

Characterization of structural variants in KOLF2.1J iPSCs and their potential impact on forebrain organoid differentiation

Lízia Branco,^{1,11} Zeynep Yentür,^{2,3,4,5,11} Madison James Yang,^{6,11} Parithi Balachandran,^{7,11} Kamilla Sedov,⁶ G. Kumar,⁶ Faria Zafar,⁶ Theresa Kagermeier,^{1,3,5} Kseniia Sarieva,^{3,4,5} Katharina Becker,³ William C. Skarnes,⁷ Birgitt Schüle,^{6,*} Simone Mayer,^{1,2,3,10,*} and Christine R. Beck^{7,8,9,12,*}

¹Zoological Institute, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

²The Heidelberg Academy of Sciences and Humanities, Heidelberg, Germany

³Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany

⁴International Max Planck Research School, Graduate Training Centre of Neuroscience, University of Tübingen, Tübingen, Germany

⁵Graduate Training Centre of Neuroscience, University of Tübingen, Tübingen, Germany

⁶Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

⁷The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

⁸Department of Genetics and Genome Sciences, University of Connecticut Health Center, Farmington, CT, USA

⁹Institute for Systems Genomics, University of Connecticut, Storrs, CT, USA

¹⁰Institute of Biological and Chemical Systems - Functional Molecular Systems, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

¹¹These authors contributed equally

¹²Lead contact

*Correspondence: bschuele@stanford.edu (B.S.), simone.mayer@kit.edu (S.M.), christine.beck@jax.org (C.R.B.)

<https://doi.org/10.1016/j.stemcr.2026.103112>

SUMMARY

The KOLF2.1J human induced pluripotent stem cell (iPSC) line was created as a reference for neurological disease modeling. Previous genotyping efforts identified several large structural variants (SVs), raising questions about their potential impact on neuronal differentiation. Here, we identified 37 additional SVs using long-read sequencing and optical genome mapping. Additionally, we found that one of the known SVs contained a previously undetected overlap with *RYBP*, a neurodevelopmental gene. To assess the functional consequences of these SVs, we performed single-cell RNA sequencing of day 90 forebrain organoids derived from KOLF2.1J and two independent control iPSC lines. While cell-type composition was preserved, multiple SV-affected genes were differentially expressed and cell-cell adhesion and synapse-related processes were enriched in KOLF2.1J organoids. These findings highlight the importance of robust genetic characterization and multi-line validation for reliable iPSC-based modeling.

INTRODUCTION

Induced pluripotent stem cell (iPSC)-derived cultures, including both two-dimensional (2D) monolayers and three-dimensional (3D) organoids, have become powerful tools for studying human development and modeling a wide range of disorders across tissues (Camp et al., 2015; Giandomenico and Lancaster, 2017; Lancaster et al., 2013; Sun et al., 2021). However, the derivation of iPSC lines and their use for various differentiation protocols can introduce variability in cellular development, potentially affecting experimental outcomes and limiting standardization across studies (Quadrato et al., 2016, 2017). KOLF2.1J is a clonal human induced pluripotent stem cell (hiPSC) line (Pantazis et al., 2022), derived from the parental HPSI0114i-kolf 2-C1 hiPSC subline (Kilpinen et al., 2017; Skarnes et al., 2019). It was edited via CRISPR/Cas9 to repair a likely pathogenic mutation in *ARID2* (Hildebrandt et al., 2019) and curated with the intent to serve as a reference cell line for modeling Alzheimer's disease and related dementias. Significant effort by the iPSC Neurodegenerative Disease Initiative (iNDI) consortium was devoted to quality control, ensuring that the KOLF2.1J line was free of deleterious mutations by

whole genome sequencing, capable of differentiating into neuronal cell types and genetically stable following gene editing (Pantazis et al., 2022).

A subsequent investigation employing high-density Illumina single-nucleotide polymorphism (SNP) arrays (Infinium GSA-24v3 BeadChip with 620,000 probes) identified five heterozygous structural variants (SVs) in KOLF2.1J, targeting SVs with lengths greater than 100 kbp (Gracia-Diaz et al., 2024). These included three heterozygous deletions and two heterozygous duplications, ranging in size from 132 to 766 kbp. These results were reaffirmed by an additional Illumina SNP array (Infinium Global Array with exonic coverage and 1.8 million probes) (Ryan et al., 2024). Two SVs, affecting the genes *JARID2* and *DTNBP1* (6p22.3 deletion), and *ASTN2* (9q33.1 deletion), have been associated with neurodevelopmental disorders. *JARID2* deletion is associated with *JARID2*-haploinsufficiency, a developmental disorder (Verberne et al., 2021); *DTNBP1* coordinates synapse maintenance and is linked to schizophrenia (Jia et al., 2014); *ASTN2* deletion is linked to varied developmental disorders through excess mitophagy and showed 50% reduced protein levels in iPSCs and neural progenitors (Hayashi et al., 2024). In contrast, while the 3p14 duplication affected several genes, none were classified as

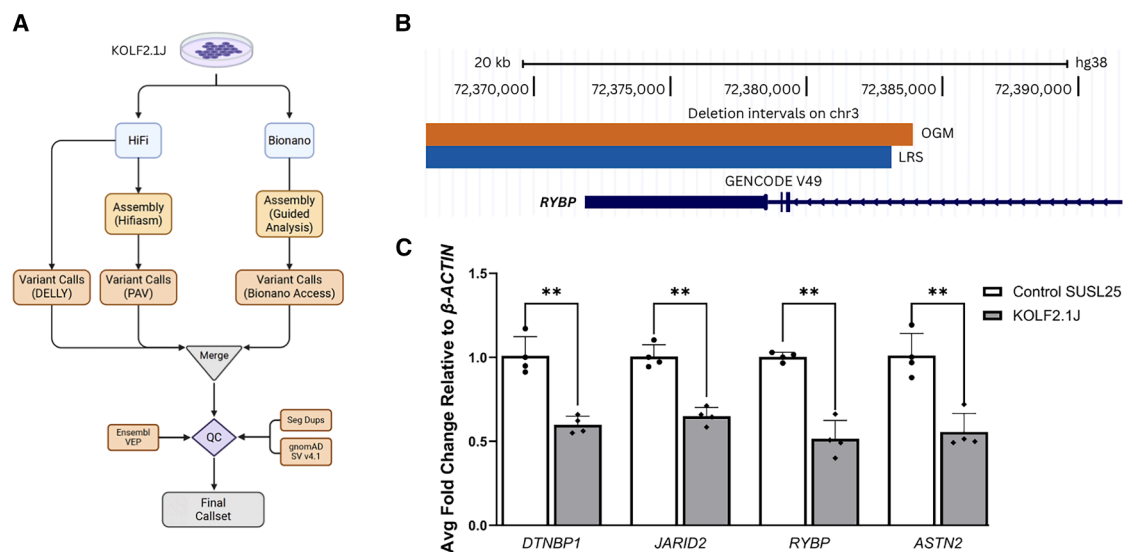


Figure 1. Detection of structural variants in KOLF2.1J

(A) Overview of the calling approach yielding raw calls from LRS and OGM.

(B) Intersection of variants with support from both LRS and OGM with genic content using VEP yielded 41 high-impact variants in KOLF2.1J; shown is a depiction of a chromosome 3 deletion impacting the *RYBP* gene.

(C) Reduced expression of *DTNBP1*, *JARID2*, *RYBP*, and *ASTN2* in KOLF2.1J cells relative to SUSL25 controls. KOLF2.1J and SUSL25 iPSCs were collected from four different passages ($n = 4$) showed significantly reduced expression of *DTNBP1* (0.598-fold), *JARID2* (0.650-fold), *RYBP* (0.515-fold), and *ASTN2* (0.556-fold) compared with SUSL25 cells. Each individual data point represents the mean of four technical replicates. Bars represent the group mean, with error bars. Unpaired t test with Welch's correction was applied with adjustment for multiple comparisons. Adjusted p values were 0.0097 for *DTNBP1*, 0.001 for *JARID2*, 0.008 for *RYBP*, and 0.0081 for *ASTN2*. $**p < 0.01$.

deleterious. Additionally, the 18q22 duplication and the 3p13 deletion did not overlap any known genes.

Due to the limited resolution of SNP arrays and the need for genetic characterization of a widely used model reference line, we interrogated the genome of KOLF2.1J with long-read sequencing (LRS) and used optical genome mapping (OGM) for orthogonal validation. With this approach, we identified SVs in KOLF2.1J that were absent from databases of common SVs in the population. Our studies curated 41 SVs (37 plus the four previously reported SVs) predicted to affect genes by variant effect predictor (VEP) and found that multiple SV-affected genes were dysregulated in KOLF2.1J-derived forebrain organoids (FOs). These genes include *ASTN2* and *DTNBP1*, which may account for the observed perturbation of cadherin-mediated cell-cell adhesion and synapse-related processes. Overall, our findings underscore the importance of thoroughly assessing individual- or cell line-specific SVs in hiPSC reference lines.

RESULTS

SV identification in KOLF2.1J using PacBio LRS

To identify KOLF2.1J cell line variants, we used PacBio HiFi sequencing to generate assembly based variant calls from

passage 4 cells (Figure 1A). We used the phased assembly variant (PAV) caller to identify variants, as it was recently found to have a higher sensitivity and overall call support than other read-based and many assembly based methods, while retaining a low Mendelian inheritance error rate (see methods) (Logsdon et al., 2025). With PAV, we identified a total of 27,665 SVs detected in KOLF2.1J compared to the GRCh38 (hg38) reference. These 27,665 SVs consist of 16,942 insertions, 10,483 deletions, 92 inversions identified by PAV (Ebert et al., 2021; Logsdon et al., 2025), and 148 duplications by DELLY (Rausch et al., 2012). Using a newer gapless Telomere to Telomere T2T-CHM13v2.0 (T2T) sequence as a reference (Nurk et al., 2022), we identified 38,909 SVs consisting of 19,459 insertions, 19,320 deletions, 57 inversions, and 73 duplications. Inversions were excluded from further analyses as they are difficult to validate and unlikely to lead to gene dosage changes. Additionally, we excluded SVs present in alternative contigs of the GRCh38 reference during orthogonal validation. All variant calls generated by PAV are deposited as VCFs in Zenodo (SVs as well as raw, unvalidated Indel and SNVs).

We then focused on the 1,974 GRCh38 insertion, deletion, and duplication SVs that were ≥ 2 kb in size, as OGM cannot reliably detect SVs that are 50–2,000 bp, and we used BioNano data for orthogonal validation



(Figure 1A). Of these 1,974 SVs (7.16% of the total SV count), 150 were identified as high impact by VEP (McLaren et al., 2016) and were absent from gnomAD-SV (filtered using gnomadSV) (Collins et al., 2020), and were therefore considered candidate KOLF2.1J-specific rare SVs. Additionally, no inversion calls were deemed to be high impact, further justifying their filtering above. All SVs with respect to either reference are reported in our raw and filtered LRS call sets.

Orthogonal validation of SV calls with Bionano OGM

To identify SVs for orthogonal validation of LRS findings, we used the Bionano OGM pipeline and generated two variant alignments for KOLF2.1J (passage 30). Aligning the reads to the hg38 reference yielded a raw list of 1,790 total SVs (973 insertions, 614 deletions, 141 inversions, and 62 duplications); alignment to the T2T reference yielded 1,564 total SVs (681 insertions, 700 deletions, 123 inversions, and 60 duplications). We focused on the 1,649 non-inversion SVs with respect to hg38 for downstream analyses; however, all SVs are reported in our raw OGM call set (<https://www.ebi.ac.uk/ena/browser/view/ERZ27120469>).

Overall, the two methodologies with similar size cut-offs identified 1,974 (LRS) and 1,649 (OGM) SVs with respect to the GRCh38 reference (Table S1). We found that 1,064 calls were shared (80% RO: reciprocal overlap) between LRS (53.9%) and OGM (64.5%) technologies. When broken down by SV type, we found that deletions (64.9%; 462 out of 712) had the highest concordance, followed by insertions (52%; 596 out of 1,146) and duplications (5.2%; 6 out of 116). Furthermore, we identified that 65 of the 150 high-impact variants were supported by both technologies. Subsequently, we dropped two of these 65 SVs as they were overlapping each other, were >1 Mbp in size, and were located in pericentromeric regions that have a higher potential false positive call rate.

Potential effects of SVs

The 63 LRS and OGM supported SVs (Table S2) were positive for high-impact VEP results and were absent from population-scale determination of SVs. Overall, these 63 SVs intersected with 120 genes (Hamosh et al., 2005; Karczewski et al., 2020; Landrum et al., 2017) (Table S2). We next intersected the 63 SVs with known segmental duplication (SD) regions, and identified 22 that were primarily within SDs, and would therefore be difficult to characterize due to gene duplications and a high likelihood of alternative haplotypes in the population. We focused on the remaining 41 SVs that were in non-SD loci (Table 1) to examine their impact on the function of the KOLF2.1J iPSCs.

These 41 LRS and OGM supported SVs overlapped 80 genes, and 4 of them were previously identified with SNP

array data (Gracia-Diaz et al., 2024). Interestingly, our OGM and LRS data revealed that the previously reported 3p13.2 SV leads to the deletion of exons 3 and 4 of the RING1 and YY1 binding protein (*RYBP*) gene, indicating that the 3p13.2 deletion is potentially larger than previously described (Gracia-Diaz et al., 2024; Ryan et al., 2024), and contains a 5' breakpoint falling within *RYBP* (Figure 1B).

qPCR validation of SV impact on gene expression in iPSCs

To assess the effects of SVs on gene expression, we performed qPCR on four SV-overlapping genes (*DTNBP1*, *JARID2*, *RYBP*, and *ASTN2*) in KOLF2.1J iPSCs collected across passages 17, 18, 19, and 29 and compared their expression to a control iPSC line (SUSL25) (Figure 1C). All four genes showed significantly reduced expression in KOLF2.1J relative to SUSL25. The mean fold changes were 0.598 for *DTNBP1*, 0.650 for *JARID2*, 0.515 for *RYBP*, and 0.556 for *ASTN2*, corresponding to reductions of approximately 40%, 35%, 49%, and 44%, respectively. These results demonstrate reproducible reductions in transcript abundance of four SV-overlapping genes in KOLF2.1J relative to SUSL25 across the tested passages, consistent with an effect of the corresponding deletions on gene expression. These findings support the conclusion that SVs can impact gene expression and highlight the importance of precise breakpoint resolution for accurate variant interpretation.

Evaluating the potential impacts of KOLF2.1J's SVs on FO differentiation

To assess how genetic variation impacts FO differentiation, we performed single-cell RNA sequencing on day 90 organoids derived from three human iPSC lines, KOLF2.1J, BIONi010-C, and HMGU1 (Velasco et al., 2019). Quality control metrics and stress levels were consistent across cell lines (Figure S1). Forebrain regional identity was confirmed by *FOXP1* expression and reference-to-query mapping against human (Braun et al., 2023) and mouse (Fleck et al., 2021) brain atlases (Figures S2A–S2D). By analyzing the expression of selected markers (Figure S2D) and performing reference-to-query integration with a meta-atlas consisting of seven published single-cell transcriptomic datasets (Nano et al., 2025) (Figure S2E), we identified 12 distinct cell populations (Figure 2). These comprise two major neuronal lineages, excitatory (*EMX1* expression) and inhibitory (*GAD2* expression) (Figure S2D). Specifically, cell populations included: radial glia/outer radial glia (RG/oRG), cycling progenitors (Cyc Prog), immature inhibitory interneurons (iINs), *FOXP1*+ and *FOXP2*+ immature inhibitory projection neurons (iPNs), Cajal-Retzius neurons (CRNs), intermediate precursor cells (IPCs), newborn (nb) and more mature callosal projection neurons (CPNs), corticofugal



Table 1. Table of genes found to be affected by KOLF2.1J SVs

Chr	POS	End	SV Type	Size	Genes
chr3	61149366	61925599	DUP	776,233	LOC105377115, LOC105377114, LOC124909387, LOC124909388, LOC124906245
chr6	15481863	15729306	DEL	247,443	LOC124901267, JARID2, LOC105374947, DTNBP1
chr15	84167503	84360490	DEL	192,987	DNM1P41, LOC105370936, LOC440300, LOC105376722, GOLGA2P7, LOC105376724, UBE2Q2P16, LOC124903544, GOLGA6L4
chr3	72240239	72383130	DEL	142,891	LOC105377159, RYBP, LOC105377158
chr9	116478811	116610424	DEL	131,613	LOC105376240, ASTN2, LOC107987014, ASTN2-AS1
chr5	686532	778595	DEL	92,063	LOC101929898, ZDHHC11B, TPPP
chr14	67268703	67349856	DEL	81,153	PALS1, ATP6V1D
chrX	3838454	3915307	DEL	76,853	FAM239B, FAM239A, LOC389906
chr7	73052215	73108363	DEL	56,148	SPDYE11, SPDYE9, SPDYE10
chr6	31984683	32023788	DEL	39,105	C4B, C4A, CYP21A1P, TNXA
chr7	112741329	112779174	DEL	37,845	TMEM168
chr22	23931955	23969108	DEL	37,153	LOC100652871, DDTL, GSTT2B
chr22	41179886	41179887	INS	31,171	EP300
chr14	73530642	73558371	DEL	27,729	LOC124903344, ACOT1
chr8	6992887	7011986	DEL	19,099	DEFT1P2, LOC124901874, DEFA1B
chr11	61205470	61224299	DEL	18,829	PGA4, PGA3
chr1	207544734	207563290	DEL	18,556	CR1
chr1	207526605	207545160	DEL	18,555	CR1
chr1	109682998	109701443	DEL	18,445	GSTM1, GSTM2
chr9	70704690	70722679	DEL	17,989	LOC105376078
chr10	37170147	37181940	DEL	11,793	ANKRD30A
chr6	170399908	170399909	INS	11,496	LOC124901475
chr7	38348792	38357949	DEL	9,157	TRG-AS1
chr11	1017981	1017982	INS	8,781	MUC6
chr7	4591607	4600110	DEL	8,503	LOC124901580
chr11	56376094	56376095	INS	7,605	OR8U1
chr1	105473194	105480775	DEL	7,581	LOC105378881
chr1	247123595	247123596	INS	6,242	ZNF124
chr5	58384165	58390274	DEL	6,109	LOC105378984
chr9	87276757	87276758	INS	6,075	LOC105376126
chr13	29641686	29647707	DEL	6,021	LOC102723345
chr1	206056136	206056137	INS	5,912	LOC105372857
chr11	7688250	7693273	DEL	5,023	OVCH2, LOC105376533

(Continued on next page)



Table 1. Continued

Chr	POS	End	SV Type	Size	Genes
chr8	143859400	143864197	DEL	4,797	EPPK1
chr10	122602824	122606883	DEL	4,059	DMBT1
chr1	234676159	234679673	DEL	3,514	LOC105373209, LOC124904555, LOC105373208
chr19	23322868	23322869	INS	2,985	ZNF91
chr12	48331683	48334597	DEL	2,914	ZNF641
chr9	86999015	86999016	INS	2,433	GAS1RR
chr13	102694504	102694505	INS	2,090	METTL21C
chrX	151915682	151917726	DEL	2,044	MAGEA4

Each of the 41 orthogonally supported SVs that was absent from gnomad SV and found to have a deleterious VEP score is tabulated along with the genes affected. For each SV, the chromosome and coordinates from hg38, the SV type, the size of the SV, and the genes affected by each variant are listed.

projection neurons (CFuPNs), and an additional neuronal group labeled FOXG1- Ns characterized by a low *FOXG1* expression (Figure S2D). Composition and pseudotime analyses revealed no significant differences in cell type proportions or neurodevelopmental trajectories between cell lines (Figures 2B and S3), indicating synchronous differentiation.

We then identified KOLF2.1J-specific differentially expressed genes (DEGs) across all cell types ($FDR < 0.05$, $|\log_2-FC| > 1.0$). iINs and FOXP2+ iPNs exhibited the highest num-

ber of DEGs, significantly exceeding the number of DEGs in BIONI010-C and HMGU1-derived FOs (Figures 3A and S4A). Four SV-affected genes were differentially expressed: *DTNBP1* (downregulated in CPNs, CFuPNs, and FOXP1+ iPNs), *GSTT2B* (downregulated in iINs, FOXP2+ iPNs, and Cyc Prog), *GSTM2* (upregulated in iINs and FOXP2+ iPNs), and *ASTN2* (upregulated in iINs) (Table S3). Gene ontology (GO) analysis associated KOLF2.1J-specific DEGs with ion transport, synaptic assembly, and cell-cell adhesion

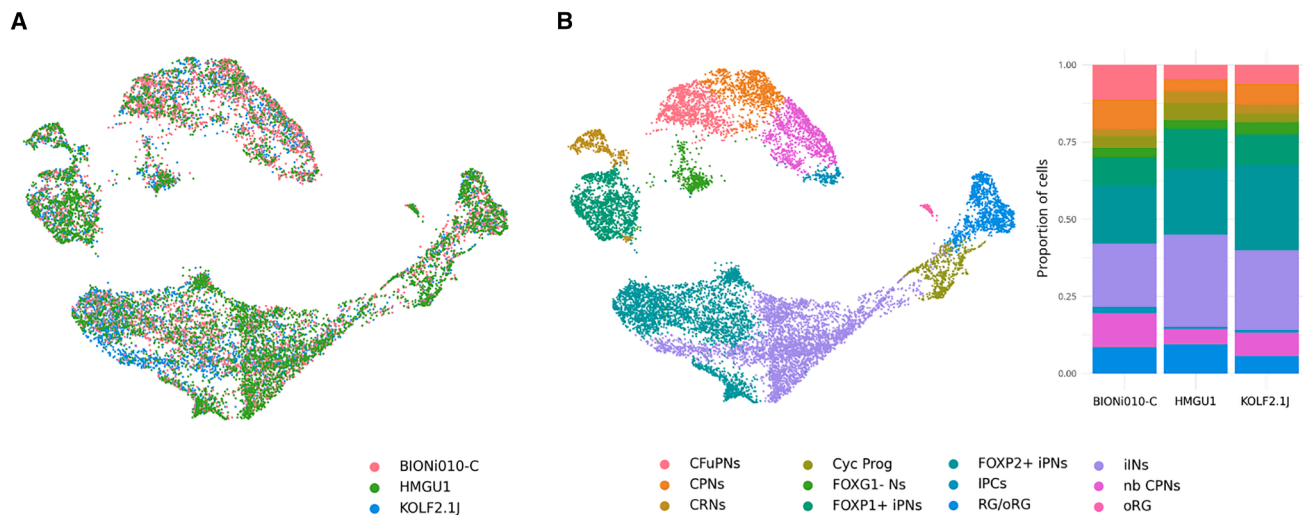


Figure 2. Cell-type composition across FOs derived from different iPSC lines

(A) UMAP embedding of FO cells colored by cell line.

(B) UMAP embedding of FO cells colored by cell type: RG/oRG, radial glia/outer radial glia; Cyc Prog, cycling progenitors; iINs, immature inhibitory interneurons; FOXP1+/FOXP2+ iPNs, FOXP1+/FOXP2+ immature inhibitory projection neurons; CRNs, Cajal-Retzius neurons; IPCs, intermediate precursor cells; nb CPNs, newborn callosal projection neurons; CFuPNs, corticofugal projection neurons; FOXG1- Ns, FOXG1- neurons. The UMAP plot was generated after removing outliers, background RNA, doublets, and stressed cells, followed by data normalization, principal component analysis based on the top 2,000 highly variable genes, and batch effect correction. Total number of cells: 13,818; $n = 3$ organoids per iPSC line.



(Figures S4B and S4C; Tables S4 and S5). Gene set enrichment analysis (GSEA) corroborated these findings across KOLF2.1J neuronal populations and revealed negative enrichment of cytoplasmic translation and ribosome biogenesis and assembly in iINs (Figure S5).

To explore the functional connectivity between differentially expressed SV-impacted genes and the DEGs associated with the up- and downregulated processes identified via GO analysis, we constructed a protein-protein interaction (PPI) network using STRING (Szklarczyk et al., 2023) (Figure 3B). High confidence interactions were sparse and predominantly unidirectional.

Furthermore, CellChat analysis revealed that the total number and strength of inferred cell-cell contact interactions were consistent across cell lines (Figure S6A), despite the observed enrichment of adhesion-related processes in KOLF2.1J-derived FOs. Although KOLF2.1J exhibited weaker synaptic signaling and lacked specific testosterone/DHT-mediated pathways, GABAergic and glutamatergic signaling were preserved and consistent across cell lines (Figures S6B and S6C).

DISCUSSION

Our findings indicate that the widely used reference iPSC line, KOLF2.1J, contains SVs that may influence neuronal gene expression, which bears consideration when modeling neurological disorders.

Interestingly, although the LRS and OGM data were generated from different passages of the same cell line (passage 4 and passage 30, respectively), there were few discrepant SVs between the two datasets. Moreover, many of the OGM-specific variants were sub-allelic in their frequencies. This indicates that overall, the KOLF2.1J cell line is stable and does not seem to be accumulating deleterious SVs at a rapid rate.

A key advantage of both OGM and LRS is their ability to resolve SV breakpoints more precisely and to detect smaller variants than SNP arrays, thereby reducing the risk of false negatives. This was exemplified by our refinement of the previously reported 3p13.2 deletion, which we found to overlap *RYBP*, a gene linked to retinoic acid signaling and regulation of *PAX6* expression. *RYBP* is a component of polycomb repressive complex 1 (PRC1), mediates histone H2A ubiquitination, and plays important roles in neuronal development (Pirity et al., 2005). Reduced *RYBP* dosage has been associated with impaired neuronal differentiation, expansion of neural progenitors, reduced dendritic complexity, and altered Notch signaling (Kovacs et al., 2016; Li et al., 2021).

Although our study focused on cortical lineages, the *RYBP* deletion may also be relevant in other neural contexts. *RYBP*

is a component of variant PRC1 complexes and has an established role in neural lineage regulation, while reduced *Rybp* dosage in mice causes severe abnormalities of CNS development, including disruption of hindbrain/cerebellar structures (Pirity et al., 2005; Liu et al., 2023). This raises the possibility that *RYBP* haploinsufficiency could influence early patterning or timing in human cerebellar organoids as well. Such an interpretation may help contextualize previous reports of reduced neuronal output in KOLF2.1J-derived cerebellar organoids relative to BIONi010-C at days 35 and 50 of differentiation (Sarieva et al., 2026), although this postulation requires direct testing.

To assess whether the identified SVs are associated with altered forebrain differentiation, we compared day 90 organoids derived from KOLF2.1J and two independent control iPSC lines. Overall, KOLF2.1J remained suitable for generating FOs, particularly for studies focused on cell-type composition and developmental trajectories, as we observed no significant differences across lines in cell-type proportions, pseudotime-associated cell densities within excitatory and inhibitory lineages, or intercellular communication patterns. This conclusion is further supported by our recent proteomic and secretomic profiling of organoids derived from the same three lines, which likewise revealed no striking inter-line differences (Yentür et al., 2025). Nevertheless, KOLF2.1J-derived organoids exhibited a disproportionately high number of DEGs, particularly in the most abundant neuronal populations. This transcriptomic signature included altered expression of four SV-affected genes, *DTNBP1*, *GSTT2B*, *GSTM2*, and *ASTN2*, together with enrichment of cell-cell adhesion- and synapse-related processes.

Among these four genes, *DTNBP1* and *ASTN2* are of particular interest because both have established roles in neuronal trafficking, migration, and synaptic function, making them plausible contributors to the synaptic and adhesion-related transcriptional shifts observed in KOLF2.1J-derived organoids. *DTNBP1*, in particular, is a critical regulator of synaptic vesicle trafficking and neurotransmitter release, and its loss of function has been linked to altered glutamatergic and GABAergic signaling in cortical interneurons and projection neurons (Jun et al., 2022). By contrast, *GSTT2B* and *GSTM2*, as members of the glutathione S-transferase family, are more readily linked to detoxification and oxidative stress defense than to direct regulation of forebrain cell identity, suggesting that their altered expression may reflect compensatory or stress-buffering responses rather than direct patterning defects (Allocati et al., 2018). Consistent with this interpretation, our PPI analysis showed that high-confidence interactions between these SV-affected genes and the broader DEG networks in KOLF2.1J-derived organoids were sparse and largely indirect. Together with the simultaneous up- and

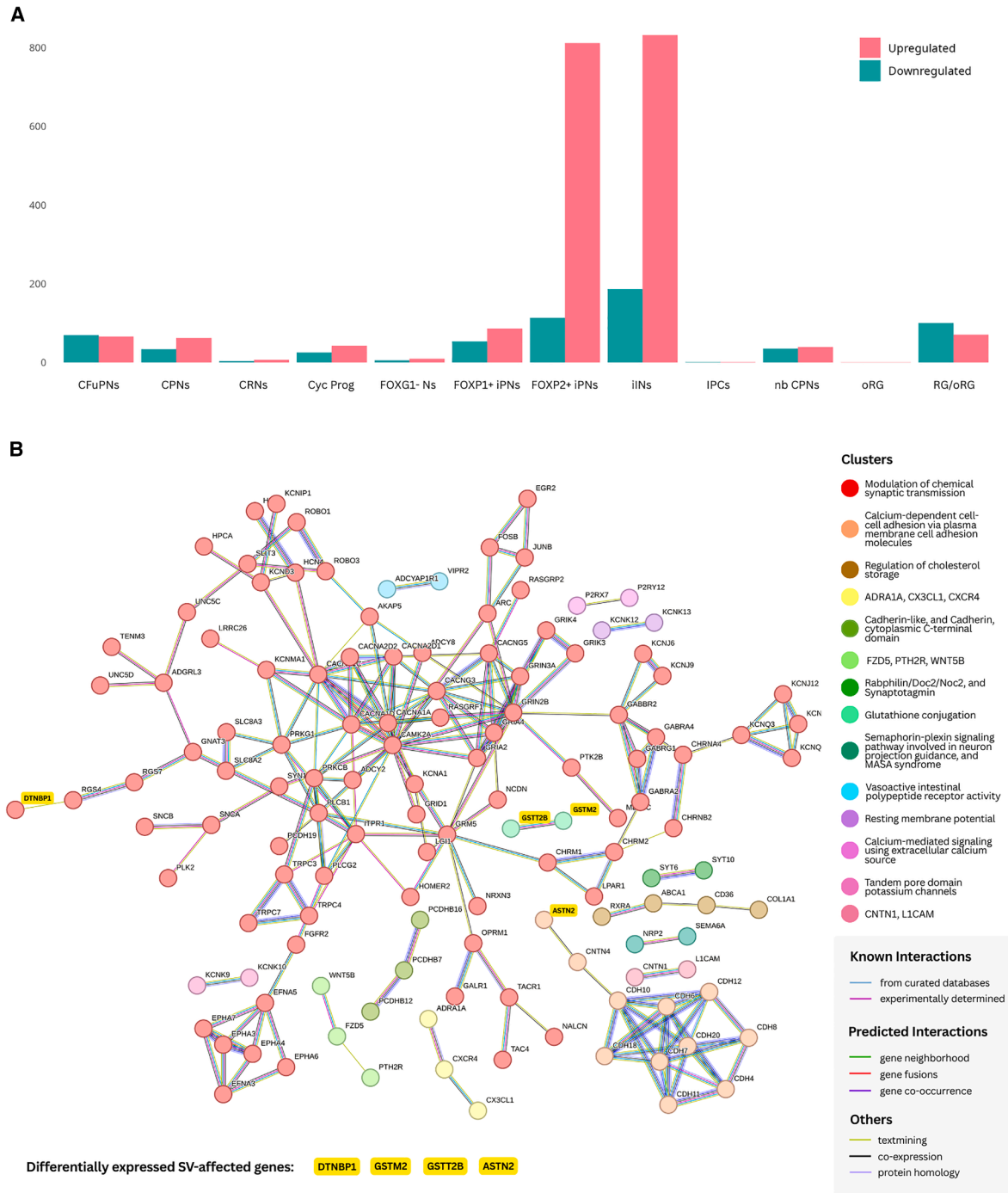


Figure 3. Differential gene expression analysis and PPI network of KOLF2.1J DEGs

(A) Number of up- and downregulated DEGs across cell types in KOLF2.1J-derived FOs.

(B) PPI network of KOLF2.1J DEGs, as assessed using STRING. Nodes include the differentially expressed SV-affected genes *DTNBP1*, *ASTN2*, *GSTT2B*, and *GSTM2*, as well as DEGs from KOLF2.1J CPNs, CFuPNs, iNs, and FOXP2+ iPNs that are associated with up- and downregulated GO terms for these cell types. Only high-confidence interactions (score > 0.7) were retained, based on evidence from text mining, experimental data, curated databases, co-expression, neighborhood, gene fusion, and co-occurrence. Disconnected nodes were omitted for clarity. Nodes are colored and labeled according to clusters identified via MCL clustering.



downregulation of genes within the same enriched biological processes, this suggests that compensatory mechanisms may buffer the downstream effects of these SV-associated transcriptional changes. Such buffering may explain why we observed only subtle differences in inferred synaptic signaling across the three lines despite enrichment of synapse-related pathways in KOLF2.1J-derived organoids.

Collectively, our findings emphasize that comprehensive genomic profiling is critical for accurately interpreting cellular phenotypes and line-specific variation in neural differentiation. While remaining suitable for FO studies, KOLF2.1J and other reference iPSC lines should be considered in the context of their individual variant profiles, and their phenotypic consequences are likely to vary depending on developmental context, lineage, and assay sensitivity. We also underscore the importance of using multiple lines for reliable iPSC-based modeling. Finally, we foresee LRS, OGM, and moving forward telomere-to-telomere sequencing as future gold standards for the characterization of reference iPSCs, ensuring that lengthy and resource-intensive differentiation protocols are carried out with physiologically relevant cell lines.

Limitations of the study

One important limitation of this study is that we do not have information on germline SVs for the individual from which KOLF2.1J was derived. Therefore, we cannot distinguish between SVs that likely represent natural variation with those SVs that have been acquired during reprogramming and subsequent manipulation. It will be important to analyze DNA from blood or skin fibroblasts to better interpret the significance and impact of SVs found in iPSC cell lines. Further, we have generated SNV and Indel calls for KOLF2.1J; however, these data need curation and validation. Additionally, our analysis could be strengthened by more comprehensive genetic characterization of the cell line (i.e., telomere to telomere assembly) and by incorporating additional organoid protocols.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Christine R. Beck (christine.beck@jax.org).

Materials availability

This study did not generate unique reagents or new iPSC lines.

Data and code availability

Seven files pertain to the KOLF OGM data: one BNX file of raw read data, and two sets of three files, one set for the hg38- and T2T-aligned analyses: CMAP (variant alignment), XMAP (alignment relative to reference), and SMAP (raw list of unfiltered SVs). These

seven files have been uploaded to the European Nucleotide Archive under analysis accession number ERZ27120469 (project number ERP173632 and sample number SAMEA118572239) and are available for download at ENA: <https://www.ebi.ac.uk/ena/browser/view/ERZ27120469>.

Long-read DNA sequencing data for KOLF2.1J has been uploaded to the NCBI sequence read archive : PRJNA1290418. Due to the cell line agreements, we have removed reads that mapped to chromosome Y. Structural variant datasets (unfiltered) generated by LRS are available to download at Zenodo: <https://doi.org/10.5281/zenodo.17955237>.

Raw data from single-cell RNA sequencing experiments are available in ArrayExpress: E-MTAB-17432, with CMO IDs 310, 311, and 312. Processed data have been deposited in CELLxGENE: <https://cellxgene.cziscience.com/collections/10d0c48e-0a15-406f-9d5e-40a2dda6e705>. All data are deposited with corresponding metadata and have been made publicly accessible according to the FAIR data principle.

All software used in this study can be found in the [key resources table](#).

All original analysis code used in single-cell RNA sequencing analysis will be accessible upon publication on GitHub.

Further requests for additional information should be directed to the [lead contact](#), Christine R. Beck (christine.beck@jax.org).

ACKNOWLEDGMENTS

We thank Elisabeth Gustafsson and Pierre Collignon for their technical support, and Charles Lee for his critical review of the manuscript. This study was funded by the WIN Program of the Heidelberg Academy of Sciences and Humanities, financed by the Ministry of Sciences, Research, and the Arts of the State of Baden-Württemberg (to S.M.) and Hertie Foundation (to S.M.). Partial funding was provided by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under Germany's Excellence Strategy via the Excellence Cluster 3D Matter Made to Order (EXC-2082/1 – 390761711 and 2082/2 390761711). Additional financial support came from the Brain & Behavior Research Foundation (NARSAD Young Investigator Grant 27026 to S.M.), the Baden-Württemberg state postgraduate fellowship (to K.S. and T.K.), and the Daimler and Benz Foundation (32-06/20 to S.M.). Further funding was provided through departmental start-up funds and philanthropic support (to B.S.). C.R.B. and P.B. were supported by start-up funds from the University of Connecticut Health Center, and long-read sequencing and analysis was supported by The Jackson Laboratory Director's Innovation Fund, Genome Technologies, Cellular Engineering, and Research IT Resources.

AUTHOR CONTRIBUTIONS

L.B., conceptualization, methodology, formal analysis, data curation, preparation of figures, and writing – original draft; Z.Y., experimentation, writing methods, and writing – review & editing; M.J.Y., conceptualization, investigation, methodology, formal analysis, data curation, preparation of OGM figures, and writing – original draft; P.B., conceptualization, LRS data curation, methodology, QC, formal analysis, visualization, and writing – original draft, review & editing; K.S., investigation, data curation,



methodology, resources, and writing – original draft, reviews & editing; G.K., investigation, writing methods, and preparation of figures; F.Z., experimentation; K.S., T.K., and K.B., experimentation and writing – review & editing; W.S., conceptualization, resources, supervision, and writing – review & editing; C.R.B., S.M., and B.S., conceptualization, funding acquisition, resources, supervision, and writing – review & editing.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Culture and clonal expansion of KOLF2.1J iPSCs for LRS
 - iPSC culture for Bionano optical genome mapping and gene expression analysis
 - iPSC culture for organoid differentiation
 - Forebrain organoid differentiation
- **METHOD DETAILS**
 - Long read sequencing
 - Bionano optical genome mapping
 - SYBR green expression analysis
 - Single-cell RNA sequencing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Long read sequencing
 - Bionano optical genome mapping
 - Concordance between LRS and OGM data
 - SYBR green expression analysis
 - Single-cell RNA sequencing data analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.stemcr.2026.103112>.

Received: July 17, 2025

Revised: September 7, 2026

Accepted: September 7, 2026

REFERENCES

Allocati, N., Masulli, M., Di Ilio, C., and Federici, L. (2018). Glutathione transferases: substrates, inhibitors and pro-drugs in cancer and neurodegenerative diseases. *Oncogenesis* 7, 8. <https://doi.org/10.1038/s41389-017-0025-3>.

Braun, E., Danan-Gotthold, M., Borm, L.E., Lee, K.W., Vinsland, E., Lönnerberg, P., Hu, L., Li, X., He, X., Andrusivová, Ž., et al. (2023). Comprehensive cell atlas of the first-trimester developing human brain. *Science* 382, eadf1226. <https://doi.org/10.1126/science.adf1226>.

Camp, J.G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Bräuninger, M., Lewitus, E., Sykes, A., Hevers, W., Lancaster, M., et al. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. USA* 112, 15672–15677. <https://doi.org/10.1073/pnas.1520760112>.

Chen, M.Y., Heinrich, L., Zafar, F., Sedov, K., and Schüle, B. (2023). Automated 384-well SYBR Green Expression Array for Optimization of Human Induced Pluripotent Stem Cell Differentiation. *Bio. Protoc.* 13, e4689. <https://doi.org/10.21769/BioProtoc.4689>.

Cheng, H., Asri, M., Lucas, J., Koren, S., and Li, H. (2024). Scalable telomere-to-telomere assembly for diploid and polyploid genomes with double graph. *Nat Methods* 21, 967–970. <https://doi.org/10.1038/s41592-024-02269-8>.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* 18, 170–175. <https://doi.org/10.1038/s41592-020-01056-5>.

Cheng, H., Jarvis, E.D., Fedrigo, O., Koepfli, K.P., Urban, L., Gemmell, N.J., and Li, H. (2022). Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology* 40, 1332–1335. <https://doi.org/10.1038/s41587-022-01261-x>.

Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., et al. (2020). A structural variation reference for medical and population genetics. *Nature* 581, 444–451. <https://doi.org/10.1038/s41586-020-2287-8>.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., et al. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10, giab008. <https://doi.org/10.1093/gigascience/giab008>.

Ebert, P., Audano, P.A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M.J., Sulovari, A., Ebler, J., Zhou, W., Serra Mari, R., et al. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* 372, eabf7117. <https://doi.org/10.1126/science.abf7117>.

Fleck, J.S., Sanchís-Calleja, F., He, Z., Santel, M., Boyle, M.J., Camp, J.G., and Treutlein, B. (2021). Resolving organoid brain region identities by mapping single-cell genomic data to reference atlases. *Cell Stem Cell* 28, 1148–1159.e8. <https://doi.org/10.1016/j.stem.2021.02.015>.

Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2022). Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res.* 10, 979. <https://doi.org/10.12688/f1000research.73600.2>.

Giandomenico, S.L., and Lancaster, M.A. (2017). Probing human brain evolution and development in organoids. *Curr. Opin. Cell Biol.* 44, 36–43. <https://doi.org/10.1016/j.cceb.2017.01.001>.

Gracia-Diaz, C., Perdomo, J.E., Khan, M.E., Roule, T., Disanza, B.L., Cajka, G.G., Lei, S., Gagne, A.L., Maguire, J.A., Shalem, O., et al. (2024). KOLF2.1J iPSCs carry CNVs associated with neurodevelopmental disorders. *Cell Stem Cell* 31, 288–289. <https://doi.org/10.1016/j.stem.2024.02.007>.

Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., and McKusick, V.A. (2005). Online Mendelian Inheritance in Man (OMIM), a



- knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 33, D514–D517. <https://doi.org/10.1093/nar/gki033>.
- Hao, Y., Stuart, T., Kowalski, M.H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C., and Satija, R. (2024). Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* 42, 293–304. <https://doi.org/10.1038/s41587-023-01767-y>.
- Hayashi, Y., Okumura, H., Arioka, Y., Kushima, I., Mori, D., Lo, T., Otgonbayar, G., Kato, H., Nawa, Y., Kimura, H., et al. (2024). Analysis of human neuronal cells carrying ASTN2 deletion associated with psychiatric disorders. *Transl. Psychiatry* 14, 236. <https://doi.org/10.1038/s41398-024-02962-4>.
- Hildebrandt, M.R., Reuter, M.S., Wei, W., Tayebi, N., Liu, J., Sharmin, S., Mulder, J., Lesperance, L.S., Brauer, P.M., Mok, R.S.F., et al. (2019). Precision Health Resource of Control iPSC Lines for Versatile Multilineage Differentiation. *Stem Cell Rep.* 13, 1126–1141. <https://doi.org/10.1016/j.stemcr.2019.11.003>.
- Jia, J.-M., Hu, Z., Nordman, J., and Li, Z. (2014). The Schizophrenia Susceptibility Gene Dysbindin Regulates Dendritic Spine Dynamics. *J. Neurosci.* 34, 13725–13736. <https://doi.org/10.1523/jneurosci.0184-14.2014>.
- Jin, S., Plikus, M.V., and Nie, Q. (2025). CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics. *Nat. Protoc.* 20, 180–219. <https://doi.org/10.1038/s41596-024-01045-4>.
- Jun, R., Zhang, W., Beacher, N.J., Zhang, Y., Li, Y., and Lin, D.T. (2022). Dysbindin-1, BDNF, and GABAergic Transmission in Schizophrenia. *Front. Psychiatry* 13, 876749. <https://doi.org/10.3389/fpsy.2022.876749>.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443. <https://doi.org/10.1038/s41586-020-2308-7>.
- Kilpinen, H., Goncalves, A., Leha, A., Afzal, V., Alasoo, K., Ashford, S., Bala, S., Bensaddek, D., Casale, F.P., Culley, O.J., et al. (2017). Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* 546, 370–375. <https://doi.org/10.1038/nature22403>.
- Kovacs, G., Szabo, V., and Pirtity, M.K. (2016). Absence of Rybp Compromises Neural Differentiation of Embryonic Stem Cells. *Stem Cells Int.* 2016, 4034620. <https://doi.org/10.1155/2016/4034620>.
- Kunze, C., Börner, K., Kienle, E., Orschmann, T., Rusha, E., Schneider, M., Radivojkov-Blagojevic, M., Drukker, M., Desbordes, S., Grimm, D., and Brack-Werner, R. (2018). Synthetic AAV/CRISPR vectors for blocking HIV-1 expression in persistently infected astrocytes. *Glia* 66, 413–427. <https://doi.org/10.1002/glia.23254>.
- Lancaster, M.A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* 501, 373–379. <https://doi.org/10.1038/nature12517>.
- Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–D1067. <https://doi.org/10.1093/nar/gkx1153>.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li, Q., Chen, J., Liang, F., Zhang, J., Qu, W., Huang, X., Cheng, X., Zhao, X., Yang, Z., Xu, S., and Li, X. (2021). RYBP modulates embryonic neurogenesis involving the Notch signaling pathway in a PRC1-independent pattern. *Stem Cell Rep.* 16, 2988–3004. <https://doi.org/10.1016/j.stemcr.2021.10.013>.
- Liu, Y., Hu, G., Yang, S., Yao, M., Liu, Z., Yan, C., Wen, Y., Ping, W., Wang, J., Song, Y., et al. (2023). Functional dissection of PRC1 subunits RYBP and YAF2 during neural differentiation of embryonic stem cells. *Nat. Commun.* 14, 7164. <https://doi.org/10.1038/s41467-023-42507-9>.
- Logsdon, G.A., Ebert, P., Audano, P.A., Loftus, M., Porubsky, D., Ebler, J., Yilmaz, F., Hallast, P., Prodanov, T., Yoo, D., et al. (2025). Complex genetic variation in nearly complete human genomes. *Nature* 644, 430–441. <https://doi.org/10.1038/s41586-025-09140-6>.
- Lotfollahi, M., Bright, C., Skorobogat, R., Dehkordi, M., George, X., Richter, S., Topalova, A., Lücken, M., and Theis, F. (2024). ArchMap: A web-based platform for reference-based analysis of single-cell datasets. Preprint at bioRxiv. <https://doi.org/10.1101/2024.09.19.613883>.
- McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122. <https://doi.org/10.1186/s13059-016-0974-4>.
- Nano, P.R., Fazzari, E., Azizad, D., et al. (2025). Integrated analysis of molecular atlases unveils modules driving developmental cell subtype specification in the human cortex. *Nat. Neurosci* 28, 949–963. <https://doi.org/10.1038/s41593-025-01933-2>.
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bizkadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2022). The complete sequence of a human genome. *Science* 376, 44–53. <https://doi.org/10.1126/science.abj6987>.
- Pantazis, C.B., Yang, A., Lara, E., McDonough, J.A., Blauwendraat, C., Peng, L., Oguro, H., Kanaujiya, J., Zou, J., Sebesta, D., et al. (2022). A reference human induced pluripotent stem cell line for large-scale collaborative studies. *Cell Stem Cell* 29, 1685–1702.e22. <https://doi.org/10.1016/j.stem.2022.11.004>.
- Pirtity, M.K., Locker, J., and Schreiber-Agus, N. (2005). Rybp/DEDAF Is Required for Early Postimplantation and for Central Nervous System Development. *Mol. Cell Biol.* 25, 7193–7202. <https://doi.org/10.1128/MCB.25.16.7193-7202.2005>.
- Quadrato, G., Brown, J., and Arlotta, P. (2016). The promises and challenges of human brain organoids as models of neuropsychiatric disease. *Nat. Med.* 22, 1220–1228. <https://doi.org/10.1038/nm.4214>.
- Quadrato, G., Nguyen, T., Macosko, E.Z., Sherwood, J.L., Min Yang, S., Berger, D.R., Maria, N., Scholvin, J., Goldman, M., Kinney, J.P., et al. (2017). Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* 545, 48–53. <https://doi.org/10.1038/nature22047>.



- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339. <https://doi.org/10.1093/bioinformatics/bts378>.
- Ryan, M., McDonough, J.A., Ward, M.E., Cookson, M.R., Skarnes, W.C., and Merkle, F.T. (2024). Large structural variants in KOLF2.1J are unlikely to compromise neurological disease modeling. *Cell Stem Cell* 31, 290–291. <https://doi.org/10.1016/j.stem.2024.02.006>.
- Shi, Y., Kirwan, P., Livesey, F.J., Castagnetti, F., Yentuer, Z., Becker, K., Matilainen, J., Casadei, N., and Mayer, S. (2026). Comparing the impact of sample multiplexing approaches for single-cell RNA-sequencing on downstream analysis using cerebellar organoids. *Nat. Protoc.* 29, 1836. <https://doi.org/10.1016/j.isci.2026.114780>.
- Skarnes, W.C., Pellegrino, E., and McDonough, J.A. (2019). Improving homology-directed repair efficiency in human stem cells. *Methods* 164–165, 18–28. <https://doi.org/10.1016/j.ymeth.2019.06.016>.
- Srinivasaraghavan, V.N., Zafar, F., and Schüle, B. (2022). Gene Expression Analysis in Stem Cell-derived Cortical Neuronal Cultures Using Multi-well SYBR Green Quantitative PCR Arrays. *Bio. Protoc.* 12, e4476. <https://doi.org/10.21769/BioProtoc.4476>.
- Sun, N., Meng, X., Liu, Y., Song, D., Jiang, C., and Cai, J. (2021). Applications of brain organoids in neurodevelopment and neurological diseases. *J. Biomed. Sci.* 28, 30. <https://doi.org/10.1186/s12929-021-00728-4>.
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., Gable, A.L., Fang, T., Doncheva, N.T., Pyysalo, S., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51, D638–D646. <https://doi.org/10.1093/nar/gkac1000>.
- Velasco, S., Kedaigle, A.J., Simmons, S.K., Nash, A., Rocha, M., Quadrato, G., Paulsen, B., Nguyen, L., Adiconis, X., Regev, A., et al. (2019). Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* 570, 523–527. <https://doi.org/10.1038/s41586-019-1289-x>.
- Verberne, E.A., Goh, S., England, J., van Ginkel, M., Rafael-Croes, L., Maas, S., Polstra, A., Zarate, Y.A., Bosanko, K.A., Pechter, K.B., et al. (2021). JARID2 haploinsufficiency is associated with a clinically distinct neurodevelopmental syndrome. *Genet. Med.* 23, 374–383. <https://doi.org/10.1038/s41436-020-00992-z>.
- Vértesy, Á., Eichmüller, O.L., Naas, J., Novatchkova, M., Esk, C., Balmaña, M., Ladstaetter, S., Bock, C., von Haeseler, A., and Knoblich, J.A. (2022). Gruffi: an algorithm for computational removal of stressed cells from brain organoid transcriptomic datasets. *EMBO J.* 41, EMBJ2022111118. <https://doi.org/10.15252/embj.2022111118>.
- Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15. <https://doi.org/10.1186/s13059-017-1382-0>.
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation* 2, 100141. <https://doi.org/10.1016/j.xinn.2021.100141>.
- Yentür, Z., Kagermeier, T., Sariyeva, K., Jarbou, M.A., Becker, K., and Mayer, S. (2025). Human dorsal forebrain organoids show differentiation-state-specific protein secretion. *iScience* 28, 112935. <https://doi.org/10.1016/j.isci.2025.112935>.
- Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience* 9, giaa151. <https://doi.org/10.1093/giga-science/giaa151>.



STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Single-cell RNA sequencing data of day 90 forebrain organoids – processed data	This study	https://cellxgene.cziscience.com/collections/10d0c48e-0a15-406f-9d5e-40a2dda6e705
Single-cell RNA sequencing data of day 90 forebrain organoids – raw data	This study	https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-17432?query=E-MTAB-17432
KOLF Optical Bionano data	This study	ERZ27120469
PacBio HiFi Sequencing data	This study	PRJNA1290418
PacBio HiFi Sequencing SV callset	This study	https://zenodo.org/records/17955237
Experimental models: Cell lines		
Human: KOLF2.1J hiPSC	The Jackson Laboratory	RRID:CVCL_B5P3
Human: BIONi010-C hiPSC	EBiSC	RRID:CVCL_1E68
Human: HMGU1 hiPSC	Human Pluripotent Stem Cell Registry	RRID:CVCL_YT30
Human: SUSL025 hiPSC	WiCell Research Institute	hPSCreg: MJFFi004-A
Software and algorithms		
HiFiasm (v0.23.0)	Cheng et al., 2021 ; Cheng et al., 2022 ; Cheng et al., 2024	https://github.com/chhylp123/hifiasm
PAV(v2.4.6)	Ebert et al., 2021	https://github.com/EichlerLab/pav
Ensembl Variant Effect Predictor (v115)	McLaren et al., 2016	https://github.com/Ensembl/ensembl-vep
bedtools (v2.31.1)	Quinlan and Hall, 2010	https://github.com/arq5x/bedtools2
Miniforge (v24.3.0)	conda-forge community	https://github.com/conda-forge/miniforge
Minimap2 (v2.28)	Li, 2018	https://github.com/lh3/minimap2
Samtools (v1.21)	Danecek et al., 2021	https://github.com/samtools/samtools
Python (v3.12.3)	Python Software Foundation	https://www.python.org
R (v4.4.1)	R Core Team	https://www.r-project.org
Scanpy (v1.10.4)	Wolf et al., 2018	https://github.com/scverse/scanpy
Seurat (v5.1.0)	Hao et al., 2024	https://satijalab.org/seurat/
SoupX (v1.6.2)	Young and Behjati, 2020	https://github.com/constantAmateur/SoupX
scDblFinder (v1.18.0)	Germain et al., 2022	https://github.com/plger/scDblFinder
Gruffi (v1.5.6)	Vértesy et al., 2022	https://github.com/jn-goe/gruffi
VoxHunt (1.1.0)	Fleck et al., 2021	https://github.com/quadbio/VoxHunt
clusterProfiler (v4.12.6)	Wu et al., 2021	https://github.com/YuLab-SMU/clusterProfiler
STRING (v12.0)	Szklarczyk et al., 2023	https://string-db.org
CellChat (v2.1.2)	Jin et al., 2025	https://github.com/jinworks/CellChat

(Continued on next page)



Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bionano Access (v1.8.1.1)	Bionano Genomics, Inc.	https://bionano.com/software-downloads
Bionano Solve (v.3.8.1.1)	Bionano Genomics, Inc.	https://bionano.com/software-downloads
Oligonucleotides		
qPCR primer sequences	This paper	See Table S6

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human donor samples used to generate the iPSC line (SUSL-25) were obtained under a protocol approved by the Stanford University Institutional Review Board (IRB-48895). Written informed consent was obtained from study participants for the collection and research use of their sample. The SUSL-25 iPSC line used in this study was reviewed and approved for use under the Stanford Stem Cell Research Oversight (SCRO) protocol SCRO-754. Human subjects research was conducted in accordance with the ethical principles of the Belmont Report and applicable federal and institutional requirements.

Culture and clonal expansion of KOLF2.1J iPSCs for LRS

The KOLF2.1J iPSC line was maintained under feeder-free conditions as described by Skarnes et al., 2019. Cells were cultured on Synthemax II-SC (Corning, Cat. no.3535)-coated tissue culture plates in StemFlex Medium (Thermo Fisher Scientific, Cat. no.A3349401) supplemented according to the manufacturer's instructions. Cultures were maintained at 37°C with 5% CO₂, and medium was replaced every 2 days. Cells were routinely passaged at approximately 80% confluence using ReLeSR (STEMCELL Technologies, Cat. no.05872) at an approximate 1:15 ratio. Briefly, cultures were rinsed with DPBS, incubated with ReLeSR for 5 min at room temperature, and detached as small multicellular aggregates by gentle scraping in StemFlex medium. Cellular aggregates were transferred onto freshly Synthemax-coated culture vessels. For genomic DNA preparation, KOLF2.1J cultures were expanded under these conditions until approximately 2×10^7 viable cells were obtained. Cells were harvested at 80–90% confluence by dissociation with Accutase (STEMCELL Technologies, Cat. no.07920) for 7–10 min at 37°C to generate a single-cell suspension.

iPSC culture for Bionano optical genome mapping and gene expression analysis

Human iPSC cell lines KOLF2.1J (male, The Jackson Laboratory) and SUSL25 (female, WiCell Research Institute) were cultured in StemFlex media (Thermo Fisher Scientific, A3349401) on Matrigel plates (ThermoFisher Scientific, CB-40230). Cells were passaged with ReLeSR (StemCell Technologies, 05872) at 80% confluence with supplementation of media with 1 μM Thiazovivin (ReproCell, 04-0017) and incubated at 37°C and 5% CO₂. Media was changed the next day after passaging and every other day during cell culture maintenance.

iPSC culture for organoid differentiation

Human male iPSC cell lines KOLF2.1J (Source: The Jackson Laboratory), BIONi010-C (Source: EBiSC), and HMGU1 (Source: Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH)) (Kunze et al., 2018) were cultured in standard conditions (37°C, 5% CO₂, and 100% humidity) using E8 flex medium (Gibco, Cat. no. A2858501) for BIONi010-C, and mTESR Plus medium (STEMCELL Technologies, Cat. no. 100-0276) for KOLF2.1J and HMGU1 cell lines. iPSCs were cultured on Geltrex-coated (Thermo Fisher Scientific, Cat. no. A1413302) plates. Subculturing of the iPSCs was performed using Gentle Dissociation Reagent (STEMCELL Technologies, Cat. no. 07174), and the culture media was supplemented with 2 μM Thiazovivin (Sigma, Cat. no. 420220). The next day, a fresh media change was performed to remove Thiazovivin, and a fresh media change was performed every other day to maintain the iPSCs. Upon thawing, all cell lines were tested for pluripotency using anti-OCT4 antibody (rabbit, 1:500, Abcam, Cat. no. ab19857) and for mycoplasma contamination (TaKaRa, Cat. no. 6601). For organoid differentiation, iPSC cell lines below passage 20 were used. More specifically, the passage number of iPSCs when we started differentiation of forebrain organoids (FOs) was 13 for KOLF2.1J, 18 for BIONi010-C, and 16 for HMGU1.

Forebrain organoid differentiation

Previously published protocol was used to generate FOs from KOLF2.1J, BIONi010-C, and HMGU1 iPSCs, with minor modifications (Velasco et al., 2019). Shortly, iPSCs were dissociated into single cells using Accutase (MERCK, Cat. no.



A6964). 9,000 cells per well in respective iPSC media were seeded in 96-well V-bottom low attachment plates (S-BIO, Cat. no. MS-9096VZ). The next day, the media was replaced with cortical differentiation media I (CDMI) supplemented with 20 μ M ROCK inhibitor Y-27632 (Cayman Chemical, Cat. no. 10005583), 5 μ M SMAD inhibitor SB 431542 (Tocris, Cat. no. 1614), and 3 μ M WNT signaling inhibitor IWR-1 (MERCK, Cat. no. 681669). From day 6 of differentiation, CDMI media was supplemented with only SB and IWR-1. On day 18, media was switched to CDMII and FOs were transferred to 24-well low-attachment plates, and placed on an orbital shaker (New Brunswick S41i, 2.5 cm throw, 56 rpm) in standard conditions. Fresh media change was performed every 3–4 days. On day 35, the media was switched to CDMIII supplemented with 1% Matrigel (Corning, Cat. no. 356234), and media change was performed twice a week. From D70 on, the media was switched to CDMIV supplemented with 1% Matrigel (Corning, Cat. no. 356234), and media change was performed twice a week. The organoids were cultured until D90, and 3 FOs per cell line were processed individually for single-cell RNA sequencing.

METHOD DETAILS

Long read sequencing

Data generation. HMW DNA was extracted from 20 million cells (Source: The Jackson Laboratory; passage 4) using the Gentra Puregene Kit (Qiagen). Briefly, the cells were resuspended in PBS and lysed. The lysate was treated with Proteinase K and RNase A. DNA was then precipitated with isopropanol, washed with 70% ethanol and resuspended in Elution Buffer (PacBio). Each library was prepared using the SMRTbell Express Template Prep Kit 2.0 (PacBio) and SMRTbell Enzyme Cleanup Kit (PacBio). First, 12 μ g of HMW DNA was sheared to 15–18 kbp size range using g-TUBE (Covaris). Sheared DNA was subjected to end repair and A-tailing treatment and ligated with V3 SMRTbell adapter (PacBio). The adapter-ligated library was treated with nuclease to remove non-intact SMRTbell templates and purified. The purified library was size selected using Pippin HT (Sage Biosciences) generating a library that has a size >10kb. The size selected library was used for sequencing on PacBio Sequel II with 30h movie time. A total of 3 SMRT cells were sequenced, generating 98G of HiFi yield.

Bionano optical genome mapping

DNA extraction. Ultra-high molecular weight DNA was extracted from cell pellets for KOLF2.1J using the Bionano Prep SP-G2 Blood & Cell Culture DNA Isolation Kit (Bionano Genomics, PN 80060) following the Bionano protocol CG-00004. **DNA labeling.** DNA samples were labeled according to the manufacturer's instructions (Bionano Genomics, PN, 20440). Chips were imaged through the Bionano Saphyr instrument's fluorescent microscopy, with a throughput target of 1500 GBp (corresponding to approximately 400x target coverage). Imaging generated a dataset of molecules with a total length of 1.504 trillion base pairs, an average molecule N50 > 150kbp of 241.4, and N50 > 20kbp 178.1.

SYBR green expression analysis

Cell culture. To identify genes where overlap with SVs affected expression in KOLF2.1J, we performed qPCR analysis utilizing a previously established 384-well SYBR Green qPCR expression array (Chen et al., 2023; Srinivasaraghavan et al., 2022), using passages 17, 18, 19, and 29 iPSCs from KOLF2.1J. To monitor consistency of growth rate between lines, lines were imaged and passaged at the same time. Media was changed every two days and cells were passaged every four days. During passaging, iPSCs were dissociated with Accutase (Thermo Fisher, NC9464543) and washed with PBS before centrifuging for 3 min at 1,000 rpm at room temperature, then resuspended in 1 mL of StemFlex (Thermo Fisher, A3349401). Cells were counted by loading 10 μ L of 1:1 resuspended cells in StemFlex solution mixed with Trypan Blue (Sigma-Aldrich, T8154-100ML) onto a hemocytometer and imaging the hemocytometer with a Corning Cell Counter (Corning, 6749). The cell counter was gated using a minimum cell size of 6 for each line. During each passage, 250,000 cells were reseeded into new wells. Cells were collected by adding ~1.5 million cells to 1.5 mL Eppendorf tubes and an equal volume of PBS. Those tubes were then centrifuged for 3 min (3,000 rpm at room temperature) to create a pellet. All media/PBS mix was aspirated away from the pellet. The pellet was then resuspended with 1 mL of PBS and again centrifuged at 3,000 rpm for 3 min at room temperature and then any excess supernatant was aspirated out. Pellets were stored in -80°C .

RNA extraction. RNA was extracted using the PurelinkTM RNA Mini Kit (Thermo Fisher, 12-183-018A) protocol with DNase treatment. The RNA concentration was quantified by NanoDrop, and 1 μ g of RNA was used to generate cDNA using the High-Capacity cDNA Reverse Transcription (Fisher Scientific, 43-688-14) kit. A control sample iPSC line SUSL25 (passages 18, 19, 20, and 29; Source: Schuele Lab, Stanford University, 70 years old, female), which does not bear the SVs found in KOLF2.1J, was also extracted and reverse transcribed. Following reverse transcription, cDNA (50ng/ μ L) was diluted 5:1 with nuclease-free water to reach a concentration of 10ng/ μ L and then stored at -20°C .



Primer plating. The day before the SYBR Green assay, both forward and reverse primers for the genes of interest (Table S6) were pre-plated in four replicates at a final concentration of 300 nM each into a 384-well optical PCR plate. Beta-actin was chosen for normalization purposes. Primers were designed according to the protocol in (Srinivasaraghavan et al., 2022). Additional wells with primers were included for the no template control. To dry out the primers, the plate was kept in a closed Ziploc bag at room temperature for 12 h.

qPCR. For the SYBR Green Assay, we added reaction mix (2.5 μ L of Applied Biosystems PowerUp SYBR Green Master Mix (Thermo Scientific, A25776), 1.5 μ L of nuclease-free water, and 1 μ L of 10ng/ μ L sample cDNA) to the primer-coated wells and sealed the plate with optical adhesive film. Following centrifugation at 1,000 rpm for 3 min, the SYBR Green PCR plate was run using a QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher, 4485691). The cycling conditions used are indicated in Table S7, while Table S8 shows the conditions used to assess the melt curve of the PCR product. The melt curve was used to validate the absence of primer dimers.

Single-cell RNA sequencing

On day 90 of differentiation, 3 organoids per cell line were dissociated into single cells following a published protocol with minor modifications (Velasco et al., 2019). Cell multiplexing was performed using Cell Multiplexing Oligos from 3' CellPlex Kit Set A (10x Genomics, Cat. no. PN-1000261), and cells were further processed using the protocol from the manufacturer. Multiplexed cells were counted and pooled in equal proportions to be processed by CeGat Tübingen for single-cell sequencing. Approximately 25,000 cells per lane were loaded to two lanes of the Chromium chip (10x Genomics) and processed through the Chromium controller. Sequencing libraries were prepared using Chromium Next GEM Single Cell 3' Kit, v3.1 (10x Genomics) and sequenced on NovaSeq 6000 (Illumina) with S4 flow cell and 200 cycles. Cell lines were treated as separate batches for downstream analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

Long read sequencing

SV calling. Raw KOLF2.1J PacBio HiFi reads were assembled using HiFiasm (v0.23.0) and SV deletions, insertions and inversions were called using PAV(v2.4.6) aligned against both GRCh38/hg38 and T2T-CHM13 references. PAV has the highest percent of supported calls across dozens of read-based and assembly-based approaches (99.7%) (Logsdon et al., 2025). Moreover, when examining trio-based Mendelian inheritance, the error rates are (with reference to the GRCh38) 2.7% for SVs, 4.5% for Indels, and 0.5% for SNVs. These error rates drop further to 1.2%, 3.7%, and 0.3%, respectively, when tandem repeats and SDs are excluded from analyses. Reference sequences and all relevant tracks from GRCh38 and T2T-CHM13 were downloaded from UCSC genome browser and (<https://github.com/marbl/CHM13>) respectively, unless noted otherwise. Additionally, raw HiFi reads were aligned to both GRCh38 and T2T-CHM13 references using pbmm2 (v1.13.1) and DELLY (v1.2.9) was used to identify tandem duplications.

Ensembl Variant Effect Predictor (VEP): We used VEP (v115) (McLaren et al., 2016) to identify SVs that impacted genes and overlapped known variants in the population (gnomAD SV database v4.1) (Collins et al., 2020).

Segmental Duplication (SD) annotation: We used bedtools (v2.31.1) to identify SVs within SD regions of the genome with $\geq 90\%$ sequence overlap.

Link to the unfiltered variant calls, and VEP results can be found in the [resource availability](#) section.

Bionano optical genome mapping

Data generation. SVs were generated from molecule reads using the Guided Assembly workflow of Bionano Solve version 3.8.1.1, using either T2Tchm13v2.0 or hg38 as the chosen reference genome. SVs generated this way were filtered based on the following criteria: valid SVs contain at least 20 molecule reads (20 being the number of reads for a 0.05 allele frequency SV at coverage 400x) that completely span all markers in the SV assembly and do not support the original reference, have a minimum allele frequency of 0.375 (0.5–0.125), and are not present in the Bionano control database, a repository of SVs from 285 healthy individuals. Between the extraction, labeling, read generation, and analysis steps, we also ran quality control measurements to ensure fidelity of reads gathered. DNA concentrations were measured with a Qubit fluorometer.

Concordance between LRS and OGM data

To compare LRS and OGM calls to identify overlapping SVs, we excluded variants identified by LRS in the alternative contigs of GRCh38 and all inversion calls (both references). This was done due to the lack of variants from OGM within the



alternative contigs of GRCh38 and the uncertainty of the breakpoints for OGM inversion calls. After applying the above filters, bedtools window (Quinlan and Hall, 2010) was used to identify concordance between the two technologies, with a 20% margin when determining the SV size overlap.

For this study, we focused on.

1. SVs identified by both LRS and OGM technologies.
2. SVs present outside SD (segmental duplication) regions of GRCh38.
3. SVs tagged as HIGH impact by VEP).
4. SVs absent in the population (gnomAD SV database v4.1).

SYBR green expression analysis

Comparisons between SUSL25 and KOLF2.1J were performed on ΔC_t values using two-tailed unpaired t-tests with Welch's correction. P-values were adjusted for four comparisons using the Bonferroni method.

Single-cell RNA sequencing data analysis

Single-cell transcriptomic data were processed using the Cell Ranger multi pipeline (v7.1.0) and aligned to the 10x Genomics hg38-2020-A reference genome. Downstream analysis was performed using Scanpy (v1.10.4) (Wolf et al., 2018) in Python (v3.12.3) and Seurat (v5.1.0) (Hao et al., 2024) in R (v4.4.1). Low-quality cells were excluded based on a median absolute deviation threshold of 3 for total counts, number of genes by counts, mitochondrial and ribosomal content. Genes expressed in fewer than 20 cells were removed. Ambient RNA contamination and doublets were filtered using SoupX (v1.6.2) (Young and Behjati, 2020) and scDblFinder (v1.18.0) (Germain et al., 2022), respectively.

Stressed cells were identified using Gruffy (v1.5.6) (Vértesy et al., 2022) based on high ER stress (GO:0034976) and glycolysis (GO:0006096), and low gliogenesis (GO:0042063) and neurogenesis (GO:0022008), applying a 0.9 quantile threshold. These cells were removed prior to re-normalization and dimensionality reduction. Batch correction was performed using Harmony. Forebrain identity was validated via ArchMap (Lotfollahi et al., 2024) using a human fetal brain transcriptomic atlas (Braun et al., 2023), and a spatial correspondence assessment with E18 mouse brain three-dimensional *in situ* hybridization data using VoxHunt (v1.1.0) (Fleck et al., 2021). Cell populations were annotated using marker gene expression and reference-to-query integration with a published meta-atlas comprising seven single-cell transcriptomic datasets (Nano et al., 2025). Cell type proportions were calculated for each organoid (3 per cell-line), and statistical significance between groups (KOLF2.1J vs. BIONi010-C/HMGU1-derived FOs) was determined using pairwise Welch's t-tests with Benjamini-Hochberg correction (FDR), and a threshold of $FDR < 0.05$ and $|\log_2 FC| > 1.0$. Pseudotime was inferred using *Diffusion Pseudotime* (DPT) as implemented in Scanpy (scanpy.tl.dpt). Root cells were defined as cycling progenitors for the inhibitory lineage and intermediate precursor cells for the excitatory lineage.

Differential gene expression analysis was performed at the single-cell level using Welch's t-tests with Benjamini-Hochberg correction (Scanpy v1.10.4). To isolate robust KOLF2.1J-specific signatures, KOLF2.1J-derived cells were compared independently against each control line (BIONi010-C and HMGU1). Genes were defined as significantly differentially expressed if they met thresholds of $FDR < 0.05$ and $|\log_2 FC| > 1.0$ in both KOLF2.1J-to-control comparisons and were either not significantly differentially expressed between control lines or exhibited a higher $|\log_2 FC|$ and lower FDR in both KOLF2.1J-to-control comparisons. No pseudobulking was performed. Functional annotation of the transcriptomic changes was performed with clusterProfiler (v4.12.6) (Wu et al., 2021) using two complementary approaches: GO analysis using the sets of DEGs identified for each cell type; GSEA using all detected genes ranked by $-\log_{10}(FDR) \times \text{sign}(\log_2 FC)$. For both GO analysis and GSEA, *p*-values were adjusted with the Benjamini-Hochberg method, and only terms with an $FDR \leq 0.05$ were considered as enriched.

A PIP network was constructed using STRING (v12.0) (Szklarczyk et al., 2023), considering high-confidence interactions and evidence from text mining, experimental data, curated databases, co-expression, neighborhood, gene fusion, and co-occurrence. Functional clusters were identified using MCL clustering (inflation 1.1).

Cell-cell communication was modeled with CellChat (v2.1.2) (Jin et al., 2025) via the tri-mean method, excluding interactions involving fewer than 10 cells.