



Long-term maturation of human cortical organoids matches key early postnatal transitions

Aaron Gordon¹, Se-Jin Yoon^{ID 2,3}, Stephen S. Tran^{4,5}, Christopher D. Makinson^{ID 6}, Jin Young Park^{2,3}, Jimena Andersen^{2,3}, Alfredo M. Valencia^{2,3}, Steve Horvath^{ID 7,8}, Xinshu Xiao^{5,9,10}, John R. Huguenard^{ID 6}, Sergiu P. Pasca^{ID 2,3} and Daniel H. Geschwind^{ID 1,8,11,12}

Human stem-cell-derived models provide the promise of accelerating our understanding of brain disorders, but not knowing whether they possess the ability to mature beyond mid- to late-fetal stages potentially limits their utility. We leveraged a directed differentiation protocol to comprehensively assess maturation *in vitro*. Based on genome-wide analysis of the epigenetic clock and transcriptomics, as well as RNA editing, we observe that three-dimensional human cortical organoids reach postnatal stages between 250 and 300 days, a timeline paralleling *in vivo* development. We demonstrate the presence of several known developmental milestones, including switches in the histone deacetylase complex and NMDA receptor subunits, which we confirm at the protein and physiological levels. These results suggest that important components of an intrinsic *in vivo* developmental program persist *in vitro*. We further map neurodevelopmental and neurodegenerative disease risk genes onto *in vitro* gene expression trajectories to provide a resource and webtool (Gene Expression in Cortical Organoids, GECO) to guide disease modeling.

The scarcity and inaccessibility of tissue from many developmental stages in patients with neurodevelopmental and neurodegenerative disorders highlights the need for advanced *in vitro* models of human brain development and maturation^{1,2}. Indeed, recent advances make it possible to differentiate human pluripotent stem cells into self-organizing, three-dimensional (3D) cellular ensembles that recapitulate several features of human brain development^{3–5}. One advantage of these organoid models is that they can be maintained for long periods of time^{5,6}. However, they have not been comprehensively shown to progress beyond stages that are equivalent to mid-fetal cortical development^{3,4,7,8}, and most organoid cultures have not been maintained for prolonged periods of time *in vitro*^{3,9}.

We previously developed a directed method of differentiation of human pluripotent stem cells in suspension that does not involve embedding into matrices⁵. This approach generates dorsal forebrain organoids named human cortical spheroid (hCS) with high reliability that can be cultured for more than 20 months progressing from neurogenesis to astrogenesis^{6,10}. However, no systematic, unbiased functional analysis has been performed to demonstrate maturation matching perinatal or postnatal stages^{3–5,7,8}. Reaching these late stages is essential to model neurodevelopmental, neuropsychiatric and neurodegenerative disorders. Moreover, it is not known whether there are intrinsic programs that underlie important physiological transitions during development and maturation, such as N-methyl-D-aspartate (NMDA) isoform shifts and RNA editing, neither of which have been studied in 3D or two-dimensional (2D) human stem-cell-based models to date^{3–5,7,8}.

Here, we leverage the ability to maintain human cortical organoids in long-term cultures and perform functional genomic assays to rigorously assess correspondence to *in vivo* developmental progression and maturation. We demonstrate substantial parallels between *in vitro* and *in vivo* neurodevelopment at the epigenetic and transcriptomic levels, as well as preservation of known physiological transitions, including NMDA receptor signaling. By mapping risk genes onto gene expression trajectories across development in this system, we also provide a searchable resource (Gene Expression in Cortical Organoids, GECO) to facilitate the choice of appropriate *in vitro* timepoints for modeling specific brain disease-causing genes.

Results

We cultured hCS differentiated from six hiPSC lines derived from five different individuals for up to 694 days *in vitro* (summarized in Extended Data Fig. 1a and Supplementary Tables 1 and 2) using a protocol that yields highly reliable hCS across experiments and cell lines^{5,11}. In total, we collected 62 samples for RNA sequencing (from four individuals, five hiPSC lines) and 50 samples for DNA methylation (from five individuals, six hiPSC lines) at 13 timepoints (Extended Data Fig. 1a and Supplementary Tables 1 and 2). Due to the challenging nature of these long-term cultures, samples were collected at various timepoints up to the following maximum differentiation day for each line: 0307-1, 280 days; 1205-4, 587 days; 2242-1, 652 days; 8858-1, 694 days; 8858-3, 652 days and H2096, 371 days.

¹Department of Neurology, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ²Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. ³Stanford Brain Organogenesis, Wu Tsai Neurosciences Institute, Stanford University, Stanford, CA, USA. ⁴Department of Psychiatry, University of California San Diego, San Diego, CA, USA. ⁵Department of Integrative Biology, University of California Los Angeles, Los Angeles, CA, USA. ⁶Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA. ⁷Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA, USA. ⁸Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA. ⁹Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA. ¹⁰Institute for Quantitative and Computational Biology, University of California Los Angeles, Los Angeles, CA, USA. ¹¹Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ¹²Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA.
✉e-mail: spasca@stanford.edu; dhg@mednet.ucla.edu

Methylation and transcriptional maturation of hCS. To assess whether hCS maturation over a long period of in vitro differentiation paralleled in vivo development, we first analyzed the epigenetic clock¹², which predicts chronological age based on the methylation status of CpG islands. We note that the epigenetic clock is unable to predict the precise culture age based on in vivo methylation, as the epigenetic clock is not calibrated for the tissue, or for the newer array type that we used^{13,14}. Nevertheless, we observed a highly significant, monotonic correlation between the length of differentiation of the hCS in vitro and their predicted methylation age (coefficient of correlation, $r = +0.76$, $P = 1.57 \times 10^{-10}$; Fig. 1a), consistent with continuous and progressive maturation of these cultures over time.

To predict culture age more precisely using an independent genome-wide approach, we next applied a validated framework that permits unbiased, quantitative statistical comparison of gene expression during cortical maturation in vitro to the BrainSpan dataset, which represents an in vivo reference for cortical development^{15–17}. To assess the quality of the data, we first sought to verify whether the main driver of variance of gene expression in the system is the time of in vitro differentiation rather than other less relevant covariates (for example batch, individual or line). Principal component (PC) analysis of gene expression revealed that the top five PCs, which together account for 57.1% of variability, were all associated with differentiation time (Fig. 1b and Extended Data Fig. 1b). Hierarchical clustering also showed that the samples clustered by differentiation day and not by other covariates (Extended Data Fig. 1c). Variance partitioning further demonstrated that the largest driver of variance was time of differentiation (median variance explained 29.2%), while the median value of variance explained for differentiation and cell line was <0.01% (Extended Data Fig. 1d). Reproducibility between samples from the same timepoint was high, both between (mean Spearman correlation 0.96, range 0.88–0.98) and within (mean Spearman correlation 0.95, range 0.88–0.98) individuals, similar to what we have previously reported^{11,18}. We note that, as differentiation progressed, this high correlation between lines decreased slightly from 0.96 at day 25 (range 0.95–0.98) to 0.92 at day 600 (range 0.88–0.95) (Fig. 1c).

Cellular stress pathways (that is endoplasmic reticulum (ER) and glycolysis pathways) have previously been proposed to be upregulated during development of in vitro 3D organoids^{19,20}. We were able to detect moderate expression levels of genes in these pathways both in vivo in BrainSpan, as well as in vitro (Extended Data Fig. 2a,b). We also detected ubiquitous expression of genes in these pathways in a large single-cell dataset consisting of 40,000 cells collected from in vivo developing human cerebral cortex, consistent with the BrainSpan results²¹ (Extended Data Fig. 2c). Although we do observe that stress pathway genes are expressed at slightly higher levels in vitro than in vivo, they follow a flat trajectory over time (Extended Data Fig. 2a,b). This is more consistent with a homeostatic state, and the absence of progressive stress or dysfunction.

We next compared changes in gene expression during the maturation of hCS to transcriptome changes observed in cortical development in vivo in humans^{16,17} using transition mapping—a quantitative method based on the rank-rank hypergeometric test¹⁵. At early timepoints (i.e., before 250 days in culture), hCS mapped to prenatal stages (stages 3–7), consistent with the observed developmental progression in vitro^{5,11} (Fig. 1d,e). By day 250, we started to observe a postnatal signature that became more evident by day 300. Between day 250 and day 300, hCS displayed a similar overlap with both prenatal and postnatal stages, whereas after day 300 they showed increasing overlap with postnatal stages (stages 8–13) (Fig. 1d,e). Based on this analysis, the predicted transition between prenatal and postnatal stages occurs around 250–300 days (~8–10 months) of in vitro differentiation.

Gene expression and gene network correspondence. Seeing this strong overall correspondence between in vitro and in vivo transcriptomes, we next compared known biological processes occurring during the maturation of hCS to those occurring during in vivo cortical development. Gene set enrichment of differentially expressed genes spanning prenatal stages in vitro (between day 25 and day 200; 1,940 downregulated genes, 2,122 upregulated genes at FDR < 0.05; Extended Data Fig. 3a and Supplementary Table 3) showed that early embryonic proliferation and developmental processes were downregulated, while neuronal and synaptic-related genes were upregulated, analogous to what was reported in vivo¹⁶ (Extended Data Fig. 3b and Supplementary Table 4). In contrast, in vitro stages that were similar to early postnatal stages (between day 200 and day 400; 992 downregulated genes and 1,369 upregulated genes at false discovery rate (FDR) < 0.05; Extended Data Fig. 3a and Supplementary Table 3) showed enrichment for processes related to translation and cortical neuronal development, including downregulation of forebrain generation of neurons, whereas terms associated with non-neuronal cell development and synaptic function (for example, regulation of long-term synaptic potentiation and neurotransmitter metabolic process) were upregulated, again corresponding to known in vivo processes¹⁶ (Extended Data Fig. 3b and Supplementary Table 4).

These parallels of in vivo biological processes being preserved in vitro were also observed when examining the expression trajectories of markers for the major hCS cell types, including progenitors, layer-specific cortical neurons and developing and maturing astrocytes (Fig. 2a–c). The expression of these cell markers followed trajectories similar to those found in vivo (Extended Data Fig. 3c), with radial glia markers peaking earliest, followed by neuronal markers and subsequently mature astrocyte markers reaching their highest level later, at around 350–400 days, when cultures start to resemble postnatal stages (Fig. 2a–c and Extended Data Fig. 3d). It is important to note that the downregulation of neuronal markers at late stages of differentiation, in both hCS and in vivo, may be due to the increase in the proportion of astrocytes, rather than an absolute reduction in the expression of these genes.

Notably, in agreement with previous reports showing that oligodendrocyte precursor cells (OPCs), GABAergic neurons and microglia were not present, or present in low proportion in hCS^{11,22,23}, we found that the expression trajectories of markers for these cell types either did not follow cortical in vivo trajectories (i.e., *GAD1* and *PLP1*) or were not detectable (i.e., *AIF1*, *ITGAM*) in hCS (Extended Data Fig. 3e). Markers for neuronal activity were only partially preserved in hCS (i.e., *NPAS4* and *ARC*; Extended Data Fig. 3f), which is likely due to lack of typical physiological inputs.

To provide a more refined view of the trajectories of specific biological processes, we leveraged co-expression modules defined from in vivo brain development to directly examine the trajectories of in vivo processes in hCS¹⁵. These in vivo modules were previously annotated based on enrichment for processes associated with specific cell types¹⁵ (Fig. 2d). We verified that in vitro modules significantly overlapped with these in vivo modules (Extended Data Fig. 4). As seen for single genes, the trajectories of these modules followed the expected in vivo sequence. Namely, the neuronal modules peaked at times matching prenatal stages (100–300 days in vitro; Fig. 2e) matching what is seen in vivo^{15,16}. The glial processes were upregulated at early stages (day 25; Fig. 2e) corresponding to the proliferation of radial glia in vivo¹⁶, and then again at postnatal stages (400–600 days of differentiation; Fig. 2e) corresponding to the development and maturation of astrocytes in vivo⁶. Overall, these analyses demonstrate that the in vivo gene expression trajectories are paralleled during long-term in vitro hCS differentiation^{11,22}.

Preservation of RNA editing. RNA editing, a major RNA processing mechanism, is dynamically regulated during in vivo brain development²⁴. RNA editing has been shown to display developmental

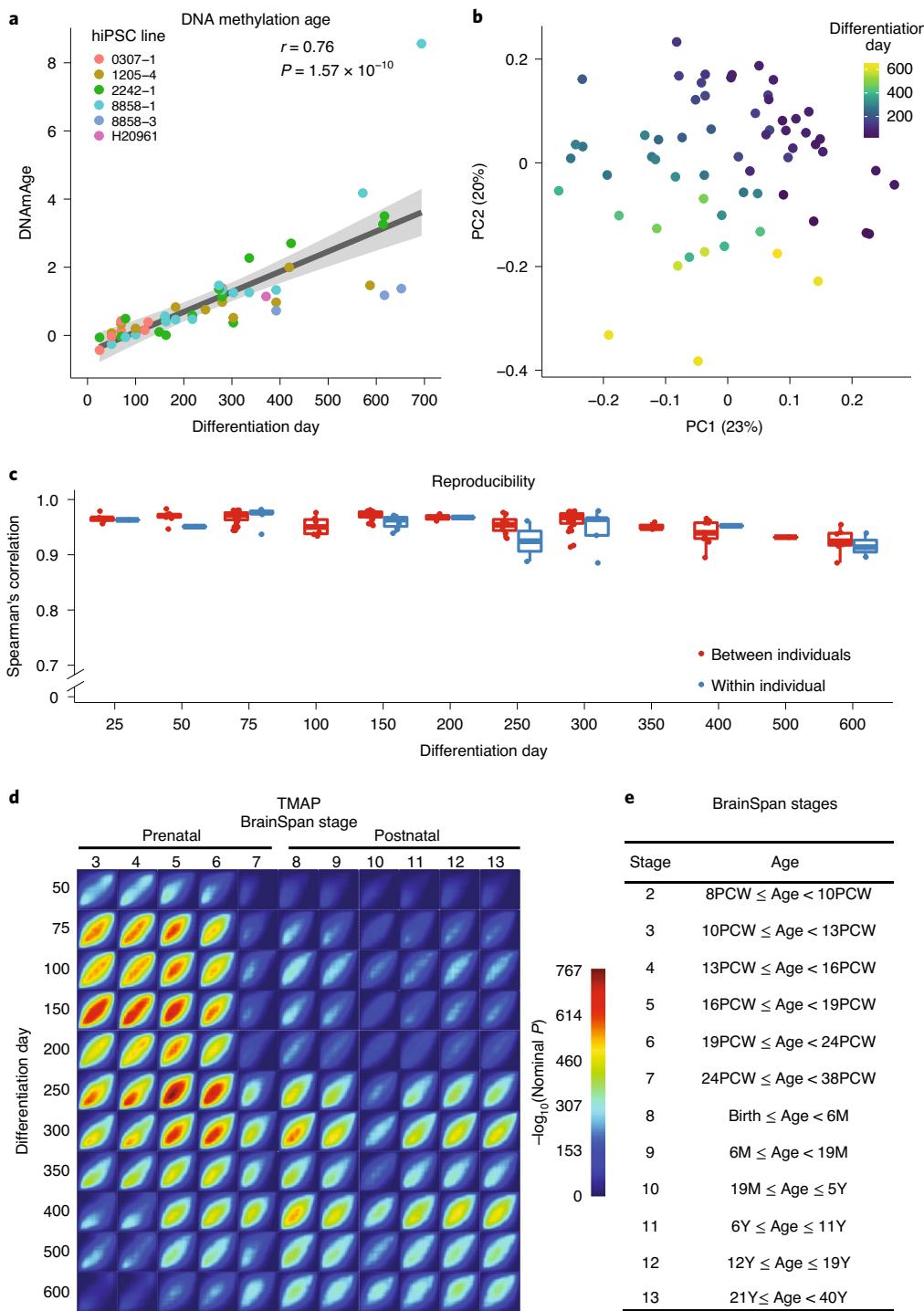


Fig. 1 | Methylation and transcriptional maturation in long-term hCS. **a**, The predicted methylation age (DNAmAge) of hCS is monotonically correlated with the in vitro differentiation day ($r = +0.76$, $P = 1.57 \times 10^{-10}$, two-sided Pearson correlation test, $n = 50$ from six hiPSC lines derived from five individuals). Colors denote individual hiPSC lines. The shaded gray area represents the 95% confidence interval. **b**, Scatterplot of the first two PCs of gene expression data. Color represents differentiation day. Numbers in brackets on axis titles are the percent of variance explained by the PC. **c**, Spearman's correlation of gene expression between samples from the same timepoint that were derived either from different individuals (red) or from the same individual (blue) ($n = 6$ samples from five hiPSC lines derived from four individuals). Boxplots: center, median; lower hinge, 25% quantile; upper hinge, 75% quantile; whiskers extend to $\pm 1.5 \times$ interquartile range. **d**, Transition mapping (TMAP) of gene expression of hCS (compared to differentiation day 25) and human primary tissue from the BrainSpan dataset (compared to stage 2). **e**, BrainSpan stages and corresponding age. PCW, post conception weeks; M, months; Y, years.

trajectories that vary with maturation in vivo and are dependent on neuronal activity²⁵. Recently, disruption of RNA editing was suggested to play a role in autism spectrum disorder (ASD), further

supporting its importance in brain function and dysfunction²⁶. Despite its importance in brain development, global patterns of RNA editing have not been explored in hiPSC-derived brain organoids.

To assess editing in vivo, we identified two modules of coordinated RNA editing in the in vivo BrainSpan data¹⁶: BSeditM1, which increased in editing rates during development, especially in late prenatal and early childhood (stages 6–9) and BSeditM2, which was slightly decreased in editing rates during development (Fig. 3a). Both in vivo modules were moderately preserved in the in vitro data ($Z_{\text{summary}}^{\text{BSeditM1}} = 9.3$, $Z_{\text{summary}}^{\text{BSeditM2}} = 5.3$) (Fig. 3b), showing preservation of in vivo RNA editing processes in hiPSC-derived brain organoids.

We next examined the expression of the major RNA editing enzymes and regulators. In vivo, *ADAR* expression increases postnatally, *ADARB1* expression increases prenatally, followed by relatively stable expression postnatally, and *FXR1* expression decreases prenatally²⁶; these followed a similar trend in vitro (Fig. 3c). Expression levels of *FMR1* increased prenatally both in vivo and in hCS. However, the subsequent postnatal increase in expression seen in vivo was not observed in hCS (Fig. 3c). Immunohistochemistry confirmed the transcript level analysis and suggested that *FXR1* predominantly localized to ventricular zone (VZ) areas and appeared downregulated at later stages in neurons (Fig. 3d).

To test whether *ADAR*, *ADARB1* and *FXR1* potentially drive RNA editing in vitro, we identified editing modules in hCS (Fig. 3e) and correlated the module eigenvalues with the expression of the editing enzymes and regulators. We found that one module, hCSeditM2, was significantly correlated with the expression of the editing related genes *FXR1* ($r = -0.32$, $\text{FDR} = 0.04$) and *ADAR* ($r = +0.60$, $\text{FDR} = 5.1 \times 10^{-6}$) (Fig. 3f). Interestingly, *FXR1* has previously been shown to inhibit editing by interacting with *ADAR*²⁶, and this interaction regulates RNA editing sites dysregulated in ASD²⁶. hCSeditM2 also significantly overlapped with the temporally increasing in vivo module BSeditM1 (odds ratio (OR) = 1.8, $\text{FDR} = 2.3 \times 10^{-4}$; Extended Data Fig. 5a). The hCSeditM2 module eigengene increased in expression over stages matching prenatal in vivo development (before 250 days) and plateaued at stages resembling postnatal periods (after 350 days) (Fig. 3e), very similar to the trajectory of the increasing in vivo module BSeditM1 (Fig. 3a).

To further support their potential to regulate RNA editing, we tested whether FMRP and FXR1P locally bind to the mRNA in close proximity to the editing sites of the genes in the module by integration with enhanced cross-linking and immunoprecipitation (eCLIP) data from the FMRP and FXR1P proteins in the adult human brain²⁶. We observed that the editing sites that we identified in both the hCS and the BrainSpan modules have highly significant proximity to both FMRP and FXR1P binding sites (Fig. 3g and Extended Data Fig. 5b). These results support the claim that FMRP and FXR1P regulate brain-related editing modules through proximity-mediated interactions²⁶. To investigate whether the target editing sites of FMRP or FXR1P were similar between BrainSpan and hCS modules, we measured the overlap of editing sites within 1 kb of a CLIP site of FMRP or FXR1P. We found that FXR1P targeted editing sites significantly overlapped between BSeditM1 and hCSeditM1 (OR = 7.71, $\text{FDR} = 6.18 \times 10^{-57}$) and hCSeditM2

(OR = 2.16, $\text{FDR} = 5.31 \times 10^{-4}$), which also showed a conserved trajectory, increasing over differentiation time in vitro (Extended Data Fig. 5c). Taken together, the methylation, gene expression and RNA editing data paint a coherent picture, indicating that hCS reach a state of postnatal maturation between 250 and 300 days.

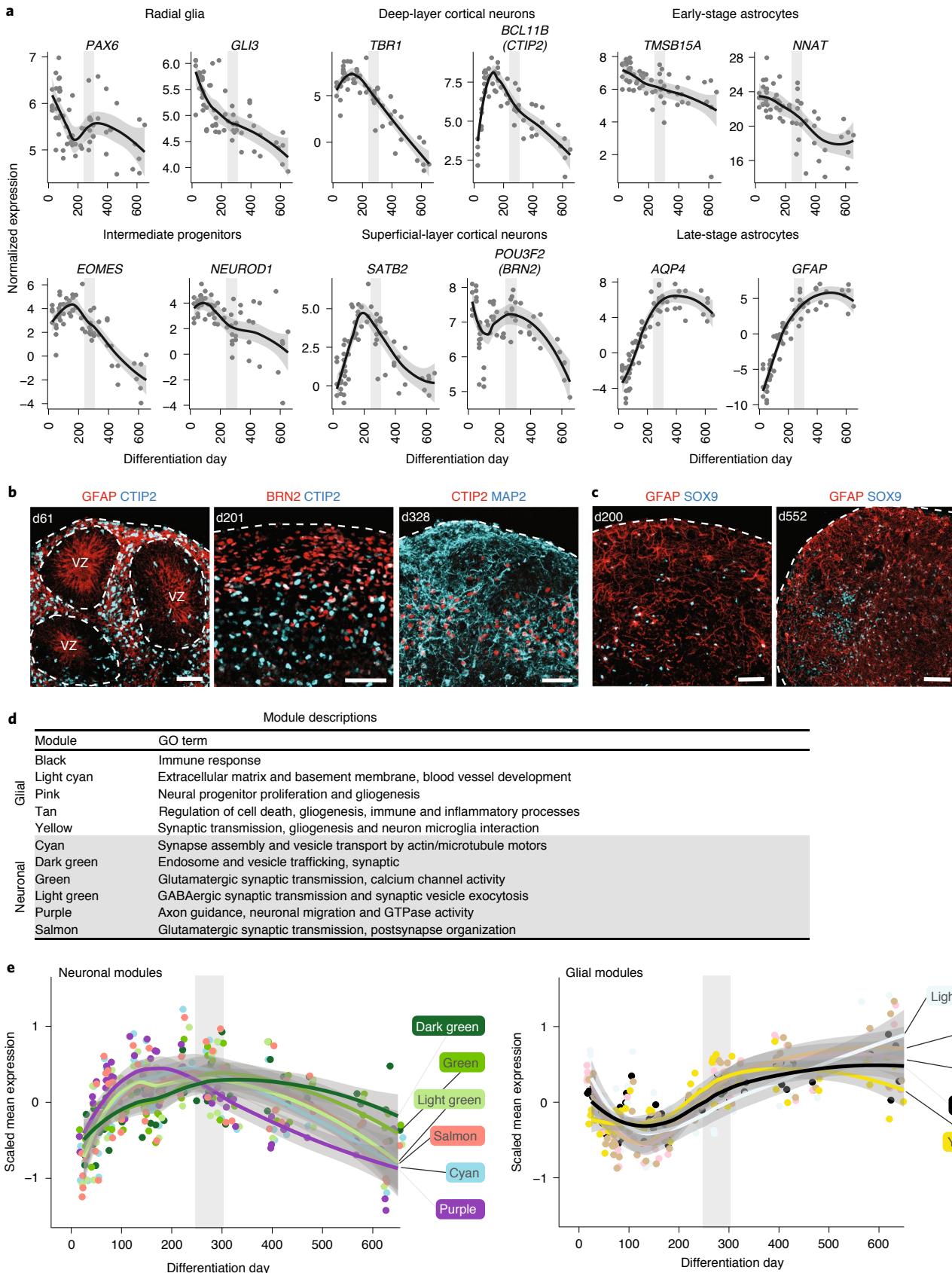
Canonical isoform switching in development. To further validate known transitions that occur with development and maturation, we assessed isoform switches in specific protein complexes related to histone modifying complexes and neurotransmitter signaling that characterize the transition from prenatal to postnatal stages of brain development^{27–29}. One canonical switch is in the histone deacetylation (HDAC) complex, in which the more highly expressed isoform *HDAC2* diminishes and isoforms *HDAC1* and *HDAC11* increase in expression^{27,30} (Fig. 4a), which plays a role in fate specification and synapse maturation²⁷. Indeed, we observed a switch in the HDACs, with an increase in *HDAC1* and *HDAC11* and a concomitant decrease in *HDAC2* in hCS. We note that, while *HDAC1* is downregulated postnatally in vivo, it remains upregulated in hCS (Fig. 4a). We speculate that perhaps cell types not present in hCS, or activity-dependent processes that occur in vivo³¹ may contribute to changes in HDAC enzyme expression in vitro at later stages. We also assessed whether these enzymes show cell-type-specific enrichment, which could contribute to the differences in postnatal expression. Using immunohistochemistry, we observed that *HDAC2* was expressed in both progenitors and neurons, although it appeared more highly expressed in progenitors, consistent with its downregulation over time (Fig. 4b and Extended Data Fig. 6a). Additionally, using a mid-fetal single-cell dataset²¹, we found that *HDAC1*, *HDAC2* and *HDAC11* did not show significant cell-type-specific enrichment at that timepoint (Extended Data Fig. 6b), which is more equivalent to the later timepoint in vitro (day 131).

Another critical neurodevelopmental switch is the change in the NMDA receptor subunits from *GRIN2B* (also known as *NR2B* or *GluN2B*) to *GRIN2A* (also known as *NR2A* or *GluN2A*) (Fig. 4c)²⁸ and from *GRIN2D* to *GRIN2C*²⁹ (Fig. 4c). The progressive change in subunit expression results in a switch around birth²⁸ and has a profound effect on channel function, including how it responds to allosteric modulators³². In vivo, we observed the subunit shift at the transcriptional level occurring soon after birth (BrainSpan stage 8; 0–6 months after birth; Fig. 4c). Interestingly, in hCS we observed a switch in NMDA receptor subunits at the time when cultures are predicted to transition from prenatal to postnatal stages based on their transcriptomic patterns (day 250–300 of differentiation; Fig. 4c). Using western blotting for *GRIN2A* and *GRIN2B* in hCS from day 51 to day 408, we found that the level of *GRIN2A* appeared to increase with time, and the level of *GRIN2B* seemed to peak at 200–250 days and decreased at later timepoints (350–400 days; Fig. 4d,e and source data). The protein trajectories mirrored the trajectories seen at the RNA level (Fig. 4c), with the switch between the two subunits occurring between 250 and 300 days (Fig. 4c–e).

Fig. 2 | Biological processes and cell-type marker changes in long-term hCS. **a**, Normalized expression (\log_2) of marker genes for neurons, intermediate progenitors, astrocytes and radial glia as well as superficial and deep layer neurons. Neuronal and intermediate progenitor markers are initially expressed at high levels and decrease after day 250. Astrocyte markers increase in expression with time and peak after day 300. Radial glia markers decrease in expression as hCS advance in differentiation. **b**, Immunohistochemistry of progenitors and neuronal markers at day 61 (d61; line 0524-1), day 201 (d201; line 8858-1) and day 328 (d328; line 2242-1) showing expression of glial fibrillary acidic protein (GFAP) in VZ-like regions and the deep and superficial layer markers CTIP2 (also known as BCL11B) and BRN2 (also known as POU3F2). **c**, Immunohistochemistry for the astrocyte markers GFAP and SOX9 at day 200 (d200; line 2242-1) and day 552 (d552; line 8858-1). Immunohistochemistry experiments were performed twice with similar results (1–3 hCS per line from at least two hiPSC lines were included). Scale bars, 50 μm (**b**, **c**). **d**, Groups and GO term annotations of in vivo WGCNA modules performed by Stein et al¹⁵. **e**, Scaled mean expression of neuronal and glial module genes. The neuronal modules peaked at ~day 200; the glial modules decreased in expression until about differentiation day 150 and then increased in expression to peak around day 500. In **a** and **e** the shaded gray areas around the trajectory lines represent the 95% confidence intervals and the vertical grays denote the shift from prenatal to postnatal gene expression based on matching to in vivo patterns. In **a** and **e**, $n = 62$ samples from five hiPSC lines derived from four individuals.

This switch, which results in a greater contribution of the GRIN2A subunit to the NMDA complex, is associated with changes in the functional properties of the channel³². To verify that this

results in physiological changes in hCS neurons, we measured the magnitude of NMDA receptor activity at early (days 54–156 of differentiation; GRIN2B predominant) and late (days 307–523 days of



differentiation; GRIN2A predominant) developmental timepoints. We performed voltage clamp recording of neurons in hCS slices while applying pulses of NMDA to measure the summation of the NMDA response throughout the cell independent of subcellular receptor localization (Fig. 4f). We found that the magnitude of NMDA responses increased over time *in vitro* ($r=+0.63$, $P=6.94\times 10^{-4}$; Fig. 4g). Importantly, there was a significant negative association between the reduction in total NMDA response after application of the GRIN2B-specific antagonist ifenprodil (IFN) and the total time of differentiation mirroring the reduction seen in the GRIN2B subunit (beta logit regression $B=-0.003$, $P=1.58\times 10^{-3}$) (Fig. 4h). This demonstrates that changes observed in hCS at the RNA and protein level result in physiological changes that are consistent with the presence of a greater proportion of GRIN2B-enriched NMDA receptors at early compared to later developmental timepoints.

Disease risk gene trajectories. One of the goals of these model systems is to guide a biological understanding of brain disorders. We therefore mapped genes associated with ASD (Fig. 5a), intellectual disability (ID; Fig. 5b) and schizophrenia (SCZ; Fig. 5c) onto these gene expression data to see if there were specific expression patterns associated with subsets of risk genes and whether they were shared between disorders.

We performed unsupervised hierarchical clustering of disorder-associated genes based on the temporal expression patterns in hCS, which identified clusters representing distinct temporal trajectories in each disorder (Fig. 5 and Supplementary Table 5). Annotation with gene ontology (GO) and a cell type enrichment analyses (Methods) revealed that each of these clusters represent different developmental trajectories, as well as specific biological processes and cell types (Fig. 5).

ASD and SCZ risk genes each clustered into five developmental trajectories, whereas ID genes formed four clusters. We found three trajectory patterns shared across disorders. One trajectory seen in ASD-C3, ID-C4 and SCZ-C2 peaked at around 150 days of differentiation (Fig. 5). These clusters were all enriched for both excitatory and inhibitory neuronal genes, as well as for synaptic related GO terms, such as modulation of chemical synaptic transmission (ASD-C3), synapse organization (ID-C4) and calcium ion transmembrane transport (SCZ-C2) (Fig. 5). Another shared trajectory was seen in ASD-C2 and ID-C1, which were highly expressed at the peak of neurogenesis (<150 days) and gliogenesis (>300 days) (Fig. 5a,b). These clusters were enriched for progenitor cell type genes (radial glia and intermediate progenitors) and for histone modification and covalent chromatin modification GO terms, and they were not conserved in SCZ. Although SCZ-C3 showed a similar trajectory to ASD-C2 and ID-C1, it was not enriched for any cell type, but was enriched for protein translation-related GO terms, such as aminoacyl-tRNA ligase activity (Fig. 5c). The third shared trajectory peaked in expression at later stages of differentiation (>250 days) and was found in ASD-C5, ID-C3 and SCZ-C1. These clusters were all enriched for astrocyte genes; however, they did not share common biological processes across

disorders, indicating that different pathways may be impacted (Fig. 5). These three trajectories were also seen in genes associated with epilepsy (Extended Data Fig. 7a). Most of the epilepsy genes (62%) were found in the cluster epilepsy-C3 that peaked at day 150. This cluster was enriched for GO terms related to ion channel activity and for excitatory and inhibitory neuronal genes. However, a substantial proportion of genes (20%; epilepsy-C2) peaked at much later stages of differentiation (>250 days), and these were not enriched for any cell type (Extended Data Fig. 7a).

We next mapped genes associated with two neurodegenerative disorders: Alzheimer's disease (AD; Fig. 6a) and Parkinson's disease (PD; Fig. 6b). We found that the majority of clusters (3 out of 4 in AD and 3 out of 4 in PD) showed peak levels of expression at late timepoints (>250 days), corresponding to predicted postnatal stages. One of these late-peaking clusters in AD (AD-C1) contained the familial Alzheimer's-associated genes *APOE*, *APP* and *PSEN2*, and was associated with amyloid-beta formation—a hallmark of AD (Fig. 6a). In PD, genes associated with Mendelian forms of the disease, such as *PRKN*, *UCHL1*, *SNCA*, *PARK7*, *PINK1* and *LRRK2*, were all found in clusters that peaked later in differentiation (>250 days; PD-C1 and PD-C2; Fig. 6b). Genes associated with two other related neurodegenerative diseases—progressive supranuclear palsy (PSP) and frontotemporal dementia (FTD)—formed two clusters that also peaked late in differentiation (Extended Data Fig. 7b).

At least one cluster in each disorder peaked in expression at later stages of neural differentiation (>250 days), which emphasizes the need for long-term maturation to study the role of these disease-related genes. These clusters can guide the selection of appropriate timepoints and cell types to model specific disorders with hiPSC-derived neural cultures. Genes in some of the clusters, such as ASD-C3, ID-C4, SCZ-C2, epilepsy-C3, AD-C2 and PD-C3, could be used in early stage hCS models, whereas genes in clusters ASD-C5, ID-C3, SCZ-C1 and AD-C1, epilepsy-C2, AD-C1, PD-C2 and FTD/PSP-C1, may require cultures beyond 250 days *in vitro*. To allow detailed comparison between *in vivo* and hCS gene trajectories, we provide a webtool named GECO that allows the concomitant examination of gene trajectories in hCS and in BrainSpan (<https://labs.dgsom.ucla.edu/geschwind/files/view/html/GECO.html>).

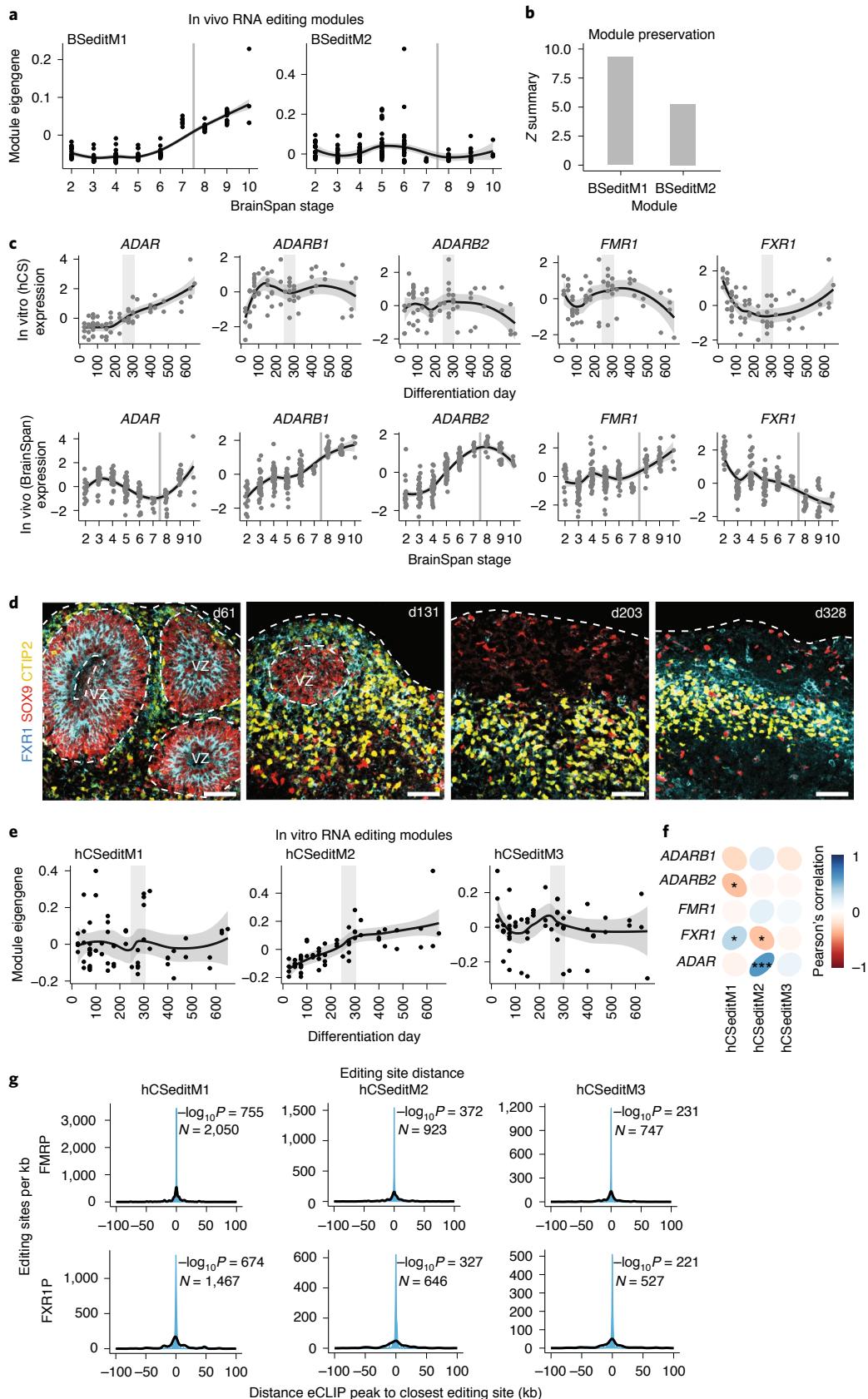
Discussion

We performed multiple independent analyses of functional genomic features to comprehensively assess to what extent *in vitro* hCS recapitulate *in vivo* cortical maturation. We observe substantial convergence across these different data types, which indicates attainment of early postnatal maturation between 250 and 300 days *in vitro*. To our knowledge, this is the first multi-level demonstration that an *in vitro* model of human neural development has matured sufficiently to acquire some postnatal features. This includes several important features of RNA editing and a major physiological transition involving the switch in the ratio of NMDA receptor subunits that occurs postnatally in mammals²⁸. Our results suggest that hCS mature both at the level of individual cells, as evident by the NMDA receptor isoform switch, as well as some aspects of cellular

Fig. 3 | RNA editing in hCS. **a**, Trajectories of *in vivo* (BrainSpan) RNA editing modules. **b**, Preservation scores (Z summary) of the *in vivo* RNA editing modules in hCS. **c**, Trajectories of RNA editing enzymes in hCS (top) and *in vivo* from BrainSpan (bottom). **d**, Immunohistochemistry of the RNA editing regulator FXR1 with the glial and neuronal markers GFAP and CTIP2 (also known as BCL11B) at day 61 (d61; line 0524-1), at day 131 (d131; line 1205-4), at day 200 (d203; line 1205-4) and at day 328 of differentiation (d328; line 2242-1). Scale bars, 50 µm. Immunohistochemistry experiments were performed once for d61, twice for d131 and d203 and three times for d328 (1–3 hCS per line from at least two hiPSC lines were included). **e**, Trajectories of the three hCS RNA editing modules. **f**, Correlation of module eigenvalues with the expression of the major known RNA editing enzymes and regulators. **g**, Distributions showing the closest distances between editing sites from hCS editing modules and FMRP or FXR1P eCLIP peaks (blue). The median of 10,000 sets of control sites (black) is depicted for comparison. See Methods for details of permutation-based two-sided P value calculation. N indicates the number of editing sites shown. *FDR < 0.05, **FDR < 0.005. In **a**, **c** and **e**, the shaded gray area around the trajectory represents the 95% confidence interval, vertical gray lines represent birth, and vertical gray bars denote the shift from prenatal to postnatal gene expression based on matching to *in vivo* patterns. In **c** (top row) and **e**, $n=62$ samples from five hiPSC lines derived from four individuals. In **a** and **c** (bottom row), $n=196$ from 24 individuals.

composition, as shown by the emergence of superficial layer neurons and astrocytes at later stages. It is important to note that some changes in gene expression are likely due to changes in cell proportions.

We also show that genes associated with neurodevelopmental and neurodegenerative disorders are not monolithic in their expression, but fall into distinct development trajectories. These trajectories



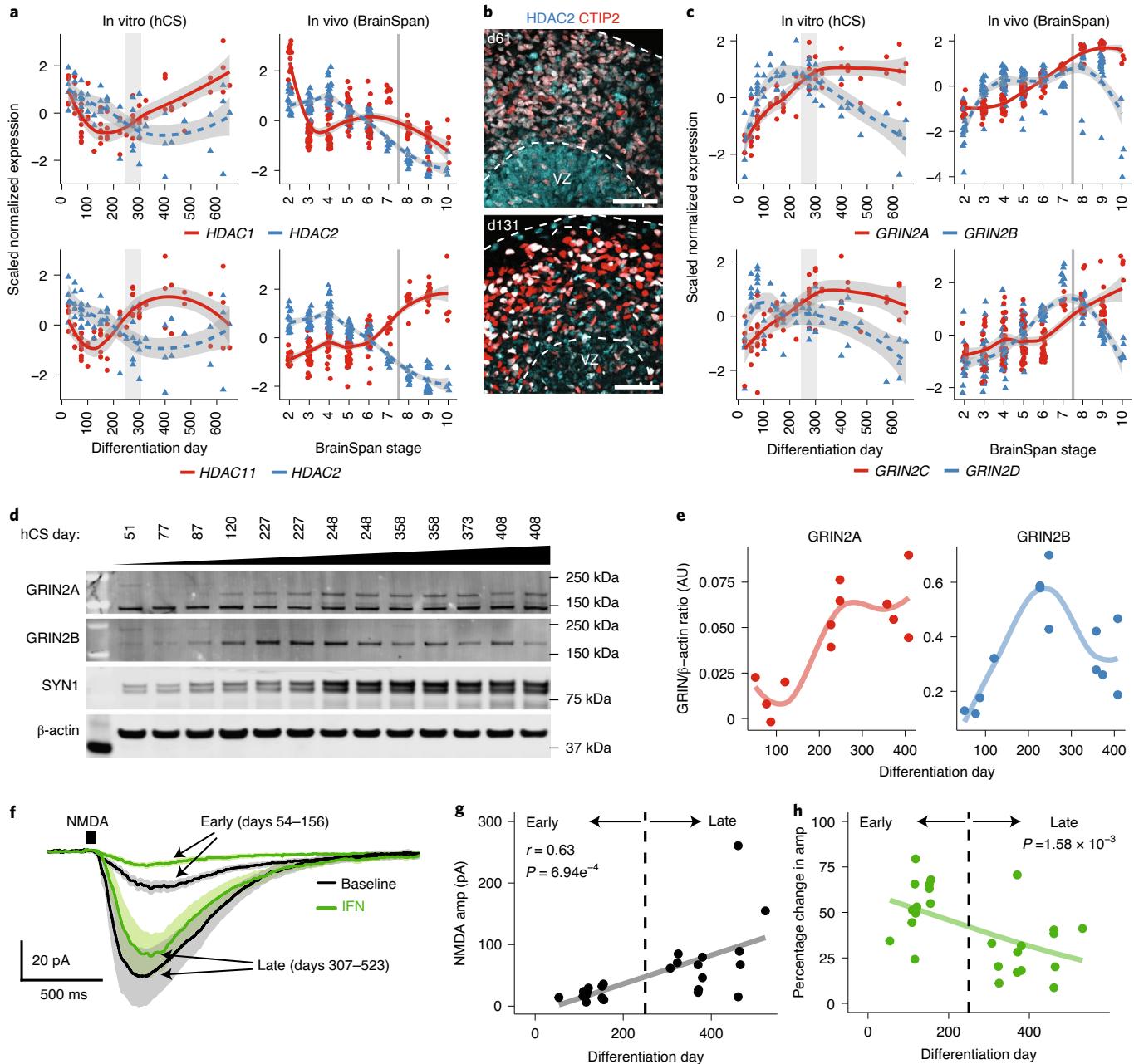


Fig. 4 | Developmental isoform switches in hCS. **a**, Expression trajectories of HDAC subunits. In vivo (right) and in vitro (left), HDAC expression decreases while the expression of both HDAC1 (top) and HDAC11 (bottom) increases. **b**, Immunohistochemistry for HDAC2 and deep layer marker CTIP2 (also known as BCL11B) at day 61 (d61; line 0524-1) and day 131 (d131; line 1205-4). Scale bars, 50 μ m. Immunohistochemistry experiments were performed once for d61 or twice for d131 (1–3 hCS per line from at least two hiPSC lines were included). **c**, Expression trajectories of NMDA receptor subunits. In vivo (right) and in vitro (left), GRIN2A (NR2A) and GRIN2B (NR2B) (top), as well as GRIN2C (NR2C) and GRIN2D (NR2D) (bottom). In **a** and **c**, the shaded gray area around the trajectory represents the 95% confidence interval, vertical gray lines represent birth and vertical gray bars denote the shift from prenatal to postnatal gene expression based on matching to in vivo patterns. For in vitro data, $n=62$ samples from five hiPSC lines derived from four individuals; for in vivo data, $n=196$ from 24 individuals. **d**, Western blots for GRIN2A, GRIN2B and Synapsin-1; β -actin was used as a loading control. The images shown were cropped (uncropped images are included in the source data). Cell lines used are 1205-4 (samples 1, 2, 3, 4, 6, 8, 10, 11 and 13) and 0524-1 (samples 5, 7, 9 and 12). Western blot experiments were run three times with similar results. **e**, Quantification of GRIN2A and GRIN2B protein levels from **d** ($n=13$ from two hiPSC lines). AU, arbitrary units. **f**, Average whole-cell voltage-clamp recordings of NMDA responses (10 mM NMDA, 50 ms pulse) at early (days 54–156) and late (days 307–523) stages of hCS development at baseline (black) and after IFN exposure (green); s.e.m. are depicted by the gray and green lines. Neurons were identified with a fluorescent reporter (Syn1::GFP). **g**, Increased maximum NMDA response amplitudes (amp) over developmental time ($r=0.63$, $P=6.94 \times 10^{-4}$). Black line represents the linear fit of the data. **h**, Percent reduction of maximum NMDA responses by the GRIN2B-containing NMDA receptor blocker IFN (10 μ M) is significantly reduced with time. Significance was measured using a beta regression with logit link function, $B=-0.003$, $P=1.58 \times 10^{-3}$. One cell was patched per hCS for a total of 25 cells from two hiPSC lines (8858-1, 1205-4).

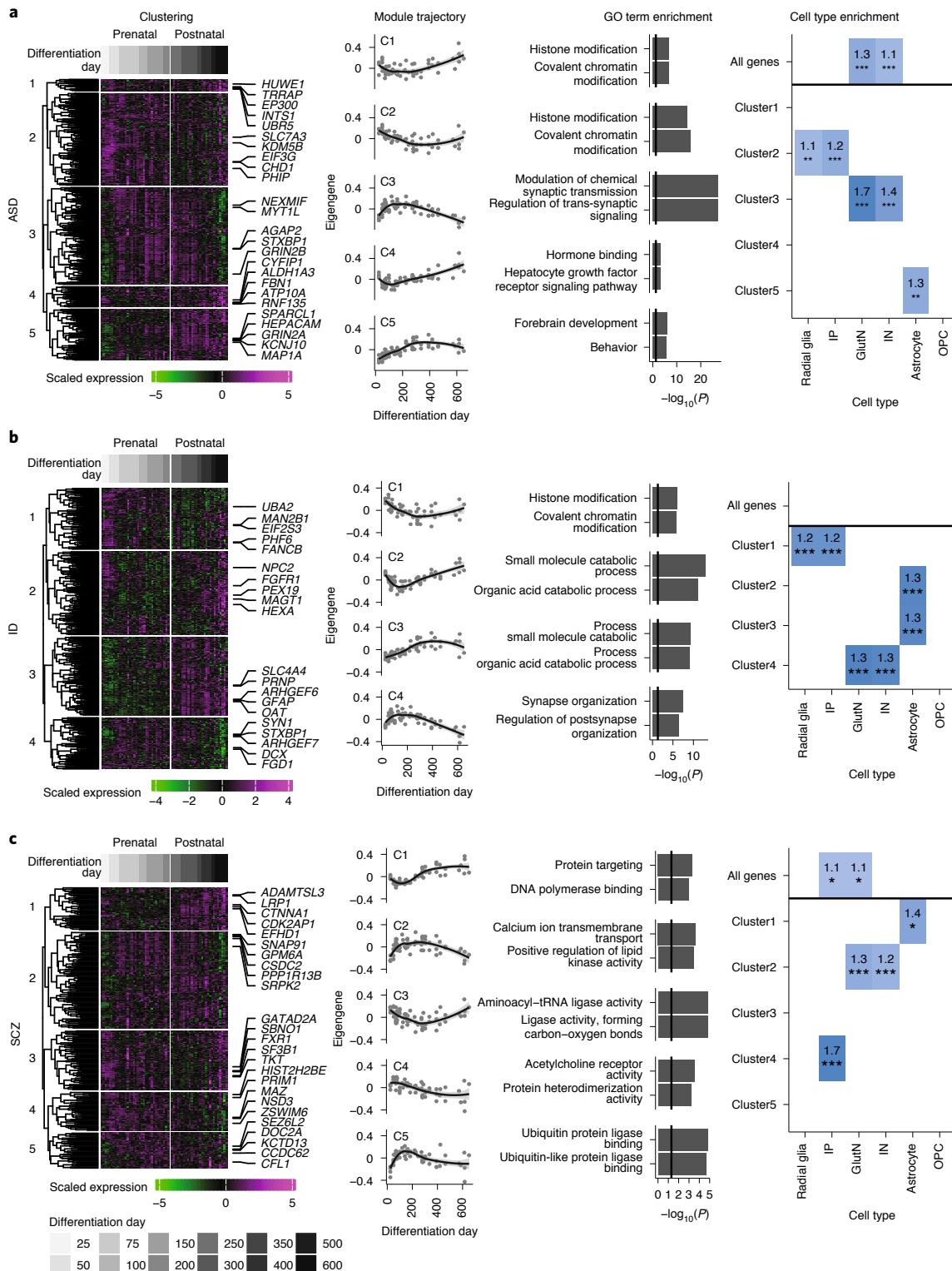


Fig. 5 | Mapping neurodevelopmental and neuropsychiatric disorder genes onto hCS differentiation. **a–c** Mapping of genes associated with ASD (a), ID (b) and SCZ (c) onto hCS differentiation trajectories. The first column shows clustering of scaled normalized expression of genes associated with a disorder. Genes (in rows) are clustered using hierarchical clustering on the Euclidean distance between genes. Samples (columns) are ordered by differentiation day (represented by gray bars), with the earliest days on the left and latest timepoints on the right. The five most representative genes (highest correlation with the cluster eigengene) are shown. The second column shows the cluster eigengenes (first PC) for the identified gene clusters. The shaded gray area around the trajectory line represents the 95% confidence interval. The third column shows the top GO terms enriched in the identified clusters. The fourth column shows cell types overexpressed in either all the genes associated with a disorder (above line) or in the genes from the identified clusters. Number and color represent the fold change. Significance was tested using a one-sided permutation test with 100,000 permutations. P values were corrected for multiple testing using the Benjamini-Hochberg (BH) method. *FDR < 0.05, **FDR < 0.01, ***FDR < 0.005; $n = 62$ samples from five hiPSC lines derived from four individuals. IP, intermediate progenitors; GlutN, glutamatergic neurons; IN, interneurons; OPC, oligodendrocyte progenitor cells.

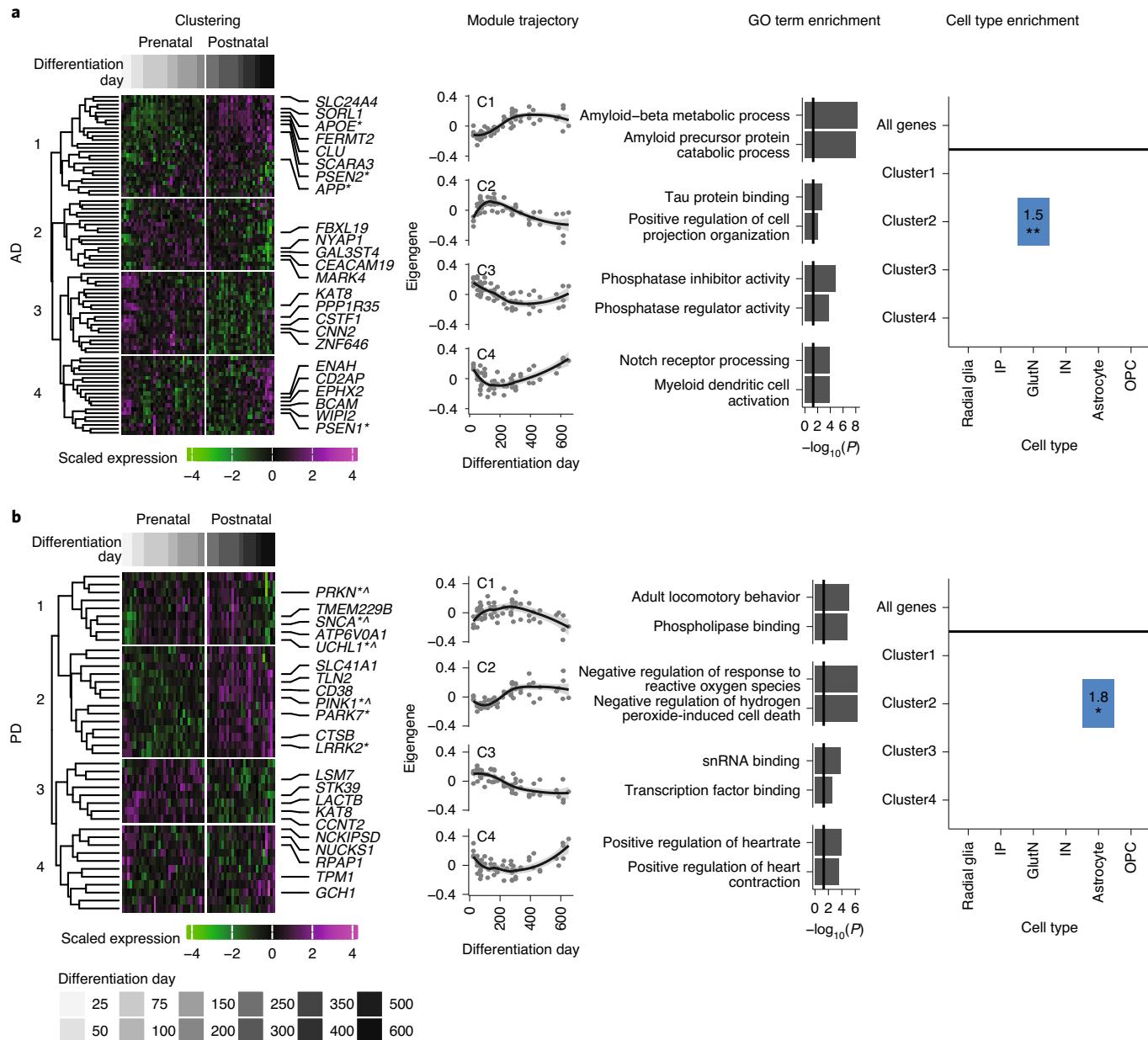


Fig. 6 | Mapping neurodegenerative disorder genes onto hCS differentiation. **a,b** Mapping of genes associated with AD (a) and PD (b) onto hCS differentiation. The first column shows clustering of scaled normalized expression of genes associated with a disorder. Genes (in rows) are clustered using hierarchical clustering on the Euclidean distance between genes. Samples (columns) are ordered by differentiation day (represented by gray bars) with the earliest days on the left and latest timepoints on the right. The five most representative genes (highest correlation with the cluster eigengene) and genes associated with familial forms of the disease are shown. Asterisks, familial gene; caret, familial gene that is also a hub gene. The second column is the cluster eigengenes (first PC) for the identified gene clusters. The shaded gray area around the trajectory line represents the 95% confidence interval. The third column is the top GO terms enriched in the identified clusters. The fourth column is cell types overexpressed in either all the genes associated with a disorder (above line) or in the genes from the identified clusters. Number and color represent the fold change. Significance was tested using a one-sided permutation test with 100,000 permutations. P values were corrected for multiple testing using the BH method. * FDR < 0.05, ** FDR < 0.01, *** FDR < 0.005. n = 62 samples from five hiPSC lines derived from four individuals. IP, intermediate progenitors; GlutN, glutamatergic neurons; IN, interneurons; OPC, oligodendrocyte progenitor cells.

include those peaking both before 100 days and after 250 days (for example, ASD-C2 and ID-C1), which represent histone modification in progenitor cells; those peaking between 100 and 150 days (for example, ASD-C3, ID-C4 and SCZ-C2), which represent synaptic structure and function in neuronal cell types; and those with late expression trajectories, which are related to astrocyte biology (for example, ASD-C5, ID-C3 and SCZ-C1). This timing

should be considered when establishing in vitro models of disease. For example, mutations in the astrocyte-related gene *HEPACAM*, which is part of the ASD-C5 cluster (Fig. 5a), should be studied at later stages of differentiation (> 250 days), while the consequences of mutations in the neuronal transcription factor *MYT1L*, which is part of ASD-C3 (Fig. 5a), can likely be probed in early stages of neural differentiation.

The majority of these cluster trajectories follow *in vivo* trajectories. One exception is the neuronal activity-dependent clusters (i.e., ASD-C3, SCZ-C2 and epilepsy-C3), which do not show strong fidelity at late stages of differentiation (>400 days) likely due to the lack of extensive network activity in hCS. We provide the GECO tool to allow detailed comparison between *in vivo* and hCS trajectories. We note that the gene trajectories presented in this webtool were derived from a directed hCS differentiation protocol, and other differentiation protocols and cell lineages from other brain regions may exhibit different timelines of maturation.

One caveat of this study is that culturing hCS for long periods of time is not trivial, leading us to collect samples as they became available. This is the reason three lines do not have samples from the entire timecourse. It is also important to note that, after 400 days of differentiation, we observed an increase in the variability of the hCS differentiation (Fig. 1c). Reducing this variability will be essential for modeling disease at these very late stages of differentiation, as will development of methods to accelerate this process.

Previous studies in brain organoids have suggested that glycolysis and ER stress are highly upregulated in these systems, reflecting a state of cell stress^{19,20}. If this were the case, we would expect to see these pathways increasing over time, reflective of progressive cellular stress *in vitro*, since cellular stress is not a homeostatic state³³. However, our analysis shows that the trajectories of these pathways remain flat during differentiation for up to 21 months *in vitro*. Moreover, we were able to detect robust levels of many ER stress and glycolysis genes *in vivo* in the BrainSpan dataset. Our ability to detect more robust *in vivo* expression of genes involved in these stress pathways may be due to higher levels of gene detection in bulk RNA sequencing compared to the single-cell RNA sequencing data used previously^{19,34}. However, we note that we were also able to detect them at similar levels in another *in vivo* single-cell dataset from fetal cortex²¹. Thus, our interpretation of these data is that the slightly higher, but relatively constant, elevation of these genes involved in glycolysis *in vitro* likely reflects the different, but homeostatic metabolic state with respect to glucose utilization that has been observed across *in vitro* cell culture systems^{35,36}.

Our findings also support the interpretation that key features of human corticogenesis are guided by an internal differentiation clock. This is consistent with findings in mouse showing that major features of the progression of cortical neurogenesis are governed by intrinsic factors and do not require extrinsic signals from other brain regions—observations that warrant further study and refinement^{37,38}. Although reaching later stages of development *in vitro* is currently time consuming (>250 days), it nevertheless extends the value of hiPSC-derived *in vitro* 3D cultures by providing a platform to study the processes occurring during late fetal and early postnatal stages of brain development. The presence of continual time-dependent aging in hCS also indicates that these cellular models could be used to model epigenetic aspects of aging, which has been shown to capture key biological features associated with aging³⁹.

It is important to note that while our model system was able to capture key features of *in vivo* human corticogenesis, some important aspects still require further investigation. Alternative splicing plays an important role in neural development⁴⁰; however, detecting significant alternative splicing will require a more deeply sequenced and extensive set of *in vivo* and *in vitro* samples. Another aspect requiring further investigation is the role of cells not born in the dorsal forebrain on the maturation of hCS. For example, ventral forebrain-derived GABAergic neurons promote synaptic maturation and network maturation³¹. These effects could be studied in forebrain assembloids, in which we have previously shown that GABAergic interneurons migrate and functionally integrate into the cortical network²². It will be important to assess how network dynamics mature in the presence of GABAergic neurons to establish the excitation–inhibition balance, which has been

linked to neurodevelopmental disorders⁴¹. Microglia, which are mesoderm-derived, also play an important role in the developmental maturation of the cerebral cortex⁴², and previous studies have shown successful integration of microglia-like cells into human brain organoids⁴³. Lastly, strategies to derive oligodendrocytes²³ or vascular-forming endothelial cells⁴⁴ in assembloids will also be useful to investigate how they modulate developmental trajectories in long-term hCS cultures.

Additionally, while we identified preservation of some aspects of RNA editing in hCS, this preservation was not complete. For instance, *ADRAB2* had a far less dynamic pattern in hCS than *in vivo*. Moreover, while the prenatal expression trajectories of *FXR1* and *FMR1* in hCS were similar to those seen *in vivo*, hCS at later stages did not track the *in vivo* trajectories. We speculate that this could be related to neuronal activity-dependent processes^{25,45}. For example, RNA editing of the AMPA receptor *GluA2* (*GRIA2* gene) is highly dependent on neuronal activity⁴⁵. Our results suggest that alterations in the balance of the RNA editing process may be used as a measure for optimizing and enhancing the functional similarities between these 3D *in vitro* models and *in vivo* brain development. Another important challenge remains to find ways to enhance this maturation speed to further facilitate more efficient *in vitro* modeling, including features that may be dependent on certain forms of neuronal activity and aging. This is especially true for modeling neurodegenerative disorders. Our approach provides a framework for comprehensive analysis of such features.

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/s41593-021-00802-y>.

Received: 13 May 2020; Accepted: 12 January 2021;
Published online: 22 February 2021

References

- Pasca, S. P. The rise of three-dimensional human brain cultures. *Nature* **553**, 437–445 (2018).
- Mertens, J., Marchetto, M. C., Bardy, C. & Gage, F. H. Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. *Nat. Rev. Neurosci.* **17**, 424–437 (2016).
- Lancaster, M. A. et al. Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
- Qian, X. et al. Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* **165**, 1238–1254 (2016).
- Pasca, A. M. et al. Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* **12**, 671–678 (2015).
- Sloan, S. A. et al. Human astrocyte maturation captured in 3D cerebral cortical spheroids derived from pluripotent stem cells. *Neuron* **95**, 779–790 e776 (2017).
- Camp, J. G. et al. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. USA* **112**, 15672–15677 (2015).
- Amiri, A., et al. Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* **362**, eaat6720 (2018).
- Quadrato, G. et al. Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* **545**, 48–53 (2017).
- Trevino, A. E., et al. Chromatin accessibility dynamics in a model of human forebrain development. *Science* **367**, eaay1645 (2020).
- Yoon, S. J. et al. Reliability of human cortical organoid generation. *Nat. Methods* **16**, 75–78 (2019).
- Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
- McEwen, L. M. et al. Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array. *Clin. Epigenetics* **10**, 123 (2018).
- Horvath, S. et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford progeria syndrome and *ex vivo* studies. *Aging (Albany NY)* **10**, 1758–1775 (2018).

15. Stein, J. L. et al. A quantitative framework to evaluate modeling of cortical development by neural stem cells. *Neuron* **83**, 69–86 (2014).
16. Li, M., et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
17. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
18. Khan, T. A., et al. Neuronal defects in a human cellular model of 22q11.2 deletion syndrome. *Nat. Med.* **26**, 1888–1898 (2020).
19. Pollen, A. A. et al. Establishing cerebral organoids as models of human-specific brain evolution. *Cell* **176**, 743–756.e717 (2019).
20. Bhaduri, A., et al. Cell stress in cortical organoids impairs molecular subtype specification. *Nature* **578**, 142–148 (2020).
21. Polioudakis, D. et al. A single-cell transcriptomic atlas of human neocortical development during mid-gestation. *Neuron* **103**, 785–801.e788 (2019).
22. Birey, F. et al. Assembly of functionally integrated human forebrain spheroids. *Nature* **545**, 54–59 (2017).
23. Marton, R. M., et al. Differentiation and maturation of oligodendrocytes in human three-dimensional neural cultures. *Nat. Neurosci.* **22**, 484–491 (2019).
24. Hwang, T. et al. Dynamic regulation of RNA editing in human brain development and disease. *Nat. Neurosci.* **19**, 1093–1099 (2016).
25. Sanjana, N. E., Levanon, E. Y., Hueske, E. A., Ambrose, J. M. & Li, J. B. Activity-dependent A-to-I RNA editing in rat cortical neurons. *Genetics* **192**, 281–287 (2012).
26. Tran, S. S. et al. Widespread RNA editing dysregulation in brains from autistic individuals. *Nat. Neurosci.* **22**, 25–36 (2019).
27. Morris, M. J., Karra, A. S. & Monteggia, L. M. Histone deacetylases govern cellular mechanisms underlying behavioral and synaptic plasticity in the developing and adult brain. *Behav. Pharmacol.* **21**, 409–419 (2010).
28. Sheng, M., Cummings, J., Roldan, L. A., Jan, Y. N. & Jan, L. Y. Changing subunit composition of heteromeric NMDA receptors during development of rat cortex. *Nature* **368**, 144–147 (1994).
29. Watanabe, M., Inoue, Y., Sakimura, K. & Mishina, M. Developmental changes in distribution of NMDA receptor channel subunit mRNAs. *Neuroreport* **3**, 1138–1140 (1992).
30. Liu, H., Hu, Q., Kaufman, A., D'Ercole, A. J. & Ye, P. Developmental expression of histone deacetylase 11 in the murine brain. *J. Neurosci. Res.* **86**, 537–543 (2008).
31. Le Magueresse, C. & Monyer, H. GABAergic interneurons shape the functional maturation of the cortex. *Neuron* **77**, 388–405 (2013).
32. Wyllie, D. J., Livesey, M. R. & Hardingham, G. E. Influence of GluN2 subunit identity on NMDA receptor function. *Neuropharmacology* **74**, 4–17 (2013).
33. Galluzzi, L., Yamazaki, T. & Kroemer, G. Linking cellular stress responses to systemic homeostasis. *Nat. Rev. Mol. Cell Biol.* **19**, 731–745 (2018).
34. Nowakowski, T. J. et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017).
35. Sunwoldt, J., Bosche, B., Meisel, A. & Mergenthaler, P. Neuronal culture microenvironments determine preferences in bioenergetic pathway use. *Front. Mol. Neurosci.* **10**, 305 (2017).
36. Ma, E. H. et al. Metabolic profiling using stable isotope tracing reveals distinct patterns of glucose utilization by physiologically activated CD8(+) T cells. *Immunity* **51**, 856–870.e855 (2019).
37. Gaspard, N. et al. An intrinsic mechanism of corticogenesis from embryonic stem cells. *Nature* **455**, 351–357 (2008).
38. Shen, Q. et al. The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. *Nat. Neurosci.* **9**, 743–751 (2006).
39. Zhang, W., Qu, J., Liu, G. H. & Belmonte, J. C. I. The ageing epigenome and its rejuvenation. *Nat. Rev. Mol. Cell Biol.* **21**, 137–150 (2020).
40. Raj, B. & Blencowe, B. J. Alternative splicing in the mammalian nervous system: recent insights into mechanisms and functional roles. *Neuron* **87**, 14–27 (2015).
41. Sohal, V. S. & Rubenstein, J. L. R. Excitation–inhibition balance as a framework for investigating mechanisms in neuropsychiatric disorders. *Mol. Psychiatry* **24**, 1248–1257 (2019).
42. Mosser, C. A., Baptista, S., Arnoux, I. & Audinat, E. Microglia in CNS development: shaping the brain for the future. *Prog. Neurobiol.* **149–150**, 1–20 (2017).
43. Lin, Y. T. et al. APOE4 causes widespread molecular and cellular alterations associated with Alzheimer's disease phenotypes in human iPSC-derived brain cell types. *Neuron* **98**, 1294 (2018).
44. Cakir, B. et al. Engineering of human brain organoids with a functional vascular-like system. *Nat. Methods* **16**, 1169–1175 (2019).
45. Balik, A., Penn, A. C., Nemoda, Z. & Greger, I. H. Activity-regulated RNA editing in select neuronal subfields in hippocampus. *Nucleic Acids Res.* **41**, 1124–1134 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Culture of hiPSC and differentiation into hCS. All hiPSC lines used in this study were validated using previously described standardized methods^{5,22,46}. Cultures were maintained mycoplasma free and were periodically tested for mycoplasma contamination. A total of six hiPSC lines were collected from five healthy subjects (four male and one female). The hiPSC H20961 line was derived by the Gilad Laboratory (University of Chicago). Approval for this study was obtained from the Stanford Institutional Review Board (IRB) panel, and informed consent was obtained from all subjects. hiPSC were cultured on inactivated mouse embryonic fibroblast feeders (EmbryoMax PMEF; Millipore) in DMEM/F12 (1:1, Life Technologies, 11330) containing 20% knockout serum (Life Technologies, 10828), 1 mM non-essential amino acids (Life Technologies, 11140), 1:200 GlutaMax (Life Technologies, 35050), 0.1 mM β -mercaptoethanol (Sigma-Aldrich, M3148) and 10 ng ml⁻¹ FGF2 (R&D Systems, 233-FB) diluted at 0.1% BSA in DPBS (Life Technologies, 14190).

hCS were generated as previously described⁵. Intact hiPSC colonies were lifted using 0.7 mg ml⁻¹ dispase and transferred to ultra-low-attachment plastic dishes (Corning) in the same hiPSC medium without FGF2 but supplemented with 5 μ M dorsomorphin (Sigma-Aldrich) and 10 μ M SB-431542 (Tocris), both of which are SMAD inhibitors, and 10 μ M Y-27632 (EMD Chemicals), which is a ROCK inhibitor. From day 2 (48 h of differentiation), the medium supplemented with dorsomorphin and SB-431542 was changed daily. From day six until day 24, neural spheroids were grown in neurobasal-A (Life Technologies, 10888) neural medium supplemented with B-27 supplement without vitamin A (Life Technologies, 12587), 1:100 GlutaMax (Life Technologies), 1:100 penicillin and streptomycin (Life Technologies, 15070) and with 20 ng ml⁻¹ EGF (R&D Systems, 236-EG) and 20 ng ml⁻¹ FGF2 (R&D Systems, 233-FB). From day 25 to 42, the neural medium was supplemented with 20 ng ml⁻¹ BDNF (Peprotech, 450-02) and 20 ng ml⁻¹ NT3 (Peprotech, 450-03) and medium was changed every other day. From day 43 onwards, hCS were maintained in un-supplemented neural medium with medium changes every 4 days. hCS of similar diameter were randomly selected for experiments.

RNA sequencing. RNA sequencing was performed as previously described¹¹. Briefly, libraries were prepared using Truseq stranded RNA RiboZero Gold (Illumina) and were sequenced using 100-bp paired end reads on an Illumina HiSeq 4000. Reads were then mapped to hg38 with Gencode v.25 annotations using STAR (v.2.5.2b)⁴⁷. Gene expression levels were quantified using RSEM (v.1.3.0)⁴⁸. Genes with low levels of expression (less than ten reads in more than 20% of the samples) were removed from the analysis. Outliers were then removed using standardized sample network connectivity (Z scores smaller than -3)⁴⁹. This method identified two samples as outliers, both of which were >600 days. To quantify the technical variation in the RNA sequencing, we calculated the first five PCs of the Picard sequencing metrics (<http://broadinstitute.github.io/picard/>; v.2.5.0). These PCs, referred to as seqPC1–seqPC5, were then included in the linear model.

To help control for variability between the individuals racial background, we used the GATK (v.3.3) haplotype caller to call single nucleotide polymorphism (SNPs) from the aligned reads⁵⁰. We filtered for sites with missing genotypes (>5%), rare minor allele frequency (<0.05) and out of Hardy-Weinberg equilibrium (<1×10⁻⁶)⁵¹. Genetic ancestry was inferred by running multidimensional scaling (MDS) on these high-quality SNPs together with HapMap3.3 (hg38). The first two MDS values, referred to as ancestryPC1/2, were then included in our linear model. For principal component analysis (PCA), as well as to visualize single gene trajectories, gene expression was normalized using CQN (without quantile normalization, sqn = FALSE) (v.1.28.0) and ancestryPC1-2 and SeqPC1-5 were regressed out before batch correction using ComBat⁵² from the sva package (v.3.30.0) in R. Single gene trajectories trends lines were fitted using the loess method⁵³ from the ggplot2 package⁵⁴ in R. PCA was calculated using the prcomp function in R on scaled normalized and batch corrected counts.

BrainSpan RNA sequencing data analysis. The BrainSpan RNA sequencing data¹⁶ was used as an *in vivo* reference for the analysis. To quantify gene expression at each developmental stage, the cortical samples were aligned to hg38 using Gencode v.25 annotations via STAR⁴⁷. Gene expression was then quantified using the union exon model in featureCounts⁵⁵. We removed low quality samples in which the RNA integrity number (RIN) was lower than 8, there were less than 25% coding bases or ribosomal bases made up more than 25% of total bases (as called by Picard tools). Genes with low levels of expression (less than ten mapped reads in more than 80% of the samples) in a given developmental stage were removed. We retained 196 samples from 24 individuals (9 female and 15 male).

Transition mapping. To compare *in vivo* and *in vitro* changes in gene expression during maturation we used transition mapping¹⁵, which utilizes a rank-rank hypergeometric test⁵⁶. To this end, both *in vivo* and *in vitro* gene expression levels were normalized using the trimmed mean of M-values (TMM) method from the edgeR package⁵⁷ (v.3.24.0). *In vitro* samples were grouped to the closest 25th day in the first 100 days, closest 50th day in between days 100 and 400 and closest 100 day until day 600. Timepoints above day 600 were included in the 600 day group

resulting in the following groups 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500 and 600. Fold change was calculated for each differentiation day (*in vitro*) or developmental stage (*in vivo*) by comparing it to the baseline values of the earliest day *in vitro* (day 25) or earliest stage *in vivo* (stage 2; 8–10 post conception weeks, PCW¹⁶) using the limma-voom method⁵⁸ from the limma package (v.3.38.2) in R. To account for multiple samples coming from the same individual, we used brainID (for the BrainSpan data) or IndividualID (for the hCS data) as blocking factors in the model. The linear model used was ~0 + Differentiation day + batch + PC1 + racePC2 + SeqPC1 + SeqPC2 + SeqPC3 + SeqPC4 + SeqPC5 for the hCS data and ~0 + Period + Ethnicity + PMI + SeqPC1 + SeqPC2 for the BrainSpan data. Genes were then ranked by logFC and the rank-rank hypergeometric test⁵⁶ was used to calculate the significance of the overlap of the gene list using a step size of 200 genes¹⁵.

DNA methylation age. To calculate the DNA methylation (DNAm) age of the samples¹², DNA was extracted using the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen, 80224). Methylation levels were measured using the Infinium MethylationEPIC BeadChip Kit (Illumina), normalized using the Noob method⁵⁹, and were then used to calculate DNAmage¹². DNAmage was averaged over technical replications. We evaluated only clocks that were designed for non-blood tissues, that is, the pan tissue clock¹² and the *in vitro* clock¹⁴. We were unable to predict the culture age using this methylation age, as these epigenetic clocks are not calibrated for the array type¹³ and those using these arrays do not perform well in brain tissue¹⁴. Results were similar between the two methods, and we present the results of the pan tissue clock.

Gene set enrichment analysis. Gene set enrichment analysis (GSEA) was performed using the fgsea package (v.1.8.0)⁶⁰ on all genes ranked by log fold change (using limma-voom as described above) at different timepoints. GO gene sets (v.7.0) were downloaded from <http://software.broadinstitute.org/gsea/msigdb/>. Sets with less than 30 or more than 500 genes were omitted. P values were calculated using 1,000,000 permutations and were corrected using the Benjamini-Hochberg (BH) method. Gene sets with $FDR < 0.05$ were considered to be significant and the top three up- and downregulated sets were plotted.

Weighted gene network analysis. To compare trajectories of genes networks between *in vivo* an *in vitro* datasets, previously described *in vivo* network modules were used¹⁵. To assure that the same networks were present *in vivo*, weighted gene network analysis (WGCNA) was performed on the *in vitro* data using a soft power of 12, minimal module size = 100, deep split = 2, cut height for creation of modules = 0.9999 and cut height for merging modules of 0.1. The modules were then tested for overlap with the *in vivo* modules using Fisher's exact test⁶¹. To visualize the trajectories of the different modules that overlapped with the *in vitro* module, the normalized average expression was calculated using the module Eigengenes function from the WGCNA package⁶² (v.1.68) in R. The trend line was fitted using the loess method⁶³ from the ggplot2 package⁶⁴ in R.

Human tissue. Human brain tissue was obtained under a protocol approved by the Research Compliance Office at Stanford University. PCW21 forebrain tissue was fixed immediately upon arrival.

Immunohistochemistry. Immunohistochemistry was performed as described^{22,63}. Briefly, hCS were fixed in 4% paraformaldehyde (PFA) for 2 h at 4°C. Samples were then washed with phosphate-buffered saline (PBS) three times, transferred to a 30% sucrose solution and, 48–72 h later, embedded and snap frozen in a 30% sucrose and OCT solution (1:1 ratio; Tissue-Tek OCT Compound, 4583, Sakura Finetek). Cryosections (16 μ m) were obtained using a cryostat (Leica). Human cortical tissue was fixed overnight in 4% paraformaldehyde and 30 μ m cryosections were made. All sections were incubated for 1 h at room temperature with blocking solution (10% normal donkey serum and 0.3% Triton-X in PBS), and then overnight with primary antibodies. The following primary antibodies were used: anti-BRN2 (Mouse, 1:500, Millipore, MABD51), anti-CTIP2 (Rat, 1:300, Abcam, ab18465), anti-FXR1 (Mouse, 1:50; Santa Cruz, sc-374148), anti-GFAP (Rabbit, 1:1,000, Dako, Z0334), anti-GFAP (Rat, 1:1,000, ThermoFisher Scientific, 13-0300), anti-HDAC2 (Mouse, 1:50, Santa Cruz, sc-9959), anti-MAP2 (1:5,000, Synaptic Systems, 188004) and anti-SOX9 (Goat, 1:500, R&D Systems, AF3075). After three PBS washes, sections were incubated with Alexa Fluor secondary antibodies (1:1,000, Life Technologies) for 1 h at room temperature. Nuclei were visualized with Hoechst 33258 (ThermoFisher Scientific, H3569). Glass coverslips were mounted on microscopy slides using Aquamount (Thermo Scientific). Images were taken using a SP8 confocal microscope and processed using ImageJ (Fiji).

Western blotting. hCS protein lysates were prepared using a radioimmunoprecipitation assay (RIPA) buffer system (Santa Cruz, sc-24948) and protein concentrations were quantified using the bicinchoninic Acid (BCA) assay (Pierce, ThermoFisher 23225). Proteins (8 μ g per sample per lane) were loaded and run on a 4–12% Bis-Tris PAGE gel (NuPAGE 4–12% Bis-Tris Protein Gel, Invitrogen) and transferred onto a polyvinylidene fluoride (PVDF) membrane (Immobilin-FL, EMD Millipore). Membranes were blocked with 5% milk in PBST

for 1 h at room temperature (RT) and incubated with primary antibodies against β -actin (mouse, 1:50,000, Sigma, A5316), synapsin-1 (rabbit, 1:1000, Cell Signaling, 5297S) overnight at 4 °C and antibodies against the NMDA receptor 2A (GRIN2A, rabbit, 1:1000, Cell Signaling, 4205S), NMDA receptor 2B (GRIN2B, rabbit, 1:1000, Cell Signaling, 4207S) for 72 h at 4 °C. Membranes were washed three times with PBST and then incubated with near-infrared fluorophore-conjugated species-specific secondary antibodies: Goat Anti-Mouse IgG Polyclonal Antibody (IRDye 680RD, 1:10,000, LI-COR Biosciences, 926-68070) or Goat Anti-Rabbit IgG Polyclonal Antibody (IRDye 800CW, 1:10,000, LI-COR Biosciences, 926-32211) for 1 h at RT. Following secondary antibody application, membranes were washed three times with PBST, once with PBS, and then imaged using a LI-COR Odyssey CLx imaging system (LI-COR). Protein band intensities were quantified using Image Studio Lite (LI-COR) with built-in background correction and normalization to β -actin controls. One sample showed unexpectedly low levels of synapsin-1 and was not included in the analyses (Fig. 4, source data).

Measurements of NMDA currents. hCS were prepared for whole cell recordings as previously described⁵. Briefly, hCS were infected with AAV-hSyn1::GFP 2 weeks before recording. Slices (200 μ m thick) were prepared using a Leica VT1200 microtome (Leica) and allowed to recover for 1 h before recordings at 32 °C in bicarbonate-buffered artificial cerebrospinal fluid (aCSF) containing 126 mM NaCl (Sigma, 5922C), 2.5 mM KCl (Sigma, P3911), 1.25 mM NaH₂PO₄ (Fisher Scientific/Acros Organics, AC424390025), 2 mM MgCl₂ (Sigma, M8266), 2 mM CaCl₂ (Sigma, C4901), 26 mM NaHCO₃ (Sigma, S5761) and 10 mM glucose (Fisher Scientific/Acros Organics, AC410950010). Whole cell patch clamp recordings were performed at room temperature (~22–25 °C). Slices were superfused with aCSF at a rate of 3 mL min⁻¹. Whole cell patch clamp recordings were collected using a MultiClamp 700 A amplifier (Molecular Devices), Axon Digidata 1550B digitizer (Molecular Devices) and Clampex 11.0 software (Molecular Devices). Borosilicate glass pipettes (3–5 M Ω) were used to obtain intracellular recordings. Fluorescently labeled neurons were randomly selected for patching. Pipettes were filled with internal solution containing 120 mM potassium gluconate, 11 mM KCl, 1 mM MgCl₂, 1 mM CaCl₂, 10 mM HEPES (Sigma, H4034) and 1 mM EGTA (Sigma, E4378), and the pH was adjusted to 7.4. NMDA (Sigma, M3262) was applied to the cell using a 2–3 M Ω borosilicate glass pipette positioned 25 μ m away from the cell body. Pulses (50 ms) of NMDA were produced using a Picopipette II (General Valve Corporation). Cell recording quality was monitored by measuring access resistance and cells that deviated by more than 15% during the course of the recording were discarded. GRIN2B-containing NMDA receptors were blocked by adding ifenprodil tartrate salt (IFN; 10 μ M; Sigma, I2892) to aCSF. hCS derived from two cell lines (8858-3 and 1205-4) were used for these recordings. One cell was patched per hCS, for a total of 25 neurons from 25 hCS (Supplementary Table 1). Association between the proportion of change in amplitude after adding IFN was tested using beta regression⁶⁴ with a logit link function using the betareg package (v.3.1-3) in R.

RNA editing identification. RNA-seq reads were mapped to GRCh37 genome and transcriptome using HISAT2 (ref.⁶⁵; v.2.1.0) with parameters accounting for the respective strand specificities of the BrainSpan and cortical spheroid datasets. Uniquely mapped reads were retained for further analysis. We then used previously developed procedures to identify RNA editing sites^{26,66,67}. In brief, first we used unmapped reads to find editing sites in hyperedited regions⁶⁸. Adenosines in unmapped reads were converted into guanosines and aligned with HISAT2 to a modified hg19 genome where adenosines were also substituted with guanosines²⁶. These hyperedited reads were then combined with the original uniquely mapped reads. Next, candidate editing sites were identified as mismatches between reads and the reference genome. A log-likelihood test and posterior filters were then applied to eliminate editing sites likely caused by sequencing errors and other technical artifacts⁶⁷. To eliminate rare genomic variants from identified RNA editing sites, we filtered for sites found across multiple individuals. Specifically, editing sites were required to be found with at least five total reads and two reads edited across five unique individuals. In our cortical spheroid dataset, we reduced the requirement to four unique individuals, due to sample size. A total of 109,487 and 19,046 editing sites were identified in the BrainSpan and cortical spheroid datasets, respectively.

Weighted co-editing network analysis for RNA editing sites. Modules of RNA editing sites were found using the WGCNA package⁶². To obtain accurate representation of topological overlap, we first filtered for editing sites with at least five total reads in at least 80% of samples, zero variance or too much missing data using the goodSamplesGenes function in the WGCNA package. To facilitate finding modules corresponding to developmental time in the BrainSpan dataset, we required nonzero editing in at least 51% of samples from at least one period. Given the limited sample size of our cortical spheroid dataset, this filter was applied only to the BrainSpan samples.

For the cortical spheroids we adjusted RNA editing levels to avoid individual modules driven by single individuals. For each editing site, a linear model was constructed between editing level against differential day, individual and batch. Editing levels were adjusted by subtracting out the maximum likelihood beta estimates for individuals and batch. A soft threshold power of ten was used to fit scale-free topology. To preclude modules driven by outlier samples, we followed

our previous bootstrapping strategy^{26,69,70}, where modules were obtained using consensus topological overlap from 100 bootstraps. In brief, for each bootstrap, samples up to the original sample size were randomly resampled with replacement. Signed topological overlap matrices were obtained from the corresponding matrix of editing sites using TOMsimilarity (adjacency(corrFn = "cor", type = "signed", power = 10, corOptions = list(method = "spearman", use = "pairwise.complete.obs"), TOMType = "signed"). The bootstrapped matrices were then recalibrated from 20,000 random matrix entries. The consensus topological overlap matrix was taken as the median across all recalibrated bootstrapped matrices.

To find associations of modules with various biological processes, we defined the eigenvalue as the first PC of each module. The timecourse trajectory of each module was determined by plotting each eigenvalue against differentiation day (for cortical spheroids) and period (for BrainSpan). Associations between the modules and RNA editing enzymes were evaluated by correlating the eigenvalue against adjusted gene expression values for established RNA editing enzymes encoded by *ADAR1*, *ADAR2* and *ADAR3* (ref.⁷¹), and against *FMR1* and *FXR1*, that encode proteins that were recently shown to also exhibit RNA editing regulation²⁶. Modules found in BrainSpan samples and cortical spheroids were tested for significant overlap of member editing sites using Fisher's exact test. Only editing sites found in both BrainSpan and cortical spheroids were considered for this test.

eCLIP analysis of RNA editing sites. To investigate whether RNA editing in modules are targetable by FMRP and FXR1P proteins, we obtained eCLIP datasets of FMRP and FXR1P binding sites in postmortem human frontal cortex²⁶. Comparison of distances between eCLIP peaks and module editing sites was also performed using published methods²⁶. Briefly, for member editing sites within each BrainSpan or hCS module, the closest distances from eCLIP peaks compared to the null background consisting of distances between peaks and gene-matched random adenosines were compared over 10,000 sets of controls. *P* values were obtained by calculating the area under the curve (AUC) of the cumulative distribution of distances of editing sites to CLIP peaks in the interval 0–100 kb. AUC values of the 10,000 sets of controls were modeled by a Gaussian distribution, which was then used to calculate a one-sided *P* value for the AUC of the module editing sites. To test overlap of FMRP- and FXR1-targeted editing sites between hCS and BS modules, an editing site was labeled "targetable" if within 1,000 bp of the nearest FMRP or FXR1P CLIP peak. Editing sites residing in genes expressed with RPKM < 5 in adult frontal cortex were not coverable by CLIP analysis and excluded from analysis. The significance of overlap of targetable editing sites between hCS and BS modules was evaluated using Fisher's exact test. Only editing sites found in both BrainSpan and hCS were considered for this test.

Mapping of disease genes. Genes associated with ASD (<https://gene.sfari.org/database/gene-scoring/>), ID⁷², epilepsy²¹, SCZ⁷³, AD^{4,75}, PD^{76,77}, PSP⁷⁸ and FTD⁷⁹ were analyzed. For the ASD genes, only high confidence genes (gene score < 2 or syndromic genes) were analyzed. For AD and PD, we combined common variants from the genome-wide association studies (GWAS) with genes associated with rare, familial forms of these diseases. For AD, these genes are *APOE*, *APP*, *PSEN1* and *PSEN2* (ref.⁷⁵). For PD these genes are *PINK1*, *SNCA*, *LRRK2*, *PRKN*, *UCHL1* and *PARK7* (ref.⁷⁷). As there are only a small number of genes associated with the neurodegenerative diseases PSP (9 genes) and FTD (13 genes), and they are considered part of a frontal lobar degeneration spectrum, these were combined. Genes were clustered by their expression in the hCS using hierarchical clustering on the Euclidean distance between the genes. Cluster eigengenes were calculated using the module Eigengenes function from the WGCNA package. The gene in each cluster were correlated to the cluster module eigengene and the top five genes were annotated on the heatmap. GO terms enrichment was performed using the enrichGO function from the clusterProfiler package⁸⁰ (v.3.12.0). Enrichment was performed on biological process and molecular function GO terms. All genes expressed in the hCS were used as background. Cell type enrichment was performed using the bootstrap.enrichment.test from the EWCE package⁸¹ (v.0.99.2) on hCS single-cell data¹¹ with 100,000 permutations. All genes expressed in both the current dataset and the single-cell dataset were used as background. This method tests whether a list of genes has a higher level of expression in a specific cell type that would be expected by chance.

Statistics. Statistical analyses, including Fisher's exact test, beta regression and Spearman correlation rank–rank hypergeometric tests, were performed as detailed in legends and Methods. As these tests do not depend on the distribution being normal, no test for normality was performed. No statistical methods were used to predetermine sample sizes, but our sample sizes per timepoint are similar to those reported in previous publications^{26,10}. Due to the nature of these long-term cultures, collection of samples was not performed blind to the differentiation stage.

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

Data availability

Gene expression data and methylation data are available in the Gene Expression Omnibus (GEO) under accession numbers GSE150122 and GSE150123. The

accompanying GECO webtool can be accessed at <https://labs.dgsom.ucla.edu/geschwind/files/view/html/GECO.html>. The BrainSpan data are available in the database of Genotypes and Phenotypes (dbGaP) under Study Accession phs000755.v2.p1. Single-cell data from human fetal cerebral cortex can be found at <http://geschwindlab.dgsom.ucla.edu/pages/codexviewer> and at dbGaP under Study accession phs001836. eCLIP data for FXR1 and FMR1 are available in GEO with accession number [GSE107895](#). Human cortical organoid single-cell sequencing data are available in GEO with accession number [GSE10771](#). Source data are provided with this paper.

Code availability

The code used in this manuscript can be found at https://github.com/dhglab/human_cortical_organoid_maturity.

References

46. Pașca, S. P. et al. Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome. *Nat. Med.* **17**, 1657–1662 (2011).
47. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
48. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 323 (2011).
49. Oldham, M. C., Langfelder, P. & Horvath, S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst. Biol.* **6**, 63 (2012).
50. McKenna, A. et al. The genome analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
51. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
52. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
53. Cleveland, W. S. & Devlin, S. J. Locally weighted regression: an approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**, 596–610 (1988).
54. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, 2016).
55. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
56. Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res.* **38**, e169 (2010).
57. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
58. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
59. Triche, T. J. Jr., Weisenberger, D. J., Van Den Berg, D., Laird, P. W. & Siegmund, K. D. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* **41**, e90 (2013).
60. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. Preprint at *bioRxiv* <https://doi.org/10.1101/060012> (2019).
61. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is my network module preserved and reproducible? *PLoS Comput. Biol.* **7**, e1001057 (2011).
62. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
63. Sloan, S. A., Andersen, J., Pasca, A. M., Birey, F. & Pasca, S. P. Generation and assembly of human brain region-specific three-dimensional cultures. *Nat. Protoc.* **13**, 2062–2085 (2018).
64. Cribari-Neto, F. & Zeileis, A. Beta regression in R. *J. Stat. Softw.* **34**, 1–24 (2010).
65. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
66. Bahn, J. H. et al. Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. *Genome Res.* **22**, 142–150 (2012).
67. Lee, J. H., Ang, J. K. & Xiao, X. Analysis and design of RNA sequencing experiments for identifying RNA editing and other single-nucleotide variants. *RNA* **19**, 725–732 (2013).
68. Porath, H. T., Carmi, S. & Levanon, E. Y. A genome-wide map of hyper-edited RNA reveals numerous new sites. *Nat. Commun.* **5**, 4726 (2014).
69. Parikshak, N. N., et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).
70. Wu, Y. E., Parikshak, N. N., Belgard, T. G. & Geschwind, D. H. Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nat. Neurosci.* **19**, 1463–1476 (2016).
71. Nishikura, K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat. Rev. Mol. Cell Biol.* **17**, 83–96 (2016).
72. Parikshak, N. N. et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
73. Wang, D., et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
74. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat. Genet.* **51**, 404–413 (2019).
75. Cacace, R., Sleegers, K. & Van Broeckhoven, C. Molecular genetics of early-onset Alzheimer's disease revisited. *Alzheimers Dement.* **12**, 733–748 (2016).
76. Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49**, 1511–1516 (2017).
77. Farrer, M. J. Genetics of Parkinson disease: paradigm shifts and future prospects. *Nat. Rev. Genet.* **7**, 306–318 (2006).
78. Chen, J. A. et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* **13**, 41 (2018).
79. Greaves, C. V. & Rohrer, J. D. An update on genetic frontotemporal dementia. *J. Neurol.* **266**, 2075–2086 (2019).
80. 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).
81. Skene, N. G. & Grant, S. G. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Front. Neurosci.* **10**, 16 (2016).

Acknowledgements

We acknowledge experimental support from members of the Pasca Lab at Stanford University: A. M. Pasca, N. Huber, T. Khan, F. Birey, A. Puno, L. Li and T. Li, and members of the Pasca lab and Geschwind lab for helpful discussions and support, including A. Elkins for setting up the GECO web browser. This work was supported by a Distinguished Investigator Award from the Paul G. Allen Frontiers Group (to D.H.G.); by grants from California Institute of Regenerative Medicine (CIRM) and the National Institute of Mental Health Convergent Neuroscience Consortium (U01 MH115745) (to D.H.G. and S.P.P.), the Stanford Human Brain Organogenesis Program in the Wu Tsai Neuroscience Institute (to S.P.P.), Stanford Bio-X (to S.P.P.), the Stanford Wu Tsai Neuroscience Institute Big Idea Grant (to S.P.P.), the Kwan Funds (to S.P.P.), the Senkut Research Fund (to S.P.P.), the Autism Science Foundation (ASF) and the Brain and Behavior Research Foundation Young Investigator award (Brain & Behavior Research Foundation) (to A.G.). S.P.P. is a New York Stem Cell Foundation (NYSCF) Robertson Stem Cell Investigator and a Chan Zuckerberg Initiative (CZI) Ben Barres Investigator.

Author contributions

A.G., S.P.P. and D.H.G. planned and directed experiments, guided analyses, and wrote the manuscript with assistance from all authors. A.G. performed RNA-seq analysis and methylation analysis. S.-J.Y. performed cell culture, DNA and RNA extraction. S.S.T. performed RNA editing analysis. C.D.M. performed electrophysiology recordings. J.A. performed immunohistochemistry. J.Y.P. and A.M.V. performed western blots. S.H. analyzed the methylation data and interpreted the findings. X.X. supervised RNA editing analysis and interpretation. J.R.H. supervised electrophysiology experiments and interpretation.

Competing interests

S.P.P. is listed on a patent held by Stanford University that covers the generation of region-specific brain organoids (US patent 62/477,858). All other authors declare no competing interests.

Additional information

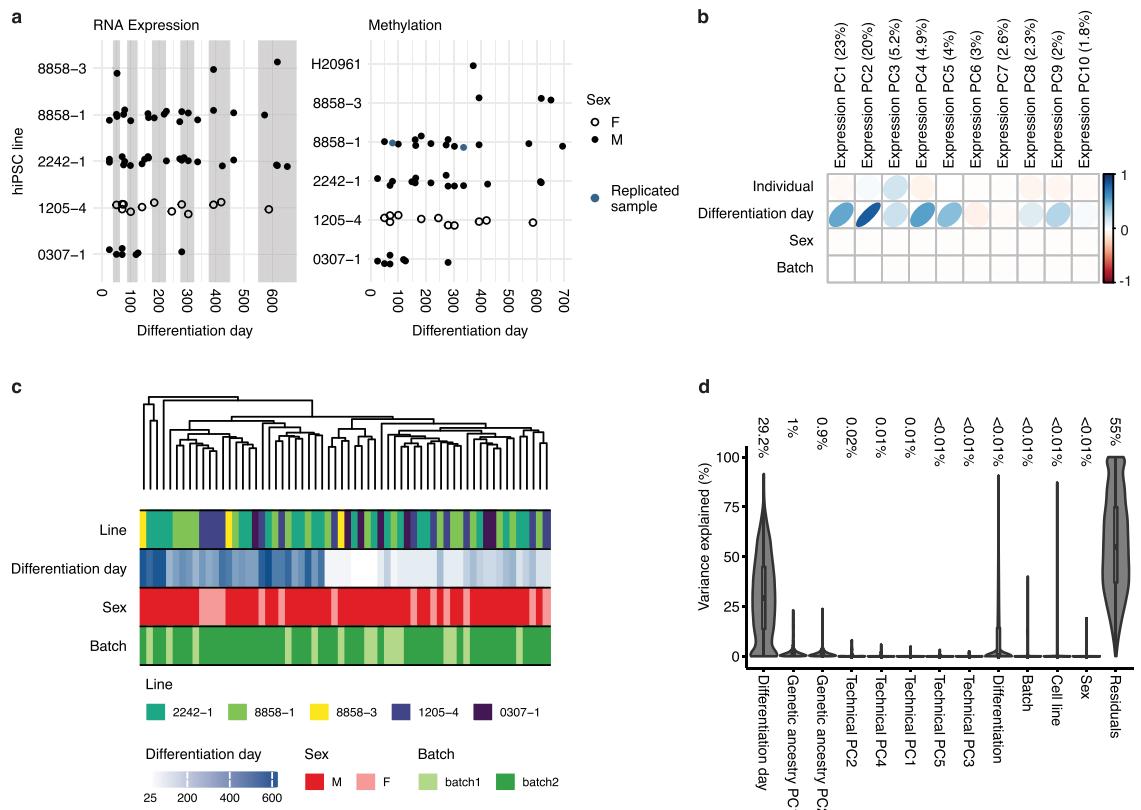
Extended data is available for this paper at <https://doi.org/10.1038/s41593-021-00802-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41593-021-00802-y>.

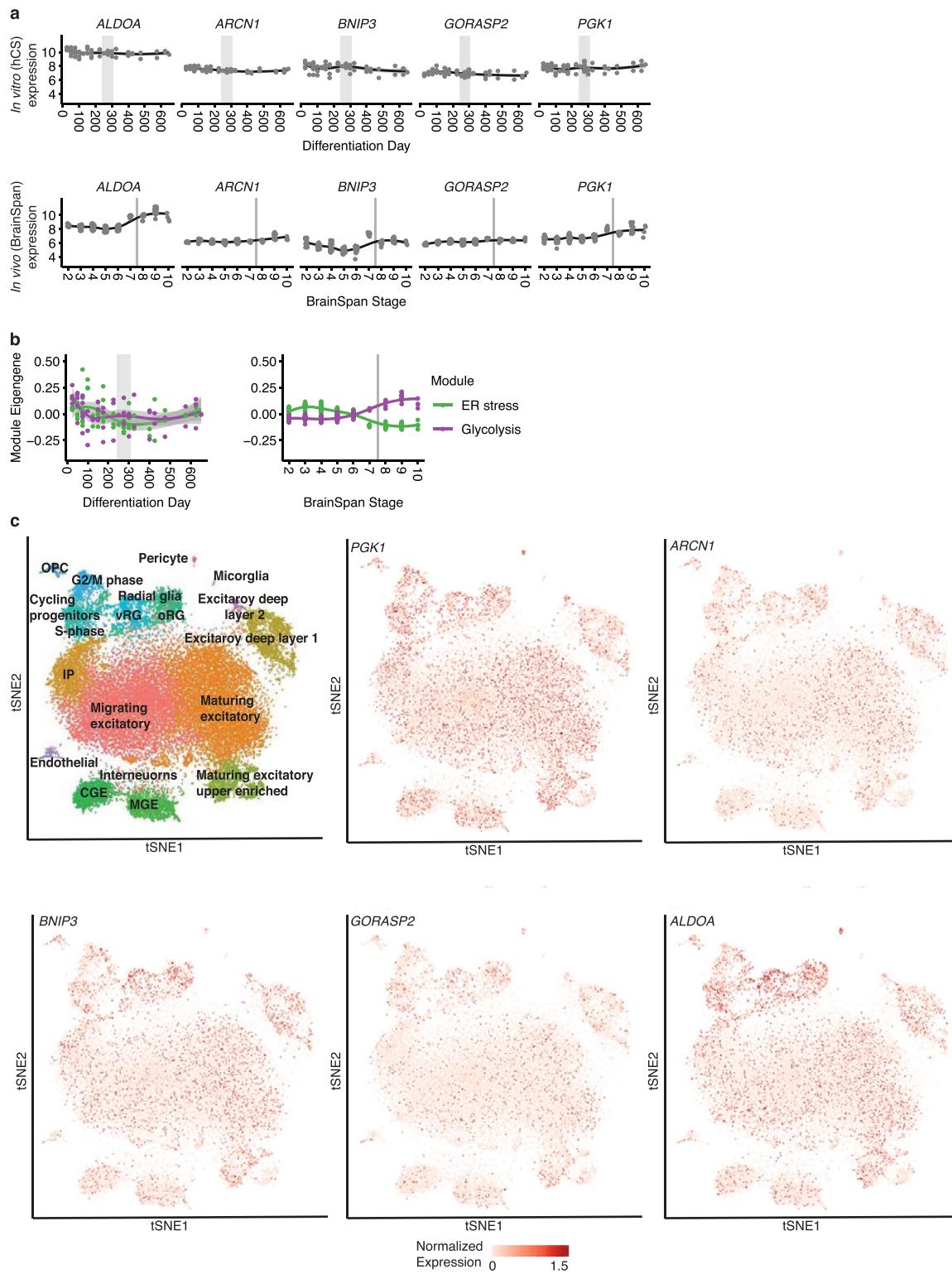
Correspondence and requests for materials should be addressed to S.P.P. or D.H.G.

Peer review information *Nature Neuroscience* thanks Kristen Brennand and the other anonymous reviewer(s) for their contribution to the peer review of this work.

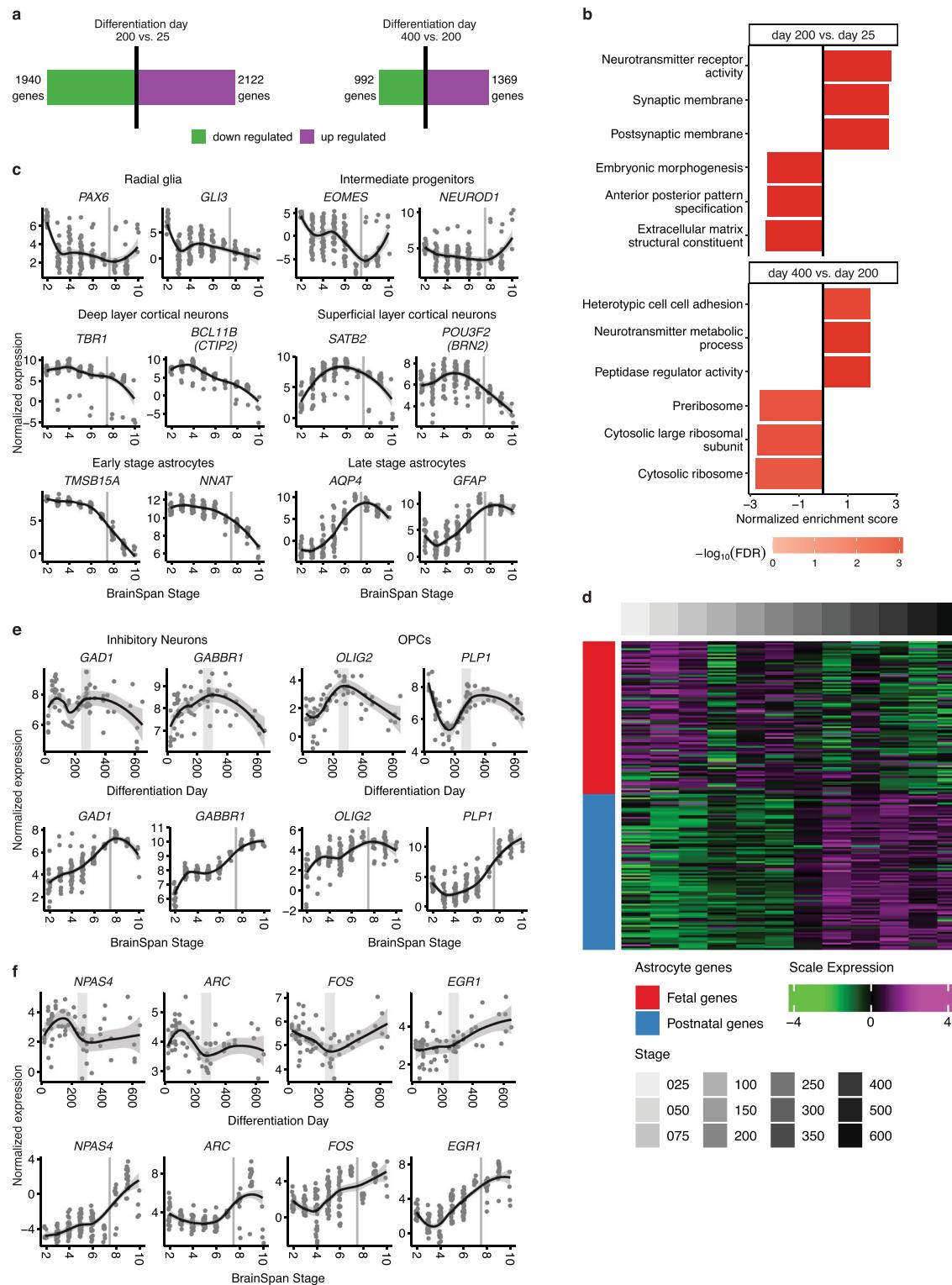
Reprints and permissions information is available at www.nature.com/reprints.



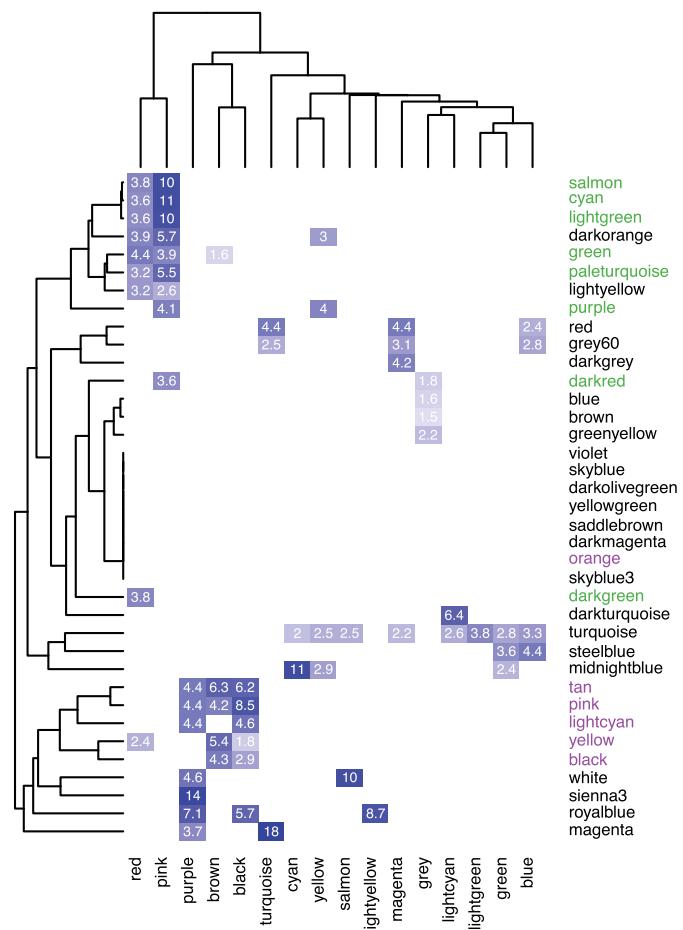
Extended Data Fig. 1 | Data description and quality. **a**, Timepoints and hiPSC line information for the 62 samples used for RNA sequencing (left). Samples were differentiated from 5 cell lines derived from 4 individuals. Timepoints and hiPSC information for the 50 samples used for DNA methylation (right). Samples were differentiated from 6 cell lines derived from 5 individuals (see Supplementary Tables 1 and 2). Two samples (blue) were hybridized in replicate for quality control purposes and their values were averaged. Each point represents one sample from a specific cell line (y-axis) and differentiation day (x-axis). Full circles represent sample coming from males and rings represent samples coming from females. Gray and white background shading show aggregation of differentiation days into stages. **b**, Principal component analysis (PCA) of the expression data. The values represent the adjusted r squared of the PC with the covariates indicated. The numbers in brackets on axis titles are the percent of variance explained by the PC. The first 5 PCs, which explain 57.1% of the total variance, show high association with differentiation day. **c**, Dendrogram of hierarchical clustering of samples demonstrating that differentiation day but no other covariates (individual, Sex, batch) is driving the clustering of samples. **d**, Violin plots of the variance explained by each of the covariates for each gene. Outlines represent the density of the percent of variance explained. The numbers are the median value of percent of explained variance for each variable. Boxplots in **d** show: center – median, lower hinge – 25% quantile, upper hinge – 75% quantile, lower whisker – smallest observation greater than or equal to lower hinge $-1.5 \times$ interquartile range, upper whisker – largest observation less than or equal to upper hinge $+1.5 \times$ interquartile range. n = 62 samples from 5 hiPSC lines derived from 4 individuals.



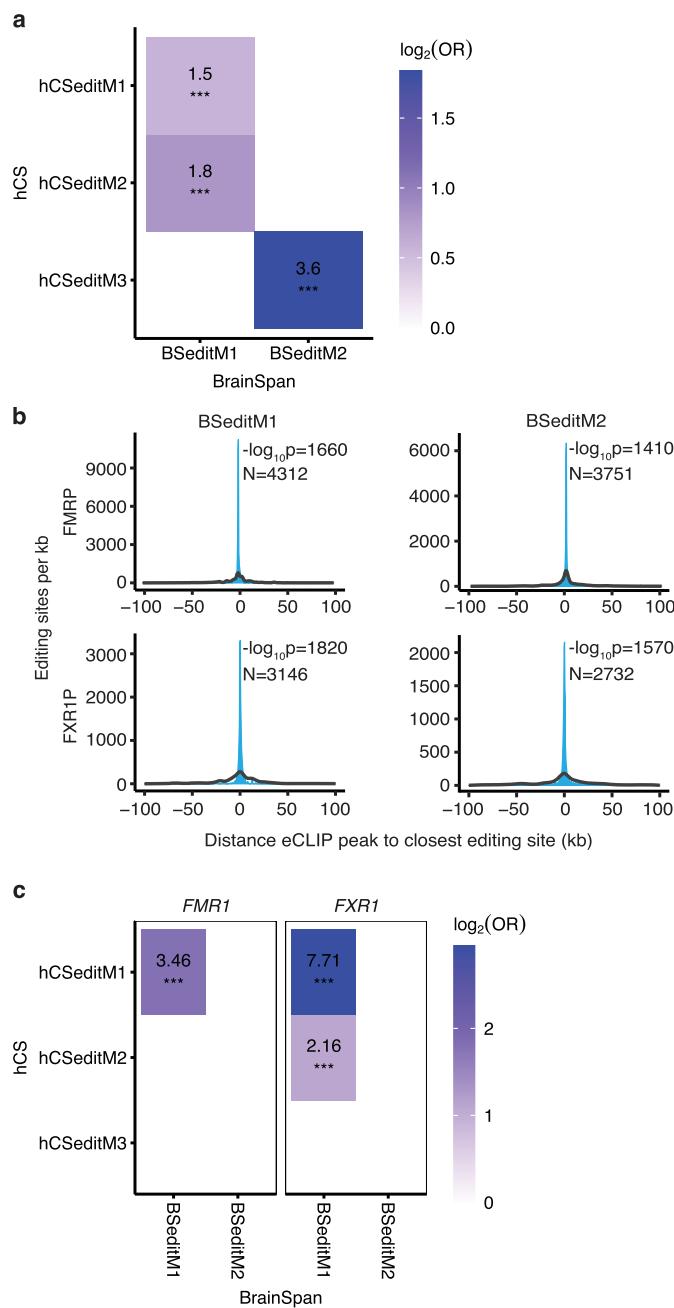
Extended Data Fig. 2 | Cell stress in hCS. **a**, Trajectories of metabolic cell stress genes²⁰ hCS (top) and *in vivo* (bottom). **b**, *In vitro* (left) and *in vivo* (right) module eigen genes of glycolysis (*organoid.Sloan.human.ME.paleturquoise*) and ER stress (*organoid.human.ME.darkred*) previously suggested to be upregulated *in vitro*²⁰. Gray areas denote time of shift from prenatal to postnatal gene expression. In (a) and (b) shaded gray area around the trajectory represents the 95% confidence interval, vertical gray lines represent birth and vertical gray bars denote the shift from prenatal to postnatal gene expression based on matching to *in vivo* patterns. For *in vitro* data n = 62 samples from 5 hiPSC lines derived from 4 individuals and for *in vivo* data n = 196 from 24 individuals. **c**, Scatterplot visualization of cells in developing fetal cortex colored by major cell types²². VRG, ventral radial glia; oRG, outer radial glia; CGE, caudal ganglionic eminence; MGE, medial ganglionic eminence; OPC, oligodendrocyte precursor cell; IP, intermediate progenitors.



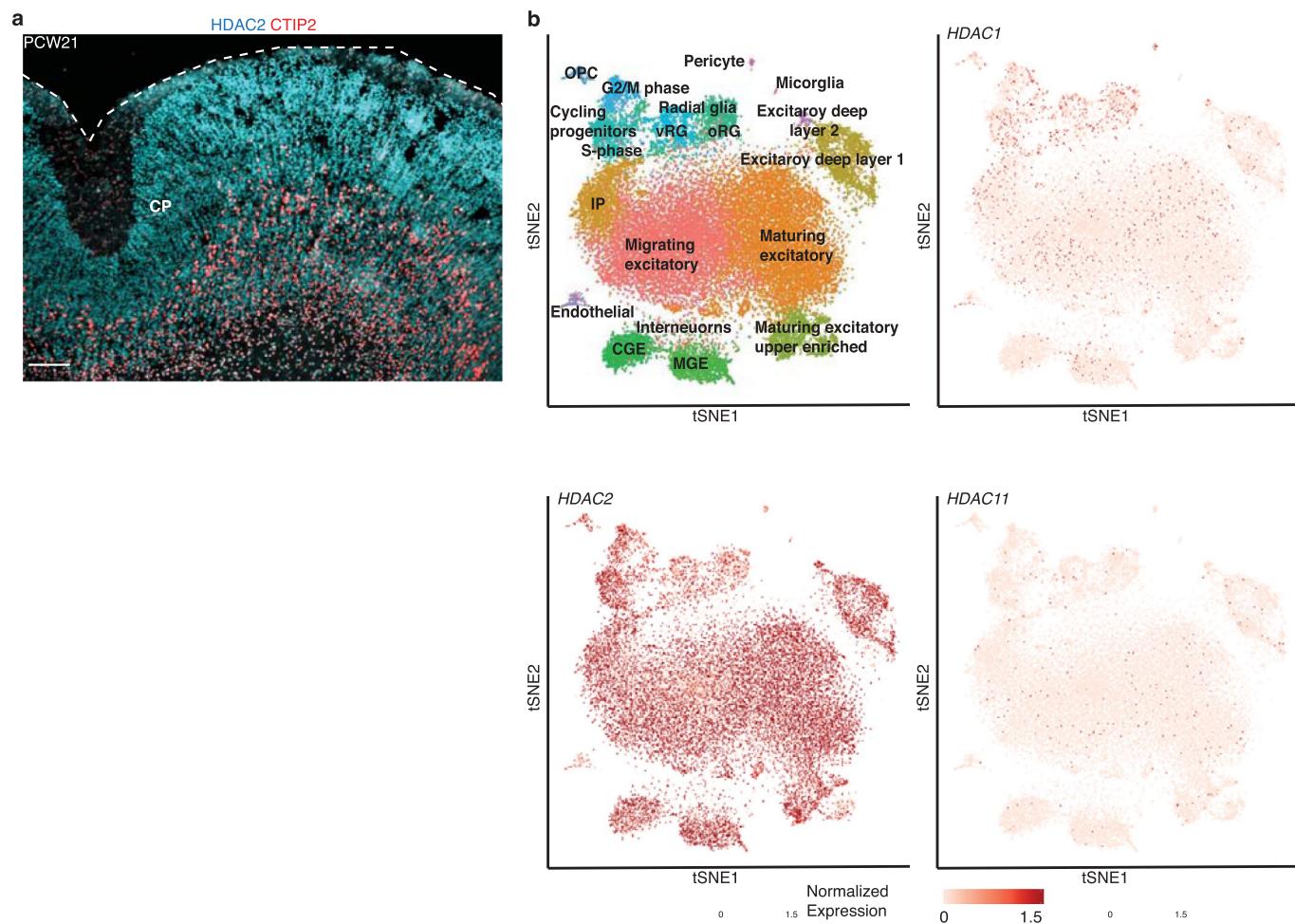
Extended Data Fig. 3 | Changes in biological processes between early and later stages of differentiation. **a**, Number of differentially expressed genes when comparing differentiation day 200 to differentiation day 25 (left) and differentiation day 400 to differentiation day 200 (right). Magenta bar represents upregulated genes and the green bar represents down-regulated genes. **b**, Top 3 up- and downregulated GO terms enriched in genes ranked by logFC using gene set enrichment analysis, (GSEA; FDR < 0.05). **c**, Normalized expression of marker genes *in vivo* for neurons, intermediate progenitors, astrocytes, and radial glia as well as superficial and deep layer cortical neurons. **d**, Scaled expression of fetal and mature astrogliial genes⁷ during differentiation. A shift between fetal and mature gene sets occurs at ~250 days of hCS differentiation. **e**, Normalized expression of marker genes for inhibitory neurons and oligodendrocyte precursor cells (OPCs) that are not preserved in hCS. **f**, Normalized expression of activity-dependent genes that are not preserved in hCS. In (c), (e) and (f) shaded gray area around the trajectory represents the 95% confidence interval, vertical gray lines represent birth and vertical gray bars denote the shift from prenatal to postnatal gene expression based on matching to *in vivo* patterns. For *in vitro* data n = 62 samples from 5 hiPSC lines derived from 4 individuals and for *in vivo* data n = 196 from 24 individuals.



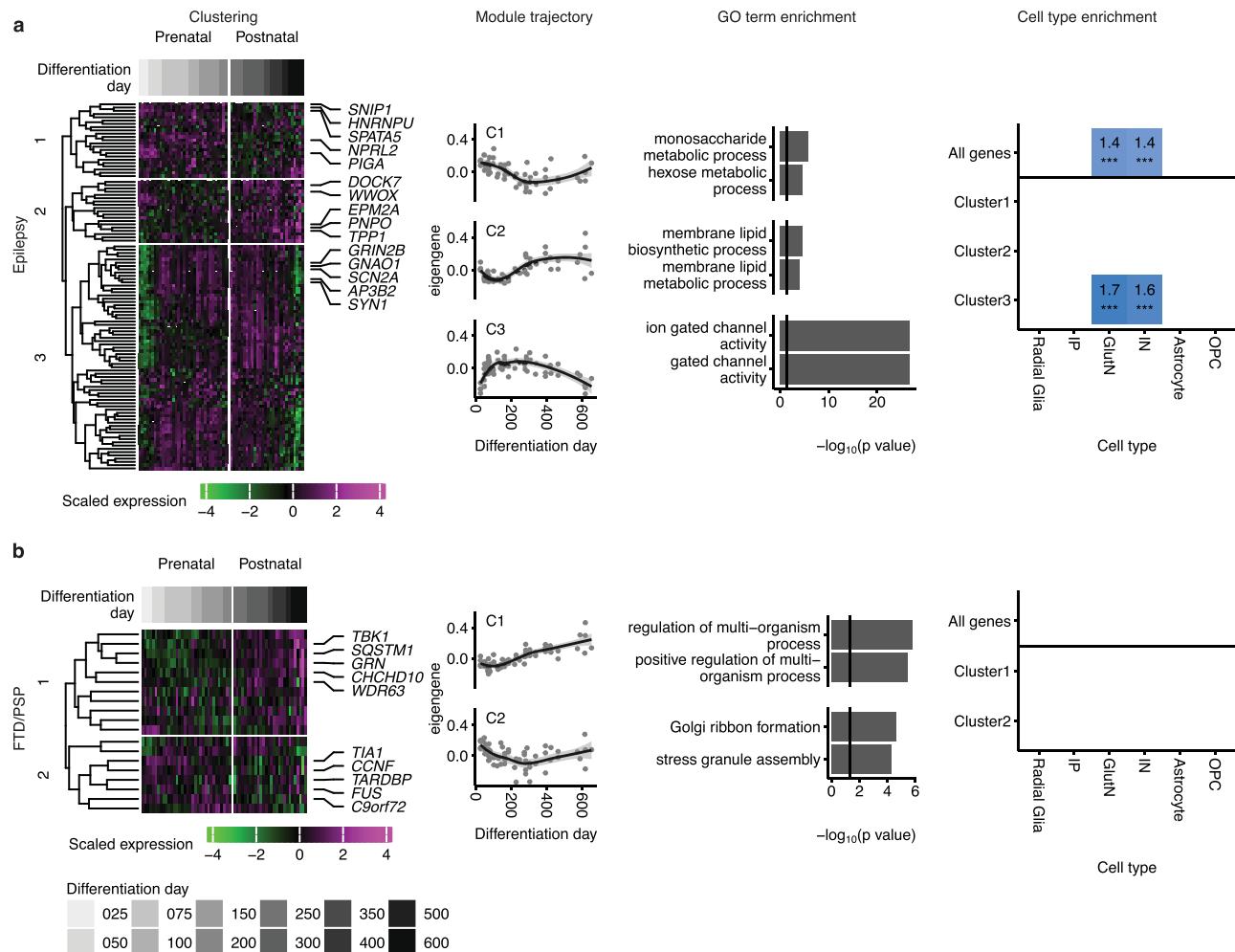
Extended Data Fig. 4 | Overlap between hCS and *in vivo* WGCNA modules. Overlap of genes in hCS and the BrainSpan *in vivo* modules. Significant ORs are presented. Modules were clustered using complete-linkage hierarchical clustering. Color represents the OR of each overlap. *In vivo* neuronal modules (green) and glial modules (purple) are marked.



Extended Data Fig. 5 | Overlap between hCS and in vivo editing modules. **a**, Overlap of editing sites in hCS and BrainSpan *in vivo* modules. Significant ORs are presented. **b**, Distributions showing the closest distances between editing sites from BrainSpan editing modules and FMRP or FXR1P eCLIP peaks (blue). The median of 10,000 sets of control sites (black) is depicted for comparison. See Methods for details of P-value calculation. N, number of editing sites shown. **c**, Overlap of editing sites within 1000bp of a CLIP site in hCS and BrainSpan *in vivo* modules. Significant ORs are presented. *** FDR < 0.005.



Extended Data Fig. 6 | Expression of select genes in the in vivo fetal cerebral cortex. **a**, Immunohistochemistry of HDAC2 and the deep layer marker CTIP2 (BCL11B) at post conception week 21 (PCW21). CP, cortical plate. Scale bars, 100 μ m. The Immunohistochemistry experiment was performed once. **b**, Scatterplot visualization of cells in developing fetal human cerebral cortex colored by major cell types²². vRG, ventral radial glia; oRG, outer radial glia; CGE, caudal ganglionic eminence; MGE, medial ganglionic eminence; OPC, oligodendrocyte precursor cell. IP, Intermediate progenitors.



Extended Data Fig. 7 | Mapping neurodegenerative and epilepsy disorder genes onto hCS differentiation. Mapping of genes associated with progressive supranuclear palsy (PSP) and frontotemporal dementia (FTD) (a), and epilepsy (b) onto hCS differentiation trajectories. The first column shows clustering of scaled normalized expression of genes associated with a disorder. Genes (in rows) are clustered using hierarchical clustering on the Euclidean distance between genes. Samples (columns) are ordered by differentiation day (represented by gray bars) with the earliest days on the left and latest timepoints on the right. The 5 most representative genes (highest correlation with the cluster eigengene) are shown. The second column shows the cluster eigengenes (first principal component) for the identified gene clusters. Shaded gray area around the trajectory line represents the 95% confidence interval. The third column shows the top GO terms enriched in the identified clusters. The fourth column shows cell types over expressed in either all the genes associated with a disorder (above line) or in the genes from the identified clusters. Number and color represent the fold change. Significance was tested using a one-sided permutation test with 100,000 permutations. P values were corrected for multiple testing using the Benjamini-Hochberg method. * FDR < 0.05, ** FDR < 0.01, *** FDR < 0.005. n = 62 samples from 5 hiPSC lines derived from 4 individuals. IP, intermediate progenitors; GlutN, glutamatergic neurons; IN, interneurons; OPC, oligodendrocyte progenitor cells.

Reporting Summary

Nature Research 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 Research 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

Hi-Seq4000
ImageJ
MultiClamp 700A amplifier (Molecular Devices)
Axon Digidata 1550B digitizer (Molecular Devices)
Clampex 11.0 software (Molecular Devices)
SP8 confocal microscope (Leica)
LI-COR Odyssey CLx imaging system (LI-COR)

Data analysis

The code used for this manuscript can be found at: https://github.com/dhglab/human_cortical_organoid_maturation
Bioinformatic tools used picard tools v2.5.0, STAR v2.5.2b, rsem v1.3.0, GATK v3.3, HISAT 2.1.0, R 3.5.1 and 3.6.0
R packages used: cqn 1.28.0, fgsea 1.8.0, sva 3.30.0, edgeR 3.24.0, limma 3.38.2, WGCNA 1.68, EWCE 0.99.2, clusterProfiler 3.12.0 , ggplot2 v3.3.2, betareg 3.1-3
Western blots were quantified using Image Studio Lite (LI-COR).

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 Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Gene expression data has been deposited in the Gene Expression Omnibus (GEO) under accession numbers GSE150122 and GSE150123. The data in this study are available on request from the corresponding author.

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	Six hiPSC lines were used in this paper (Supplementary Table 1 and 2). A total of 62 samples for RNA sequencing (from 4 individuals, 5 hiPSC lines) and a total of 50 samples for DNA methylation (from 5 individuals, 6 hiPSC lines) at 13 time periods were collected. For WB, 3 hiPSC were used, for electrophysiological experiments 2 hiPSC lines were used and for immunocytochemistry 4 hiPSC lines were used. No statistical methods were used to pre-determine sample sizes, but our sample sizes per time point are similar to those reported in previous publications (Pasca et al., Nature Methods, 2015; Sloan et al., Neuron 2017; Trevino et al., Science, 2020)
Data exclusions	The predetermined exclusion criteria were: For RNA-seq, samples were excluded if 3 standard deviations away from the mean standardized sample network connectivity. For electrophysiology experiments, cells which deviated by more than 15% in their access resistance during the course of the recording were not used in the analysis.
Replication	hiPSC lines were differentiated from 6 hiPSC lines derived from 5 individuals to assess reliability of the methods and maintained long term (previously described in Yoon et al., Nature Methods, 2019).
Randomization	hCS of similar diameter were randomly selected for experiments. Neurons for patching were randomly selected using a fluorescent reporter.
Blinding	This study did not include case-control comparison. For western blotting the investigators were not blinded to the differentiation status of the hCS samples. Patch clamping was performed blindly to the stage of differentiation of hCS.

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.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

anti-BRN2 (Mouse, 1:500, Millipore, MABD51, clone 8C4.2, Lot#2375594)
 anti-CTIP2 (Rat, 1:300, Abcam, ab18465, Lot#GR322373-7)
 anti-FXR1 (B-2, Mouse, 1:50; Santa Cruz, sc-374148, Lot#B2018)
 anti-GFAP (Rabbit, 1:1,000, Dako, Z0334, Lot#20073982)
 anti-GFAP (Rat, 1:1,000, Thermo Fisher Scientific, Clone 2.2B10, 13-0300)
 anti-HDAC2 (C-8, Mouse, 1:50, Santa Cruz, sc-9959, Lot#G1320),
 anti-MAP2 (Guinea pig, 1:5,000, Synaptic Systems, 188004, Lot#2-26)
 anti-SOX9 (Goat, 1:500, R&D Systems, AF3075)
 anti-NR2A(GRIN2A) (Rabbit, 1:1000, Cell Signaling #4205)
 anti-NR2B(GRIN2b) (Rabbit, 1:1000, Cell Signaling #4207)
 anti-β-actin (Mouse, 1:50,000, Sigma, A5316)
 anti-Synapsin1 (Rabbit, 1:1000, Cell Signaling, 5297S)
 anti-Mouse IgG Polyclonal Antibody Goat, IRDye 680RD, 926-68070
 anti-Rabbit IgG Polyclonal Antibody (Goat, IRDye 800CW, 926-32211)

Validation

All antibodies were commercially available.

We have previously used and/or validated some of these antibodies (Pasca et al., *Nature Methods* 2015; Sloan et al., *Neuron* 2017; Trevino et al., *Science* 2020).

The anti-BRN2 antibody was validated in human neural cells (Trevino et al., *Science* 2020).

The anti-CTIP2 has been referenced in 401 publications according to the manufacturer's website, and validated in human neural cells (Pasca et al., *Nature Methods* 2015; Trevino et al., *Science* 2020).

The anti-FXR1 antibody has been referenced in 2 publications according to the manufacturer's website.

The anti-GFAP (Rabbit) has been used in 8 studies according to the manufacturer's website, and has been validated in human neural cells (Pasca et al., *Nature Methods* 2015; Sloan et al., *Neuron* 2017; Trevino et al., *Science* 2020).

The anti-GFAP (Rat) has been used in 130 publications according to the manufacturer's website and has been validated in human neural cells (Trevino et al., *Science* 2020).

The anti-HDAC2 has been used in 64 publications according to the manufacturer's website.

The anti-MAP2 has also been used in 71 studies for ICC and 13 studies for IHC according to the manufacturer's website, and has been validated in human neural cells (Pasca et al., *Nature Methods* 2015; Trevino et al., *Science* 2020).

The anti-SOX9 has been used in 36 studies according to the manufacturer's website, and has been validated in human neural cells (Trevino et al., *Science* 2020).

The rabbit anti-NMDAR-2A (Cell Signaling, #4205) was used in 38 publications according manufacturer's website.

The mouse anti-NMDAR-2B (Cell Signaling, #4207) was used in 31 publication according manufacturer's website.

The mouse anti-β-actin (Sigma, A5316) has been used in 1,670 publications according manufacturer's website, and tested for western blot analysis in human cells (Gabriel-Salazar et al., 2018).

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

Five hiPSC lines from were derived at Stanford University with IRB approval and following written consent; one line (H20961) was obtained from the Gilad lab at University of Chicago under an MTA.
 Inactivated mouse fibroblasts EmbryoMax PMEF were purchased from EMD Millipore.

Authentication

All hiPSC lines were assessed for pluripotency, like previously described in Pasca et al, *Nature Medicine* 2011 or Yazawa et al., *Nature* 2011.
 hiPSC lines were assessed for genomic integrity by SNP microarray "GSAMD-24v2-0" (with 759,993 probes).

Mycoplasma contamination

All cell lines and P-MEF cells were tested for Mycoplasma contamination and tested negative

Commonly misidentified lines (See [ICLAC](#) register)

No commonly misidentified cell lines were used