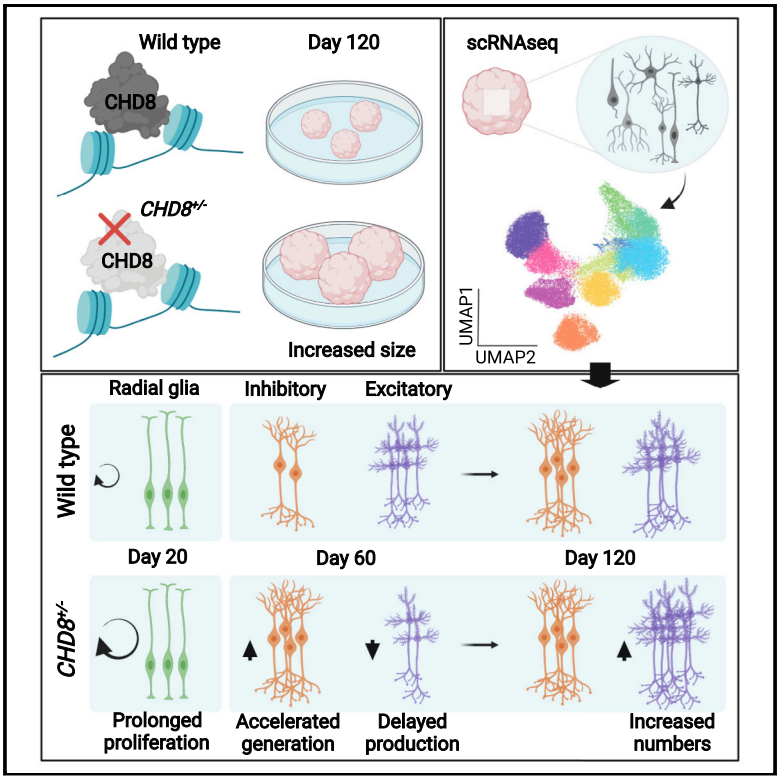
Cell Reports

Article

CHD8 haploinsufficiency links autism to transient alterations in excitatory and inhibitory trajectories

Graphical abstract



Authors

Carlo Emanuele Villa, Cristina Cheroni, Christoph P. Dotter, ..., Johann G. Danzl, Giuseppe Testa, Gaia Novarino

Correspondence

giuseppe.testa@fht.org (G.T.), gnovarino@ist.ac.at (G.N.)

In brief

Using cerebral organoids, Villa et al. show that *CHD8* haploinsufficiency disrupts neurodevelopmental trajectories by promoting an accelerated generation of inhibitory neurons and a delayed production of excitatory neurons with a temporally restricted cell-type-specific effect on proliferation of radial glial cells.

Highlights

- *CHD8*-mutant cerebral organoids recapitulate a macrocephaly-like phenotype
- Proliferation and differentiation dynamics are affected by *CHD8* mutations
- *CHD8* haploinsufficiency disrupts neurodevelopmental trajectories
- mRNA processing in post-mitotic neurons is altered in *CHD8* mutants

Villa et al., 2022, Cell Reports *39*, 110615
April 5, 2022
https://doi.org/10.1016/j.celrep.2022.110615

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

CHD8 haploinsufficiency links autism to transient alterations in excitatory and inhibitory trajectories

Carlo Emanuele Villa,[1,2] Cristina Cheroni,[1,2,3,6] Christoph P. Dotter,[4,6] Alejandro López-Tóbon,[1,2,3,6] Bárbara Oliveira,[4,6] Roberto Sacco,[4] Aysan Çerağ Yahya,[4] Jasmin Morandell,[4] Michele Gabriele,[1] Mojtaba R. Tavakoli,[4] Julia Lyudchik,[4] Christoph Sommer,[4] Mariano Gabitto,[5] Johann G. Danzl,[4] Giuseppe Testa,[1,2,3,*] and Gaia Novarino[4,7,*]

[1]Department of Experimental Oncology, IEO, European Institute of Oncology, IRCCS, 20139 Milan, Italy
[2]Human Technopole, Viale Rita Levi Montalcini 1, 20157 Milan, Italy
[3]Department of Oncology and Hemato-oncology, University of Milan, 20122 Milan, Italy
[4]Institute of Science and Technology (IST) Austria, 3400 Klosterneuburg, Austria
[5]Allen Institute for Brain Science, Seattle, WA 98109, USA
[6]These authors contributed equally
[7]Lead contact
*Correspondence: giuseppe.testa@fht.org (G.T.), gnovarino@ist.ac.at (G.N.)
https://doi.org/10.1016/j.celrep.2022.110615

SUMMARY

Mutations in the chromodomain helicase DNA-binding 8 (*CHD8*) gene are a frequent cause of autism spectrum disorder (ASD). While its phenotypic spectrum often encompasses macrocephaly, implicating cortical abnormalities, how *CHD8* haploinsufficiency affects neurodevelopmental is unclear. Here, employing human cerebral organoids, we find that *CHD8* haploinsufficiency disrupted neurodevelopmental trajectories with an accelerated and delayed generation of, respectively, inhibitory and excitatory neurons that yields, at days 60 and 120, symmetrically opposite expansions in their proportions. This imbalance is consistent with an enlargement of cerebral organoids as an *in vitro* correlate of patients' macrocephaly. Through an isogenic design of patient-specific mutations and mosaic organoids, we define genotype-phenotype relationships and uncover their cell-autonomous nature. Our results define cell-type-specific *CHD8*-dependent molecular defects related to an abnormal program of proliferation and alternative splicing. By identifying cell-type-specific effects of *CHD8* mutations, our study uncovers reproducible developmental alterations that may be employed for neurodevelopmental disease modeling.

INTRODUCTION

Autism spectrum disorder (ASD) defines heterogeneous conditions characterized by difficulties in establishing social contacts and the manifestation of repetitive behaviors. There is no cure for the core symptoms of ASD, reflecting a lack of understanding of the events leading to these neuropsychiatric disorders. Functional genomic studies and single-cell RNA sequencing (RNA-seq) of postmortem brain samples obtained from patients with autism have highlighted alterations of transcription in pyramidal cells and local interneurons at early to mid-fetal developmental stages (Chen et al., 2014; Parikshak et al., 2013; Willsey et al., 2013); however, evidence from tractable experimental models of human neurodevelopment is still lacking. One of the major limitations is the inaccessibility of the affected tissue during development, when the bases of these disorders are eventually laid out. The possibility to generate human cerebral organoids from human embryonic stem cells (hESCs) or induced pluripotent stem cells (iPSCs) represents an opportunity to study the cellular and molecular consequences of ASD-associated mutations (Birey et al., 2021; Catlett et al., 2021; de Jong et al., 2021; Urresti et al., 2021; Wegscheid et al., 2021). Cerebral organoids represent the closest *in vitro* model of a human cortex at early embryonic stages (Camp et al., 2015; Lancaster and Knoblich, 2014), thus providing a tool to unravel the impact of gene impairment during time windows of ASD vulnerability (Willsey et al., 2013).

Genes coding for chromatin remodelers are frequently mutated in ASD patients (De Rubeis et al., 2014; Gabriele et al., 2018) and, among them, chromodomain helicase DNA-binding protein 8 (*CHD8*) is one of the most frequently mutated and most penetrant (Bernier et al., 2014; O'Roak et al., 2012a, 2012b). *CHD8*, initially identified in a screen for interactors within the canonical Wnt/beta-catenin pathway (Sakamoto et al., 2000), is part of the chromodomain helicase DNA-binding protein family, characterized by a SFN2-like ATPase and two chromo-domains. The majority of ASD-associated *CHD8* *de novo* mutations lead to loss of function (LoF) and result in gene haploinsufficiency (Beighley et al., 2020; Bernier et al., 2014; O'Roak et al., 2012a, 2012b). Patients with such mutations, in addition to autism-relevant behaviors, present with gastrointestinal complaints, intellectual disability, and macrocephaly, linking *CHD8* haploinsufficiency to abnormal cortical development (Beighley et al., 2020; Bernier et al., 2014).

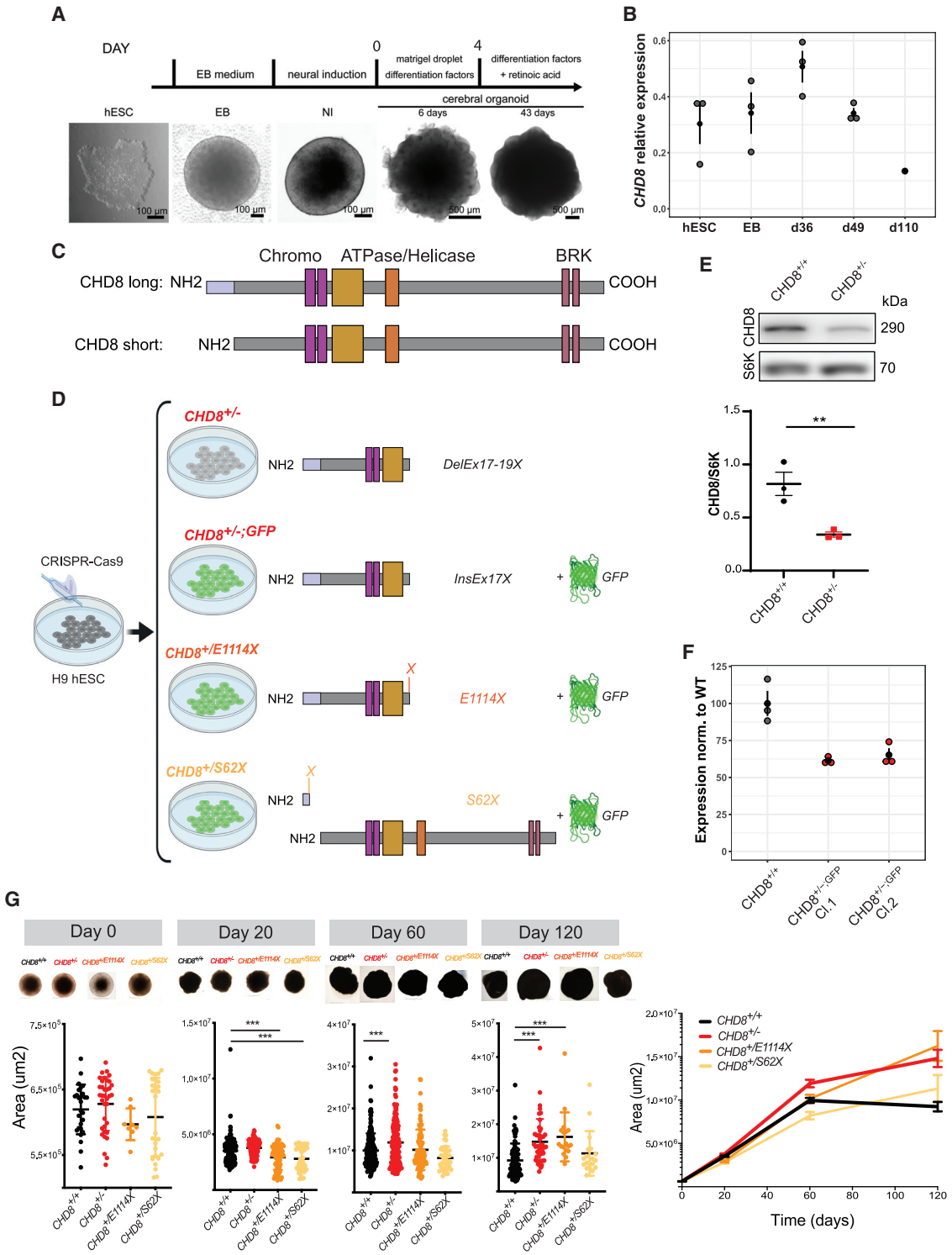


Figure 1. Increased cerebral organoid size linked to *CHD8* heterozygous mutations

(A) Outline for the generation of cerebral organoids from hESCs.

(B) Results of quantitative RT-PCR were normalized to *TBP* and relative expression values are shown for each time point. Results are presented as mean ± SEM.

(C) Schematics of the human long (L) and short (S) CHD8 isoforms. The N-terminal part specific for the long isoform is highlighted in light purple.

(D) Schematics illustrating the CRISPR/Cas9-mediated generation of the hESC lines employed in this study, the resulting genotypes, and predicted proteins. DelEx17-19X, deletion of part of exon 17, whole exon 18, and part of exon 19 and premature stop codon; InsEx17X, hrGFP was inserted into exon 17, thereby

(legend continued on next page)

Previous studies have explored the role of *CHD8* haploinsufficiency in brain development and ASD manifestations by employing animal models (Durak et al., 2016; Gompers et al., 2017; Katayama et al., 2016; Suetterlin et al., 2018) or iPSC-derived models, including cerebral organoids, profiled at a single time point and in bulk (Wang et al., 2015, 2017), thus precluding the investigation of neurodevelopmental trajectories. Importantly, while, in humans, *CHD8* LoF mutations lead to a number of severe neurological problems, in mouse they are associated with rather mild neurological abnormalities (Gompers et al., 2017; Katayama et al., 2016; Suetterlin et al., 2018). In addition, while *Chd8* haploinsufficient mice show a slight increase of brain size, *in utero* short hairpin RNA (shRNA)-mediated *Chd8* knockdown results in reduced neural progenitor cell proliferation and premature differentiation (Durak et al., 2016), a phenotype usually associated with reduced brain size. Thus, the cellular and molecular mechanisms underlying *CHD8*-associated phenotypes are still elusive, and conflicting results suggest possible technique- and species-specific issues, making it difficult to find a consensus on the role of *CHD8* in the pathogenesis of ASDs.

Here we employed a cerebral organoid model system to determine the effects of *CHD8* haploinsufficiency on human cortical development. We found that *CHD8* haploinsufficiency leads to a significant enlargement of cerebral organoids with associated features of abnormal cortical development. A time-course single-cell RNA-seq analysis of control and *CHD8* mutant organoids revealed a disrupted differentiation dynamic of excitatory and inhibitory neurons, with a longer phase of excitatory neuron progenitor proliferation and an accelerated production of inhibitory neurons. At the molecular level, our analysis uncovers cell-type-specific phenotypes that underlie differentially affected developmental trajectories and point to temporally discrete mechanisms.

RESULTS

Generation and characterization of hESC-derived cerebral organoids

We and others proposed that the combination of disease-specific organoid models and single-cell omic resolution are materializing the vision of interceptive medicine, whereby deviations from development and homeostasis are intercepted and recast as problems of cellular dynamics within tissues (Rajewsky et al., 2020). A prerequisite for this approach is the reproducibility of organoid generation in longitudinal designs, both within and across batches, so as to enable robust phenotyping. Reproducibility issues have led to the proposition of a number of protocols for the generation of cerebral organoids, each aiming to improve in- and inter-batch reproducibility and/or specifically driving the generation of defined neuronal cell types (Luo et al., 2016; Qian et al., 2016; Quadrato et al., 2017). We considered that reproducibility problems may likely arise during the ectoderm and neuroectoderm specification in embryoid bodies (EBs). Indeed, suboptimal timing in the induction of the neuroectoderm radial organization at this stage compromises the subsequent formation of the neuroepithelium in stereotypical rosette-like structures (Lancaster and Knoblich, 2014), likely indicating unstable early ectoderm formation in the EB right after hESC seeding. Building on other successful methods (Lancaster and Knoblich, 2014), we reasoned that optimizing culture conditions during the hESC seeding step might be key to further increase the robustness of EB formation. Specifically, we took inspiration from work (Camp et al., 2015) showing strong positive correlation between human fetal and organoid cell types and introduced slight variations to the initial culturing steps of the original protocol (see STAR Methods). Indeed, in our culture conditions (Figure 1A), we observed a high degree of reproducibility in the quality of EB ectoderm and neuroectoderm differentiation (Figure S1A), which, as expected, resulted in a high number of successfully differentiated cerebral organoids per number of Matrigel embedded EBs (>90%). We verified by immunostaining that the reproducibility observed at the macroscopic level was matched by a reproducible cell composition in the cortical structures (Figures S1B–S1C). Control cerebral organoids displayed the progressive features of cortical development starting with the formation of ventricular zone (VZ)-like structures, the appearance of subventricular-zone (SVZ)-like areas, and the development of organized neuronal cell layers (Figure S1D). In agreement with the developing human cortical VZ, VZ-like structures of cortical organoids at day 10 are constituted almost exclusively by Sox2-positive radial glia cells (Figure S1E). At later points, an increasing proportion of intermediate progenitors (TBR2 positive) and lower (CTIP2 positive) and upper (SATB2 positive) layer neurons can be observed, indicating the formation of structures reminiscent of the cortical plate and a layered structure (Figure S1F). As development progresses, the size of the organoids increases accordingly (Figure 1A). Finally, we checked *CHD8* expression at multiple stages and found its mRNA levels to be highest in 36-day old cerebral organoids (Figure 1B), suggesting that *CHD8* may play essential roles around that developmental stage.

introducing a premature stop codon; E1114X, the codon encoding for the glutamate in position p.1114 was mutated to generate a stop codon; S62X, the serine in position 62 was mutated to a stop codon. This affects only the long isoform, as indicated by the presence of the full protein for the short isoform. For both E1114X and S62X, hrGFP was inserted downstream and adjacent to the mutations.

(E) Western blot of $CHD8^{+/+}$ and $CHD8^{+/-}$ hESCs revealing reduction in CHD8 protein upon deletion of exon 20, normalized to S6K. p value = 0.013, unpaired t test, results are presented as mean ± SEM.

(F) Reduction in *CHD8* mRNA levels in day 60 $CHD8^{+/-;GFP}$ organoids as shown by qPCR, relative to $CHD8^{+/+}$. Error bars display technical variation from triplicates represented as mean ± SEM.

(G) Representative pictures and area of the cerebral organoids of the four different genotypes at day 0, 20, 60, and 120. For this analysis, $CHD8^{+/-}$ and $CHD8^{+/-;GFP}$ organoids were pooled together. Day 0, n($CHD8^{+/+}$) = 27, n($CHD8^{+/-}$) = 36, n($CHD8^{+/E1114X}$) = 11, n($CHD8^{+/S62X}$) = 27; day 20, n($CHD8^{+/+}$) = 196, n($CHD8^{+/-}$) = 119, n($CHD8^{+/E1114X}$) = 80, n($CHD8^{+/S62X}$) = 45; day 60, n($CHD8^{+/+}$) = 202, n($CHD8^{+/-}$) = 163, n($CHD8^{+/E1114X}$) = 106, n($CHD8^{+/S62X}$) = 41; day 120, n($CHD8^{+/+}$) = 85, n($CHD8^{+/-}$) = 46, n($CHD8^{+/E1114X}$) = 18, n($CHD8^{+/S62X}$) = 18 (n indicates number of organoids). ***p < 0.001, ordinary one-way ANOVA followed by Dunnett's multiple comparisons test, results are presented as mean ± SEM.

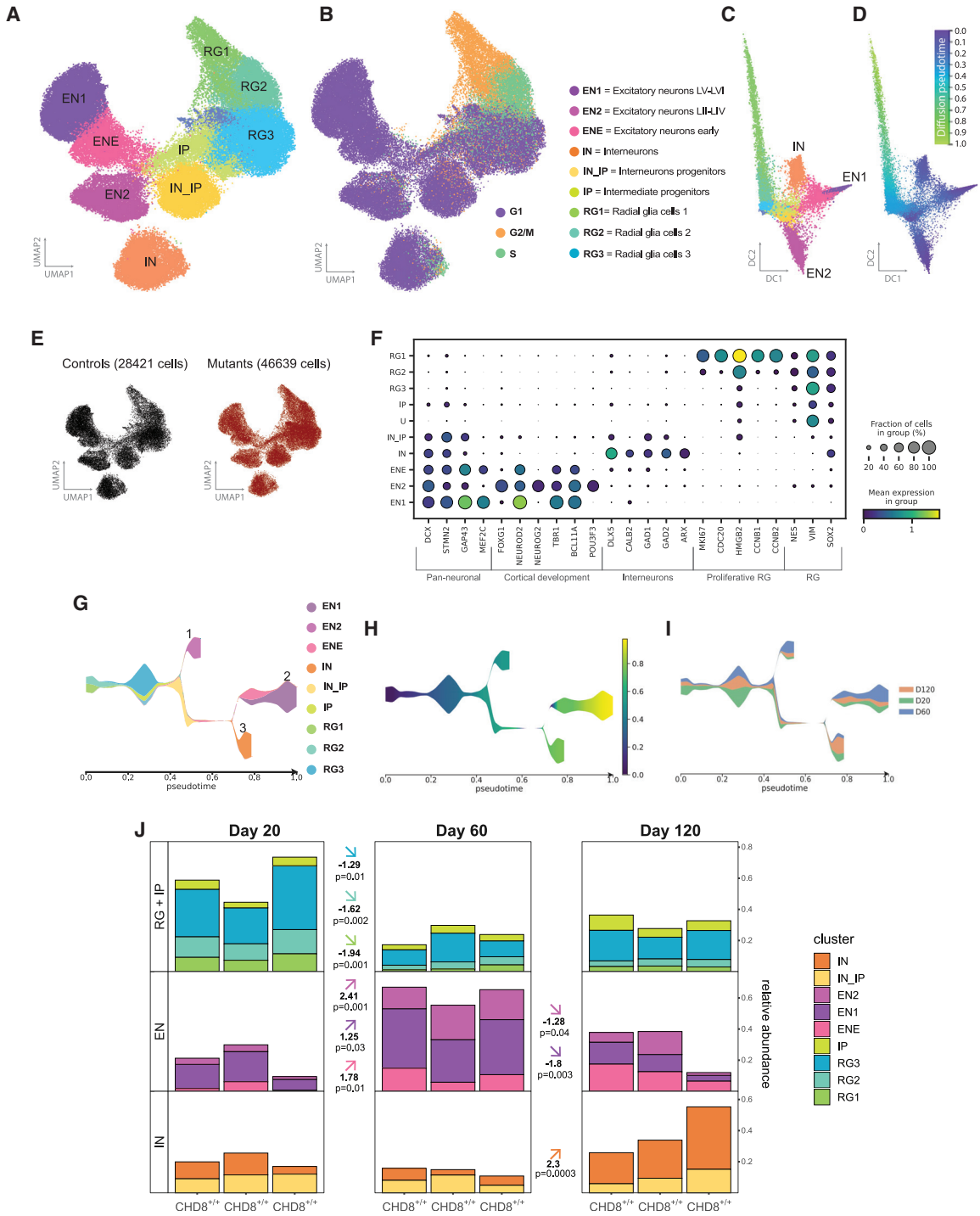


Figure 2. Cerebral organoids recapitulate neurogenic lineages

(A) UMAP projection of droplet-based scRNA-seq from cerebral organoids profiled at day 20, day 60, and day 120 *in vitro*, with colored clusters for: EN1, excitatory neurons layers V–VI; EN2, excitatory neurons layers II–IV; ENE, excitatory neurons early; IN, interneurons; IN_IP, interneuron intermediate progenitors; IP, intermediate progenitor; RG1, radial glia cells 1; RG2, radial glia cells 2; RG3, radial glia cells 3.

(B) The same UMAP plot color coded to highlight the cell cycle stages; purple for G1, orange for G2/M, and green for S.

(C and D) Diffusion map depicting the evolution of the cerebral organoid cell populations. Color code is set according to (C) cell populations and (D) pseudotemporal inferred trajectories, with the three terminal populations colored in blue, while the common origin is colored in green.

(E) Total number of cells profiled per condition, three control (CTL) and four mutant (MUT) samples for each time point. Each sample contained a pool of three cerebral organoids obtained from at least two independent differentiation batches.

(legend continued on next page)

Generation of isogenic *CHD8* haploinsufficient hESC lines

We built on our previous transcriptomics-based disease modeling benchmark (Germain and Testa, 2017) and adopted an isogenic experimental design introducing *CHD8* mutations in an isogenic hESC line employing the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 genome engineering technology (Ran et al., 2013a, 2013b).

In humans the *CHD8* gene has a short and a long isoform, with the former lacking the first 279 N-terminal amino acids (Figure 1C). To study the effect of *CHD8* LoF mutations, we generated two separate hESC lines ($CHD8^{+/-}$ and $CHD8^{+/-;GFP}$) carrying a deletion in the genomic portion of *CHD8* encoding for the C-terminal helicase domain, as well as two cell lines carrying patient-specific mutations resulting in premature stop codons (i.e., S62X and E1114X) (Figure 1D). Importantly, while both patient mutations have been linked to the autism core symptoms, the S62X patient does not present with macrocephaly and intellectual disability (Beighley et al., 2020; Bernier et al., 2014), indicating that the S62X mutation results in only a subset of the *CHD8* clinical phenotypes. In all but one case (i.e., $CHD8^{+/-}$) we coupled *CHD8* mutations with the expression of enhanced green fluorescence protein (eGFP) (Figure 1D). hESCs carrying the selected *CHD8* mutations on a single allele (i.e., $CHD8^{+/-}$, $CHD8^{+/-;GFP}$, $CHD8^{+/E1114X;GFP}$, $CHD8^{+/S62X;GFP}$) were isolated, expanded, and analyzed for potential off-target effects (Figure S2 and STAR Methods). Next, we tested the effect of the mutations on *CHD8* protein and mRNA levels by immunoblotting and qPCR. We confirmed that all the mutants show reduced CHD8 protein (Figures 1E and S3C) and mRNA levels (Figure 1F).

CHD8 haploinsufficiency leads to a macrocephaly-like phenotype in human cerebral organoids

We employed the *CHD8* mutant and control hESC lines to generate human cerebral organoids. Similar to the mutant hESCs, *CHD8* mutant organoids show a 50% reduction of CHD8 protein levels compared with $CHD8^{+/+}$ samples (Figure S3C). At very early stages (day 0–20), control ($CHD8^{+/+}$) and mutant ($CHD8^{+/-}$, $CHD8^{+/E1114X}$, $CHD8^{+/S62X}$) cerebral organoids are mostly indistinguishable, with equal appearance, comparable neuroectoderm thickness, and identical or slightly different size (Figure 1G). This observation was in contrast to what was expected given that LoF mutations in *CHD8* are associated with macrocephaly. We thus let the cerebral organoids grow longer and found that, by day 60, $CHD8^{+/-}$ and $CHD8^{+/E1114X}$ organoids are comparable with or significantly larger than wild-type samples (Figure 1G). This trend becomes even more obvious by day 120, when $CHD8^{+/-}$ and $CHD8^{+/E1114X}$, but not $CHD8^{+/S62X}$, cerebral organoids are about 50% larger than controls (Figure 1G). Thus, *CHD8* haploinsufficiency results in enlargement of human cortical organoids between day 20 and day 120 (Figure 1G). Importantly, S62X has hardly any effect on organoid size by day 120, confirming that this mutation is not associated with macrocephaly and underscoring the sensitivity of our organoid system in capturing pathophysiological readouts of patient-specific mutations.

Effects of *CHD8* haploinsufficiency at single-cell resolution

To define the cellular mechanisms underlying *CHD8* haploinsufficiency phenotypes, we comprehensively analyzed control and mutant organoids by droplet-based single-cell RNA-seq (scRNA-seq). Specifically, we profiled gene expression of 75,060 cells obtained from a longitudinal cohort amounting to a total of 63 control and mutant cerebral organoids (Figures 2A and 2E) at three different developmental stages (i.e., day 20, day 60, and day 120) (Figure 2A). After data integration (see STAR Methods), application of Leiden, a community detection algorithm optimized to identify communities guaranteed to be connected and hence to reliably score cell population based phenotypes (Traag et al., 2019), yielded 10 different cell clusters. To characterize the resulting cell clusters, we next adopted an integrative annotation strategy combining (1) supervised plotting of known cell identity markers (Figure S4A); (2) cluster-wise overlap of cluster identity markers in organoids versus human fetal cortex (Nowakowski et al., 2017); and (3) integration, in the same dimensionality reduction space (Uniform Manifold Approximation and Projection [UMAP]), of the cell populations from our cerebral organoid dataset with the ones from a large public dataset of fetal human primary cells spanning several developmental stages (Pollen et al., 2019) (Figures S4C–S4D). This annotation strategy led to the discrimination of 10 cell populations, including three clusters of radial glia cells comprising different cell cycle phases (Figures 2A–2D; RG1 includes mostly cells in G2M phase, RG2 in S phase, and RG3 in G1 phase), intermediate progenitors (IPs), interneuron progenitors (IN_IP), interneurons (INs), early excitatory neurons (ENEs), excitatory neurons of upper (EN1) and lower (EN2) layers, along with a minor number (2.9%) of unidentified cells (Figures 2A and S4). To determine the evolution of the system in terms of temporal relationships and ordering across cell populations, we performed a diffusion pseudotime (DPT) analysis, uncovering a hierarchical organization starting from actively proliferating radial glia cells and later

(F) Expression of canonical markers for cortical neuronal classes across clusters. The size of the dots represents the percentage of cells that express the gene (x axis) in the given population (y axis), while the color gradient depicts the normalized level of expression.

(G–I) Tree graphs, calculated in control lines, indicating how cells are distributed in a pseudo-temporal trajectory (x axis); the thickness of each branch represents the percentage of cells in that pseudotime frame. The graph's structure highlights the division into three main lineages connecting the origin (on the left at time 0), identified as cycling radial glia, with endpoints (1) excitatory neurons LII–IV, (2) excitatory neurons LV–VI, and (3) interneurons. The color code is according to (G) cell populations, as identified in UMAP clusters; (H) total pseudotime calculated on all lineages; or (I) cell distribution along pseudotime for each stage (day 20 in green, day 60 in blue, and day 120 in orange), normalized considering the total number of cells in control lines profiled at that time point.

(J) Stacked bar plot depicting, for each stage, the proportion of cells in each cluster across three control lines. Cell types are grouped by lineage. Populations with significant changes between time points are highlighted with colored arrows and p values and log2(fold changes) are reported. Positive and negative fold changes indicate increase and reduction in the older time point, respectively.

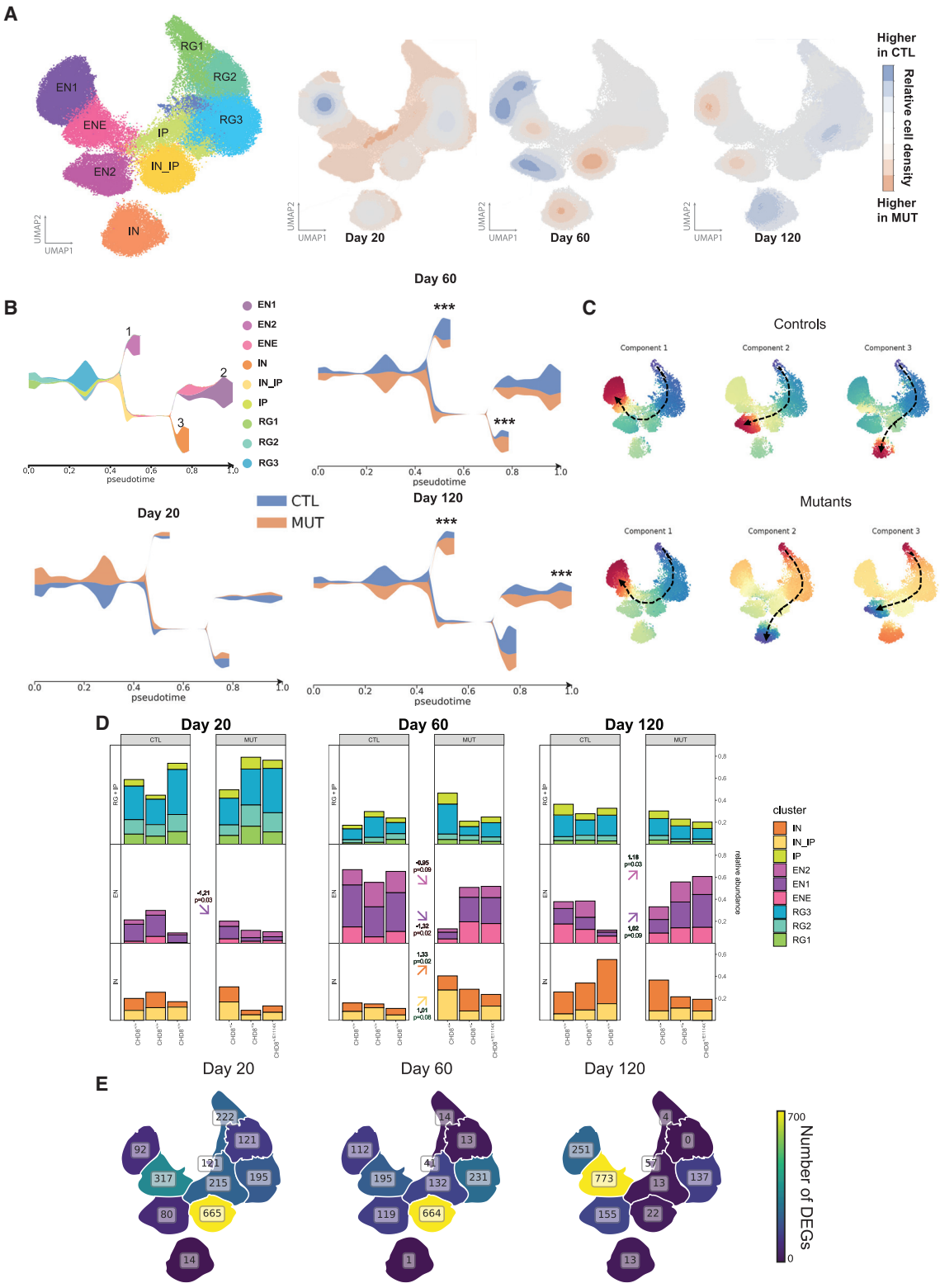


Figure 3. *CHD8* haploinsufficiency causes transient aberrations in cell population proportions and precocious interneuron differentiation

(A) (Left) UMAP projection of droplet-based scRNA-seq from cerebral organoids profiled at day 20, day 60, and day 120, with colored clusters for: EN1, excitatory neurons layers V–VI; EN2, excitatory neurons layers II–IV; ENE, excitatory neurons early; IN, interneurons; IN_IP, interneuron intermediate progenitors; IP,

(legend continued on next page)

segregating into three main branches (i.e., IN, EN1, and EN2) (Figures 2B–2D).

Analysis of population reproducibility across control organoids (obtained from independently grown clones of the original hESC line) corroborated the presence of comparable proportions of all identified subpopulations (Figures S5A–S5B), each of which followed the same reproducible pseudo-temporal trajectory starting from early proliferating progenitors and bifurcating into inhibitory and excitatory lineages (Figures 2C and 2D and S5C), confirming that both population composition and developmental trajectories are robustly recapitulated across control organoid lines.

To corroborate with a complementary approach the reconstruction of the main developmental trajectories in control cerebral organoids (three lines, n = 28,421 cells), we applied tree graph analysis (Chen et al., 2019), which recapitulated the three main developmental trajectories (EN1, EN2, IN). Each maturation trajectory is driven by the expression of specific transcription factors: (1) excitatory neurons Foxg1-Ngn2, (2) excitatory neurons NeuroD2 and Tbr1, and (3) interneurons Dlx1/2 (Figures 2G and S4B). Although, at the neural progenitors stage, excitatory and inhibitory lineages were not resolved as separate populations by this approach, these lineages went on to display along their differentiation a stereotypical pseudotime progression from progenitors to neurons (Figures 2B, 2G, and 2H) with each neuronal population matching the expected increase in abundance throughout stages (Figures 2I and 2J).

To investigate the impact of *CHD8* haploinsufficiency on cortical development, we focused on the three *CHD8* mutant lines whose phenotypes consistently resulted in macrocephaly-matching organoid overgrowth ($CHD8^{+/-}$, $CHD8^{+/-;GFP}$, and $CHD8^{+/E1114X}$). Analysis of relative population densities across conditions revealed stage-specific alterations in the abundance of different populations, including a delayed production of EN1-EN2 together with a robust anticipated generation of IN and IN_IP (Figure 3A). Following the developmental branches through pseudotime, we confirmed a higher representation of mutant cells in the interneuron branch at day 60 and in the excitatory branch at day 120 (Figure 3B). Differential abundance analysis of cluster frequencies confirmed the statistical significance of these differences and indicated that the increase at day 60 in the number of interneurons of the parvalbumin lineage (Figures 3B and S5D–S5E) is the most robust population change caused by *CHD8* haploinsufficiency. We also evaluated the transcriptional impact of *CHD8* LoF in each population by performing a stage- and cluster-wise differential expression analysis. Remarkably, when plotting the number of differentially expressed genes (DEGs) per cluster per developmental stage, we noticed that ENEs and inhibitory neuron progenitors (IN_IPs) showed the highest number of DEGs, consistent with the downstream changes in frequencies of cortifugal excitatory neurons (EN1) and INs (Figure 3E).

To understand whether differences in cell population frequencies result from aberrant developmental trajectories, we calculated, separately for mutants and controls, the first three pseudotime components at each stage and thereby the three most likely trajectories in each condition at each stage. This analysis revealed that while, in control organoids, the first two most likely trajectories yield two distinct excitatory neuronal clusters, in *CHD8* mutant organoids the second most likely trajectory anticipates interneuron production (Figure 3C), underscoring how *CHD8* haploinsufficiency accelerates interneuron development, resulting in a higher proportion of interneurons specifically at day 60.

CHD8 haploinsufficiency causes sustained proliferation in human brain organoids

Next, we employed immunohistochemistry to both validate and dissect the mechanism of the delayed generation of excitatory neurons. Thus, we established an unbiased method to quantify the amount of different cell types in cerebral organoids *in situ* by adapting a deep-learning-based approach for semi-automatic counting of immunolabeled cells. In short, we performed automated cell body segmentation in manually curated regions of interest using Cellpose (Stringer et al., 2020), a state-of-the-art pre-trained deep neural network for cell segmentation that enabled us to analyze large portions of the organoids instead of focusing on small manual counting windows (Figure S6).

To evaluate if the increased size and delayed neurogenesis are due to a proliferative imbalance of neural progenitor cells, we pulsed $CHD8^{+/+}$ and $CHD8^{+/-}$ cerebral organoids at d10 and d20 with 5-Ethynyl-2′-deoxyuridine (EdU) for 1 h and checked

intermediate progenitors; RG1, radial glia cells 1; RG2, radial glia cells 2; RG3, radial glia cells 3. (Right) Density counter plot highlighting stage-wise differences in cell frequencies. Specific sub-areas of the UMAP representation are highlighted and display regions where more mutant (blue) or more control (red) cells are observed, normalized by total number of cells in mutant and control respectively.

(B) Tree graph, divided by stages, indicating how cells are distributed in a pseudo-temporal trajectory (x axis): the origin (time 0, left), comprising cells identified as cycling radial glia branches into three main lineages with the following endpoints: top, excitatory neurons LII–IV; middle, excitatory neurons LV–VI; bottom, interneurons. The thickness of the graph represents percentage of cells in that pseudotime frame. Top left shows distribution of different populations across the graph in controls for reference. Color code on the other three plots displays stage-wise differences in proportion of cells between control and mutant per branch (lineage). Asterisks indicate areas where frequency differences are significant ($p < 0.05$), calculated as reported in (D).

(C) First three diffusion pseudotime components (see STAR Methods) from control and mutant cerebral organoids at day 60, resulting from an unsupervised analysis and highlighting the main components ordered by likelihood (the first component has a stronger effect than the second and the second has a stronger effect than the third).The first three main trajectories within the populations, defined between the two ends (blue and red arrows), are indicated and highlight a preference for interneuron differentiation in mutant organoids.

(D) Stacked bar plots depicting proportion of cells in each cluster across conditions divided by stage. Changes in abundance were tested by a differential abundance analysis for each population at each stage. Populations with significant changes ($p < 0.1$) are highlighted with colored arrows and p values and log2(fold changes) are reported. Positive and negative fold changes indicate, respectively, increase and reduction in the mutant.

(E) UMAP outline illustrating for each cluster and stage the number of mutant versus control DEGS (false discovery rate [FDR] <0.05 and log2 fold change |log2FC| >|1).

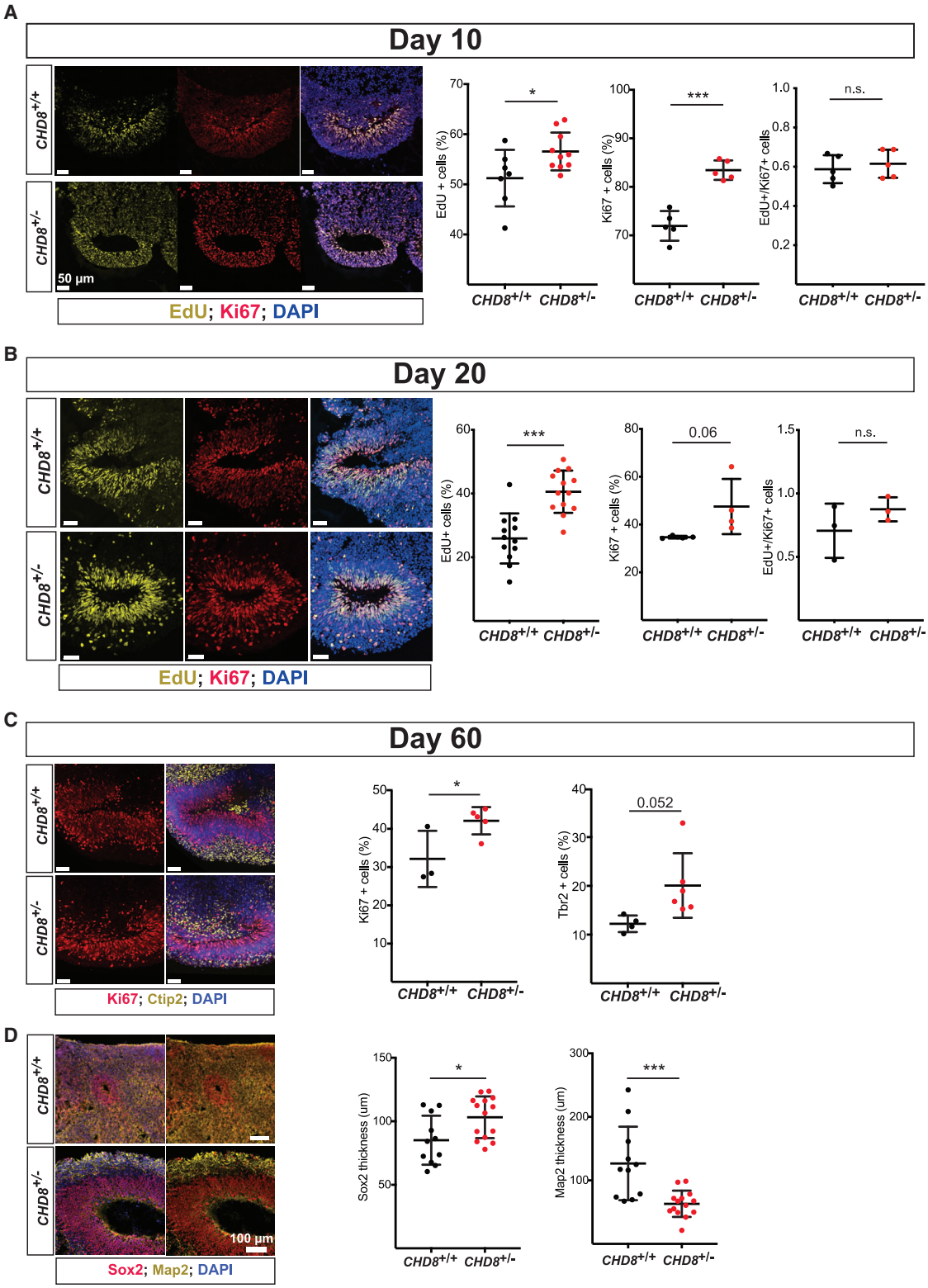


Figure 4. Increased and sustained proliferation in cerebral organoids carrying *CHD8* mutations

(A) Representative images and quantification (right) of ventricular zone (VZ)-like structures of 10-day-old *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ cerebral organoids, showing a higher number of EdU- and Ki67-positive cells, but no difference in their ratio, in *CHD8*$^{+/-}$ organoids. Number of EdU- or Ki67-positive cells represented as

(legend continued on next page)

its incorporation after 16 h. Interestingly, *CHD8*$^{+/-}$ organoids, at both developmental stages, displayed significantly more EdU- and Ki67-positive cells than in the *CHD8*$^{+/+}$ samples (Figures 4A, 4B, S7A, and S7B), indicating that already at day 10, and consistently at day 20, *CHD8* mutations are associated with an increased number of proliferating cells. This difference was not due to changes in cell cycle length as suggested by an equal ratio of Edu/Ki67-positive cells in mutant and control organoids 16 h after the EdU pulse (Figures 4A, 4B, S7A, and S7B). Interestingly, at day 10, *CHD8*$^{+/S62X}$ organoids do not show any difference in EdU-positive cell number compared with control organoids (Figures S7C and S7D), a finding maintained through day 20 (with only a non-significant trend toward increased EdU-labeled cells). This observation is in line with the much milder phenotype observed in S62X mutant organoids and patients, indicating that the early increase in proliferating cells drives the observed increase in organoid size and macrocephaly in patients.

In line with the expansion of the proliferative compartment, we found that, at the same early stages, the number of Tbr2-positive IPs is decreased in the *CHD8*$^{+/-}$ organoids (Figure S7E). This reduced neuronal output was reflected at later time points in decreased thickness of the neuronal layer, as identified by Map2 staining, and increased amounts of Sox2-positive radial glia progenitor cells. Together, our data argue that macrocephaly-associated *CHD8* haploinsufficient mutations lead to increased brain size in humans by extending the phase of neural stem cell proliferation, increasing the progenitor pool, and eventually resulting in the generation of a higher number of neurons at later stages.

CHD8 haploinsufficiency affects neural stem cell proliferation cell autonomously

To understand whether *CHD8* haploinsufficiency effects are cell autonomous while controlling for possible confounders due to the inherent variability of the model, we employed the stable eGFP-expressing *CHD8* LoF mutant hESC line (*CHD8*$^{+/-;GFP}$) (Figure 1D) to generate mosaic organoids allowing us to track mutant cells through maturation. We generated control (*CHD8*$^{+/+}$), GFP-*CHD8* mutant (*CHD8*$^{+/-;GFP}$), and mixed (1:1) control/GFP-mutant (*CHD8*$^{+/+}$/*CHD8*$^{+/-;GFP}$, hereinafter referred to as *CHD8*^{mix}) organoids (Figure 5A). We next counted the proportion of eGFP-positive or -negative progenitor (Sox2 positive) and neuronal (NeuN positive) cells at day 60 in control, mutant, and mixed organoids. In line with the thicker Sox2 layer observed in the *CHD8*$^{+/-}$ organoids (Figure 4D) and indicating a cell-autonomous phenotype in mixed organoids at day 60, we found a higher proportion of Sox2+/GFP+ cells (i.e., *CHD8*$^{+/-}$), compared with Sox2+/GFP− cells (i.e., *CHD8*$^{+/+}$) (Figures 5B and 5C). Conversely, at the same stage, we observed a lower proportion of *CHD8* mutant neuronal cells (i.e., NeuN+/GFP+) compared with wild-type cells (i.e., NeuN+/GFP−) (Figures 5B and 5C).

Together, these data indicate that heterozygous *CHD8* LoF mutations lead to a cell-autonomous sustained proliferation of neural precursors in human cerebral organoids.

CHD8 mutations alter proliferative gene expression program

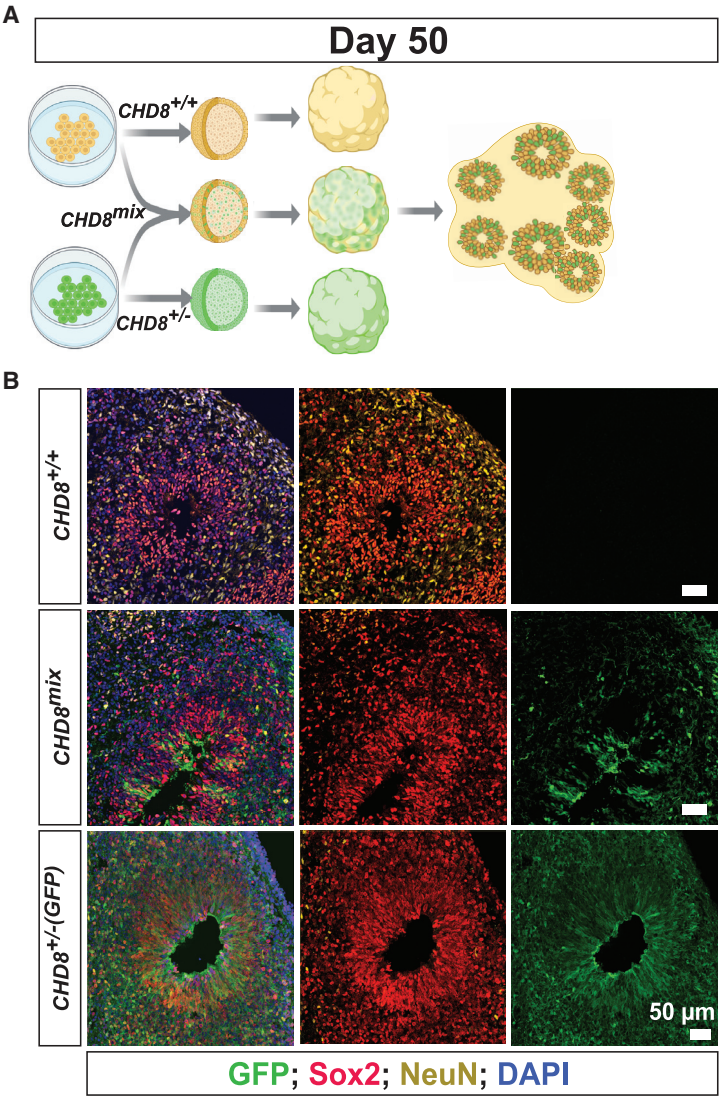
Given the function of *CHD8* (Thompson et al., 2008) and the phenotypes observed in *CHD8* mutant organoids, we carried out a transcriptomic analysis on 10-day-old cerebral organoids, a stage enriched for proliferating cells that marks the first deviation between *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ cerebral organoid development (Figure 4A). This uncovered 868 genes dysregulated as a result of *CHD8* LoF (Table S2). Consistent with the functional convergence of ASD genes along specific developmental stages of human neocortical development (Parikshak et al., 2013; Willsey et al., 2013), as well as with co-expression networks around *CHD8* targets in human mid-fetal brain, we observed a significant enrichment in top ASD-associated genes among the DEGs (Figure 6A; Table S2), including other chromatin regulators such as *ASH1L* and two direct targets of CHD8, *ARID1B*, and *TBL1XR1*. Gene Ontology (GO) analysis showed upregulated genes to be enriched for terms related to cell cycle regulation, RNA splicing, and regulation of transcription (Figure 6B; Table S2). In contrast, downregulated genes were enriched for terms related to neuronal differentiation and brain development (Figure 6B; Table S2), underscoring how, already at early stages of development, *CHD8* haploinsufficiency alters the regulation of neural progenitor proliferation and differentiation. Among the down- and upregulated genes, respectively 28% and 47% are reported as CHD8 bound (Sugathan et al., 2014) (Figure 6C), pointing to a direct effect of *CHD8* haploinsufficiency. Interestingly, GO term enrichment analysis on downregulated genes showed a functional partitioning between direct and indirect targets, with indirect targets enriched in neurogenesis regulation and direct targets enriched for chromatin modifiers and transcriptional regulators (Figure 6D and Table S2). This suggests that, in progenitor cells, CHD8 acts directly on gene regulation, and indirectly on neurogenesis, reflecting either intermediate effectors or the proliferative skew of the neural stem cell

percentage of DAPI-stained cells. Unpaired t test, *p < 0.01; ***p < 0.001; *CHD8*$^{+/+}$, n = 7; *CHD8*$^{+/-}$, n = 10 for EdU; *CHD8*$^{+/+}$, n = 5; *CHD8*$^{+/-}$, n= 5 for Ki67. Organoids were pulsed with EdU for 1 h and fixed 16 h later. Sections were stained for EdU (yellow), Ki67 (red), and DAPI (blue).

(B) Representative images and quantification of EdU (yellow) and Ki67 (red) double labeling of 20-day-old *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ organoids, showing an increased proportion of EdU- and Ki67-positive cells in *CHD8*$^{+/-}$ VZ structures (unpaired t test, ***p < 0.001 for EdU and p = 0.06 for Ki67; *CHD8*$^{+/+}$, n = 12; *CHD8*$^{+/-}$, n = 13 for EdU; *CHD8*$^{+/+}$, n = 4; *CHD8*$^{+/-}$, n= 4 for Ki67).

(C) Representative images of 60-day-old organoids stained for Ki67 (red) and Ctip2 neurons (yellow). Quantification (right) revealed an increased number of Ki67+ and Tbr2+ cells in *CHD8* mutant organoids (unpaired t test, *CHD8*$^{+/+}$, n = 3; *CHD8*$^{+/-}$, n= 5 for Ki67; *CHD8*$^{+/+}$, n = 4; *CHD8*$^{+/-}$, n= 6 for Tbr2).

(D) Representative images for cortical structures from day 60 *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ cerebral organoids stained for Sox2 (red) and Map2 (yellow), showing (right) a significant increase in Sox2+ layer thickness and a significant decrease in Map2+ layer thickness (unpaired t test, *CHD8*$^{+/+}$, n= 11; *CHD8*$^{+/-}$, n= 13).*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; n.s., non-significant. n, number of organoids investigated (multiple rosettes per organoids were analyzed) obtained from at least two independent batches. Results are presented as mean ± SD. Scale bars, 50 μm for (A)–(C), 100 μm for (D).



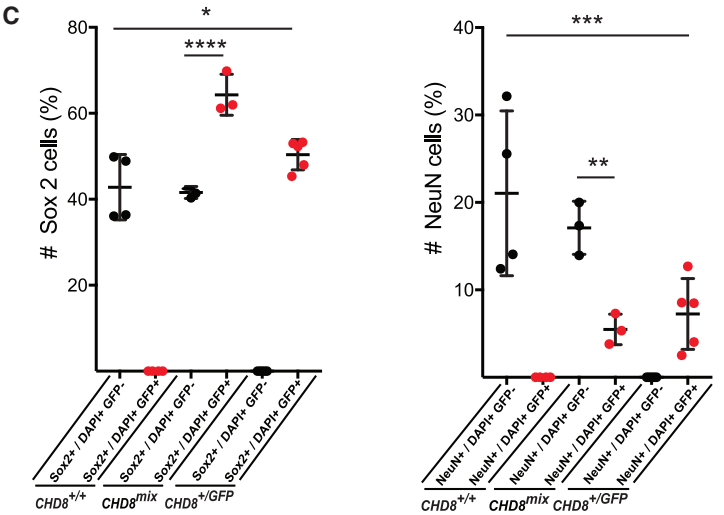


Figure 5. *CHD8* haploinsufficiency leads to cell-autonomous defects

(A) Cartoon explaining the strategy for the generation of $CHD8^{+/+}$, $CHD8^{+/+}$ 1:1 mixed with $CHD8^{+/GFP}$ ($CHD8^{mix}$), and $CHD8^{+/GFP}$ cerebral organoids.

(B) Representative images of $CHD8^{+/+}$ and $CHD8^{mix}$ and $CHD8^{+/-;GFP}$ developing cortical structures, stained for NeuN (yellow) and Sox2 at day 60.

(C) Significant increase of Sox2+ cells in $CHD8^{+/GFP}$ and a larger proportion of Sox2+GFP+ cell population in $CHD8^{mix}$ cortical structures. Ordinary one-way ANOVA followed by Dunnett's multiple comparisons test; $CHD8^{+/+}$, n= 4; $CHD8^{+GFP}$; n = 3; $CHD8^{mix}$, n = 5) and reduction in NeuN+ neurons among $CHD8^{+/GFP}$ cells in the $CHD8^{mix}$ and in $CHD8^{+/GFP}$ cortical structures. Ordinary one-way ANOVA followed by Dunnett's multiple comparisons test; $CHD8^{+/+}$, n= 4; $CHD8^{+/GFP}$, n = 3; $CHD8^{mix}$, n= 5). *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001. n, number of organoids investigated (multiple rosettes per organoids were analyzed). Results are presented as mean ± SD.

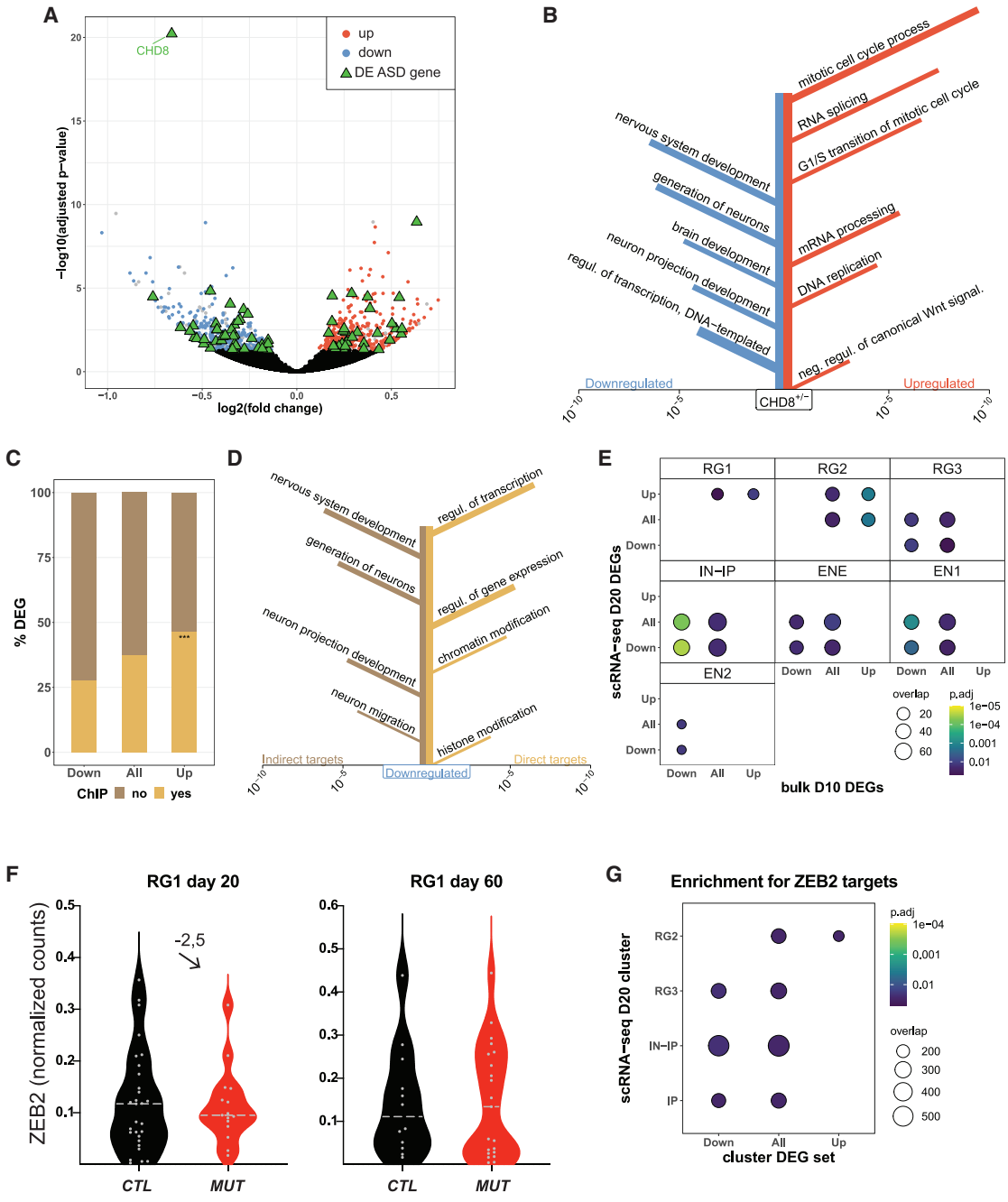


Figure 6. Cell-type-specific gene expression defects in *CHD8* mutant cerebral organoids

(A) Volcano plot displaying results from the differential expression analysis between *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ organoids as negative decadian logarithm of the FDR adjusted p value on the y axis and the binary logarithm of the fold change on the x axis. Genes displaying significant differential expression between *CHD8*$^{+/+}$ and *CHD8*$^{+/-}$ (adjusted p value $\leq$ 0.05, 868 genes) are shown in blue and red for genes down- and upregulated in *CHD8*$^{+/-}$ organoids, respectively. Genes excluded because of significant differences between the two wild-type controls are shown in gray and genes associated with ASD based on the SFARI list are displayed as green triangles (Table S2).

(B) The graph shows selected Gene Ontology (GO) biological processes enriched in genes downregulated (left) or upregulated (right) in *CHD8*$^{+/-}$ organoids. Branch length represents significance of the enrichment, while branch thickness displays the number of dysregulated genes associated with each term.

(C) Overlap between our list of 868 dysregulated genes and CHD8 targets based on published ChIP-seq data (Sugathan et al., 2014), with significant enrichment for CHD8 binding in upregulated genes (***p < 0.001, Fisher's exact test, Table S2).

(D) GO term enrichment for downregulated genes split into direct and indirect targets.

(legend continued on next page)

population. Upregulated genes, on the other hand, showed no marked difference in GO term enrichment between direct and indirect targets (Table S2).

We extended the analysis to the single-cell level by comparing bulk results with stage-specific differential gene expression analysis, comparing macrocephaly-associated mutants versus control lines and thereby cross-validating bulk and single-cell RNA-seq. We found significant overlap between the DEGs identified through the two transcriptomic readouts and, importantly, the dichotomy between proliferation and neurogenesis enrichments observed in bulk data at day 10 was corroborated by the findings that day 10 upregulated and downregulated genes mapped, at day 20, onto DEGs from distinct cell clusters (Figure 6E), with the former enriched in cycling radial glia cells while the latter were in clusters further along differentiation.

We proceeded to annotate the function of genes aberrantly expressed in *CHD8* mutants in specific cell types, performing GO enrichment analysis for each cluster at each developmental stage (Table S1). This revealed an upregulation of genes belonging to GO terms related to cell cycle, mRNA metabolism, ribosome biogenesis, and translation initiation (Table S1) in proliferating cells (RG1–3). Among the top 20 DEGs in radial glia cell clusters at day 20, *ZEB2* caught our attention due its recent implication in cortical expansion (2.5-fold below control levels; Figure 6F). *ZEB2* encodes a transcription factor promoting neuroepithelial differentiation into radial glia, thus determining the initial number of proliferating cells and thereby controlling cortical tissue expansion (Benito-Kwiecinski et al., 2020). Consistently, we found enrichment for ZEB2 chromatin immunoprecipitation sequencing (ChIP-seq)-defined targets among the DEGs of multiple clusters at day 20 (Figure 6G), whose specificity (enrichment in upregulated genes for cycling cells and for downregulated genes in more differentiated cells) further indicates ZEB2 as a potential mediator of the impact of *CHD8* haploinsufficiency on the proliferative dynamics of the neural stem cell compartment. We supplemented this targeted analysis with an unbiased approach searching for other transcription factors that may contribute to the observed phenotypes affecting gene expression in RG cells. We identified three additional genes fulfilling the same criteria of *ZEB2* (i.e., differential expression and enrichment of targets among the DEGs in the RG clusters) (Figure S8A; Table S3). Of note, one of them is *RCOR1*, a co-repressor of REST, a gene previously linked to *CHD8* (Katayama et al., 2016). *RCOR1* has itself been linked to the regulation of proliferation/differentiation balance in the developing brain (Monaghan et al., 2017).

To place our DE results in the context of previous studies, we looked at the overlap of our results with published DEGs (Cotney et al., 2015; Durak et al., 2016; Gompers et al., 2017; Katayama et al., 2016; Mariani et al., 2015; Sugathan et al., 2014; Wang et al., 2017) (Figures S8B and S8C). We found significant overlap with most datasets for our bulk data, with direction-specific overlaps mostly in studies investigating neural stem cells (NSCs) or neural progenitor cells (NPCs), in line with the early time point of our experiment. For our scRNA-seq data, the results reflect the sample origin with most enrichments across clusters found in comparisons with the two organoid studies, especially at day 20. At later time points, the enrichment is more cell type restricted, with most significant overlaps in clusters of excitatory neurons and IP. At day 120, for ENE and EN2, these enrichments also hold true when only comparing upregulated genes, indicating a higher degree of concordance.

CHD8 mutations affect mRNA processing in post-mitotic neurons

Single-cell differential expression analysis highlighted a preferential impact of *CHD8* haploinsufficiency on specific cell populations in terms of number of genes modulated at the transcriptional level (Figure 3E). We thus focused on those clusters with the highest number of DEGs to find indications about the potential downstream impact of *CHD8* haploinsufficiency on human cortex developmental programs, noting that two of the most affected clusters in terms of DEGs (ENE and IN_IP) are the ones immediately preceding the most affected clusters in terms of cell frequencies (EN1 and IN) (Figure 3), thereby linking upstream transcriptional changes to downstream population imbalances. In light of previous findings from murine fetal brain that associated alternative splicing alterations to CHD8-linked developmental delay (Gompers et al., 2017), we investigated whether RNA processing could also be affected in the human setting and related to specific developing subpopulations. We thus probed the delta percentage spliced-in index (dPSI) with a pseudo-bulk approach (Wang et al., 2008) examining stage- and population-specific changes in splicing patterns. As a result, we found significantly different dPSI scores in a subset of genes in IN_IP, IN, and ENE/EN1 populations when comparing $CHD8^{+/+}$ and $CHD8^{+/-}$ and benchmarking against the physiological regulation of alternative splicing across control clusters (Table S4). We complemented this analysis with an approach at the single-cell level, by applying to the same dataset a recently introduced algorithm (SpliZVD; Olivieri et al., 2021). Interestingly, when applying functional enrichment analysis on the identified genes, both approaches (pseudo-bulk and single cell) showed consistent canonical pathways, especially at later time points, with most of the top-10 categories (ranked according to single-cell analysis) shared between day 60 and day 120 (Table S4). Among the enriched pathways, EIF2 signaling was the top scoring, with several of the components showing significantly different dPSI scores in mutant organoids (Figure S9). This mapping revealed the downstream part of the pathway, related to ribosome assembly and translational initiation, as the most affected. Interestingly, both ribosomal proteins and translation

(E) Overlap between DEGs identified in bulk RNA-seq from day 10 and DEGs from scRNA-seq at day 20 (no FC cutoff). Comparisons are for all DEGs as well as down- and upregulated DEGs separately. Only significant overlaps (Fisher's exact test FDR adjusted p value <0.05) are shown.
(F) ZEB2 expression in RG1 at day 20 and day 60 from scRNA-seq data. log2(fold change) is indicated for this comparison.
(G) Overlap between ZEB2 target genes (see STAR Methods) and DEGs in scRNA-seq at day 20 (no FC cutoff). Different clusters are shown on the y axis, and the x axis displays the set of DEGs that was used for the overlap (all, upregulated, or downregulated). Only significant overlaps are included (Fisher's exact test FDR adjusted p value <0.05).

initiation factors have previously been involved in alternative splicing and splicing rate dynamics, especially of their own transcripts, outlining a self-regulatory circuit between transcriptional and translational dynamics (Bohnsack and Bohnsack, 2019; Kachaev et al., 2021) whose key components now emerge to be vulnerable to *CHD8* haploinsufficiency. Moreover, we were able to pinpoint population-specific effects on the *CHD8*-dependent alternative splicing alterations with the excitatory neurons (EN2) showing these alterations throughout the examined time points (Figures S9B–S9D).

DISCUSSION

CHD8 haploinsufficiency is a high-confidence risk factor for ASD. Despite significant progress in genotype-phenotype relationships, the human neurodevelopmental trajectories that underlie them are still poorly understood. In particular, while ASD features are invariably present, albeit with variable severity, macrocephaly is present in only up to one-half to two-thirds of cases (Hanly et al., 2020), affording the opportunity to dissect the impact of different mutations in terms of shared versus distinct neurodevelopmental outcomes. Here we pursued a first genotype-phenotype dissection of *CHD8* haploinsufficiency in terms of human neurodevelopmental trajectories at single-cell resolution.

First, we focused on the initial stages of neuroectoderm specification and modified culture conditions to improve the efficiency and reproducibility of cerebral organoid derivation as a prerequisite for robust phenotypic scoring. We next built on our previous transcriptomic benchmarking of iPSC-based disease modeling (Germain and Testa, 2017) and adopted a longitudinal isogenic design aimed at (1) comparing in the same genetic background LoF of the critical CHD8 helicase domain with patient-specific truncating mutations differentially associated to macrocephaly; (2) tracking *CHD8* dosage-dependent neurodevelopmental trajectories by sampling multiple time points corresponding to key morphological and molecular milestones; (3) resolving the cell-autonomous component of *CHD8* haploinsufficiency.

Our findings enable us to draw the following conclusions. First, *CHD8* haploinsufficiency alters the proliferation/differentiation dynamics, with an expansion of the neural progenitors' compartment. Control-mutant mosaic organoids show this effect to be cell autonomous, while both bulk and single-cell transcriptomics converged on the identification of its molecular underpinnings in terms of a dysregulation, in radial glia cells, of neuronal differentiation and cell cycle regulatory pathways. These were, respectively, down- and upregulated and included a significant fraction of *bona fide* CHD8 direct targets. Importantly, these early CHD8 targets overlap significantly with *bona fide* ASD-causative genes (De Rubeis et al., 2014), enabling the systematic benchmark of this convergent subset of ASD syndromes against the neurodevelopmental trajectories uncovered in this work.

Importantly, while the relevance of the genetic background even for highly penetrant mutations is being increasingly recognized (Ardhanareeswaran et al., 2017), the observation that the patient-specific mutation not associated to macrocephaly did not induce overgrowth underscores the sensitivity of our organoid system in capturing pathophysiological readouts of patient-specific mutations.

Second, the *CHD8* mutations causing cerebral organoid overgrowth (and associated to macrocephaly in patients) disrupt the balance between excitatory and inhibitory neuronal production, with a delayed production of excitatory neurons (both lower and upper layers) and a major increase in interneuron output at day 60 that emerged from single-cell transcriptomics as the most conspicuous subpopulation change in mutant lines. This temporal and lineage specificity of *CHD8* impact acquires even greater salience through the partitioning of gene expression dysregulation revealed by single-cell transcriptomics. Indeed, only selected cell populations appear vulnerable to *CHD8* dosage, and among these only some significantly so throughout development, as in the case of the RG3 cluster of radial glia (i.e., radial glia in the G1 phase of the cell cycle) and the corticofugal neurons. The contrast is indeed particularly striking when comparing the excitatory and inhibitory trajectories, with interneuron progenitors and early intermediates greatly affected only through day 60, while both lower- and upper-layer excitatory neurons become increasingly affected over time. These findings uncover a bimodal developmental impact of *CHD8* haploinsufficiency, with an early switch that accelerates interneuron fate acquisition but with relatively little lasting impact on interneuron transcriptional identity, accompanied by a delay in excitatory fate that leaves a major lasting legacy (i.e., high number of DEGs) on the transcriptional identity, especially of lower-layer neurons.

Finally, by probing cell-type-specific transcriptional dysregulation, we uncover RNA processing as a layer of *CHD8* dosage-dependent vulnerability. Mouse studies had previously reported splicing alterations in *Chd8* heterozygous mutants (Gompers et al., 2017). This is in line with evidence from ASD patient cohorts, with alterations in alternative splicing patterns reported in blood samples of young ASD patients (Stamova et al., 2013) as well as in a large cohort of ASD individuals by transcriptomics analyses on postmortem brain specimens (Parikshak et al., 2013). Here we found that mutants displayed differences in the pattern of alternatively spliced transcripts, with interneuron and excitatory compartments being the most affected. Strikingly, these alternatively processed transcripts specifically recovered from the cell clusters most affected by *CHD8* haploinsufficiency code in turn for proteins involved in mRNA processing and splicing, pointing to an amplifying cascade of *CHD8* dosage-dependent dysregulation in RNA processing (Table S4).

Categories found concordantly enriched encompass relevant biological functions such as EIF2 signaling (Cagnetta et al., 2019; Tomoiaga et al., 2020) and mitochondrial dysfunction (Legido et al., 2013), both already associated with ASD.

Importantly, having included in the analysis the physiological dynamics of developmentally regulated alternative splicing in controls, we can rule out that the aberrations in mutants are a byproduct of developmental skews rather than a primary effect of *CHD8* haploinsufficiency. In conclusion, our work exposes the neurodevelopmental impact of *CHD8* haploinsufficiency, highlighting its exquisite specificity in terms of affected pathways and temporal windows that underlie patient-relevant endophenotypes.

Limitations of the study

While our analysis highlights a lengthening of the proliferative phase of excitatory neuron progenitors as a potential factor leading to the increased brain size seen in patients with mutations in *CHD8*, it cannot exclude that additional phenotypes are also at play. In particular, changes affecting brain areas or cellular features that are not recapitulated in the cerebral organoids may have not been captured in our study. In addition, future studies are warranted to probe the excitatory and inhibitory neuron ratio changes at the functional level and understand how the observed changes affect the activity of the circuitry. Furthermore, while we identified changes in the expression of certain transcription factors, and their targets, as potential critical changes linking *CHD8* haploinsufficiency to macrocephaly, we did not perform rescue experiments to assess the contribution of each of these factors on the phenotype. Finally, our analysis on the effect of *CHD8* mutations in alternative splicing highlights their impact on RNA processing, which will have to be further dissected at the mechanistic level through follow-up *ad hoc* studies. Although our approach included the use of two different algorithms (dPSI and SpliZVD), these results warrant further functional validation to address the consequences of changes in alternative isoform ratios in terms of their impact on critical neurodevelopmental periods.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHODS DETAILS
  - Batch information
  - Oligonucleotide sequences
  - Plasmids
  - Generation of *CHD8*$^{+/-}$ and *CHD8*$^{+/gfp}$ hESC
  - RT-qPCR
  - Western blot
  - Southern blot
  - Generation of cerebral organoids from hESC
  - Histology and immunofluorescence
  - Image processing and quantification
  - Semi-automated cell counting
  - RNA sequencing
  - Transcriptomic analysis
  - Single cell transcriptomics
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.110615.

ACKNOWLEDGMENTS

We thank Farnaz Freeman for technical assistance. This research was supported by the Scientific Service Units (SSU) of IST Austria through resources provided by the Bioimaging Facility (BIF) and the Life Science Facility (LSF). This work supported by the European Union's Horizon 2020 research and innovation program (ERC) grant 715508 to G.N. (REVERSEAUTISM) and grant 825759 to G.T. (ENDpoiNTs); the Fondazione Cariplo 2017-0886 to A.L.T.; E-Rare-3 JTC 2018 IMPACT to M. Gabriele; and the Austrian Science Fund FWF I 4205-B to G.N. Graphical abstract and figures were created using BioRender.com.

AUTHOR CONTRIBUTIONS

B.O., R.S., A.C.Y., and J.M. generated and biochemically characterized cell lines and cerebral organoids. C.E.V., C.C., and M. Gabitto performed the computational analysis of the single-cell datasets. C.E.V., A.L.T., and C.C. annotated single-cell datasets. A.L.T. and M. Gabriele performed single-cell experimental procedures. C.P.D. performed and analyzed bulk transcriptomic experiments. C.S. developed semi-automatic cell counting. M.R.T., J.L., and J.G.D. contributed to imaging analysis. C.E.V., A.L.T., C.C., G.N., and G.T. developed the interpretive framework of the study. G.N. and G.T. led the study.

DECLARATION OF INTERESTS

G.N. is a co-founder of Neurolentech. All the other authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community.

Received: April 1, 2021
Revised: November 18, 2021
Accepted: March 13, 2022
Published: April 5, 2022

REFERENCES

Alexa, A., and Rahnenfuhrer, J. (2021). topGO: Enrichment Analysis for Gene Ontology. https://rdrr.io/bioc/topGO/.

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

Barkas, N., Petukhov, V., Nikolaeva, D., Lozinsky, Y., Demharter, S., Khodosevich, K., and Kharchenko, P.V. (2019). Joint analysis of heterogeneous single-cell RNA-seq dataset collections. Nat. Methods *16*, 695–698.

Becht, E., McInnes, L., Healy, J., Dutertre, C.A., Kwok, I.W.H., Ng, L.G., Ginhoux, F., and Newell, E.W. (2018). Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. *37*, 38–44.

Beighley, J.S., Hudac, C.M., Arnett, A.B., Peterson, J.L., Gerdts, J., Wallace, A.S., Mefford, H.C., Hoekzema, K., Turner, T.N., O'Roak, B.J., et al. (2020). Clinical phenotypes of carriers of mutations in CHD8 or its conserved target genes. Biol. Psychiatry *87*, 123–131.

Benito-Kwiecinski, S., Giandomenico, S.L., Sutcliffe, M., Riis, E.S., Freire-Pritchett, P., Kelava, I., Wunderlich, S., Martin, U., Wray, G., and Lancaster, M.A. (2020). An early cell shape transition drives evolutionary expansion of the human forebrain. Cell *184*, 2084–2102.

Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K., Gerdts, J., Baker, C., Vulto-van Silfhout, A.T., et al. (2014). Disruptive CHD8 mutations define a subtype of autism early in development. Cell *158*, 263–276.

Birey, F., Li, M.Y., Gordon, A., Thete, M.V., Valencia, A.M., Revah, O., Pasca, A.M., Geschwind, D.H., and Pasca, S.P. (2021). Dissecting the molecular basis of human interneuron migration in forebrain assembloids from Timothy syndrome. Cell Stem Cell *29*, 248–264.e7.

Bohnsack, K.E., and Bohnsack, M.T. (2019). Uncovering the assembly pathway of human ribosomes and its emerging links to disease. EMBO J. *38*, e100278.

Cagnetta, R., Wong, H.H., Frese, C.K., Mallucci, G.R., Krijgsveld, J., and Holt, C.E. (2019). Noncanonical modulation of the eIF2 pathway controls an increase in local translation during neural wiring. Mol. Cell *73*, 474–489.e475.

Camp, J.G., Badsha, F., Florio, M., Kanton, S., Gerber, T., Wilsch-Brauninger, 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. U S A *112*, 15672–15677.

Catlett, T.S., Onesto, M.M., McCann, A.J., Rempel, S.K., Glass, J., Franz, D.N., and Gomez, T.M. (2021). RHOA signaling defects result in impaired axon guidance in iPSC-derived neurons from patients with tuberous sclerosis complex. Nat. Commun. *12*, 2589.

Chen, E.S., Gigek, C.O., Rosenfeld, J.A., Diallo, A.B., Maussion, G., Chen, G.G., Vaillancourt, K., Lopez, J.P., Crapper, L., Poujol, R., et al. (2014). Molecular convergence of neurodevelopmental disorders. Am. J. Hum. Genet. *95*, 490–508.

Chen, H., Albergante, L., Hsu, J.Y., Lareau, C.A., Lo Bosco, G., Guan, J., Zhou, S., Gorban, A.N., Bauer, D.E., Aryee, M.J., et al. (2019). Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. Nat. Commun. *10*, 1903.

Concordet, J.P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. Nucleic Acids Res. *46*, W242–W245.

Cotney, J., Muhle, R.A., Sanders, S.J., Liu, L., Willsey, A.J., Niu, W., Liu, W., Klei, L., Lei, J., Yin, J., et al. (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat. Commun. *6*, 6404.

de Jong, J.O., Llapashtica, C., Genestine, M., Strauss, K., Provenzano, F., Sun, Y., Zhu, H., Cortese, G.P., Brundu, F., Brigatti, K.W., et al. (2021). Cortical overgrowth in a preclinical forebrain organoid model of CNTNAP2-associated autism spectrum disorder. Nat. Commun. *12*, 4087.

De Rubeis, S., He, X., Goldberg, A.P., Poultney, C.S., Samocha, K., Cicek, A.E., Kou, Y., Liu, L., Fromer, M., Walker, S., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. Nature *515*, 209–215.

Dehghannasiri, R., Olivieri, J.E., Damljanovic, A., and Salzman, J. (2021). Specific splice junction detection in single cells with Sicilian. Genome Biol. *22*, 219.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Durak, O., Gao, F., Kaeser-Woo, Y.J., Rueda, R., Martorell, A.J., Nott, A., Liu, C.Y., Watson, L.A., and Tsai, L.H. (2016). Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. Nat. Neurosci. *19*, 1477–1488.

Falcon, S., and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258.

Gabriele, M., Lopez Tobon, A., D'Agostino, G., and Testa, G. (2018). The chromatin basis of neurodevelopmental disorders: rethinking dysfunction along the molecular and temporal axes. Prog. Neuropsychopharmacol. Biol. Psychiatry *84*, 306–327.

Germain, P.L., and Testa, G. (2017). Taming human genetic variability: transcriptomic meta-analysis guides the experimental design and interpretation of iPSC-based disease modeling. Stem Cell Rep. *8*, 1784–1796.

Gompers, A.L., Su-Feher, L., Ellegood, J., Copping, N.A., Riyadh, M.A., Stradleigh, T.W., Pride, M.C., Schaffler, M.D., Wade, A.A., Catta-Preta, R., et al. (2017). Germline Chd8 haploinsufficiency alters brain development in mouse. Nat. Neurosci. *20*, 1062–1073.

Hanly, C., Shah, H., Au, P.Y.B., and Murias, K. (2020). Description of neurodevelopmental phenotypes associated with 10 genetic neurodevelopmental disorders: a scoping review. Clin. Genet. *99*, 335–346.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

Kachaev, Z.M., Ivashchenko, S.D., Kozlov, E.N., Lebedeva, L.A., and Shidlovskii, Y.V. (2021). Localization and functional roles of components of the translation apparatus in the eukaryotic cell nucleus. Cells *10*, 3239.

Katayama, Y., Nishiyama, M., Shoji, H., Ohkawa, Y., Kawamura, A., Sato, T., Suyama, M., Takumi, T., Miyakawa, T., and Nakayama, K.I. (2016). CHD8 haploinsufficiency results in autistic-like phenotypes in mice. Nature *537*, 675–679.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44*, W90–W97.

Lancaster, M.A., and Knoblich, J.A. (2014). Generation of cerebral organoids from human pluripotent stem cells. Nat. Protoc. *9*, 2329–2340.

Legido, A., Jethva, R., and Goldenthal, M.J. (2013). Mitochondrial dysfunction in autism. Semin. Pediatr. Neurol. *20*, 163–175.

Li, K., Ouyang, Z., Lin, D., Mingueneau, M., Chen, W., Sexton, D., and Zhang, B. (2020). Cellxgene VIP unleashes full power of interactive visualization, plotting and analysis of scRNA-seq data in the scale of millions of cells. Preprint at bioRxiv, 2020.2008.2028.270652.

Lopez-Tobon, A., Villa, C.E., Cheroni, C., Trattaro, S., Caporale, N., Conforti, P., Iennaco, R., Lachgar, M., Rigoli, M.T., Marco de la Cruz, B., et al. (2019). Human cortical organoids expose a differential function of GSK3 on cortical neurogenesis. Stem Cell Rep. *13*, 847–861.

Lotfollahi, M., Wolf, F.A., and Theis, F.J. (2019). scGen predicts single-cell perturbation responses. Nat. Methods *16*, 715–721.

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

Luo, C., Lancaster, M.A., Castanon, R., Nery, J.R., Knoblich, J.A., and Ecker, J.R. (2016). Cerebral organoids recapitulate epigenomic signatures of the human fetal brain. Cell Rep. *17*, 3369–3384.

Mariani, J., Coppola, G., Zhang, P., Abyzov, A., Provini, L., Tomasini, L., Amenduni, M., Szekely, A., Palejev, D., Wilson, M., et al. (2015). FOXG1-Dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. Cell *162*, 375–390.

Monaghan, C.E., Nechiporuk, T., Jeng, S., McWeeney, S.K., Wang, J., Rosenfeld, M.G., and Mandel, G. (2017). REST corepressors RCOR1 and RCOR2 and the repressor INSM1 regulate the proliferation-differentiation balance in the developing brain. Proc. Natl. Acad. Sci. U S A *114*, E406–E415.

Moore, J.C., Atze, K., Yeung, P.L., Toro-Ramos, A.J., Camarillo, C., Thompson, K., Ricupero, C.L., Brenneman, M.A., Cohen, R.I., and Hart, R.P. (2010). Efficient, high-throughput transfection of human embryonic stem cells. Stem Cell Res. Ther. *1*, 23.

Nowakowski, T.J., Bhaduri, A., Pollen, A.A., Alvarado, B., Mostajo-Radji, M.A., Di Lullo, E., Haeussler, M., Sandoval-Espinosa, C., Liu, S.J., Velmeshev, D., et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science *358*, 1318–1323.

O'Roak, B.J., Vives, L., Fu, W., Egertson, J.D., Stanaway, I.B., Phelps, I.G., Carvill, G., Kumar, A., Lee, C., Ankenman, K., et al. (2012a). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science *338*, 1619–1622.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., et al. (2012b). Sporadic autism exomes reveal a

highly interconnected protein network of de novo mutations. Nature *485*, 246–250.

Olivieri, J.E., Dehghannasiri, R., and Salzman, J. (2021). The spliZ generalizes "percent spliced In" to reveal regulated splicing at single-cell resolution. Nat. Methods *19*, 307–310.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., and Geschwind, D.H. (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008–1021.

Pollen, A.A., Bhaduri, A., Andrews, M.G., Nowakowski, T.J., Meyerson, O.S., Mostajo-Radji, M.A., Di Lullo, E., Alvarado, B., Bedolli, M., Dougherty, M.L., et al. (2019). Establishing cerebral organoids as models of human-specific brain evolution. Cell *176*, 743–756.e717.

Qian, X., Nguyen, H.N., Song, M.M., Hadiono, C., Ogden, S.C., Hammack, C., Yao, B., Hamersky, G.R., Jacob, F., Zhong, C., et al. (2016). Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. Cell *165*, 1238–1254.

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.

Rajewsky, N., Almouzni, G., Gorski, S.A., Aerts, S., Amit, I., Bertero, M.G., Bock, C., Bredenoord, A.L., Cavalli, G., Chiocca, S., et al. (2020). LifeTime and improving European healthcare through cell-based interceptive medicine. Nature *587*, 377–386.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y., et al. (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Sakamoto, I., Kishida, S., Fukui, A., Kishida, M., Yamamoto, H., Hino, S., Michiue, T., Takada, S., Asashima, M., and Kikuchi, A. (2000). A novel beta-catenin-binding protein inhibits beta-catenin-dependent Tcf activation and axis formation. J. Biol. Chem. *275*, 32871–32878.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Setty, M., Kiseliovas, V., Levine, J., Gayoso, A., Mazutis, L., and Pe'er, D. (2019). Characterization of cell fate probabilities in single-cell data with palantir. Nat. Biotechnol. *37*, 451–460.

Srivastava, A., Malik, L., Smith, T., Sudbery, I., and Patro, R. (2019). Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. Genome Biol. *20*, 65.

Stamova, B.S., Tian, Y., Nordahl, C.W., Shen, M.D., Rogers, S., Amaral, D.G., and Sharp, F.R. (2013). Evidence for differential alternative splicing in blood of young boys with autism spectrum disorders. Mol. Autism *4*, 30.

Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2020). Cellpose: a generalist algorithm for cellular segmentation. Preprint at bioRxiv, 2020.2002.2002.931238.

Suetterlin, P., Hurley, S., Mohan, C., Riegman, K.L.H., Pagani, M., Caruso, A., Ellegood, J., Galbusera, A., Crespo-Enriquez, I., Michetti, C., et al. (2018). Altered neocortical gene expression, brain overgrowth and functional over-connectivity in Chd8 haploinsufficient mice. Cereb. Cortex *28*, 2192–2206.

Sugathan, A., Biagioli, M., Golzio, C., Erdin, S., Blumenthal, I., Manavalan, P., Ragavendran, A., Brand, H., Lucente, D., Miles, J., et al. (2014). CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. Proc. Natl. Acad. Sci. U S A *111*, E4468–E4477.

Thompson, B.A., Tremblay, V., Lin, G., and Bochar, D.A. (2008). CHD8 is an ATP-dependent chromatin remodeling factor that regulates beta-catenin target genes. Mol. Cell Biol. *28*, 3894–3904.

Tomoiaga, D., Aguiar-Pulido, V., Shrestha, S., Feinstein, P., Levy, S.E., Mason, C.E., and Rosenfeld, J.A. (2020). Single-cell sperm transcriptomes and variants from fathers of children with and without autism spectrum disorder. NPJ Genom Med. *5*, 14.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci. Rep. *9*, 5233.

Trincado, J.L., Entizne, J.C., Hysenaj, G., Singh, B., Skalic, M., Elliott, D.J., and Eyras, E. (2018). SUPPA2: fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. Genome Biol. *19*, 40.

Urresti, J., Zhang, P., Moran-Losada, P., Yu, N.K., Negraes, P.D., Trujillo, C.A., Antaki, D., Amar, M., Chau, K., Pramod, A.B., et al. (2021). Cortical organoids model early brain development disrupted by 16p11.2 copy number variants in autism. Mol. Psychiatry *26*, 7560–7580.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470–476.

Wang, P., Lin, M., Pedrosa, E., Hrabovsky, A., Zhang, Z., Guo, W., Lachman, H.M., and Zheng, D. (2015). CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in neurodevelopment. Mol. Autism *6*, 55.

Wang, P., Mokhtari, R., Pedrosa, E., Kirschenbaum, M., Bayrak, C., Zheng, D., and Lachman, H.M. (2017). CRISPR/Cas9-mediated heterozygous knockout of the autism gene CHD8 and characterization of its transcriptional networks in cerebral organoids derived from iPS cells. Mol. Autism *8*, 11.

Wegscheid, M.L., Anastasaki, C., Hartigan, K.A., Cobb, O.M., Papke, J.B., Traber, J.N., Morris, S.M., and Gutmann, D.H. (2021). Patient-derived iPSC-cerebral organoid modeling of the 17q11.2 microdeletion syndrome establishes CRLF3 as a critical regulator of neurogenesis. Cell Rep. *36*, 109315.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., et al. (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997–1007.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

Zhang, Z., Luo, D., Zhong, X., Choi, J.H., Ma, Y., Wang, S., Mahrt, E., Guo, W., Stawiski, E.W., Modrusan, Z., et al. (2019). SCINA: a semi-supervised subtyping algorithm of single cells and bulk samples. Genes (Basel) *10*, 531.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-CHD8 C-terminus	Novus Biologicals	Cat#NB100-60418; RRID:AB_905325
Rabbit monoclonal anti-CHD8 N-terminus	Cell Signaling	Cat#77694; RRID:AB_2799904
Rabbit polyclonal anti- P70 S6 Kinase	Cell Signaling	Cat#2708; RRID:AB_390722
Rabbit polyclonal anti-GAPDH	Millipore	Cat#ABS16; RRID:AB_10806772
Rat monoclonal anti-BrdU	Bio-Rad	Cat#MCA2060T; RRID:AB_10015293
Mouse monoclonal anti-N-cadherin	ThermoFisher	Cat#33-3900; RRID:AB_2313779
Rabbit polyclonal anti-SOX2	Millipore	Cat# AB5603; RRID:AB_2286686
Rat monoclonal anti-Laminin	Abcam	Cat# ab44941; RRID:AB_775971
Chicken polyclonal anti-TBR2	Millipore	Cat# AB15894; RRID:AB_10615604
Rabbit polyclonal anti-Cleaved Caspase-3	Cell Signaling	Cat# 9661; RRID:AB_2341188
Mouse monoclonal anti-MAP2	Sigma	Cat# M4403; RRID:AB_477193
Rabbit polyclonal anti-phospho-Histone H3	Millipore	Cat# 06-570; RRID:AB_310177
Mouse monoclonal anti-NeuN	Abcam	Cat# ab104224; RRID:AB_10711040
Mouse monoclonal anti-Parvalbumin	Millipore	Cat# MAB1572; RRID:AB_2174013
Rabbit polyclonal anti-FOXG1	Abcam	Cat# ab18259; RRID:AB_732415
Rabbit polyclonal anti-EMX1	Sigma	Cat# HPA006421; RRID:AB_1078739
Rat monoclonal anti-CTIP2	Abcam	Cat# ab18465; RRID:AB_2064130
Critical commercial assays		
QuantSeq 3′ mRNA-Seq FWD Library Prep Kit	Lexogen	Cat#015 N/A
Deposited data		
Raw bulk RNA sequencing data	This paper	ArrayExpress:E-MTAB-9933
Raw single cell sequencing data	This paper	ArrayExpress: E-MTAB-11406
ZEB2 ChIP-seq data	ENCODE Project Consortium, 2012	GEO: GSE91749
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/
GENCODE Genome annotation Release 24	The GENCODE Project	https://www.gencodegenes.org/human/release_24.html
Experimental models: Cell lines		
H9 ES cells	WiCell	Cat#WA09; RRID:CVCL_9773
Oligonucleotides		
See Table S5B	This paper	N/A
Recombinant DNA		
pX330-U6-Chimeric_BB-CBh-hSpCas9	Ran et al. (2013b)	Addgene Cat#42230; RRID:Addgene_42230
pCR2.1 vector	ThermoFisher	Cat#K202020 N/A
CHD8_E19KO_PGK-Neo_pCAG-GFP	This paper	N/A
Software and algorithms		
CRISPR Design tool	Hsu et al. (2013)	http://crispr.mit.edu/
CRISPOR	Concordet and Haeussler et al. (2018)	http://crispor.tefor.net/
Fiji/ImageJ	Schindelin et al. (2012)	https://imagej.nih.gov/ij/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cellpose version 0.1.0.1	Stringer et al. (2020)	https://www.cellpose.org/
BBDuk – BBMap package	Joint Genome Institute	http://jgi.doe.gov/data-and-tools/bbtools/
STAR version 2.5.4	Dobin et al. (2013)	https://github.com/alexdobin/STAR
DESeq2 version 1.12.4	Love et al. (2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
GOstats version 2.36.0	Falcon and Gentleman (2007)	https://bioconductor.org/packages/release/bioc/html/GOstats.html
GOTreeVis	GitHub	https://github.com/dottercp/GOTreeVis
HOMER version 4.11	Heinz et al. (2010)	http://homer.ucsd.edu/homer/ngs/quantification.html
enrichR version 3.0	Kuleshov et al. (2016)	https://github.com/wjawaid/enrichR
Enrichr	Kuleshov et al. (2016)	https://maayanlab.cloud/Enrichr/
ggh4x version 0.2.1	GitHub	https://github.com/teunbrand/ggh4x
scGen	Lotfollahi et al. (2019)	https://github.com/theislab/scgen
Scanpy version 1.6.0	Wolf et al. (2018)	https://pypi.org/project/scanpy/
SCINA semi-supervised annotations algorithm	Zhang et al. (2019)	https://github.com/jcao89757/SCINA
Palantir algorithm	Setty et al. (2019)	https://github.com/dpeerlab/Palantir
STREAM algorithm	Chen et al. (2019)	https://stream.pinellolab.partners.org/
TopGo 2.42.0	Rahnenfuhrer 2020	https://bioconductor.org/packages/release/bioc/html/topGO.html
Alevin	Srivastava et al. (2019)	https://salmon.readthedocs.io/en/latest/alevin.html
SUPPA2	Trincado et al. (2018)	https://github.com/comprna/SUPPA
SICILIAN	Dehghannasiri et al. (2021)	https://cgc.sbgenomics.com/public/apps/jordanski.milos/deepest-fusion/sicilian/
SpliZ pipeline	Olivieri et al. (2021)	https://github.com/juliaolivieri/SpliZ_pipeline
GraphPad Prism 6.0	GraphPad Software	www.graphpad.com
Python module and cellpose filter	This paper	https://doi.org/10.5281/zenodo.6327683

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Gaia Novarino (gnovarino@ist.ac.at).

Materials availability

Human H9 embryonic stem (ES) cells (WA09) are available from commercial sources (WiCell). All *CHD8* mutant cell lines used in this study are available from the lead contact subject to institutional MTA regulations. Newly generated plasmid constructs (CHD8_E19KO_PGK-Neo_pCAG-GFP) are available upon request.

Data and code availability

- Single-cell and bulk RNA-seq data have been deposited at ArrayExpress and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Original western blot images and microscopy data reported in this paper will be shared by the lead contact upon request.
- All code for the cell counting pipeline have been deposited at Github and is available as of the date of publication, DOI is listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

One human female ES cell line was used in this study. Feeder-independent human H9 ES cells (WA09) were purchased from WiCell at passage 24, and maintained on growth factor reduced Matrigel basement membrane (Corning, #356230) in mTeSR1 (later changed to StemFlex (Gibco)) medium (Stemcell Technologies) according to WiCell indications. Cells were passaged using 0.5 mM EDTA (pH 8.0). All *CHD8*$^{+/-}$ and *CHD8*$^{+/gfp}$ lines were generated using this cell line as described in the methods details. Human ESCs used in this project were approved for use in this project by an ERC ethics committee. Commercial H9 ES cell line was authenticated by the provider.

METHODS DETAILS

Batch information

Detailed information on the number of samples and batches used for each experiment is summarized on Table S5.

Oligonucleotide sequences

Sequences referred to as sN can be found in Table S5.

Plasmids

gRNA/Cas9-coding plasmids for nucleofection experiments were obtained by cloning gRNA sequences targeting PAM regions in exon 17 (s1/2) and exon 19 (s3/4) of *CHD8* into the pX330-U6-Chimeric_BB-CBh-hSpCas9 (Addgene #42230) according to the protocol from Zhang group (Ran et al., 2013b). Guide RNA sequences were designed through the CRISPR Design tool (http://crispr.mit.edu/) and had reported score of 94 (target in exon 17) and 96 (target in exon 19). For E1114X and S62X gRNA sequences were designed using CRISPOR (http://crispor.tefor.net/) and cloned as described above (E1114X: s5/6, S62X: s7/8).

The donor plasmid for homology-driven recombination (HDR)-mediated repair (CHD8_E19KO_PGK-Neo_pCAG-GFP) was obtained by multiple sequence subcloning into the pCR2.1 vector (ThermoFisher) as follows: the humanized Renilla green fluorescent protein (hrGFP) coding sequence and the pCAG promoter sequence were inserted by restriction enzyme cloning. In the resulting vector, the left homology arm (amplification primers: s9/10) and the right arm (amplification primers: s11/12) were cloned by In-Fusion cloning (Clontech); finally, the PGK-Neo sequence was cloned by In-Fusion cloning. For E1114X and S62X the same vector was used but the PGK-Neo cassette was removed as it was no longer needed (HA amplification primers: s13–16 and s17-20, respectively). Here the homology arms included the premature stop codon mutations found in patients (E1114X and S62X).

Generation of *CHD8*$^{+/-}$ and *CHD8*$^{+/gfp}$ hESC

At a passage number <35, cell colonies were dissociated to single cells by StemPro Accutase treatment (Gibco) for 4 min at 37°C and gentle resuspension in mTeSR1 supplemented with 10 μM Y-27632 (Stemcell Technologies) (later changed to RevitaCell (Gibco). Cells were counted with Trypan Blue with an automated cell counter (BioRad Tc-20), then 1.2 × 10^6 cells were resuspended into Nucleofection buffer (Human Stem Cell Nucleofection Kit 1, Lonza) in the presence of gRNA/Cas9-coding plasmid DNA, and nucleofected in Nucleofector 2B (Lonza, program A-023).

For the generation of deletion *CHD8*$^{+/-}$ mutants, 5 μg each of two gRNA/Cas9-coding plasmids, targeting PAM sequences in exon 17 and 19 of *CHD8* respectively, were used. After nucleofection cells were resuspended in mTeSR1 supplemented with ROCK inhibitor and replated on a well of a 24-well culture plate for 48 h in order to promote recovery, after which time the medium was replaced with mTeSR1 or StemFlex in absence of ROCK inhibitor and changed daily until a confluence of 80% was reached. Cells were then again dissociated with Accutase treatment as indicated above and a fifth of the total amount was plated on a 10cm culture dish in mTeSR1 supplemented with ROCK inhibitor or StemFlex supplemented with RevitaCell to obtain a cell density which would favor the formation of single cell-derived colonies. After 3-4 days, the medium was replaced with mTeSR1 or StemFlex and changed daily until colonies reached a diameter compatible with colony picking. For the screening of deletion-mutant colonies, around 200 colonies were picked and expanded in 96-well culture plates, then passaged to obtain replica plates to be used for genomic DNA extraction and screening of heterozygous deletion mutants. Analysis of off-targets predicted by the CRISPR Design tool was performed on positive clones. Confirmed positive clones were then expanded and subjected to at least three subcloning steps to achieve monoclonality.

For the generation of *CHD8*$^{+/-;GFP}$ insertion mutants, 3 × 10^6 cells were nucleofected as described above, in the presence of 2.5 μg of exon 17-targeting gRNA/Cas9-coding plasmid and 7.5 μg of the linearized CHD8_E19KO_PGK-Neo_pCAG-GFP donor plasmid. After nucleofection cells were plated on a 10 cm culture dish. After 96 h, the medium was replaced and 50 μg/mL G418 added to select neomycin-resistant cells. Since H9 single cell survival is affected by cell density, selective medium was changed daily until elimination of an amount of non-resistant cells compatible with viability of resistant colonies, as described in Moore et al. (Moore et al., 2010). GFP-positive resistant clones were monitored and picked under an EVOS FL imaging system (ThermoFisher). Picked colonies were expanded, screened for heterozygous insertion mutations, subjected to off-target analysis and subcloned to reach monoclonality as described above. The possible off-targets for deletion lines were determined using the CRISPR Design tool (http://crispr.mit.edu/). There were three possible off-target sites in coding regions and all are checked by PCR (primers: s21 – s26).

For E1114X and S62X the premature stop mutation was introduced via mutations in the homology arms and the GFP cassette for selection was inserted downstream of this mutation. Nucleofection was performed as described above for *CHD8*$^{+/-;GFP}$. We checked the top E1114X off-targets: two exons (primers: s27-30) and four introns (primers: s31-38), as well as S62X off-targets: four exons (primers: s39-46) and in three introns (primers: s47-52). CRISPOR offtarget predictions for all guides can be found in Table S5.

RT-qPCR

Total mRNA was isolated from hESCs, EBs and organoids at different stages in triplicate using Trizol (Ambion). Potential contaminating DNA was removed using DNase1 treatment and 1 μg RNA was used for cDNA synthesis using SuperScript III (Life Technologies). qPCR was performed using the LightCycler® 480 SYBR Green I Master on a Roche LightCycler® 480, according to the manufacturer's instructions with primers for *hCHD8* (s53/54) and *hTBP* (s55/56) for normalization. Plots of the results were generated in R.

Western blot

Whole-cell protein extracts from hESCs were prepared by lysing cells in RIPA buffer (50 mM Tris-HCl, pH 7.6, 1% Triton X-100, 0.5% Na-deoxycholate, 0.1% SDS, 150 mM NaCl, 2 mM EDTA, 1x Protease inhibitor cocktail (Roche). Extracts were denatured by adding Laemmli buffer and boiling for 5 min at 95°C, then 20 μg were resolved on a 6% SDS-PAGE (10% for GAPDH) and transferred to 0.45 μM Immobilon®-P PVDF membranes (Millipore). Protein extracts from day 50 cerebral organoids were obtained by disrupting organoids through pipetting in lysis buffer (50 mM Tris-HCl pH 8, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, Protease inhibitor cocktail (Roche) and sonicating (5 cycles, 30 sec each cycle). 50 μg aliquots were then resolved as described above. Membranes were blocked in 0.05% Tween-PBS, 5% milk for 6 hours at 4°C and incubated with a CHD8 primary antibody (rabbit, C-terminus Novus Biologicals NB100-60418; N-terminus Cell Signaling #77694; 1:2000) in blocking solution overnight at 4°C. The next day membranes were washed and incubated for one hour at room temperature with anti-Rabbit IgG (H+L)-conjugated secondary antibody (donkey, Pierce SA1-200, 1: 1000), then washed and signal detected by enhanced chemiluminescence (Pierce, #32209 and #34095). CHD8 levels were normalized to P70 S6 Kinase (rabbit, Cell Signaling 2708, 1:1000) or GAPDH (rabbit, Millipore, 1:1000).

Southern blot

Genomic DNA was digested with the chosen restriction enzymes (New England Biolabs) (*CHD8*$^{+/-}$: EcoRV and NsiI; *CHD8*$^{+/S62X}$: EcoRv and EcoRI; *CHD8*$^{+/E1114X}$: EcoRV and NsiI), separated in a 0.8% agarose gel, transferred to a positively charged membrane (Roche, 11209299001) and hybridized with DIG labeled probes (Roche, 11218590910). Membranes were imaged with Molecular Imaging GE Healthcare Amersham 600.

Generation of cerebral organoids from hESC

At a confluency of 70-80%, *CHD8*$^{+/+}$, *CHD8*$^{+/-}$ or *CHD8*$^{+/-,GFP}$ cells were dissociated to single cells using Accutase as described above. A total of 9000 cells, later optimized to 2500 cells, was transferred to each well of an ultra-low-binding-96-well plate (Corning) in mTeSR1 media supplemented with 50 μM Y-27632. Cells were allowed to aggregate to EBs and fed every second day. At day 3 supplements were removed and from day 6 the generation of cerebral organoids was performed according to Lancaster and Knoblich (Lancaster and Knoblich, 2014). Briefly, EBs were transferred to neural induction medium (NIM) in low-adhesion 24-well plates (Corning), and fed every second day for 5 days until formation of neuroepithelial tissue (day 0 of cerebral organoid formation). Neuroepithelial tissue-displaying organoids were embedded in Matrigel droplets (Corning, #356234) and grown in cerebral organoid medium (COM) supplemented with B27 without vitamin A (Gibco) and fed every other day. After 4 days tissues were transferred to COM supplemented with B27 containing vitamin A and placed on a horizontal shaker at 70-100rpm. Cerebral organoids were fed twice a week.

For BrdU and EdU incorporation experiments, organoids were incubated with COM supplemented with 10μM BrdU (Sigma) or 10μM EdU for 1h at 37°C, then washed with 1x PBS and either processed directly for immunostaining or incubated for 16h with COM, washed with 1x PBS and processed for immunostaining.

Histology and immunofluorescence

Cerebral organoids were fixed in 4% paraformaldehyde for 20 min at 4°C and washed in 1x phosphate buffered saline (PBS) two times for 10 min. Tissues were dehydrated in 30% sucrose until they sunk and then embedded in O.C.T. (Tissue Tek). Cerebral organoids harvested at day 10, 20, and 50 were cryosectioned at 14-18 μm, and cerebral organoids harvested at day 60-62 were cryosectioned at 18-20 μm. For immunofluorescence, sections were washed 3 × 5min in 1x PBS and blocked and permeabilized in 0.5% Triton X-100 and 4% normal goat serum in PBS. Sections were then incubated with primary antibodies in 0.3% Triton X-100, 4% normal goat serum at 4°C overnight. Sections stained for BrdU were first incubated with 2N HCl solution (pH 0.6–0.9) at RT for 30 min followed by washing 3 × 5 min in PBS before blocking. Primary antibody dilutions used: BrdU (rat, AbD Serotec MCA2060T, 1:500), N-cadherin (mouse, ThermoFisher 33-3900, 1:500), Sox2 (rabbit, Millipore AB5603, 1:300), Laminin (rat, Abcam ab44941, 1:500), Tbr2 (chicken, Millipore AB15894, 1:300), Cleaved Caspase-3 (rabbit, Cell Signaling 9661, 1:600), Map2 (mouse, Sigma M4403-50UL, 1:500), phosphor-Histone H3 (rabbit, Millipore 06-570, 1:500), NeuN (mouse, Abcam,

ab104224, 1:300), Parvalbumin (mouse, Millipore, MAB1572, 1:250), Foxg1 (rabbit, Abcam, ab18259), Emx1 (rabbit, Sigma, HPA00641), Ctip2 (rat, Abcam, ab18465). For double labeling with Ki67 (rabbit, Leica, NCL-Ki67p, 1:300 and rat, Abcam, ab156956) and BrdU sections underwent heat induced antigen retrieval with Dako target retrieval solution (Dako, S1699). In brief, sections were washed 3 × 5 min in 1x PBS, emerged in antigen retrieval solution, heated up from 65° to 95°C, left at 95°C for 10 min, allowed to cool to RT for 30 min, washed 3 × 5 min in PBS followed by normal goat serum blocking and antibody incubation as described above. The next day sections were washed 3 × 5 min in PBS and incubated for 1 hour at RT with fluorophore-conjugated secondary antibodies. For EdU labeling, the Click-iT A647 (Invitrogen) kit was used. Secondary antibodies were donkey or goat Alexa Fluor A488, 555, 594, 647 conjugates (ThermoFisher, 1:1000). DNA nuclear counterstaining was performed using DAPI (ThermoFisher) or Hoechst (Invitrogen, 33,342) in 1x PBS before mounting in DAKO fluorescent mounting medium. Z-stack tile scans of whole organoid sections or VZ-like structures were acquired using LSM700 and LSM800 confocal systems (Zeiss).

Image processing and quantification

Images for analysis and Figures were adjusted for brightness/contrast in Fiji/ImageJ (Schindelin et al., 2012). Thickness of the Map2-positive layer was assessed only in developing cortical regions facing the outer border of the cerebral organoids. For each structure, three measurements were taken at 45° angles to obtain the mean value.

Semi-automated cell counting

Extraction of cell counting regions: Regions of interest (ROIs) (pixel size: 0.312 μm; voxel depth: 1.5 μm) were extracted with a custom ImageJ (ver. 1.52p) Python macro. After manual selection of polygon ROIs, cropped 2D images are extracted for segmentation in the selected reference channel (Hoechst 33342) and stored in the output directory. For all other channels, cropped 3D images are extracted with the number of z-slices specified by the user (±1 z-slices in reference to the ROIs z-position) and stored to the output directory. Pixels outside of the polygon ROI are set to 0.

Cell segmentation with Cellpose: For segmenting cell bodies in 2D in the extracted ROIs, we use the pre-trained deep neural network *cyto* from Cellpose (vers. 0.1.0.1) in single channel mode. We set the only required input parameter for Cellpose, the approximate cell diameter in pixel, to 18px (5.62 μm). Cellpose cell segmentations were done in batches of folders containing ROI reference images. Segmented nuclei are visually inspected to verify the quality of segmentation and the correct identification of cellular structures. Segmented structures that do not correspond to cell nuclei were removed from the analysis.

Cell filtering and counting: A custom Python module *cellpose_filter* then loads segmentation masks and previously extracted reporter channels for each ROI. It filters segmented cell masks based on well-established cell morphology measures (see below), displays them together with all extracted reporter channels for visual inspection, and computes mean fluorescence intensity values per cell in each reporter channels for all ROIs. Average and single z-slice intensity values are exported (Excel.xlsx) for subsequent thresholding.

To exclude potential cell segmentation with irregular shape and morphology, all the cell masks were filtered to have a minimum area of 19.46 μm^2 (200 px) and a maximum area of 97.32 μm^2 (1000 px). In addition, cell masks were discarded if their circularity is below 0.5, roundness below 0.2, and solidity below 0.8, where:

$$\text{Circularity} = 4\pi \frac{Area}{Perimeter^2} \in (0, 1]$$

$$\text{Roundness} = 4\pi \frac{Area}{Major\ axis^2} \in (0, 1]$$

$$\text{Solidity Convex} = \frac{Area}{Convex\ area} \in (0, 1]$$

With Major axis being the length of the ellipse fit having identical secondary moments of inertia.

All cell counts are computed by defining thresholds for each reporter channel manually for batches of images.

RNA sequencing

For the bulk RNA-seq experiment at day 10 after Matrigel droplet embedding we used twelve samples from a single batch (6x wild-type, two independent lines and 6x *CHD8*$^{+/-}$, two independent clones). Each sample was comprised of a pool of three organoids at day 10 were dissociated in 700 μL Trizol (Invitrogen) by passing them 20–30 times through a 20G needle. RNA extraction was performed by adding 140 μL chloroform (Sigma), followed by centrifugation for 15 min at 12,000 g at 4°C. The aqueous upper phase was transferred to a new tube and mixed with 1.5 volumes of 100% EtOH. RNA was extracted using Zymo-Spin IC columns (Zymo Research). In short, the aqueous phase/ethanol mixture was loaded onto the column, washed with 400 μL 70% EtOH before being treated with RQ1 DNAseI (Promega, 5μL + 5 μL reaction buffer +40 μL 70% EtOH) for 15 min at RT. After two washes with 70% EtOH the sample was eluted from the column with DEPC-treated H_2O. RNA concentration was measured using the NanoDrop spectrophotometer (Thermo Scientific) and quality was assessed with the Bioanalyzer 2100 employing the RNA 6000 Nano kit

(Agilent). 1 μg of RNA was used for cDNA library preparation using the QuantSeq 3′ mRNA-Seq FWD Library Prep Kit (Lexogen) according to the manual. Library quality and size distribution were assessed with the High Sensitivity DNA Analysis Kit for the Bioanalyzer 2100 (Agilent). Concentration was measured on a Qubit® 2.0 fluorometer using the Qubit® dsDNA HS Assay Kit and libraries were pooled in equimolar amounts. Sequencing was performed by the NGS facility of the Vienna Biocenter Core Facilities.

Transcriptomic analysis

Demultiplexed raw sequencing reads were trimmed before alignment according to the guidelines provided by Lexogen, i.e. random primer sequence, adapter contamination and low quality tails were filtered using BBDuk from the BBMap package (http://jgi.doe.gov/data-and-tools/bbtools/). Trimmed reads were aligned to the human genome using STAR version 2.5.4 (Dobin et al., 2013) (genome: hg38/GRCh38, gene annotation: Gencode release 24). Only uniquely aligning reads were kept for further analysis. Read counts per gene were also derived using STAR. Differential expression analysis was done in R 3.3.1 using the Bioconductor package DESeq2 (Love et al., 2014) version 1.12.4, employing an FDR threshold of 0.05, local fit for estimating dispersion and default parameters otherwise. The comparison was done at the genotype level, wildtype with mutant (6 samples vs 6 samples). To be more stringent regarding variability between lines differential expression analysis was also performed for the two wildtype lines alone and genes significantly dysregulated between wildtype lines were removed from the DEGs between wildtype and mutant. Gene Ontology enrichment analysis was done with the Bioconductor package GOstats (Falcon and Gentleman, 2007) version 2.36.0 with a p value cutoff of 0.001 and conditional testing enabled. GO enrichment results were visualized using the GOTreeVis package (https://github.com/dottercp/GOTreeVis) for CHD8 binding was performed using a published set of genes with CHD8 bound near their transcription start site based on ChIP-seq (Sugathan et al., 2014). (5601 genes after filtering for duplicate gene symbols). Data from the Simons Foundation Autism Research Initiative (SFARI) was downloaded in November 2020. For overlap with scRNA-seq data DEGs for each cluster were selected as p.adj $\leq$ 0.05 with no fold change cutoff. Statistical analysis for overlap between gene sets was done in R using Fisher's exact test and p values were adjusted for multiple testing with the Benjamini and Hochberg method. Visualization of differential expression, correlation, and CHD8 binding was done in R (ggplot2 package v3.3.5 (Wickham, 2016)).

Single cell transcriptomics

Organoids collected at days 20, 60 and 120 were dissociated by incubation with a solution of Papain (30U/mL, Worthington LS03126) and DNAseI (3U/uL, Zymo Research) for 30–45 min depending on organoid size. Dissociated suspensions were passed once through 0.4 mm Flowmi™ cell strainers, resuspended in PBS and counted using TC20 automatic cell counter (Biorad). Resulting single-cell suspension was mixed with RT-PCR master mix at a density of 1000 cells/μl and loaded together with Chromium Single-Cell 3′ gel beads and partitioning oil into a Chromium Single Cell 3′ Chip. The gel beads were coated with unique primers bearing 10x cell barcodes, unique molecular identifiers (UMI) and poly(dT) sequences. The chip was then loaded onto a Chromium instrument (10x Genomics) for single-cell GEM generation and barcoding. Amplified cDNAs were fragmented, and adapter and sample indices were incorporated into finished libraries, following manufacturer's instructions. The final libraries were quantified by Qubit system (Thermo) and calibrated with an in-house control sequencing library. The size profiles of the pre-amplified cDNA and sequencing libraries were examined by Agilent Bioanalyzer 2100 using a High Sensitivity DNA chip (Agilent). Two indexed libraries were equimolarly pooled and sequenced on Illumina NOVAseq 6000 platform using the v2 Kit (Illumina, San Diego, CA) with a customized paired-end, dual indexing format according to the recommendation by 10x Genomics. Using proper cluster density, a coverage around 250 M reads per sample (2000–5000 cells) were obtained corresponding to at least 50,000 reads/cell.

Reads were aligned to the hg38 human reference genome with Cell Ranger. Low-quality cells were discarded by imposing the following thresholds: Mitochondrial RNA more than 5% per cell and number of genes less than 450 per cell.

Before downstream analyses, data deriving from the 21 samples was integrated by Conos (Barkas et al., 2019); the resulting clusters, common among all samples, were considered as shared populations and used as such for the scGen variational autoencoder algorithm (Lotfollahi et al., 2019). On the integrated dataset, UMAP dimensionality reduction as implemented in Scanpy (Wolf et al., 2018) (version 1.6.0) was applied. Clusters were identified by applying Leiden algorithm from Scanpy, which is a community detection algorithm that has been optimized to identify communities that are guaranteed to be connected. This resulted in clusters of cells that are more coherent with the biological phenotype and more reliably identify cell populations. The resolution parameter value was optimized by surveying the stability of the resulting clusters, choosing a final value of 0.7. This resulted in the identification of 10 clusters.

Cluster annotation in cell populations was obtained by a combination of the following approaches: (I) scanpy's rank_genes_groups to identify the most characterizing genes per clusters; (II) SCINA semi-supervised annotations algorithm (Zhang et al., 2019), using as source labels coming from three publicly-available datasets (Lopez-Tobon et al., 2019; Nowakowski et al., 2017; Pollen et al., 2019), (III) Overlap of cluster marker genes (log2FC > 0.5, qval <0.05 compared to all other clusters) with cell type marker genes identified in a published single cell dataset (Nowakowski et al., 2017). Significance of overlap between marker sets was determined using Fisher's exact test; (IV) projection of a single cell fetal cortex data in the UMAP by ingest algorithm. The obtained information was finally manually curated using a self-hosted cellxgene (chanzuckerberg/cellxgene: An interactive explorer for single-cell transcriptomics data (github.com)), cellxgeneVIP (Li et al., 2020) and cellxgeneGateway (https://github.com/Novartis/cellxgene-gateway). Dotplots showing a panel of cell population specific genes across cluster was produced using Scanpy DotPlot function.

Pseudotime analysis was performed on clusters identified by Leiden algorithm with a resolution of 9. The resulting clusters were stratified by conditions/stage and then aggregated by their median to generate a new structure composed of 'supercells'. Supercells

with low contribution of reads for any mutation line were discarded. The Palantir algorithm (Setty et al., 2019) was applied on control supercells (CTL). An optimized version of the STREAM (Chen et al., 2019) algorithm was employed to visualize the result and calculate the different ratio.

Sample-wise frequency of cells in each cluster was visualized by mosaic plots, representing either the three control lines across stages or comparing at each stage controls versus mutants. For the comparisons of interest, the distribution of cells across clusters was analyzed by differential abundance analysis, performed following the procedure described in (http://bioconductor.org/books/release/OSCA/): edgeR was employed to test for differential abundance across stages in control lines (contrasts: day60 versus day20; day120 versus day60) or to compare mutant (excluding S62X) to control lines at each time point. The 'Unknown' cluster was excluded from this analysis. A threshold of conventional p Value <0.05 was imposed to select significant modulations.

Density plots were produced by dividing UMAP into sub-regions through a grid (Becht et al., 2018) and then by adding a Gaussian kernel to identify the density of cells per sub-region using python sklearn 0.23.2. For each stage, the plot shown in Figure 3 has been produced by subtracting the plot of the control from the mutant. Differential expression analysis was performed with diffxpy 0.7.4 with Mann-Whitney U test on the cells after removing the outlier cells identified by the Leiden algorithm. Resulting genes were filtered by qval<0.05 and |log2FC| > 1 unless otherwise specified in the text. Gene ontology enrichment analysis was performed in R with the TopGO package (Alexa and Rahnenfuhrer, 2021). *topGO: Enrichment Analysis for Gene Ontology*. R package version 2.42.0), considering the Biological Component domain of the ontology and using the Fisher statistics and the Weight01 algorithm. Highly variable genes were used as universe to calculate the enrichment. Terms were considered significant with a p Value <0.01 and an enrichment value of at least 1.75.

For alternative splicing analysis, single cell sequencing reads were aligned using Alevin (Srivastava et al., 2019). Cluster-wise pseudo-bulk samples were then generated for each line (excluding S62X) at each stage aggregating transcriptomics indexed reads by their median on the previously identified clusters and conditions. SUPPA2 (Trincado et al., 2018) was used for the calculation of PSI (proportion spliced-in) and dPSI for relative changes between mutant and control. Given the shallow sequencing characteristics of scRNA-Seq, a series of controls were implemented to check for possible spuriously significant dPSI related to lowly expressed genes and dropouts: (I) cells from control lines were randomized to *in-silico* generate four control samples to each stage, that were then compared two against two; (II) from the four *in-silico* controls, differential splicing analysis was applied comparing cluster RG1 vs cluster EN1. The common transcripts identified in the two control analyses were removed from the main MUT against CTL comparison to ensure the specificity of the transcripts identified.

For analyzing overlap with ZEB2 targets ChIP-seq data for ZEB2 in HEK293T cells was downloaded from GEO (GSE91749) in form of a bed file of called peaks. Peaks were then associated with genes using the annotatePeaks function from HOMER version 4.11 (Heinz et al., 2010) with default parameters. The results were filtered for peaks associated with TSS leading to a set of 8209 genes (after filtering for duplicate gene symbols). Enrichment analyses for ZEB2 targets and DEGs from previous studies was performed in R using Fisher's exact test. Results were visualized using ggplot2.

Enrichment for transcription factor targets in Radial Glia Cell clusters was done using the R interface package enrichR v3.0 for the Enrichr gene set enrichment analysis web server (Kuleshov et al., 2016). The dataset selected was ENCODE_and_ChEA_Consensus_TFs_from_ChIP-X. In case of multiple significant results for the same transcription factor we reported the result with the lower adjusted p value. We used a conservative adjusted p value cutoff of 0.005 for the supplementary figure. The results were visualized using R (packages: ggplot2 v3.3.5, ggh4x v0.2.1)

To complement our alternative splicing analysis we also used a different algorithm based on Olivieri et al. (Olivieri et al., 2021). Following the published pipeline we aligned the data with the pipeline of Dehghannasiri et al. (Dehghannasiri et al., 2021) and then calculated the SpliZVD statistics to identify alternative splicing events per cell. Differences in the number of significant transcripts is to be expected given the differences of the two methods. SUPPA2 (Trincado et al., 2018) pools together multiple cells resulting in a deeper characterization of the differences in splicing, a common approach to mitigate the dropout problem of single cells RNA-seq. The SplizVD method instead keeps single cell resolution but reduces the possibility to resolve small differences in transcript percentage.

Pathway analysis is performed with IPA (Ingenuity Pathway Analysis QIAGEN Digital Insights) using p value as observable and subselecting as source information Neurons and Nervous System. For each stage and each method the union of genes identified in any cluster is used. The canonical pathways listed in the table include all significant pathways derived from the results of either AS method. For Figure S9 we selected the EIF2 Signaling pathway with IPA (Ingenuity Pathway Analysis QIAGEN Digital Insights). Next we generated lists of population specific differentially spliced genes from the SUPPA2 results and used the overlay tools to highlight the genes/complexes included in these gene sets.

QUANTIFICATION AND STATISTICAL ANALYSIS

All biological experiments were performed with multiple independent biological samples as detailed in the figure legends. Numbers of samples are detailed in figure legends and Table S5 for each quantification performed, including numbers of cells, organoids, and batches for each cell line. Statistical test details can be found in the figure legends. All plots (unless otherwise indicated) were generated and statistical analyses were performed using GraphPad Prism 6.0 for Windows (GraphPad Software, La Jolla California USA, www.graphpad.com).