

***NEUROG2* regulates a human-specific neurodevelopmental gene regulatory program**

Vorapin Chinchalongporn^{1,2}, Lakshmy Vasan^{1,3}, Fermisk Saleh^{1,2}, Dawn Zinky^{1,2}, Hussein Ghazale^{1,2}, Ana-Maria Oproescu^{1,3}, Shruti Patel^{4,5}, Matthew Rozak^{4,5}, Yutaka Amemiya^{1,3}, Sisu Han^{1,2}, Alexandra Moffat^{1,3}, Sandra E Black^{6,7}, JoAnne McLaurin^{1,3}, Jamie Near^{4,5}, Arun Seth^{1,3}, Maged Goubran^{4,5}, Orly Reiner⁸, Satoshi Okawa⁹, Carol Schuurmans^{1,2,3*}

Running title: *NEUROG2-mCherry* lineage analysis

Keywords: cerebral organoid, CRISPR gene editing, lineage tracing, proneural gene, *NEUROG2*, extracellular matrix

¹ Sunnybrook Research Institute, Biological Sciences Platform, Hurvitz Brain Sciences Program, 2075 Bayview Ave, Toronto, ON, Canada, M4N 3M5

² Department of Biochemistry, 1 King's College Cir, University of Toronto, ON, Canada, M5S 1A8

³ Department of Laboratory Medicine and Pathobiology, 1 King's College Cir, University of Toronto, ON, Canada, M5S 1A8

⁴ Department of Medical Biophysics, 101 College St Suite 15-701, Toronto General Hospital, University of Toronto, ON, Canada, M5G 1L7

⁵ Sunnybrook Research Institute, Physical Sciences Platform, Hurvitz Brain Sciences Program, 2075 Bayview Ave, Toronto, ON, Canada, M4N 3M5

⁶ Dr. Sandra Black Centre for Brain Resilience & Recovery, LC Campbell Cognitive Neurology Unit, Sunnybrook Research Institute, Toronto, Ontario, Canada; Hurvitz Brain Sciences Program

⁷ Department of Medicine (Neurology) (SEB), University of Toronto, Toronto, Ontario, Canada.

⁸ Departments of Molecular Genetics and Molecular Neuroscience, Weizmann Institute of Science, 76100, Rehovot, Israel

⁹ University of Pittsburgh School of Medicine, Vascular Medicine Institute, Department of Computational and Systems Biology, McGowan Institute for Regenerative Medicine, Pittsburgh, PA, USA

*Correspondence: carol.schuurmans@sunnybrook.ca

SUMMARY STATEMENT

Transcriptomic analyses of *NEUROG2-mCherry* knock-in human embryonic stem cell-derived cerebral organoids reveal a link between *NEUROG2* and extracellular matrix remodeling during human cortical development.

ABSTRACT

Unique hallmarks of human neocortical development include slower rates of neurogenesis and the establishment of an extracellular matrix-rich, outer-subventricular zone that supports basal neural progenitor cell expansion. How gene regulatory networks have evolved to support these human-specific neurodevelopmental features is poorly understood. Mining single cell data from cerebral organoids and human fetal cortices, we found that *NEUROG2* expression is enriched in basal neural progenitor cells. To identify and purify *NEUROG2*-expressing cells and trace their short-term lineage, we engineered two *NEUROG2-mCherry* knock-in human embryonic stem cell lines to produce cerebral organoids. Transcriptomic profiling of mCherry-high organoid cells revealed elevated expression of *PPP1R17*, associated with a fast-evolving human-accelerated regulatory region, oligodendrocyte precursor cell and extracellular matrix-associated gene transcripts. Conversely, only neurogenic gene transcripts were enriched in mCherry-high cortical cells from *Neurog2:mCherry* knock-in mice. Finally, we show that *Neurog2* is sufficient to induce *Ppp1r17*, which slows human neural progenitor cell division, and *Col13a1*, an extracellular matrix gene, in P19 cells. *NEUROG2* thus regulates a human neurodevelopmental gene regulatory program implicated in supporting a pro-proliferative basal progenitor cell niche and tempering the neurogenic pace.

INTRODUCTION

The rodent neocortex is smooth-surfaced, or lissencephalic, whereas, in larger mammals, including humans, neuronal layers are folded to create a gyrencephalic cortical structure marked by gyri and sulci. Understanding how these structural differences arise during development is an intensive area of investigation (Amin and Borrell, 2020; Moffat and Schuurmans, 2023). A key distinction between human and rodent corticogenesis is the evolution of unique neural progenitor cell (NPC) pools along with temporal differences in neuronal differentiation and maturation processes (Wallace and Pollen, 2023). Human NPCs have an overall longer neurogenic period, in part due to increased protein half-lives (Rayon et al., 2020) and longer cell cycle times, which are recapitulated in differentiation assays *in vitro* (Herculano-Houzel, 2012; van den Ameele et al., 2014). Additionally, a multitude of new genes have evolved to drive human brain development and function, originating either via gene duplications, structure/function alterations of ancestral genes or their gene regulatory regions, or from *de novo* origins (Hodge et al., 2019). Several human genes involved in neurodevelopmental processes are associated with fast-evolving regulatory regions, termed Human Accelerated Regions (HARs) (Girskis et al., 2021). One example of a critical HAR-regulated gene is *PPP1R17*, which is expressed in human cortical NPCs and not in rodent or ferret NPCs, and which lengthens the cell cycle when misexpressed in murine cortical cells (Girskis et al., 2021). However, it has yet to be determined how the expression of *PPP1R17* and other HAR-regulated genes are differentially regulated to support the unique developmental trajectory of human neocortices.

Initially, the cortical ventricular zone (VZ) is populated by primary NPCs, termed apical radial glia (aRG), which undergo direct neurogenesis to form deep-layer neurons. Later, aRG give rise to a secondary pool of intermediate progenitor cells (IPCs) that form a more basally-located subventricular zone (SVZ) (Vaid and Huttner, 2022). In rodent cortices, IPCs have a limited proliferative capacity and divide only 1-2 times before differentiating into late-born, upper-layer neurons (Haubensak et al., 2004; Miyata et al., 2004; Noctor et al., 2004). In contrast, in gyrencephalic species, NPCs have diversified and multiplied to drive cortical folding. In these folded cortices, early-born, deep-layer neurons are derived from aRG as well as from an expanded pool of basal radial glia (bRG) that lose their apical attachments and populate an enlarged outer subventricular zone (oSVZ) (Betizeau et al., 2013; Hansen et al., 2010; LaMonica et al., 2013; Martinez-Cerdeno et al., 2012; Shitamukai et al., 2011). Conversely, bRG are rare in

rodent cortices, found mainly in dorsomedial domains (Vaid and Huttner, 2022) and representing only ~5% of all NPCs (Martinez-Cerdeno et al., 2012; Shitamukai et al., 2011; Wang et al., 2011). In species with folded cortices, IPCs act as transit-amplifying cells, dividing multiple times before differentiating through a process termed indirect neurogenesis (Fietz et al., 2010). These rapidly expanding IPCs give rise to many more upper-layer neurons, with the increase in neuronal mass accommodated by the formation of cortical folds that expand the overall surface area (Fernandez and Borrell, 2023; Fietz et al., 2010). To support the expanded proliferative properties of basal progenitors in gyrencephalic species, basal NPCs produce an extracellular matrix (ECM)-rich niche that supports cell division (Amin and Borrell, 2020; Arai et al., 2011; Fietz et al., 2012; Florio et al., 2015; Martinez-Martinez et al., 2016; Pollen et al., 2015).

In invertebrate (e.g., *Drosophila*) and vertebrate (e.g., mouse and man) species alike, proneural genes encode basic-helix-loop-helix (bHLH) transcription factors that are evolutionarily-conserved drivers of neurogenesis (Oproescu et al., 2021; Quan et al., 2016). The proneural gene *Neurog2* is the main driver of murine neocortical development. In the developing rodent neocortex, *Neurog2* is required and sufficient to specify a glutamatergic neuronal identity (Fode et al., 2000; Mattar et al., 2008; Oproescu et al., 2021; Schuurmans et al., 2004). The related gene, *Neurog1*, also specifies a glutamatergic neuronal fate in the developing neocortex, but functions at a slower rate, and tempers the pace of early cortical neurogenesis by heterodimerizing with *Neurog2* to suppress its potent neurogenic activity (Han et al., 2018).

Due to the inaccessibility of fetal brain tissue, uncovering human-specific features of proneural gene function during human neocortical development is a major challenge. To circumvent these issues, 3D brain organoids can be generated from human embryonic stem cells (hESCs) or induced pluripotent stem cells (iPSCs). These organoids can be directed to differentiate into specific neural tissues, such as the cerebral cortex, optic vesicles, midbrain, spinal cord and choroid plexus using activators and inhibitors of signaling pathways that drive normal central nervous system development (Amin et al., 2023; Karzbrun and Reiner, 2019). Cerebral organoids are advantageous as they can be grown for a year or more *in vitro* (Lancaster et al., 2013; Qian et al., 2016) or transplanted in rodent brains for further maturation (Revah et al., 2022; Wilson et al., 2022). Further features of cerebral organoids that show a resemblance to the neocortex; they are comprised of neural progenitor cells (NPCs) that differentiate into

glutamatergic and GABAergic neurons, astrocytes and oligodendrocytes (Karzbrun and Reiner, 2019; Lancaster and Knoblich, 2014; Lancaster et al., 2013; Qian et al., 2016)

Here, we have begun to dissect the role of *NEUROG2* in human cortical neurogenesis using a cerebral organoid model system. We found that *NEUROG2* is expressed in human fetal cortices and in cerebral organoids, with the highest transcript counts detected in IPCs and aRG, followed by smaller subsets of bRG and neurons. By engineering *NEUROG2-mCherry* knock-in (KI) hESC lines, and creating cerebral organoids from these cells, we were able to compare the transcriptomes of mCherry-high versus mCherry-low cells. Unexpectedly, mCherry-high cells were enriched in oligodendrocyte precursor cell (OPC) and ECM-related transcripts, contrasting to the *Neurog2-mCherry*-high lineage in rodent cortices, in which neurogenic gene expression was instead elevated. *PPP1R17* transcripts were also enriched in mCherry-high cells in *NEUROG2-mCherry* KI-hESC-derived organoids, which is of interest given the link between *PPP1R17* and human neocortical evolution (Girskis et al., 2021). Finally, *Neurog2* was sufficient to induce the expression of *Ppp1r17* and *Col3a1*, an ECM gene, in P19 cells. Given the significance of the ECM to basal progenitor cell expansion (Amin and Borrell, 2020; Arai et al., 2011; Fietz et al., 2012; Florio et al., 2015; Martinez-Martinez et al., 2016; Pollen et al., 2015), and human *PPP1R17* to slowing down cell cycle progression and overall rates of neurogenesis (Girskis et al., 2021), our study highlights important ways in which *NEUROG2* function has adapted to regulate human-specific features of neocortical development.

RESULTS

***NEUROG2* is the predominant proneural gene in human fetal cortices and cerebral organoids**

In rodent cortices, *Neurog2* is the predominant proneural gene, whereas the related gene *Neurog1* is expressed at lower levels and in a shorter time window (Han et al., 2018; Moffat et al., 2023). To determine whether *NEUROG2* similarly predominates during human cortical development, we assayed *NEUROG1* and *NEUROG2* expression profiles in existing scRNA-seq data from human fetal cortical cells collected between gestational week (GW) 08 to GW26, during the period of active neurogenesis (Zhong et al., 2018). Of the cells assigned an NPC identity, the majority expressed *NEUROG2* (72.06%), either together with *NEUROG1* (38.62%)

or alone (33.4%) (Fig. 1A). In contrast, *NEUROG1* was expressed in fewer human fetal cortical cells (47.93%) and only 9.31% of cortical NPCs expressed *NEUROG1* alone (Fig. 1A). The expression of *NEUROG1* and *NEUROG2* remained elevated from GW 09 and persisted until GW 26 (Fig. 1A), contrasting to the reduction in *Neurog1* expression and double⁺ cells observed at the midpoint (*i.e.*, embryonic day 15.5) of murine cortical development (Han *et al.*, 2018).

We circumvented the inaccessibility of fetal brain tissues by using human cerebral organoids, as a surrogate model of human corticogenesis, to assay proneural gene function. To confirm that hESC-derived cerebral organoids expressed *NEUROG1* and *NEUROG2*, and to identify expressing cell types, we mined scRNA-seq data from 90-day old cerebral organoids generated with a Lancaster protocol (Sivitilli *et al.*, 2020). Unbiased Seurat clustering produced 19 cell clusters, which were then stratified into distinct cell types based on the expression of cell-type specific markers (Fig. 1B-E, Table S1), as previously described (Sivitilli *et al.*, 2020). We found that *NEUROG1* and *NEUROG2* expression was highest in IPCs, of which, more than 50% expressed one of these proneural genes, with *NEUROG2* transcripts detected in more cells than *NEUROG1* mRNA (Fig. 1C-E). *NEUROG1* and/or *NEUROG2* transcripts were also detected in ~40% of aRG and ~10% of bRG, with *NEUROG2* transcripts similarly more frequently detected in these cells compared to *NEUROG1* mRNA (Fig. 1C-E). Finally, we also detected *NEUROG2* transcripts in a small percentage of cortical neurons, whereas *NEUROG1* expression was nearly absent in neuronal populations (Fig. 1C-E). Thus, *NEUROG2* expression is more widespread among human fetal cortical cells and human cerebral organoid cells than *NEUROG1*.

To understand the lineage dynamics associated with human proneural gene expression during corticogenesis, we computationally isolated cortical cells expressing *NEUROG1* and/or *NEUROG2* from scRNA-seq data of 90-day-old human cerebral organoids (Sivitilli *et al.*, 2020) and generated pseudotime trajectories. The resultant pseudotime ordering of cells along a lineage trajectory revealed a single branch point and three cell states (Fig. 2A,B). State 1 cells had the earliest pseudotime identity and included the highest fraction of *NEUROG1/NEUROG2* double⁺ and *NEUROG1* single⁺ cells, and the lowest fraction of *NEUROG2* single⁺ cells (Fig. 2B-D). We determined that state 1 cells had the highest levels of aRG, bRG, and proliferating cell-associated transcripts by mapping the expression of cell identity markers onto the lineage trajectory (Fig. 2E, F; Fig. S1A; Table S1). In state 2 cells, which had an intermediate pseudotime identity, *NEUROG1* expression and *NEUROG1/NEUROG2* co-expression declined, while the fraction of

NEUROG2 single⁺ cells increased slightly (Fig. 2B-D). In state 2 cells, aRG/bRG-marker expression declined, while there was an increase in IPC markers, especially in *NEUROG2* single⁺ cells (Fig. 2E, F). Finally, state 3 cells, which had the latest pseudotime identities, predominantly expressed *NEUROG2* alone, with very low fractions of the *NEUROG1/NEUROG2* double⁺ and *NEUROG1* single⁺ pools (Fig. 2B-D). State 3 cells displayed elevated levels of early neuronal marker transcripts, including deep-layer 5 and 6 markers (Fig. 2B-F, Table S1). Thus, *NEUROG1* expression is confined to early stages of neural lineage development in human cerebral organoids, and is enriched in aRG and bRG. In contrast, while *NEUROG2* is also expressed in aRG and bRG, the proportion of *NEUROG2*-expressing cortical cells increases as they mature into IPCs, and transcripts are also detected in neurons in later pseudotime states 2 and 3.

CRISPR engineering to generate *NEUROG2-mCherry*-KI hESCs for cerebral organoid production

For the remainder of this study, we focused on *NEUROG2*, given its predominant expression in human cerebral lineages. To isolate and characterize *NEUROG2*-expressing cells from cerebral organoids, we used CRISPR/Cas9 to engineer *NEUROG2-mCherry* KI reporter hESC lines (Fig. S2A,B). Individual clones were screened for homology-directed repair using digital droplet PCR (Fig. S2B,C). Of the 87 clones that were selected for screening, 4 clones had a correctly targeted *NEUROG2-mCherry* insertion (4.6% target efficiency). To ensure that the gene targeting process did not introduce any unintended mutations in the *NEUROG2* locus, we sequenced the knock-in clones at the 5' and 3' junction of the Cas9 targeted site. No mutations were detected in the two correctly targeted clones, which we designated as lines #105 and #117 (Fig. S2D). In addition, both *NEUROG2-mCherry* KI hESC lines expressed pluripotency genes (*OCT4*, *SOX2*, *NANOG*), and did not show genomic abnormalities in the most common mutated regions reported in hESCs (Fig. S3A-C). We thus followed through with the two correctly targeted *NEUROG2-mCherry* KI hESC clones (#105, #117) for downstream experiments.

To investigate *NEUROG2* function, we employed a modified Lancaster protocol incorporating dual Smad inhibitors to generate hESC-derived cerebral organoids that are enriched in cortical gene expression (Qian *et al.*, 2018) (Fig. 3A,B). To confirm neural induction

and organoid maturation with this approach, cerebral organoids were isolated on day 42, optically cleared, co-immunostained with SOX2 to label NPCs and TUJ1 (β 3-tubulin) to mark newborn neurons, and imaged in wholmount using lightsheet microscopy (Fig. 3C). 3D imaging revealed a network of TUJ1⁺ axonal fibers encasing the 42-day old cerebral organoids. In some instances, neural rosettes comprised of an inner SOX2⁺ NPC layer and outer TUJ1⁺ neuronal layer appeared as external protrusions (Fig. 3C). In 2D sections, day 30-35 organoids contained neural rosettes that expressed the pan-NPC markers SOX2 and NESTIN, as well as PAX6, a regional marker of a cortical NPC identity (Fig. 3D). Newborn neuronal markers, such as DCX and TUJ1, labeled neurons and their processes on the outside of the neural rosettes (Fig. 3D).

To confirm the competence of *NEUROG2-mCherry* KI hESC clones to generate cerebral organoids, we followed the same dual SMAD inhibitor protocol. Notably, mCherry expression was more widespread than *NEUROG2* due to the inherent characteristics of the proteins involved: mCherry is a stable protein with a long intracellular half-life, while *NEUROG2* has a relatively short half-life of approximately 30 minutes (Li *et al.*, 2012). The long half-life of mCherry makes it suitable for lineage tracing of cells derived from *NEUROG2*⁺ NPCs, at least in the short-term. In day 18 organoids, we detected robust co-expression of *NEUROG2* and mCherry in neural rosettes encircling the periphery of the organoid (Fig. 4A). mCherry was co-expressed with additional NPC markers, including SOX2 (Fig. 4B) and PAX6 (Fig. 4C), consistent with a neural and cortical NPC identity, respectively. By day 45, SOX2 expression remained widespread in neural rosettes, whereas *NEUROG2* and PAX6 were expressed in relatively fewer cells by this stage of organoid development (Fig. 4D-F). In contrast, mCherry expression persisted in scattered rosettes throughout the day 45 organoids, albeit with limited co-localization between mCherry and NPC markers (Fig. 4D-F). In the BrainSpan Atlas, *NEUROG2* transcript levels similarly decline between GW 9 and 12 in human fetal cortices and are reduced over time in cortical brain organoids (Cheroni *et al.*, 2022). Overall, these data demonstrate that mCherry expression persists after *NEUROG2* expression declines, allowing us to use mCherry to profile the phenotype of organoid cells derived from *NEUROG2*⁺ NPCs.

Transcriptomic analyses reveal a link between *NEUROG2*, the ECM and oligodendrocyte precursor cells.

To examine the molecular phenotype of cells in the *NEUROG2*⁺ lineage, cerebral organoids were generated from two *NEUROG2-mCherry* KI hESC lines (#105 and #117) in two independent experiments. Cerebral organoids were harvested after 45-47 days *in vitro*, at a stage when cerebral organoids, generated with the Qian et al. (2016, 2018) protocol, share a transcriptional signature with GW 9-12 fetal cortices, according to the BrainSpan Atlas (Cheroni et al., 2022). mCherry-high and mCherry-low cells were isolated by FACS from pools of 7-8 cerebral organoids and the enrichment of *mCherry* transcripts in mCherry-high versus -low cells was verified by qPCR (Fig. 4G). To profile gene expression, we used targeted transcriptome sequencing with a human gene expression assay that covers 20,802 human genes, representing >95% of the UCSC reference genome (Table S2). Principal component analysis (PCA) of gene expression datasets revealed that mCherry-high versus mCherry-low cells were transcriptionally divergent and segregated from each other irrespective of the initiating cell line or experimental day (Fig. 4H). A comparative analysis of differentially expressed genes (DEGs) in mCherry-high versus mCherry-low cells identified 1204 genes enriched in mCherry-high cells and 263 genes enriched in mCherry-low cells (Fig. 4I, Table S3). Biological Process gene ontology (GO) analysis of DEGs revealed an enrichment of terms associated with the extracellular matrix (ECM) in mCherry-high cells, including '*extracellular matrix organization*' and '*collagen fibril organization*' (Fig. 4J; Table S4). Conversely, neuronal differentiation GO terms such as '*generation of neurons*' and '*neuron differentiation*' were over-represented in mCherry-low cells (Fig. 4K; Table S5).

To characterize the two populations of cerebral organoid cells more thoroughly, we selectively assayed transcript counts for known genes involved in cortical development, focusing on transcripts with a log₂-fold-change (log₂FC) greater than 1 (*i.e.*, a doubling in the original scaling) and with a padj<0.05, corresponding to a false discovery rate cutoff of 0.05. Markers of apical RG, basal RG and proliferating cells were expressed at relatively equivalent levels in mCherry-high and mCherry-low cerebral organoid cells (Fig. S4A-C, Table S2). Of the previously characterized human IPC markers, only *PPP1R17* (Pebworth et al., 2021) and *NHLH1* were expressed at detectable levels in day 45-47 cerebral organoids, and strikingly, there was a 3.2 log₂FC in *PPP1R17* transcripts, which were elevated in mCherry-high cells (Fig. S4D). Thus, *NEUROG2-mCherry* KI hESC-derived cerebral organoids are comprised of both mCherry-high and low cells that express relatively equivalent levels of aRG, bRG and IPC

markers, with the exception of an enrichment in *PPP1R17* gene transcripts in mCherry-high IPCs.

We next assessed transcript levels for pan-neuronal markers in the two cerebral organoid populations to characterize lineage preference (Table S2). Of these pan-neuronal markers, only *GAP43* was enriched in mCherry-high versus mCherry-low cerebral organoid cells at day 45-47 (Fig. S4E). In addition, transcripts for two glutamatergic-specific neuronal genes, *SLC17A6/VGLUT2* (1.03 log2FC) and *SLC17A7/VGLUT1* (2.36 log2FC), were more abundant in mCherry-high cerebral organoid cells (Fig. S4F). This finding was anticipated given that *Neurog2* is necessary and sufficient to specify a glutamatergic neuronal fate (Dixit et al., 2014; Li et al., 2012; Schuurmans et al., 2004). Unexpectedly, *NEUROD4* and *NEUROD6*, which act downstream of *NEUROG2* in a neurogenic cascade, were slightly enriched in mCherry-low cells, possibly due to the reduction in mCherry expression as NPCs differentiate (Fig. S4F). In addition, two GABAergic neural lineage markers (e.g., *ASCL1*, *GAD2*) were elevated in the mCherry-low lineage (Fig. S4G). These findings are in keeping with the known cross-repressive interactions between *Neurog2* and *Ascl1*, a different proneuronal gene that specifies GABAergic neuronal fate in cortical NPCs (Han et al., 2021; Schuurmans et al., 2004).

Next, we examined layer 1-6 neuronal markers to conclude our characterization of neuronal diversity in the cerebral organoid populations (Fig. S4I, Table S2). There were relatively equal transcript levels for most laminar markers in mCherry-high and mCherry-low cerebral organoid cells, except for increased transcript counts for the layer 5 marker *PCP4* (1.68 log2FC) and the layer 1 marker *RELN* (2.05 log2FC) in mCherry-high cerebral organoid cells. Conversely, the layer 5 marker *BCL11B* (1.45 log2FC) and layer 6 marker *FOXP2* (1.21 log2FC) transcript counts were higher in mCherry-low cells (Fig. S4I). These data suggest that there may be unique laminar biases for neurons derived from *NEUROG2* positive versus negative NPCs.

Lastly, we examined the representation of glial cells amongst the mCherry-high and mCherry-low populations. We detected transcripts for several astroglial cell markers at equivalent levels in mCherry-high versus low cells (Fig. S4H). In contrast, 28/221 genes expressed in OPCs and/or involved in oligodendrocyte development (GO:0048709) were among the DEGs that were upregulated in mCherry-high cerebral organoid cells (Fig. 4L, Table S6). Taken together, these data reveal an unexpected enrichment of *PPP1R17*, an IPC-expressed gene

that is associated with a fast-evolving human-accelerated regulatory region, and OPC-related genes in *NEUROG2-mCherry* KI-high cerebral organoid cells.

ECM protein deposition adjacent to mCherry⁺ cells in *NEUROG2-mCherry* KI hESC-derived cerebral organoids

A pro-proliferative ECM niche that supports integrin signaling is produced by basal NPCs in human and not in rodent cortices (Amin and Borrell, 2020; Arai *et al.*, 2011; Fietz *et al.*, 2012; Florio *et al.*, 2015; Martinez-Martinez *et al.*, 2016; Pollen *et al.*, 2015). However, even though ECM genes are transcribed in germinal compartments in human fetal cortices (Fietz *et al.*, 2012), ECM proteins are deposited in the cortical plate, such that transcripts and proteins are not detected in the same compartments (Long *et al.*, 2018). Given the enrichment of GO terms associated with the ECM in *NEUROG2-mCherry* KI-high cerebral organoid cells, we further mined transcript read count data for ECM-related genes from GO: 0030198, of which 76/327 genes were among the DEGs (Table S7). ECM genes that were expressed at elevated levels by mCherry-high cerebral organoid cells included collagens, and ADAM and matrix metalloproteases, which are involved in ECM remodeling (Fig. 4M, Table S2).

To examine whether ECM proteins were detected in *NEUROG2-mCherry* KI hESC-derived cerebral organoids, we performed immunostaining with COL4 and LAM. In day 18 cerebral organoids, robust expression of COL4 and LAM was detected in circular formations surrounding the necrotic core, as well as in protrusions into the organoid periphery, where mCherry⁺ cells were located (Fig. 5A). In higher magnification images, there was limited overlap between COL4 or LAM and mCherry (Fig. 5A), which was not necessarily surprising given that ECM proteins are produced by NPCs but finally deposited in the cortical plate. At day 45, there was a change in the distribution of ECM proteins in *NEUROG2-mCherry* KI hESC-derived cerebral organoids. At this stage, COL4⁺ and LAM⁺ fibrils invaded the neural rosettes, and LAM was enriched in the periphery of the organoids (Fig. 5B). A necrotic core was no longer visible at this stage (Fig. 5B). To better visualize the distribution of ECM proteins at a later stage, tissue clearing, immunolabeling, and lightsheet fluorescence microscopy was performed on day 119 organoids. In these cleared organoids, mCherry⁺ cells were scattered throughout the organoid

interior, COL4 was similarly detected in the center of the organoids, and LAM proteins encased the outer surface of the organoid (Fig. 5C,D).

Since the enrichment of ECM gene expression in mCherry-high cerebral organoid cells was unexpected, we confirmed this enrichment by performing qPCR on the FACS-enriched mCherry-high and -low cells, focusing on *COL1A2* (Fig. 5E). To independently validate the significance of the association between *NEUROG2* and ECM gene expression, we mined a bulk RNA-seq dataset from a study in which the human fetal cortex was dissected into germinal zones, leading to the conclusion that ECM genes support the proliferation of human NPCs (Fietz et al., 2012). In this dataset, several collagen genes, especially *COL1A2*, *COL4A1*, and *COL4A2*, were expressed at elevated levels in the VZ as well as inner and outer SVZ, in compartments in which *NEUROG2* transcript levels were also elevated (Fig. 5F). In conclusion, while mCherry⁺ cells do not directly co-localize with the ECM proteins, *NEUROG2*-expressing NPCs transcribe ECM genes that are then translated into proteins that are deposited throughout the cerebral organoid.

***Neurog2* does not lineage trace ECM or OPC genes in the early embryonic murine cortex**

The unforeseen enrichment of OPC and ECM genes in *NEUROG2-mCherry*-high organoid cells prompted us to compare to murine *Neurog2-mCherry* cortical lineages. We previously generated a *Neurog2-mCherry* KI transgenic line, from which we performed RNA-seq on mCherry-high versus proneural gene-‘negative’ NPCs collected from embryonic day (E) 12.5 cortices (Fig. 6A) (Han et al., 2021). E12.5 murine cortices are roughly at the same stage as GW10 human cortices, and have the equivalent of a day 45-47 cerebral organoid gene signature (Cheroni et al., 2022). ‘Proneural-negative’ NPCs also did not express *Ascl1*, which were negatively selected against using GFP expression from E12.5 *Neurog2-mCherry;Ascl1-GFP* double KI transgenic mice (Han et al., 2021). FACS was used to select mCherry⁺ cells that co-expressed CD15, an NPC marker, and ‘proneural-negative’ NPCs that expressed CD15 without mCherry or GFP expression were also collected (Han et al., 2021).

Comparisons of bulk RNA-seq datasets from mCherry-high and proneural negative NPCs identified 1112 DEGs (padj<0.05; Table S8). Within the “*positive regulation of neurogenesis*” GO: 0050769 category, 61/329 genes were DEGs when comparing mCherry-high to proneural

negative cortical cells (Fig. 6B; Table S9). Further GO term enrichment analysis identified multiple “Biological Process” GO terms associated with neurogenesis that were enriched in *Neurog2-mCherry* high cortical NPCs, including “*axonogenesis*”, “*generation of neurons*”, and “*nervous system development*” (Fig. 6C). Conversely, “*canonical Wnt signaling*” and “*extracellular matrix organization*” GO terms were under-represented in *Neurog2-mCherry*-high cortical NPCs (Fig. 6D), which is the opposite of the *NEUROG2-mCherry* data collected from cerebral organoids. Moreover, neither Collagen nor Laminin expression was detected in E14.5 *Neurog2-mCherry* KI cortices, despite their enrichment in the basal lamina at the pial surface (Fig. S6A-C). These data suggest that the association between *NEUROG2* and ECM gene expression is unique to human cortical lineages and are consistent with the finding that ECM production by human basal NPCs is a newly evolved driver of their increased proliferative potential (Amin and Borrell, 2020; Arai et al., 2011; Fietz et al., 2012; Florio et al., 2015; Martinez-Martinez et al., 2016; Pollen et al., 2015).

To further characterize *Neurog2*-expressing lineages in the early embryonic murine cortex, we next examined whether known determinants of murine neural cell fates and drivers of cortical neurogenesis were amongst the DEGs. A subset of aRG markers, including *Fabp7* (1.6 log2FC), *Hes1* (2.03 log2FC) and *Gli3* (1.26 log2FC), were enriched in *Neurog2-mCherry*-negative cortical NPCs (Fig. S5A). These data are consistent with the idea that *Neurog2* expression is turning on just as aRG transition to IPCs (Moffat et al., 2023), and/or suggest that there is a *Neurog2*-negative aRG lineage. Conversely, mCherry-high NPCs were enriched in the expression of IPC markers, including *Nhlh1* (6.48 log2FC), *Neurod1* (5.21 log2FC) and *Eomes* (5.19 log2FC) (Fig. S5B). Notably, *Ppp1r17* transcripts, which were elevated in mCherry-high cerebral organoid cells, and mark human IPCs, were not detected in this data set. Moreover, proliferation markers were expressed at similar levels in the mCherry-high/-negative cells, as observed in human cerebral organoids (Fig. S5C).

An expected enrichment of transcripts for *Tbr1* (5.97 log2FC), a layer 6 marker, and *Pcp4* (1.73 log2FC) and *Bcl11b* (2.12 log2FC); layer 5 markers, was observed in *Neurog2-mCherry*-high cells (Fig. 6B). In contrast, most upper-layer markers were expressed at low levels, with the exception of *Pou3f2* (Fig. S5D). The enrichment of early born neuronal markers is consistent with E12.5 as the stage of analysis, when early-born neurons are generated. Transcripts for pan-neuronal markers, including *Tubb3* (3.65 log2FC), *Dll1* (2.87 log2FC),

Stmn2 (4.65 log2FC), and *Gap43* (2.41 log2FC) were all also enriched in mCherry-high cortical cells (Fig. S5E). Notably, at this early stage of development, glial genes were not expressed at high enough levels to compare. Thus, although mCherry is preferentially enriched in IPC subpopulations in human and murine cortices, ECM gene expression is only elevated in mCherry-high cells from human cerebral organoids.

Murine *Neurog2* induces *Ppp1r17* and *Col3a1* gene expression in P19 cells

Neurog2 is expressed in the vast majority of basal NPCs during early murine corticogenesis (Moffat et al., 2023), yet ECM gene expression is not detected in the germinal compartments. However, NEUROG2 transcriptional activity is dosage sensitive (Kovach et al., 2013), with phosphorylation of SP/TP sites by proline-directed serine-threonine kinases (e.g., GSK3, CDKs) inhibiting the ability of NEUROG2 to transactivate its target genes (Ali et al., 2011; Hardwick et al., 2015; Li et al., 2012). Thus, we asked whether *Neurog2* was sufficient to induce ECM gene expression when overexpressed at supraphysiological levels, using a native *Neurog2* gene or a mutated version (*Neurog2*^{SA9TA1}) containing nine serine-to-alanine and a threonine-to-alanine mutations. We specifically asked whether *Neurog2* and/or *Neurog2*^{SA9TA1} were able to induce the expression of ECM and OPC genes, as well as *Ppp1r17*, a critical regulator of the timing of human cortical neurogenesis. P19 cells were transfected with pCIG2-GFP, pCIG2-*Neurog2*-GFP or pCIG2-*Neurog2*^{SA9TA1}-GFP expression vectors. After 2 days *in vitro*, the ability of *Neurog2* and *Neurog2*^{SA9TA1} to induce the expression of known target genes was first assessed by qPCR. *Neurog2*^{SA9TA1} was significantly better at turning on the expression of *Dll3*, *Rnd2* and *Neurod1* than native *Neurog2*, as anticipated (Fig. 6E). *Neurog2*^{SA9TA1} also induced the ectopic expression of *Ppp1r17*, a human IPC gene (Fig. 6F) and *Col3a1*, an ECM gene (Fig. 6G), at higher levels than *Neurog2*. However, the overexpression of *Neurog2* or *Neurog2*^{SA9TA1} could not induce OPC gene expression in P19 cells (Fig. 6H). Thus, even though the *Neurog2*⁺ murine cortical lineage does not normally express ECM genes or *Ppp1r17*, *Neurog2* is sufficient to induce the expression of at least a subset of these genes in murine P19 cells, especially when inhibitory phosphoacceptor sites are mutated.

DISCUSSION

In this study, we engineered *NEUROG2-mCherry* KI reporter hESCs to facilitate cerebral organoid modeling of *NEUROG2* function during human cortical development. We found that in derivative organoids, *NEUROG2* expression is enriched in IPCs, basal NPCs that occupy an expanded oSVZ during human neocortical development. Using short-term lineage tracing, we further found that *NEUROG2* regulates a human-specific neurodevelopmental gene regulatory program. Key human-specific genes regulated by *NEUROG2* include *PPPIR17*, a phosphatase-encoding gene that slows down cell cycle progression and overall rates of neurogenesis (Girsikis et al., 2021), and ECM remodeling genes, which drive basal progenitor cell expansion (Amin and Borrell, 2020; Arai et al., 2011; Fietz et al., 2012; Florio et al., 2015; Martinez-Martinez et al., 2016; Pollen et al., 2015). These findings provide important new insights into the molecular mechanisms that distinguish the gene regulatory pathways that control murine versus human neurogenesis, allowing for the formation of lissencephalic versus gyrencephalic cortices, respectively.

The ECM regulates tissue stiffness, creating biophysical signals that operate together with biochemical (*e.g.*, morphogens) cues in the microenvironment (*i.e.*, niche) to control NPC proliferation and differentiation (Chan et al., 2017; Chaudhari et al., 2014; Gattazzo et al., 2014; Guilak et al., 2009; Mohyeldin et al., 2010). These biophysical and biochemical signals differ across species and are continually remodeled as the brain develops and matures into adulthood (Nelson, 2009; Rozario and DeSimone, 2010). In rodent cortices, ECM expression levels are high in the VZ, where proliferative, aRG reside, but not in the SVZ, where IPCs have a limited proliferative potential (Arai et al., 2011; Florio et al., 2015; Pollen et al., 2015). In contrast, in gyrencephalic cortices, bRG and IPCs express high levels of ECM genes to support integrin signaling and create a pro-proliferative, oSVZ niche (Amin and Borrell, 2020; Arai et al., 2011; Fietz et al., 2012; Martinez-Martinez et al., 2016). As a result, bRG and IPCs, which have lost constraining attachments to the ventricular surface, proliferate extensively to support cortical folding. By demonstrating that *NEUROG2* is expressed in IPCs and regulates ECM gene expression, we have found a potential explanation for the expanded ECM production by basal NPCs during human cortical development.

The link between *NEUROG2-mCherry* expression and the oligodendrocyte lineage was also unexpected but may relate to the existence of Neurog2-Ascl1 co-expressing cortical NPCs in the rodent cortex. These cortical NPCs have a mixed potential to give rise to neurons or

oligodendrocytes, as we previously demonstrated (Han et al., 2021). Indeed, OPC marker expression was observed in *Neurog2-mCherry* cortical cells from rodents at later stages of development (postnatal day – 2) (Han et al., 2021), which is the period of oligodendrocyte differentiation in rodents (Kessaris et al., 2006). In gyrencephalic cortices, neurogenic and oligodendrogenic NPCs co-exist at the same time during the expansion phase of the oSVZ (Martinez-Cerdeno et al., 2012; Rash et al., 2019; Zecevic et al., 2005). In this study, we showed that Neurog2 regulates ECM production, and ECM molecules promote OPC proliferation (Hu et al., 2009), suggesting that NEUROG2 may support the expansion of OPCs in an autocrine fashion. SOX9 is similarly expressed in basal NPCs in the human and ferret cortex, and its overexpression in murine basal NPCs induces ECM gene expression, downregulates Eomes, and induces the expression of gliogenic genes such as Olig2 (Guven et al., 2020). The relationship between SOX9 and NEUROG2 during ECM deposition in human cortical development will be an area of future study.

A marked number of new genes have evolved to drive human brain development and functioning, originating either via gene duplications, structural/functional alterations of ancestral genes or their gene regulatory regions, or from *de novo* origins (Hodge et al., 2019). Our study suggests that human *NEUROG2* has acquired new target genes that contribute to the diversification of the developmental trajectories of cortical neural stem and progenitor cells. Altered target selection may arise due to modifications to the regulatory regions of target genes, new protein-protein interactions that change NEUROG2 activity, or other post-translational alterations. However, alterations to the core *NEUROG2* coding sequence seem unlikely to be the central driver of target switching. Indeed, human NEUROG2 protein has 84.2% homology with murine NEUROG2, with 99% homology in the bHLH domain, in keeping with the high degree of conservation of bHLH activity across invertebrate and vertebrate species (Guillemot and Hassan, 2017). Such conservation means that bHLH genes can compensate for each other to some extent. For instance, even though *Neurog2* is normally expressed in the dorsal telencephalon (i.e., the neocortical anlage), *Neurog2* can rescue ventral telencephalic neurogenesis defects when knocked-into the *Ascl1* locus (Parras et al., 2002). Moreover, when expressed at physiological levels with this knock-in approach, *Neurog2* does not respecify ventral NPCs, which still acquire a GABAergic identity (Parras et al., 2002). However, when *Neurog2* is overexpressed via electroporation under the control of a strong CAGG promoter, the

supraphysiological levels allow *Neurog2* to respecify ventral telencephalic NPCs to a dorsal, glutamatergic identity (Kovach et al., 2013; Mattar et al., 2008). Thus, the proneural genes are dosage sensitive, raising the possibility that levels of expression distinguish target gene selection in murine and human cortical NPCs.

Proneural transcription factors induce neural stem cells to differentiate and sustain neural lineage fidelity in two ways: they directly transactivate downstream genes and they act as pioneer transcription factors, meaning they bind target sites in ‘closed’ chromatin and unwind tightly packed nucleosomes so other neurogenic transcription factors can bind (Aydin et al., 2019; Chanda et al., 2014; Fernandez Garcia et al., 2019; Park et al., 2017; Păun et al., 2023; Raposo et al., 2015; Smith et al., 2016; Soufi et al., 2015; Wapinski et al., 2017; Wapinski et al., 2013). By altering chromatin accessibility, proneural transcription factors may be central architects of the widespread epigenomic remodeling that occurs as NPCs traverse neurodevelopment and adulthood (Yang et al., 2023). Since pioneer transcription factor activity can be facilitated by interactions with chromatin remodeling complexes (Minderjahn et al., 2020) (e.g., ASCL1 protein interacts with chromatin remodelers) (Păun et al., 2023) or by interactions with transcription factor partners (Cernilogar et al., 2019), changes to binding partners may account for differences in how murine and human NEUROG2 operate. Endogenous *Ascl1* has bona fide pioneer functions *in vivo* (Păun et al., 2023), but *Neurog2* has only been studied via over-expression (Aydin et al., 2019). Since over-expressed TFs can gain non-physiologic pioneer activity (Hansen et al., 2022), it is important that future studies examine pioneering functions of endogenous proneural transcription factors. The *NEUROG2-mCherry* KI hESCs generated in this study will facilitate this future area of study.

We and others found that proneural transcription factors are inhibited by phosphorylation in embryonic mouse brain and spinal cord, in *Xenopus* and fruit fly primary neurogenesis, and in cancer cells (Ali et al., 2011; Ali et al., 2020; Azzarelli et al., 2015; Azzarelli et al., 2022; Ge et al., 2006; Hand et al., 2005; Hindley et al., 2012; Li et al., 2014; Li et al., 2012; Quan et al., 2016; Sun et al., 2001). We used this knowledge to mutate *Ascl1* so that it cannot be phosphorylated and remains active, and showed that this designer proneural transcription factor is more efficient at lineage conversion of an astrocyte to an induced neuron than native *Ascl1* (Ghazale et al., 2022). In this study, we used a similar strategy to mutate SP and TP sites in *Neurog2* to generate *Neurog2^{SA9TA1}*, and we demonstrated that this mutated version has an

enhanced capacity to transactivate known *Neurog2* target genes, *Ppp1r17* and *Col3a1*. Thus, target gene selection may be differentially modulated in human versus rodent cortical NPCs depending on phosphorylation status of *NEUROG2*. Intriguingly, *SOX9* is similarly expressed in basal NPCs in the human and ferret cortex, and its overexpression in murine basal NPCs induces ECM gene expression cell autonomously and non-autonomously (Guven *et al.*, 2020). Thus, *NEUROG2*-expressing basal NPCs may be similarly influenced by signals emitted by other cells in the environment that alters, for example, phosphorylation status.

Taken together, these findings reveal that human *NEUROG2* and murine *Neurog2* have distinct transcriptional targets and altered functions during neocortical development. Future studies will be aimed at more comprehensively elucidating human-specific alterations to *NEUROG2* regulation and selection of downstream transcriptional targets during human neocortical development.

MATERIALS AND METHODS

Animal sources and maintenance. All animal procedures were performed with the approval of the SRI Animal Care Committee (AUP 21-769) and followed the Canadian Council on Animal Care guidelines. Animals were housed in the Sunnybrook Research Institute (SRI) Comparative Research Unit (CRU, SRI, Toronto, ON, Canada) with *ad libitum* access to food and water. *Neurog2*^{Flag-mCherry}^{KI/+} mice were maintained on a CD1 background and genotyped as described (Han *et al.*, 2021). The day of the vaginal plug was considered E0.5 for embryonic staging.

hESC maintenance. Human embryonic stem cells (hESCs) (H1/WA01) were purchased from WiCell Research Institute, Wisconsin, USA. hESC usage for this project was approved by the Canadian Stem Cell Oversight Committee (SCOC application to CS and to CS and JN) as well as by SRI's Research Ethics Board (REB Project Identification Number: 5003). Briefly, hESCs were cultured under feeder-free conditions in mTeSR Plus media (Stem Cell Technologies, 100-0276) on plates coated with Matrigel (Corning, 354277) and maintained in 5% CO₂ incubators at 37°C. Versene (Thermo Fisher Scientific, 15040-066) was used to dissociate hESCs by manual pipetting every 4-5 days for maintenance. hESC cultures were monitored daily for differentiated

cells, which were removed by manual scraping. Prior to generating cerebral organoids, quality control tests were routinely performed on hESC cultures, using a human stem cell pluripotency detection qPCR kit (Sciencecell, 0853) and hPSC genetic analysis kit (Stem Cell Technologies, 07550), according to the manufacturer's protocols.

CRISPR-Cas9 Gene-Editing. Clustered regularly interspaced short palindromic repeats (CRISPR) genome editing was used to insert an mCherry reporter gene into the 3' untranslated region (UTR) of the *NEUROG2* locus. The vector pSpCas9(BB)-2A-GFP (PX458) (Addgene, Plasmid 48138), was purchased to target *NEUROG2*. To promote homology-directed repair of the Cas9-cleaved *NEUROG2* target locus, we co-electroporated: (1) a CRISPR plasmid containing SpCas9-2A-eGFP and a single guide RNA (sgRNA) to *NEUROG2* that was a fusion of a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA); and (2) a repair template containing homology arms flanking an mCherry reporter cassette. 1.5×10^6 cells were transfected with 1 μ g DNA (500 ng Cas9 plasmid and 500 ng linearized donor plasmid) by nucleofection (pulse code CA137) using P3 Primary Cell 4D-Nucleofector X kit (Lonza, V4XP-3024) in a 4D-Nucleofector (Lonza, AAF-1003B) and were plated in a 6-well plate containing 10 μ M Rock inhibitor Y-27632 (Stem Cell Technologies, 72302). Twenty-four hours post transfection, cells were harvested for quantification of transfection efficiency by flow cytometry for EGFP expression, indicative of Cas9 transfection. To isolate individual clones, positive GFP cells were sorted and replated at clonal density in multiple 10-cm plates. Each clone was expanded for 7-10 days and frozen while some cells were lysed for gDNA extraction for HDR screening with digital droplet PCR (ddPCR).

Individual clones were then picked and further expanded, followed by genomic DNA extraction to screen for clones that underwent homology-directed repair (HDR) using ddPCR. For ddPCR, we employed a forward primer-probe upstream to the starting point of the homology arm region and a reverse primer-probe that only bound to a site inside the exogenous mCherry sequence. The selected two correctly targeted clones were used for downstream experiments.

Digital Droplet PCR. QX200 droplet digital PCR (ddPCR) system (Bio-Rad) was used for all ddPCR reactions. Detailed information for all primers is in Table S10. For HDR screening, the

absolute number of *NEUROG2-mCherry* KI gene copies per cell was quantified and normalized to *RPP30* (Bio-Rad, 10031243). 20 ng of genomic DNA was used in 20 μ l PCR reaction containing 900 nM of the forward and reverse *NEUROG2-mCherry* KI and *RPP30* primers, 250 nM of *NEUROG2-mCherry* KI and *RPP30* probes, and 10 μ l of 2X ddPCR supermix for probes (Bio-Rad). Assay mixtures were loaded into a droplet generator cartridge (Bio-Rad), followed by the addition of 70 μ L of droplet generation oil for probes (Bio-Rad) into each of the eight oil wells. The cartridge was then placed inside the QX200 droplet generator (Bio-Rad). Generated droplets were transferred to a 96-well PCR plate (Eppendorf), which was heat-sealed with foil and placed in C1000 Touch Thermal Cycler (Bio-Rad). Thermal cycling conditions were as follows: 95°C for 10 min, 44 cycles of 94°C for 30 s, 53°C for 1 min, and 98°C for 10 min. FAM fluorescent signal, which labeled the *NEUROG2-mCherry* KI DNA sequence, and HEX fluorescent signal which labeled the *RPP30* DNA sequence, were counted by a QX200 digital droplet reader and analyzed by QuantaSoft analysis software ver.1.7.4.0917 (Bio-Rad).

Identified positive clones were expanded and underwent further quality checks.

To quantify the absolute number of *mCherry* and *COLIA2* transcripts, RNA from mCherry-high and mCherry-low cell populations were collected from cerebral organoids. 10ng of total RNA was reverse-transcribed in a 10 μ L reaction using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). The resulting cDNA was diluted to either 1:5 (*mCherry*) or 1:1500 (*COLIA2*) before amplification. The ddPCR reaction was performed in a 20 μ L volume containing 10 μ L of 2X QX200 ddPCR EvaGreen Supermix (Bio-Rad), 5 μ L of diluted cDNA, and 1 μ L each of 4 μ M forward and reverse primers and 3 μ L of nuclease-free water. Droplet generation was completed as above but with the addition of 70 μ L of droplet generation oil for EvaGreen (Bio-Rad). Thermal cycling conditions were as follows: 95°C for 5 minutes, then 44 cycles of 96°C for 30 seconds and 56°C (*mCherry*) or 60°C (*COLIA2*) for 1 minute, and then 4°C for 5 minutes, 90°C for 5 minutes and 4°C for indefinite hold for dye stabilization. EvaGreen fluorescent signal in each droplet were counted and analyzed as described. The copy number of *mCherry* and *COLIA2* transcripts were normalized to the copies per ng of total RNA. All ddPCR analyses were performed at the SRI Genomics Core Facility.

Cerebral organoid generation with dual SMAD inhibitors. We adapted our cerebral organoid differentiation protocol according to a previously described protocol (Qian et al., 2018; Qian et al., 2016). For embryoid body (EB) formation on day 0, hESC colonies were dissociated with Gentle cell dissociation reagent (GCDR, Stem Cell Technologies, 07174) for 7 mins at 37 °C and 12,000 cells in 100 µl STEMdiff kit EB formation media (Stem Cell Technologies, 08570) supplemented with 50 µM Rock inhibitor Y-27632 (Stem Cell Technologies, 72302) were plated in 96 well V-bottom plate (low-binding) (Greiner bio-one, 651970). On day 1 and 3, media was added with 2 µM Dorsomorphine, an inhibitor of BMP type I receptors (ALK2, ALK3, ALK6), (Stem Cell Technologies, 72102) and 2 µM A83-01, an inhibitor of TGFβ type I receptors (ALK4, ALK5, ALK7) (Stem Cell Technologies, 72022) for 5 days *in vitro* to induce embryoid body (EB) formation. On day 5, single EBs that reached ~ 400- 600 um diameter were selected for neural induction and were transferred to individual wells of 24-well ultra-low attachment plate (Corning, 3473) and media was replaced with STEMdiff kit Induction media consisting 1 µM SB431542, an ALK4, ALK5, ALK7 inhibitor (Stem Cell Technologies, 72234) and 1 µM CHIR99021, a GSK3β inhibitor that activates Wnt signaling and limits apoptosis (Delepine et al., 2021; Qian et al., 2016) (Stem Cell Technologies, cat. no. 72054) and cultured 4 more days. On day 9, EBs of ~500-800 um diameter that had translucent edges, a sign of neuroepithelial induction were placed onto a single dimple on embedding sheet (Stem Cell Technologies, cat. no. 08579) and 15 µl of Matrigel, an undefined ECM preparation that contains collagens, laminins, other ECM molecules and growth factors (Corning, cat. 354277) was added to each cerebral organoid to encapsulate it. The Matrigel droplets were incubated at 37 °C for 30 mins before they were washed into 6-well ultra-low attachment plate (Stem Cell Technologies, 38071) containing STEMdiff kit Expansion media with 1 µM SB431542 and 1 µM CHIR99021. On day 13, individual EBs with clear neuroepithelial cell buds were transferred to each well of a 12-well miniature spinning bioreactor (Qian et al., 2018; Qian et al., 2016) containing STEMdiff kit Maturation media. From day 30, extracellular matrix (ECM) proteins were supplemented in Maturation media by dissolving Matrigel at 1% (v/v) containing human recombinant brain-derived neurotrophic factor (BDNF; PeproTech, AF-450-02). The aggregated cells were referred to as cerebral organoids from this stage onward and were allowed to further develop in maturation media until the experimental endpoints, as described.

Cryosectioning and immunostaining. Cerebral organoids were rinsed with ice-cold phosphate-buffered saline (PBS, without Ca^{2+} and Mg^{2+}) (Wisent, 311-010-CL), fixed in 4% paraformaldehyde in PBS (PFA, Electron Microscopy Sciences, 19208) overnight, and immersed in 20% sucrose (Sigma, 84097)/1X PBS overnight after three washes for 5 mins in PBS. Embryonic rodent brains were dissected from the skull before following the same procedures. Tissues were embedded in optimal cutting temperature compound (OCT), and 10-micron sections were collected with a Leica CM3050 cryostat (Leica Microsystems Canada Inc., Richmond Hill, ON, Canada). Samples were collected on FisherbrandTM SuperfrostTM Plus Microscope Slides (Thermo Fisher Scientific, 12-550-15). Cryosections of fixed cerebral organoids were washed in 0.1% Triton X-100 (Sigma, T8787) in PBS (PBST), then blocked for 1 h at room temperature in 10% horse serum (HS, Wisent, 065-150) in PBST. Primary antibodies were diluted in blocking solution as follows: SOX2 (1:500, Abcam #ab97959), PAX6 (1:500, Biolegend #901301), NEUROG2 (1:500, Invitrogen #PA5-78556), NESTIN (1:500, R&D Systems #MAB1259), DCX (1:500, Abcam #ab6586), COL4 (1:200, Abcam #ab6586), LAM (1:200, Sigma #L9393), mCherry (1:500, SICGEN #AB0040-200) and TUJ1 (1:500, BioLegend #802001). After one hour blocking at room temperature, slides were incubated with primary antibodies at 4°C overnight. The next day, slides were washed 5 times for 5 mins in PBST, followed by incubation with 1:500 dilutions of species-specific secondary antibodies (Invitrogen Molecular Probes) for 1 h at room temperature. Slides were washed five times in PBST and counterstained with 4',6-diamidino-2-phenylindole (DAPI, Invitrogen, D1306) and mounted in Aqua-polymount (Polysciences Inc., 18606-20). All images were taken using a Leica DMi8 Inverted Microscope (Leica Microsystems CMS, 11889113).

Tissue clearing, immunolabeling, and lightsheet fluorescence microscopy. hESC-derived cerebral organoids at 42 and 93 days old were fixed in 4% PFA overnight at 4°C. To preserve the tissue protein architecture, samples were cleared using the SHIELD (Stabilization to Harsh conditions via Intramolecular Epoxide Linkages to prevent Degradation) method (Park *et al.*, 2018). Specifically, organoids were incubated in SHIELD OFF solution at 4°C with shaking for 24hrs. Subsequently, they were incubated in a mixture of SHIELD ON-Buffer and the SHIELD-Epoxy solution (7:1 ratio) at 37°C with shaking for 6 hrs. Lastly, the samples were incubated in SHIELD ON-Buffer at 37°C with shaking overnight. To carry out tissue delipidation, the

samples were passively run down in the Delipidation buffer (Lifecanvas Technologies) for 3 days at room temperature (RT). Samples were washed in 0.1% PBST 3 times over a span of 3 hours following each incubation and PFA fixation. Organoids were incubated in primary antibodies diluted in 0.1% PBST (see above) at RT for 48hrs. Samples were then incubated in conjugated secondary antibodies diluted in 0.1% PBST (Alexa Fluor 488 goat anti-mouse IgG2a, Alexa Fluor 568 donkey anti-rabbit, Alexa Fluor 488 donkey anti-rabbit IgG(H+L); 1:250) at RT for 48hrs. For index matching and to make samples optically transparent they were incubated in the EasyIndex medium (Lifecanvas Technologies, RI= 1.52) at RT overnight. Cerebral organoid images were acquired using the UltraMicroscope Blaze lightsheet fluorescence microscope (Miltenyi Biotech) with a 4X objective. The samples were mounted on a small sample stage using photoactivated adhesive (Bodic CNA) and placed into a custom organic imaging medium in the microscope's chamber (Cargille Immersion Liquid, RI= 1.52). Two channels were acquired with a 488 nm wavelength and 85 mW power, and 639nm with 70 mW, for mCherry and SOX2/TUJ1/LAM or COL4, respectively. A 1.67x magnification post-objective lens was employed generating an in-plane resolution of 1.95 μ m and a step size of 3.55 μ m (scanning protocol parameters: laser sheet thickness = 7.1 μ m, NA=0.050, and laser width = 30% single sided multi-angle excitation).

Bulk RNA-seq. Cerebral organoids were dissociated into a single-cell suspension using Worthington Papain System kit (Worthington, LK003150) according to the manufacturer's instruction. We collected single-cell suspensions from organoids at 45-47 days old and multiple organoids (7-8) were pooled together per sample to obtain a sufficient number of cells. Briefly, prewarmed papain solution with DNase (2.5 ml) was added to the organoids in a 60mm dish. Cerebral organoids were minced with a sterile razor blade into smaller pieces and incubated for 30-45 mins at 37°C on an orbital shaker (70 rpm). The tissue suspension was triturated 8-10 times with P1000 pipette tip to assist the release of single cells. Cell suspension was transferred to a 15 ml centrifuge tube and added with ovomucoid protease inhibitor solution to stop papain activity. The cells were centrifuged at 300g for 7 mins and filtered through 40 μ m strainer to remove remaining cell aggregates. Single cells were resuspended in PBS containing FACS buffer (1 mM EDTA, 0.1% BSA and $\text{Ca}^{2+}/\text{Mg}^{2+}$ free PBS) with DAPI before sending for flow

cytometry and each sample was sorted into mCherry-positive and mCherry-negative groups. Total RNA was extracted from FACS-isolated cells using MagMAX-96 total RNA isolation kit (ThermoFisher, AM1830). The extracted total RNA was quantified by Qubit 3 Fluorometer with Qubit RNA HS Assay kit. The integrity of total RNA (RIN value) was measured by Agilent 2100 Bioanalyzer with RNA 6000 Pico kit.

Targeted transcriptome analysis. The targeted transcriptome sequencing was performed on the Ion S5XL Next Generation Sequencing system with the Ion AmpliSeq Transcriptome Human Gene Expression Assay (ThermoFisher Scientifics Inc). This assay covers 20,802 human RefSeq genes (>95% of UCSC refGene) with a single amplicon designed per gene target. The gDNA in the RNA sample was digested by ezDNase and the cDNA was synthesized from 10ng of total RNA using SuperScript IV VILO Master Mix with ezDNase Enzyme kit (Thermofisher Scientifics). The cDNA libraries were constructed by Ion Ampliseq Library Kit Plus. The targeted areas were amplified by polymerase chain reaction for 12 cycles. The resulting amplicons were treated with FuPa reagent to partially digest primers. Amplicons were ligated to Ion P1 and IonCode barcode adapters and purified using Agencourt AMPure XP reagent (Beckman Coulter). Barcoded libraries were quantified using the Ion Library TaqMan Quantitation Kit (ThermoFisher Scientific Inc) and diluted to a final concentration of 80pM. The sequencing template preparation was done using Ion Chef with Ion 540 Chef Kits. Sequencing was performed for 500 flows on an Ion S5XL Sequencer with Ion 540 chip.

Next generation sequencing data analysis. The Ion Torrent platform-specific pipeline software, Torrent Suite version 5.18.1 (Thermo Fisher Scientific Inc) was used to separate barcoded reads and to filter and remove polyclonal and low-quality reads. Ion Torrent platform-specific plugin, ampliseqRNA (v5.18.0.0) was used for the alignment of the raw sequencing reads and quantitation of normalized gene expression level (reads-per-million:RPM). DESeq2 was used to do differential expression. Principal component analysis, generation of scatter plot and volcano plot, hierarchical clustering and pathway analysis were performed by Transcriptome Analysis Console (TAC) 4.0 software using the CHP files.

Single cell RNA-seq data analysis in human fetal cortices and cerebral organoids. A Seurat v.3.2.3 R package was used to analyze the scRNA seq data in human cerebral organoids available from GSE137877 (Sivitilli et al., 2020). The scRNA seq data of human fetal cortices was available from GSE104276 (Zhong et al., 2018). Cells with more than 4000 genes and less than 500 genes were filtered out as doublets or low-quality cells, respectively. SCTransform function was used for data transformation followed by clustering with RunPCA, FindNeighbours and FindClusters, using the first 30 principal components. RunUMAP function was used to generate 2D projections of cell clustering, and cell type annotation was done using the described genes (Sivitilli et al., 2020). Proneural negative, *NEUROG2* or *NEUROG1* single and double positive cells were identified with an expression threshold greater than 0. Monocle3 R package was used for a pseudotime analysis and differentially expressed genes (DEGs) were reported using differential Gene Test function with an adjusted p value less than 0.001.

P19 cell transfection. P19 cells were transfected using Lipofectamine 3000 with pCIG2, pCIG2-Neurog2 and pCIG2-Neurog2SA9TA1 DNA. The cell growth media was changed to fresh media 24 hrs post transfection and the cells were harvested 48 hrs post transfection.

RNA isolation, cDNA preparation and qPCR. Total RNA extraction was performed from P19 cells using Qiagen RNA isolation mini kit (#74104), followed by cDNA preparation using RT2 First strand reverse transcription kit (#330401). qPCR using specific primers targeting known and predicted targets of *Neurog2* was performed using RT2 SYBR green qPCR Kit (#330513). The Qiagen primers used were *Dll3* (#PPM25734G-200), *Neurod1* (#PPM05527D-200), *Col1a1* (#PPM03845F-200), *Col3a1* (#PPM04784B-200), *Sox10* (#PPM04723A-200) and *Pdgfra* (#PPM03640D-200). qPCR analysis was performed using $2^{-\Delta\Delta CT}$ method by normalizing to the CT value of pCIG2 control. This experiment was repeated 3 independent times (N=3 biological replicates) with 3 technical replicates per sample in each experiment. The trend was the same in all experiments. A single representative result is shown.

Imaging, quantification, and statistics. Imaging was performed using epifluorescence microscope followed by analysis and quantification using Photoshop. Statistical analysis was performed using GraphPad Prism software.

DATA AVAILABILITY

We mined scRNA seq data in human cerebral organoids available from GSE137877 (Sivitilli et al., 2020) and scRNA seq data of human fetal cortices was available from GSE104276 (Zhong et al., 2018). RNA-seq data from this study is presented in Supplemental Table 2.

ACKNOWLEDGEMENTS

The authors acknowledge the support of the Sunnybrook Research Institute (SRI) Genomics Core Facility, the SRI Histology Core Facility (Petia Stefanova), and the Sunnybrook Centre for Cytometry and Scanning Microscopy (Kevin Conway). We also acknowledge The Imaging Facility at The Hospital for Sick Children for assistance with lightsheet microscopy.

COMPETING INTERESTS

No competing interests are declared.

FUNDING

This work was supported by operating grants from the Canadian Institutes of Health Research (CIHR) to CS (PJT – 162108) and to CS and JN (PJT-183715). It was also supported by a Canada-Israel Health Research Initiative, jointly funded by CIHR, the Israel Science Foundation, the International Development Research Centre, Canada and the Azrieli Foundation (IDRC 108875) to CS and OR. We acknowledge and are grateful for philanthropic support from the Sunnybrook Hospital Foundation (Community Contributors: Martha Billes, Catherine Rogers, Brian Prendergast). CS holds the Dixon Family Chair in Ophthalmology Research. AMO was supported by a *Canada Graduate Scholarship – Master’s NSERC Studentship*, AM was supported by a *Canada Graduate Scholarship –Master’s Canadian Institutes of Health Research*

(CGS-M/CIHR) and SH by a Cumming School of Medicine, Ontario Graduate Scholarship (OGS), University of Toronto Vision Science Research Program, Peterborough K.M. HUNTER Charitable Foundation and Margaret and Howard GAMBLE Research Grant.

DATA AVAILABILITY

All relevant data can be found within the article and its supplementary information.

REFERENCES

Ali, F., Hindley, C., McDowell, G., Deibler, R., Jones, A., Kirschner, M., Guillemot, F. and Philpott, A. (2011). Cell cycle-regulated multi-site phosphorylation of Neurogenin 2 coordinates cell cycling with differentiation during neurogenesis. *Development* **138**, 4267-4277.

Ali, F. R., Marcos, D., Chernukhin, I., Woods, L. M., Parkinson, L. M., Wylie, L. A., Papkovskaia, T. D., Davies, J. D., Carroll, J. S. and Philpott, A. (2020). Dephosphorylation of the Proneural Transcription Factor ASCL1 Re-Engages a Latent Post-Mitotic Differentiation Program in Neuroblastoma. *Mol Cancer Res* **18**, 1759-1766.

Amin, N. D., Kelley, K. W., Hao, J., Miura, Y., Narazaki, G., Li, T., McQueen, P., Kulkarni, S., Pavlov, S. and Paşa, S. P. (2023). Generating human neural diversity with a multiplexed morphogen screen in organoids. *bioRxiv*.

Amin, S. and Borrell, V. (2020). The Extracellular Matrix in the Evolution of Cortical Development and Folding. *Front Cell Dev Biol* **8**, 604448.

Arai, Y., Pulvers, J. N., Haffner, C., Schilling, B., Nusslein, I., Calegari, F. and Huttner, W. B. (2011). Neural stem and progenitor cells shorten S-phase on commitment to neuron production. *Nat Commun* **2**, 154.

Aydin, B., Kakumanu, A., Rossillo, M., Moreno-Estelles, M., Garippler, G., Ringstad, N., Flames, N., Mahony, S. and Mazzoni, E. O. (2019). Proneural factors Ascl1 and Neurog2 contribute to neuronal subtype identities by establishing distinct chromatin landscapes. *Nat Neurosci* **22**, 897-908.

Azzarelli, R., Hardwick, L. J. and Philpott, A. (2015). Emergence of neuronal diversity from patterning of telencephalic progenitors. *Wiley interdisciplinary reviews. Developmental biology* **4**, 197-214.

Azzarelli, R., McNally, A., Dell'Amico, C., Onorati, M., Simons, B. and Philpott, A. (2022). ASCL1 phosphorylation and ID2 upregulation are roadblocks to glioblastoma stem cell differentiation. *Sci Rep* **12**, 2341.

Betizeau, M., Cortay, V., Patti, D., Pfister, S., Gautier, E., Bellemin-Menard, A., Afanassieff, M., Huisoud, C., Douglas, R. J., Kennedy, H., et al. (2013). Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442-457.

Cernilogar, F. M., Hasenöder, S., Wang, Z., Scheibner, K., Burtscher, I., Sterr, M., Smialowski, P., Groh, S., Evenroed, I. M., Gilfillan, G. D., et al. (2019). Pre-marked chromatin and transcription factor co-binding shape the pioneering activity of Foxa2. *Nucleic Acids Res* **47**, 9069-9086.

Chan, C. J., Heisenberg, C. P. and Hiiragi, T. (2017). Coordination of Morphogenesis and Cell-Fate Specification in Development. *Curr Biol* **27**, R1024-R1035.

Chanda, S., Ang, C. E., Davila, J., Pak, C., Mall, M., Lee, Q. Y., Ahlenius, H., Jung, S. W., Sudhof, T. C. and Wernig, M. (2014). Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem Cell Reports* **3**, 282-296.

Chaudhari, P., Ye, Z. and Jang, Y. Y. (2014). Roles of reactive oxygen species in the fate of stem cells. *Antioxid Redox Signal* **20**, 1881-1890.

Cheroni, C., Trattaro, S., Caporale, N., Lopez-Tobon, A., Tenderini, E., Sebastiani, S., Troglio, F., Gabriele, M., Bressan, R. B., Pollard, S. M., et al. (2022). Benchmarking brain organoid recapitulation of fetal corticogenesis. *Transl Psychiatry* **12**, 520.

Delepine, C., Pham, V. A., Tsang, H. W. S. and Sur, M. (2021). GSK3ss inhibitor CHIR 99021 modulates cerebral organoid development through dose-dependent regulation of apoptosis, proliferation, differentiation and migration. *PLoS One* **16**, e0251173.

Dixit, R., Wilkinson, G., Cancino, G. I., Shaker, T., Adnani, L., Li, S., Dennis, D., Kurrasch, D., Chan, J. A., Olson, E. C., et al. (2014). Neurog1 and Neurog2 control two waves of neuronal differentiation in the piriform cortex. *J Neurosci* **34**, 539-553.

Fernandez Garcia, M., Moore, C. D., Schulz, K. N., Alberto, O., Donague, G., Harrison, M. M., Zhu, H. and Zaret, K. S. (2019). Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol Cell* **75**, 921-932.e926.

Fernandez, V. and Borrell, V. (2023). Developmental mechanisms of gyration. *Curr Opin Neurobiol* **80**, 102711.

Fietz, S. A., Kelava, I., Vogt, J., Wilsch-Brauninger, M., Stenzel, D., Fish, J. L., Corbeil, D., Riehn, A., Distler, W., Nitsch, R., et al. (2010). OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat Neurosci* **13**, 690-699.

Fietz, S. A., Lachmann, R., Brandl, H., Kircher, M., Samusik, N., Schroder, R., Lakshmanaperumal, N., Henry, I., Vogt, J., Riehn, A., et al. (2012). Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc Natl Acad Sci U S A* **109**, 11836-11841.

Florio, M., Albert, M., Taverna, E., Namba, T., Brandl, H., Lewitus, E., Haffner, C., Sykes, A., Wong, F. K., Peters, J., et al. (2015). Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **347**, 1465-1470.

Fode, C., Ma, Q., Casarosa, S., Ang, S. L., Anderson, D. J. and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes Dev* **14**, 67-80.

Gattazzo, F., Urciuolo, A. and Bonaldo, P. (2014). Extracellular matrix: a dynamic microenvironment for stem cell niche. *Biochimica et biophysica acta* **1840**, 2506-2519.

Ge, W., He, F., Kim, K. J., Blanchi, B., Coskun, V., Nguyen, L., Wu, X., Zhao, J., Heng, J. I., Martinowich, K., et al. (2006). Coupling of cell migration with neurogenesis by proneural bHLH factors. *Proc Natl Acad Sci U S A* **103**, 1319-1324.

Ghazale, H., Park, E., Vasan, L., Mester, J., Saleh, F., Trevisiol, A., Zinky, D., Chinchalongporn, V., Liu, M., Fleming, T., et al. (2022). Ascl1 phospho-site mutations enhance neuronal conversion of adult cortical astrocytes in vivo. *Front Neurosci* **16**, 917071.

Girsakis, K. M., Stergachis, A. B., DeGennaro, E. M., Doan, R. N., Qian, X., Johnson, M. B., Wang, P. P., Sejourne, G. M., Nagy, M. A., Pollina, E. A., et al. (2021). Rewiring of human neurodevelopmental gene regulatory programs by human accelerated regions. *Neuron* **109**, 3239-3251 e3237.

Guilak, F., Cohen, D. M., Estes, B. T., Gimble, J. M., Liedtke, W. and Chen, C. S. (2009). Control of stem cell fate by physical interactions with the extracellular matrix. *Cell stem cell* **5**, 17-26.

Guillemot, F. and Hassan, B. A. (2017). Beyond proneural: emerging functions and regulations of proneural proteins. *Curr Opin Neurobiol* **42**, 93-101.

Guven, A., Kalebic, N., Long, K. R., Florio, M., Vaid, S., Brandl, H., Stenzel, D. and Huttner, W. B. (2020). Extracellular matrix-inducing Sox9 promotes both basal progenitor proliferation and gliogenesis in developing neocortex. *eLife* **9**.

Han, S., Dennis, D. J., Balakrishnan, A., Dixit, R., Britz, O., Zinyk, D., Touahri, Y., Olander, T., Brand, M., Guillemot, F., et al. (2018). A non-canonical role for the proneural gene Neurog1 as a negative regulator of neocortical neurogenesis. *Development* **145**.

Han, S., Okawa, S., Wilkinson, G. A., Ghazale, H., Adnani, L., Dixit, R., Tavares, L., Faisal, I., Brooks, M. J., Cortay, V., et al. (2021). Proneural genes define ground-state rules to regulate neurogenic patterning and cortical folding. *Neuron* **109**, 2847-2863 e2811.

Hand, R., Bortone, D., Mattar, P., Nguyen, L., Heng, J. I., Guerrier, S., Boutt, E., Peters, E., Barnes, A. P., Parras, C., et al. (2005). Phosphorylation of Neurogenin2 specifies the migration properties and the dendritic morphology of pyramidal neurons in the neocortex. *Neuron* **48**, 45-62.

Hansen, D. V., Lui, J. H., Parker, P. R. and Kriegstein, A. R. (2010). Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554-561.

Hansen, J. L., Loell, K. J. and Cohen, B. A. (2022). A test of the pioneer factor hypothesis using ectopic liver gene activation. *eLife* **11**, e73358.

Hardwick, L. J., Ali, F. R., Azzarelli, R. and Philpott, A. (2015). Cell cycle regulation of proliferation versus differentiation in the central nervous system. *Cell Tissue Res* **359**, 187-200.

Haubensak, W., Attardo, A., Denk, W. and Huttner, W. B. (2004). Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proc Natl Acad Sci U S A* **101**, 3196-3201.

Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc Natl Acad Sci U S A* **109 Suppl 1**, 10661-10668.

Hindley, C., Ali, F., McDowell, G., Cheng, K., Jones, A., Guillemot, F. and Philpott, A. (2012). Post-translational modification of Ngn2 differentially affects transcription of distinct targets to regulate the balance between progenitor maintenance and differentiation. *Development* **139**, 1718-1723.

Hodge, R. D., Bakken, T. E., Miller, J. A., Smith, K. A., Barkan, E. R., Graybuck, L. T., Close, J. L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-68.

Hu, J., Deng, L., Wang, X. and Xu, X. M. (2009). Effects of extracellular matrix molecules on the growth properties of oligodendrocyte progenitor cells in vitro. *J Neurosci Res* **87**, 2854-2862.

Karzbrun, E. and Reiner, O. (2019). Brain Organoids-A Bottom-Up Approach for Studying Human Neurodevelopment. *Bioengineering* **6**.

Kessaris, N., Fogarty, M., Iannarelli, P., Grist, M., Wegner, M. and Richardson, W. D. (2006). Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. *Nat Neurosci* **9**, 173-179.

Kovach, C., Dixit, R., Li, S., Mattar, P., Wilkinson, G., Elsen, G. E., Kurrasch, D. M., Hevner, R. F. and Schuurmans, C. (2013). Neurog2 simultaneously activates and represses alternative gene expression programs in the developing neocortex. *Cereb Cortex* **23**, 1884-1900.

LaMonica, B. E., Lui, J. H., Hansen, D. V. and Kriegstein, A. R. (2013). Mitotic spindle orientation predicts outer radial glial cell generation in human neocortex. *Nat Commun* **4**, 1665.

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

Lancaster, M. A., Renner, M., Martin, C. A., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P. and Knoblich, J. A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373-379.

Li, S., Mattar, P., Dixit, R., Lawn, S. O., Wilkinson, G., Kinch, C., Eisenstat, D., Kurrasch, D. M., Chan, J. A. and Schuurmans, C. (2014). RAS/ERK signaling controls proneural genetic programs in cortical development and gliomagenesis. *J Neurosci* **34**, 2169-2190.

Li, S., Mattar, P., Zinyk, D., Singh, K., Chaturvedi, C. P., Kovach, C., Dixit, R., Kurrasch, D. M., Ma, Y. C., Chan, J. A., et al. (2012). GSK3 temporally regulates neurogenin 2 proneural activity in the neocortex. *J Neurosci* **32**, 7791-7805.

Long, K. R., Newland, B., Florio, M., Kalebic, N., Langen, B., Kolterer, A., Wimberger, P. and Huttner, W. B. (2018). Extracellular Matrix Components HAPLN1, Lumican, and Collagen I Cause Hyaluronic Acid-Dependent Folding of the Developing Human Neocortex. *Neuron* **99**, 702-719.e706.

Martinez-Cerdeno, V., Cunningham, C. L., Camacho, J., Antczak, J. L., Prakash, A. N., Cziep, M. E., Walker, A. I. and Noctor, S. C. (2012). Comparative analysis of the subventricular zone in rat, ferret and macaque: evidence for an outer subventricular zone in rodents. *PLoS One* **7**, e30178.

Martinez-Martinez, M. A., De Juan Romero, C., Fernandez, V., Cardenas, A., Gotz, M. and Borrell, V. (2016). A restricted period for formation of outer subventricular zone defined by Cdh1 and Trnp1 levels. *Nat Commun* **7**, 11812.

Mattar, P., Langevin, L. M., Markham, K., Klenin, N., Shivji, S., Zinyk, D. and Schuurmans, C. (2008). Basic helix-loop-helix transcription factors cooperate to specify a cortical projection neuron identity. *Molecular and cellular biology* **28**, 1456-1469.

Minderjahn, J., Schmidt, A., Fuchs, A., Schill, R., Raithel, J., Babina, M., Schmidl, C., Gebhard, C., Schmidhofer, S., Mendes, K., et al. (2020). Mechanisms governing the pioneering and redistribution capabilities of the non-classical pioneer PU.1. *Nature communications* **11**, 402.

Miyata, T., Kawaguchi, A., Saito, K., Kawano, M., Muto, T. and Ogawa, M. (2004). Asymmetric production of surface-dividing and non-surface-dividing cortical progenitor cells. *Development* **131**, 3133-3145.

Moffat, A., Oproescu, A.-M., Okawa, S., Han, S., Vasan, L., Ghazale, H., Dennis, D. J., Zinyk, D., Guillemot, F., Sol, A. d., et al. (2023). Proneural genes form a combinatorial code to diversify neocortical neural progenitor cells. *bioRxiv*, 2023.2007.2029.551096.

Moffat, A. and Schuurmans, C. (2023). The Control of Cortical Folding: Multiple Mechanisms, Multiple Models. *Neuroscientist*, 10738584231190839.

Mohyeldin, A., Garzon-Muvdi, T. and Quinones-Hinojosa, A. (2010). Oxygen in stem cell biology: a critical component of the stem cell niche. *Cell stem cell* **7**, 150-161.

Nelson, C. M. (2009). Geometric control of tissue morphogenesis. *Biochimica et biophysica acta* **1793**, 903-910.

Noctor, S. C., Martinez-Cerdeno, V., Ivic, L. and Kriegstein, A. R. (2004). Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci* **7**, 136-144.

Oproescu, A. M., Han, S. and Schuurmans, C. (2021). New Insights Into the Intricacies of Proneural Gene Regulation in the Embryonic and Adult Cerebral Cortex. *Front Mol Neurosci* **14**, 642016.

Park, N. I., Guilhamon, P., Desai, K., McAdam, R. F., Langille, E., O'Connor, M., Lan, X., Whetstone, H., Coutinho, F. J., Vanner, R. J., et al. (2017). ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell stem cell* **21**, 209-224.e207.

Park, Y. G., Sohn, C. H., Chen, R., McCue, M., Yun, D. H., Drummond, G. T., Ku, T., Evans, N. B., Oak, H. C., Trieu, W., et al. (2018). Protection of tissue physicochemical properties using polyfunctional crosslinkers. *Nat Biotechnol*.

Parras, C. M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D. J. and Guillemot, F. (2002). Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. *Genes Dev* **16**, 324-338.

Păun, O., Tan, Y. X., Patel, H., Strohbuecker, S., Ghanate, A., Cobolli-Gigli, C., Llorian Sopena, M., Gerontogianni, L., Goldstone, R., Ang, S. L., et al. (2023). Pioneer factor ASCL1 cooperates with the mSWI/SNF complex at distal regulatory elements to regulate human neural differentiation. *Genes Dev* **37**, 218-242.

Pebworth, M. P., Ross, J., Andrews, M., Bhaduri, A. and Kriegstein, A. R. (2021). Human intermediate progenitor diversity during cortical development. *Proc Natl Acad Sci U S A* **118**.

Pollen, A. A., Nowakowski, T. J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C. R., Shuga, J., Liu, S. J., Oldham, M. C., Diaz, A., et al. (2015). Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55-67.

Qian, X., Jacob, F., Song, M. M., Nguyen, H. N., Song, H. and Ming, G. L. (2018). Generation of human brain region-specific organoids using a miniaturized spinning bioreactor. *Nat Protoc* **13**, 565-580.

Qian, X., Nguyen, H. N., Song, M. M., Hadiono, C., Ogden, S. C., Hammack, C., Yao, B., Hamersky, G. R., Jacob, F., Zhong, C., et al. (2016). Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell* **165**, 1238-1254.

Quan, X. J., Yuan, L., Tiberi, L., Claeys, A., De Geest, N., Yan, J., van der Kant, R., Xie, W. R., Klisch, T. J., Shymkowitz, J., et al. (2016). Post-translational Control of the Temporal Dynamics of Transcription Factor Activity Regulates Neurogenesis. *Cell* **164**, 460-475.

Raposo, A. A., Vasconcelos, F. F., Drechsel, D., Marie, C., Johnston, C., Dolle, D., Bithell, A., Gillotin, S., van den Berg, D. L., Ettwiller, L., et al. (2015). Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis. *Cell Rep*.

Rash, B. G., Duque, A., Morozov, Y. M., Arellano, J. I., Micali, N. and Rakic, P. (2019). Gliogenesis in the outer subventricular zone promotes enlargement and gyration of the primate cerebrum. *Proc Natl Acad Sci U S A* **116**, 7089-7094.

Rayon, T., Stamatakis, D., Perez-Carrasco, R., Garcia-Perez, L., Barrington, C., Melchionda, M., Exelby, K., Lazaro, J., Tybulewicz, V. L. J., Fisher, E. M. C., et al. (2020). Species-specific pace of development is associated with differences in protein stability. *Science* **369**.

Revah, O., Gore, F., Kelley, K. W., Andersen, J., Sakai, N., Chen, X., Li, M.-Y., Birey, F., Yang, X., Saw, N. L., et al. (2022). Maturation and circuit integration of transplanted human cortical organoids. *Nature* **610**, 319-326.

Rozario, T. and DeSimone, D. W. (2010). The extracellular matrix in development and morphogenesis: a dynamic view. *Dev Biol* **341**, 126-140.

Schuurmans, C., Armant, O., Nieto, M., Stenman, J. M., Britz, O., Klenin, N., Brown, C., Langevin, L. M., Seibt, J., Tang, H., et al. (2004). Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways. *The EMBO Journal* **23**, 2892-2902.

Shitamukai, A., Konno, D. and Matsuzaki, F. (2011). Oblique radial glial divisions in the developing mouse neocortex induce self-renewing progenitors outside the germinal zone that resemble primate outer subventricular zone progenitors. *J Neurosci* **31**, 3683-3695.

Sivitilli, A. A., Gosio, J. T., Ghoshal, B., Evstratova, A., Trcka, D., Ghiasi, P., Hernandez, J. J., Beaulieu, J. M., Wrana, J. L. and Attisano, L. (2020). Robust production of uniform human cerebral organoids from pluripotent stem cells. *Life Sci Alliance* **3**.

Smith, D. K., Yang, J., Liu, M. L. and Zhang, C. L. (2016). Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. *Stem Cell Reports* **7**, 955-969.

Soufi, A., Garcia, M. F., Jaroszewicz, A., Osman, N., Pellegrini, M. and Zaret, K. S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568.

Sun, Y., Nadal-Vicens, M., Misono, S., Lin, M. Z., Zubiaga, A., Hua, X., Fan, G. and Greenberg, M. E. (2001). Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell* **104**, 365-376.

Vaid, S. and Huttner, W. B. (2022). Progenitor-Based Cell Biological Aspects of Neocortex Development and Evolution. *Front Cell Dev Biol* **10**, 892922.

van den Ameele, J., Tiberi, L., Vanderhaeghen, P. and Espuny-Camacho, I. (2014). Thinking out of the dish: what to learn about cortical development using pluripotent stem cells. *Trends in neurosciences* **37**, 334-342.

Wallace, J. L. and Pollen, A. A. (2023). Human neuronal maturation comes of age: cellular mechanisms and species differences. *Nature reviews. Neuroscience*.

Wang, X., Tsai, J. W., LaMonica, B. and Kriegstein, A. R. (2011). A new subtype of progenitor cell in the mouse embryonic neocortex. *Nat Neurosci* **14**, 555-561.

Wapinski, O. L., Lee, Q. Y., Chen, A. C., Li, R., Corces, M. R., Ang, C. E., Treutlein, B., Xiang, C., Baubet, V., Suchy, F. P., et al. (2017). Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell Rep* **20**, 3236-3247.

Wapinski, O. L., Vierbuchen, T., Qu, K., Lee, Q. Y., Chanda, S., Fuentes, D. R., Giresi, P. G., Ng, Y. H., Marro, S., Neff, N. F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* **155**, 621-635.

Wilson, M. N., Thunemann, M., Liu, X., Lu, Y., Puppo, F., Adams, J. W., Kim, J.-H., Ramezani, M., Pizzo, D. P., Djurovic, S., et al. (2022). Multimodal monitoring of human cortical organoids implanted in mice reveal functional connection with visual cortex. *Nature communications* **13**, 7945.

Yang, J.-H., Hayano, M., Griffin, P. T., Amorim, J. A., Bonkowski, M. S., Apostolides, J. K., Salfati, E. L., Blanchette, M., Munding, E. M., Bhakta, M., et al. (2023). Loss of epigenetic information as a cause of mammalian aging. *Cell* **186**, 305-326.e327.

Zecevic, N., Chen, Y. and Filipovic, R. (2005). Contributions of cortical subventricular zone to the development of the human cerebral cortex. *The Journal of comparative neurology* **491**, 109-122.

Zhong, S., Zhang, S., Fan, X., Wu, Q., Yan, L., Dong, J., Zhang, H., Li, L., Sun, L., Pan, N., et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524-528.

FIGURE LEGENDS

Figure 1. Transcriptomic assessment of NEUROG1 and NEUROG2 expression in gestational week 8-26 human fetal cortices and 3-mo old cerebral organoids. (A)
Distribution of NEUROG1/NEUROG2 single and double⁺ cells in scRNA-seq datasets from human fetal cortices between gestational week 8 to 26. (B-E) UMAP of scRNA-seq data from 3-month-old cerebral organoids (B), showing proportions of cortical cell types expressing NEUROG1 and/or NEUROG2 (C). Key genes made for cluster assignments are mapped onto the UMAP, including NEUROG1 and NEUROG2 (D) and cell type-specific markers (E). aRG,

apical radial glia; bRG, basal radial glia; ch-plexus, choroid plexus; DL, deep layer; Gly, high glycolytic cells; IN, interneuron; IPC, intermediate progenitor cell; UL, upper layer.

Figure 2. Pseudotime trajectory analysis of *NEUROG1* and *NEUROG2* expressing cerebral organoid cells. (A-F) Monocle3 lineage trajectory analysis of *NEUROG2⁺*, *NEUROG1⁺* and double⁺ cells in 3-mo cerebral organoid scRNA-seq data (Sivitilli et al., 2020) (A), showing a pseudotime trajectory (B), and the distributions of proneural-expressing cells in states 1, 2 and 3 (C). *NEUROG1* and/or *NEUROG2* expressing cells were mapped onto the pseudotime trajectory (D), along with other key markers of cortical cell types (F). The distributions of proneural-expressing cells in each cell state expressing cell type-specific markers were plotted (E).

Figure 3. Generation of hESC-derived cerebral organoids using dual Smad inhibition.

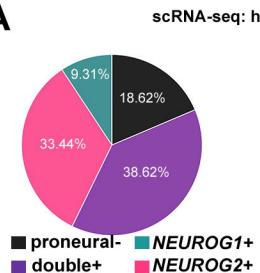
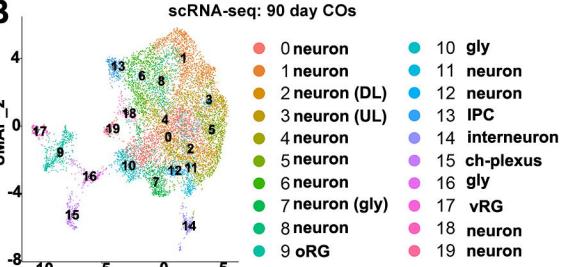
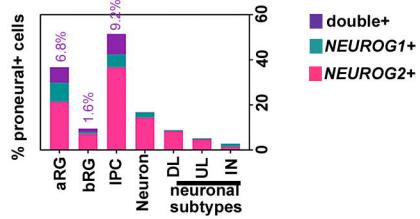
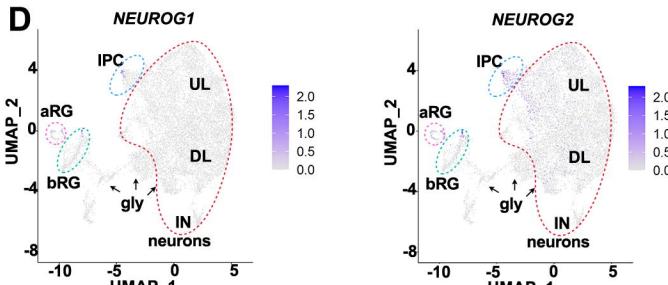
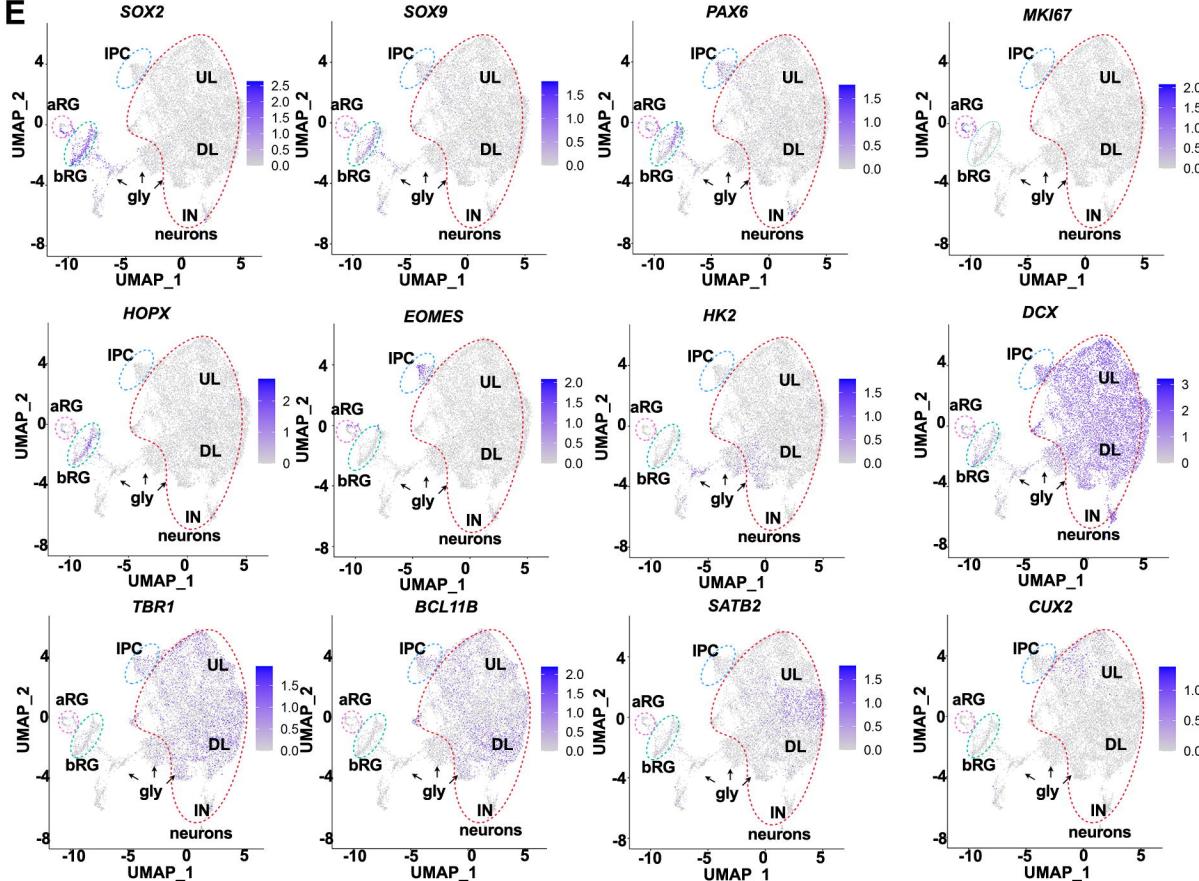
(A,B) Experimental protocol used for cerebral organoid generation (A) and photomicrographs of derivative cells and cerebral organoids at various stages of cerebral organoids formation (B). (C) 3D light-sheet imaging of two-day 42 cerebral organoids that were cleared and immunolabeled for SOX2 (red) and TUJ1 (green). (D) Immunostaining of cryosections of day 30-35 cerebral organoids immunostained for SOX2/TUJ1, PAX6, NESTIN/TUJ1 and DCX. Scale bars in B, 100 μ m; in C, 200 μ m; in D, 400 μ m in low magnification images and 100 μ m in high magnification images.

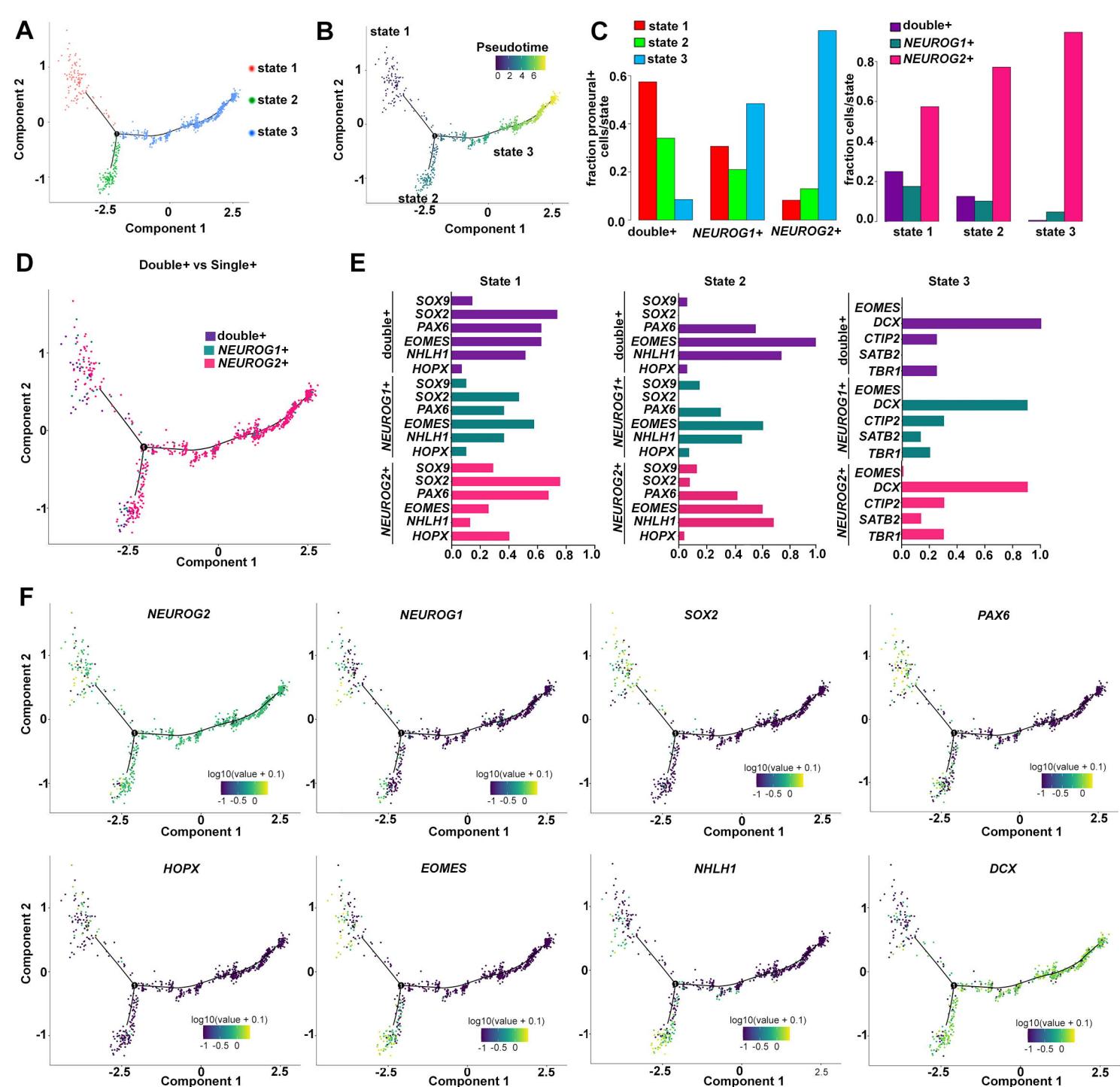
Figure 4. Analyses of mCherry-high and mCherry-low cells in *NEUROG2-mCherry* KI hESC-derived cerebral organoids. (A-F) Immunostaining of *NEUROG2-mCherry* KI hESC-derived cerebral organoids collected at day 18 (A-C) and day 45 (D-F) and analysed for the expression of mCherry with *NEUROG2*, SOX2 and PAX6. (G) qPCR to validate FACS-enrichment of mCherry (mCh)-high and mCh-low cells collected from two *NEUROG2* mCherry KI hESC cell lines (#105 and #117) from different sets of cerebral organoids (-1, -2). (H-K) PCA analysis of bulk RNA-seq data collected from mCherry-high and mCherry-low cells (H). Volcano plot showing an enrichment of more genes in mCherry-high versus mCherry-low cells (I). Differentially expressed genes (DEGs) in mCherry-high cells were identified compared to

mCherry-low cells (I). Biological Process Gene Ontology (GO) terms enriched in DEGs that were upregulated (J) or downregulated (K) in mCherry-high cells. (L-M) Bar graphs showing the log2FC values of the differentially expressed genes, OPC lineage markers (L) and ECM proteins and remodelers (M). DEGs, differentially expressed genes; ECM, extracellular matrix; mCh, mCherry; OPC, oligodendrocyte precursor cells.

Figure 5. Expression of ECM markers in *NEUROG2-mCherry* KI hESC derived cerebral organoid. (A,B) Expression of mCherry with ECM markers Collagen IV (COL 4) and Laminin (LAM) in day 18 (A) and day 45 (B) *NEUROG2-mCherry* derived cerebral organoids. (C,D) 3D light-sheet imaging of two-clarified day 119 cerebral organoids immunolabeled for mCherry (red) and COL4 (green, C) or LAM (green, D). (E) ddPCR to look at the relative expression of COL1A2/GAPDH in mCherry-high versus mCherry-low cells isolated from day 45 cerebral organoids derived from the two *NEUROG2-mCherry* KI cell lines (#117 and #105). (F) Mining a bulk RNA-seq dataset from human fetal cortical zones (Fietz *et al.*, 2012), revealed an enrichment of *NEUROG2* and ECM gene transcripts in the same cell populations.

Figure 6. Characterization of mCherry positive cells in embryonic *Neurog2-mCherry* KI cortices. (A) mCherry expression in E12.5 *Neurog2-mCherry* KI mouse brain. (B) Transcriptomic analysis to compare the log2FC values of cell type specific markers in mCherry-high, CD15⁺ cortical NPCs to mCherry low cells (data from (Han *et al.*, 2021)). Biological Process GO terms enriched in DEGs that were upregulated (C) or downregulated (D) in mCherry-high cells. (E-H) P19 cells transfected with pCIG2, pCIG2-*Neurog2* or pCIG2-*Neurog2-SA9* and analysed for the expression of known target genes (E), OPC genes (F), ECM genes (G) and IPC marker (H). Student t-tests were used to individually compare expression levels. p-values: ns - not significant, <0.05 *, <0.01 **, <0.001 ***. Significance was defined as p-values less than 0.05. DEGs, differentially-expressed genes; ECM, extracellular matrix; IPC, intermediate progenitor cell; mCh, mCherry; OPC, oligodendrocyte precursor cells.

A**B****C****D****E**



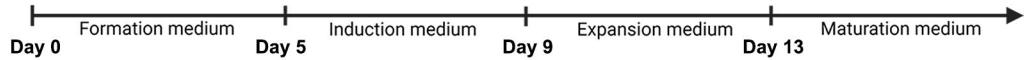
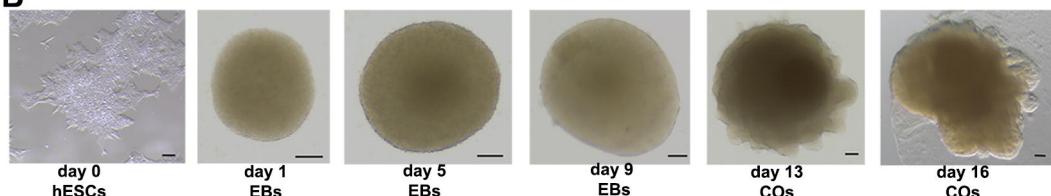
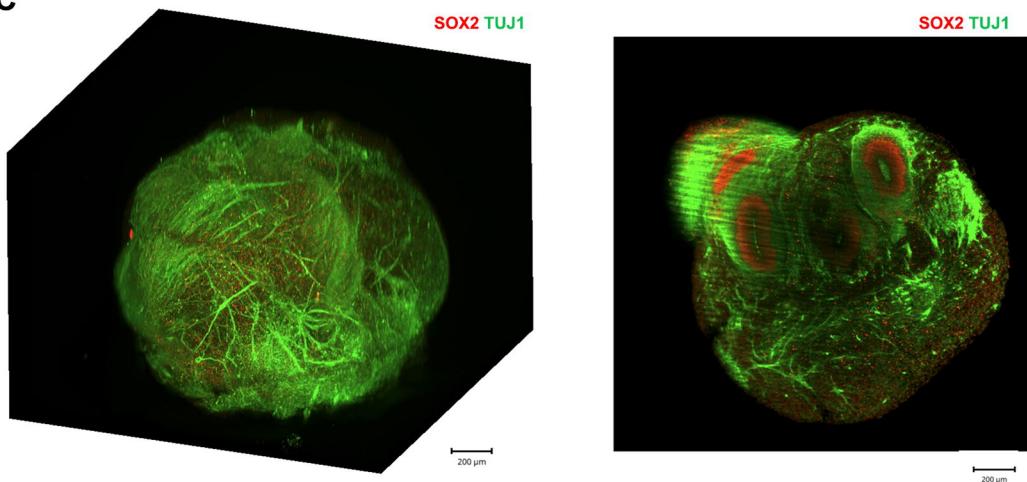
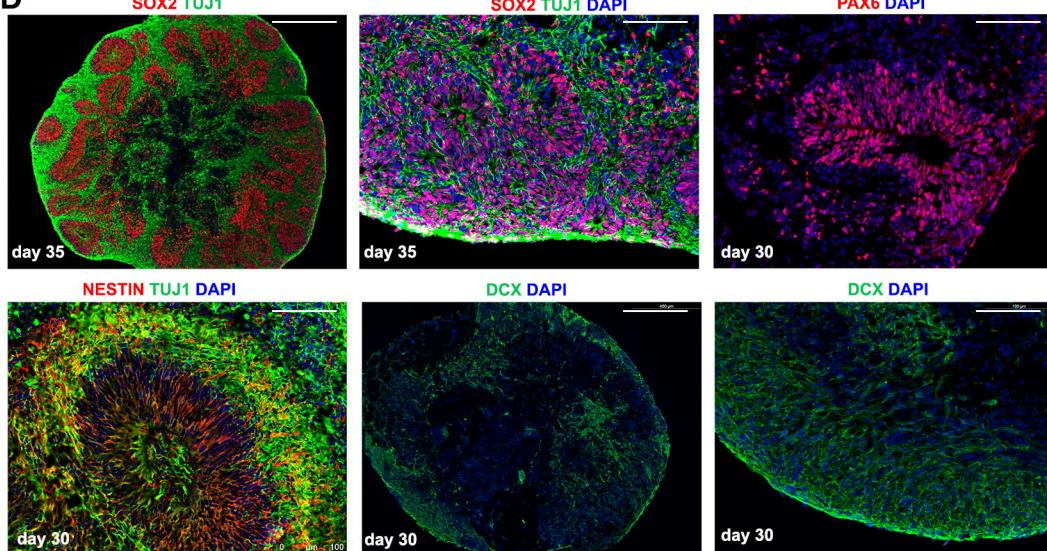
A

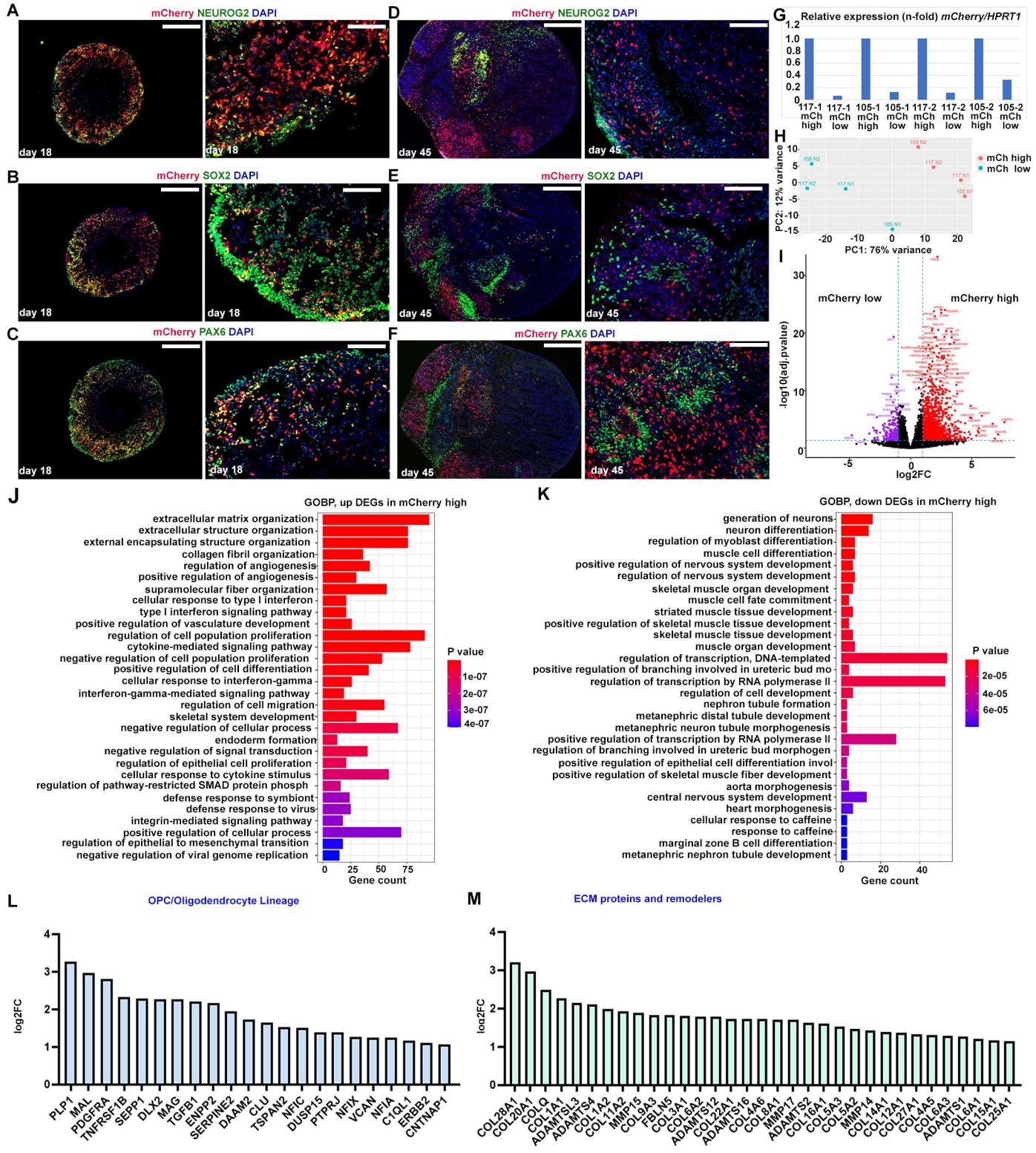
Dorsomorphin, A-83

CHIR-99021, SB-431542

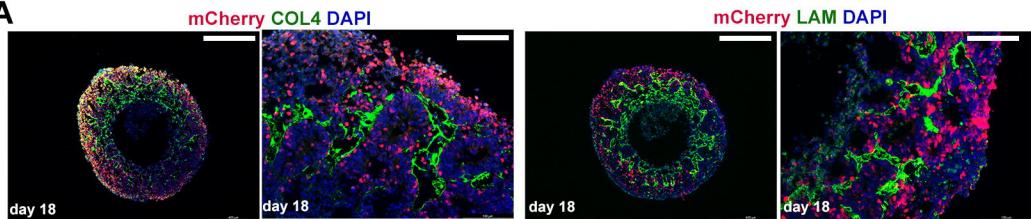
embed EBs in Matrigel

culture in mini-bioreactor

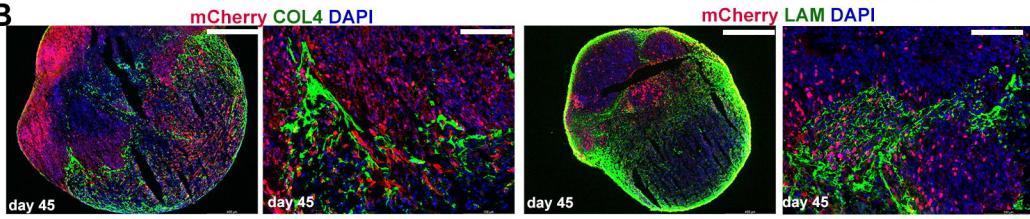
**B****C****D**



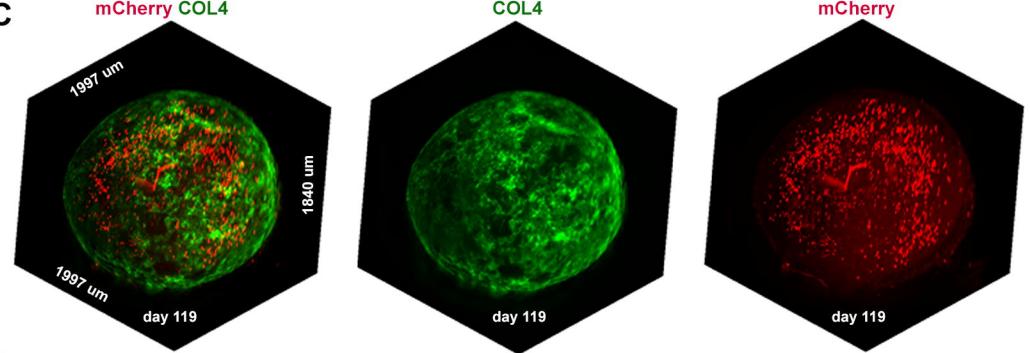
A



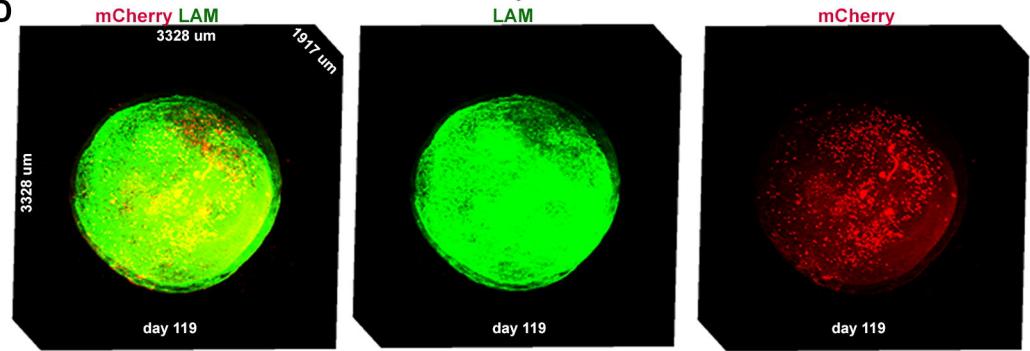
B



C



D



E

