Article

Autism genes converge on asynchronous development of shared neuron classes

Bruna Paulsen[1,2,24], Silvia Velasco[1,2,16,24 ✉], Amanda J. Kedaigle[1,2,3,24], Martina Pigoni[1,2,24], Giorgia Quadrato[1,17], Anthony J. Deo[1,2,18,19,20], Xian Adiconis[2,3], Ana Uzquiano[1,2], Rafaela Sartore[1,2], Sung Min Yang[1,2], Sean K. Simmons[2,3], Panagiotis Symvoulidis[4], Kwanho Kim[1,2,3], Kalliopi Tsafou[2], Archana Podury[4,5], Catherine Abbate[1,2], Ashley Tucewicz[1,2], Samantha N. Smith[1,2], Alexandre Albanese[6], Lindy Barrett[1,2], Neville E. Sanjana[2,21,22], Xi Shi[2], Kwanghun Chung[6,7,21], Kasper Lage[2,8], Edward S. Boyden[4,5,9,10,11,12,13,14], Aviv Regev[3,15,23], Joshua Z. Levin[2,3] & Paola Arlotta[1,2 ✉]

Genetic risk for autism spectrum disorder (ASD) is associated with hundreds of genes spanning a wide range of biological functions[1–6]. The alterations in the human brain resulting from mutations in these genes remain unclear. Furthermore, their phenotypic manifestation varies across individuals[7,8]. Here we used organoid models of the human cerebral cortex to identify cell-type-specific developmental abnormalities that result from haploinsufficiency in three ASD risk genes—*SUV420H1* (also known as *KMT5B*), *ARID1B* and *CHD8*—in multiple cell lines from different donors, using single-cell RNA-sequencing (scRNA-seq) analysis of more than 745,000 cells and proteomic analysis of individual organoids, to identify phenotypic convergence. Each of the three mutations confers asynchronous development of two main cortical neuronal lineages—γ-aminobutyric-acid-releasing (GABAergic) neurons and deep-layer excitatory projection neurons—but acts through largely distinct molecular pathways. Although these phenotypes are consistent across cell lines, their expressivity is influenced by the individual genomic context, in a manner that is dependent on both the risk gene and the developmental defect. Calcium imaging in intact organoids shows that these early-stage developmental changes are followed by abnormal circuit activity. This research uncovers cell-type-specific neurodevelopmental abnormalities that are shared across ASD risk genes and are finely modulated by human genomic context, finding convergence in the neurobiological basis of how different risk genes contribute to ASD pathology.

ASD is a childhood-onset neurodevelopmental disorder that is characterized by cognitive, motor and sensory deficits[1]. ASD has a strong genetic component, with risk contribution from hundreds of genes[2–6]. Furthermore, the same mutation can result in varied clinical manifestations, probably reflecting a modulatory effect of the overall genetic and epigenetic background[7,8]. The shared developmental effects that cause this large and heterogeneous collection of genes to converge on the phenotypic features of ASD remain poorly understood.

Here we used reproducible organoid models of the developing human cerebral cortex[9] to investigate the roles of three ASD risk genes across multiple human stem cell lines. *SUV420H1*, *ARID1B* and *CHD8* have emerged repeatedly as top hits in studies of ASD genetic risk[6,10–13]. All three of these

[1]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [2]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [4]MIT Center for Neurobiological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [5]Harvard-MIT Health Sciences & Technology Program (HST), Harvard Medical School, Boston, MA, USA. [6]Institute for Medical Engineering and Science, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [7]Picower Institute for Learning and Memory, Departments of Chemical Engineering and Brain and Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [8]Department of Surgery and Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [9]McGovern Institute for Brain Research, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [10]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [11]Howard Hughes Medical Institute, MIT, Cambridge, MA, USA. [12]Department of Brain of Cognitive Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [13]Department of Media Arts and Sciences, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [14]Department of Biological Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [15]Department of Biology, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. [16]Present address: Murdoch Children's Research Institute, The Royal Children's Hospital, Parkville, Victoria, Australia. [17]Present address: Eli and Edythe Broad CIRM Center for Regenerative Medicine and Stem Cell Research, Department of Stem Cell Biology and Regenerative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [18]Present address: Department of Psychiatry, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [19]Present address: Department of Psychiatry, Rutgers-Robert Wood Johnson Medical School, Piscataway, NJ, USA. [20]Present address: Rutgers University Behavioral Health Care, Piscataway, NJ, USA. [21]Present address: New York Genome Center, New York, NY, USA. [22]Present address: Department of Biology, New York University, New York, NY, USA. [23]Present address: Genentech, South San Francisco, CA, USA. [24]These authors contributed equally: Bruna Paulsen, Silvia Velasco, Amanda J. Kedaigle, Martina Pigoni. ✉e-mail: silvia.velasco@mcri.edu.au; paola_arlotta@harvard.edu

genes are associated with severe neurodevelopmental abnormalities, including high frequencies of macrocephaly[11,14–16]. We show that mutations in these genes converge on asynchronous development of shared neuronal classes, rather than on shared molecular mechanisms. The degree of expressivity varies depending on the risk gene and phenotype, highlighting the nuanced interactions between the genetic variants and the genomic contexts that produce the phenotypic manifestation of ASD.

Organoids as models of ASD risk genes

To investigate whether mutations in different ASD risk genes converge on shared phenotypes, we generated cortical organoids[9] from different human induced pluripotent stem (iPS) cell lines (Methods) and profiled individual organoids using scRNA-seq at three stages: 1 month, when organoids contain mostly progenitors and neurogenesis is beginning; 3 months, when progenitor diversity increases and multiple subtypes of cortical excitatory neurons emerge; and 6 months, when interneurons and astroglia are present. We first verified that these organoids initiate appropriate neurodevelopment and express known ASD risk genes[6] (Supplementary Notes and Extended Data Figs. 1 and 2).

We next selected three ASD risk genes, *SUV420H1*, *ARID1B* and *CHD8* (Supplementary Notes), and engineered heterozygous protein-truncating indel mutations in multiple parental lines, targeting protein domains that are mutated in patients (Methods, Supplementary Table 1 and Extended Data Fig. 3a–c). Notably, for all genes, the different parental lines showed substantial differences in endogenous expression of the risk proteins, consistent with documented interindividual variability[7,8], which in turn influenced the absolute amount of protein remaining in the heterozygote (Supplementary Notes and Extended Data Fig. 3d–f). These differences underscore the importance of investigating risk genes across multiple genomic contexts.

As all three genes are linked to macrocephaly and/or microcephaly in patients, we quantified organoid size in each background, at two weeks and 1 month (Supplementary Table 2). Mutant lines showed size defects resembling the abnormalities seen in patients, with varying severity between different genomic contexts (Supplementary Notes and Extended Data Figs. 3g–j and 4a). These data indicate that organoids can capture clinically relevant features of ASD pathology.

Asynchronous development in *SUV420H1*-mutant organoids

We profiled early stages of *SUV420H1*+/– and control Mito294 (30,733 cells, 35 days in vitro (d.i.v.)), PGP1 (37,510 cells, 35 d.i.v.) and Mito210 (two independent differentiations: 57,941 cells, 28 d.i.v.; and 33,313 cells, 35 d.i.v.) organoids using scRNA-seq. Strikingly, mutants showed a consistent presence of GABAergic neurons in all backgrounds (Fig. 1a–c and Extended Data Fig. 4b–e), although these neurons are rare or absent in controls until approximately 3.5 months[9]. The GABAergic population in mutant organoids at 1 month expressed broad markers of GABAergic identity (hereafter, GABAergic neurons).

Despite the consistency of this phenotype across lines and differentiations, there were noticeable differences in phenotypic severity (expressivity) across genomic contexts. Specifically, the Mito294 *SUV420H1* line showed the most substantial increase in GABAergic neurons, with over 50% of the cells in all of the mutant organoids belonging to the GABAergic lineage, and <5% belonging to the excitatory projection neuron lineage (n = 3 organoids per genotype, adjusted *P* = 0.002, logistic mixed models; Fig. 1a and Extended Data Fig. 4b, c). PGP1 *SUV420H1* organoids showed intermediate severity, with up to 35% of cells in the mutants belonging to the GABAergic lineage (n = 2–3 organoids per genotype, adjusted *P* = 0.004, logistic mixed models; Fig. 1b and Extended Data Fig. 4d). Finally, Mito210 *SUV420H1* organoids showed the mildest phenotype, with no more than 5% of cells in the mutants belonging to the GABAergic lineage in one batch (28 d.i.v.,

n = 3 organoids per genotype, adjusted *P* = 0.017, logistic mixed models; Fig. 1c and Extended Data Fig. 4e) and no GABAergic neurons in a second differentiation batch at 35 d.i.v. (Extended Data Fig. 4f). This suggests that, although these cell lines converge on the same phenotype of premature generation of GABAergic neurons, the genetic and epigenetic context of each cell line modulates phenotypic expressivity.

We next investigated whether the increase in GABAergic neurons persisted at later stages. We profiled organoids from the two lines that showed the greatest difference in phenotypic severity (Mito294 and Mito210) at 3 months in vitro. At 3 months and beyond, GABAergic populations expressed clear molecular features of cortical interneurons (therefore indicated as GABAergic interneurons). The Mito294 *SUV420H1*+/– organoids still showed a disproportionately large GABAergic population (32,276 cells, n = 3 single organoids per genotype; Extended Data Fig. 5a). However, two distinct batches of Mito210 *SUV420H1* organoids showed no GABAergic interneurons in the mutant or control (Extended Data Fig. 5b, c). This indicates that, depending on its expressivity, the GABAergic phenotype can resolve over development (Mito210) or persist (Mito294).

We next sought to examine the changes in other cell types. Owing to the pronounced overgrowth of the GABAergic lineage in the Mito294 *SUV420H1*+/– organoids, most other cell types had reduced proportions (Fig. 1a and Extended Data Figs. 4c and 5a). However, in 1 month Mito210 *SUV420H1*+/– organoids, the milder GABAergic phenotype enabled us to detect an increase in immature deep-layer projection neurons, the first-born neurons of the cortical plate[17,18], in two differentiation batches (batch I, 28 d.i.v., adjusted *P* = 0.027; batch II, 35 d.i.v., adjusted *P* = 0.001; logistic mixed models, n = 3 single organoids per genotype; Fig. 1c and Extended Data Fig. 4e, f). Earlier cell types of the deep-layer projection neuron lineage (intermediate progenitor cells and early-postmitotic newborn deep-layer projection neurons) were also increased (Extended Data Fig. 4f). Similar to the transient GABAergic phenotype in this line, the deep-layer projection neuron phenotype was rescued after 3 months in vitro (two batches, 92 and 90 d.i.v.; Extended Data Fig. 5b, c).

In the PGP1 background, although the GABAergic phenotype was consistently observed, we did not observe an increase in the number of deep-layer projection neurons at 1 month (35 d.i.v.; Fig. 1b and Extended Data Fig. 4d). However, genes that were upregulated in the deep-layer projection neuron lineage in mutants from both the PGP1 and Mito210 lines were enriched in gene ontology (GO) terms related to neuronal differentiation and maturation (Methods, Supplementary Notes, Extended Data Fig. 4g and Supplementary Table 3), indicating a molecular profile that is consistent with more advanced neuronal maturation in both backgrounds. Interestingly, although the Mito210 line showed a more severe phenotype for the deep-layer projection neurons compared with the PGP1 line, the PGP1 line showed a more severe phenotype for the GABAergic neurons, indicating that different features of the mutant phenotype can be differentially modulated by the same genomic context.

Accelerated maturation of neuron classes

We next examined the developmental dynamics of the affected cell types within a specific lineage. Owing to the low numbers of GABAergic neurons in the control lines at 1 month, we focused on the deep-layer projection neuron phenotype. We calculated pseudotime trajectories, and identified the portion of the trajectory corresponding to the development of the affected cell types (the partition of interest; Methods and Extended Data Fig. 5d). The deep-layer projection neuron lineage in the combined 35 d.i.v. Mito210 *SUV420H1*+/– and control organoids (batch II) showed an increased distribution of mutant cells towards the end point of the trajectory ($P < 2.2 \times 10^{-16}$, one-sided Kolmogorov–Smirnov test; Fig. 1d, e), supporting accelerated development of these neurons in the mutants. Co-expression analysis using WGCNA[19] (Methods,
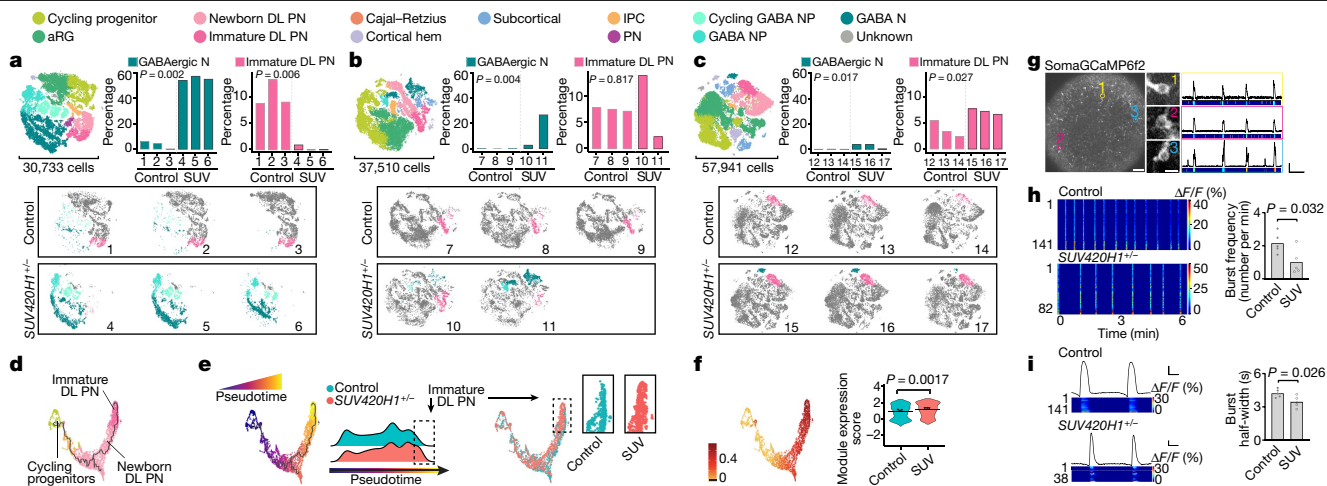
Fig. 1 | The *SUV420H1*⁺ᐟ⁻ genotype induces the asynchronous generation of GABAergic neurons and deep-layer projection neurons, and changes in circuit activity. a–c, Combined *t*-distributed stochastic neighbor embedding (*t*-SNE) analysis of all organoids (top left, with total cells per dataset) and the percentage of the indicated cell types per organoid (top right), colour-coded by cell type, for Mito294 (**a**; 35 d.i.v.), PGP1 (**b**; 35 d.i.v.) and Mito210 batch I (**c**; 28 d.i.v.) organoids. Adjusted *P* values were determined using logistic mixed models, comparing the difference in cell-type proportions between genotypes (Methods). Bottom, *t*-SNE analysis of individual organoids; cell types of interest are coloured. **d**, **e**, Pseudotime uniform manifold approximation and projection (UMAP) analysis of the Mito210 organoids (batch II, 35 d.i.v.), colour-coded by cell type (**d**), pseudotime (early, blue; late, yellow; **e**, left) or genotype (**e**, right). Insets: cells indicated by the dotted boxes, separated by genotype. **f**, Neuronal maturation and synapse formation module of highly correlated genes in Mito210 batch II cells from **d** and **e**, showing a UMAP plot of module scores (left) and the score distribution across the genotypes (right). The horizontal bars show the median scores, and the dots show the average

score per organoid. Adjusted *P* values were determined using linear mixed models, comparing differences between the control and mutant organoids (Methods). **g–i**, Calcium imaging of neuronal activity in intact PGP1 organoids (128 d.i.v.). **g**, Left, representative organoid transduced with SomaGCaMP6f2. Scale bar, 100 μm. Insets: high-magnification image of individual cells (1–3). Scale bar, 10 μm. Right, spontaneous calcium signal for each example cell as Δ*F/F* (top) and a pseudocolour heat map (bottom). Scale bars, 10% (vertical), 30 s (horizontal). **h**, Representative heat maps of calcium signal for each condition (left). Right, spontaneous network burst frequency. The dots show the average values per organoid and the bars show the mean across all organoids. **i**, The population-averaged calcium transients (top left) and heat map for individual cells (bottom left). Scale bars, 2% (vertical), 5 s (horizontal). Right, spontaneous network burst duration. The dots show the average values per organoid and the bars show the mean across all organoids. aRG, apical radial glia; DL, deep layer; GABA, GABAergic; IPC, intermediate progenitor cells; N, neurons; NP, neuron progenitors; PN, projection neurons; SUV, *SUV420H1*⁺ᐟ⁻.

Supplementary Table 4 and Extended Data Fig. 5e) identified a module containing multiple genes associated with neuronal maturation and synapse formation that was positively correlated with pseudotime progression (Fig. 1f; Pearson correlation *r* = 0.94, *P* < 2.2 × 10⁻¹⁶), and was significantly upregulated in mutant organoids (adjusted *P* = 0.0017, linear mixed models; Fig. 1f). These results support an accelerated differentiation phenotype in deep-layer projection neurons induced by *SUV420H1*⁺ᐟ⁻.

We next examined the mechanisms for the premature expression of maturation-associated genes in this mutant. As *SUV420H1* is a histone-lysine *N*-methyltransferase[20], we examined changes in chromatin accessibility. We performed a single-cell assay for transposase-accessible chromatin with high-throughput sequencing (scATAC-seq) on Mito210 *SUV420H1* organoids at 1 and 3 months (31 d.i.v., 84,696 nuclei; 93 d.i.v., 23,669 nuclei; *n* = 3 single organoids per genotype and timepoint). Co-embedding scATAC-seq and scRNA-seq data showed that chromatin accessibility captures most of the cell types identified by gene expression (Extended Data Fig. 6a).

At 1 month, most of the significant differentially accessible regions (DARs) between the mutant and control overlapped across cell types (Supplementary Table 5). We therefore combined all cells, and identified 414 DARs (Methods). The genes that were nearest (within 10 kb) to DARs with increased accessibility in mutant organoids were enriched for GO terms associated with synaptic transmission and neuronal maturation, whereas the genes that were nearest to DARs with reduced accessibility were enriched for negative regulation of neuronal maturation and connectivity (Extended Data Fig. 6b–d and Supplementary Table 5), consistent with the phenotypes observed in the scRNA-seq analysis.

At a later developmental stage (93 d.i.v.), we detected only 43 significant DARs (adjusted *P* < 0.1) across all cells (Supplementary Table 5). However, regions that were more accessible in the mutant were enriched for GO terms linked to synaptic function (Extended Data Fig. 6b), suggesting that differential regulation of genes that are important to neuronal maturation and function remains.

Regions with increased accessibility in the mutant were enriched for transcription factor (TF)-binding sites for regulators of neurogenesis and patterning of the developing nervous system, including multiple genes involved in the development of the GABAergic lineage (Methods, Supplementary Table 5 and Extended Data Fig. 6e).

Our results show that, in *SUV420H1*⁺ᐟ⁻ organoids, both GABAergic and deep-layer projection neurons exhibit accelerated development, and that the genomic context can differentially modulate phenotypic abnormalities affecting distinct cell types.

Reduced spontaneous circuit activity

The early developmental abnormalities in GABAergic and deep-layer projection neurons, along with the changes in expression and accessibility of synapse-associated genes, prompted us to investigate circuit activity.

We analysed spontaneous neuronal activity in a line with an intermediate phenotype (PGP1 *SUV420H1*; Fig. 1b), using adeno-associated viruses driving GCaMP (Methods) to record intracellular calcium dynamics in intact 4 month organoids (128 d.i.v.; Fig. 1g, Extended Data Fig. 7a and Supplementary Video 1). The predominant form of activity was a tetrodotoxin (TTX)-sensitive calcium signal (*n* = 10/10 organoids; Extended Data Fig. 7b), of which the large amplitude, slow kinetics
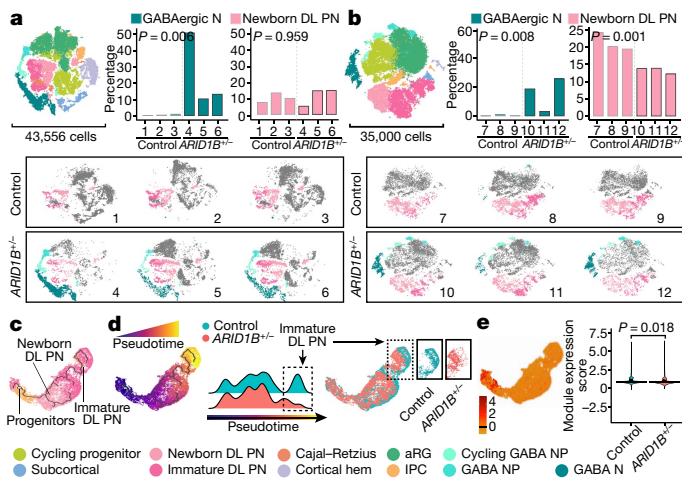
Fig. 2 | The *ARID1B*[+/−] genotype induces asynchronous generation of GABAergic neurons and deep-layer projection neurons. a, **b**, Combined *t*-SNE analysis of all organoids (top left), the percentage of the indicated cell types per organoid (top right) and *t*-SNE analysis for individual organoids (bottom), as in Fig. 1a–c, for Mito210 batch I (**a**; 35 d.i.v.) and Mito210 batch II (**b**; 35 d.i.v.) organoids. **c**, **d**, Pseudotime UMAP analysis of Mito210 batch II (35 d.i.v.) organoids as in Fig. 1d, e, colour-coded by cell type (**c**) and pseudotime (**d**). **e**, Cell division and proliferation module of highly correlated genes by co-expression network analysis of the Mito210 batch II cells in **c** and **d**, as in Fig. 1f.

and multipeak structure suggested that it was mediated by trains of action potentials (Extended Data Fig. 7c); this result was confirmed by extracellular single-unit recordings using a multielectrode array (MEA) (Extended Data Fig. 7d). These large calcium spikes occurred in periodic, synchronized bursts across most neurons (Extended Data Fig. 7e, f), resembling early network activity observed in the developing brain[21,22]. Network bursting was abolished after bath application of NBQX, an antagonist of non-NMDA glutamate receptors (Methods and Extended Data Fig. 7g), suggesting that coordinated network activity was driven by excitatory synaptic transmission.

Notably, after blockade of excitatory synapses with NBQX, only controls displayed calcium transients (Extended Data Fig. 7g, h), indicating that the control cells were more excitable, and probably more immature compared with the mutants. This is consistent with the accelerated molecular differentiation observed in *SUV420H1*[+/−] organoids.

Mutants showed a relative reduction in both frequency ($P = 0.032$, *t*-test; Fig. 1h) and duration ($P = 0.026$, *t*-test; Fig. 1i) of network bursts (Extended Data Fig. 7i, j), indicating that *SUV420H1*[+/−] organoids have reduced spontaneous activity, consistent with mouse models[23]. These data suggest that, beyond the molecular and cellular changes observed in mutant organoids, the *SUV420H1*[+/−] genotype can also induce long-term abnormalities in circuit activity.

ARID1B and *SUV420H1* share target populations

We next investigated whether changes in the production of neuronal classes were a shared feature of ASD risk genes. We profiled individual Mito210 *ARID1B*[+/−] and control organoids from two independent differentiations at 1 month (35 d.i.v.) using scRNA-seq (batch I: 43,556 cells; batch II: 35,000 cells; Fig. 2a, b and Extended Data Fig. 8a–c).

Although the controls had few or no GABAergic lineage cells at this age, Mito210 *ARID1B*[+/−] organoids showed a consistent increase in the proportions of GABAergic neurons and their progenitors (batch I and II, respectively; GABAergic neurons: adjusted $P = 0.0057$, $P = 0.0076$; GABAergic neuron progenitors: adjusted $P = 0.0004$, $P = 0.0128$; cycling GABAergic neuron progenitors: adjusted $P = 0.0004$, $P = 0.0001$;

logistic mixed models, $n = 3$ single organoids per genotype; Fig. 2a, b and Extended Data Fig. 8b, c). In the first batch, GABAergic neurons constituted up to 50% of the profiled cells, making it difficult to draw conclusions about other cell types (Fig. 2a and Extended Data Fig. 8b). In the second batch, the GABAergic phenotype was less severe, enabling us to detect a significant reduction in newborn deep-layer projection neurons (adjusted $P = 0.001$, logistic mixed models; Fig. 2b and Extended Data Fig. 8c). Notably, these are the same two neuron populations that are affected in *SUV420H1*[+/−] organoids. Although the *ARID1B*[+/−] genotype had an opposite effect on the deep-layer projection neuron lineage, the data show convergence of two previously unrelated risk genes in the classes of cells that they affect. This phenotype of increased GABAergic populations was still detectable at a later developmental stage, 3 months, although it was less severe (Extended Data Fig. 8d–f).

To test the effect of genomic context, we generated control and *ARID1B*[+/−] organoids in the Mito294 background, and profiled 50,081 cells at 35 d.i.v. ($n = 3$ per genotype). Consistent with the Mito210 *ARID1B* phenotype, there was a decreased number of newborn deep-layer projection neurons in the mutant ($P = 0.025$, logistic mixed models; Extended Data Fig. 8g). However, there was no significant increase in the GABAergic population in this background ($P = 0.24$, logistic mixed models; Extended Data Fig. 8g). This line, Mito294, had the most pronounced increase in GABAergic neurons in *SUV420H1*[+/−] organoids, showing that the genomic context modifies the expressivity of each mutation differently.

Pseudotime analysis (Extended Data Fig. 8h) showed a decreased distribution of cells towards the end point of the trajectory, progressing from progenitors to deep-layer projection neurons in Mito210 *ARID1B*[+/−] organoids at 1 month (35 d.i.v., batch II; $P < 2.2 \times 10^{-16}$, one-sided Kolmogorov–Smirnov test; Fig. 2c, d). Gene module analysis (Extended Data Fig. 8i) identified a module containing multiple DNA-replication and cell-cycle genes (Supplementary Table 3) that was enriched in progenitor cells and was significantly upregulated in mutants (adjusted $P = 0.018$, linear mixed models; Fig. 2e). These data indicate delayed differentiation of deep-layer projection neurons in *ARID1B*[+/−] organoids.

In sum, similarly to *SUV420H1*[+/−] organoids, *ARID1B*[+/−] organoids exhibit both a phenotype of premature expansion of the GABAergic neuron lineage, and asynchronous development of deep-layer projection neurons. Notably, as in *SUV420H1*[+/−] organoids, these two phenotypes were differentially modulated in distinct parental lines in the *ARID1B*[+/−] organoids.

CHD8[+/−] genotype promotes interneuron development

To further examine the hypothesis of convergent phenotypes among ASD risk genes, we profiled HUES66 *CHD8*[+/−] and control organoids at 3.5 months (109 d.i.v., 67,024 cells, $n = 3$ single organoids per genotype; Fig. 3a and Extended Data Fig. 9a). Mutants had an increased number of GABAergic interneurons and their progenitors (GABAergic interneurons: adjusted $P = 0.079$; cycling GABAergic interneuron progenitors: adjusted $P = 0.031$; GABAergic interneuron progenitors: adjusted $P = 0.0012$, logistic mixed models; Fig. 3a and Extended Data Fig. 9a). A second independent batch of HUES66 *CHD8* organoids showed an even more substantial increase ($n = 2–3$ single organoids per genotype; cycling GABAergic interneuron progenitors: adjusted $P = 7.2 \times 10^{-5}$; GABAergic interneuron progenitors: adjusted $P = 1.8 \times 10^{-5}$; GABAergic interneurons: adjusted $P = 8.9 \times 10^{-6}$, logistic mixed models; Fig. 3b and Extended Data Fig. 9b–d). Notably, at 6 months (190 d.i.v., 39,285 cells, $n = 3$ individual organoids per genotype), the GABAergic interneuron phenotype was still present (adjusted $P = 0.002$, logistic mixed models; Extended Data Fig. 9d–f).

This increase in GABAergic populations is consistent with two recent reports showing that *CHD8*[+/−] affects the expression of GABAergic interneuron marker genes in two additional human parental lines[24,25]. However, as we found for *SUV420H1*[+/−] and *ARID1B*[+/−] organoids, the genomic context was able to modulate the expressivity of the *CHD8*[+/−] phenotype. We compared *CHD8*[+/−] and control organoids generated
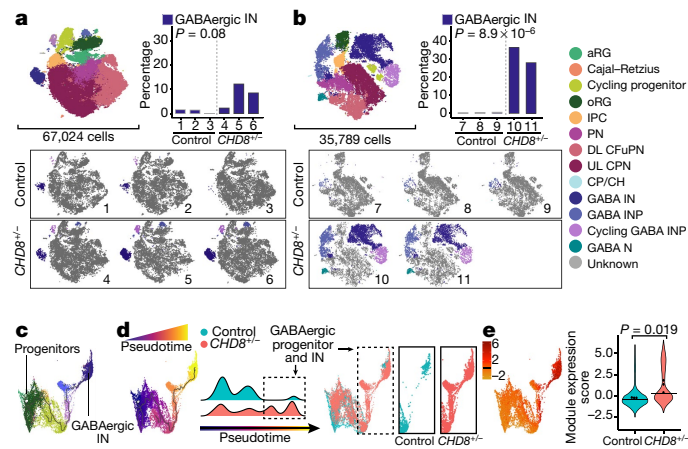
Fig. 3 | The *CHD8*[+/-] genotype leads to asynchronous generation of GABAergic interneurons. a, **b**, Combined *t*-SNE analysis of all organoids (top left), the percentage of the indicated cell types for each organoid (top right) and *t*-SNE analyses for individual organoids (bottom), as in Fig. 1a–c, for HUES66 batch I (**a**; 109 d.i.v.) and HUES66 batch II (**b**; 107 d.i.v.) organoids. **c**, **d**, Pseudotime UMAP analysis of HUES66 batch I organoids (109 d.i.v.), as in Fig. 1d, e, colour-coded by cell type (**c**) and pseudotime (**d**). **e**, Interneuron differentiation module of highly correlated genes by co-expression network analysis of the HUES66 batch I cells in **c** and **d**, as in Fig. 1f. CFuPN, corticofugal projection neurons; INP, interneuron progenitors; IN, interneurons; oRG, outer radial glia; UL CPN, upper layer callosal projection neurons.

from four different parental lines spanning different basal levels of CHD8 protein expression (Extended Data Fig. 3c, f and Supplementary Table 1). Bulk RNA-seq analysis of 35 d.i.v. organoids showed that, although differentially expressed genes (DEG) between the mutant and control did not significantly overlap between lines, DEGs from three out of the four lines (HUES66, GM08330 and H1) shared GO terms related to neurodevelopment and neuronal maturation (Extended Data Fig. 10a and Supplementary Table 6). However, scRNA-seq analysis of *CHD8*[+/-] and control organoids from GM08330 and H1 lines at 3.5 months showed no significant difference in the number of GABAergic interneurons (105–108 d.i.v., *n* = 3 individual organoids per genotype, 107,490 cells; Extended Data Fig. 10b–d).

Pseudotime analysis of the GABAergic lineage (progressing from radial glia to GABAergic interneurons) in 3.5 month HUES66 *CHD8*[+/-] and control organoids showed an increased distribution of mutant cells towards the end point of the developmental trajectory ($P < 2.2 \times 10^{-16}$, one-sided Kolmogorov–Smirnov test; Fig. 3c, d and Extended Data Fig. 9g). Gene module analysis of the GABAergic lineage (Extended Data Fig. 9h and Supplementary Table 4) identified a module of interneuron differentiation genes that was upregulated in HUES66 *CHD8*[+/-] organoids (adjusted $P = 0.019$, linear mixed models; Fig. 3e), and two modules related to progenitor biology that were downregulated in the mutant (Extended Data Fig. 9h).

Thus, similar to the *SUV420H1*[+/-] and *ARID1B*[+/-] genotypes, the *CHD8*[+/-] genotype leads to an accelerated development of the GABAergic lineage that, for *CHD8*, leads to a persistent increase in the proportion of these cell types. For all three risk genes, this phenotype occurs in multiple parental lines, but with different degrees of phenotypic expressivity.

Convergence through distinct mechanisms

We next investigated whether *SUV420H1*[+/-], *ARID1B*[+/-] and *CHD8*[+/-] organoids converged on asynchronous development of the same neuronal lineages by acting through common molecular pathways. We compared gene expression changes across the three ASD risk genes in cell lines that showed a strong phenotype (Mito210 *SUV420H1*, Mito210 *ARID1B* and HUES66 *CHD8*). Although mutations shared enrichment
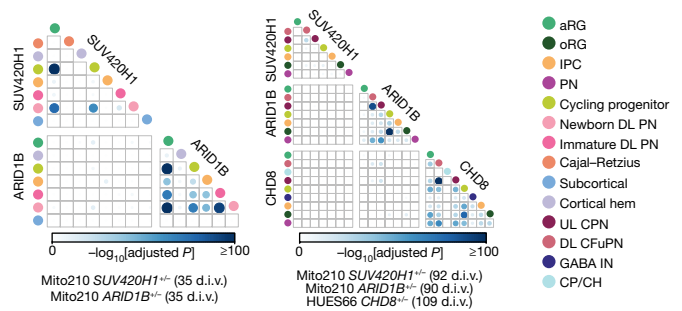

Fig. 4 | The *SUV420H1*[+/-], *ARID1B*[+/-] and *CHD8*[+/-] genotypes act through distinct gene targets. Overlap between the sets of DEGs (mutant versus control) in individual cell types from the scRNA-seq datasets. The colour and size of the boxed circles indicates the significance of the number of overlapping genes between the corresponding cell populations (Bonferroni-adjusted *P* value, determined using hypergeometric tests). CP/CH, choroid plexus/cortical hem.

for GO categories, DEGs from bulk or pseudobulk RNA-seq analysis showed little overlap (Supplementary Notes, Supplementary Table 7 and Extended Data Fig. 11a–d). Similarly, although related cell types within the same mutation shared overlapping DEGs, DEGs caused by different mutations rarely overlapped, even for identical or closely related cell types (Fig. 4, Supplementary Notes and Supplementary Table 7). Thus, although these three mutants share a degree of convergence in altered neurodevelopmental processes, they affect largely distinct genes.

Whole-proteome mass spectrometry (MS) analysis of mutant and control single organoids (Methods) identified 233 significantly differentially expressed proteins (DEPs; false-discovery rate (FDR) < 0.1) for *SUV420H1*[+/-] (≥4,000 proteins detected per sample), 24 for *ARID1B*[+/-] (≥900 proteins) and 34 for *CHD8*[+/-] (≥2,800 proteins; Extended Data Fig. 12a–c and Supplementary Table 8) organoids. DEPs had a very low overlap between mutations, with only five proteins significantly dysregulated in at least two mutations (Supplementary Table 8). DEPs and enriched biological processes (gene set enrichment) for all mutations resembled the gene modules that were identified by scRNA-seq analysis (Supplementary Notes and Extended Data Fig. 12d–f).

To evaluate whether the affected proteins in the three mutants were predicted to interact with one another, or with shared target proteins, the top 50 DEPs (adjusted *P*) for each mutation were used to create a network of interacting proteins[26,27], followed by clustering to identify subnetworks (Methods). Each subnetwork contained DEPs from multiple mutations (Supplementary Notes and Extended Data Fig. 12g–i), indicating that these three risk genes affect shared processes, albeit by dysregulating different proteins.

Discussion

The process by which mutations in ASD risk genes converge on the neurobiology of this multifaceted disorder remains unclear. Our results define two neuronal classes of the local cortical circuit (GABAergic and deep-layer projection neurons) as specifically affected populations. Excitatory/inhibitory imbalance of the cortical microcircuit is a major hypothesis for the aetiology of ASD[28–30], and previous studies have implicated the dysregulation of GABAergic and glutamatergic cortical neurons in ASD patients and experimental models[31–37]. Notably, we show that different human genomic contexts modulate phenotypic expressivity, based on both the risk gene and the specific abnormalities caused by each mutation. This is interesting, as many ASD risk gene mutations show variable clinical manifestations in humans[7,8,14,38].

Our finding that different ASD risk genes converge on a phenotype of asynchronous neuronal development but mostly diverge at the level

of molecular targets suggests that a shared clinical pathology of these genes may derive from higher-order processes of neuronal differentiation and circuit wiring. These results encourage future investigation of therapeutic approaches aimed at the modulation of shared dysfunctional circuit properties in addition to shared molecular pathways.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04358-6.

1. Lord, C. et al. Autism spectrum disorder. *Nat. Rev. Dis. Primers* **6**, 5 (2020).
2. Rosenberg, R. E. et al. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch. Pediatr. Adolesc. Med.* **163**, 907–914 (2009).
3. Sanders, S. J. et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
4. Ruzzo, E. K. et al. Inherited and de novo genetic risk for autism impacts shared networks. *Cell* **178**, 850–866 (2019).
5. Grove, J. et al. Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
6. Satterstrom, F. K. et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell* **180**, 568–584 (2020).
7. Cooper, D., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130 (2013).
8. Zlotogora, J. Penetrance and expressivity in the molecular age. *Genet. Med.* **5**, 347–352 (2003).
9. Velasco, S. et al. Individual brain organoids reproducibly form cell diversity of the human cerebral cortex. *Nature* **570**, 523–527 (2019).
10. de Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).
11. Stessman, H. A. F. et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* **49**, 515–526 (2017).
12. Yuen, R. K. C. et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* **20**, 602–611 (2017).
13. Sanders, S. J. et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015).
14. Bernier, R. et al. Disruptive *CHD8* mutations define a subtype of autism early in development. *Cell* **158**, 263–276 (2014).
15. Faundes, V. et al. Histone lysine methylases and demethylases in the landscape of human developmental disorders. *Am. J. Hum. Genet.* **102**, 175–187 (2018).
16. Vals, M. et al. Coffin-Siris syndrome with obesity, macrocephaly, hepatomegaly and hyperinsulinism caused by a mutation in the *ARID1B* gene. *Eur. J. Hum. Genet.* **22**, 1327–1329 (2014).
17. Lodato, S. & Arlotta, P. Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015).
18. Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H. & Macklis, J. D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
19. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).
20. Wickramasekara, R. N. & Stessman, H. A. F. Histone 4 lysine 20 methylation: a case for neurodevelopmental disease. *Biology* **8**, 11 (2019).
21. Garaschuk, O., Linn, J., Eilers, J. & Konnerth, A. Large-scale oscillatory calcium waves in the immature cortex. *Nat. Neurosci.* **3**, 452–459 (2000).
22. Adelsberger, H., Garaschuk, O. & Konnerth, A. Cortical calcium waves in resting newborn mice. *Nat. Neurosci.* **8**, 988–990 (2005).
23. Wang, Z.-J. et al. Autism risk gene *KMT5B* deficiency in prefrontal cortex induces synaptic dysfunction and social deficits via alterations of DNA repair and gene transcription. *Neuropsychopharmacology* **46**, 1617–1626 (2021).
24. Villa, C. E. et al. CHD8 haploinsufficiency alters the developmental trajectories of human excitatory and inhibitory neurons linking autism phenotypes with transient cellular defects. Preprint at *bioRxiv* https://doi.org/10.1101/2020.11.26.399469 (2020).
25. Wang, P. et al. 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 (2017).
26. Tuncbag, N. et al. Network-based interpretation of diverse high-throughput datasets through the Omics Integrator software package. *PLoS Comput. Biol.* **12**, e1004879 (2016).
27. Szklarczyk, D. et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).
28. Rubenstein, J. L. R. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
29. Gogolla, N. et al. Common circuit defect of excitatory-inhibitory balance in mouse models of autism. *J. Neurodev. Disord.* **1**, 172–181 (2009).
30. Dani, V. S. et al. Reduced cortical activity due to a shift in the balance between excitation and inhibition in a mouse model of Rett syndrome. *Proc. Natl Acad. Sci. USA* **102**, 12560–12565 (2005).
31. Mariani, J. et al. FOXG1-dependent dysregulation of GABA/glutamate neuron differentiation in autism spectrum disorders. *Cell* **162**, 375–390 (2015).
32. Marchetto, M. C. et al. Altered proliferation and networks in neural cells derived from idiopathic autistic individuals. *Mol. Psychiatry* **22**, 820–835 (2017).
33. Adhya, D. et al. Atypical neurogenesis in induced pluripotent stem cells from autistic individuals. *Biol. Psychiatry* **89**, 486–496 (2020).
34. Wade, A. A., Lim, K., Catta-Preta, R. & Nord, A. S. Common CHD8 genomic targets contrast with model-specific transcriptional impacts of CHD8 haploinsufficiency. *Front. Mol. Neurosci.* **11**, 481 (2019).
35. Moffat, J. J., Smith, A. L., Jung, E. M., Ka, M. & Kim, W. Y. Neurobiology of ARID1B haploinsufficiency related to neurodevelopmental and psychiatric disorders. *Mol. Psychiatry* https://doi.org/10.1038/s41380-021-01060-x (2021).
36. Velmeshev, D. et al. Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685–689 (2019).
37. Willsey, A. J. et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).
38. Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 551–563 (2015).

Article

Methods

PS cell culture

The HUES66 *CHD8* parental hESC line[39] and *CHD8* mutant line (HUES66 AC2), a clone that has a heterozygous 13-nucleotide deletion, resulting in a stop codon at amino acid 1248 (*CHD8* gRNA: 5′-TTCTTACTGTGTACCCGGGC-3′ (TGG)), were provided by N. Sanjana, X. Shi, J. Pan and F. Zhang (Broad Institute of MIT and Harvard). The psychiatric control Mito210 and Mito294 parental iPS cell lines were provided by B. Cohen (McLean Hospital); the parental PGP1 iPS cell line by G. Church (Harvard University)[40]; the GM08330 iPS cell line (also known as GM8330-8) by M. Talkowski (MGH) and was originally from Coriell Institute; and the H1 parental hESC line (also known as WA01) was purchased from WiCell. Cell lines were cultured as previously described[9,41]. Among these cell lines, we included iPS cell lines from individuals with no known history of ASD or other psychiatric condition (Mito210 and Mito294 confirmed by structured psychiatric interview, PGP1 with publicly available records). All human pluripotent stem (PS) cell lines were maintained below passage 50, were negative for mycoplasma (assayed with MycoAlert PLUS Mycoplasma Detection Kit, Lonza) and karyotypically normal (G-banded karyotype test performed by WiCell Research Institute). The HUES66 and PGP1 lines were authenticated using STR analysis completed by GlobalStem (in 2008) and TRIPath (in 2018), respectively. The Mito210 and Mito294 lines were authenticated by genotyping analysis (Fluidigm FPV5 chip) performed by the Broad Institute Genomics Platform (in 2017). The H1 and GM08330 lines were authenticated using STR analysis completed by WiCell (in 2021). In the Mito294 *ARID1B* control line, a CNV smaller than 0.5 Mb on chromosome 19 was detected by single-nucleotide polymorphism array analysis. The GM08330 parental line and edited lines all have an interstitial duplication in the long (q) arm of chromosome 20. All of the experiments involving human cells were performed according to ISSCR 2021 guidelines[42], and approved by the Harvard University IRB and ESCRO committees.

CRISPR guide design

The CRISPR guides for *SUV420H1* and *ARID1B* were designed using the Benchling CRISPR Guide Design Tool (Benchling Biology Software, 2017). The guides were designed to maximize on-target efficiency and minimize off-target sites in intragenic regions[43,44]. For *SUV420H1*, a guide was designed to target the N-terminal domain to create a protein truncation early in the translated protein in all known protein coding transcripts (*SUV420H1* gRNA: 5′-CAAGAACCAAACTGGTTGCT-3′ (AGG)). The *ARID1B* guide was chosen to induce a stop codon immediately before the ARID domain (*ARID1B* gRNA: 5′-CTCTAGCCTGATGAACACGC-3′ (AGG)). For *CHD8*, all of the mutant lines were generated using the same gRNA previously used for the generation of the HUES66 AC2 (*CHD8* gRNA: 5′-TTCTTACTGTGTACCCGGGC-3′ (TGG)).

CRISPR-mediated gene editing

Mutations in *SUV420H1* were introduced in the Mito210, Mito294 and PGP1 iPS cell lines. For the Mito210 and Mito294 *SUV420H1* mutant lines, nanoblades that were generated as previously described[45] were mixed with 300 μl of mTeSR1 and 4 μg ml⁻¹ of polybrene and added to 80% confluent cells. For selection of the edited clones, cells were enzymatically detached and plated at a ratio of ~5,000 cells per 60 mm dish with 10 μM of ROCK inhibitor (Y-27632, Millipore-Sigma) to increase single-cell survival. When the colonies started to appear, each clone was manually collected and transferred into a single well of a 96-well plate. During colony picking, some of the cells were reserved for DNA extraction and clonal screening. The PGP1 *SUV420H1*-mutant line was generated in collaboration with the Harvard Stem Cell Institute (HSCI) iPS Core Facility. In brief, parental cells were transfected using the Neon system (1,000 V, 1,100 V or 1,200 V, 30 ms, 1 pulse). For 100,000 cells, 6.25 pmol TrueCut Cas9 Protein v2 (Thermo Fisher Scientific, A36496)

and 12.5 pmol of sgRNA (Synthego) were used. After transfection, the pools of cells were collected to test knock-out efficiency. The best pool was then selected for low-density plating (600 to 2,000 cells per 10 cm dish). A week later, colonies were picked into one 96-well plate. Clones were screened by PCR and Sanger sequencing. Heterozygous clones were expanded and the genotypes were reconfirmed after expansion.

Mito210 and Mito294 *ARID1B*-edited lines were generated by the Broad Institute Stem Cell Facility. The guide RNA and Cas9 (EnGen Cas9 NLS from New England Biolabs) were transfected by using the NEON transfection system (Thermo Fisher Scientific, 1,050 V, 30 ms, 2 pulses and 2.5×10^5 cells).

Mutations in *CHD8* were introduced in the Mito210 and Mito294 lines using the Amaxa 4D-Nucleofector (Lonza), using the protocol optimized for PS cell lines. Parental cell lines were transfected with gRNA-*CHD8*-Cas92APuro and immediately plated in mTeSR1 for 24 h. Selection of transfected cells was performed by adding 0.25–0.5 μg ml⁻¹ of puromycin after 48 h of transfection, for 2 days. Selection of the edited clones was performed according to the protocol described for the Mito210 and Mito294 *SUV420H1* clones. The H1 and GM08330 *CHD8*-mutant lines were generated in collaboration with the HSCI iPS Core Facility according to the protocol that was used to generate the PGP1 *SUV420H1*-mutant line.

Sequence confirmation of edits

Insertions and deletions in individual clones were screened by PCR amplification using primers flanking the guide. Further details about the insertions/deletions are provided in Supplementary Table 1.

Organoid differentiation

Cortical organoids were generated as previously described[9,41]. Embryoid bodies were formed in the same pluripotent medium in which they were maintained for 1–2 days to better enable the formation of embryoid bodies from each line (except for the lines Mito210 *SUV420H1* and HUES66 *CHD8* where cells were plated directly in CDM1 as previously described[9,41]).

Immunohistochemistry

Samples were prepared as previously described[9]. Cryosection thickness varied from 14 μm to 18 μm. A list of the primary and secondary antibodies is provided in Supplementary Table 9.

Whole-organoid imaging

Organoids in Extended Data Fig. 4a were processed using the SHIELD protocol[46]. In brief, organoids were fixed for 30 min in 4% paraformaldehyde (PFA) at room temperature and were then treated with 3% (w/v) polyglycerol-3-polyglycidyl ether (P3PE) for 48 h in ice cold 0.1 M phosphate buffer (pH 7.2) at 4 °C then transferred to 0.3% P3PE in 0.1 M sodium carbonate (pH 10) for 24 h at 37 °C. The samples were rinsed and cleared in 0.2 M SDS in 50 mM phosphate-buffered saline (pH 7.3) for 48 h at 55 °C. Organoids were stained with Syto16 (Thermo Fisher Scientific, S7578) and anti-SOX2 antibodies using the SmartLabel system (Lifecanvas). A list of the primary antibodies is provided in Supplementary Table 9. Tissues were washed extensively for 24 h in phosphate-buffered saline + 0.1% Triton X-100 and antibodies were fixed to the tissue using a 4% PFA solution overnight at room temperature. Tissues were refractive-index-matched in PROTOS solution (RI = 1.519) and imaged using a SmartSPIM axially swept light-sheet microscope (LifeCanvas Technologies). 3D image datasets were acquired using a ×15/0.4 NA objective (ASI-Special Optics, 54-10-12). Optical sections from whole-organoid datasets are shown in Extended Data Fig. 4a.

Microscopy and organoid size analysis

Images of organoids in culture were taken with an EVOS FL microscope (Invitrogen), Lionheart FX Automated Microscope (BioTek Instruments), or with an Axio Imager.Z2 (Zeiss). Immunofluorescence images

were acquired with the latter two and analysed with the Gen5 (BioTek Instruments) or Zen Blue (ZEN 2.6 – blue edition, Zeiss) image processing software. ImageJ[47] (v.2.0) was used to quantify organoid size. Area values were obtained by tracing individual organoids on ImageJ, which measured area pixels. Measurements were plotted as a ratio to the average value for control organoids of each experimental batch. A summary of the number of organoids and differentiations used for the measurements is provided in Supplementary Table 2.

Western blotting
Proteins were extracted from iPS cells using N-PER Neuronal Protein Extraction Reagent (Thermo Fisher Scientific) supplemented with protease (cOmplete Mini Protease Inhibitor Cocktail, Roche) and phosphatase (PhosSTOP, Sigma-Aldrich) inhibitors. Lysates were centrifuged for 10 min at 13,500 rpm at 4 °C. Protein concentration was quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Protein lysates (15–20 μg) were separated on a NuPAGE 4–12%, Bis-Tris Gel (Invitrogen) or Mini-PROTEAN 4–15% Gels (Bio-Rad) and transferred onto a polyvinylidene difluoride membrane. Blots were blocked with 5% non-fat dry milk (Bio-Rad) and incubated with primary antibodies overnight (Supplementary Table 9). The blots were then washed and incubated at room temperature with secondary horseradish peroxidase-conjugated antibodies (Abcam) for 1 h. The blots were developed using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) or ECL Prime Western Blotting System (Millipore), and the ChemiDoc System (Bio-Rad). Densitometry band quantification was performed using Fiji software[48] v.2.0 and normalized to housekeeping genes (*GAPDH* or *ACTB*). The bands used for quantification are marked with an asterisk in Extended Data Fig. 3d–f. Uncropped gel images of western blots are provided in Supplementary Fig. 1.

Calcium imaging
Organoids were transduced with pAAV-CAG-SomaGCaMP6f2 (Addgene, 158757) by pipetting 0.2 μl of stock virus into 500 μl Cortical Differentiation Medium IV (CDMIV, without Matrigel) in a 24-well plate containing a single organoid. The next day, each organoid was transferred to a 6-well plate filled with 2 ml of fresh medium. On the third day after transduction, organoids were transferred to low-attachment 10 cm plates and, on the seventh day, the medium was switched to BrainPhys (5790, STEMCELL Technologies) supplemented with 1% N2 (17502-048, Thermo Fisher Scientific), 1% B27 (17504044, Thermo Fisher Scientific), GDNF (20 ng ml^{-1}, 78139, STEMCELL Technologies), BDNF (20 ng ml^{-1}, 450-02, Peprotech), cAMP (1 mM, 100-0244, STEMCELL Technologies), ascorbic acid (200 nM, 72132, STEMCELL Technologies) and laminin (1 μg ml^{-1}, 23017015, Life Technologies). The organoids were cultured in BrainPhys for at least 2 weeks before imaging.

Brain organoids were randomly selected and transferred to a recording chamber containing BrainPhys. Imaging was performed using a confocal scanner (CSU-W1, Andor confocal unit attached on an inverted microscope (Ti-Eclipse and NIS-elements imaging software (NIS-Elements Advance Research (v.4.51.01)), both from Nikon)), while the organoids were kept at 37 °C using a heating platform and a controller (TC-324C, Warner Instruments). The use of a ×10 objective (Plan Apo λ, ×10/0.45 NA) resulted in a field of view of 1.3 × 1.3 mm^2 and a pixel size of 0.6 μm. Imaging took place in fast-time-lapse mode, with an exposure time of 100 ms, resulting in an acquisition rate of 8 f.p.s. Spontaneous activity was recorded in three different z planes, for at least 22 min of baseline activity in total (with no pharmacology treatment).

Stock solutions of 2,3-dioxo-6-nitro-1,2,3,4-tetrahydrobenzo[*f*] quinoxaline-7-sulfonamide disodium salt (NBQX disodium salt, Abcam; 100 mM) and tetrodotoxin citrate (TTX, Abcam; 10 mM) were prepared in double-distilled H$_2$O. Bath application of NBQX (antagonist of AMPA/kainate glutamate receptors) and TTX (voltage-gated sodium-channel antagonist) was applied to achieve a final bath concentration of 20 μM and 2 μM, respectively.

Data were converted from nd2 format to tiff, and automated motion correction and cell segmentation were performed using Suite2p[49], followed by manual curation of segmented cells (we examined the spatial footprint and temporal characteristics of each candidate cells, as well as manually adding neurons with clear cell body morphology; Fig. 1g). The mean raw fluorescence for each cell was then measured as a function of time.

Analysis of calcium imaging data
Analysis was performed using custom MATLAB scripts. Raw calcium signals for each cell, $F(t)$, were converted to represent changes from the baseline level, $\Delta F/F(t)$ defined as $(F(t) - F_0(t))/F_0(t)$. The time varying baseline fluorescence, $F_0(t)$, for each cell was a smoothed fluorescence trace obtained after applying a 10-s-order median filter centred at t. Calcium events elicited by action potentials were detected based on a threshold value given by their peak amplitude (5 times the s.d. of the noise value) and their first time derivative (2.5 times the s.d. of the noise value).

The analysis of network bursting was performed on the basis of the population-averaged calcium signal along all of the segmented cells. A peak in the population signal was considered to be a network burst if it met the following criteria: (1) the peak amplitude was greater than 10 times the s.d. of the noise value; (2) a set of bursting cells composed of at least 20% of total cells were active during that population calcium transient; and (3) a cell was considered part of the set of bursting cells only if it participated in at least 50% of the network bursts. Under these criteria, 89.3 ± 14% (range from 60.5% to 100%) and 95.5 ± 6.8% (range from 77.6% to 100%) of segmented cells participated in network bursting in control and mutant organoids, respectively.

The peaks of the network bursts were used to measure the interspike interval (ISI), and the burst frequency was obtained from the average ISI. The burst half-width was also measured from the population-averaged calcium signal by calculating the width of the transient at 50% of the burst peak amplitude.

For analyses of the synchronicity, the $\Delta F/F(t)$ signal was used to calculate the cross-correlation between all pairs of cells at zero lag (Extended Data Fig. 7e) as well as the cross-correlogram between a reference cell and the rest of the cells (Extended Data Fig. 7f). Along with the original signal, we randomly selected ten active cells, circularly shifted their $\Delta F/F(t)$ signal by random phases (keeping their internal temporal structure but altering their temporal relationship with the network) and used them as control.

Multi-electrode array
Extracellular neurophysiological signals were recorded using the Maxwell Biosystems CMOS-HD-MEA system[50] (MaxOne System, MaxWell Biosystems). The MaxOne chip contains 26,400 platinum electrodes in a sensing area of 3.85 × 2.10 mm^2 with 17.5 μm centre-to-centre pitch, 3,265 electrodes per mm^2 density, and 1,024 configurable low-noise readout channels (2.4 μV root mean square (r.m.s.) in the 300 Hz–10 kHz band) with a sampling rate of 20 kHz s^{-1} at 10-bit resolution. Acute recordings were performed at room temperature, with the intact organoid in fresh BrainPhys.

For the recordings, we used MaxLab Live Software (v.20.1.6. MaxWell Biosystems). In brief, spontaneous activity of neurons was measured using the Activity Scan Assay whereby the whole chip area was scanned with a sparse recording (30 s per configuration, seven configurations). Active neurons were automatically identified on the basis of the firing rate and spike amplitude obtained from the Activity Scan. On the basis of the activity of the neurons, the most active electrodes were routed for the creation of the network configuration based on units of 4 × 3 electrodes each, with 1,024 recording electrodes in total (Extended Data Fig. 7d (top)). Selected electrodes were then simultaneously recorded using the network assay to investigate the spontaneous neuronal network activity.

Article

For spike detection, the software used a finite impulse response bandpass filter between 300–3,000 Hz to preprocess the raw data (Extended Data Fig. 7d (middle)). The r.m.s. noise per electrode was calculated and every negative peak larger than 6 r.m.s. was considered to be a spike.

When extracting the waveform of the electrodes in a single unit (set of neighbouring 4 × 3 electrodes; Extended Data Fig. 7d (bottom)), we used the spike time of one selected electrode as a reference to extract the simultaneous signal across the different electrodes (instead of using their individual spike times).

All descriptive statistics and statistical tests were performed in MAT-LAB (v.9.5, R2018b, MathWorks), using the Statistics Toolbox (v.11.4, R2018b, MathWorks). The Lilliefors test was used to test for normality of data distributions. All datasets met the assumptions of the applied statistical tests. When comparing groups, the equality of the variance was tested at the 5% significance level using a two-tailed squared-rank test. All statistical tests applied to the electrophysiological data were two-tailed, with a 5% significance level.

Cell lysis and filter-aided sample preparation digestion for MS

For *SUV420H1*, 4 mutant and 4 control organoids were used; for *CHD8*, 3 mutant and 3 control organoids were used; and, for *ARID1B*, 5 mutant and 4 control organoids were used. Cells were placed into microTUBE-15 (Covaris) microtubes with TPP buffer (truXTRAC Protein Extraction Buffer TP, Covaris, 520103) and lysed using a S220 Focused-ultrasonicator instrument (Covaris) with 125 W power over 180 s at 10% max peak power. After precipitation with chloroform/methanol, extracted proteins were weighed and digested according to the filter-aided sample preparation protocol[51,52] (100 μg for *ARID1B* and *CHD8*; 150 μg for *SUV420H1*). In brief, the 10 kDa filter was washed with 100 μl of 50 mM triethylammonium bicarbonate (TEAB). Each sample was added, centrifuged and the supernatant was discarded. Then, 100 μl of 20 mM Tris (2-carboxyethyl) phosphine at 37 °C was added for 1 h, centrifuged and the supernatant was discarded. While shielding from light, 100 μl of 10 mM IAcNH$_2$ was added for 1 h followed by spinning and discarding the supernatant. Next, 150 μl of 50 mM TEAB + 3 μg of Sequencing Grade Trypsin (Promega) was added to each sample and left overnight at 38 °C. The samples were then centrifuged and the supernatants were collected. Finally, 50 μl of 50 mM TEAB was added to the samples, followed by spinning and supernatant collection. The samples were then transferred to HPLC.

TMT mass tagging protocol peptide labelling

Tandem mass tag (TMT) label reagents (TMTPro, Thermo Fisher Scientific, 16plex Label Reagent Set, A44521) were equilibrated to room temperature and resuspended in anhydrous acetonitrile or ethanol (for the 0.8 mg and 5 mg vials, 41 μl and 256 μl were added, respectively). The reagent was dissolved for 5 min with occasional vortexing. TMT label reagent (41 μl, equivalent to 0.8 mg) was added to each 100–150 μg sample. The reaction was incubated for 1 h at room temperature. The reaction was quenched after adding 8 μl of 5% hydroxylamine to the sample and incubating for 15 min. The samples were combined, dried in a speedvac (Eppendorf) and stored at −80 °C.

Hi-pH separation and MS analysis

Before submission to liquid chromatography with tandem MS (LC–MS/MS), each experiment sample was separated on a Hi-pH column (Thermo Fisher Scientific) according to the vendor's instructions. After separation into 40 (20 for the *ARID1B* experiment) fractions, each fraction was submitted for a single LC–MS/MS experiment performed on a Lumos Tribrid (Thermo Fisher Scientific) system equipped with 3000 Ultima Dual nanoHPLC pump (Thermo Fisher Scientific). The peptides were separated onto a microcapillary trapping column (inner diameter, 150 μm) packed first with approximately 3 cm of C18 Reprosil resin (5 μm, 100 Å, Dr. Maisch) followed by PharmaFluidics

micropack analytical 50 cm column. Separation was achieved by applying a gradient of 5–27% acetonitrile in 0.1% formic acid over 90 min at 200 nl min^{-1}. Electrospray ionization was enabled by applying a voltage of 1.8 kV using a custom-made electrode junction at the end of the microcapillary column and sprayed from stainless-steel tips (PepSep). The Lumos Orbitrap was operated in data-dependent mode for the MS methods. The MS survey scan was performed in the Orbitrap in the range of 400–1,800 *m/z* at a resolution of 6 × 104, followed by the selection of the 20 most intense ions (TOP20) for CID-MS2 fragmentation in the ion trap using a precursor isolation width window of 2 *m/z*, AGC setting of 10,000 and a maximum ion accumulation of 50 ms. Singly charged ion species were not subjected to CID fragmentation. Normalized collision energy was set to 35 V and an activation time of 10 ms. Ions in a 10 ppm *m/z* window around ions selected for MS2 were excluded from further selection for fragmentation for 90 s. The same TOP20 ions were subjected to HCD MS2 events in the Orbitrap part of the instrument. The fragment ion isolation width was set to 0.8 *m/z*, AGC was set to 50,000, the maximum ion time was 150 ms, normalized collision energy was set to 34 V and an activation time of 1 ms for each HCD MS2 scan.

MS data generation

Raw data were submitted for analysis in Proteome Discoverer 2.4 (Thermo Fisher Scientific). Assignment of MS/MS spectra was performed using the Sequest HT algorithm by searching the data against a protein sequence database including all entries from the Human UniProt database[53,54] and other known contaminants such as human keratins and common laboratory contaminants. Sequest HT searches were performed using a 10 ppm precursor ion tolerance and requiring the N/C termini of each peptide to adhere with Trypsin protease specificity, while allowing up to two missed cleavages. 16-plex TMTpro tags on peptide N termini and lysine residues (+304.207 Da) were set as static modifications while methionine oxidation (+15.99492 Da) was set as a variable modification. A MS2 spectra assignment FDR of 1% on the protein level was achieved by applying the target–decoy database search. Filtering was performed using a Percolator (64 bit version)[55]. For quantification, a 0.02 *m/z* window centred on the theoretical *m/z* value of each of the 6 reporter ions and the intensity of the signal closest to the theoretical *m/z* value was recorded. Reporter ion intensities were exported in the result file of Proteome Discoverer 2.4 search engine as Excel tables. The total signal intensity across all peptides quantified was summed for each TMT channel, and all intensity values were normalized to account for potentially uneven TMT labelling and/or sample handling variance for each labelled channel.

MS data analysis

Potential contaminants were filtered out and proteins supported by at least two unique peptides for the *SUV420H1* and *CHD8* experiment and at least one for the *ARID1B* experiment were used for further analysis. We retained proteins that were missing in at most one sample per condition. Data were transformed and normalized using variance stabilizing normalization using the DEP package of Bioconductor[56]. To perform statistical analysis, data were imputed for missing values using random draws from a Gaussian distribution with 0.3 width and a mean that was down-shifted from the sample mean by 1.8. To detect statistically significant differential protein abundance between conditions, we performed a moderated *t*-test using the LIMMA package of Bioconductor[57], using an FDR threshold of 0.1. Gene set enrichment analysis (GSEA) was performed using the GSEA software[58]. GO and KEGG pathway annotation were used to perform functional annotation of the significantly regulated proteins. GO terms and KEGG pathways with FDR-adjusted *q* < 0.05 were considered to be statistically significant.

To build protein interaction networks, we used the prize-collecting Steiner forest algorithm[26,59] using the top 50 DEPs (ranked by adjusted *P* value) from each mutation as terminals, with the absolute value of

their log-transformed fold change as prizes. This algorithm optimizes the network to include high-confidence protein interactions between protein nodes with large prizes. We used the PCSF R package (v.0.99.1)[60] to calculate networks, with the STRING database as a background protein–protein interactome[27], using the parameters $n = 10$, $r = 0.1$, $w = 2$, $b = 40$ and mu = 0.01. As by default in that package, the network was subclustered using the edge-betweenness clustering algorithm from the igraph package, and functional enrichment was performed on each cluster using the ENRICHR API. Cytoscape software (v.3.8.2) was used for network visualization[61]. To assess relationships between the three sets of differential proteins, a protein–protein interaction (PPI)-weighted gene distance (pMM)[62] was calculated between each pair of protein sets. A background distribution was calculated by drawing size-matched random lists of proteins from all of the detected proteins in each dataset and calculating the pMM between these sets. This was repeated 1,000 times, and an empirical P value was calculated by evaluating the number of times randomized pMMs were lower than the value calculated using DEPs.

Dissociation of brain organoids and scRNA-seq

Organoids were dissociated as previously described[41,63]. Volumes of reagents were scaled down 25× for one-month-old organoids. Cells were loaded onto either a Chromium Single Cell B or G Chip (10x Genomics, PN-1000153, PN-1000120), and processed through the Chromium Controller to generate single-cell gel beads in emulsion. scRNA-seq libraries were generated using the Chromium Single Cell 3′ Library & Gel Bead Kit v3 or v3.1 (10x Genomics, PN-1000075, PN-1000121), with the exception of a few libraries in the earlier experiments that were prepared using the v2 kit (10x Genomics, PN-120237). Information on the estimated number of cells loaded and the version of kit used is provided in Supplementary Table 10. Libraries were pooled from different samples based on molar concentrations and sequenced them on a NextSeq 500 or NovaSeq instrument (Illumina) with 28 bases for read 1 (26 bases for v2 libraries), 55 bases for read 2 (57 bases for v2 libraries) and 8 bases for index 1. If necessary, after the first round of sequencing, libraries were repooled on the basis of the actual number of cells in each and resequenced with the goal of producing an approximately equal number of reads per cell for each sample.

scRNA-seq data analysis

Reads from scRNA-seq were aligned to the GRCh38 human reference genome and the cell-by-gene count matrices were produced using the Cell Ranger pipeline (10x Genomics)[64]. Cell Ranger v.2.0.1 was used for experiments using the GM08330 control 'single cell map' and for HUES66 *CHD8*-mutant and control organoids at 3.5 months, batch I, while v.3.0.2 was used for all of the other experiments. The default parameters were used, except for the '--cells' argument. Data were analysed using the Seurat R package v.3.1.5[65] using R v.3.6. Cells expressing a minimum of 500 genes were retained, and UMI counts were normalized for each cell by the total expression multiplied by 10^6 and log-transformed. Variable genes were found using the mean.var.plot method, and the ScaleData function was used to regress out variation due to differences in total UMIs per cell. Principal component analysis (PCA) was performed on the scaled data for the variable genes, and the top principal components were chosen based on Seurat's ElbowPlots (at least 15 PCs were used in all cases). Cells were clustered in PCA space using Seurat's FindNeighbors on top principal components, followed by FindClusters with resolution = 1.0 (in brief, a 20-nearest-neighbor graph was constructed and modularity optimization using the Louvain algorithm was performed to identify clusters). Variation in the cells was visualized by t-SNE analysis of the top principal components.

In the case of the GM08330 1 month organoids (single-cell map), cells were demultiplexed using genotype clustering from cells from a different experiment that were sequenced in the same lane. To demultiplex, SNPs were called from CellRanger BAM files using the cellSNP tool v.0.1.5, and then the vireo function was used with the default parameters and n_donor = 2, from the cardelino R library (v.0.4.0)[66,67] to assign cells to each genotype.

In two cases, one organoid was excluded from the analysis as outliers. See the 'Statistics and reproducibility' section for details.

For each dataset, upregulated genes in each cluster were identified using the VeniceMarker tool from the Signac package v.0.0.7 from Bio-Turing (https://github.com/bioturing/signac). Cell types were assigned to each cluster by looking at the top most significant upregulated genes. In a few cases, clusters were further subclustered to assign identities at higher resolution. At 1 month, the excitatory projection neurons included a gradient of immature neurons, which were split into two clusters: we labelled the cluster representing the earlier developmental stage 'newborn deep-layer projection neurons' and the cluster representing the later stage 'immature deep-layer projection neurons'. At 3 months and beyond, excitatory projection neuron clusters could be identified as deep-layer corticofugal neurons and upper-layer callosal projection neurons. For the GABAergic populations, 1 month organoids included neurons expressing broad markers of GABAergic identity (labelled GABAergic neurons), progenitor cells expressing markers of GABAergic lineage identity (GABAergic neuron progenitors) and progenitor cells with high expression of cell cycle markers in addition to the progenitor identity markers (cycling GABAergic neuron progenitors). At 3 months and beyond, GABAergic neurons expressed more specific markers of cortical interneurons (thus labelled GABAergic interneurons), and GABAergic lineage progenitors at these ages were divided into 'GABAergic interneuron progenitors' and 'cycling GABAergic interneuron progenitors' on the basis of the level of expression of cell cycle markers.

To assess gene expression of ASD risk genes in GM08330 and Mito210 control organoids across timepoints, datasets from 1, 3 and 6 months were merged using Seurat v.3.1.5, and then batch-corrected using Harmony v.1.0 with the default parameters[68]. As the 1 month data are dominated by cell cycle signal, the ScaleData function was used to regress out variation due to both total UMI count per cell and cell cycle stage differences, calculated using Seurat's CellCycleScore. Variation was visualized using t-SNE on the first 30 harmony dimensions. Broad cell types were assigned as described above, and mutual information was calculated between cell type assignments and individual organoids using the mpmi R package[69]. Expression of the 102 ASD risk genes identified in the Satterstrom et. al.[6] study was evaluated using Seurat's AddModuleScore function using the default parameters. This function calculates the average expression level per cell of the set of genes (based on log-normalized, unscaled data), and then subtracts the average expression of a randomly selected expression-matched control set of genes. A resulting score of greater than zero indicates that the ASD risk gene set is expressed more highly in that cell than would be expected, given the average expression of the gene set across the dataset.

To compare cell type proportions between control and mutant organoids, for each cell type present in a dataset, the glmer function from the R package lme4 (v.1.1-23)[70] was used to estimate a mixed-effect logistic regression model[71]. The output was a binary indicator of whether cells belong to this cell type, the control or mutant state of the cell was a fixed predictor, and the organoid that the cell belonged to was a random intercept. Another model was fit without the control-versus-mutant predictor, and the ANOVA function was used to compare the two model fits. P values for each cell type were then adjusted for multiple-hypothesis testing using Benjamini–Hochberg correction.

Pseudotime, gene module and differential expression analysis

Pseudotime analysis was performed using the Monocle3 v.0.2.0 software package[72] with the default parameters. The cells were first subset to contain an equal amount from control and mutant. A starting point for the trajectory was chosen manually by finding an endpoint of the tree located in the earliest developmental cell type (generally, cycling progenitors). In cases in which the cells were split into more than one partition, the starting point was chosen within the partition of interest,

Article

and a new UMAP was calculated using just these cells. To test whether mutant trajectories were accelerated compared with the control, a one-sided Kolmogorov–Smirnov test was applied comparing the distribution of psuedotime values of control versus mutant cells, using the stats R package.

To learn patterns of coordinated gene regulation across the cells, we applied WGCNA[19] to each dataset. In cases in which cells were split into partitions in the above pseudotime analysis, only cells belonging to the partition of interest were used. Normalized gene expression data were further filtered to remove outlying genes, mitochondrial and ribosomal genes. Outliers were identified by setting the upper (>9) and lower (<0.15) thresholds to the average normalized expression per gene. After processing, blockwiseModules function from the WGCNA v.1.69 library was performed in R with the parameters networkType="signed", minModuleSize=4, corType="Bicor", maxPOutliers=0.1, deepSplit=3, trapErrors=T and randomSeed=59069. Other than power, the remaining parameters were left as the default setting. To pick an adequate power for each dataset, we used the pickSoftThreshold function from WGCNA to test values from 1 to 30. The final resolution was determined by choosing the resolution that captured most variation in the fewest total number of modules– this resulted in a power of 3 for *SUV420H1* 35 d.i.v., 9 for *ARID1B* 35 d.i.v. and 12 for *CHD8* 109 d.i.v.

To calculate differential expression of modules, Seurat objects were downsampled to have an equal number of cells per organoid, and then the AddModuleScore function was used, using gene lists from WGCNA results. For each module, linear mixed-effect models were fit to the data, with the modules scores as the output, the organoid the cell belongs to as a random intercept, and with or without the control-versus-mutant state as a predictor. The ANOVA function was used to compare the models, and *P* values were then adjusted across modules using Benjamini–Hochberg correction.

DEGs between control and mutant organoids were assessed after datasets were subset to the cells from the partition of interest in the above pseudotime analysis, to the cells from each individual cell type, or not subset at all for pseudobulk analysis. Reads were then summed across cells in each organoid. Genes with less than 10 total reads were excluded, and DESeq2 (ref. [73]) was used to calculate DEGs, with each organoid as a sample[74]. The clusterProfiler[75] R package was used to find enriched biological processes in these gene sets, with the enrichGO function and the compareCluster function to highlight processes the gene sets might have in common.

Single-nucleus isolation and single-cell ATAC-seq

Nuclei from 1 month and 3 month organoids were extracted with two types of procedures according to their size differences. For the 1 month organoids, nuclei were extracted according to a protocol provided by 10x Genomics[76] to minimize material loss, while a sucrose-based nucleus isolation protocol[77] was used for the 3 month organoids to better remove debris. Single-nucleus ATAC-seq libraries were prepared using the Chromium Single Cell ATAC Library & Gel Bead v1 Kit (10x Genomics, PN-1000110) and around 15,300 nuclei per channel were loaded to give an estimated recovery of 10,000 nuclei per channel. Libraries from different samples were pooled on the basis of molar concentrations and sequenced with 1% PhiX spike-in on the NextSeq 500 instrument (Illumina) with 33 bases each for read 1 and read 2, 8 bases for index 1 and 16 bases for index 2.

Single-cell ATAC-seq data analysis

Reads from scATAC-seq were aligned to the GRCh38 human reference genome and the cell-by-peak count matrices were produced using the Cell Ranger ATAC pipeline v.2.0.0 (10x Genomics) with the default parameters. Data were analysed using the Signac R package (v.1.2.1)[78] using R v.4.0. Annotations from the EnsDb.Hsapiens.v86 package[79] were added to the object. After consideration of the quality control metrics recommended in that package, cells with 1,500–20,000 fragments in

peak regions, at least 35% of reads in peaks, a nucleosome signal of less than 4 and a TSS enrichment score of greater than 2 were retained for further analysis. Latent semantic indexing (LSI) was performed to reduce data dimensionality (counts were normalized using term frequency inverse document frequency, all features were set as top features, and singular value decomposition was performed). The top LSI component was discarded as it correlated strongly with sequencing depth, and components 2–30 were used for downstream analysis. Cells were clustered using Seurat's FindNeighbors, followed by FindClusters with the SLM algorithm (a 20-nearest-neighbor graph was constructed and modularity optimization using the smart local moving algorithm was performed to identify clusters). Variation in the cells was visualized using UMAP analysis of the top LSI components.

scATAC-seq data were integrated with scRNA-seq data from the corresponding Mito210 dataset for each timepoint, using Seurat's TransferData to predict cell type labels for the ATAC profiles. Concurrently, differentially accessible (DA) peaks per cluster were called using FindMarkers using the logistic regression framework with the number of fragments in peak regions as a latent variable. These DA peaks were mapped to the closest genes. The top genes per cluster were used to confirm and refine cluster cell type assignments from those based on transferring RNA labels.

DA peaks between control and *SUV420H1*-mutant organoids were calculated per cell type, using the same method as described above. We noticed that most cell types had very few significantly DARs (range 6–34, except for apical radial glia cells, the most prevalent and, therefore, the most powered cell type at this time point, which had 515 DARs), and that the DARs were almost entirely overlapping in all cell types. Therefore, DARs were calculated using all cells together to improve power. DARs were visualized using Signac's CoveragePlot function with the default parameters.

To find transcription factor motifs enriched in DARs, the top 400 up- and downregulated peaks for each time point differentially accessible peaks were supplied to the HOMER software (v.4.11.1)[80], using a 300 bp fragment size and masking repeats. In the case of upregulated regions in 3 month mutant organoids, only 341 regions were supplied, as that was the total number of regions with log[FC] > 0.1 and *P* > 0.1. The top 5 de novo motifs per cell type found by HOMER with $P \leq 10^{-10}$ are reported, along with all TFs of which the known binding sites match that motif with a score of ≥0.59.

Statistics and reproducibility

Organoid size analysis. Information about the number of organoids used is provided in Supplementary Table 2. In summary, for *SUV420H1*: *n* = 132 for total control organoids, *n* = 132 for total mutant organoids, from 6 experimental batches. For *ARID1B*: *n* = 109 for total control organoids, *n* = 122 for total mutant organoids, from 4 experimental batches. For *CHD8*: *n* = 472 for total control organoids, *n* = 482 for total mutant organoids, from 7 experimental batches. *P* values were calculated using two-sided *t*-tests and then adjusted using Bonferroni correction.

Proteomic analysis. Four mutant and four control organoids were used for *SUV420H1*. Three mutant and three control, and five mutant and four control organoids were used for *CHD8* and *ARID1B*, respectively. To detect statistically significant differential protein abundance between conditions, moderated *t*-tests were performed as described in 'MS data analysis' (FDR threshold of 0.1; Extended Data Fig. 12a–c). GO terms and KEGG pathways were calculated using the GSEA software (Extended Data Fig. 12d–f) and FDR-adjusted *q* < 0.05 was considered to be statistically significant. For each pair of protein set distances between pairs of DEP sets (Extended Data Fig. 12h, i), a PPI-weighted protein set distance was calculated between all significant DEPs (FDR < 0.1). To determine whether this distance was smaller than would be expected by chance, size-matched sets were randomly chosen from the proteins detected in each experiment, and the distance between these random sets was

calculated 1,000 times per pair. *P* values were assigned by counting the fractions of times that this random distance was less than the actual distance value between differential sets.

scATAC-seq analysis. Detailed information is provided in Supplementary Table 10. In summary, three *SUV420H1* mutant and three control organoids were used for each of the 1 month and 3 month timepoints, for a total of twelve individually sequenced organoids. The total number of cells sequenced was 45,988.

scRNA-seq analysis. Detailed information is provided in Supplementary Table 10. In summary, in each dataset, three individual organoids per genotype were profiled. In two cases, one organoid was excluded from the analysis as an outlier; in PGP1 *SUV420H1* organoids at 1 month, a mutant organoid was excluded due to very low average nUMI and nGene in that sequencing lane, and in the HUES66 *CHD8* organoids at 3.5 months batch II, a mutant organoid was excluded because it contained mostly interneuron lineage cells, with very few projection neuron cells. Although an increase in interneuron-lineage cells was seen in all mutant organoids, this organoid was excluded to be conservative. This left a total of 112 single organoids that passed quality control and were considered in downstream analysis, with a total of 749,370 cells. Adjusted *P* values for differences in cell type proportions between control and mutant organoids (Figs. 1a–c, 2a, b and Fig. 3a, b and Extended Data Figs. 4c–f, 5a–c; 8b, c, e, g, 9a, b, e and 10b–d) were based on logistic mixed models (see the 'scRNA-seq data analysis' section). Adjusted *P* values for differences in the distribution of module scores between control and mutants (Figs. 1f, 2e and 3e and Extended Data Figs. 5e, 8i and 9h) were based on linear mixed models (see the 'Pseudotime, gene module and differential expression analysis' section). In Fig. 4, for each comparison of two gene lists, the circles inside the box are coloured and sized according to the significance of the number of overlapping genes in those two lists, reported as the Bonferroni-adjusted *P* value determined using a hypergeometric test.

Bulk RNA-seq analysis. Three organoids were sequenced per genotype for a total of 30 individual organoids.

Calcium imaging analysis. Five organoids were analysed per genotype. Spontaneous activity was recorded in three different *z* planes (120 ± 803 neurons per plane (range from 25 to 294 neurons per plane) in control organoids, and 107 ± 75 neurons per plane (range from 32 to 255 neurons per plane) in *SUV420H1*$^{+/-}$ organoids). *P* values were calculated from two-tailed t-tests (Fig. 1h, i). *P* values for cumulative frequency distribution (Extended Data Fig. 7j) of ISI for control and *SUV420H1*$^{+/-}$ organoids were determined using two-sided Kolmogorov–Smirnov tests. Representative images in Fig. 1g and Extended Data Fig. 7a show one control organoid out of five control and five *SUV420H1*$^{+/-}$ organoids.

Immunohistochemistry. At least three organoids of each condition were used for verifying the expression of the indicated markers in Extended Data Figs. 1a–c, 3g, 4a, b, 8a, d, f and 9c, d, f.

Western blotting. Each control and mutant protein lysate was blotted at least twice in Extended Data Fig. 3d–f.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability
Read-level data from scRNA-seq and scATAC-seq, along with proteomics data, supporting the findings of this study have been deposited in a controlled access repository at https://www.synapse.org with accession number project ID syn26346373 Count-level data and metadata have been deposited at the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1129). The electrophysiology materials and data are available from the corresponding authors on request. Public data used in this paper include the GRCh38 human reference genome and the EnsDb.Hsapiens.v86 annotation package.

Code availability
The code used for data analysis is available at GitHub (https://github.com/AmandaKedaigle/mutated-brain-organoids).

39. Chen, A. E. et al. Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* **4**, 103–106 (2009).
40. Church, G. M. The personal genome project. *Mol. Syst. Biol.* **1**, 2005.0030 (2005).
41. Velasco, S., Paulsen, B. & Arlotta, P. Highly reproducible human brain organoids recapitulate cerebral cortex cellular diversity. *Protoc. Exchange* https://doi.org/10.21203/rs.2.9542/v1 (2019).
42. Lovell-Badge, R. et al. ISSCR guidelines for stem cell research and clinical translation: the 2021 update. *Stem Cell Rep.* **16**, 1398–1408 (2021).
43. Doench, J. G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
44. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
45. Mangeot, P. E. et al. Genome editing in primary cells and in vivo using viral-derived nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun.* **10**, 45 (2019).
46. Park, Y.-G. et al. Protection of tissue physicochemical properties using polyfunctional crosslinkers. *Nat. Biotechnol.* **37**, 73–83 (2019).
47. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
48. Ohgane, K. Quantification of gel bands by an Image J macro, band/peak quantification tool. *protocols.io* https://doi.org/10.17504/protocols.io.7vghn3w (2019).
49. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at *bioRxiv* https://doi.org/10.1101/061507 (2017).
50. Müller, J. et al. High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels. *Lab Chip* **15**, 2767–2780 (2015).
51. Ostasiewicz, P., Zielinska, D. F., Mann, M. & Wiśniewski, J. R. Proteome, phosphoproteome, and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin embedded tissue and analyzable by high-resolution mass spectrometry. *J. Proteome Res.* **9**, 3688–3700 (2010).
52. Wiśniewski, J. R. Quantitative evaluation of filter aided sample preparation (FASP) and multienzyme digestion FASP protocols. *Anal. Chem.* **88**, 5438–5443 (2016).
53. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **27**, 49–54 (1999).
54. Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2018).
55. Käll, L., Storey, J. D., MacCoss, M. J. & Noble, W. S. Posterior error probabilities and false discovery rates: two sides of the same coin. *J. Proteome Res.* **7**, 40–44 (2008).
56. Zhang, X. et al. Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat. Protoc.* **13**, 530–550 (2018).
57. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
58. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
59. Tuncbag, N. et al. Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. *J. Comput. Biol.* **20**, 124–136 (2013).
60. Akhmedov, M. et al. PCSF: an R-package for network-based interpretation of high-throughput data. *PLoS Comput. Biol.* **13**, e1005694 (2017).
61. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
62. Yoon, S. et al. GScluster: network-weighted gene-set clustering analysis. *BMC Genom.* **20**, 352 (2019).
63. Quadrato, G., Sherwood, J. L. & Arlotta, P. Long term culture and electrophysiological characterization of human brain organoids. *Protoc. Exchange* https://doi.org/10.1038/protex.2017.049 (2017).
64. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
65. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
66. McCarthy, D. J. et al. Cardelino: computational integration of somatic clonal substructure and single-cell transcriptomes. *Nat. Methods* **17**, 414–421 (2020).
67. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
68. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
69. Pardy, C. mpmi: mixed-pair mutual information estimators (2020).
70. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
71. Fonseka, C. Y. et al. Mixed-effects association of single cells identifies an expanded effector CD4$^+$ T cell subset in rheumatoid arthritis. *Sci. Transl. Med.* **10**, eaaq0305 (2018).

Article

72. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
73. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
74. Lun, A. T. L. & Marioni, J. C. Overcoming confounding plate effects in differential expression analyses of single-cell RNA-seq data. *Biostatistics* **18**, 451–464 (2017).
75. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
76. *Nuclei Isolation from Mouse Brain Tissue for Single Cell ATAC Sequencing* Rev B (10x Genomics, 2019).
77. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
78. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
79. Rainer, J. EnsDb.Hsapiens.v86: ensembl based annotation package (2017).
80. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

Author contributions P.A., B.P., S.V., A.J.K., M.P., and G.Q. conceived the experiments. A.J.D. designed *SUV420H1* and *ARID1B* gRNAs and generated the Mito210 and Mito294 *SUV420H1*-edited lines with B.P.; L.B. generated the *ARID1B*-edited lines. N.E.S. and X.S. designed *CHD8* gRNAs and generated the HUES66 *CHD8* line. A.J.D. and B.P. generated the Mito210 and Mito294 *CHD8*-edited lines. S.V., B.P., M.P., R.S., C.A., A.T and S.N.S. generated, cultured and characterized all of the organoids used in this study and P.A. supervised their work. X.A. performed scRNA-seq experiments with help from B.P., S.V., M.P., R.S. and G.Q. under the supervision of P.A. and J.Z.L.; A.J.K., K.K., S.K.S. and J.Z.L. performed scRNA-seq analysis and J.Z.L. and A.R. supervised their work. S.V., B.P., M.P., R.S., A.U., G.Q. and A.J.K. worked on cell type assignments and data analysis. K.T., M.P. and A.J.K. performed proteomic analysis, supervised by K.L.; S.M.Y., P.S. and A.P. performed the calcium imaging experiments and analysis, supervised by E.S.B. and P.A.; S.M.Y. and R.S. performed the MEA recordings and analysis supervised by P.A.. A.A. performed whole-organoid imaging under the supervision of K.C. P.A., B.P., S.V., A.J.K. and M.P. wrote the manuscript with contributions from all of the authors. All of the authors read and approved the final manuscript.

Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | Cortical organoids cultured for one, three and six months generate the cellular diversity of the human cerebral cortex with high organoid-to-organoid reproducibility. a, scRNA-seq and immunohistochemistry analysis of organoids cultured for one month (32 d.i.v.), three months (98 d.i.v.), and six months (190 d.i.v.). Left, *t*-SNE plots (n = 3 organoids per timepoint, co-clustered). Cells are coloured by cell-type. Right, immunohistochemistry for specific markers. Neural progenitor marker SOX2 (magenta) and postmitotic neuronal marker TBR1 (green) are shown at one month. CPN marker SATB2 (magenta) and CFuPN marker CTIP2 (green) are shown at three months. The astroglia markers S100B (magenta) and GFAP (green) are shown at six months. Below, schematic images of brain organoids in each timepoint. Scale bars are 100 μm. **b**, Immunohistochemistry for neuronal (MAP2), dorsal forebrain neural progenitor (EMX1, SOX2), CFuPN (CTIP2), and CPN (SATB2) markers in GM08330 organoids at one, three, and six months. Scale bars: whole organoids (leftmost column), 200 μm; others, 50 μm. **c**, Immunohistochemistry for cell-type specific markers in Mito210 organoids, as in **b**. **d**, Top, *t*-SNE plots of the scRNA-seq data from individual replicates from three organoids at one month, three organoids at three months, and three organoids at six months from the GM08330 cell line shown in **b**. Bottom, bar charts showing the cell-type composition of each individual organoid. On top of the bar charts, mutual information (MI) scores between cell-type proportions and organoid identities are displayed. A MI score of 0 would indicate identical cell type proportions between organoids, while a score of 1 would indicate completely divergent profiles. In previous work, MI scores for endogenous brain datasets were reported to range from 0.008 to 0.064[9]. **e**, scRNA-seq data of organoids from the Mito210 cell line at one month (35 d.i.v.), three months (92 d.i.v.), and six months (178 d.i.v.), as in **d**. Organoids for the one and three month timepoints are the same as the control organoids in Extended Data Fig. 4f and Extended Data Fig. 5b. **f**, Expression of selected marker genes used in cell-type identification. Violin plots show distribution of normalized expression in cells from GM08330 organoids at one, three and six months (n = 3 individual organoids per timepoint). **g**, Expression of marker genes in Mito210 organoids, as in **f**. Number of organoids used for each analysis can be found in the Methods under "Statistics and reproducibility". aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons.

GM08330 iPSC line

Cell Types
- Cycling Progenitor
- RG
- Cortical Hem
- Subcortical
- IPC
- EN
- GABA IN

a Cell Types — harmonized data GM08330 — 58,568 cells

Timepoints — 1 month, 3 months, 6 months — 9 organoids

b ASD Gene Set — Enrichment compared to random gene sets

c CHD8, ARID1B, SUV420H1, SCN1A, GRIN2B, STXBP1, RELN, PTEN, TCF4, POGZ, FOXP1, ANK2

d Timepoint / Cell Types heatmap

Mito210 iPSC line

Cell Types
- Cycling Progenitor
- RG
- Cortical Hem
- Subcortical
- IPC
- EN
- Astroglia
- GABA IN

e Cell Types — harmonized data Mito210 — 50,952 cells

Timepoints — 1 month, 3 months, 6 months — 9 organoids

f ASD Gene Set — Enrichment compared to random gene sets

g CHD8, ARID1B, SUV420H1, SCN1A, GRIN2B, STXBP1, RELN, PTEN, TCF4, POGZ, FOXP1, ANK2

h Timepoint / Cell Types heatmap

Extended Data Fig. 2 | See next page for caption.

Article

Extended Data Fig. 2 | Expression of selected ASD risk genes in cortical organoids cultured for one, three, and six months. a, *t*-SNE plots of 58,568 cells from nine organoids from the GM08330 cell line, shown in Extended Data Fig. 1d, after Harmony batch correction. Cells are coloured according to cell-type (left) and timepoint (right). **b**, Gene set expression scores for a set of 102 genes associated with ASD risk[6] across cell-types, in cells from **a**. Scores above 0 indicate enriched expression over similar sets of randomly chosen genes. **c**, *t*-SNE plots showing normalized expression of selected ASD risk genes in cells from **a**. **d**, Average expression of 102 genes associated with ASD risk across cell-types and timepoints in the GM08330 cell line. **e**, *t*-SNE plots of nine organoids from the Mito210 cell line, shown in Extended Data Fig. 1e, after Harmony batch correction. Cells are coloured according to cell-type (left) or timepoint (right). **f**, Gene set scores for the set of ASD risk genes as in **b**, in cells from **e**. Scores above 0 indicate higher expression than similar modules of randomly chosen genes. **g**, *t*-SNE plots showing normalized expression of selected ASD risk genes in cells from **e**. **h**, Expression of 102 genes associated with ASD risk across cell-types and timepoints in Mito210 cell line. RG, radial glia (aRG, oRG, and oRG/Astroglia), IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; EN, Excitatory neurons (CPN, CFuPN and PN); GABA IN, GABAergic interneurons.

SUV420H1

a

SUV420H1 — N-domain, SET, Post-SET

Name of the line	Type of line	Sex	InDel	Stop Codon
Mito 210	hiPSC	♂	10bp del	90 aa
PGP1	hiPSC	♂	1bp in	90 aa
Mito 294	hiPSC	♂	8bp del	87 aa

ARID1B

b

ARID1B — ARID, BAF250C

Name of the line	Type of line	Sex	InDel	Stop Codon
Mito 210	hiPSC	♂	11bp del + 10 bp in	977 aa
Mito 294	hiPSC	♂	22bp del	984 aa

CHD8

c

CHD8 — Chromo, DEX, HELC, BRK

Name of the line	Type of line	Sex	InDel	Stop Codon
HUES66	hESC	♀	13bp del	1248 aa
H1	hESC	♂	1bp in	1199 aa
GM08330	hiPSC	♂	1bp in	1199 aa
Mito294	hiPSC	♂	17bp del	1211 aa
Mito210	hiPSC	♂	7bp del	1250 aa

d Mito294 Mito210 PGP1

Control / SUV420H1+/-

50 kDa — SUV420H1 *
15 kDa — H4K20me3
15 kDa — H4 total
37 kDa — GAPDH

● normalized to GAPDH

SUV420H1 / housekeeping gene — Mito294 Mito210 PGP1
Control / SUV420H1+/-

● normalized to H4 total

H4K20me3 / housekeeping gene — Mito210 PGP1
Control / SUV420H1+/-

e Mito294 Mito210 Mito294 Mito210

Control / ARID1B+/-

250 kDa — ARID1B **
longer exposure time

37 kDa — GAPDH

● normalized to GAPDH

ARID1B / housekeeping gene — Mito210
Control / ARID1B+/-

f GM08330 H1 Mito210 HUES66 Mito294

Control / CHD8+/-

250 kDa — CHD8 *
50 kDa — β-Actin
37 kDa — GAPDH

● normalized to β-actin   ● normalized to GAPDH

CHD8 / housekeeping gene — GM08330 H1 Mito210 HUES66 Mito294
Control / CHD8+/-

g

Mito210, 35 d.i.v. — Control / SUV420H1+/-
DAPI, MAP2, EMX1, SOX2, CTIP2

Mito210, 35 d.i.v. — Control / ARID1B+/-
DAPI, MAP2, EMX1, SOX2, CTIP2

HUES66, 35 d.i.v. — Control / CHD8+/-
DAPI, MAP2, EMX1, SOX2, CTIP2

h Mito294 Mito210 PGP1

Control / SUV420H1+/-
b.I, b.II, b.III, b.IV

Organoid size (fold change)

*p=5.2e-13 *p=9.6e-06 *p=0.0042 *p=2.3e-05 *p=3.4e-04 *p=0.029

18 29 18 27 17 30 d.i.v.

i Mito294 Mito210

Control / ARID1B+/-
b.I, b.II, b.III

Organoid size (fold change)

*p=0.033 *p=0.0004 *p<8.8e-16 *p=1.2e-12

15 35 18 35 d.i.v.

j GM08330 H1 HUES66 Mito294

Control / CHD8+/-
b.I, b.II

Organoid size (fold change)

p=0.82 p=1.0 p=1.0 p=0.27 p=1.0 *p=1.8e-04 *p=0.008 *p=1.2e-06
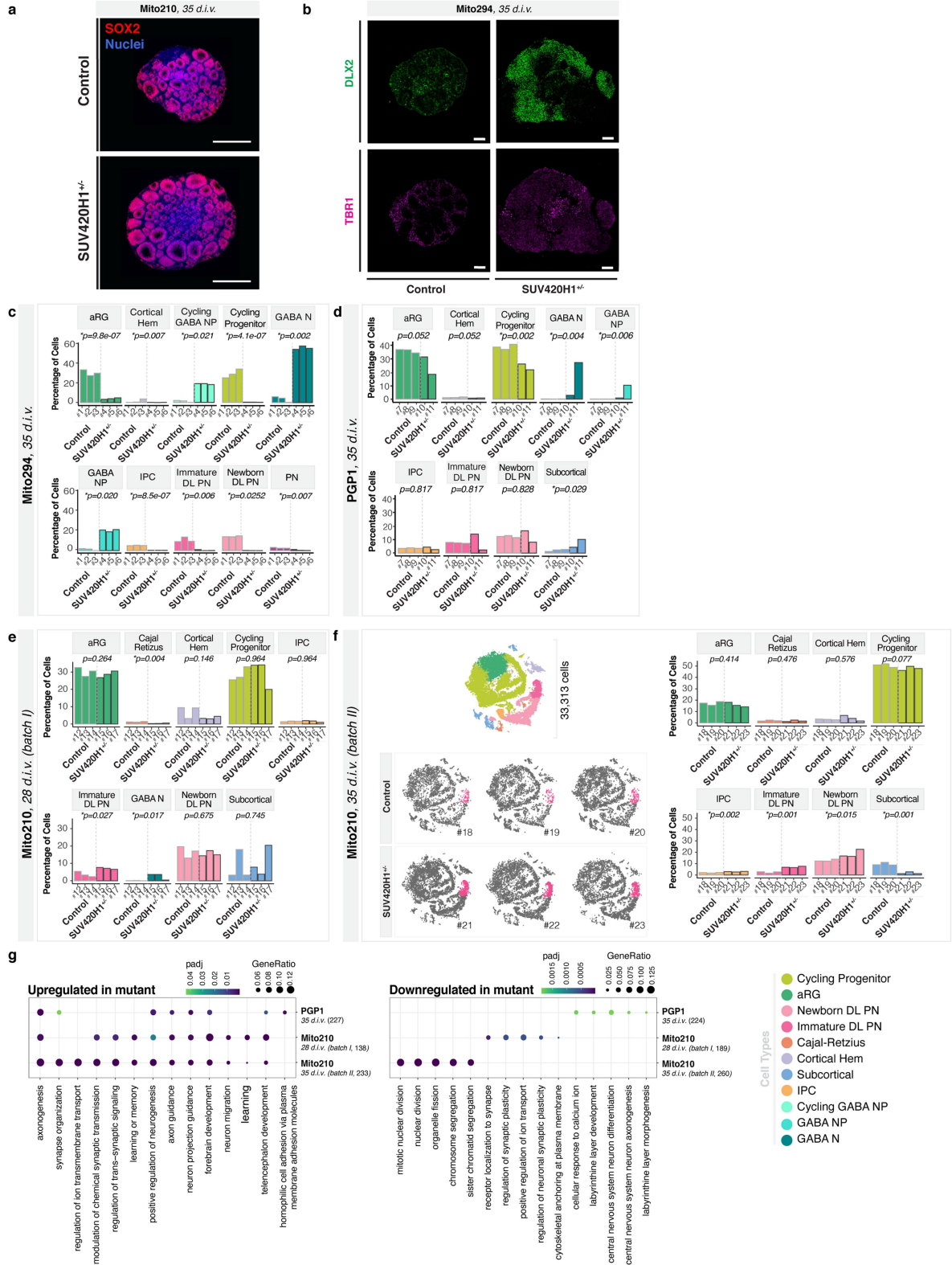
16 25 15 24 16 25 15 25 d.i.v.

Extended Data Fig. 3 | See next page for caption.

Article

Extended Data Fig. 3 | Generation and characterization of *SUV420H1*, *ARID1B*, and *CHD8* mutant organoids. a, Protein domain structure of SUV420H1. Arrow indicates the region (N-domain) mutated in the Mito210, PGP1 and Mito294 parental lines (bottom). **b**, Protein domain structure of ARID1B. Arrow indicates the region before the ARID domain mutated in the Mito210 and Mito294 parental lines (bottom). **c**, Protein domain structure of CHD8. Arrows indicates the helicase C-terminal (HELC) domain mutated in the HUES66, H1, GM08330, Mito294 and Mito210 lines (bottom). **d**–**f**, Western blot analysis showing presence of SUV420H1 (**d**), ARID1B (**e**) and CHD8 (**f**) protein expression in control lines, and its reduction in the mutant lines. Molecular weight in kDa is shown on the left of the gel. H4K20me3, a hallmark of SUV420H1 activity, and total levels of histone H4 were also detected in control and in *SUV420H1*⁺/⁻ lines (**d**). ARID1B was not detectable in Mito294 even with a longer exposure of the blotted membrane (**e**, right). Asterisks indicate the bands used for quantification. Bottom, protein levels in control and mutant lines were quantified and normalized for housekeeping genes β-Actin or GAPDH. For gel source data, see Supplementary Fig. 1. **g**, Immunohistochemistry for neuronal (MAP2), dorsal forebrain neural progenitor (EMX1, SOX2) and CFuPN (CTIP2) markers in organoids at 35 d.i.v. derived from the Mito210 *SUV420H1*⁺/⁻, Mito210 *ARID1B*⁺/⁻ and HUES66 *CHD8*⁺/⁻ and isogenic control cell lines. Scale bar, 300 μm. **h**–**j**, Size quantification of control and *SUV420H1*⁺/⁻ (**h**), *ARID1B*⁺/⁻ (**i**) and *CHD8*⁺/⁻ (**j**) organoids across lines and at different timepoints. The ratio of organoid size compared to the average of control organoids in each batch is plotted. Differentiation batch (b.) is indicated by the shade of the points. Lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and middle hinge is the median (50th). Both whiskers extends from the hinge to the largest or smallest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). *P*-values from a two-sided t-test, after Bonferroni adjustment within each mutation. Number of organoids and differentiations used for the measurement are summarized in Supplementary Table 2 and in the Methods under "Statistics and reproducibility".
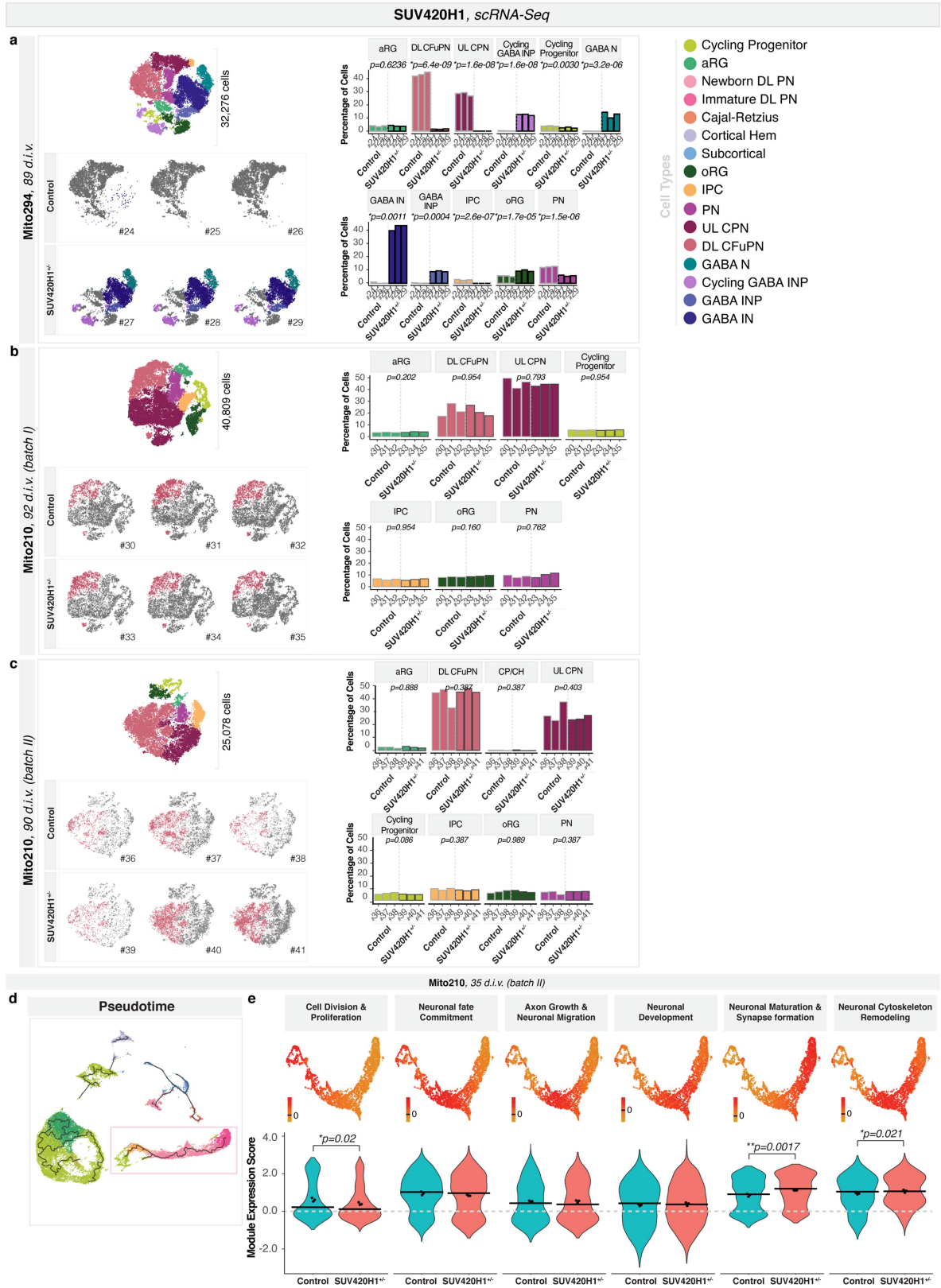
Extended Data Fig. 4 | See next page for caption.

Article

Extended Data Fig. 4 | Cell-type composition of _SUV420H1_[+/−] and isogenic control organoids. a, Immunohistochemistry of Mito210 _SUV420H1_[+/−] and control organoids cultured for one month (35 d.i.v.). Optical section from the middle of whole-organoid dataset. Scale bars are 500 µm. SOX2, a marker of neuronal progenitors, is shown in red, and nuclei (Syto16) are shown in blue. **b**, Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 and GABAergic marker DLX2 in Mito294 control and _SUV420H1_[+/−] organoids at one month (35 d.i.v.). Scale bars: 200 µm. **c−e**, scRNA-seq data from one month (Mito294 35 _d.i.v._ (**c**), PGP1 35 d.i.v. (**d**) and Mito210 28 d.i.v., batch I (**e**)) control and _SUV420H1_[+/−] organoids. Bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted _p_-values for a difference in cell type proportions between control and mutant, based on logistic mixed models (see Methods) are shown. **f**, scRNA-seq data from Mito210 35 d.i.v. (batch II) control and _SUV420H1_[+/−] organoids. Left top shows combined _t_-SNE plots of control and mutant organoids (n = 3 single organoids per genotype). Cells are coloured by cell-type, and the total number of cells per plot is indicated. Left bottom, _t_-SNE plots for control and mutant

individual organoids. Immature deep-layer projection neuron populations are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as in **c−e**. **g**, Enriched GO terms for genes upregulated and downregulated in _SUV420H1_[+/−] vs. control across lines. Genes were calculated using cells from the partition of interest. The top 5 most significant terms per dataset are shown. Size of dot indicates the proportion of genes belonging to each term found in the list of dysregulated genes ("GeneRatio"). Colour indicates enrichment adjusted _p_-value. Numbers in parentheses along the y axis indicate the number of DEGs in that dataset. As control, we calculated GO term enrichment for 100 random gene sets of the same size sampled from genes expressed in each dataset, and found no significant enrichment of these terms (see Methods). Number of organoids used for each analysis can be found in the Methods under "Statistics and reproducibility". aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA N, GABAergic neurons.

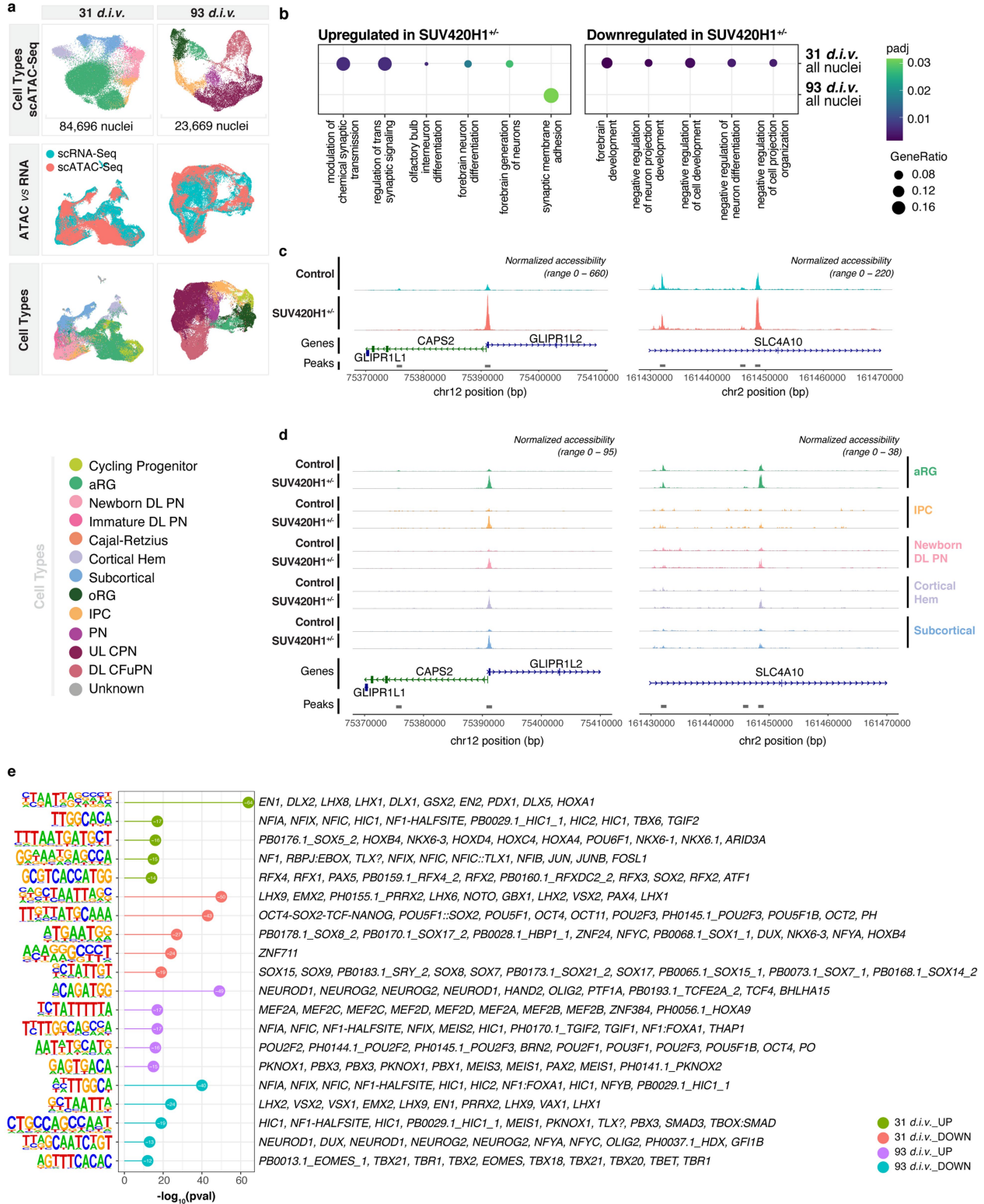**Extended Data Fig. 5** | See next page for caption.

Article

Extended Data Fig. 5 | Cell-type composition, full pseudotime trajectories, and gene modules in *SUV420H1*$^{+/-}$ and isogenic control organoids. a–c, scRNA-seq data from three month Mito294 89 d.i.v. (**a**), Mito210 92 d.i.v. batch I (**b**), and 90 d.i.v. batch II (**c**) control and SUV420H1$^{+/-}$ organoids. Left top shows combined *t*-SNE plots of control and mutant organoids (n = 3 single organoids per genotype). Cells are coloured by cell type, and the number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids. Cell-types of interest are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell type proportions between control and mutant, based on logistic mixed models (see Methods) are shown. **d**, Pseudotime trajectory from the full dataset of Mito210 *SUV420H1*$^{+/-}$ 35 d.i.v. (batch II) and control organoids, calculated with Monocle3. The partition highlighted by a box was subsetted and the trajectory is shown in Fig. 1d. **e**, Module scores (top) and their distribution across mutant and control cells (bottom) for all modules resulting from WGCNA analysis of the partition of interest from Mito210 *SUV420H1*$^{+/-}$ and control organoids at 35 d.i.v. (batch II). Cells were downsampled to have an equal number of cells per organoid. Names were assigned to each module based on the known functions of the genes included in each one. Horizontal bars show median scores, and dots show average score per organoid. Adjusted *p*-values show differences between control and mutant based on linear mixed models (see Methods). aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; CP/CH, Choroid Plexus/Cortical Hem; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons; GABA N, GABAergic neurons.

SUV420H1, *scATAC-Seq*

Mito210, *31 d.i.v. and 93 d.i.v.*

a

31 d.i.v. / **93 d.i.v.**

Cell Types scATAC-Seq — 84,696 nuclei / 23,669 nuclei

ATAC *vs* RNA — scRNA-Seq / scATAC-Seq

Cell Types

b

Upregulated in SUV420H1+/- | Downregulated in SUV420H1+/-

31 *d.i.v.* all nuclei
93 *d.i.v.* all nuclei

padj 0.03 / 0.02 / 0.01

Upregulated terms: modulation of chemical synaptic transmission; regulation of trans synaptic signaling; olfactory bulb interneuron differentiation; forebrain neuron differentiation; forebrain generation of neurons; synaptic membrane adhesion

Downregulated terms: forebrain development; negative regulation of neuron projection development; negative regulation of cell development; negative regulation of neuron differentiation; negative regulation of cell projection organization

GeneRatio 0.08 / 0.12 / 0.16

c

Control / SUV420H1+/-
Normalized accessibility (range 0 – 660) / (range 0 – 220)
Genes: GLIPR1L1, CAPS2, GLIPR1L2 / SLC4A10
Peaks
chr12 position (bp): 75370000 – 75410000 / chr2 position (bp): 161430000 – 161470000

d

Control / SUV420H1+/- : aRG, IPC, Newborn DL PN, Cortical Hem, Subcortical
Normalized accessibility (range 0 – 95) / (range 0 – 38)
Genes: GLIPR1L1, CAPS2, GLIPR1L2 / SLC4A10
Peaks
chr12 position (bp): 75370000 – 75410000 / chr2 position (bp): 161430000 – 161470000

Cell Types
- Cycling Progenitor
- aRG
- Newborn DL PN
- Immature DL PN
- Cajal-Retzius
- Cortical Hem
- Subcortical
- oRG
- IPC
- PN
- UL CPN
- DL CFuPN
- Unknown

e

Motif	Genes
	EN1, DLX2, LHX8, LHX1, DLX1, GSX2, EN2, PDX1, DLX5, HOXA1
	NFIA, NFIX, NFIC, HIC1, NF1-HALFSITE, PB0029.1_HIC1_1, HIC2, HIC1, TBX6, TGIF2
	PB0176.1_SOX5_2, HOXB4, NKX6-3, HOXD4, HOXC4, HOXA4, POU6F1, NKX6-1, NKX6.1, ARID3A
	NF1, RBPJ:EBOX, TLX?, NFIX, NFIC, NFIC::TLX1, NFIB, JUN, JUNB, FOSL1
	RFX4, RFX1, PAX5, PB0159.1_RFX4_2, RFX2, PB0160.1_RFXDC2_2, RFX3, SOX2, RFX2, ATF1
	LHX9, EMX2, PH0155.1_PRRX2, LHX6, NOTO, GBX1, LHX2, VSX2, PAX4, LHX1
	OCT4-SOX2-TCF-NANOG, POU5F1::SOX2, POU5F1, OCT4, OCT11, POU2F3, PH0145.1_POU2F3, POU5F1B, OCT2, PH
	PB0178.1_SOX8_2, PB0170.1_SOX17_2, PB0028.1_HBP1_1, ZNF24, NFYC, PB0068.1_SOX1_1, DUX, NKX6-3, NFYA, HOXB4
	ZNF711
	SOX15, SOX9, PB0183.1_SRY_2, SOX8, SOX7, PB0173.1_SOX21_2, SOX17, PB0065.1_SOX15_1, PB0073.1_SOX7_1, PB0168.1_SOX14_2
	NEUROD1, NEUROG2, NEUROG2, NEUROD1, HAND2, OLIG2, PTF1A, PB0193.1_TCFE2A_2, TCF4, BHLHA15
	MEF2A, MEF2C, MEF2C, MEF2D, MEF2D, MEF2A, MEF2B, MEF2B, ZNF384, PH0056.1_HOXA9
	NFIA, NFIC, NF1-HALFSITE, NFIX, MEIS2, HIC1, PH0170.1_TGIF2, TGIF1, NF1:FOXA1, THAP1
	POU2F2, PH0144.1_POU2F2, PH0145.1_POU2F3, BRN2, POU2F1, POU3F1, POU2F3, POU5F1B, OCT4, PO
	PKNOX1, PBX3, PBX3, PKNOX1, PBX1, MEIS3, MEIS1, PAX2, MEIS1, PH0141.1_PKNOX2
	NFIA, NFIX, NFIC, NF1-HALFSITE, HIC1, HIC2, NF1:FOXA1, HIC1, NFYB, PB0029.1_HIC1_1
	LHX2, VSX2, VSX1, EMX2, LHX9, EN1, PRRX2, LHX9, VAX1, LHX1
	HIC1, NF1-HALFSITE, HIC1, PB0029.1_HIC1_1, MEIS1, PKNOX1, TLX?, PBX3, SMAD3, TBOX:SMAD
	NEUROD1, DUX, NEUROD1, NEUROG2, NEUROG2, NFYA, NFYC, OLIG2, PH0037.1_HDX, GFI1B
	PB0013.1_EOMES_1, TBX21, TBR1, TBX2, EOMES, TBX18, TBX21, TBX20, TBET, TBR1

-log₁₀(pval): 0 / 20 / 40 / 60

- 31 *d.i.v._*UP
- 31 *d.i.v._*DOWN
- 93 *d.i.v._*UP
- 93 *d.i.v._*DOWN

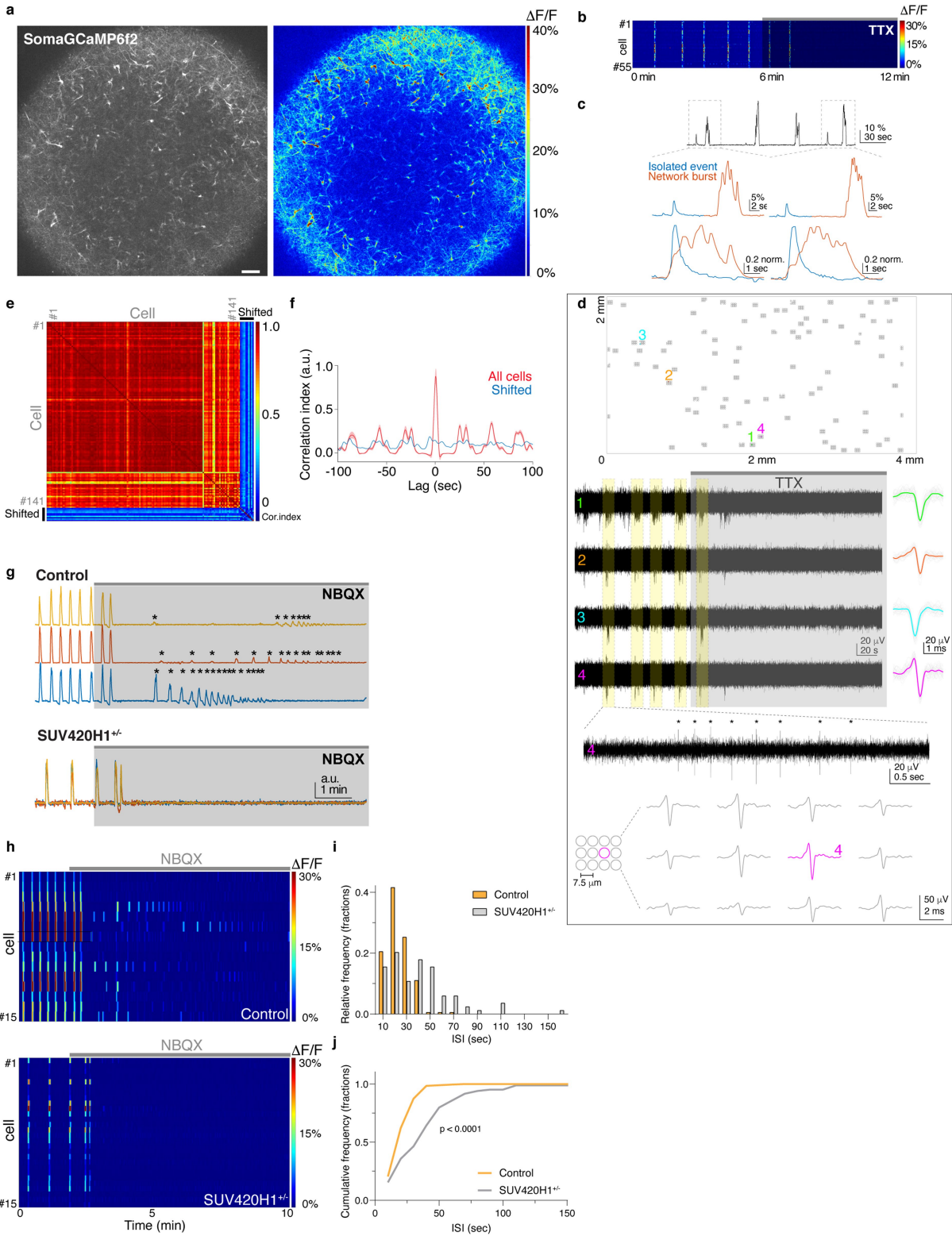**Extended Data Fig. 6** | See next page for caption.

Article

Extended Data Fig. 6 | scATAC-seq analysis in *SUV420H1*$^{+/-}$ and isogenic control organoids. a, UMAPs of scATAC-seq data in Mito210 *SUV420H1*$^{+/-}$ and control organoids at one month (31 d.i.v., upper left) and three months (93 d.i.v., upper right), and coembedded UMAPs with scRNA-seq in Mito210 *SUV420H1*$^{+/-}$ and control organoids at one month (28 d.i.v., middle bottom left) and three months (90 d.i.v., middle bottom right). Number of nuclei per plot is indicated. **b**, Enriched GO terms for the nearest genes to regions with increased and decreased accessibility in *SUV420H1*$^{+/-}$ compared to control organoids. **c**, Genome tracks showing read coverage for representative regions with increased accessibility between *SUV420H1*$^{+/-}$ and control organoids. Scales for the y axes (normalized counts) are displayed on the top right. **d**, Genome tracks showing read coverage for representative regions with increased accessibility between Mito210 *SUV420H1*$^{+/-}$ and control organoids, split by cell-type. Scales for the y axes (normalized counts) are displayed on the top right. **e**, Top 5 *de novo* motifs enriched in the regions with altered accessibility in Mito210 *SUV420H1*$^{+/-}$ compared to control organoids at one month (31 d.i.v.) and three months (93 d.i.v.), as calculated with HOMER (see Methods). Regions that showed increased accessibility in mutant compared to control organoids are in green and purple, while those with decreased accessibility are in red and blue. Transcription factors with known binding sites resembling the discovered motifs are shown.
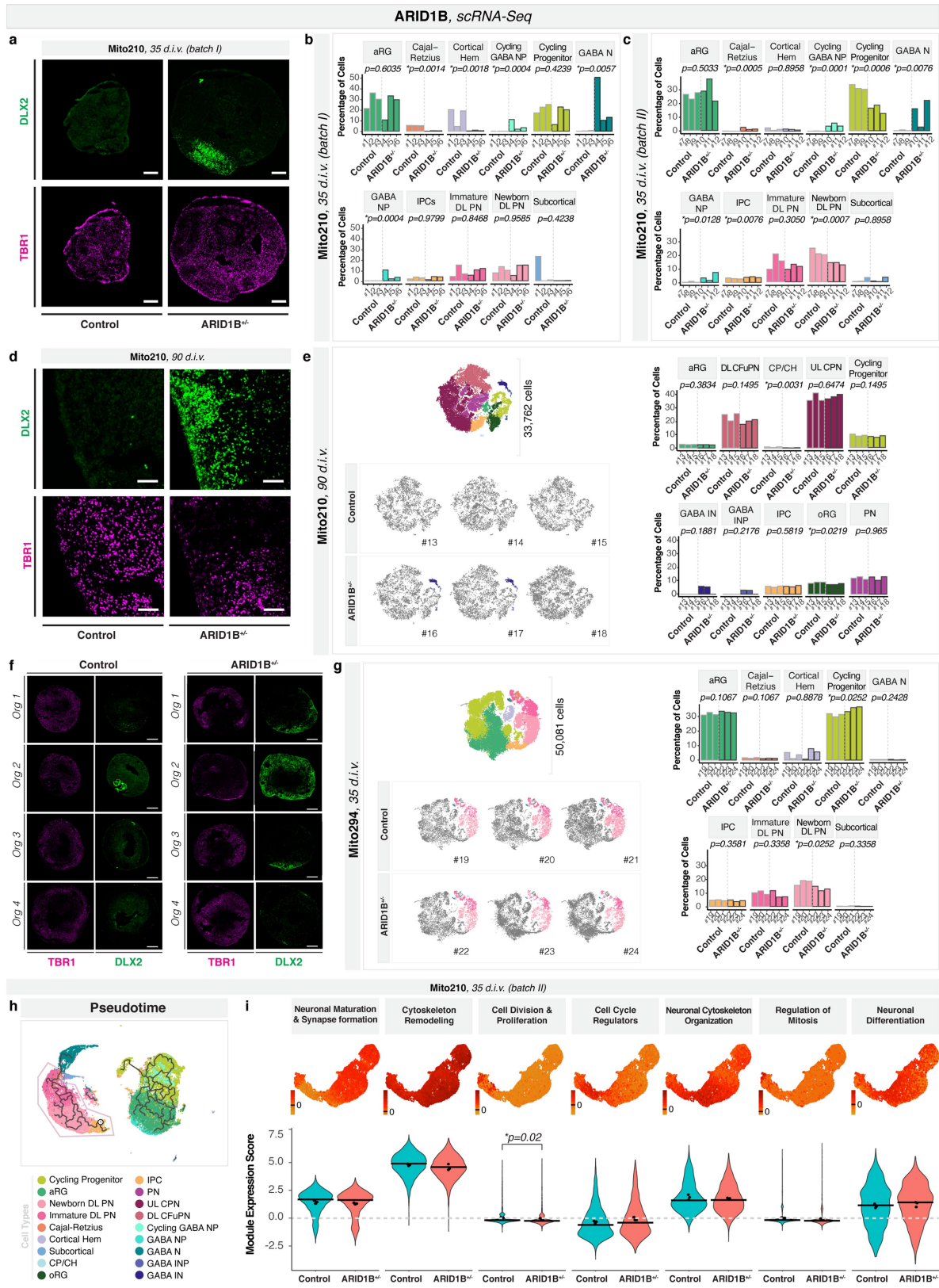
PGP1, *128 d.i.v.*

Extended Data Fig. 7 | See next page for caption.

Article

Extended Data Fig. 7 | Neuronal spontaneous activity in *SUV420H1*[+/−] and isogenic control organoids. a, Left, Representative image of a PGP1 *SUV420H1* organoid infected with SomaGCaMP6f2. Right, ΔF/F signal at the peak of a network burst. Scale bar: 100 μm. **b**, Heat map of calcium signal from individual cells (rows), showing the effect of 2 μM TTX. **c**, Top, representative trace of spontaneous calcium signal (corresponding to cell #3 in Fig. 1g). Bottom, high magnification traces of calcium transients, displaying the difference in amplitude between the isolated event and the network burst (top), and normalized traces (bottom) showing their kinetics and the multiple peaks of the burst signal. **d**, MEA recordings. Top, Spatial configuration of recording electrodes. Middle, example raw traces for the numbered electrodes shown at the top, and the effect upon 2 μM TTX application. Yellow columns indicate the network bursts. Right, individual (grey) and average (colour) spike waveforms for each electrode. High magnification of the trace #4 showing the individual spikes (asterisk) during a burst event. Bottom, average spike waveform (right) in a unit of electrodes (left), extracted at the time points determined by the spikes in electrode #4. **e, f**, Synchronous network activity in human brain organoids. Heat map of cross-correlation index (**e**) and cross-correlogram against a reference signal (cell #135) for a representative recording. As a control, the signal of 10 cells were circularly shifted by a random phase and the cross-correlation was then calculated. In **f**, the average value was plotted, and the synchronous activations as well as the periodic bursting can be seen ("All cells" in red). **g**, Effect of NBQX on neuronal activity. Representative traces for individual cells were normalized (3 traces for *SUV420H1*[+/−] are superimposed) and post-NBQX residual/isolated calcium transients are indicated by asterisks. **h**, Effect of NBQX on calcium signal. Heat map of ΔF/F signal for 15 representative cells in control (top) and *SUV420H1*[+/−] (bottom) organoids. **i, j**, Inter-spike interval (ISI) analysis for the network bursting. Relative frequency (top) and cumulative frequency distribution (bottom) of ISI for control and *SUV420H1*[+/−] organoids. In **j**, two-sided Kolmogorov-Smirnov test (n = 5 organoids per genotype). Number of organoids used for each analysis can be found in the Methods under "Statistics and reproducibility".

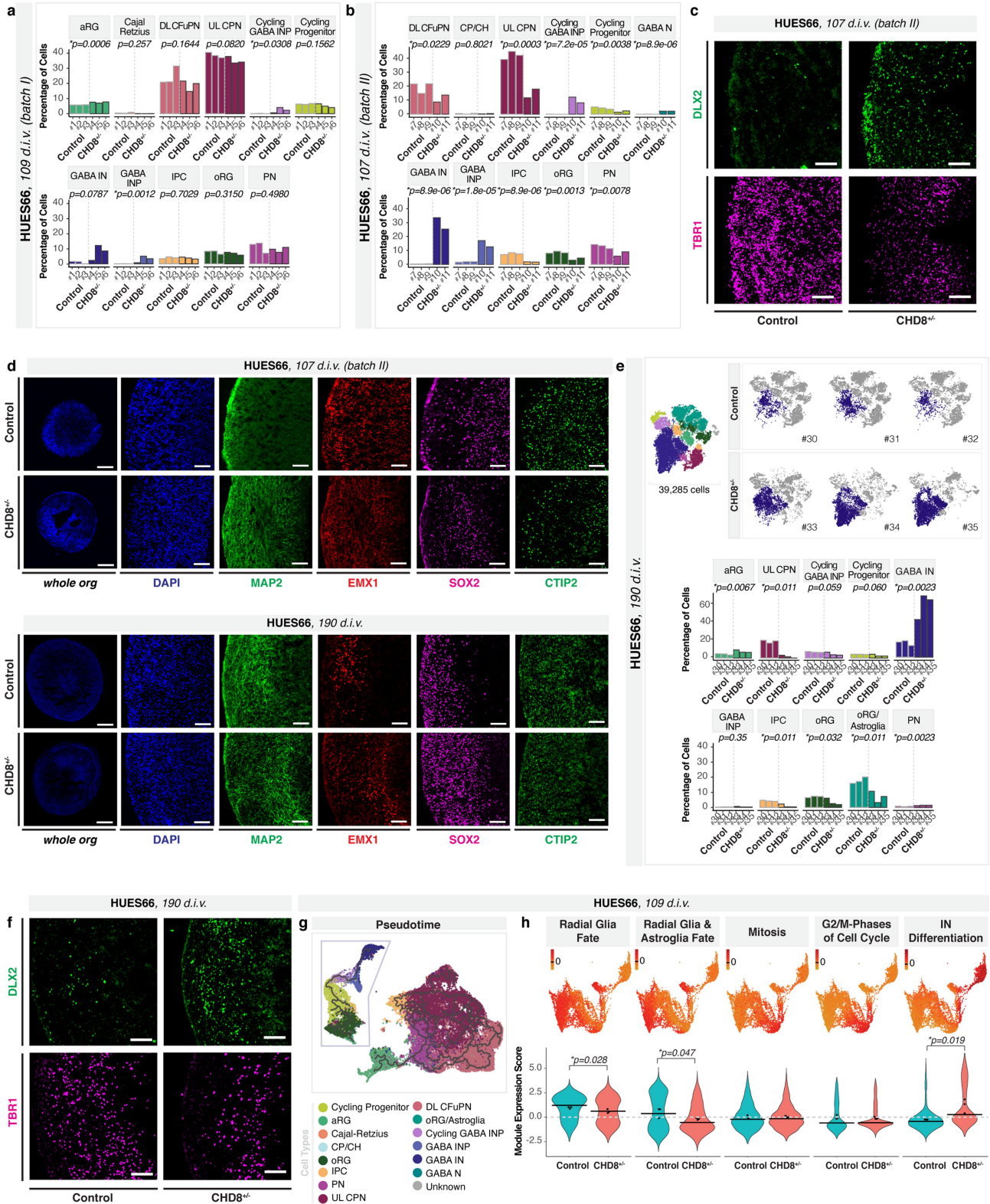**Extended Data Fig. 8** | See next page for caption.

Article

Extended Data Fig. 8 | Cell-type composition, full pseudotime trajectories, and gene modules of *ARID1B*⁺/⁻ and isogenic control organoids.
a, Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in Mito210 control and *ARID1B*⁺/⁻ organoids at one month (35 d.i.v.). Scale bars: 200 μm. **b**, **c**, scRNA-seq data from Mito210 one month (35 d.i.v. batch I in **b**, batch II in **c**) control and *ARID1B*⁺/⁻ organoids. Bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell-type proportions between control and mutant, based on logistic mixed models (see Methods) are shown. **d**, Immunohistochemistry for TBR1 (magenta) and DLX2 (green) in Mito210 control and *ARID1B*⁺/⁻ organoids at three months (90 d.i.v.). Scale bars: 100 μm. **e**, scRNA-seq data from Mito210 three months (90 d.i.v.) control and *ARID1B*⁺/⁻ organoids. Left top shows combined *t*-SNE plots of control and mutant organoids (n = 3 single organoids per genotype). Cells are coloured by cell-type, and the number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids. GABAergic interneurons are highlighted in colour. Left, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as shown in **b**, **c**. Two out of three mutant organoids show an increase in GABAergic interneurons, vs. zero out of three controls (adjusted p = 0.19, logistic mixed models). **f**, Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in Mito210 control and *ARID1B*⁺/⁻ organoids at three months (90 d.i.v.). Three out of four mutant organoids contain DLX2-positive cells, while no DLX2 expression is detected in the four controls. Scale bars: 500 μm. **g**, scRNA-seq data from Mito294 one month (35 d.i.v.) *ARID1B*⁺/⁻ and control organoids. Left top shows combined *t*-SNE plots of control and mutant organoids (n = 3 single organoids per genotype). Cells are coloured by cell type, and the number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids. GABAergic neurons, newborn deep-layer projection neurons and immature deep-layer projection neuron populations are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as in **b**, **c**, **e**.
h, Pseudotime trajectories from the full dataset of Mito210 *ARID1B*⁺/⁻ 35 d.i.v. *batch II* and control organoids, calculated with Monocle3. The partition highlighted by a box was subsetted and the trajectory is shown in Fig. 2c.
i, Module scores (top) and their distribution across mutant and control cells (bottom) for all modules resulting from WGCNA analysis of the partition of interest from Mito210 *ARID1B1*⁺/⁻ and control organoids at 35 d.i.v. Cells were downsampled to have an equal number of cells per organoid. Names were assigned to each module based on the known functions of the genes included in each one. Horizontal bars show median scores, and dots show average score per organoid. Adjusted *p*-values show differences between control and mutant based on linear mixed models (see Methods). Number of organoids used for each analysis can be found in the Methods under "Statistics and reproducibility". aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; CP/CH, Choroid Plexus/Cortical Hem; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA NP, GABAergic neuron progenitors; GABA N, GABAergic neurons; GABA INP; GABAergic interneuron progenitors; GABA IN, GABAergic interneurons.
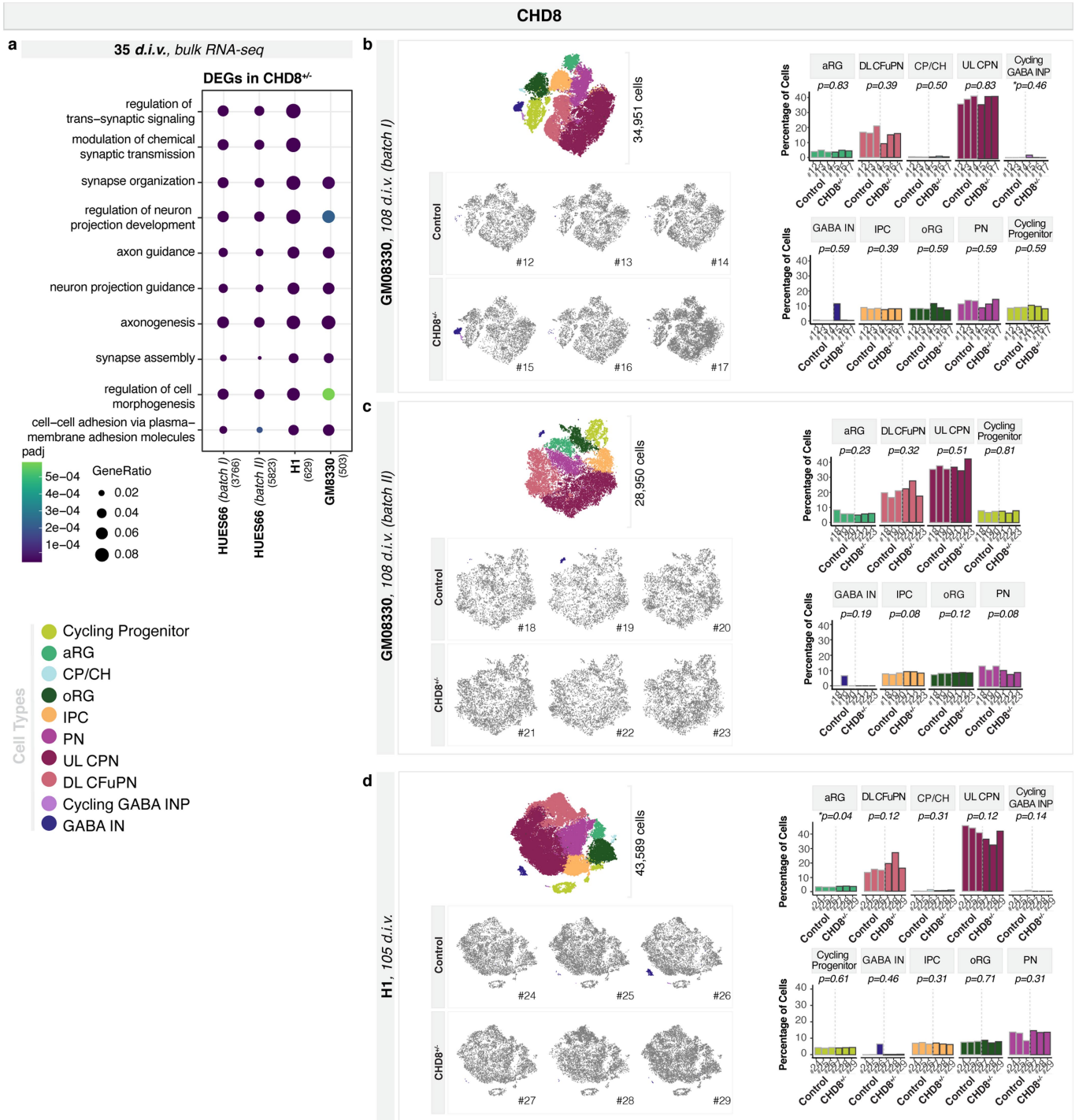
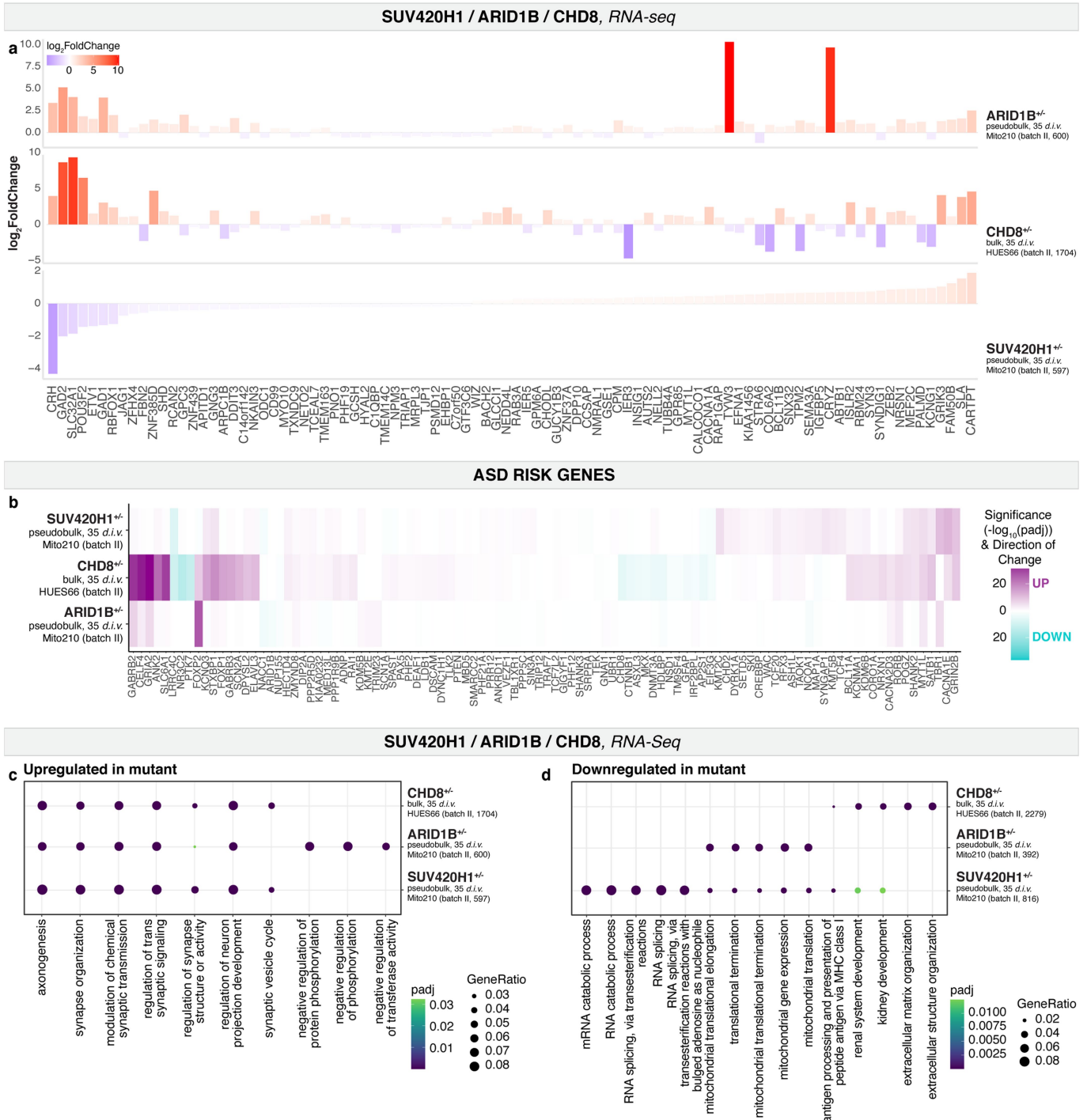**Extended Data Fig. 9** | See next page for caption.

Article

Extended Data Fig. 9 | Cell-type composition, immunohistochemistry, and full pseudotime trajectories and gene modules of *CHD8*$^{+/-}$ **and isogenic control HUES66 organoids. a, b,** scRNA-seq data from HUES66 3.5-month (109 d.i.v. (**a**), batch I and 107 d.i.v. (**b**). batch II) *CHD8*$^{+/-}$ and control organoids. Bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell-type proportions between control and mutant, based on logistic mixed models (see Methods) are shown. **c,** Immunohistochemistry for the postmitotic excitatory neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in HUES66 control and *CHD8*$^{+/-}$ organoids at 3.5 months (107 d.i.v.). Scale bars: 100 μm. **d,** Immunohistochemistry for neuronal (MAP2), dorsal forebrain neural progenitor (EMX1, SOX2) and CFuPN (CTIP2) markers in HUES66 *CHD8*$^{+/-}$ and control organoids at 3.5 months (107 d.i.v., top), and six months (190 d.i.v., bottom). Scale bars: whole organoids, 500 μm; others, 100 μm. **e,** scRNA-seq data from HUES66 *CHD8*$^{+/-}$ and control organoids at six months (190 d.i.v.). Top left shows combined *t*-SNE plots of control and mutant organoids (n = 3 single organoids per genotype). Cells are coloured by cell-type, and the number of cells per plot is indicated. Top right, *t*-SNE plots for control and mutant individual organoids. GABAergic interneurons are highlighted in colour. Bottom, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid, as in **a, b. f,** Immunohistochemistry for the postmitotic neuronal marker TBR1 (magenta) and GABAergic marker DLX2 (green) in HUES66 control and *CHD8*$^{+/-}$ organoids at six months (190 d.i.v.). Scale bars: 100 μm. **g,** Pseudotime trajectories from the full dataset of HUES66 batch I *CHD8*$^{+/-}$ and control organoids at 109 d.i.v., calculated with Monocle3. The partition highlighted by a box was subsetted and the trajectory is shown in Fig. 3c. **h,** Module scores (top) and their distribution across mutant and control cells (bottom) for all modules resulting from WGCNA analysis of the partition of interest from HUES66 *CHD8*$^{+/-}$ and control organoids at 109 d.i.v. Cells were downsampled to have an equal number of cells per organoid. Names were assigned to each module based on the known functions of the genes included in each one. Horizontal bars show median scores, and dots show average score per organoid. Adjusted *p*-values show differences between control and mutant based on linear mixed models (see Methods). Number of organoids used for each analysis can be found in the Methods under "Statistics and reproducibility". aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons.
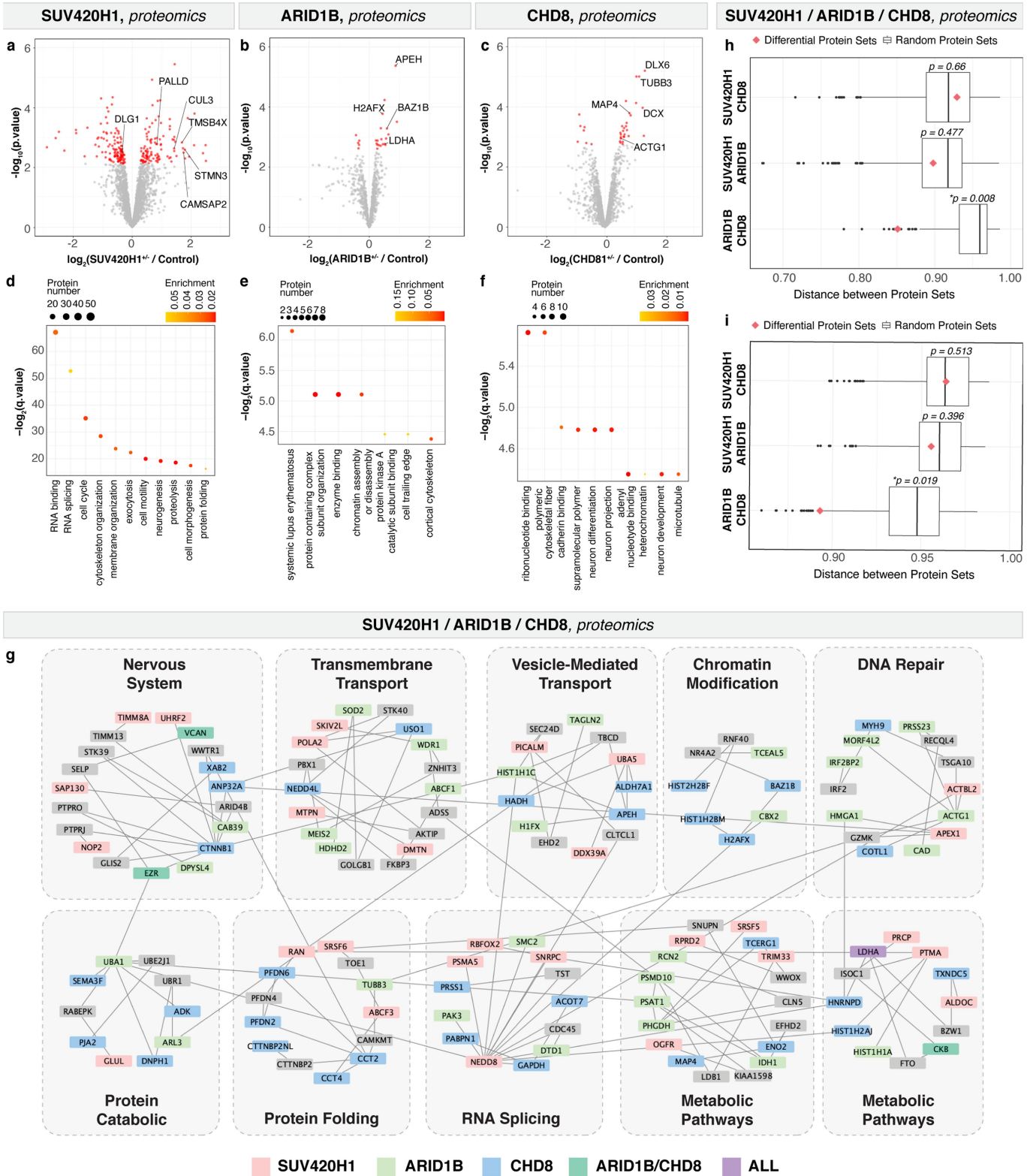
Extended Data Fig. 10 | Bulk RNA-seq and scRNA-seq of *CHD8*[+/−] and isogenic control organoids from multiple cell lines. a, Bulk RNA-seq data from HUES66, GM83330 and H1 35 d.i.v. organoids. Enriched GO terms for genes differentially expressed in *CHD8*[+/−] vs. control organoids. The top 5 most significant terms per dataset are shown. Size of dot indicates the proportion of genes belonging to each term found in the list of dysregulated genes ("GeneRatio"). Colour indicates enrichment adjusted *p*-value. Numbers in parentheses along the y axis indicate the number of DEGs in that dataset. **b**–**d**, scRNA-seq data from control and *CHD8*[+/−] organoids at 3.5 months (GM83330 108 d.i.v., batch I (**b**), GM83330 108 d.i.v., batch II (**c**) and H1 105 d.i.v. (**d**)). Left top shows combined *t*-SNE plots of control and mutant organoids (n = 3 single organoids per genotype). Cells are coloured by cell type, and the number of cells per plot is indicated. Left bottom, *t*-SNE plots for control and mutant individual organoids. GABAergic interneurons are highlighted in colour. Right, bar charts show the percentage of cells for all the cell populations in each control and mutant organoid. Adjusted *p*-values for a difference in cell-type proportions between control and mutant, based on logistic mixed models (see Methods) are shown. aRG, apical radial glia; DL, deep layer; UL, upper layer; PN, projection neurons; CP/CH, Choroid Plexus/Cortical Hem; oRG, outer radial glia; IPC, intermediate progenitor cells; CPN, callosal projection neurons; CFuPN, corticofugal projection neurons; GABA INP, GABAergic interneuron progenitors; GABA IN, GABAergic interneurons; GABA N, GABAergic neurons.

Extended Data Fig. 11 | Convergent differential expressed genes for the three mutations. a, Log fold change of all genes which showed significant change (adjusted *p* < 0.05) in all three of the 1 month datasets: Mito210 *SUV420H1*[+/-] 35 d.i.v., Mito210 *ARID1B*[+/-] 35 d.i.v., and HUES66 *CHD8*[+/-] 35 d.i.v. DEGs were calculated using all cells as a pseudobulk for Mito210 *SUV420H1*[+/-] and Mito210 *ARID1B*[+/-]. **b**, Differential expression of all 102 genes associated with ASD risk[6] in the three datasets Mito210 *SUV420H1*[+/-] 35 d.i.v., Mito210 *ARID1B*[+/-] 35 d.i.v. and in HUES66 *CHD8*[+/-] 35 d.i.v. compared to relative controls. Expression of risk genes was calculated using all cells (pseudobulk) for Mito210 *SUV420H1*[+/-] and Mito210 *ARID1B*[+/-]. Boxes are coloured according to

-log[10](adjusted *p* value) according to whether they are upregulated (purple), or downregulated (turquoise) in mutant vs. control. Genes are ordered according to hierarchical clustering (using Euclidean distance) of those values. **c**, **d**, Enriched GO terms for genes upregulated (**c**) and downregulated (**d**) in mutant vs. control. Genes were calculated using the cells as in **a**, **b**. The top 5 most significant terms per dataset are shown. Size of dot indicates the proportion of genes belonging to each term found in the list of dysregulated genes ("GeneRatio"). Colour indicates enrichment adjusted *p*-value. Numbers in parentheses along the x axis indicate the number of DEGs in that dataset.
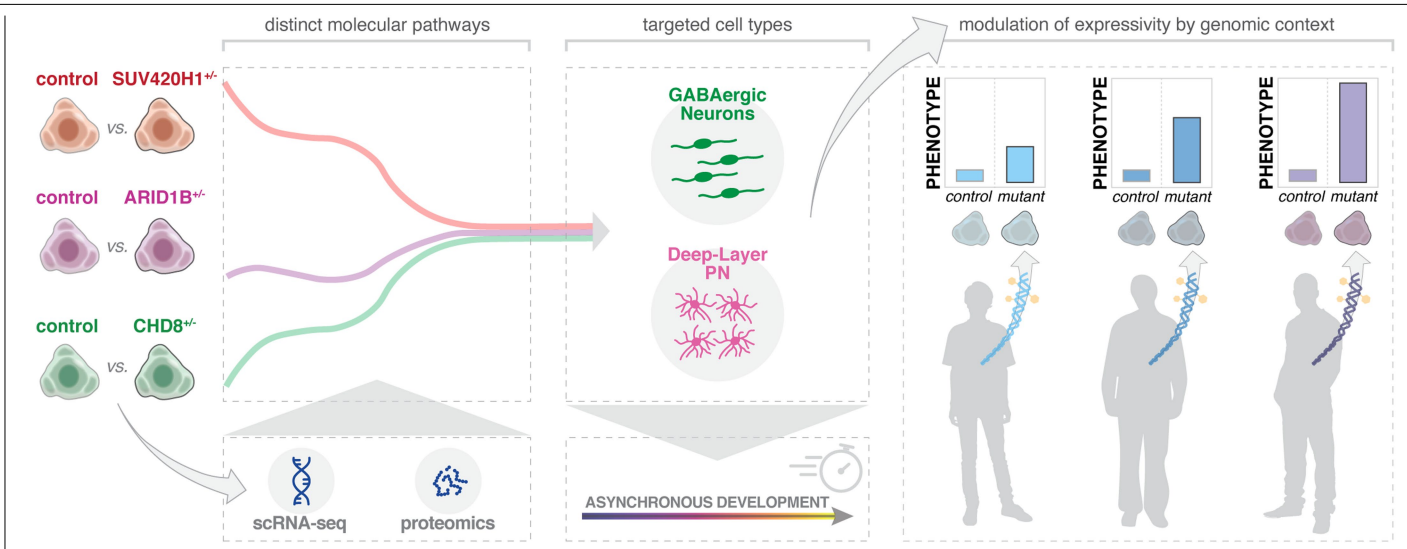
SUV420H1, *proteomics*

a

b ARID1B, *proteomics*

c CHD8, *proteomics*

d

e

f

SUV420H1 / ARID1B / CHD8, *proteomics*

h

i

SUV420H1 / ARID1B / CHD8, *proteomics*

g

Nervous System · Transmembrane Transport · Vesicle-Mediated Transport · Chromatin Modification · DNA Repair

Protein Catabolic · Protein Folding · RNA Splicing · Metabolic Pathways · Metabolic Pathways

Legend: SUV420H1 · ARID1B · CHD8 · ARID1B/CHD8 · ALL

Extended Data Fig. 12 | See next page for caption.

Article

Extended Data Fig. 12 | Convergent differentially expressed proteins for the three mutations. a–c, Volcano plot showing fold change versus adjusted *p*-value of measured proteins in MS experiments on Mito210 *SUV420H1*[+/−] (**a**), Mito210 *ARID1B*[+/−] (**b**), and HUES66 *CHD8*[+/−] (**c**) vs. control organoids at 35 d.i.v. (n = 4 single organoids per genotype for *SUV420H1*, 4 controls and 5 mutants for *ARID1B*, and n = 3 single organoids per genotype for *CHD8*). To detect statistically significant differential protein abundance between conditions a moderated t-test was performed (see Methods, FDR threshold of 0.1). Significant DEPs are shown in red (FDR < 0.1). **d–f**, Selected enriched GO terms for DEPs in Mito210 *SUV420H1*[+/−] (**d**), Mito210 *ARID1B*[+/−] (**e**), and HUES66 *CHD8*[+/−] (**f**) vs. control organoids cultured for 35 d.i.v. GO terms and KEGG pathways were calculated using the GSEA software (see Methods) and FDR q-values < 0.05 were considered statistically significant. **g**, Protein-protein interaction network using the top 50 DEPs from the three sets of mutant *versus* control organoids, created using the prize-collecting Steiner forest algorithm (see Methods). Protein nodes are coloured by the mutant in which they were differentially expressed. Gray nodes indicate "Steiner nodes", proteins that did not result from any screen but were included by the algorithm to connect DEPs. Lines between nodes indicate physical protein-protein interactions from the STRING database, where line thickness correlates with interaction confidence. Subclusters of the network and significantly enriched terms for those subclusters are highlighted with gray rectangles and black text. **h**, Protein set distances between pairs of differentially expressed protein sets. For each pair of mutations, a PPI-weighted protein set distance was calculated between all significant DEPs (FDR < 0.1, pink diamond). To determine if this distance was smaller than would be expected by chance, size-matched sets were randomly chosen from the proteins detected in each experiment, and distance between these random sets was calculated 1000 times per pair. *P*-values were assigned by counting the fraction of times this random distance was less than the actual distance value between differential sets. Lower and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles) and middle hinge is the median (50th). Both whiskers extends from the hinge to the largest or smallest value no further than 1.5 * IQR from the hinge (where IQR is the inter-quartile range, or distance between the first and third quartiles). **i**, Protein set distances between the top 50 DEPs per mutation. For each pair of mutations, a PPI-weighted protein set distance was calculated as in **h**. Number of organoids used for the analyses are summarized in the Methods under "Statistics and reproducibility". DEPs: differentially expressed proteins. MS: mass spectrometry.

Extended Data Fig. 13 | Mutations in ASD risk genes in human brain organoids converge on asynchronous development of shared neuronal classes. Conceptual schematics highlighting main results.

Corresponding author(s):	Silvia Velasco, Paola Arlotta
Last updated by author(s):	Dec 8, 2021

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Imaging software Gen5 (BioTek Instruments) or Zen Blue (ZEN 2.6 – blue edition, Zeiss) were used. Calcium imaging was performed using Nikon's NIS-Elements imaging software (NIS-Elements Advance Research (Ver.4.51.01). Extracellular neurophysiological signals were recorded using MaxLab Live Software (v.20.1.6. MaxWell Biosystems AG, Switzerland).
Data analysis	The CRISPR guides for SUV420H1 and ARID1B were designed using the Benchling CRISPR Guide Design Tool (Benchling Biology Software, 2017). ImageJ72 (v.2.0) was used to quantify organoid size. Raw proteomics data was analyzed with Proteome Discoverer 2.4 software (Thermo Fisher Scientific), and Cytoscape v. 3.8.2 was used for network visualization. Cell Ranger versions 2.0.1 and 3.0.2 were used for analysis of raw scRNA-seq data. In one case, scRNA-seq demultiplexing was performed with with the cellSNP tool v. 0.1.5 and the cardelino R library v. 0.4.0. Downstream analysis of scRNA-seq and proteomics data was performed using R v. 3.6 packages including DEP, LIMMA, GSEA, PCSF, Seurat v. 3.1.5, Harmony v. 1.0, lme4 v. 1.1-23, Monocle3 v. 0.2.0, stats, clusterProfiler, and WGCNA v. 1.69, as well as Signac v. 0.0.7 from BioTuring. Downstream analysis of scATAC-seq data was performed using Signac v. 1.2.1 in R v. 4.0.0, and TF motifs were called using HOMER v. 4.11.1. Calcium imaging was analyzed using Suite2P and Matlab. Multi-electrode array analysis was done with MaxLab Live Software (v. 20.1.6. MaxWell Biosystems AG, Switzerland) and Matlab Toolbox provided by MaxWell Biosystems. Code written by authors used for data analysis is available on GitHub at https://github.com/AmandaKedaigle/mutated-brain-organoids.

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

Data

Read-level data from scRNA-seq and scATAC-seq, as well as proteomic data, that support the findings of this study have been deposited in a controlled access repository at https://www.synapse.org with accession number project ID syn26346373, while count-level data and metadata have been deposited at the Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP1129). The electrophysiology materials and data are available upon request.

Public data used in this paper includes the GRCh38 human reference genome and the EnsDb.Hsapiens.v86 annotation package.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size | For organoid size analysis, detailed information can be found in Extended Data Table 2 - in summary SUV420H1+/- organoids: n = 132 for total control organoids, n = 132 for total mutant organoids, from 6 experimental batches. For ARID1B+/- organoids: n = 109 for total control organoids, n = 112 for total mutant organoids, from 4 experimental batches. For CHD8+/- organoids: n = 472 for total control organoids, n = 482 for total mutant organoids, from 7 experimental batches.

For proteomic analysis, all organoids were collected at 35 d.i.v. For SUV420H1: 4 organoids were used per genotype; for ARID1B: 4-5 organoids were used per genotype; and for CHD8: 3 organoids were used per genotype.

For RNA-seq experiments, detailed information can be found on Extended Data Table 10. In summary, 3 organoids per genotype were initially sequenced for scRNA-seq, for a total of 114 single organoids sequenced individually. The number of cells sampled were chosen to maximize the availability of the sequencing lane per organoid. For bulk RNA-seq, 3 organoids were sequenced per genotype, being a total of 30 individual organoids.

For scATAC-seq experiments, detailed information can be found on Extended Data Table 10. In summary, 3 organoids were sequenced per genotype, being a total of 12 individually sequenced organoids.

For Calcium imaging experiments, 5 organoids were analyzed per genotype. Spontaneous activity was recorded in three different z-planes (120 ± 83 neurons/plane [range from 25 to 294 neurons/plane] in control organoids, and 107 ± 75 neurons/plane [range from 32 to 255 neurons/plane] in SUV420H1+/- organoids).

No methods were used to predetermine sample size, rather at least three individual organoids were used for each experiment, based on analysis from previously published work (https://www.nature.com/articles/s41586-019-1289-x) which showed that organoids were reproducible and 3 organoids captured variation.

Data exclusions | In two cases, one organoid was excluded from the analysis as an outlier: in PGP1 SUV420H1 organoids at one month, a mutant organoid was excluded due to very low average nUMI and nGene in that sequencing lane, and in the HUES66 CHD8 organoids at 3.5 month batch II, a mutant organoid was excluded because it mostly contained interneuron lineage cells, with very few projection neuron cells. Although an increase in interneuron-lineage cells was seen in all mutant organoids, this organoid was excluded to be conservative. This left a total of 124 single organoids that passed quality control and were considered in downstream analysis, with a total of 795,358 cells, for both scATAC-seq and scRNA-seq.

Replication | Where indicated in the paper and in Extended Data Table 10, scRNA-seq experiments were replicated in multiple differentiation batch and cell lines.

For scATAC-seq, calcium imaging, and proteomics, one experiment each was performed at each time point, with three or more individual organoids per genotype. Bulk RNA-seq experiments were replicated in five batches.

Immunohistochemistry: at least three organoids of each condition were used for verifying the expression of the indicated markers in Extended Data Fig. 1a-c; Extended Data Fig. 3g; Extended Data Fig. 4a-b; Extended Data Fig. 8a,d,f; Extended Data Fig. 9c,d,f.
Western Blot: each control and mutant protein lysate was botted at least twice.

Detailed information can be found in Methods under "Statistics and reproducibility".

Randomization	Samples were not randomized to experimental groups because treatment groups (mutants) were paired with their isogenic controls from the same differentiation batch. Organoids used for analysis within each experimental group were chosen from each batch to be representative of the morphology seen in that differentiation.
Blinding	Investigators were not blinded. All bioinformatic analyses were applied the same to all samples without adjustment for genotype.

Reporting for specific materials, systems and methods

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

<table>
<tr><td colspan="2">Materials & experimental systems</td><td colspan="2">Methods</td></tr>
<tr><td>n/a</td><td>Involved in the study</td><td>n/a</td><td>Involved in the study</td></tr>
<tr><td>☐</td><td>☒ Antibodies</td><td>☒</td><td>☐ ChIP-seq</td></tr>
<tr><td>☐</td><td>☒ Eukaryotic cell lines</td><td>☒</td><td>☐ Flow cytometry</td></tr>
<tr><td>☒</td><td>☐ Palaeontology and archaeology</td><td>☒</td><td>☐ MRI-based neuroimaging</td></tr>
<tr><td>☒</td><td>☐ Animals and other organisms</td><td></td><td></td></tr>
<tr><td>☒</td><td>☐ Human research participants</td><td></td><td></td></tr>
<tr><td>☒</td><td>☐ Clinical data</td><td></td><td></td></tr>
<tr><td>☒</td><td>☐ Dual use research of concern</td><td></td><td></td></tr>
</table>

Antibodies

Antibodies used	Primary antibodies: Rat CTIP2 [25B6] (Abcam AB18465,1/100) Rabbit EMX1 (Atlas Antibodies HPA006421, Lot: E114426, 1/50) Mouse GFAP (Sigma-Aldrich G3893, Lot: 038M4864V, 1/400) Chicken MAP2 (Abcam AB5392, 1/5,000) Rabbit S100B (Abcam AB41548, 1/2,000) Mouse SATB2 [SABTA4B10] (Abcam AB51502, 1/50) Goat SOX2 (RD Systems AF2018, 1/50) Rabbit TBR1 (Abcam AB31940, Lot: GR3182037-1, 1/500) Mouse DLX2 (Santa Cruz sc-393879, Lot: A2320, 1/100) Rabbit Histone H4 tri-methyl K20 (Abcam ab9053, 1/000) Rabbit CHD8 (Bethy Lab, A301-225A, 1/1000) Rabbit ARID1B (Bethyl Lab, A301-046A, 1/1000) Rabbit SUV420H1 (ProSci, 28-129, 1/1000) Rabbit Histone 4 (Abcam, ab7311, 1/500) Mouse GAPDH (Sigma, G8795, 1/30,000) Rabbit b-Actin (D6A8, Cell Signaling, 8457T, 1/10,000) Secondary antibodies: All secondary antibodies were diluted 1/1,200 Alexa Fluor donkey anti-rabbit 647, 546 (Life Technologies A31573, A10040) Alexa Fluor donkey anti-mouse 546, 488 (Life Technologies A10036, A21202) Alexa Fluor donkey anti-goat 647 (Life Technologies A21447) Alexa Fluor donkey anti-rat 488 (Life Technologies A21208) Alexa Fluor donkey anti-chicken 488,647 (Jackson ImmunoResearch Laboratories 703-545-155, Millipore AP194SA6) Alexa Fluor donkey anti-guinea pig 647 (Jackson ImmunoResearch Laboratories 706-605-148) Alexa Fluor goat anti-mouse 647 (ThermoFisher Scientific, A21241) Secondary antibodies Wbs: All secondary antibodies were diluted 1/10,000 HRP Goat Anti-Mouse IgG H&L (Abcam, ab205719) HRP Goat Anti-Rabbit IgG H&L (Abcam, ab205718)
Validation	According to the manufacturer's website: The rat CTIP2 antibody (Abcam, AB 18465) has been cited in 295 publications, including Quadrato et al., 2017, from our lab, which used this antibody on human brain organoids. The rabbit EMX1 antibody (Atlas Antibodies, HPA006421) has been cited in 8 publications, including Lancaster et al., 2013 and Quadrato et al., 2017, which used this antibody on human brain organoids. The mouse GFAP antibody (Sigma-Aldrich, G3893) is reactive to human, and has been cited in 541 publications. The chicken MAP2 antibody (Abcam, AB5392) is reactive to human, and has been cited in 277 publications. The rabbit S100B antibody (Abcam, 41548) has been cited in 32 publications, including Bedner et al., 2015, which used this antibody in human hippocampal tissue. The mouse SATB2 antibody (Abcam, AB51502) is reactive to human and has been cited in 99 publications. The goat SOX2 antibody (RD Systems, AF2018) is reactive to human and has been cited in 72 publications.

The rabbit TBR1 antibody (Abcam AB31940) is reactive to human and has been cited in 190 publications.
The mouse DLX2 antibody (Santa Cruz sc-393879) is reactive to human and has been cited in 3 publications.
The Histone H4 tri-methyl K20 antibody (Abcam ab9053) is reactive to human and has been cited in 173 publications.
The CHD8 antibody (Bethy Lab, A301-225A) is reactive to human and has been cited in 10 publications.
The ARID1B antibody (Bethy Lab, A301-046A) is reactive to human and has been cited in 4 publications.
The SUV420H1 antibody (ProSci, 28-12) is reactive to human and has tested in Jurkat cells, human skin, human fetal lung, liver, heart and brain.
The Histone 4 antibody (Abcam, ab7311) is reactive to human and has been cited in 82 publications.
The GAPDH antibody (Sigma, G8795) is reactive to human and has been cited in 1139 publications.
The b-Actin antibody (Cell Signaling, 8457T) is reactive to human and has been cited in 528 publications.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The HUES66 ESC line and mutant CHD8 line (HUES66 AC2) were provided by N. Sanjana, X. Shi, J. Pan, and F. Zhang (Broad Institute of MIT and Harvard); the Mito 210 human male psychiatric control iPSC line by B. Cohen Lab (McLean Hospital); and the GM08330 iPSC line (a.k.a. GM8330-8) by M. Talkowski Lab (MGH Hospital) and was originally from Coriell Institute and the H1 parental hESC line (a.k.a. WA01) was purchased from WiCell.
Authentication	The HUES66 and PGP1 lines were authenticated using STR analysis completed by GlobalStem (in 2008) and TRIPath (in 2018), respectively. The Mito210 and Mito294 lines were authenticated by genotyping analysis (Fluidigm FPV5 chip) performed by the Broad Institute Genomics Platform (in 2017). The H1 and GM08330 lines were authenticated using STR analysis completed by WiCell (in 2021).
Mycoplasma contamination	All human PSC cultures were negative for mycoplasma (assayed with MycoAlert PLUS Mycoplasma Detection Kit, Lonza)
Commonly misidentified lines (See ICLAC register)	No misidentified lines were used in the study.