTC samples, plotted as mean +/- standard error of the mean (SEM). Statistical analysis was performed using ordinary two-way ANOVA with Dunnett's multiple comparisons test (GraphPad Prism). Full details regarding sample numbers and statistical results are

749 available in **Supplementary Table 6**.
750
751 **Assessment of FLICK KD Efficiency**
752 To determine KD efficiency using FLICK constructs, three distinct samples were analyzed per
753 condition. Control 280 iPSCs were nucleofected with constructs encoding the NTC1 control gRNA array
754 or a gRNA array targeting the gene *ENSG00000261840*, along with a plasmid expressing piggyBac
755 transposase[92] to genomically integrate the FLICK constructs. To induce KD, the cells were treated with
756 2ug/ml doxycycline (Sigma D9891-1G) for 24 hours. Cells that had recombined the FLICK construct
757 were then isolated through fluorescence-activated cell sorting (FACS) (BD FACSymphony™ S6) for
758 EGFP+ cells, which were re-plated in triplicate and grown for an additional two days. RNA was then
759 isolated using the Direct-zol RNA Microprep Kit (Zymo Research #R2062). To generate cDNA, 500ng
760 of RNA was reverse transcribed using SuperScript™ IV VILO™ Master Mix (ThermoFisher Scientific
761 #11756050). Quantitative PCR (qPCR) was performed by combining cDNA at an RNA-equivalent
762 concentration of 1ng/uL (10ng per 10uL reaction), primers at a final concentration of 0.5uM, and
763 PowerUp™ SYBR™ Green Master Mix for qPCR (ThermoFisher Scientific #A25742). The qPCR
764 reactions were run in technical triplicate for each sample on the CFX384 Touch Real-Time PCR
765 Detection System (BioRad #1855484) using the "fast cycling mode" conditions described in the Master
766 Mix protocol. Expression of the KD target gene (*ENSG00000261840*) was assessed using the delta-
767 delta Ct method, with expression first normalized to that of the housekeeping gene *HPRT1* for each
768 sample, and then displayed relative to the average of the NTC control samples, plotted as mean +/-
769 SEM. Statistical significance was assessed using a two-tailed t test (GraphPad Prism).
770
771 **Processing of Cerebral Organoids for Histological Analysis**
772 Cerebral organoids were fixed in 4% paraformaldehyde (PFA), diluted in PBS from 16% PFA
773 (Electron Microscopy Sciences # 15710), at 4°C overnight. Following fixation, cerebral organoids were
774 rinsed with PBS and then equilibrated in 30% sucrose (w/v in PBS) at least overnight at 4°C. For large
775 organoids, the 30% sucrose solution was replaced at least once following overnight incubation to
776 ensure complete cryoprotection. For cryosectioning, cerebral organoids were frozen in Tissue-Tek®
777 O.C.T. Compound (Sakura #4583) on dry ice. A Leica CM3050 S Cryostat was used to obtain 30um
778 sections, which were collected on Fisherbrand™ Superfrost™ Plus Microscope Slides (Fisher Scientific
779 #12-550-15) and stored at -80°C.
780 For immunohistochemical analysis, blocking buffer was prepared by combining 0.3M glycine
781 (Sigma G7403-1KG), 1% bovine serum albumin (Sigma A4503-50G), 0.3% Triton X-100 (Sigma
782 T8787-50ML), 10% normal goat serum (Fisher Scientific # NC9660079), and PBS. Slides were rinsed
783 in PBS to remove O.C.T. and then incubated in blocking buffer for 1 hour at room temperature (RT).
784 Blocking buffer was removed and replaced with fresh blocking buffer containing primary antibodies of
785 interest (see below), and the slides were incubated in a slide moisture chamber (Avantor # 76278-848)
786 at 4°C overnight. Slides were then rinsed with PBS three times for five minutes each. Secondary
787 antibodies (see below) and the nuclear stain DAPI (ThermoFisher Scientific #62248) were diluted in
788 blocking buffer (1:500 and 1:1000, respectively) and added to the slides for 30 minutes at RT, and the
789 slides were protected from light from this point forward. The slides were then rinsed in PBS three times
790 for five minutes each. Cover slips (Epredia #152455) were mounted using Fluoromount-G® (Fisher
791 Scientific #OB100-01), which was allowed to dry at least overnight at RT. After drying, slides were
792 stored at 4°C.
793
794 **Antibodies**
795 • PAX6 (BD Pharmingen #561462): diluted 1:300
796 • Nestin (Millipore Sigma #HPA026111-25UL): diluted 1:500
797 • TBR2 (Millipore Sigma #HPA028896-100UL): diluted 1:300
798 • CTIP2 (Abcam #ab18465): diluted 1:300
799 • EGFP (Aves #GFP-1010): diluted 1:1000
800 • Secondary antibodies (from ThermoFisher Scientific, e.g. #A-21247): diluted 1:500
801

802 **Confocal Imaging and Processing**
803 Cerebral organoid sections were imaged using a Zeiss LSM 980 confocal microscope. Standard
804 image processing was performed using Adobe Photoshop.
805
806 **Cerebral Organoid scRNA-seq**
807 For scRNA-seq, cells were collected from:
808 - 1 month: 10 organoids, pooled together and collected as a single batch
809 - 2 months: 14 organoids collected in two batches, with 7 organoids collected together on day 60
810 and an additional 7 organoids collected together on day 61
811 - acute KD: 8 organoids, pooled together and collected as a single batch

812 To collect cells, mosaic cerebral organoids were cut into pieces using spring scissors, rinsed twice
813 with PBS to remove debris, and then dissociated using the Neural Tissue Dissociation Kit – Postnatal
814 Neurons (Miltenyi Biotec # 130-094-802) using the enzyme mixes described for brain tissue. Gentle
815 dissociation was performed by trituration with pipette tips of decreasing diameter, starting with wide
816 orifice tips (Rainin # 30389218) and progressing to standard tips. Cells with recombined FLICK
817 constructs were isolated through FACS (BD FACSymphony™ S6) for EGFP+ cells. Transcriptome-
818 wide scRNA-seq libraries were prepared using Chromium Next GEM Single Cell 3' Reagent Kits v3.1
819 (Dual Index) from 10X Genomics, following user guide CG000316 Rev A.
820
821 **Construction of FLICK Barcode Sequencing Libraries**
822 Please see full details on protocols.io. Briefly:
823 For quantification of FLICK constructs from the initial pool of iPSCs used to generate mosaic
824 organoids (d0), RNA was extracted from triplicate samples of 42k cells using the Direct-zol RNA
825 Microprep Kit (Zymo Research #R2062). To generate cDNA, 50-250ng of RNA was reverse-transcribed
826 using the SuperScript™ IV First-Strand Synthesis System (Invitrogen 18091050) along with an
827 oligo(dT)VN primer containing UMIs. Sequencing libraries were then generated using Q5® Hot Start
828 High-Fidelity 2X Master Mix (NEB #M0494L) and custom primers.
829 To generate FLICK barcode libraries from scRNA-seq libraries, 5uL of amplified and cleaned cDNA
830 (produced by following the 10X User Guide CG000316 Rev A through the end of Step 2) was used.
831 Barcode-containing cDNA was amplified for sequencing using Q5® Hot Start High-Fidelity 2X Master
832 Mix (NEB #M0494L) and custom primers.
833 Libraries were cleaned using SPRIselect DNA Size Selection Reagent (Beckman Coulter #
834 B23317) and assessed using TapeStation High Sensitivity D1000 reagents (Agilent #5067-5603 and
835 #5067-5584).
836
837 **Alignment and Quantification of Organoid scRNA-seq Data**
838 scRNA-seq data was aligned to the pre-built CellRanger (v.9.0.0) hg38 reference using *cellranger-*
839 *count* in feature-barcode mode to perform simultaneous gene expression and guide abundance
840 quantification. Introns were not included in gene expression quantification (--include-introns=false) to
841 avoid quantifying partially degraded transcripts that had been targeted by CRISPR-CasRx. CRISPR
842 guide sequencing library FASTQs for 2-month samples were downsampled to 1.25% of reads using
843 seqtk (v1.4) to match CRISPR library sequencing depth for one-month samples and to minimize
844 background for guide assignment using the cellranger-CRISPR guide quantification pipeline. All other
845 default parameters were used. CellRanger guide assignments were used to filter to only cells with
846 exactly one assigned guide.
847
848 **Pre-Processing and Quality Control of Organoid scRNA-seq Data**
849 Long-term and acute KD samples were processed separately but equivalently. CellBender (v0.3.2)
850 was used to perform detection and filtering of ambient RNA contamination, while scrublet (v0.2.3) was
851 used to perform automatic doublet detection and filtering. All other downstream analysis was performed
852 with Seurat (v5.0.0). High-quality cells were retained according to the following quality-control metrics:
853 1) percent mitochondrial (percent.mt) gene expression < 20% and 2) 500 < number of expressed genes
854 (nFeature_RNA) < 6000.

855
856 **Cell-Type Annotation of Organoid scRNA-seq Data**
857 Cell-type annotation was first performed on only non-targeting control (NTC) cells. NTC cells were
858 extracted to create a control-only Seurat object. This object subsequently underwent normalization
859 (*NormalizeData*), variable feature identification (*FindVariableFeatures*), scaling (*ScaleData*; regressed
860 covariates of percent.mt, percent_ribo, nCount_RNA, and nFeature_RNA), and dimensionality
861 reduction (*RunPCA* and *RunUMAP*). Integration was performed using harmony (v1.2.0) across
862 organoid batches and this integrated object was used for unsupervised clustering (*FindNeighbors* and
863 *FindClusters*, resolution = 0.6 and 1.2). A small number of unsupervised clusters (<1000 cells) were
864 filtered due to low cluster-wide QC metrics (i.e. exceptionally high or low nCount_RNA, high
865 percent.mt). All subsequent 2-D visualization of this control-only Seurat object was performed using a
866 UMAP representation of the integrated data. Cell type annotation was performed using canonical brain
867 marker genes. A Seurat object was next constructed for all cells, following the same processing
868 pipeline (e.g., normalization, scaling, dimensionality reduction, and integration) as described above. To
869 perform cell-type annotation, label transfer was performed from the control-only Seurat object
870 (*TransferData*) to the full Seurat object. Expression patterns of canonical brain marker genes were then
871 used to refine and finalize these cell type annotations.
872
873 **Analysis of Relative Abundance per KD in Cerebral Organoids**
874 The relative abundance of cells for each KD target was determined for each time point (d0, d30,
875 d60, or d61). For d0, relative abundance was determined from bulk RNA-seq libraries (see
876 "Construction of FLICK Barcode Sequencing Libraries") using a custom pipeline (see **Code**
877 **Availability**). For d30, d60, and d61, relative abundance was determined from scRNA-seq using the
878 CRISPR guide assignments generated by CellRanger (see "Alignment and Quantification of Organoid
879 scRNA-seq Data"). For each sample, results were normalized to the starting proportion at d0 and
880 displayed relative to NTCs from the same timepoint. Results were plotted as mean +/- SEM. Statistical
881 analysis was performed using repeated measures one-way ANOVA with Dunnett's multiple
882 comparisons test (GraphPad Prism), considering all long-term samples (d30, d60, and d61).
883
884 **Analysis of Relative Cell Type Proportions in Cerebral Organoids**
885 The relative ExN/RG proportion was determined for each KD target at d30, d60, and d61. Results
886 were normalized to the NTC sample from the same timepoint. Statistical analysis was performed using
887 ordinary one-way ANOVA with Dunnett's multiple comparisons test (GraphPad Prism), considering all
888 long-term samples (d30, d60, and d61).
889
890 **Differential Expression Analysis of Cerebral Organoid scRNA-seq Data**
891 Differential expression analysis was performed between perturbed cells per guide and cells that
892 received NTCs, stratified by major cell types (radial glia, intermediate progenitor cells, excitatory
893 neurons, or inhibitory neurons). Seurat *FindMarkers* was used to perform differential expression with
894 logfc.threshold = 0, min.pct = 0.10, min.cells.group = 3 and otherwise all default parameters.
895 Enrichment analyses of differentially expressed genes (DEGs) were performed using hypergeometric
896 tests, and p-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing.
897 The universe was defined as all genes expressed in both comparisons of interest.
898
899 **Visualization of Shared DEGs Between Organoid Cell Types**
900 Shared DEGs were visualized between organoid cell types (**Fig. 3c**) using the R package igraph
901 (v2.0.3). The Fruchterman-Reingold layout algorithm (*layout_with_fr*) was used to determine final node
902 placement.
903
904 **MAGIC Analyses**
905 MAGIC[52] (v1.1) was used to identify transcription factors and cofactors whose targets are enriched
906 in the DEGs from individual perturbations. The strategy of "gene body (promoter to the end of the last
907 exon) plus 1Kb flanking sequence either side of the gene body" was used. For NPC analyses, the
908 universe was set as all genes with baseMean > 10 in NTC samples. For organoid analyses, the

universe was set as all genes expressed in at least 10% of cells of that cell type (Seurat min.pct = 0.10; no log2 fold-change cutoff), as these were the genes assessed for differential expression.

Re-Analysis of REST ChIP-Seq Data

Data was extracted from Table S2 from Rockowitz and Zheng 2015. Mouse genes were mapped to human genes using Orthology.eg.db (v3.20.0), org.Mm.eg.db (v3.20.0), and org.Hs.eg.db (v3.20.0). Enrichment analysis was performed using a hypergeometric test with all BrainSpan genes with normalized expression >= 1 as the universe (version "RNA-Seq Gencode v10 summarized to genes"; https://brainspan.org; accessed 1/6/24). P-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing. Results were displayed as bar plots, with main bars indicating odds ratio and error bars indicating lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown).

Gene Regulatory Network (GRN) Construction and Benchmarking

A GRN was constructed using the bioinformatic tool ARACNe-AP[57]. The input was the RNA-seq data from our 3 NTCs and the successful KD samples with at least 50 DEGs. We used curated lists of transcription factors[104] and epigenetic regulators[59] and selected genes expressed in NPCs (baseMean >10 in NTC samples) as the candidate genetic regulators. To create the GRN, we used default parameters and 100 bootstraps that were then aggregated together to form the final network. We then used VIPER[62] (v1.36.0) to identify differentially active regulons following standard operating procedures. In brief, we identified regulons with significant differential expression between NTCs and KDs using a t-test, with a null distribution defined by sample permutations.

To benchmark the predictive power of the ARACNe-derived GRNs, we generated additional ARACNe networks after individually holding out perturbation samples for five target genes that were identified as regulators of at least 250 genes (range: 478-1219 genes) in their 2nd degree regulons (i.e., their direct targets, which constitute their 1st degree regulon, as well as the direct targets of those genes). For each held-out regulator, we assessed overlap using a hypergeometric test between 1) the predicted positively regulated genes within its 2nd degree regulon, and 2) the set of downregulated DEGs resulting from perturbation of that same regulator. P-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing. P-values across the five held-out perturbations were combined using Fisher's method.

PageRank Analysis of GRN

To identify central regulators within our GRN, we converted the GRN to a graph using the R package igraph (v2.1.4) and ran the PageRank algorithm[60] using the igraph method *page_rank*. We then rank-ordered genes by their PageRank score and ran gene set enrichment analysis (GSEA)[105] using the R package fgsea (v1.32.4), testing the following ontologies: SFARI genes, HC-SFARI genes, ID genes (exclusive of SFARI genes), and the 50 Molecular Signature Database Hallmark Pathways[61] (obtained from https://www.gsea-msigdb.org/gsea/msigdb).

Identification of Top Candidate Regulators for Driving Convergence in Autism

VIPER[62] (v1.36.0) was used to identify GRN regulators whose regulons were significantly differentially active across the KDs compared to NTC controls. From these, we prioritized 70 regulators whose regulons were significantly enriched for both SFARI genes (hypergeometric test with Benjamini-Hochberg correction for multiple hypothesis testing) and high PageRank scores (gene set enrichment analysis with Benjamini-Hochberg correction for multiple hypothesis testing).

Additional GRN Analyses

Mean normalized degree for the indicated set of regulators and their 1st degree regulons (**Fig. 5d**) was calculated using the R package igraph (v2.1.4) by running the function *mean(degree(graph, normalized = TRUE))*. We then created an empirical null distribution of mean normalized degrees for each indicated set of regulators by creating 1M subnetworks of the same set size of randomly selected GRN regulators. Relative mean normalized degree was calculated as the observed mean normalized degree for a given set of regulators divided by the mean of the values for the corresponding empirical

null distribution. A one-sided empirical p-value was then calculated as the proportion of null values greater than or equal to the observed mean normalized degree, with a detection limit of p = 1e-6.

To evaluate known or predicted protein-protein interactions between the top 70 regulators, we utilized the STRING database GUI[63]. For the universe, we used the full set genes that were considered as candidate genetic regulators in the GRN: transcription factors[104] and epigenetic regulators[59] that were expressed in NPCs (baseMean >10 in NTC samples). All other default parameters were used.

For clustering of our 70 regulators into modules, we used iterative walktrap clustering using the R package igraph (v2.0.3) by running the function *cluster_walktrap* with default parameters.

Re-Analysis of CHD8 ChIP-Seq Data

For enrichment analysis of CHD8 targets, we used Supplementary Data 1 from Sugathan et al., 2014[64]. Enrichment of the indicated gene sets was tested using hypergeometric tests, with the universe set to NPC-expressed genes (baseMean > 10 in NTCs). P-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing. Results were displayed as a bar plot, with main bars indicating odds ratio and error bars indicating lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown).

Identification of Female-Biased Genes

To identify genes that exhibit female-biased expression in the neurotypical brain, we combined DEG call sets from the following works:

(1) Table S4 from Kissel et al., 2024[68]: Identified as significantly female-biased in either the UCLA cohort, BrainVar cohort, or via meta-analysis.
(2) Oliva et al., 2020[69] (https://storage.cloud.google.com/adult-gtex/bulk-qtl/v8/sb-eqtl/GTEx_Analysis_v8_sbgenes.tar.gz): Identified as significantly female-biased in either BRNCTXA or BRNCTXB.
(3) Table S4 from Naqvi et al., 2019[70]: Identified as significantly female-biased in the brains of all assessed species (human, mice, rats, dogs, and macaques).

Re-Analysis of ZFX Targets from Rhie et al., 2018 and San Roman et al., 2023

ZFX ChIP-seq targets were obtained from Table S3E from Rhie et al. 2018. ZFX targets were also obtained from San Roman et al. 2024, which previously performed integrative analysis of ZFX ChIP-seq data (ENCODE) and *ZFX* KD RNA-seq data from Rhie et al. 2018. Enrichment was analyzed using hypergeometric tests and p-values were adjusted using Benjamini-Hochberg correction for multiple hypothesis testing. For analysis of common ZFX targets from Rhie et al. 2018, the universe was defined as genes expressed in both of the cell lines used for ChIP-seq for which RNA-seq was also available (C4-2B and MCF-7). For analysis of ZFX targets from San Roman et al. 2024, the universe was defined as genes expressed in the corresponding cell type from Rhie et al. 2018. Results were displayed as a bar plot, with main bars indicating odds ratio and error bars indicating lower 95% confidence interval for one-sided Fisher's test (upper bound of infinity not shown).

Data Availability

All raw data has been deposited to GEO (GSE305954). GEO data can currently be accessed anonymously by reviewers using the information provided to the editor. Please note that this data is currently embargoed and should only be accessed by reviewers for the purpose of manuscript review. This data should not be distributed or used for other purposes until the embargo is lifted on the date of publication, at which point all GEO data will be publicly available.

Code Availability

Code has been provided to reviewers as a zip file (REA_MT_CAW_code.zip) and will be publicly available via Github on the date of publication.

1013 **ACKNOWLEDGEMENTS**

1014 We thank members of the Walsh Lab for helpful discussions. We are also grateful to Dr. Norma K.
1015 Hylton and Evan M. Bushinsky of the Walsh Lab for assistance with cerebral organoid culture, Dr.
1016 August Yue Huang for technical feedback regarding statistical analyses, and Dr. Adrianna K. San
1017 Roman and Dr. Marla E Tharp of the Page Lab for their insights regarding *ZFX*. This work was
1018 facilitated by the BCH Flow Cytometry Research Core (funded by 5U54DK110805), the BCH IDDRC
1019 Molecular Genetics Core Facility (funded by NIH P50 HD105351), the BCH Cellular Imaging Core
1020 (funded by NIH P50 HD105351), and Lingsheng Dong of the HMS Research Computing Core. This
1021 work was supported by Autism Speaks Postdoctoral Fellowship 13008 and NIGMS T32 GM007748
1022 (R.E.A.); NIH T32 GM007753 and T32 GM144273 (M.T.); the Simons Foundation Autism Research
1023 Initiative, the Howard Hughes Medical Institute, The Brit Jepson d'Arbeloff Center on Women's Health,
1024 Arthur W. and Carol Tobin Brill, Matthew Brill, Charles Ellis, Carla Knobloch, The Brett Barakett
1025 Foundation, the Howard P. Colhoun Family Foundation, and the Seedlings Foundation (M.T. and
1026 D.C.P.); the Harvard Program for Research in Science and Engineering and the Harvard College
1027 Research Program (T.S.); a Howard Hughes Medical Institute Fellowship from the Helen Hay Whitney
1028 Foundation and Grant K99MH136290 from NIMH/NIH (J.H.T.S); K99 NS135123 from the NINDS/NIH
1029 (X.Q.); the Howard Hughes Medical Institute and Grant K99EY034603 from NEI/NIH (R.N.D); the
1030 Harvard-Amgen Scholars Program (G.E.); and grants from the Simons Foundation and Autism
1031 BrainNet (953759), NIMH (U01MH106883), the Hock E. Tan and K. Lisa Yang Center for Autism
1032 Research at Harvard University, and the Allen Discovery Center program, a Paul G. Allen Frontiers
1033 Group advised program of the Paul G. Allen Family Foundation (C.A.W). D.C.P. and C.A.W. are
1034 Investigators of the Howard Hughes Medical Institute.

1035 **AUTHOR CONTRIBUTIONS**

1036 R.E.A and M.T. conceived the experiments and led the analyses, with supervision from C.A.W. M.T. led
1037 all bioinformatic analyses. R.E.A. designed all plasmid constructs with support from T.S. and S.L.
1038 Molecular cloning and lentivirus production were performed by R.E.A., T.S., and S.L. The CasRx
1039 knock-in iPSC line was generated by J.H.T.S. Cell culture was performed by R.E.A., T.S., S.L., and
1040 N.M., with contributions from S.Z., G.E., and J.H. Bulk RNA-seq libraries were prepared by R.E.A. and
1041 S.L. Analysis of bulk RNA-seq results was performed by M.T., with code contributions from R.N.D. NPC
1042 proliferation assays were performed by R.E.A., S.L., S.Z., G.E., and J.H. Cerebral organoids were
1043 cultured by R.E.A., T.S., and N.M., with advice and support from X.Q. Cerebral organoid scRNA-seq
1044 libraries were prepared by R.E.A. and N.M. Analysis of scRNA-seq libraries was performed by M.T.
1045 Cryosectioning and immunohistochemical labeling of cerebral organoids was performed by R.E.A. and
1046 N.M. Confocal imaging and histological analyses of cerebral organoids were performed by D.E.-A.
1047 D.C.P. contributed to the interpretation of *ZFX* as a contributor to the female protective effect in ASD.
1048 R.E.A., M.T, and C.A.W. wrote the manuscript with contributions from all of the authors. All of the
1049 authors read and approved the final manuscript.

1050

1051 **COMPETING INTERESTS**

1052 C.A.W. consults for Maze Therapeutics, Mosaica Medicines, Bioskryb Genomics (cash, equity). The
1053 remaining authors declare no competing interests.

1054

1055 **SUPPLEMENTARY INFORMATION**

1056 Supplementary Information is available for this paper.

1057

1058 **MATERIALS & CORRESPONDENCE**

1059 Correspondence and requests for materials should be addressed to Christopher A. Walsh,
1060 christopher.walsh@childrens.harvard.edu.

1061 **SUPPLEMENTARY INFORMATION**

1062 **Supplementary Table 1.** Guide RNA (gRNA) sequences for use with CasRx.

1063 **Supplementary Table 2.** Data associated with **Extended Data Fig. 1a**.

1064 **Supplementary Table 3.** Data associated with **Fig. 1c**.

1065 **Supplementary Table 4.** Median LOEUF scores used for assessing constraint. Highly constrained
1066 genes were defined as having a median LOEUF score of less than 0.35.

1067 **Supplementary Table 5.** Genes that meet the criteria used for identifying candidate ASD risk genes: 1)
1068 statistically significant downregulation by at least three of the ASD KDs, with at least 80% concordance
1069 in the direction of significant differential expression across all KDs (consistently downregulated DEGs);
1070 and 2) intolerance to loss of function variants (highly constrained genes).

1071 **Supplementary Table 6.** Data associated with NPC proliferation assays depicted in **Fig. 1f**.

1072 **Supplementary Table 7.** Top significant MAGIC regulators identified in different cell populations. For
1073 each population, MAGIC regulators were scored by the number of ASD gene KDs that implicated the
1074 regulator. The 25 top scoring regulators are included here, along with any additional regulators with a
1075 tied score.

1076 **Supplementary Table 8.** Top drivers of convergence in ASD, as identified through GRN analysis.

1077 **Supplementary Table 9.** Oligo sequences used to generate libraries for bulk RNA-sequencing.

1078 **Supplementary Table 10.** Primers for ddPCR/qPCR.