An integrated framework for functional dissection of variable expressivity in genetic disorders

Jiawan Sun[1]*, Serena Noss[1]*, Deepro Banerjee[1], Venkata Hemanjani Bhavana[1], Corrine Smolen[1], Maitreya Das[1], Belinda Giardine[1], Anisha Prabhu[1], David J Amor[2,3], Kate Pope[2], Paul J Lockhart[2,3], and Santhosh Girirajan[1,4].

1. Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802
2. Department of Paediatrics, University of Melbourne, Melbourne, Australia
3. Bruce Lefroy Centre, Murdoch Children's Research Institute, Melbourne, Australia
4. Huck Institutes of the Life Sciences, University Park, PA 16802

Correspondence:
Santhosh Girirajan
205A Huck Life Sciences Building
Pennsylvania State University
University Park, PA 16802
E-mail: sxg47@psu.edu

ABSTRACT

Disease-associated variants can lead to variable phenotypic outcomes, but the biological mechanisms underlying this variability remain poorly understood. We developed a framework to investigate this phenomenon using the 16p12.1 deletion as a paradigm of variable expressivity. Using induced pluripotent stem cell models from affected families and CRISPR-edited lines with the 16p12.1 deletion, we found that the deletion and secondary variants in the genetic background jointly influenced chromatin accessibility and expression of neurodevelopmental genes. Cellular analyses identified family-specific phenotypes, including altered inhibitory neuron production and neural progenitor cell proliferation, which correlated with head-size variation. CRISPR activation of individual 16p12.1 genes variably rescued these defects by modulating key signaling pathways such as TGF-β and PI3K-AKT. Integrative analyses further identified regulatory hubs, including transcription factors FOXG1 and JUN, as mediators of these effects. Our study provides a functional framework for investigating how individual genetic architectures contribute to phenotypic variability in genetic disorders.

INTRODUCTION

In contrast to Mendelian disorders with straightforward genotype to phenotype relationships, variants associated with complex disorders often lead to variable clinical features[1,2]. This phenomenon of variable expressivity is particularly true for rare copy number variants (CNVs), such as deletions and duplications within chromosomal regions 1q21.1, 15q13.3, and 22q11.2, which have been associated with variable neurodevelopmental outcomes and have also been identified in population controls[3]. For example, the 22q11.2 deletion, classically known to cause DiGeorge syndrome, is associated with intellectual disability and developmental delay (ID/DD), autism, and congenital heart defects, and it accounts for about 1% of individuals with schizophrenia[4,5]. The approximately 500-kbp 16p12.1 deletion, encompassing eight genes, represents an ideal model to study variable expressivity of disease-associated variants for several reasons. *First*, the deletion is associated with a wide range of phenotypic outcomes, including ID/DD, autism, congenital anomalies, and epilepsy in affected children, as well as psychiatric features such as schizophrenia, depression, and anxiety in adolescents and adults, with varying levels of severity[6–12]. *Second*, unlike syndromic CNVs such as 17p11.2 deletion in Smith-Magenis syndrome or 5q35 deletion in Sotos syndrome[3], that mostly occur *de novo*, the 16p12.1 deletion is inherited in over 90% of affected individuals from parents with milder cognitive or neuropsychiatric features[13]. The high inheritance rate allows us to assess phenotypic outcomes in multiple carriers from the same family. *Third*, specific patterns of rare variants in the genetic background ("secondary variants") correlated with distinct phenotypic trajectories associated with the deletion[14]. In fact, modifier roles of secondary variants have also been reported as contributing to phenotypic variability in both monogenic and complex neurodevelopmental disorders, including in individuals affected by 7q11.23 duplication and 16p11.2 duplication [2,3,15,16]. However, the molecular basis by which these CNVs confer disease susceptibility and interact with secondary variants to produce variable outcomes remains poorly understood.

Previous studies have shown the utility of induced pluripotent stem cells (iPSC) to recapitulate complex biological processes *in vitro*, enabling investigations into the molecular etiology of neurodevelopmental disorders[17–19]. However, these studies have typically focused on aggregate analysis of subjects with the same primary variant, such as the 16p11.2 and 22q11.2 deletion, or the same phenotype, such as autism and schizophrenia, without investigating the variability among these subjects[20–22]. Thus, the effects of disease-associated variants within the

78 context of an individual's genetic background remain unexplored. Here, we developed a
79 framework to dissect the mechanistic basis of variable expressivity in iPSC models of the
80 16p12.1 deletion **(Figure 1)**. By integrating multi-omics profiling, CRISPR/dCas9-mediated
81 gene activation, and family-based analyses, we demonstrate that phenotypic trajectories
82 associated with 16p12.1 deletion are shaped by the combinatorial effects of the deletion and
83 secondary variants through the modulation of key signaling pathways. We further identified
84 regulatory hubs that mediate these effects, including transcription factors (TFs) FOXG1 and
85 JUN, along with genes in the regulatory network. We propose a conceptual shift from a focus on
86 single causal variants to a system-level understanding of individual genetic architecture,
87 advancing more effective precision medicine strategies.
88

RESULTS

Generation and characterization of iPSC models for 16p12.1 deletion

To model the effects of the deletion on neuronal development in an isogenic setting, we created a 465-kbp 16p12.1 deletion in the HD_01 line, derived from a healthy donor, using a CRISPR/Cas-9-based strategy **(Supplementary Figure 1A-E)**. To study the effects of the 16p12.1 deletion under different family-specific genetic backgrounds, we reprogrammed iPSC lines from peripheral blood mononuclear cells derived from 12 individuals from three families **(Figure 1A-B).** As controls, we used iPSC lines (HD_01 and HD_02; obtained from NINDS repository) from unrelated healthy male and female donors, as well as a HD_01 line transfected with empty vector as a CRISPR control. These 16 iPSC lines were then differentiated to neural progenitor cells (NPC) and immature (iMN) and mature neurons (MN) using dual-SMAD inhibition[18] **(**see **Methods, Figure 1B)**. RNA sequencing was performed at all four cell stages, while ATAC-seq was performed for iPSC and NPC lines (**Figure 1A**). We confirmed reduced expression of 16p12.1 deletion genes including *UQCRC2, POLR3E, MOSMO*, and *CDR2*, as well as reduced ATAC-seq signals across the region **(Supplementary Figure 2A-B, Supplementary Table 1-2).** The expression of expected gene markers were robust for each differentiation cell state, such as expression of *SOX2*, *POU5F1*, and *FUT4* for iPSC; *SOX2*, *HES1*, *NES*, and *EMX2* for NPC; and *MAP2*, *DLG4*, *NCAM1*, and *RBFOX3* for neurons **(Supplementary Figure 2C)**, and ATAC-seq peak revealed differences in chromatin accessibility near cell-type-specific genetic markers such as *POU5F1*, *SOX2,* and *HES1*, confirming the expected cell identities across the lines **(Supplementary Figure 2D).**

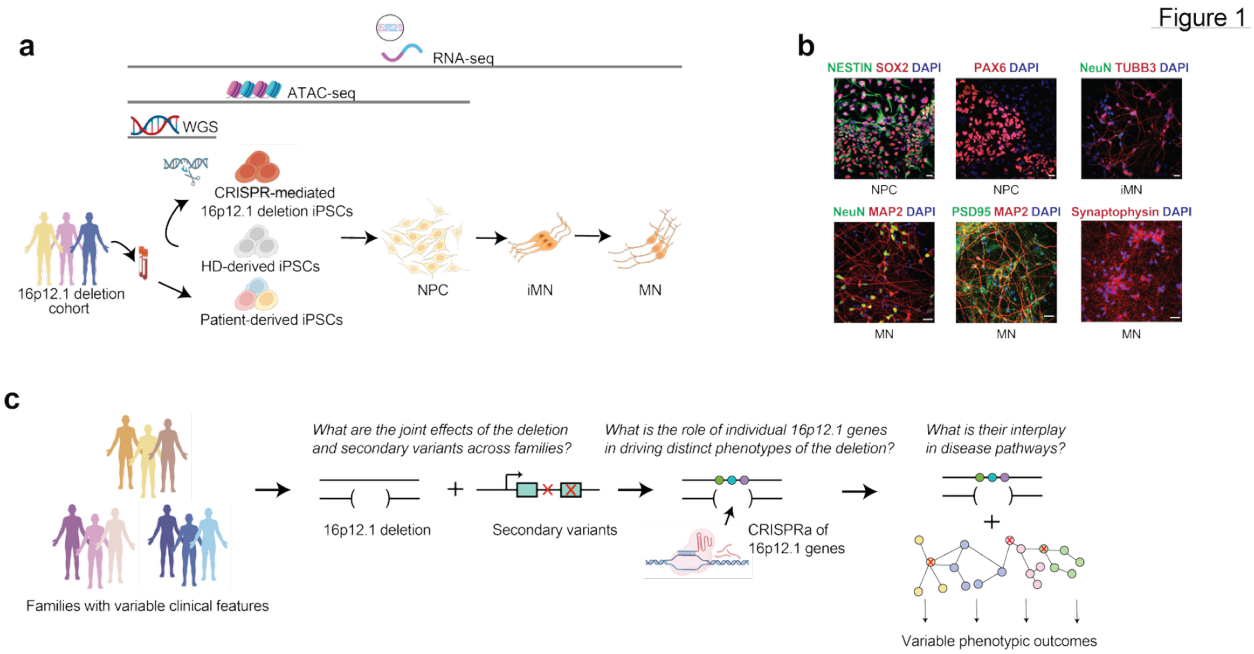


Figure 1. Overview of the integrated framework using 16p12.1 deletion iPSC models. (**a**) Experimental design. WGS was performed on DNA isolated from blood or iPSCs derived from healthy donors and individuals in the three families. RNA-seq was performed across all four differentiation stages and ATAC-seq was performed at iPSC and NPC stages. (**b**) Representative images of neural-converted cells stained with different markers for distinct cell stages (scale bars, 20 µm). **(c)** Overview of experimental strategy used to investigate how the 16p12.1 deletion and secondary variants contribute to phenotypic variability. The core questions driving the experiments within the framework are also shown.

Combined effects on gene expression related to neurodevelopment

We compared all deletion lines to nondeletion lines at each cell stage and found that the most significantly differentially expressed genes (DEGs) were those within the deleted region (**Supplementary Figure 2A, Supplementary Table 1**). This suggested that a combined analysis may obscure the downstream effects of the deletion due to variability between samples, likely due to differences in their genetic background. To isolate the direct impact of the deletion, we compared the CRISPR-edited 16p12.1 deletion line to its isogenic control. This approach is commonly used in iPSC models of disease[23] to control for the effects of the genetic background. DEGs identified in NPC, iMN, and MN were enriched for neurodevelopmental and psychiatric disorder genes from the DisGeNET and published datasets[24–29] **(Supplementary Figure 1F-G)**. To assess the extent to which the effects of genetic background on DEGs were controlled in the isogenic setting, we performed whole genome sequencing and identified 6,659 rare variants

133 (referred to as *secondary variants*) in the HD_01 line (see **Methods**). We classified these
134 secondary variants into distinct classes, including CNVs, simple tandem repeat (STR)
135 expansions, and SNVs within exonic, intronic, upstream, downstream, 5', and 3' UTR regions of
136 genes, as previously described[14,30], and found that 2%, 31%, 21%, and 10% of secondary
137 variants overlapped DEGs identified in iPSC, NPC, iMN, and MN, respectively **(Figure 2A,**
138 **Supplementary Table 1)**. Rare noncoding intronic variants were also significantly associated
139 with altered usage of isoforms in genes without expression changes (**Figure 2A, Supplementary**
140 **Table 1)**. Over-representation analysis of genes and isoforms with secondary variants revealed
141 significant enrichment for neurodevelopment, neuronal function, and signaling pathways such as
142 axon guidance, synaptic signaling, and PI3K-AKT signaling pathway (**Supplementary Figure**
143 **1H, Supplementary Table 1)**. These results suggest that secondary variants, though individually
144 rare but collectively prevalent in healthy populations[31], can alter gene expression when combined
145 with the 16p12.1 deletion[32]. This observation also supports the idea that secondary variants may
146 exert latent effects that become unmasked in the context of the deletion[33,34].
147 Given the significant overlap between rare noncoding variants and DEGs (**Figure 2A**),
148 we evaluated their potential to alter chromatin accessibility. We identified six and 130 rare
149 noncoding variants that intersected with differentially accessible chromatin (i.e., differential
150 peaks or "Diffpeak") regions from ATAC-seq analysis of iPSCs and NPCs, respectively **(Figure**
151 **2B, Supplementary Table 2).** Motif enrichment analysis of these intersecting Diffpeak regions
152 revealed significant enrichment for binding motifs of TFs such as FOXA1, MYB, and TCF4,
153 which are involved in signaling pathways such as TGF-β, Wnt, and PI3K-AKT that are relevant
154 to neurodevelopment[35–38] (**Figure 2C, Supplementary Table 2)**. We also observed differential
155 expression of target genes affected by these noncoding secondary variants, such as *WWC1* and
156 *TRIP13*, which are likely involved in a broader regulatory network of those TFs (**Figure 2C**).
157 We next tested the combined effects of the 16p12.1 deletion and the secondary variants
158 using iPSC models derived from families with variable clinical features **(Figure 1C).** The DEGs
159 identified in deletion lines compared to nondeletion lines from these families were enriched for
160 genes associated with neurodevelopmental and psychiatric disorders; however, the enrichment
161 patterns varied by family (**Supplementary Figure 3A-B, Supplementary Table 8)**. We
162 analyzed each family separately and found that genes with secondary variants, including
163 *MAPK10* and *SLC35A2*, showed significant changes in chromatin accessibility and gene

expression when both the deletion and secondary variant were present, suggesting potential nonadditive effects. (**Figure 2C, Supplementary Figure 3C Supplementary Table 3).** This phenomenon has been reported previously, where nonadditive functional effects were observed among multiple schizophrenia-associated common variants in iPSC models[39,40]. Overall, our findings support a modifier role for secondary variants, which act in combination with the deletion, to shape gene expression patterns within neurodevelopmental pathways.

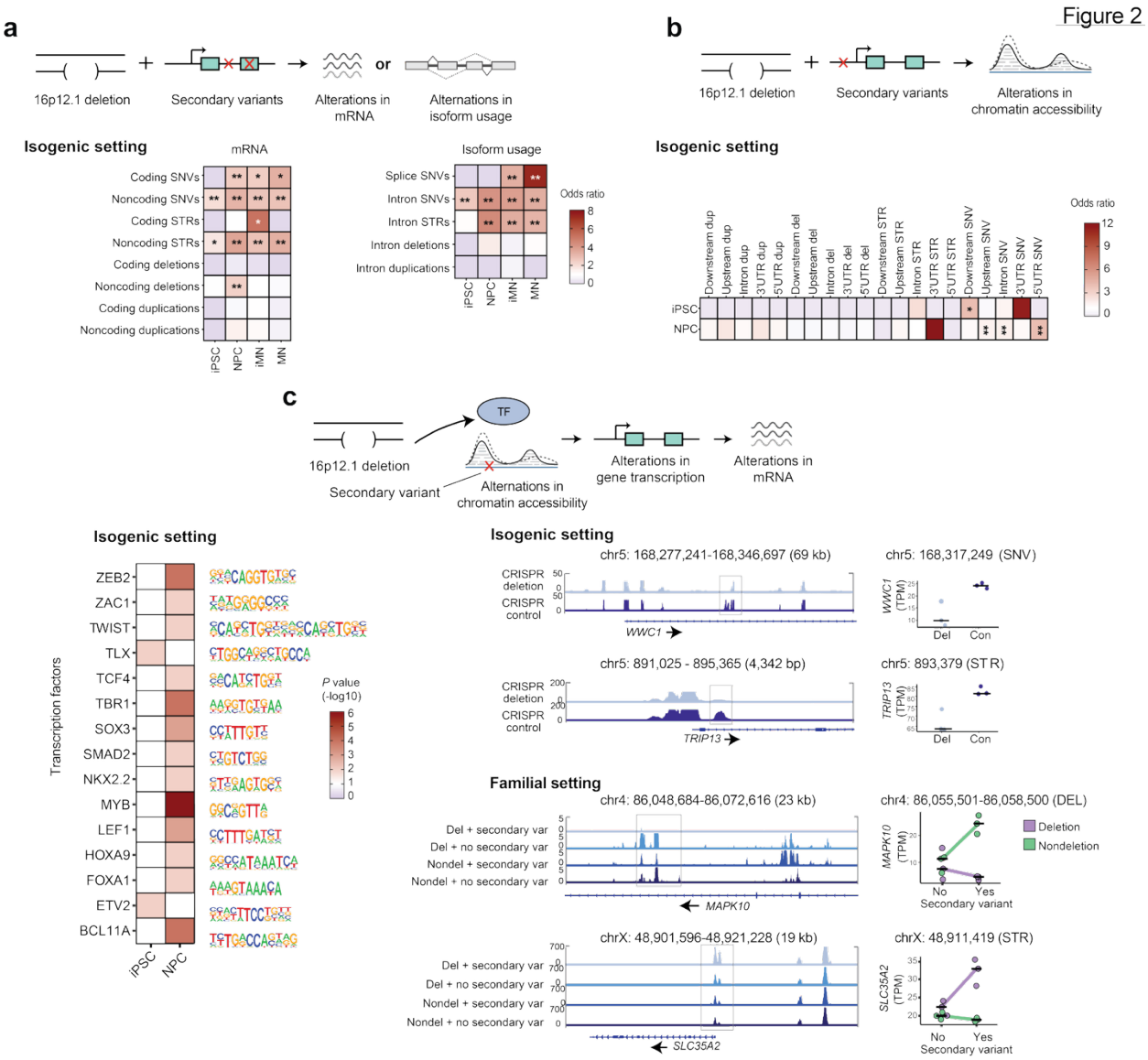


Figure 2. **Evaluating the impact of secondary variants on chromatin accessibility and gene expression in the context of 16p12.1 deletion**. (**a**) Overlap between secondary variants and DEGs (padj<0.05) or isoforms with altered usages (q<0.05) in the isogenic setting. Color

175 represents odds ratio from Fisher's exact test. *p<0.05, **Benjamini-Hochberg FDR<0.05,
176 Fisher's exact test. (**b**) Overlap between secondary variants and differential peaks obtained from
177 ATAC-seq (padj<0.01) in the isogenic setting. Color represents odds ratio from Fisher's exact
178 test. *p<0.05, **Benjamini-Hochberg FDR<0.05, Fisher's exact test. **(c)** Representative motifs
179 (left) of transcription factor binding sites enriched within ATAC-seq differential peak regions that
180 intersect with secondary variants. Examples (right) of altered chromatin accessibility peaks that
181 intersect with secondary variants as well as associated differential expression of the
182 corresponding gene in the isogenic and familial settings (transcripts per million, TPM, with
183 median values) at the NPC stages. ATAC-seq peaks were visualized using Integrative Genome
184 Viewer (IGV). From the DESeq2 results, the adjusted p-value (padj) for *WWC1* was 0.00041,
185 and for *TRIP13*, padj was 0.018 **(Supplementary Table 8)**. Con, CRISPR control; Del, CRISPR
186 deletion; var, variant; DEL, deletion; SNV, single nucleotide variant; STR, short tandem repeat;
187 Del, 16p12.1 deletion lines; Nondel, nondeletion lines. Data in the familial setting were derived
188 from the GL_079 samples **(Supplementary Table 3).** For both RNA-seq and ATAC-seq, we
189 used n=3 independent experiments.

190

191 **Variable neuronal phenotypes across genetic backgrounds**

192 Beyond transcriptomic changes, we also evaluated the combined effects of the deletion and
193 secondary variants on neuronal phenotypes. At the NPC stage, we found no differences in
194 NESTIN and SOX2 levels between the deletion and nondeletion lines. However, we observed a
195 decrease in PAX6 (a dorsal telencephalic marker) (**Supplementary Figure 4 A-B)** and a highly
196 variable increase in NKX2.1 (a ventral telencephalic marker) in the deletion lines compared to
197 nondeletion lines **(Figure 3A),** suggesting altered neuronal lineage commitment during
198 differentiation**.** To further investigate this variability, we performed family-specific analysis
199 comparing the deletion lines to HD lines and, when available, to nondeletion lines from the same
200 family. We found a marked increase in NKX2.1-positive cells specifically in GL_007 and the
201 CRISPR deletion lines **(Figure 3A-B)**. These results are consistent with previous findings in
202 other iPSC models of neurodevelopmental disorders[41]. In addition to NKX2.1, which is critical
203 for the development of inhibitory neurons, we examined VGAT, the vesicular GABA transporter,
204 in mature neurons expressing NeuN and MAP2 (**Supplementary Figure 4C)** and found
205 opposing trends across families, with a substantial increase in VGAT intensity in the deletion
206 lines of GL_007 and a variable decrease in the deletion lines of GL_077 and GL_079 compared
207 to controls **(Figure 3C-D)**. We also assessed other genes involved in inhibitory neuron
208 differentiation and migration, including *DLX1*, *DLX2*, *LHX6* and *LHX8*, and found them to be
209 highly expressed across NPCs, iMNs, and MNs in both the CRISPR deletion and GL_007
210 deletion lines compared to controls (**Supplementary Figure 4D, Supplementary Tables 8)**.

211 Although NKX2.1 level was elevated in the CRISPR deletion line compared to its isogenic
212 control, the absence of a significant difference in VGAT intensity suggests that this alteration
213 may not persist into the MN stage or may be too subtle to be detected with this assay.
214 Additionally, glutamatergic neuron production did not appear to be affected, as expression levels
215 of glutamatergic markers (*SLC17A6*, *SLC17A7*, *GRIN1*, *GRIN2B*, and *GLS*) were not
216 consistently altered across the deletion lines, and mature neurons showed no overt change in
217 VGLUT1 expression **(Supplementary Figure 4C)**. These results suggest that the deletion
218 contributes to altered production of inhibitory neurons under specific genetic backgrounds.
219 Aberrant neuronal proliferation and apoptosis, which alter the proportions of distinct
220 brain cell types, represent another convergent phenotype in neurodevelopmental disorders[42]. We
221 evaluated NPC proliferation by measuring EdU-labeled cells at 0, 24, and 48 hours, as well as
222 Ki-67-positive cells, and found no differences in proliferation rates between the deletion and
223 nondeletion lines (**Supplementary Figure 5A-B)**. Similarly, TUNEL assays revealed no
224 differences in apoptosis between the two groups (**Supplementary Figure 5C**). However, when
225 grouped by family, we observed opposite trends in the deletion lines from GL_079 and GL_007
226 compared to controls: GL_079 deletion lines showed increased proliferation, while GL_007
227 deletion lines showed decreased proliferation and increased apoptosis compared to HD controls
228 (**Figure 3E-F, Supplementary Figure 5D-G)**. GL_079 deletion lines also showed decreased
229 apoptosis, but only when compared to nondeletion lines from the same family **(Supplementary**
230 **Figure 5F)**. Notably, these results correspond with the head-size phenotypes in P2C_079 and
231 FNC_079 from family GL_079 (macrocephaly) and P1C_007 from family GL_007
232 (microcephaly).
233 We also found evidence of altered developmental timing in deletion lines, a phenomenon
234 also reported in iPSC models of autism and schizophrenia[43,44]. A significant increase in TUBB3-
235 positive cells (a marker for newborn neurons) at the NPC stage was detected in the CRISPR
236 deletion and GL_007 deletion lines compared to controls **(Supplementary Figure 5H-J).** To
237 further investigate this trend, we performed weighted gene co-expression network analysis
238 (WGCNA) on RNA-seq data from NPC, iMN, and MN stages to identify gene modules
239 correlated with specific stages of differentiation and examine changes in these modules within
240 each family (**Supplementary Figure 6A, Supplementary Table 7**). These gene modules were
241 enriched for differentiation stage-specific Gene Ontology (GO) terms such as cell division and

242 chemical synaptic transmission (**Supplementary Figure 6B)**. The CRISPR deletion and GL_007
243 deletion lines exhibited significantly lower eigengene scores for the NPC-associated module
244 (*blue*) at the NPC stage and significantly higher eigengene scores for the mature neuron-
245 associated module (*turquoise*) at both the NPC and iMN stages **(Supplementary Figure 6C)**.
246 These results suggest premature differentiation of NPCs. Furthermore, gene set enrichment
247 analysis (GSEA) comparing deletion lines to nondeletion lines from the isogenic setting and each
248 family across four differentiation stages revealed significant enrichment for GABAergic neuron
249 and DNA replication-related terms, as well as for signaling pathways associated with PI3K-AKT,
250 TGF-β, and Wnt **(Figure 3G, Supplementary Figure 6D, Supplementary Table 8).** Together,
251 these findings suggest that the combined effect of the deletion and secondary variants lead to
252 family-specific alterations in neurogenesis and associated signaling pathways.
253

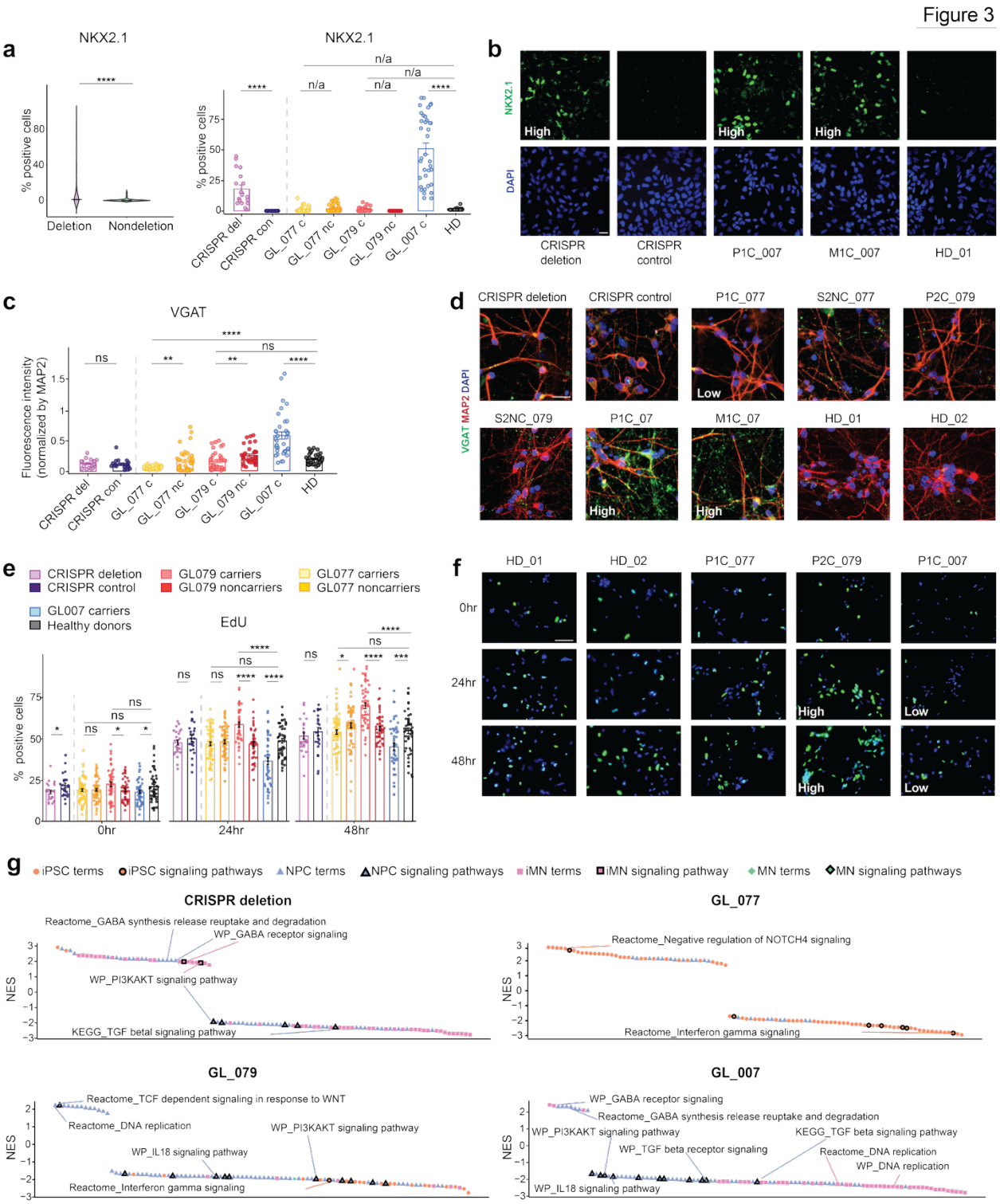


Figure 3. Assessing neuronal phenotypes across genetic backgrounds. (a) Violin plot (left) with median values shows the percentage of NKX2.1-positive cells at the NPC stage in all deletion lines compared to all nondeletion lines. Bar plot (right) with mean values shows the percentage of NKX2.1-positive cells within family-specific lines. **(b)** Representative images of

NPCs stained for NKX2.1. Scale bars, 20 µm. **(c)** Bar plot with mean values shows VGAT fluorescence intensity normalized by MAP2 fluorescence intensity in each image. **(d)** Representative images of neurons stained for VGAT and MAP2. Scale bars, 20 µm. **(e)** Bar plot with mean values shows the percentage of Edu-positive cells in specific groups. **(f)** Representative images of EdU-labeled NPCs across three timepoints. Scale bars, 70 µm. **(g)** Waterfall plot of unbiasedly selected top pathways from each comparison of deletion versus nondeletion samples using GSEA ($q<0.05$) annotated in the Reactome, KEGG and WikiPathway (WP) (see **Methods**). NES, normalized enrichment score; isogenic, CRISPR deletion versus CRISPR control. Comparisons involve the grouped deletion carriers of each family versus the two healthy donor lines. The bold border indicates terms that belong to signaling pathways. Data are mean±s.e.m and two-tailed Welch's t-tests for **(a), (c), (e)**. *$p<0.05$; **$p<0.01$; ***$p<0.001$; ****$p<0.0001$. n=3 independent experiments for each line, except for RNA-seq and downstream analyses where n=2 for iMNs and MNs of P1C_077, and MNs of HD_01. "n/a" indicates that Welch's t-tests were not performed due to high zero inflation. For **(a)** and **(c)**, n=6 images per experiment and for **(e)** n=8 images per experiment were quantified. For **(a)** and **(e)**, for GL_079, GL_077, GL_007, "c" contains all deletion lines and "nc" contains all nondeletion lines, for their respective families (see Figure 1C). For **(c)**, the same cell lines were used, excluding GMC_077 and GFNC_077 lines. In the family-based analyses, we compared deletion lines from GL_077 and GL_079 to both nondeletion lines within the same family and HD lines. For GL_007, due to the absence of nondeletion family lines, we compared the deletion lines to HD lines.

CRISPR activation of 16p12.1 genes reverses transcriptomic and neuronal defects

We previously found that the 16p12.1 genes exhibit diverse molecular functions and show limited connectivity within a brain-specific co-expression network, especially when compared to genes within the more penetrant 16p11.2 deletion[13,45,46]. Therefore, we sought to investigate the role of individual 16p12.1 genes in driving distinct functional changes across genetic backgrounds. We designed single-guide RNAs (sgRNAs) targeted to the promoters of *MOSMO*, *POLR3E*, and *UQCRC2* and used CRISPR activation (dCas9-VP64 and MS2-P65-HSF1-mediated CRISPRa[47]**)** to increase their expression in the deletion lines **(**see **Methods, Figure 4A, Supplementary Figure 7A)**. We selected these genes as they showed significantly reduced expression in the deletion samples across neuronal differentiation cell states **(Supplementary Figure 2A)**. Previous studies have also implicated these genes in cellular functions such as Sonic hedgehog (Shh) signaling, immune response, and mitochondrial homeostasis[48–50]. We performed CRISPRa experiments at the iPSC stage in three proband lines to assess the contribution of individual 16p12.1 genes to their more severe and varied cellular phenotypes and confirmed sustained overexpression of the targeted genes in NPCs, the cell state in which most cellular effects were observed in this study **(Supplementary Figure 7A)**. However, neurosphere

298 formation failed for CRISPRa of *MOSMO* in P2C_079. This could be potentially due to a
299 dosage-sensitive role of *MOSMO* in the genetic background of P2C_079 in regulating Shh
300 signaling[51,52], which is known to influence neurosphere formation[53].
301 Using RNA-seq in NPCs, we identified DEGs comparing each CRISPRa line to its
302 corresponding empty vector (EV) control. We found that most DEGs were unique to each
303 proband, with overlaps primarily within the same proband, rather than across different probands,
304 even when the same 16p12.1 deletion gene was activated **(Figure 4B, Supplementary Table 4)**.
305 Furthermore, CRISPRa of each one of these genes did not change the expression of other
306 16p12.1 genes **(Supplementary Figure 7B, Supplementary Table 4)**. These findings suggest
307 that CRISPRa-mediated restoration of 16p12.1 genes leads to outcomes modulated by the
308 genomic background and that these deletion genes are not strongly interconnected, consistent
309 with previous observations[46].
310 We investigated whether CRISPR activation of individual 16p12.1 genes could reverse
311 transcriptomic changes initially observed in probands with the deletion. By comparing DEGs
312 obtained from CRISPRa experiments with those from probands, we identified 5,058 genes with
313 opposing directions of gene expression change (i.e., reversed genes) **(Supplementary Figure**
314 **7C, Supplementary Table 4).** For example, restoring *POLR3E* reversed expression of 16 genes
315 across all probands, including cytoskeleton-related (*PLS3* and *LIMA1*), cell adhesion
316 (*PCDHA12*, *JAM2*, and *CD24*), and signaling transduction (*ZYX*, *S1PR1*, and *PDE1A*) genes
317 **(Supplementary Figure 7D, Supplementary Table 4)**. Expression of specific phenotype-
318 related genes were also reversed by activating distinct 16p12.1 genes. For example, the epilepsy
319 risk gene *NNAT*[54] was upregulated in P2C_079 and P1C_007, who both have epilepsy, and was
320 reversed with CRISPRa of *POLR3E* in both probands. CRISPRa also reversed altered isoform
321 usage; activating *MOSMO* restored isoforms of *MECP2* (ENST00000627864) and *EPS8L2*
322 (ENST00000318562), genes that have been associated with autism and deafness, respectively, in
323 P1C_077, who manifested these clinical features **(Supplementary Table 4)**[55,56].
324 Next, we investigated whether CRISPR activation of the 16p12.1 genes rescued cellular
325 phenotypes. In P1C_007, overexpression of NKX2.1 was reversed with CRISPRa of *POLR3E*
326 and *MOSMO*, while proliferation and apoptosis phenotypes were rescued by CRISPRa of
327 *UQCRC2* and *MOSMO* **(Figure 4C-D, Supplementary Figure 8A**). Additionally,
328 overexpression of *UQCRC2* lead to further enhanced premature differentiation of NPCs to

329 neurons in this proband **(Supplementary Figure 8B)**. In P2C_079, CRISPRa of *POLR3E*
330 rescued hyperproliferation phenotypes **(Figure 4E, Supplementary Figure 8C)**. These results
331 underscore the direct role of 16p12.1 genes towards the neuronal defects observed in deletion
332 carriers, with some functional effects observed among multiple genes and others being gene
333 specific.
334 We further investigated the pathways altered by CRISPR activation of 16p12.1 genes and
335 found enrichment for multiple signaling pathways such as TGF-β**,** PI3K-AKT, Wnt, and MAPK
336 **(Supplementary Figure 8D, Supplementary Table 4).** Dysregulation of these pathways is
337 known to disrupt cell-fate determination and cell-cycle progression[57–59], as observed in the
338 deletion lines. Therefore, we performed GSEA on the genes with altered expression from
339 CRISPRa lines to assess the direction of change in associated signaling pathways compared to
340 EV controls in proband lines **(Figure 4F, Supplementary Table 4).** We found that CRISPRa of
341 *POLR3E* and *MOSMO* led to the restoration of TGF-β signaling, which corresponded to the
342 rescue of the NKX2-1 expression level in P1C_007. Furthermore, activation of *POLR3E* also
343 resulted in the upregulation of immune-related pathways, including interferon and cytokine
344 signaling, which were downregulated in multiple deletion lines**.** Specific genes within these
345 pathways, such as *STAT3* and *TGFB3* within immune pathways, and *FZD4*, *BMP7* and *NOG*
346 within Wnt and TGF-β pathways, were also found to be dysregulated in these probands,
347 potentially contributing to the reversal of proliferation and apoptosis phenotypes[60] (**Figure 4G**).
348 Together, these results suggest that each 16p12.1 gene modulates distinct signaling cascades in a
349 genetic background-dependent manner, thereby contributing to variable functional trajectories
350 across individuals.

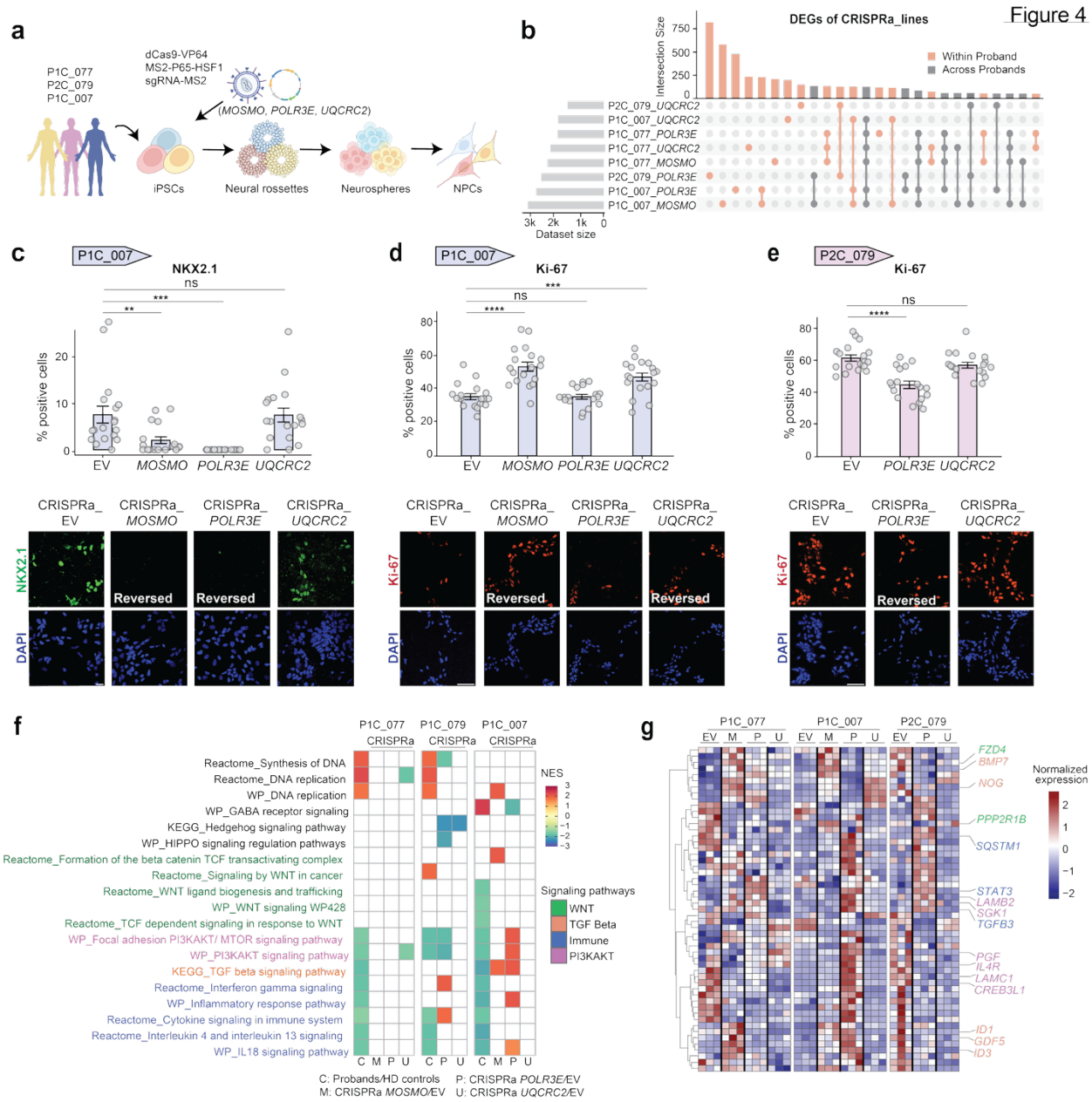


Figure 4. CRISPRa-mediated activation of 16p12.1 genes across probands (a) Schematic of the CRISPRa approach (using dCas9-VP64 and MS2-P65-HSF1) in probands from the three families. **(b)** Upset plot shows overlaps among DEGs ($\log_2|FC|\geq0.5$, padj<0.05) for each comparison (CRISPRa sgRNAs of deletion genes vs. empty sgRNA-MS2 vector) in the three proband lines. **(c)** Quantification and representative images of NKX2.1-positive cells with CRISPR activation of 16p12.1 genes in NPCs from P1C_007. Scale bars, 20 µm. **(d)** Quantification and representative images of Ki-67-positive cells with CRISPR activation of 16p12.1 genes in NPCs from P1C_007. Scale bars,70 µm. **(e)** Quantification and representative images of Ki-67-positive cells with CRISPR activation of 16p12.1 genes in NPCs from P2C_079. Scale bars, 70 µm. **(f)** Select pathways (q<0.05) enriched by GSEA across

363 comparisons of different sets of DEGs (q<0.05) annotated in the Reactome, WP, and KEGG
364 databases. Enrichment analysis with GSEA were made between DEGs obtained from deletion
365 NPCs versus controls compared to CRISPRa versus EV controls for each proband. NES,
366 normalized enrichment score; M, pathways derived from CRISPRa_*MOSMO* DEGs; P, pathways
367 derived from CRISPRa_*POLR3E* DEGs; U, pathways derived from CRISPRa *UQCRC2* DEGs.
368 White color fill indicates absence of terms in GSEA results. (**g**) Heatmap of normalized gene
369 expression in selected pathways (colored in (**f**)) across CRISPRa lines, clustering using Ward's
370 D2. Color scale represents z-scores calculated from TPM within each group. EV, empty sgRNA-
371 MS2 vector; M, CRISPRa of *MOSMO*; P, CRISPRa of *POLR3E*; U, CRISPRa of *UQCRC2*.
372 Data are mean±s.e.m and one-way ANOVA followed by Dunnett's post hoc test were used for
373 **(c)**, **(d)**, and **(e)**. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. n=3 independent experiments
374 for each line, except for RNA-seq and downstream analyses where n=2 for iMNs and MNs of
375 P1C_077, and MNs of HD_01. For **(c)**, **(d)**, and **(e),** n=6 images per experiment were quantified.
376

377

378 **Connectivity of genes with secondary variants in PPI network**

379 To further investigate the role of secondary variants and explain phenotypic differences across
380 probands, we analyzed how genes with secondary variants whose expression was altered
381 following CRISPRa of individual 16p12.1 genes in NPCs, were connected within a protein-
382 protein interaction network[61] **(Figure 5A, Supplementary Figure 9A, Supplementary Table 5)**.
383 In the *POLR3E* and *MOSMO* CRISPRa lines, we observed a higher network connectivity for
384 P1C_007 and P2C_079 compared to P1C_077, correlating with the presence of neuronal defects
385 in NPCs of P1C_007 and P2C_079 but not P1C_077 **(Figure 5A).** These findings were further
386 validated by analysis of PPI networks specifically within signaling pathways, which revealed
387 higher network connectivity in P1C_007 and P2C_079 **(Figure 5B)**. Interestingly, in the
388 *UQCRC2* CRISPRa lines, P1C_077 showed higher connectivity compared to P1C_007 and
389 P2C_079 **(Supplementary Figure 9A-B).** The highly connected nodes within this network
390 included proteins such as ADCY1, PDE4A, and GNG10, which are involved in cAMP signaling
391 and G protein signaling pathways that modulate synaptic functions in mature neurons[62,63]. This
392 may explain the absence of cellular phenotypes in P1C_077 at the NPC stage. These findings
393 indicate that the extent to which modifiers of individual 16p12.1 genes are connected within
394 signaling pathways is a determinant of phenotypic variation across probands.

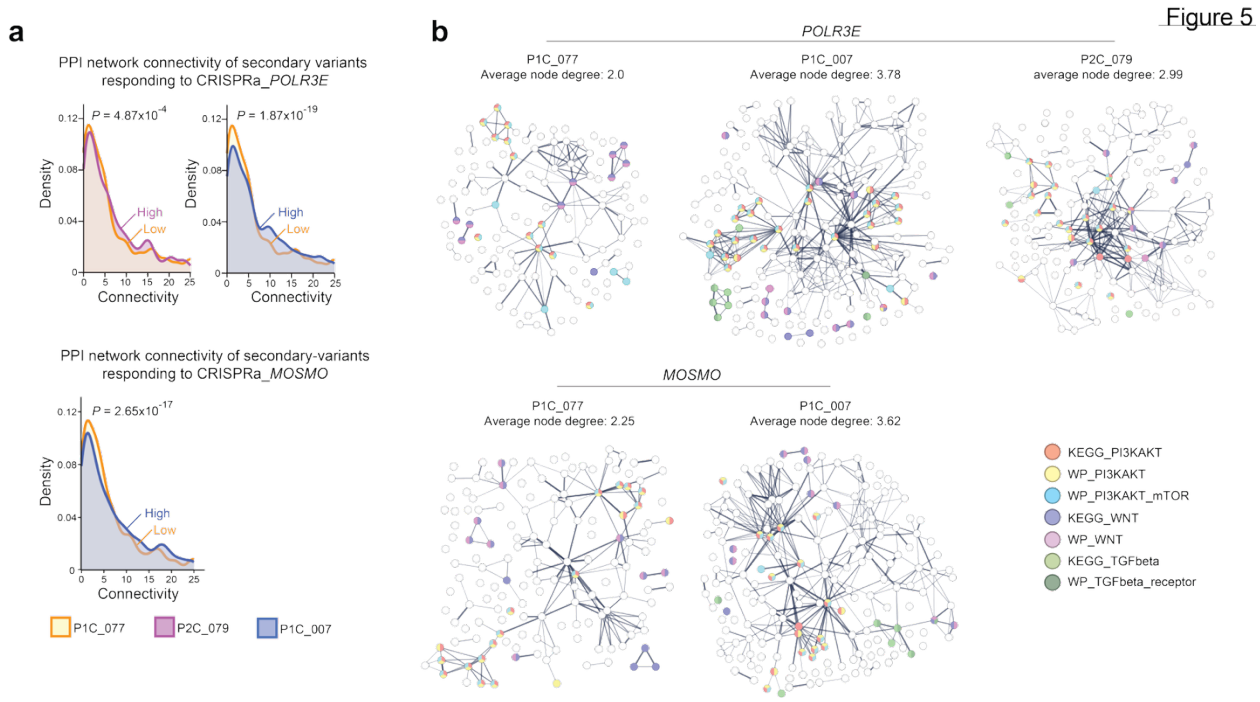


Figure 5. **Connectivity of secondary variants across probands in response to CRISPR activation of 16p12.1 genes. (a)** Density plots show the connectivity (node degree, i.e., the number of connections a node has to other nodes in the network) of *POLR3E* and *MOSMO* CRISPRa DEGs ($\log_2|FC|>0.5$, padj< 0.05) that overlap with secondary variants, as measured using the STRING database. Anderson-Darling k-sample test was used to calculate p values. **(b)** Visualization of the PPI network of DEGs ($\log_2|FC|\geq 0.5$, padj<0.05) that overlap with secondary variants and were restored by CRISPRa of *POLR3E* in P1C_077 (top left), P1C_007 (top middle), and P2C_079 (top right), as well as by CRISPRa of *MOSMO* in P1C_077 (bottom left) and P1C_007 (bottom right). DEGs shown were selected based on their involvement in signaling pathways annotated in the Reactome database. p-value $<10^{-16}$ for all networks. Protein nodes involved in specific pathways are color-coded based on annotations in the STRING database. Average node degrees and PPI enrichment p values were also calculated using the STRING database. n=3 independent experiments for each line.

Modulation of FOXG1 and JUN mediates phenotypic variability

We sought to identify functional convergence across samples through which the deletion confers disease susceptibility, modulation of which by secondary variants lead to diverse outcomes. As the network of genes within signaling pathways typically converge on TFs[64], we sought to identify key TFs that regulate downstream effects in the deletion lines. We found that the sequences within Diffpeak regions across the deletion NPCs were enriched for binding motifs of TFs involved in developmental processes, such as FOS, HOXD13 and NR4A1[65–67]

418 **(Supplementary Table 6)**. Furthermore, co-regulatory networks constructed from these TFs
419 whose binding motifs showed the most significant enrichment in Diffpeak regions across the
420 deletion lines identified FOXG1 and JUN as two of the top-ranked TFs using ChEA3 **(Figure**
421 **6A, Supplementary Table 6**, see **Methods).** Both of these TFs have been previously implicated
422 in multiple signaling pathways, including Wnt, BMP, and JNK[68–70]. Furthermore, we observed
423 altered expression of FOXG1 and JUN across the deletion lines compared to controls, along with
424 changes in both expression and chromatin accessibility of genes within their regulatory
425 networks. These included genes such as *HES1*, *KLF6*, *DLX2*, and *FZD8,* which are known to
426 play roles in neural processes such as inhibitory neuron production and NPC proliferation[71–74]
427 (**Figure 6A, Supplementary Fig. 10A-B, Supplementary Table 6).** Therefore, we hypothesized
428 that the 16p12.1 deletion affects regulatory hubs within signaling pathway interaction networks,
429 and that modulating the expression of individual 16p12.1 genes would, in turn, alter the
430 expression of these hub genes, such as FOXG1 and JUN, as well as their connected genes,
431 leading to changes in cellular phenotypes (**Figure 7**). To test this, we further investigated the
432 roles of FOXG1 and JUN towards the observed cellular dysregulations in the deletion lines.
433 Since altered activity of the TFs reflects perturbations across multiple associated signaling
434 pathways, we did not expect a direct one-to-one correspondence between specific TF activity and
435 cellular phenotypes. We found that the expression levels of FOXG1 and JUN varied across
436 probands, and these levels were differentially modulated with CRISPR activation of different
437 16p12.1 genes **(Figure 6B-C)**. Changes in FOXG1 and JUN expression in the P1C_007
438 CRISPRa lines, and changes in JUN expression in the P2C_079 lines, were observed alongside
439 the reversal of cellular phenotypes upon CRISPR activation **(Figure 4, Figure 6D)**. Furthermore,
440 we found that the expression of genes within their regulatory networks, including *DLX2* and
441 *KLF6*, mirrored changes in *FOXG1* and *JUN* expression across CRISPRa lines **(Figure 6E)**.
442 These findings suggest that key TFs and the regulatory network within individual 16p12.1 gene-
443 associated signaling pathways mediate the phenotypic trajectories of the deletion (**Figure 7**).
444

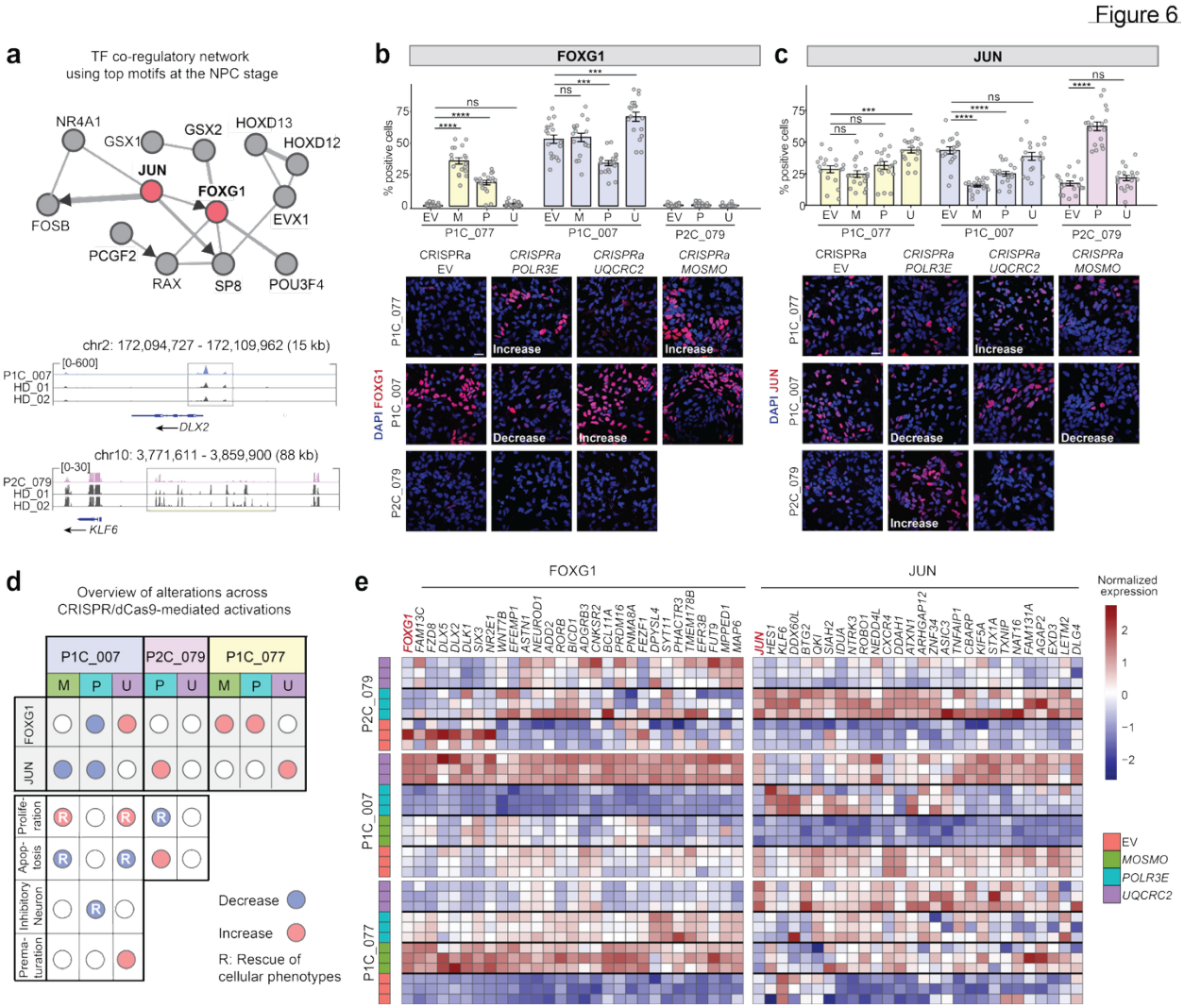
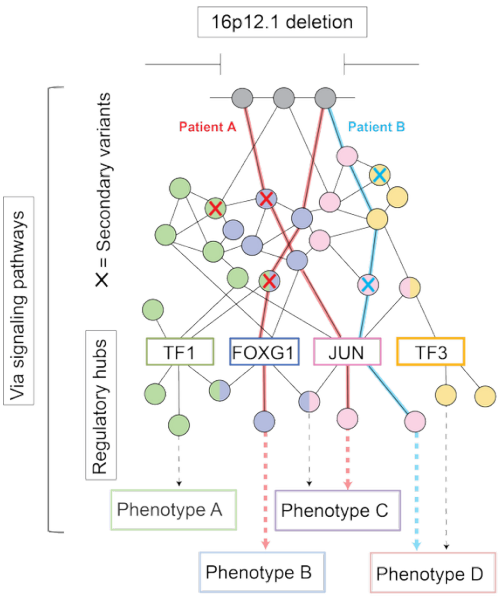


Figure 6. Modulation of gene regulatory hubs corresponding with variable phenotypes. (a) TF co-regulatory network constructed by ChEA3 using TFs whose binding motifs were most significantly enriched in ATAC-seq Diffpeak regions from comparison of deletion lines vs. healthy donor lines at the NPC stage. Motif lists are provided in **Supplementary Table 6** (see **Methods**). The panel below depicts IGV snapshots of ATAC-seq peaks showing increased (*DLX2*) or decreased (*KLF6*) chromatin accessibility for select genes in the regulatory network of FOXG1 and JUN, respectively. **(b)** Quantification and representative images of FOXG1-positive NPCs in CRISPR activated lines from P1C_077, P1C_007, and P2C_079. Scale bars, 20 µm. EV, empty sgRNA-MS2 vector; M, CRISPRa of *MOSMO*; P, CRISPRa of *POLR3E*; U, CRISPRa of *UQCRC2*. **(c)** Quantification and representative images of JUN-positive NPCs in CRISPR activated lines from P1C_077, P1C_007, and P2C_079. Scale bars, 20 µm. **(d)** Overview of CRISPRa results at the NPC stage. Columns represent CRISPRa lines from the three probands; rows indicate the direction of change based on phenotype quantification. Empty

460 bubbles denote no detectable change. Rescue of cellular phenotype refers to the reversal of a
461 cellular phenotype observed in the proband lines (compared to healthy donor controls). The
462 phenotypes of proliferation, apoptosis, inhibitory neuron, and premature maturation were,
463 respectively, measured with EdU, TUNEL assay, VGAT, NKX2.1 and TUBB3 positive cells. **(e)**
464 Representative heatmap of normalized expression of genes within the regulatory networks
465 (identified using ChEA3) of FOXG1 (left) and JUN (right) across the CRISPRa lines. Color
466 scale represents the z-scores calculated from TPM within each group. Data are mean±s.e.m and
467 one-way ANOVA followed by Dunnett's post hoc test for **(b)** and **(c)**. *p<0.05; **p<0.01;
468 ***p<0.001; ****p<0.0001. n=3 independent experiments for each line. 6 images per
469 experiment were quantified for **(b)** and **(c)**.

470

Figure 7



471

472

473 **Figure 7. Conceptual model illustrating how the interplay between the 16p12.1 deletion and**
474 **secondary variants contributes to variable expressivity**. The interactive effects are mediated
475 through regulatory hubs, such as TFs (e.g., FOXG1, JUN), within signaling pathways. The
476 network is colored for the TF regulatory hub, and two-toned colors indicate genes shared among
477 multiple pathways. Patient-specific secondary variants or combinations of variants lead to
478 distinct phenotypic trajectories, shown by the red or blue connecting lines traversing distinct
479 genetic pathways.
480

481

482 **DISCUSSION**

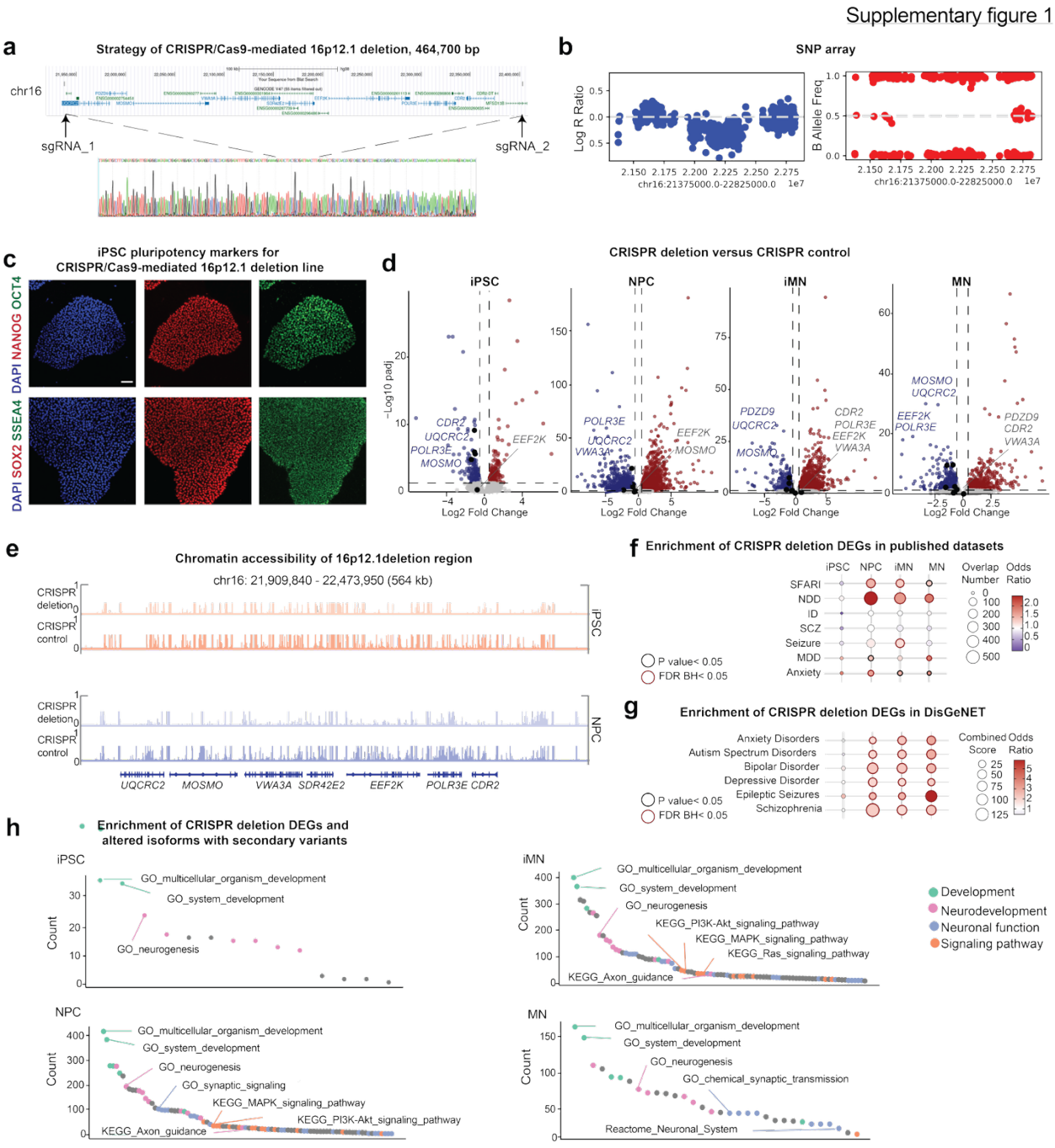
483 Our study provides a framework to understand how variable expressivity in neurodevelopmental
484 disorders arises from the complex interplay between disease-associated variants and secondary
485 variants in the genetic background. Using the 16p12.1 deletion as a model, we found that
486 primary and secondary variants jointly modulate key signaling pathways and shape distinct
487 phenotypic trajectories through transcription factors within gene regulatory networks (**Figure 7**).
488 Several themes have emerged from our work, contributing to the emerging picture that network-
489 level genetic interactions underlie variable expressivity in complex genetic disorders.

490 Previous studies have linked CNVs to changes in chromatin accessibility and gene
491 expression in neurodevelopmental disorders[75,76]. Here, we observed potential nonadditive effects
492 of primary and secondary variants on both chromatin accessibility and gene expression. Notably,
493 even the ostensibly healthy control used to generate the CRISPR deletion line did not fully
494 control the genetic background effects and therefore cannot isolate the effect of the deletion
495 alone. The presence of secondary variants interacting with the deletion highlights the need to
496 carefully account for genetic background effects in future studies, even when using isogenic
497 models, as introducing the same variant into distinct lines may yield different results when
498 characterizing variant function.

499 Although observed within specific genetic backgrounds, the 16p12.1 deletion led to
500 defects in inhibitory neuron production, abnormal cell proliferation and apoptosis, and premature
501 neuronal differentiation. These findings reflect disruptions in excitatory/inhibitory balance,
502 neurogenesis dynamics, resulting in altered cell type proportions during brain development, a
503 hallmark of multiple neurodevelopmental disorders including ASD, schizophrenia and
504 intellectual disability[19,41,77]. Furthermore, CRISPRa-based rescue of 16p12.1 genes supports our
505 previous findings[46] that each gene within this region functions independently, as rescue of
506 different genes reversed distinct sets of dysregulated genes and cellular phenotypes without
507 affecting the expression of other deleted genes. Future work using this framework could
508 incorporate CRISPR interference-based gene knockdown and be extended to other variably
509 expressive CNVs, such as those at 1q21.1, 15q13.3, 16p11.2, and 22q11.2, to link genes within
510 these regions to cellular phenotypes across the diverse genetic backgrounds of individual
511 patients.

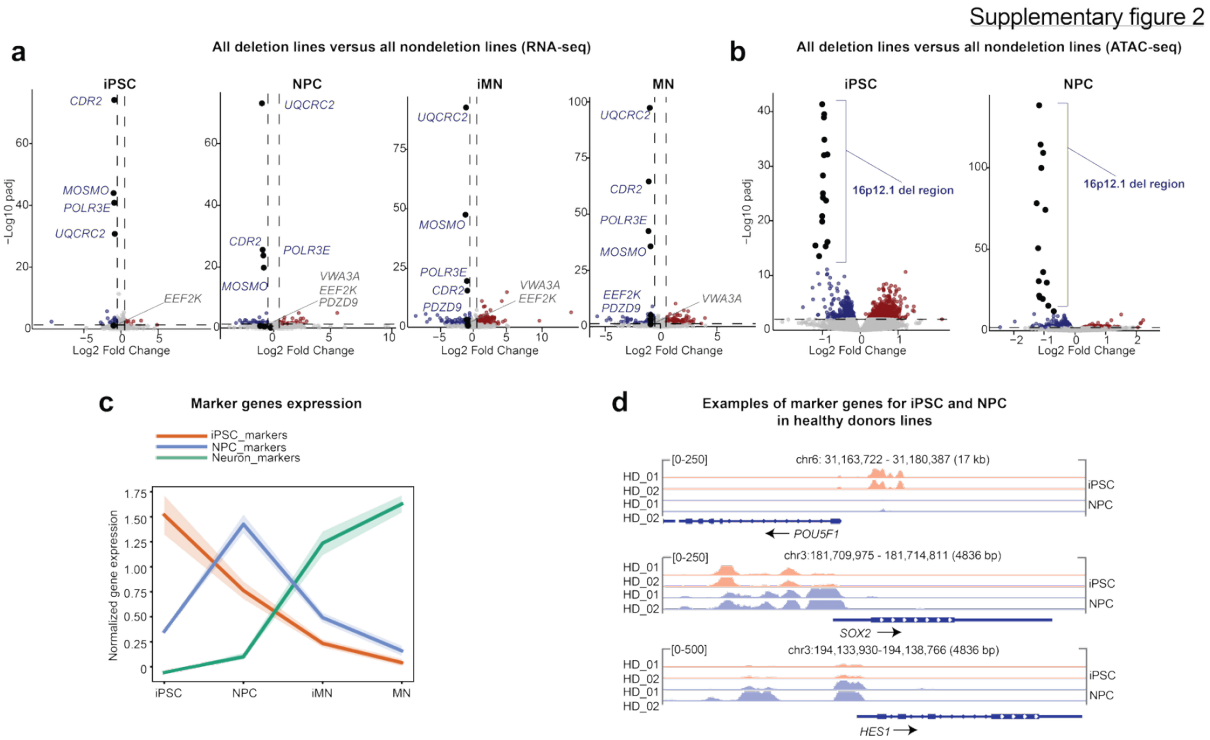
512 We observed alterations across multiple signaling pathways, including Wnt, TGF-β,
513 PI3K-AKT, and immune-related pathways, several of which have been implicated in
514 neurodevelopmental disorders[58,78–81]. These findings reflect the extensive crosstalk and network-
515 level interconnectivity among these pathways[78]. Furthermore, we found that the severity of
516 phenotypic outcomes correlated with the PPI connectivity of secondary variants in specific
517 signaling pathways. In addition, consistent with prior studies identifying convergent mechanisms
518 underlying heterogeneity in autism[82], we found differential modulation of transcription factors
519 such as FOXG1 and JUN, which serve as regulatory hubs across deletion lines with distinct
520 genetic backgrounds. TFs have been shown to be key regulators of cellular differentiation, and
521 their dysregulation has been linked to developmental disorders, including craniofacial
522 anomalies[83]. Together, these results underscore the importance of examining both inter-
523 individual variability and the convergent mechanisms that emerge across diverse genetic
524 contexts.
525 While our study provides key insights into the functional impact of secondary variants,
526 several aspects could be further refined or expanded in future work. *First*, while we observed
527 strong associations between secondary variants and changes in gene expression and chromatin
528 accessibility, the small sample size limits our ability to infer causality without further
529 experimental validation. High-throughput approaches such as Perturb-seq and CROP-seq could
530 help confirm these interactions and elucidate how they modulate signaling pathways[84,85]. *Second*,
531 we used 2D iPSC differentiation combined with bulk RNA-seq, which may overlook cell type–
532 specific effects, particularly in neuronal subtypes and non-neuronal brain cells relevant to
533 neurodevelopmental disorders[38]. Future work could leverage advanced models such as neural
534 organoids, assembloids, chimeroids, or other emerging platforms, which provide greater cellular
535 diversity and spatial resolution[86]. Overall, our findings lay the groundwork for future functional
536 studies to explore genotype–phenotype relationships, inter-individual variability, and potential
537 personalized therapeutic strategies.
538 In summary, our study provides a framework for understanding how the combined effects
539 of disease-associated variants and genetic background drive variable expressivity in
540 neurodevelopmental disorders. By uncovering both shared and background-specific impacts on
541 gene regulation and signaling pathways, we underscore the importance of accounting for genetic
542 context in functional studies and in developing personalized therapeutic strategies.

Supplementary figure 1

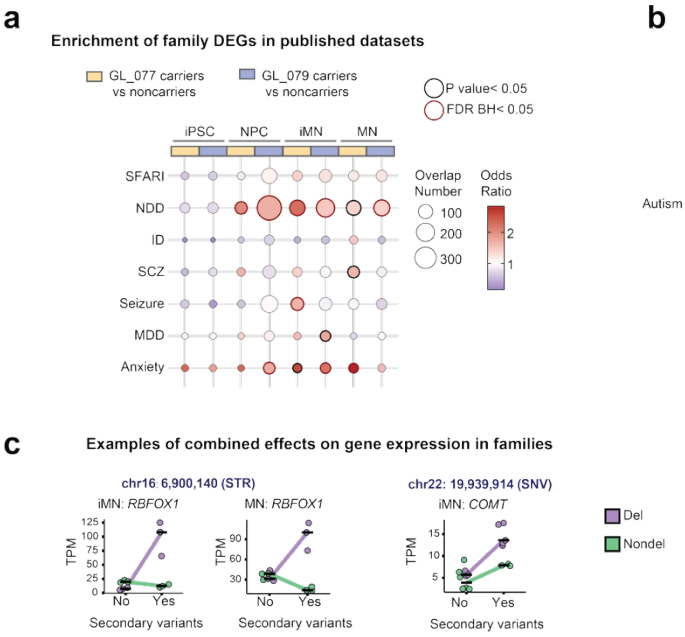


Supplementary Figure 1. Generation of CRISPR/Cas9-mediated 16p12.1 deletion. (a) CRISPR/Cas9 strategy for generating 16p12.1 deletion. Arrows demonstrate the locations of the designed sgRNAs. The engineered 465-kbp deletion was validated by Sanger sequencing. **(b)** Confirmation of the deletion by SNP array. **(c)** Representative images of the CRISPR edited line stained for pluripotency markers (SOX2, SSEA4, NANOG and OCT4). Scale bars, 70 µm. **(d)** Volcano plots show the differential expression of 16p12.1 deletion genes (*UQCRC2*, *MOSMO*, *VWA3A*, *EEF2K*, *POLR3E* and *CDR2*) in the isogenic setting across neuronal differentiation stages (iPSC, NPC, IMN and MN). Dashed lines represent the threshold ($\log_2|FC|\geq 0.5$,

553 padj<0.05). Red, upregulated genes; blue, downregulated genes; grey, non-significant genes.
554 Deletion genes that are not labeled demonstrate extremely low expression levels and were not
555 detected. **(e)** IGV snapshots show the ATAC-seq peaks of chromatin accessibility within the
556 16p12.1 deletion region from the CRISPR-edited line. **(f)** Enrichment analysis of DEGs
557 ($\log_2|FC|\geq1$, padj <0.05) obtained from RNA-seq of the CRISPR-edited line compared to
558 isogenic controls queried within published datasets. Gene lists from selected papers are provided
559 in **Supplementary Table 8**. **(g)** Enrichment analysis of DEGs ($\log_2|FC|\geq1$, padj <0.05) obtained
560 from RNA-seq of the CRISPR-edited line compared to isogenic controls within DisGeNET. For
561 **(f)** and **(g)**, Fisher's exact test, bold black circles indicate p<0.05, bold red circles indicate
562 Benjamini-Hochberg FDR<0.05. **(h)** Waterfall plot shows enrichment of DEGs overlapping
563 secondary variants across the 4 neuronal differentiation stages. Top 30 terms (q<0.05) were
564 unbiasedly selected from GO (Biological Process, Gene Ontology), KEGG (Kyoto Encyclopedia
565 of Genes and Genomes) and Reactome databases. RNA-seq and downstream analyses: n=3
566 independent experiments; ATAC-seq and downstream analyses: n=3 independent experiments.
567

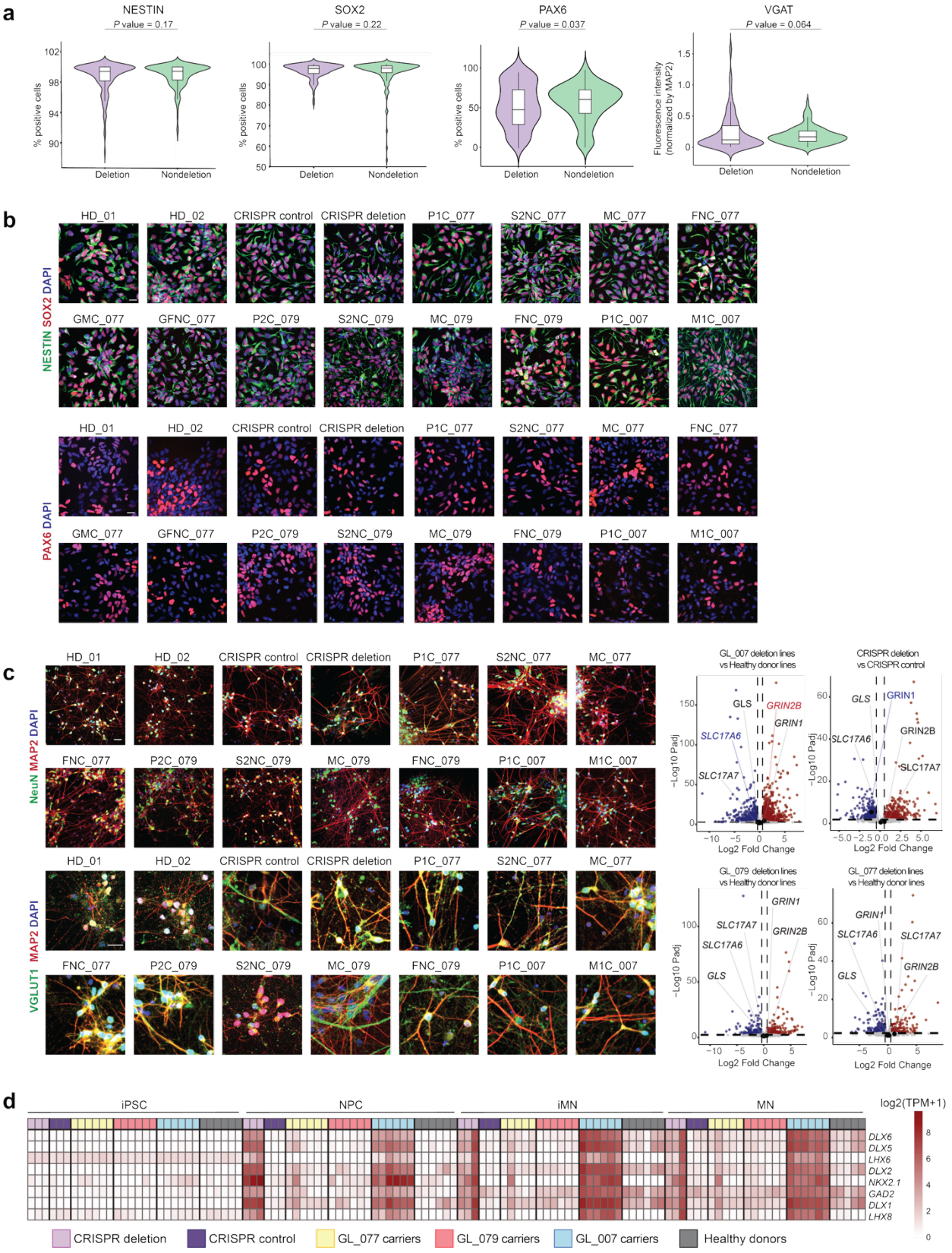


Supplementary Figure 2. Validation of cell identities across stages of neuronal differentiation. **(a)** Volcano plots show the differential expression of 16p12.1 deletion genes (*UQCRC2*, *MOSMO*, *VWA3A*, *EEF2K*, *POLR3E* and *CDR2*) comparing all the deletion lines to nondeletion lines across neuronal differentiation stages (iPSC, NPC, IMN and MN). Dashed lines represent the statistical threshold for significance ($\log_2|FC|\geq0.5$, padj<0.05). Red, upregulated genes; blue, downregulated genes; grey, non-significant genes. Deletion genes that are not labeled demonstrate extremely low expression levels and were not detected. **(b)** Volcano plots show the differential peaks in 16p12.1 deletion when comparing all the deletion lines to nondeletion lines at iPSC and NPC stages. Dashed lines represent the threshold (padj<0.01). Red, upregulated peaks; blue, downregulated peaks; grey, non-significant peaks. Black dots demonstrate differential peaks within 16p12.1 regions. A list of all ATAC-seq peak regions is provided in **Supplementary Table 2**. **(c)** Line plot shows the expression of select marker genes with shading representing 95% confidence interval of the mean across all the lines and differentiation stages. Y-axis represents z-scores of TPM. iPSC markers: SOX2, POU5F1, and FUT4; NPC markers: SOX2, HES1, NES, and EMX2; neuron markers: MAP2, DLG4, NCAM1, and RBFOX3. **(d)** IGV snapshots show the ATAC-seq peaks of chromatin accessibility from HD_01 and HD_02 in the 16p12.1 deletion regions. For RNA-seq, ATAC-seq and downstream analyses: n=3 independent experiments, except for RNA-seq and downstream analyses where n=2 for iMNs and MNs of P1C_077, and MNs of HD_01.



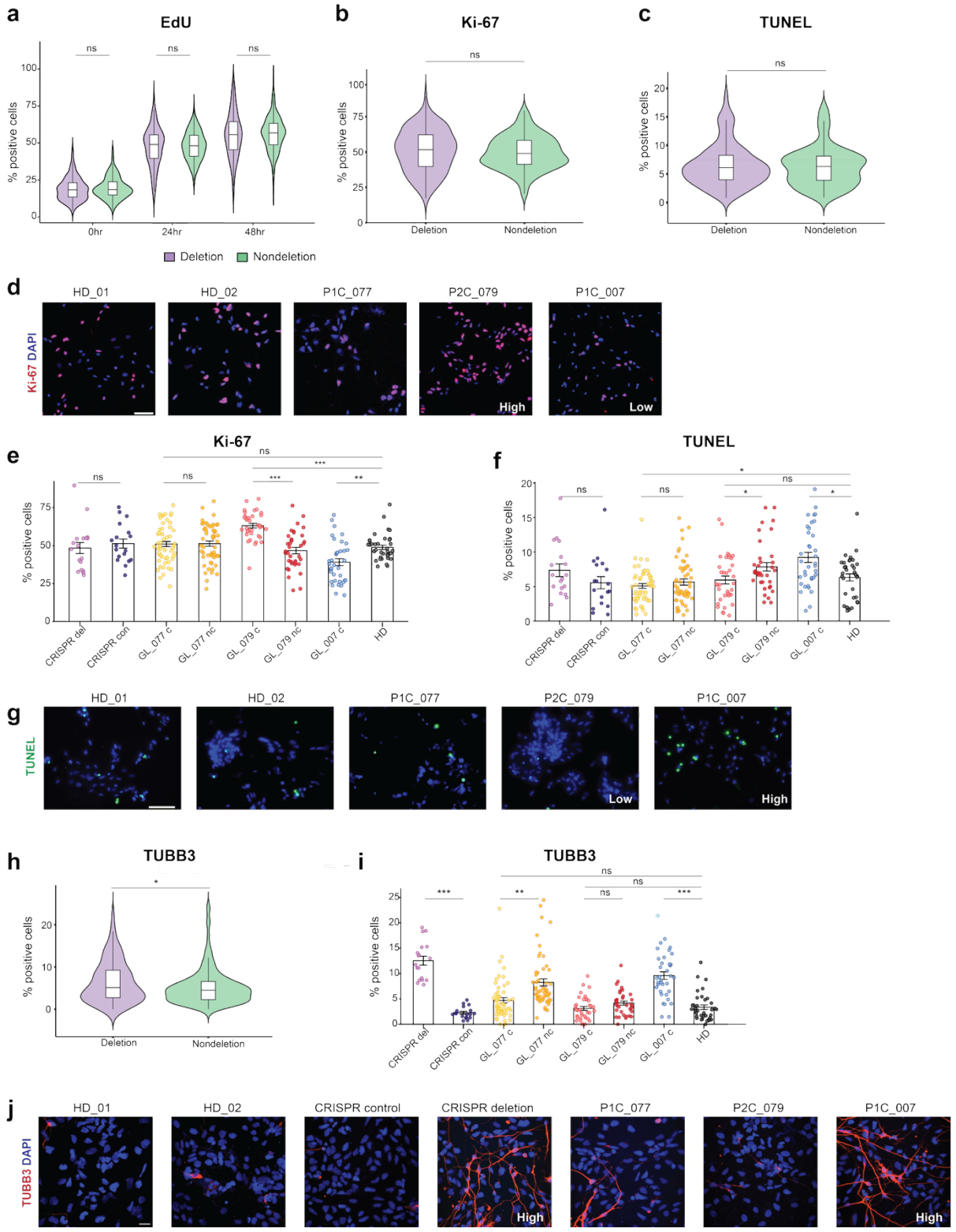
Supplementary Figure 3. Family-specific effects on gene expression. (a) Enrichment of DEGs ($\log_2|FC|\geq1$, padj <0.05) from deletion versus nondeletion within-family comparisons in published datasets. A complete list of genes from select papers is provided in **Supplementary Table 8**. **(b)** Enrichment analysis of DEGs ($\log_2|FC|\geq1$, padj <0.05) from deletion versus nondeletion within-family comparisons in DisGeNET. For **(a)** and **(b)**, Fisher's exact test, bold black circles indicate p<0.05, bold red circles indicate Benjamini-Hochberg FDR<0.05. GL_007 carriers are MC_077 and P1C_077, noncarriers are FNC_007 and S2NC_077, and GL_079 carriers are MC_079 and P2C_079, noncarriers are FNC_079 and S2NC_079. **(c)** In the same family, examples of genes and isoforms overlapping with secondary variants showing altered (TPM and mean values) gene expression at iMN and MN stages (**Supplementary Table 3**). *indicates the carriers of secondary variants. SNV, single nucleotide variant; STR, short tandem repeat; Del, 16p12. 1 deletion lines; Nondel, nondeletion lines. Two-way ANOVA tests for all the examples using TPM are interaction p<0.05. For RNA-seq and downstream analyses: n=3 independent experiments, except for iMNs and MNs of P1C_077, where n=2.

Supplementary figure 4

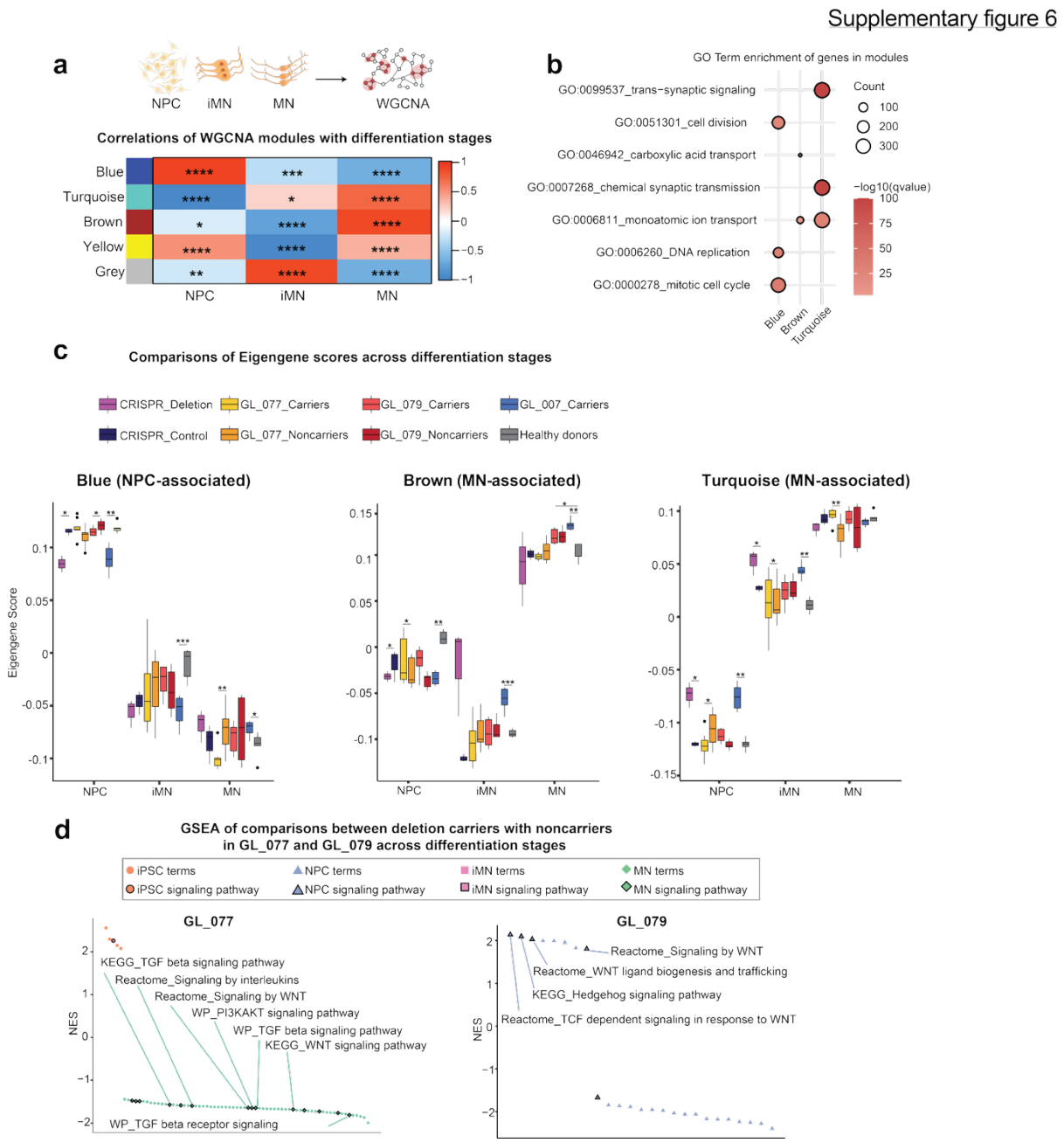


615 **Supplementary Figure 4. Characterization of NPCs and neuronal lineage markers**. (a)
616 Violin plots show the percentage of NPCs that were positive for NESTIN, SOX2, PAX6 and
617 quantification of VGAT staining in neurons from all deletion lines versus all nondeletion lines.
618 The embedded box plot indicates the interquartile range (IQR) and median (black horizontal
619 line). p-values calculated using two-tailed Welch's t-test. **(b)** Representative images of NPCs
620 stained for NESTIN, SOX2, and PAX6. **(c)** Representative images (left) of MNs stained for
621 NeuN and MAP2 (mature neuron markers) and VGLUT (excitatory neuron marker), and volcano
622 plots (right) showing the differential expression of five marker genes for glutamatergic neurons
623 between deletion lines in each family and healthy donor lines, as well as the CRISPR deletion
624 compared to the CRISPR control. Red, upregulated genes; blue, downregulated genes; grey, non-
625 significant genes. **(d)** Heatmap shows the expression levels of inhibitory neuron-related genes
626 across neuronal differentiation stages. Note that the differences in gene expression between the
627 CRISPR-edited line and controls become more subtle as neuronal differentiation progresses
628 (from iPSC to MN). Color scale represents $\log_2$(TPM+1). RNA-seq and downstream analyses:
629 n=3 independent experiments, except for MNs of HD_01, where n=2, and iM and MNs of
630 P1C077, where n=2. Scale bars for **(b)** and **(d)**, 20 μm.

Supplementary figure 5



Supplementary Figure 5. Altered proliferation, apoptosis, and neuronal maturation at the NPC stage. (a) Violin plot shows the percentage of EdU-positive cells at the NPC stage in all deletion lines versus all nondeletion lines across three time points. The embedded box plot indicates the interquartile range (IQR) and median (black horizontal line). **(b)** Violin plot shows the percentage of Ki-67 positive cells at the NPC stage in all the deletion lines versus all the nondeletion lines. The embedded box plot indicates the IQR and median (black horizontal line). **(c)** Violin plot shows the percentage of TUNEL-positive cells at the NPC stage in all the deletion lines versus all the nondeletion lines. The embedded box plot indicates the interquartile range (IQR) and median (black horizontal line). **(d)** Representative images of NPCs stained for Ki-67. Scale bars, 70 µm. **(e)** Bar plot (with mean values) shows the percentage of Ki-67-positive cells in different samples. **(f)** Bar plot shows the percentage of TUNEL-positive cells in different samples. **(g)** Representative images of TUNEL-labeled NPCs. Scale bars, 70 µm. **(h)** Violin plot shows the percentage of TUBB3-positive cells at the NPC stage in all the deletion lines versus all the nondeletion lines. The embedded box plot indicates the interquartile range (IQR) and median (black horizontal line). **(i)** Bar plot shows the percentage of TUBB3- positive cells in different samples. **(j)** Representative images of TUBB3-positive NPCs. Scale bars, 20 µm. CRISPR_del, CRISPR deletion; CRISPR_con, CRISPR control; GL_077_c, 16p12.1 deletion lines in GL_077 (GMC_077, MC_077, P1C_077); GL_077_nc, nondeletion lines in GL_077 (GFNC_077, FNC_077, S2NC_077); GL_079_c, deletion lines in GL_077 (MC_079, P2C_079); GL_079_nc (FNC_079, S2NC_079); GL_007_c (M1C_007, P1C_007); HD, healthy donor lines (HD_01, HD_02). Two-tailed Welch's t-test, *$p<0.05$; **$p<0.01$; ***$p<0.001$; ****$p<0.0001$, for **(a)**, **(b)**, **(c)**, **(e)**, **(f)**, **(h)**, **(i)**. n=3 independent experiments for each line. 6 images per experiment were quantified.

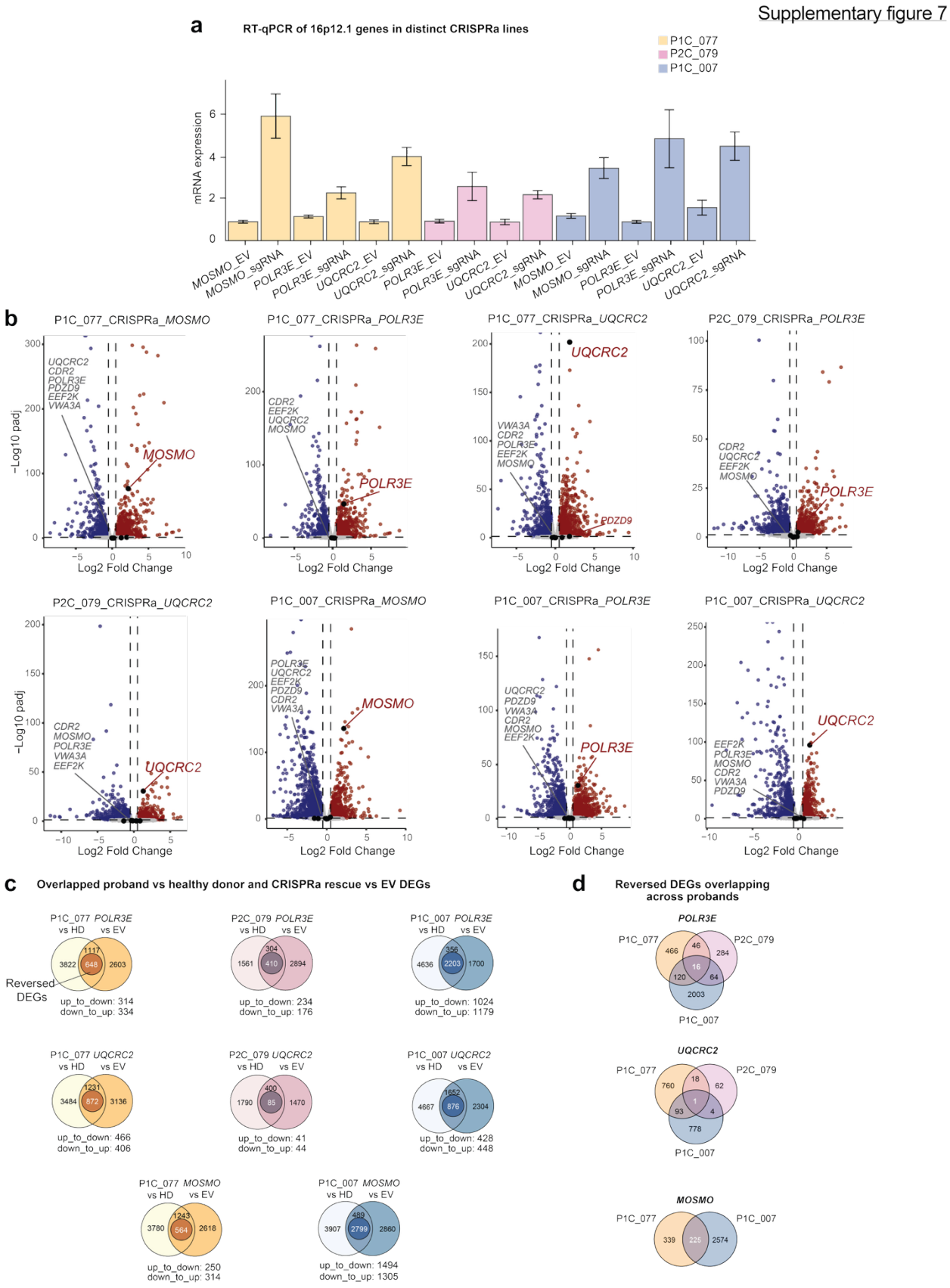


Supplementary Figure 6. Transcriptome dynamics during neuronal differentiation. (a) Heatmap shows Pearson's correlation coefficients between WGCNA modules and differentiation stages. Student asymptotic p-value for correlation for statistics of Pearson correlation. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. Color bar represents direction and strength of correlation. **(b)** GO term enrichment of genes in each module. **(c)** Box plots show the eigengene scores across the differentiation stages. Two-tailed Welch's t-test. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. Each box shows the IQR, with a horizontal line indicating the median. **(d)** Waterfall plot of total unbiasedly selected top 100 pathways enriched for the DEGs identified across differentiation stages comparing deletion to nondeletion lines in GL_077 and GL_079 by GSEA (q<0.05) annotated in Reactome, KEGG and WP (WikiPathway) databases. The bold

672 border indicates terms that belong to signaling pathways. NES, normalized enrichment score;
673 GL_077, GMC_077, MC_077, and P1C_077 versus GFNC_077, FNC_077, and S2NC_077;
674 GL_079, MC_079, and P2C_079 versus FNC_079 and S2NC_079. n=3 independent
675 experiments, except for RNA-seq and downstream analyses where n=2 for iMNs and MNs of
676 P1C_077, and MNs of HD_01.
677

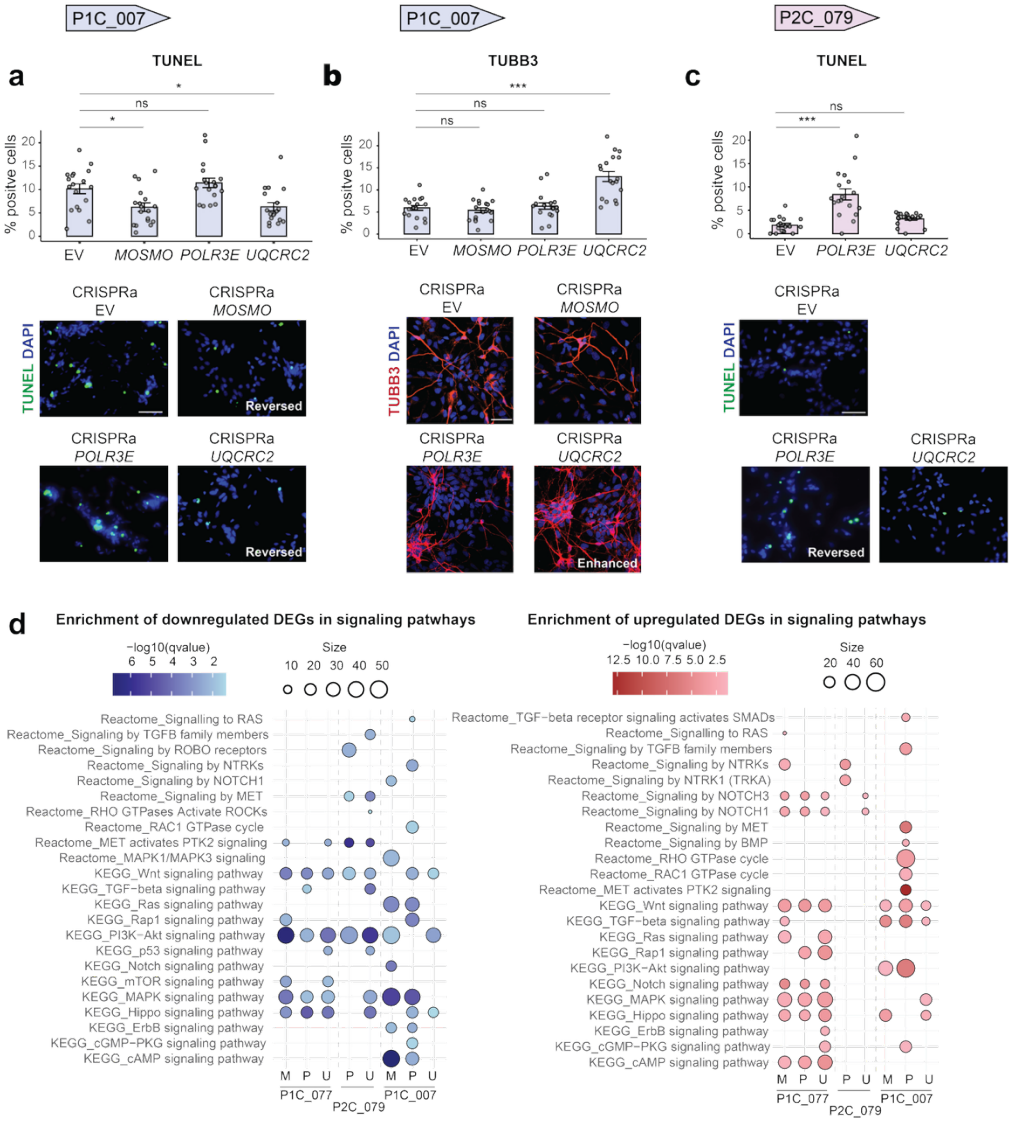
678

Supplementary figure 7

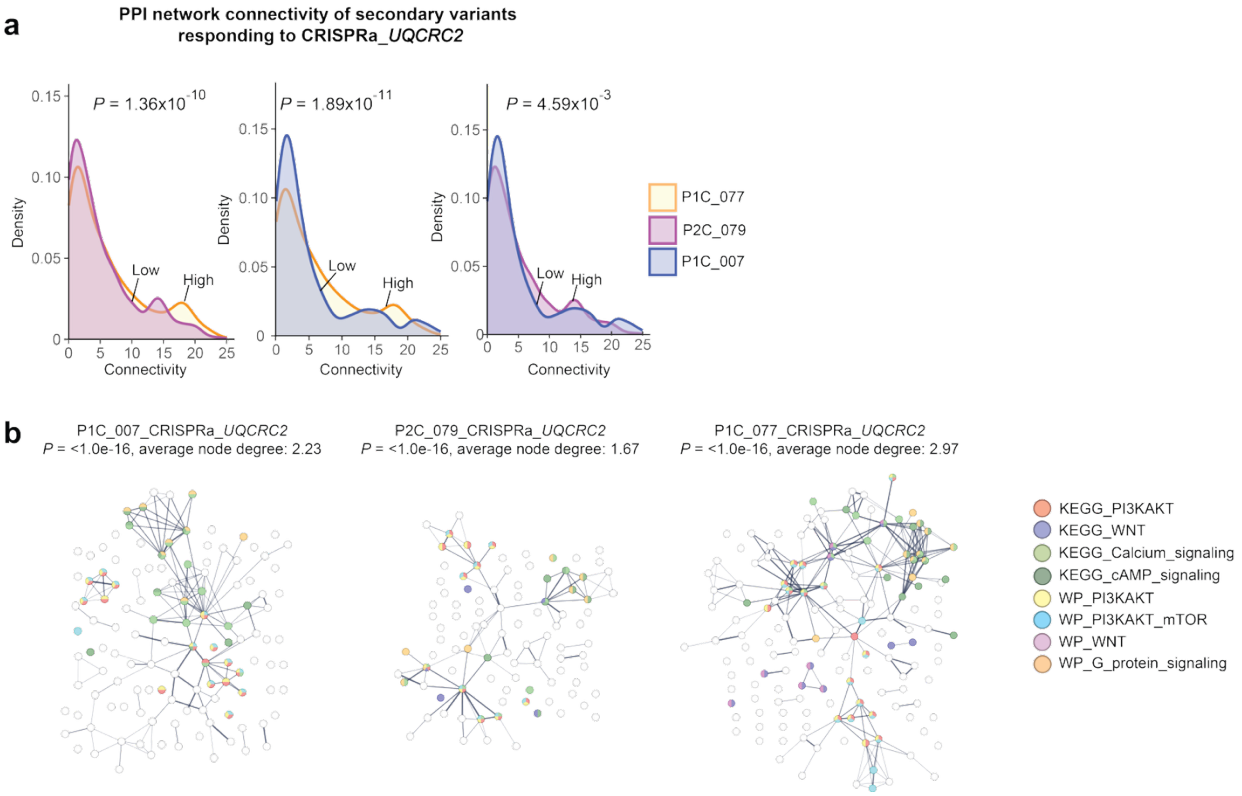


679

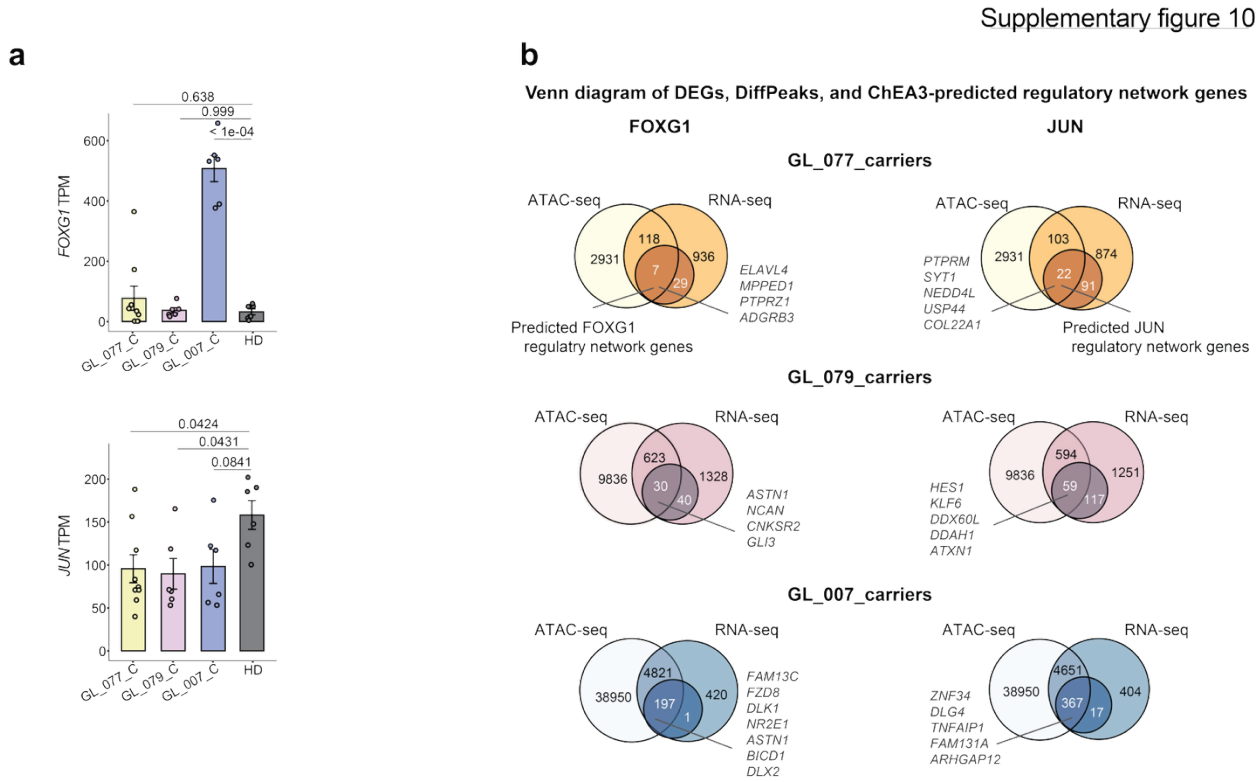
Supplementary Figure 7. Transcriptomic alterations induced by CRISPRa-mediated activation of individual 16p12.1 genes. (a) Bar plot shows relative mRNA expression levels of *MOSMO*, *POLR3E*, and *UQCRC2* across CRISPRa lines derived from the three probands (P1C_077, P2C_079, and P1C_007), measured by RT-qPCR. Each bar represents the mean±standard error of three independent experiments. **(b)** Volcano plots show the differential expression of 16p12.1 deletion genes (*UQCRC2*, *MOSMO*, *VWA3A*, *EEF2K*, *POLR3E*, and *CDR2*) comparing CRISPRa of individual 16p12.1 genes to EV lines. Dashed lines represent the statistical significance threshold ($\log_2|FC|\geq 0.5$, padj<0.05). Red, upregulated genes; blue, downregulated genes; grey, non-significant genes. Deletion genes that are not labeled demonstrate extremely low expression levels and were not detected. **(c)** Venn diagrams show the overlaps (reversed DEGs) of DEGs (probands with the deletion versus HDs) and DEGs (padj<0.05) from CRISPRa of individual genes versus EV in the proband lines. Gene lists are provided in **Supplementary Table 4**. **(d)** Venn diagrams show the patterns of all DEGs that were reversed by CRISPR activation of each 16p12.1 gene across probands. Gene lists are provided in **Supplementary Table 4**. Up_to_down, upregulated in deletion lines but downregulated by CRISPRa; down_to_up, downregulated in deletion lines but upregulated by CRISPRa; HD, healthy donors (HD_01 and HD_02); EV, empty sgRNA-MS2 vector.



Supplementary Figure 8. Alterations in cellular phenotypes following CRISPR activation of 16p12.1 genes. **(a)** Quantification and representative images of TUNEL-labeled NPCs in CRISPR activated lines from P1C_007. Scale bars, 70 µm. **(b)** Quantification and representative images of NPCs stained for TUBB3 in CRISPR activated lines from P1C_007. Scale bars, 20 µm. **(c)** Quantification and representative images of TUNEL-labeled NPCs in CRISPR activated lines from P2C_079. Scale bars, 70 µm. **(d)** Plots show enrichment of downregulated (left, blue) and upregulated (right, red) DEGs within signaling pathways ($\log_2|FC|\geq 0.5$, padj <0.05) obtained from RNA-seq of CRISPR activated versus EV lines in probands (M: *MOSMO*; P: *POLR3E*, and U: *UQCRC2*). The size of the circles represents gene counts within each pathway, and the color intensity corresponds to significance of enrichment. Data are mean±s.e.m and one-way ANOVA followed by Dunnett's post hoc test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001, for **(a)**, **(b)**, and **(c)**. EV, empty sgRNA-MS2 vector.



Supplementary Figure 9. Connectivity of secondary variants in response to CRISPR activation of *UQCRC2* within protein-protein interaction networks. (a) Density plots show the connectivity (node degree) of DEGs that overlap with secondary variants ($\log_2|FC|\geq0.5$, padj<0.05) and are reversed with CRISPR activation of *UQCRC2* in three probands, as measured using the STRING database. Anderson-Darling k-sample tests were used to calculate p-values. **(b)** PPI network of DEGs ($\log_2|FC|\geq0.5$, padj<0.05) involved in signaling pathways (from Reactome database) that overlap with secondary variants and were reversed by CRISPRa in the probands. For all networks, $p<10^{-16}$. Protein nodes involved in specific pathways were differently colored using the STRING database. Average node degrees and PPI enrichment p values were also calculated using the STRING database.



Supplementary Figure 10. Expression changes of *FOXG1*, *JUN* and their regulatory network genes in NPCs. (a) Bar plots with mean values show the TPM of gene expression for *FOXG1* and *JUN* across the proband lines (bottom panel). One-way ANOVA followed by Dunnett's post hoc test. **(b)** Venn diagrams show the overlap of genes from DEGs ($\log_2|FC|\geq 0.5$, padj<0.05), ATAC-seq (padj<0.01), and genes within regulatory networks of FOXG1 and JUN as predicted by ChEA3 based on DEGs. Select example genes are also labeled. GL_077 carriers (GMC_077, MC_077 and P1C_077); GL_079 carriers (MC_079 and P2C_079); GL_007 carriers (M1C_07 and P1C_07); healthy donor (HD_01 and HD_02). n=3 independent experiments.

742 **MATERIALS AND METHODS**

743 **Patient recruitment and clinical phenotype analysis.**

744 Patient recruitment and clinical phenotyping were performed as previously described[14]. Families
745 were recruited through clinics by obtaining de-identified samples and clinical data based on an
746 approved protocol (IRB #STUDY00017269). Detailed medical histories, including clinician-
747 reported, guardian-reported (for children), or self-reported (for adults), were collected, and
748 standardized questionnaires were administered to assess developmental phenotypes in children
749 and psychiatric features in adults. Questionnaires for children assessed neuropsychiatric and
750 developmental features, anthropometric measurements, congenital anomalies in multiple organ
751 systems, and family history of medical or psychiatric conditions. For assessing quantitative
752 phenotypes, we conducted quantitative assessment using the Hansen Research Services Matrix
753 Adaptive Test (HRS-MAT)[87] to evaluate nonverbal IQ and the Social Responsiveness Scale
754 (SRS)[88] to assess autism-related social behavior. HRS-MAT was self-administered by
755 participants via an online platform. SRS was delivered through a RedCap-based survey platform
756 maintained by the Geisinger Autism and Developmental Medicine Institute. For participants aged
757 18 or older, the SRS was self-reported; for those under 18, responses were provided by parents or
758 guardians. Body Mass Index (BMI) and head circumference were obtained either from medical
759 records or self/guardian-reports. BMI was calculated from height and weight data obtained from
760 medical records or self/guardian-reports where necessary. Both BMI and head circumference
761 were converted to age- and sex-adjusted z- scores. For the assessment of developmental
762 milestones, we followed the CDC guidelines[89]. For detailed phenotypic data and related
763 information, please contact the corresponding author.

764

765 **Induced pluripotent stem cell reprogramming and maintenance**

766 Induced pluripotent stem cells (iPSCs) were reprogrammed from peripheral blood mononuclear
767 cells using the Cytotune-iPS 2.0 Sendai Programing kit, including the four Yamanaka
768 reprogramming factors POU5F1 (OCT4), SOX2, KLF4, and MYC (ThermoFisher Scientific) as
769 described previously[90], and validated by flow cytometry (EPCAM, TRA-1-81, SSEA4 and
770 CD9) and immunofluorescence (NANOG, OCT4, SOX2 and SSEA4). HD_01 and HD_02 were
771 derived from CD34+ cord blood from healthy male and female donors, respectively (BID00274
772 and BID00271, NINDS). These lines were grown in mTESR1 (#85850 StemCell Technologies)

773 or mTESR plus medium (#100-0276 StemCell Technologies), supplemented with 1%
774 Penicillin/Streptomycin (# P4333 Sigma-Aldrich) on dishes coated with Geltrex (#A1413302
775 Gibco), and were passaged using 0.5mM EDTA (#AM9260G Invitrogen) or ReLeSR (#100-
776 0483, StemCell Technologies), with Rock inhibitor (Y-27632) to improve cell survival (# 72304
777 StemCell Technologies). Cells were incubated at 37ºC and 5% CO_2. Mycoplasma tests were
778 performed on iPSCs and NPCs using mycoplasma PCR detection kit (#MP0035, Sigma-
779 Aldrich).
780
781 **Generation of CRISPR/Cas9-mediated 16p12.1 deletion iPSC line.**
782 The sequences of two sgRNAs are: (1) TCGGTGCTTAGGATCAGCCT, (2)
783 GCCACTAGCTGACATGGTTG. Two separate vectors were used for delivering the sgRNAs:
784 pSpCas9(BB)-2A-Puro (PX459) V2.0 was a gift from Feng Zhang (Addgene plasmid # 62988)
785 and pGH020_sgRNA_G418-GFP was a gift from Michael Bassik (Addgene plasmid # 85405).
786 Two vectors with sgRNAs were transfected into the HD_01 line using Lipofectamine stem
787 transfection reagent (# STEM00001, Invitrogen), and the CRISPR control line was generated via
788 transfection of empty vectors. Puromycin selection (0.5ng/ml) was performed 24 hours after
789 transfection and lasted an additional 24 hours before the selection medium was replaced with
790 mTESR1 medium. The cells recovered for several days before being diluted and split into 96-
791 well plates manually to isolate single-cell clones. Each clone was cultured for two weeks and
792 genotyped by the following two set of primers: WT-F: GACTTCCTCCACATCTTCCTCTA,
793 WT-R: TCAAATAGAGGGGCAGGAGC (546bp); Deletion-F:
794 TCCTCAGACTCAATAATTGCCA, Deletion-R: TGACCTTTACTCTGTGACATTGC (942bp).
795 The desired PCR products were purified, and sanger sequencing was used to confirm that the
796 correct sequence was present. For further conformation, genomic DNA of the selected clones
797 was extracted by GenElute Mammalian Genomic DNA Miniprep kits (#G1N350, Sigma-
798 Aldrich) and PureLink Genomic DNA Mini kit (#K182001, Invitrogen) for SNP-array.
799
800 **RNA isolation and RT-qPCR**
801 Total RNA was isolated using the TRIzol reagent (#15596026, Invitrogen) and PureLink RNA
802 mini kit (#12183018A, invitrogen), following a modified version of this kit's protocol. Cells
803 were collected directly in TRIzol and frozen until RNA extraction. Chloroform was added to the

804 TRIzol in a 1:5 ratio, shaken vigorously for 30 seconds, then incubated for 3 minutes before
805 being centrifuged at 12,500 g for 15 minutes at 4°C. The upper phase was collected in a fresh
806 tube and an equal amount of ethanol was added before vortexing. The mixed reagent was
807 transferred to a spin column, and washing and recovery steps proceeded as described in the kit
808 protocol. Isolated RNA was treated with TURBO DNase (catalog #AM1907, Thermo Fisher
809 scientific, MA, USA). RT-qPCR was performed in three independent experiments, each with
810 three replicates, for each sample. 1000ng RNA was used in reverse transcription reaction by
811 qScript cDNA Synthesis Kit (Quantabio, catalog#95047-100). The cDNA product was diluted
812 100 times and 2ul cDNA was used in 10ul RT-qPCR reaction, using PowerTrack SYBR Green
813 Master Mix (catalog#A46109, Applied Biosystems). RT-qPCR was performed using the
814 QuantStudio 3 (Applied Biosystems) through the following steps: initial denaturation step at
815 95 °C for 2 min, followed by 40 cycles of denaturation at 95 °C for 15 s and
816 annealing/elongation at 60°C for 1 min. A melt curve analysis was conducted at the end of the
817 amplification. The results were analyzed using QuantStudio Design & Analysis software. The
818 following primers were used: GAPDH-F: ATGGGGAAGGTGAAGGTCGG, GAPDH-R:
819 TGACGGTGCCATGGAATTTG; POLR3E-F: GGAGCAGATTGCGCTGAA, POLR3E-R:
820 TTACTGGTGGTCTGGGAAGA; UQCRC2-F: TTCAGCAATTTAGGAACCACCC;
821 UQCRC2-R: GGTCACACTTAATTTGCCACCAA**;** MOSMO-F:
822 CTGTCACATGTGGTTTGCTGG**;** MOSMO-R: GGGCAGCCATACAGAAAAGGA.
823
824 **Neural conversion of iPSCs to NPCs**
825 The neural conversion of iPSCs was performed as previously described with modifications[18]. On
826 day 0, iPSCs were treated with Collagenase Type IV (StemCell Technologies #07909). Then,
827 cells were scraped and transferred to an uncoated 60mm dish with mTESR1 media supplied with
828 10uM Y-27632 for embryoid body (EB) formation. After 48 hours in suspension, mTESR1
829 media was switched to neural differentiation media (N2 media) containing DMEM/F-12
830 (catalog# 10565018, Gibco) with 1% N-2 Supplement (catalog#17502001, Gibco) and 1% MEM
831 Nonessential amino acids (catalog#11140050, Gibco), 2ug/ml Heparin (catalog#07980, StemCell
832 Technologies), 1% Penicillin/Streptomycin (catalog#P4333, Sigma), 5uM SB431542
833 (catalog#72232, StemCell Technologies) and 0.25uM LDN193189 (catalog#72147, stem cell
834 technologies). The EBs were cultured in neural differentiation media for 4 days and the media

835 was changed every two days. On day 6, EBs were seeded on Geltrex-coated plates for 8 days to
836 form rosettes. From days 6- 8, rosettes were cultured in N2 media without Dual-SMAD
837 inhibitors, and from days 8- 14, they were cultured in N2B27 media containing 1:1 DMEM/F-12
838 and Neurobasal Medium (catalog#21103049, Gibco) with GlutaMax supplement
839 (catalog#35050061, Gibco), 1% N-2 Supplement, 2% B-27 supplement minus vitamin A
840 (catalog#12587010, Gibco), 1% MEM Nonessential amino acids, 2ug/ml Heparin, 1%
841 Penicillin/Streptomycin. On day 14, rosettes were treated with Collagenase Type IV and then
842 transferred to uncoated 60mm dish with N2B27 media for neurosphere formation in suspension
843 culture for 12 days. On day 26, neurospheres were dissociated by Accutase (catalog#A1110501,
844 Gibco) then plated on Geltrex-coated plates. Neural progenitor cells were maintained in Stemdiff
845 neural progenitor medium (catalog#05833, StemCell Technologies) and passaged 1:2-1:3 using
846 Accutase. This protocol is expected to create a population of NPCs which express SOX2 and
847 NESTIN, with the majority of cells expressing Pax6.
848
849 **Neuronal differentiation of NPCs to neurons**
850 Plates were prepared for neuronal differentiation with Poly-D-Lysine (PDL, catalog#A3890401,
851 Gibco) and Laminin (catalog#L2020, Sigma-Aldrich). PDL was diluted 1:1 with PBS to a
852 concentration of 50ug/ml and left in the wells for 1 hour at room temperature. PDL was aspirated
853 and the well was washed 4-5 times with sterile DI water and then allowed to dry for 15 minutes.
854 Laminin was diluted to 20ug/ml in chilled dPBS and added to the wells. After characterization,
855 NPCs were plated 800,000 cells/per well on the 6-well plates. The next day, 0.2uM Compound E
856 (catalog#73952, StemCell Technologies) was added to the media. In the first two weeks,
857 neuronal differentiation media was used, containing Neurobasal medium, 1% GlutaMAX, 1%
858 Penicillin/Streptomycin, 1% N-2 Supplement, 2% B-27 supplement (#17504044, Gibco), 0.2%
859 MycoZap Plus-PR (catalog#195263, LONZA) supplied with 20ng/ml BDNF (#450-02-1mg,
860 Pepro Tech), 20ng/ml GDNF (450-10-1mg, Pepro Tech), 200uM Ascorbic acid(#72132,
861 StemCell Technologies), 1uM cAMP (#A9501) and 1ug/ml Laminin. In the last two weeks, the
862 neuronal differentiation media was switched to BrainPhys Neuronal medium (catalog#05790,
863 StemCell Techonologies), supplemented with 40uM 5-Fluoro-2′-deoxyuridine (FUDR)
864 (catalog#F0503, Sigma-Aldrich) to kill mitotic cells. Half media changes were performed every
865 3 days.

866

867 **Cell proliferation and apoptosis**

868 Proliferation and apoptosis assays were adapted from a previous study[18]. On day 0, NPCs at
869 passage 6 were seeded at $1x10^5$ on 18mm diameter coverslips coated in PDL (1:1 in PBS) and
870 Geltrex (1:100 in DMEM/F12) on 12-well cell culture plates in neuronal differentiation medium
871 containing Neurobasal media, 1% N-2 supplement, 2% B-27 supplement minus vitamin A, 1%
872 GlutaMax, 1% Penicillin/Streptomycin, and were incubated overnight. On the morning of day 1,
873 neural progenitor cells were treated with 10uM 5- ethynyl-2'-deoxyuridine (EdU)
874 (catalogC10337, Invitrogen) for four hours and before the media was replaced with fresh
875 medium without Edu. Coverslips were collected and fixed at three timepoints, of no chase (0
876 hours), 24-hour chase, and 48-hour chase. The Edu label was detected via Click-iT chemistry
877 following manufacturers' instructions. At the 24-hour chase timepoint, coverslips were also
878 stained for Ki-67 markers (1:200 catalog#12202, Cell Signaling TECHNOLOGY). On day 5,
879 TUNEL assay for apoptosis detection was performed using Click-iT Plus TUNEL Assay kit
880 (catalog#C10617, Invitrogen).

881

882 **iPSC and NPC immunostaining**

883 NPCs at passage 6 or iPSC clones were seeded on coverslips coated in PDL (1:1 in PBS) and
884 Geltrex (1:100 in DMEM/F12) in 6-well plates. Media was aspirated and cells were washed
885 using PBS three times. Cells were fixed in 4% paraformaldehyde (PFA) (catalog#158127,
886 Sigma-Aldrich) for 15 minutes, washed three times with PBS, then permeabilized in 0.2% Triton
887 X-100 (catalog#X100, Sigma-Aldrich) in PBS for 15 minutes. Blocking was done with a
888 solution of 2% BSA (catalog#A7906, Sigma-Aldrich), 0.2% Triton X-100, and 5% goat serum
889 (catalog# G9023, Sigma-Aldrich) in PBS for 1 hour at room temperature or overnight at 4 °C.
890 The coverslips were incubated with primary antibodies at room temperature for 2 hours or
891 overnight in the 4°C. Coverslips were washed with PBS three times for 5 minutes each, then
892 incubated with the secondary antibody for 2 hours at room temperature. Coverslips were again
893 washed in PBS three times and mounted using DAPI antifade mounting reagent. All antibodies
894 were diluted in blocking buffer. Antibodies: SOX2 (1:200, catalog#3579S Cell Signaling
895 Technology); NESTIN (1:800, catalog#33475S Cell Signaling Technology); NKX2.1 (1:200,
896 catalog#MAB5460, Millipore Sigma); PAX6 (1:800, catalog#60433S, Cell Signaling

897 Technology); TUBB3 (1:400, catalog#5568T, Cell Signaling Technology); FOXG1 (1:400,
898 calalog#29642S, Cell Signaling Technology); c-JUN (1:200, catalog#9165T, Cell Signaling
899 Technology); Ki-67 (1:200, catalog#9129T, Cell signaling Technology); SSEA4 (1:200,
900 catalog#4755T, Cell Signaling Technology); NANOG (1:500, catalog#4893T, Cell signaling
901 Technology); OCT4 (1:200, catalog#2750S, Cell Signaling Technology); Anti-mouse IgG (H+L),
902 F(ab')2 Fragment (1:1000,catalog #4408S, Cell Signaling Technology); Anti-rabbit IgG (H+L),
903 F(ab')2 Fragment (1:1000, catalog#8889S, Cell Signaling Technology).
904
905 **Neuron immunostaining**
906 Neurons were differentiated for 28 days then fixed with 20% PFA added directly to media to
907 reach a final concentration of 4% and incubated at room temperature for 15 minutes. PFA was
908 removed and cells were washed in PBS once for 5 minutes. 0.2% Triton X-100 in PBS was
909 added and incubated room temperature for 15 minutes to permeabilize. Triton X-100 was
910 removed, and cells were blocked at room temperature for 1 hour in 2% BSA, 0.2% Triton X-100,
911 and 5% goat serum in PBS. The coverslips were incubated with primary antibodies at room
912 temperature for 2 hours or overnight in the 4°C refrigerator. Coverslips were washed with PBS
913 two times for 5 minutes each, then incubated with the secondary antibody for 2 hours at room
914 temperature. Coverslips were again washed in PBS 2 times for 5 minutes each, then mounted
915 using DAPI antifade mounting reagent. All antibodies were diluted in blocking buffer.
916 Antibodies: MAP2 (1:500, catalog#4542S, Cell Signaling Technology); VGAT (1:250, Synaptic
917 Systems); VGLUT1 (1:500, Synaptic Systems). NeuN (1:400, catalog#94403S, Cell Signaling
918 Technology); Anti-mouse IgG (H+L), F(ab')2 Fragment (1:1000, catalog #4408S, Cell signaling
919 Technology); Anti-rabbit IgG (H+L), F(ab')2 Fragment (1:1000, catalog#8889S, Cell signaling
920 Technology).
921
922 **Lentivirus packaging**
923 HEK293T cells were grown in media containing DMEM (catalog#D6429, Sigma-Aldrich) and
924 10% FBS (catalog#F2442, Sigma-Aldrich), 1% Penicillin/Streptomycin. On day 0, cells were
925 seeded 18-24 hours prior to transfection to reach 80%-95% confluency on the day of
926 transfection. On day 1, fresh media was added 30 minutes prior to transfection. Transfection was
927 performed using TransIT-Lenti transfection Reagent (catalog#MIR6604, Mirus Bio), Opti-MEM

928 (catalog#31985070, Gibo), and LV-MAX Lentiviral Packaging Mix (catalog#A43237, Thermo
929 Fisher). On day 2, fresh media was added with ViralBoost Reagent (catalog#VB100, ALSTEM
930 CELL ADVANCEMENTS). On day 3, the supernatant was collected and centrifuged at room
931 temperature for 10 min at 300g to pellet debris and then the supernatant was passed through a
932 0.45um filter. Lentivirus Precipitation Solution (catalog#VC100, ALSTEM CELL
933 ADVANCEMENTS) was added to supernatant and a ratio of 1:4, mixed thoroughly, and
934 incubated overnight at 4°C, and then centrifuged for 30 min at 1500g at 4°C. The virus pellet
935 was diluted using mTESR1 medium and stored at -80°C.
936
937 **RNA-seq sample processing and analysis**
938 Each iPSC line was grown in three replicates, which were differentiated separately. RNA was
939 extracted from iPSCs, neural progenitor cells (NPCs; passage 5), and neurons differentiated from
940 NPCs at day 10 (iMNs) and day 28 (MNs). Only samples with RNA integrity number (RIN) ⩾
941 6.0, as measured by the Agilent TapeStation 4200 (Agilent Technologies), and purity (A260/280
942 ratio >1.8) were subjected to RNA sequencing. RNA sequencing libraries were prepared using
943 the NEBNext Ultra II RNA library Prep kit (#E7770S, NEB) for Illumina according to the
944 manufacturer's instructions. 150 base pair (bp) paired-end sequencing were performed using the
945 Illumina NovaSeq 6000 platform by Genewiz from Azenta Life Sciences (South Plainfield, NJ)
946 for 30 million reads. Trimmomatic v0.39[91] were used to trim adapters and to remove low-quality
947 reads (leading:3, trailing:3, slidingwindow:4:15, and minlen:36 parameters). Kallisto v.050.0[92]
948 with n=100 bootstrap samples was used to quantify the abundances of transcripts and hg38 cdna
949 was used to build the Kallisto index. Differential gene expression was analyzed using the
950 DEseq2 v1.44 package in R v.4.3.1[93]. Batch effects were corrected by Combat-seq using sva
951 v.3.35.2 package in R v.4.3.1[94].
952
953 **Detection of alternative isoform usage**
954 Alternative isoform usage was analyzed using the IsoformSwitchAnalyzeR v.2.4.0.[95] package in
955 R. Abundance files from Kallisto were input using the importIsoformExpression function. Genes
956 or isoforms under the cutoff were removed by the prefilter function at default settings. Isoform
957 switches were identified by isoformSwitchTestDEXseq function. To identify alternative isoform

usage events independent of overall gene expression changes, isoforms belonging to genes that were also identified as DEGs were excluded.

Weighed gene co-expression network analysis (WGCNA)

WGCNA was performed using the WGCNA 1.72 package in R[96]. For WGCNA analysis using the transcriptome data of NPC, iMN, and MN in Supplementary Figure 6A-B, first, raw counts were normalized by variance stabilizing transformation (VST) from DEseq2 1.44 package in R. The top 10,000 variable genes were extracted by median absolute deviation (MAD). The power (β) was chosen by pickSoftThreshold function with an R.sq cutoff of 0.8. Signed hybrid co-expression network (TOMType = "unsigned", minModuleSize = 100, mergeCutHeight =0.25) was built by $a_{ij} = [cor(x_i, x_j)]^{\beta}$, for $cor(x_i, x_j) > 0$; $a_{ij} = 0$ for $cor(x_i, x_j) \leq 0$. Module eigengene was calculated as the first principal component to represent the expression profiles in each model. Pearson's correlation coefficient was used to assess correlations between modules and traits.

Whole genome sequencing (WGS) sample processing and analysis

DNA was isolated from blood samples from 22 participants and Illumina TruSeq DNA PCR-free libraries (San Diego, CA, USA) were constructed for 150bp paired-end whole-genome sequencing using Illumina HiSeq X by Macrogen Labs (Rockville, MD, USA). DNA from five samples (HD_01, HD_02, CRISPR deletion, P2C_079, and FNC_079) were isolated from iPSC cell lines using GenElute Mammalian Genomic DNA miniprep kits (#G1N79-1KT SIGMA). The genomic DNA sample was broken into short fragments. These DNA fragments were then end-polished, A-tailed, and ligated with full-length adapters for Illumina sequencing before further size selection. PCR amplification was then conducted unless specified as PCR-free. Purification was performed through the AMPure XP system. The resulting library was assessed on the Agilent Fragment Analyzer System and quantified to 1.5nM through Qubit and qPCR. Whole genome sequencing was conducted using Illumina Novaseq X by Novogene (Durham, NC, USA). Samples were sequenced at an average 37.9X coverage, with an average of 810M reads per sample and 98.0% of reads mapping to the human genome. We followed the GATK Best Practices pipeline[97] to identify single nucleotide variants (SNVs) and small indels using GATK v.4.5.0. GATK's MarkIlluminaAdapters was used to first mark adapter sequences and then align the reads to the GRCh38 reference genome using BWA v.0.7.173[98]. Duplicate reads were

989 removed using the Picard implementation of MarkDuplicates and base quality score recalibration
990 was performed using the BaseRecalibrator and ApplyBQSR. HaplotypeCaller was used to call
991 variants for each sample, merge all variants into a single GVCF, and perform joint genotyping
992 using GenotypeGVCFs. We then used the Python API for Hail and Ensembl Variant Effect
993 Predictor (VEP) v.109[99] to annotate variants. After splitting multi-allelic sites, variants were
994 filtered for those with (i) ⩾90% call rate, (ii) Hardy-Weinberg equilibrium p value ⩾10-15, (iii)
995 read depth ⩾8, and (iv) allele balance ⩾0.2. Variants were further filtered for those with a
996 gnomAD v.2.17[100] frequency <0.1%. We annotated variant consequences based on transcript
997 consequences annotated by VEP and filtered for variants that were LOF ("transcript_ablation",
998 "stop_gained", "frameshift_variant", "stop_lost", "start_lost"), missense ("missense_variant"),
999 splice LOF ("splice_acceptor_variant", "splice_donor_variant"), splice
1000 ("splice_donor_5th_base_variant", "splice_region_variant", "splice_donor_region_variant",
1001 "splice_polypyrimidine_tract_variant"), upstream ("upstream_gene_variant"), downstream
1002 ("downstream_gene_variant"), 5' UTR ("5_prime_UTR_variant"), 3' UTR
1003 ("3_prime_UTR_variant"), or intronic ("intron_variant"). We used dbNSFP8–10 v.4 annotations
1004 to filter missense variants for those predicted to be deleterious by at least five of nine selected
1005 tools (SIFT, LRT, FATHMM, PROVEAN, MetaSVM, MetaLR, PrimateAI, DEOGEN2, and
1006 MutationAssessor). Variants were finally filtered for those private to related individuals to
1007 account for any technical differences between our data and gnomAD. STRs were initially called
1008 using GangSTR v2.5.0[101], followed by quality control and filtering with DumpSTR v6.0.1. STRs
1009 with low-quality calls and those with read depths outside the range of 20 to 1000 were excluded
1010 using the --gangstr-min-call-DP and --gangstr-max-call-DP parameters. Additional quality
1011 control measures included filtering out all reads except spanning and bounding (--gangstr-filter-
1012 spanbound-only) and filtering regions with poorly estimated confidence intervals (--gangstr-
1013 filter-badCI). Individual VCF files were then merged into a multi-sample VCF using MergeSTR
1014 v.6.0.1. To identify rare STRs, loci with an allele frequency of <0.1% in EnsembleTR
1015 (https://github.com/gymrek-lab/EnsembleTR) were initially retained. Summary statistics were
1016 calculated using StatSTR v.6.0.1[102]. For frequency >0.1%, only loci with repeat lengths
1017 exceeding two standard deviations above the population mean were kept. Additionally, rare STRs
1018 defined as those present in fewer than eight individuals in the iPSC cohort were selected for
1019 downstream analysis. Final STR annotations were performed using ANNOVAR[103]. CNVs were

called using CNVpytor v.1.3.1[104] with bin size 500. Rare CNVs were defined as those with a population frequency of <5% based on gnomAD v.4.1. CNV annotations were performed manually, consistent with annotations generated by VEP and ANNOVAR, considering only upstream, downstream, 3' UTR, 5' UTR, exonic, and intronic regions.

Analysis of combined effects on gene expression

We assessed for combined effects by first extracting differentially expressed genes (padj<0.05) by DESeq2 in individuals carrying both 16p12.1 deletion and the specific secondary variant compared to all other individuals in the same family. Then, rare variants in those DEGs were extracted and filtered by those which were inherited in the family. Next, two-way ANOVA (using the statsmodels.formula.api and statsmodels.api modules in Python v.3.11.7) was performed using TPM (Transcripts Per Million) of genes annotated from the filtered variants. Genes and variants were kept only when the interaction p value (16p12.1 deletion: secondary hit", "PR(>F)") was <0.05 using two-way ANOVA tests. The intersection of the variants with Diffpeak regions was identified using Pybedtools[105]v.0.12.0.

ATAC-seq sample processing and analysis

Live cell samples were collected and run through a 30 µm filter to obtain a single cell suspension before cryopreservation. Thawed cells were washed and treated with DNAse I (Life Tech, Cat. #EN0521) to remove genomic DNA contamination. Live cell samples were quantified and assessed for viability using a Countess Automated Cell Counter (ThermoFisher Scientific, Waltham, MA, USA). After cell lysis and cytosol removal, nuclei were treated with Tn5 enzyme (Illumina, Cat. #20034197) for 30 minutes at 37°C and purified with Minelute PCR Purification Kit (Qiagen, Cat. #28004) to produce tagmented DNA samples. Tagmented DNA was barcoded with Nextera Index Kit v2 (Illumina, Cat. #FC-131-2001) and amplified via PCR prior to a SPRI Bead cleanup to obtain purified DNA libraries. The sequencing libraries were multiplexed and clustered onto a flow cell on the Illumina NovaSeq X and NovaSeq 6000 according to manufacturer's instructions by Azenta Life Sciences (South Plainfield, NJ, USA). The samples were sequenced using a 2x150bp Paired End (PE) configuration. Image analysis and base calling were conducted by the NovaSeq Control Software (NCS). Raw sequence data (.bcl files) generated from Illumina NovaSeq was converted into fastq files and de-multiplexed using

1051 Illumina bcl2fastq 2.20 software. One mismatch was allowed for index sequence identification.
1052 Quality control and processing of fastq sequencing files were performed using the ENCODE
1053 ATAC-seq Data Standards and Processing Pipeline (https://github.com/ENCODE-DCC/atac-seq-
1054 pipeline). DESeq2 was used to identify differential peaks. The cutoff for adjusted p value (padj)
1055 was 0.01. Motif enrichment analysis was performed using the findMotifsGenome.pl from the
1056 HOMER[106] on regions of Diffpeaks.
1057
1058 **ATAC-seq peak Annotation**
1059 We annotated ATAC-seq peaks using the union of three approaches: ChIPseeker[107] the Activity-
1060 by-Contact (ABC) model[108], and nearest gene assignment. ChIPseeker analysis was performed
1061 via Galaxy (version 1.28.3), using GENCODE Release 46 (GRCh38.p14) as the reference
1062 genome annotation. The ABC model was applied using RNA-seq and ATAC-seq data from
1063 HD_01 and HD_02 at the iPSC and NPC stages. Publicly available Hi-C and H3K27ac ChIP-seq
1064 datasets were used as additional inputs for contact estimation and enhancer activity scoring[109,110].
1065 Expressed genes were defined as those with expression >1 TPM and promoter activity
1066 quantile >0.4. For each chromosome, enhancer–gene pairs within 5 Mb of a transcription start
1067 site (TSS) were scored, and those with an ABC score above 0.022 (default threshold) were
1068 retained as predicted regulatory interactions. Peak-to-gene associations were also established by
1069 linking the nearest gene to peaks using the closest function from the Pybedtools.
1070
1071 **CRISPRa experiments**
1072 iPSCs derived from probands were first infected with lenti dCas9-VP64_Blast and lenti MS2-
1073 P65-HSF1_Hygro. Lenti dCAS-VP64_Blast (Addgene plasmid # 61425;
1074 http://n2t.net/addgene:61425; RRID: Addgene_61425) and lenti MS2-P65-HSF1_Hygro
1075 (Addgene plasmid # 61426; http://n2t.net/addgene:61426; RRID: Addgene_61426)) were both
1076 gifts from Feng Zhang. Then, cells were infected with lenti sgRNA (MS2) _Zeo. The EV line
1077 was infected with only the backbone. To insert sgRNA, the backbone was digested with BsmBI-
1078 v2 (NEB) and ligation was performed using quick ligase (NEB #M2200S) with phosphorylated
1079 and annealed sgRNA. Lenti sgRNA (MS2) _Zeo backbone was a gift from Feng Zhang
1080 (Addgene plasmid # 61427; http://n2t.net/addgene:61427; RRID: Addgene_61427). sgRNA
1081 sequence: *POLR3E*: CACGGCCTGCATGAATGGCG; *MOSMO*:

1082 GAGCCGGGAGGACGGAGCTG; *UQCRC2*: ATAAAGAGAGCAGTAGAGCG. All the
1083 transductions were performed with 4μg/ml polybrene to improve efficiency. After 24hr
1084 incubation, the medium was switched to fresh culture medium and incubated for another 48hr.
1085 Selection was then performed at the iPSC stage using 10ng/ml Blasticidin, 250ng/ml
1086 Hygromycin, 250ng/ml Zeocin until no further cell death was observed. Selection was repeated
1087 using a half dose concentration from day 10 to day 12 of neural rosette formation during the
1088 NPC differentiation protocol. Comparing CRISPRa lines to their corresponding EV controls
1089 allowed us to identify DEGs resulting from the restoration of specific deletion genes in an
1090 isogenic background. These DEGs were subsequently used in downstream functional analyses.
1091 Additionally, we defined "reversed genes" as those DEGs from CRISPRa experiments that
1092 exhibited changes in the opposite direction compared to DEGs identified in the proband line
1093 compared to nondeletion sibling or healthy donor lines.
1094

1095 **PPI network analysis**

1096 Protein–protein interaction (PPI) network analysis was performed using the STRING database
1097 with default settings (network type: full STRING network; required confidence score: ⩾0.400;
1098 FDR stringency: medium, 5%). To create the density plot, the active interaction sources were text
1099 mining, experiments, databases, co-expression, neighborhood, gene fusion, co-occurrence.
1100 Anderson-Darling k-sample test was performed by ad.test function in kSamples v1.2-10 package
1101 in R. To visualize the proteins involved in signal transduction based on Reactome data, active
1102 interaction sources were experiments, databases and co-expression. Average node degree, PPI
1103 enrichment p-value and functional enrichment were obtained from the STRING database.
1104

1105 **Enrichment analysis**

1106 We performed over-representation analysis (ORA) and gene set enrichment analysis (GSEA)
1107 using clusterProfilter v.4.12.6 package in R, following the described guidelines[111]. For ORA, we
1108 used enrichGO, entichKEGG, and enrichPathway functions. For GSEA, analysis was performed
1109 against curated gene sets obtained using the msigdbr v.10.0.1 package in R. Pathway terms
1110 belonging to Reactome, WP, and KEGG databases were extracted in an unbiased manner. For
1111 generation of the waterfall plots in Figure 3G and Supplementary Figure 6D, the top 30 pathways
1112 were unbiasedly selected from each database at the iPSC, NPC, iMN, and MN stages. These

1113 pathways were then pooled, and the top 100 were selected based on q-value. If the number of
1114 significantly enriched terms was fewer than 30 at each stage or fewer than 100 in total after
1115 pooling, all enriched terms were included for generating waterfall plots. Enrichment analysis of
1116 DEGs in published gene lists was performed using the GeneOverlap v.1.36.0[112] package in R.
1117 Enrichment analysis of DEGs in DisGeNET was performed using enrichR v.3.4[113] package in R.
1118

1119 **Construction of TF regulatory networks for JUN and FOXG1**

1120 We used the ChEA3 (https://maayanlab.cloud/chea3/) web application to construct TF–TF co-
1121 regulatory networks using the top 15 TFs whose binding motifs were significantly enriched in
1122 Diffpeak regions from deletion lines versus HD lines comparisons at the NPC stage. In
1123 comparisons where fewer than 15 motifs were enriched, only motifs with p-value less than 0.01
1124 were included. FOXG1 and JUN ranked highest among the enriched transcription factors based
1125 on the Integrated Scaled Rank. TF-TF co-regulatory networks were constructed by ChEA3 using
1126 the top results, and edges between TFs defined by the evidence from libraries which supports the
1127 interactions. To identify regulatory network genes for each TF, we input DEGs from each
1128 comparison into ChEA3 and used the "Overlapping Genes" column. The top integrated rank
1129 across ChEA3 libraries was used in our analysis.
1130

1131 **Imaging and quantification**

1132 Revolve (Discover Echo) was used to detect Click-iT signals. LSM 800 upright and LSM 800
1133 inverted (ZEISS) were used to detect fluorescence signals for immunofluorescence assays.
1134 Images were analyzed using Fiji (ImageJ).
1135

1136 **Statistics and reproducibility**

1137 Investigators were blinded to experimental condition when quantifying fluorescence intensity.
1138 Comparisons between deletion carrier and noncarrier lines within the same family were used
1139 when possible to control for the independent effects of shared secondary variants and to
1140 specifically analyze the joint effects caused by both the 16p12.1 deletion and secondary variants.
1141 Comparisons with healthy donor lines provided a consistent baseline across families and served
1142 as appropriate controls for cellular assays and functional analysis, as many noncarrier lines also
1143 exhibited clinical features due to the underlying genetic liability. The same bioinformatics

pipelines were applied across all conditions for each assay. Statistical analyses were conducted using R v4.3.1 and Python v3.11.7. Details of the statistical tests and number of replicates are provided in the corresponding figure legends and Supplementary Tables.

Acknowledgments

The authors thank Dr. Yingwei Mao for valuable advice on iPSC culture and Dr. Melissa Rolls for providing imaging resources. We are grateful to the National Institute of Neurological Disorders and Stroke (NINDS) for supplying iPSC lines derived from healthy donors. The authors thank Dr. Matthew Jensen, Dr. Francisca Canzar and Johnathan Ray for their assistance and insights in data analysis. The following funding supported this study: NIH R01GM121907 and R21NS122398 (to S.G.)

Author contributions

J.S. and S.G. conceived of and designed the study. S.G. supervised the experiments and analyses. J.S., S.N. and M.D. maintained iPSC culture, performed the neural conversion and extracted RNA. J.S., M.D. and A.P. generated the CRISPR-edited lines. J.S. and S.N. performed immunostaining, imaging and image quantification. J.S., and V.H.B. performed assays on cell proliferation and apoptosis. J.S. designed CRISPRa experiments and J.S., V.H.B. performed CRISPRa experiments. J.S., D.B., C.S., and B.G. performed analyses on ATAC-seq, RNA-seq and WGS. C.S. summarized the clinical phenotype information. D.A. and K.P. recruited 16p12.1 deletion families. P.L. reprogrammed iPSCs from PBMCs derived from recruited patients. J.S., S.N. and S.G. wrote the paper.

Data and code availability

Whole genome sequencing, RNA-seq and ATAC-seq generated in this study are available at NCBI dbGaP phs002403.v1.p1. All code generated for this project, including pipelines for running bioinformatic software and custom analysis scripts, are available at https://github.com/Jiawan1023/iPSC_integrated_framework. Detailed information of statistical analyses is available in Supplementary Tables.

Supplementary Table 1:

Lists of DEGs and altered isoforms with secondary variants in the isogenic setting; Detailed results of Fisher's exact tests in Figure 1C and volcano plots in Supplementary Figure 2A.

Supplementary Table 2:

Lists of Diffpeaks with secondary variants in the isogenic setting; Detailed results of Fisher's exact tests in Figure 1D and volcano plots in Supplementary Figure 2B.

Supplementary Table 3:

Lists of TPM values of combined effects on gene expression across iPSC, NPC, iMN and MN stages in the familial setting; Detailed results for Figure 1C.

Supplementary Table 4:

Lists of DEGs, reversed genes, GSEA results and ORA results in CRISPRa experiments.

Supplementary Table 5:

Lists of DEGs with secondary variants, and DEGs with secondary variants involved in signal transductions for Figure 5.

Supplementary Table 6:

Lists of enriched motifs at the NPC stage, DEGs and Diffpeaks for Supplementary Figure 10B and ChEA3 result.

Supplementary Table 7:

Lists of detailed results for WGCNA.

Supplementary Table 8:

Lists of DESeq2 results and GSEA results for each comparison, enrichment results using published datasets and DisGeNET.

Supplementary Table 9:

Summary of all statistical tests.

1197 1. Sun, J., Noss, S., Banerjee, D., Das, M. & Girirajan, S. Strategies for dissecting the
1198 complexity of neurodevelopmental disorders. *Trends Genet* **40**, 187–202 (2024).

1199 2. Kingdom, R., Beaumont, R. N., Wood, A. R., Weedon, M. N. & Wright, C. F. Genetic
1200 modifiers of rare variants in monogenic developmental disorder loci. *Nat Genet* **56**, 861–868
1201 (2024).

1202 3. Girirajan, S. *et al.* Phenotypic heterogeneity of genomic disorders and rare copy-number
1203 variants. *N Engl J Med* **367**, 1321–1331 (2012).

1204 4. Bassett, A. S. & Chow, E. W. C. Schizophrenia and 22q11.2 deletion syndrome. *Curr*
1205 *Psychiatry Rep* **10**, 148–157 (2008).

1206 5. McDonald-McGinn, D. M. *et al.* 22q11.2 deletion syndrome. *Nat Rev Dis Primers* **1**, 15071
1207 (2015).

1208 6. Girirajan, S. *et al.* A recurrent 16p12.1 microdeletion supports a two-hit model for severe
1209 developmental delay. *Nat Genet* **42**, 203–209 (2010).

1210 7. Collins, R. L. *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**,
1211 3041-3055.e25 (2022).

1212 8. Rees, E. *et al.* Analysis of Intellectual Disability Copy Number Variants for Association With
1213 Schizophrenia. *JAMA Psychiatry* **73**, 963–969 (2016).

1214 9. Auwerx, C. *et al.* Rare copy-number variants as modulators of common disease susceptibility.
1215 *Genome Med* **16**, 5 (2024).

1216 10. Stefansson, H. *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in
1217 controls. *Nature* **505**, 361–366 (2014).

1218 11. Montanucci, L. *et al.* Genome-wide identification and phenotypic characterization of
1219 seizure-associated copy number variations in 741,075 individuals. *Nat Commun* **14**, 4392
1220 (2023).

1221 12. Rees, E. *et al.* CNV analysis in a large schizophrenia sample implicates deletions at
1222 16p12.1 and SLC1A1 and duplications at 1p36.33 and CGNL1. *Hum Mol Genet* **23**, 1669–
1223 1676 (2014).

1224 13. Pizzo, L. *et al.* Rare variants in the genetic background modulate cognitive and
1225 developmental phenotypes in individuals carrying disease-associated variants. *Genet Med* **21**,
1226 816–825 (2019).

1227 14. Jensen, M. *et al.* Genetic modifiers and ascertainment drive variable expressivity of
1228 complex disorders. 2024.08.27.24312158 Preprint at
1229 https://doi.org/10.1101/2024.08.27.24312158 (2024).

1230 15. Qaiser, F. *et al.* Rare and low frequency genomic variants impacting neuronal functions
1231 modify the Dup7q11.23 phenotype. *Orphanet J Rare Dis* **16**, 6 (2021).

1232 16. Taylor, C. M. *et al.* Phenotypic shift in copy number variants: Evidence in 16p11.2
1233 duplication syndrome. *Genet Med* **25**, 151–154 (2023).

1234 17. Ardhanareeswaran, K., Mariani, J., Coppola, G., Abyzov, A. & Vaccarino, F. M. Human
1235 induced pluripotent stem cells for modelling neurodevelopmental disorders. *Nat Rev Neurol*
1236 **13**, 265–278 (2017).

1237 18. Deshpande, A. *et al.* Cellular Phenotypes in Human iPSC-Derived Neurons from a
1238 Genetic Model of Autism Spectrum Disorder. *Cell Reports* **21**, 2678–2687 (2017).

1239 19. Wang, M. *et al.* Increased Neural Progenitor Proliferation in a hiPSC Model of Autism
1240 Induces Replication Stress-Associated Genome Instability. *Cell Stem Cell* **26**, 221-233.e6
1241 (2020).

1242 20. Nehme, R. *et al.* The 22q11.2 region regulates presynaptic gene-products linked to
1243 schizophrenia. *Nat Commun* **13**, 3690 (2022).

1244 21. Habela, C. W., Song, H. & Ming, G.-L. Modeling synaptogenesis in schizophrenia and
1245 autism using human iPSC derived neurons. *Mol Cell Neurosci* **73**, 52–62 (2016).

1246 22. Schafer, S. T. *et al.* Pathological priming causes developmental gene network
1247 heterochronicity in autistic subject-derived neurons. *Nat Neurosci* **22**, 243–255 (2019).

1248 23. Hockemeyer, D. & Jaenisch, R. Induced Pluripotent Stem Cells Meet Genome Editing.
1249 *Cell Stem Cell* **18**, 573–586 (2016).

1250 24. Peña, C. J. *et al.* Early life stress alters transcriptomic patterning across reward circuitry
1251 in male and female mice. *Nat Commun* **10**, 5098 (2019).

1252 25. Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic
1253 basis of autism spectrum disorder. *Nat Neurosci* **19**, 1454–1462 (2016).

1254 26. Parikshak, N. N. *et al.* Integrative functional genomic analyses implicate specific
1255 molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).

1256 27. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in
1257 schizophrenia. *Nature* **604**, 502–508 (2022).

1258 28. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102
1259 independent variants and highlights the importance of the prefrontal brain regions. *Nat*
1260 *Neurosci* **22**, 343–352 (2019).

1261 29. Pfisterer, U. *et al.* Identification of epilepsy-associated neuronal subtypes and gene
1262 expression underlying epileptogenesis. *Nat Commun* **11**, 5038 (2020).

1263 30. Jensen, M. *et al.* Combinatorial patterns of gene expression changes contribute to
1264 variable expressivity of the developmental delay-associated 16p12.1 deletion. *Genome Med*
1265 **13**, 163 (2021).

1266 31. Backman, J. D. *et al.* Exome sequencing and analysis of 454,787 UK Biobank
1267 participants. *Nature* **599**, 628–634 (2021).

1268 32. Campbell, R. F., McGrath, P. T. & Paaby, A. B. Analysis of Epistasis in Natural Traits
1269 Using Model Organisms. *Trends Genet* **34**, 883–898 (2018).

1270 33. Queitsch, C., Carlson, K. D. & Girirajan, S. Lessons from model organisms: phenotypic
1271 robustness and missing heritability in complex disease. *PLoS Genet* **8**, e1003041 (2012).

1272 34. Gibson, G. Decanalization and the origin of complex disease. *Nat Rev Genet* **10**, 134–140
1273 (2009).

1274 35. Liu, Y. *et al.* Neuronal IFN-beta-induced PI3K/Akt-FoxA1 signalling is essential for
1275 generation of FoxA1+Treg cells. *Nat Commun* **8**, 14709 (2017).

1276 36. Bu, C. *et al.* c-Myb protects cochlear hair cells from cisplatin-induced damage via the
1277 PI3K/Akt signaling pathway. *Cell Death Discov* **8**, 78 (2022).

1278 37. Li, Y. *et al.* c-Myb Enhances Breast Cancer Invasion and Metastasis through the Wnt/β-
1279 Catenin/Axin2 Pathway. *Cancer Res* **76**, 3364–3375 (2016).

1280 38. Liu, C., Oikonomopoulos, A., Sayed, N. & Wu, J. C. Modeling human diseases with
1281 induced pluripotent stem cells: from 2D to 3D and beyond. *Development* **145**, dev156166
1282 (2018).

1283 39. Zhang, S. *et al.* Multiple genes in a single GWAS risk locus synergistically mediate
1284 aberrant synaptic development and function in human neurons. *Cell Genom* **3**, 100399 (2023).

1285 40. Schrode, N. *et al.* Synergistic effects of common schizophrenia risk variants. *Nat Genet*
1286 **51**, 1475–1485 (2019).

1287 41. Mariani, J. *et al.* FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron
1288 Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375–390 (2015).

1289 42. Ernst, C. Proliferation and Differentiation Deficits are a Major Convergence Point for
1290 Neurodevelopmental Disorders. *Trends Neurosci* **39**, 290–299 (2016).

1291 43. Schafer, S. T. *et al.* Pathological priming causes developmental gene network
1292 heterochronicity in autistic subject-derived neurons. *Nat Neurosci* **22**, 243–255 (2019).

1293 44. Jaaro-Peled, H. *et al.* Neurodevelopmental mechanisms of schizophrenia: understanding
1294 disturbed postnatal brain maturation through neuregulin-1-ErbB4 and DISC1. *Trends Neurosci*
1295 **32**, 485–495 (2009).

1296 45. Iyer, J. *et al.* Pervasive genetic interactions modulate neurodevelopmental defects of the
1297 autism-associated 16p11.2 deletion in Drosophila melanogaster. *Nat Commun* **9**, 2548 (2018).

1298 46. Pizzo, L. *et al.* Functional assessment of the 'two-hit' model for neurodevelopmental
1299 defects in Drosophila and X. laevis. *PLoS Genet* **17**, e1009112 (2021).

1300 47. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-
1301 Cas9 complex. *Nature* **517**, 583–588 (2015).

1302 48. Ramanathan, A. *et al.* A mutation in POLR3E impairs antiviral immune response and
1303 RNA polymerase III. *Proc Natl Acad Sci U S A* **117**, 22113–22121 (2020).

1304 49. Kong, J. H. *et al.* Gene-teratogen interactions influence the penetrance of birth defects by
1305 altering Hedgehog signaling strength. *Development* **148**, dev199867 (2021).

1306 50. Gaignard, P. *et al.* UQCRC2 mutation in a patient with mitochondrial complex III
1307 deficiency causing recurrent liver failure, lactic acidosis and hypoglycemia. *J Hum Genet* **62**,
1308 729–731 (2017).

1309 51. Jing, J. *et al.* Hedgehog signaling in tissue homeostasis, cancers, and targeted therapies.
1310 *Signal Transduct Target Ther* **8**, 315 (2023).

1311 52. Pusapati, G. V. *et al.* CRISPR Screens Uncover Genes that Regulate Target Cell
1312 Sensitivity to the Morphogen Sonic Hedgehog. *Dev Cell* **44**, 113-129.e8 (2018).

1313 53. Palma, V. *et al.* Sonic hedgehog controls stem cell behavior in the postnatal and adult
1314 brain. *Development* **132**, 335–344 (2005).

1315 54. Sharma, J. *et al.* Neuronatin-mediated Aberrant Calcium Signaling and Endoplasmic
1316 Reticulum Stress Underlie Neuropathology in Lafora Disease *. *Journal of Biological*
1317 *Chemistry* **288**, 9482–9490 (2013).

1318 55. Furness, D. N. *et al.* Progressive hearing loss and gradual deterioration of sensory hair
1319 bundles in the ears of mice lacking the actin-binding protein Eps8L2. *Proceedings of the*
1320 *National Academy of Sciences* **110**, 13898–13903 (2013).

1321 56. Zhubi, A. *et al.* Increased binding of MeCP2 to the GAD1 and RELN promoters may be
1322 mediated by an enrichment of 5-hmC in autism spectrum disorder (ASD) cerebellum. *Transl*
1323 *Psychiatry* **4**, e349–e349 (2014).

1324 57. Jansen, L. A. *et al.* PI3K/AKT pathway mutations cause a spectrum of brain
1325 malformations from megalencephaly to focal cortical dysplasia. *Brain* **138**, 1613–1628
1326 (2015).

1327 58. Kwan, V., Unda, B. K. & Singh, K. K. Wnt signaling networks in autism spectrum
1328 disorder and intellectual disability. *J Neurodev Disord* **8**, 45 (2016).

1329 59. Meyers, E. A. & Kessler, J. A. TGF-β Family Signaling in Neural and Neuronal
1330 Differentiation, Development, and Function. *Cold Spring Harb Perspect Biol* **9**, a022244
1331 (2017).

1332 60. Duronio, R. J. & Xiong, Y. Signaling pathways that control cell proliferation. *Cold Spring*
1333 *Harb Perspect Biol* **5**, a008904 (2013).

1334 61. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks
1335 and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*
1336 **51**, D638–D646 (2023).

1337 62. Hollenbeck, P. J. Mitochondria and neurotransmission: evacuating the synapse. *Neuron*
1338 **47**, 331–333 (2005).

1339 63. Zenisek, D. & Matthews, G. The role of mitochondria in presynaptic calcium handling at
1340 a ribbon synapse. *Neuron* **25**, 229–237 (2000).

1341 64. Weidemüller, P., Kholmatov, M., Petsalaki, E. & Zaugg, J. B. Transcription factors:
1342 Bridge between cell signaling and gene regulation. *Proteomics* **21**, e2000034 (2021).

1343 65. Perkins, K. K., Admon, A., Patel, N. & Tjian, R. The Drosophila Fos-related AP-1 protein
1344 is a developmentally regulated transcription factor. *Genes Dev* **4**, 822–834 (1990).

1345 66. Freitas, R., Gómez-Marín, C., Wilson, J. M., Casares, F. & Gómez-Skarmeta, J. L.
1346 Hoxd13 contribution to the evolution of vertebrate appendages. *Dev Cell* **23**, 1219–1229
1347 (2012).

1348 67. Nowyhed, H. N. *et al.* The nuclear receptor nr4a1 controls CD8 T cell development
1349 through transcriptional suppression of runx3. *Sci Rep* **5**, 9059 (2015).

1350 68. Rahman, M. S., Akhtar, N., Jamil, H. M., Banik, R. S. & Asaduzzaman, S. M. TGF-
1351 β/BMP signaling and other molecular events: regulation of osteoblastogenesis and bone
1352 formation. *Bone Res* **3**, 15005 (2015).

1353 69. Liao, G. *et al.* Jun NH2-terminal kinase (JNK) prevents nuclear beta-catenin
1354 accumulation and regulates axis formation in Xenopus embryos. *Proc Natl Acad Sci U S A*
1355 **103**, 16313–16318 (2006).

1356 70. Robertson, F. L. *et al.* Elevated FOXG1 in glioblastoma stem cells cooperates with
1357 Wnt/β-catenin to induce exit from quiescence. *Cell Rep* **42**, 112561 (2023).

1358 71. Boyd, J. L. *et al.* Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle
1359 dynamics in the developing neocortex. *Curr Biol* **25**, 772–779 (2015).

1360 72. Maeda, Y., Isomura, A., Masaki, T. & Kageyama, R. Differential cell-cycle control by
1361 oscillatory versus sustained Hes1 expression via p21. *Cell Rep* **42**, 112520 (2023).

1362 73. Narla, G. *et al.* KLF6, a candidate tumor suppressor gene mutated in prostate cancer.
1363 *Science* **294**, 2563–2566 (2001).

1364 74. Yang, N. *et al.* Generation of pure GABAergic neurons by transcription factor
1365 programming. *Nat Methods* **14**, 621–628 (2017).

1366 75. Zhang, X. *et al.* Local and global chromatin interactions are altered by large genomic
1367 deletions associated with human brain development. *Nat Commun* **9**, 5356 (2018).

1368 76. Zhang, S. *et al.* Network Effects of the 15q13.3 Microdeletion on the Transcriptome and
1369 Epigenome in Human-Induced Neurons. *Biol Psychiatry* **89**, 497–509 (2021).

1370 77. Wada, M. *et al.* Dopaminergic dysfunction and excitatory/inhibitory imbalance in
1371 treatment-resistant schizophrenia and novel neuromodulatory treatment. *Mol Psychiatry* **27**,
1372 2950–2967 (2022).

1373 78. Kumar, S. *et al.* Impaired neurodevelopmental pathways in autism spectrum disorder: a
1374 review of signaling mechanisms and crosstalk. *J Neurodev Disord* **11**, 10 (2019).

1375 79. Kalkman, H. O. A review of the evidence for the canonical Wnt pathway in autism
1376 spectrum disorders. *Mol Autism* **3**, 10 (2012).

1377 80. Gao, R. & Penzes, P. Common mechanisms of excitatory and inhibitory imbalance in
1378 schizophrenia and autism spectrum disorders. *Curr Mol Med* **15**, 146–167 (2015).

1379 81. Nussinov, R., Tsai, C.-J. & Jang, H. Neurodevelopmental disorders, immunity, and cancer
1380 are connected. *iScience* **25**, 104492 (2022).

1381 82. Willsey, A. J. *et al.* Coexpression networks implicate human midfetal deep cortical
1382 projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).

1383 83. Naqvi, S. *et al.* Precise modulation of transcription factor levels identifies features
1384 underlying dosage sensitivity. *Nat Genet* **55**, 841–851 (2023).

1385 84. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat*
1386 *Methods* **14**, 297–301 (2017).

1387 85. Replogle, J. M. *et al.* Combinatorial single-cell CRISPR screens by direct guide RNA
1388 capture and targeted sequencing. *Nat Biotechnol* **38**, 954–961 (2020).

1389 86. Pașca, S. P. *et al.* A framework for neural organoids, assembloids and transplantation
1390 studies. *Nature* **639**, 315–320 (2025).

1391 87. Hansen, J. A. Development and Psychometric Evaluation of the Hansen Research
1392 Services Matrix Adaptive Test: A Measure of Nonverbal IQ. *J Autism Dev Disord* **49**, 2721–
1393 2732 (2019).

1394 88. Constantino, J. N. Social Responsiveness Scale. in *Encyclopedia of Autism Spectrum*
1395 *Disorders* 2919–2929 (Springer, New York, NY, 2013). doi:10.1007/978-1-4419-1698-3_296.

1396 89. Zubler, J. & Whitaker, T. CDC's Revised Developmental Milestone Checklists. *Am Fam*
1397 *Physician* **106**, 370–371 (2022).

1398 90. Bozaoglu, K. *et al.* Generation of seven iPSC lines from peripheral blood mononuclear
1399 cells suitable to investigate Autism Spectrum Disorder. *Stem Cell Res* **39**, 101516 (2019).

1400 91. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
1401 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

1402 92. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq
1403 quantification. *Nat Biotechnol* **34**, 525–527 (2016).

1404 93. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
1405 for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

1406 94. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for
1407 RNA-seq count data. *NAR Genom Bioinform* **2**, lqaa078 (2020).

1408 95. Vitting-Seerup, K. & Sandelin, A. IsoformSwitchAnalyzeR: analysis of changes in
1409 genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*
1410 **35**, 4469–4471 (2019).

1411 96. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network
1412 analysis. *BMC Bioinformatics* **9**, 559 (2008).

1413 97. Van der Auwera, G. A. *et al.* From FastQ data to high confidence variant calls: the
1414 Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**, 11.10.1-
1415 11.10.33 (2013).

1416 98. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
1417 transform. *Bioinformatics* **25**, 1754–1760 (2009).

1418 99. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol* **17**, 122 (2016).

1419 100. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**,
1420 285–291 (2016).

1421 101. Mousavi, N., Shleizer-Burko, S., Yanicky, R. & Gymrek, M. Profiling the genome-wide
1422 landscape of tandem repeat expansions. *Nucleic Acids Res* **47**, e90 (2019).

1423 102. Mousavi, N. *et al.* TRTools: a toolkit for genome-wide analysis of tandem repeats.
1424 *Bioinformatics* **37**, 731–733 (2021).

1425 103. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic
1426 variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).

1427 104. Suvakov, M., Panda, A., Diesh, C., Holmes, I. & Abyzov, A. CNVpytor: a tool for copy
1428 number variation detection and analysis from read depth and allele imbalance in whole-
1429 genome sequencing. *Gigascience* **10**, giab074 (2021).

1430 105. Dale, R. K., Pedersen, B. S. & Quinlan, A. R. Pybedtools: a flexible Python library for
1431 manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).

1432 106. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
1433 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576–589
1434 (2010).

1435 107. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak
1436 annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

1437 108. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from
1438 thousands of CRISPR perturbations. *Nat Genet* **51**, 1664–1669 (2019).

1439 109. Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation.
1440 *Nature* **518**, 331–336 (2015).

1441 110. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic
1442 stem cells. *Cell* **153**, 1134–1148 (2013).

1443 111. Xu, S. *et al.* Using clusterProfiler to characterize multiomics data. *Nat Protoc* **19**, 3292–
1444 3320 (2024).

1445 112. GeneOverlap. *Bioconductor* http://bioconductor.org/packages/GeneOverlap/.

1446 113. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment
1447 analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

1448