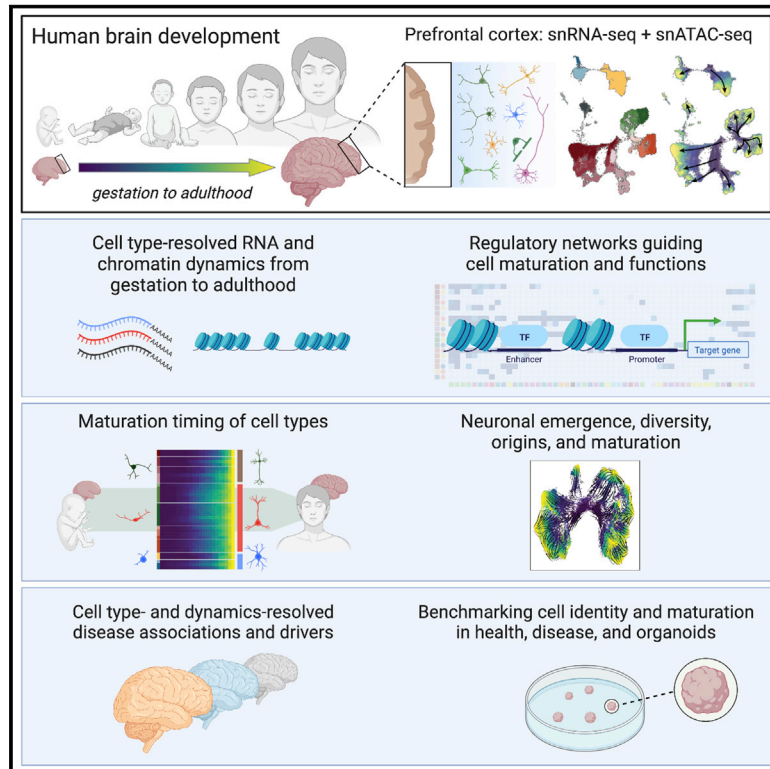# Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution

Graphical abstract

Authors

Charles A. Herring, Rebecca K. Simmons, Saskia Freytag, Daniel Poppe, ..., Irina Voineagu, Luciano Martelotto, Ryan Lister

Correspondence

ryan.lister@uwa.edu.au

In brief

A resource charting the gene expression and chromatin accessibility landscapes of human prefrontal cortex reveals gene expression reconfiguration at the prenatal-to-postnatal transition followed by continuous reconfiguration into adulthood.

Highlights

- Single-nucleus RNA and chromatin dynamics from gestation to adulthood in humans

- Protracted and diverse timing of cell-type development to a mature state

- RNA dynamics and regulators predict cellular and temporal origins of brain diseases

- Atlas integration reveals few postnatal maturity neurons in long-term brain organoids

Resource

Human prefrontal cortex gene regulatory dynamics from gestation to adulthood at single-cell resolution

Charles A. Herring,[1,2,10] Rebecca K. Simmons,[1,2,10] Saskia Freytag,[1,2,10] Daniel Poppe,[1,2,10] Joel J.D. Moffet,[1]
Jahnvi Pflueger,[1,2] Sam Buckberry,[1,2] Dulce B. Vargas-Landin,[1,2] Olivier Clément,[1,2] Enrique Goñi Echeverría,[1]
Gavin J. Sutton,[3] Alba Alvarez-Franco,[4] Rui Hou,[1] Christian Pflueger,[1,2] Kerrie McDonald,[5] Jose M. Polo,[6,7]
Alistair R.R. Forrest,[1] Anna K. Nowak,[8] Irina Voineagu,[3] Luciano Martelotto,[6,9] and Ryan Lister[1,2,11,*]

[1]Harry Perkins Institute of Medical Research, QEII Medical Centre and Centre for Medical Research, The University of Western Australia, Perth, WA 6009, Australia
[2]ARC Centre of Excellence in Plant Energy Biology, School of Molecular Sciences, The University of Western Australia, Perth, WA 6009, Australia
[3]School of Biotechnology and Biomolecular Sciences, Cellular Genomics Futures Institute, and the RNA Institute, University of New South Wales, Sydney, NSW 2052, Australia
[4]Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid 28029, Spain
[5]Lowy Cancer Research Centre, University of New South Wales, Sydney, NSW 2052, Australia
[6]Adelaide Centre for Epigenetics and the South Australian Immunogenomics Cancer Institute, Faculty of Health and Medical Sciences, The University of Adelaide, Adelaide, SA 5000, Australia
[7]Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, VIC 3000, Australia
[8]Medical School, University of Western Australia, Perth, WA 6009, Australia
[9]University of Melbourne Centre for Cancer Research, Victoria Comprehensive Cancer Centre, Melbourne, VIC 3000, Australia
[10]These authors contributed equally
[11]Lead contact
*Correspondence: ryan.lister@uwa.edu.au
https://doi.org/10.1016/j.cell.2022.09.039

SUMMARY

Human brain development is underpinned by cellular and molecular reconfigurations continuing into the third decade of life. To reveal cell dynamics orchestrating neural maturation, we profiled human prefrontal cortex gene expression and chromatin accessibility at single-cell resolution from gestation to adulthood. Integrative analyses define the dynamic trajectories of each cell type, revealing major gene expression reconfiguration at the prenatal-to-postnatal transition in all cell types followed by continuous reconfiguration into adulthood and identifying regulatory networks guiding cellular developmental programs, states, and functions. We uncover links between expression dynamics and developmental milestones, characterize the diverse timing of when cells acquire adult-like states, and identify molecular convergence from distinct developmental origins. We further reveal cellular dynamics and their regulators implicated in neurological disorders. Finally, using this reference, we benchmark cell identities and maturation states in organoid models. Together, this captures the dynamic regulatory landscape of human cortical development.

INTRODUCTION

The human prefrontal cortex (PFC) does not fully mature until well into adulthood (Molnár et al., 2019; Sydnor et al., 2021), in a highly protracted process unique among primates (Miller et al., 2012). Neurological executive functions rely on complex spatial, electrical, and chemical interactions of diverse neural cell types that mature through a coordinated process spanning development from *in utero* to adulthood (Silbereis et al., 2016). Essential processes of synaptogenesis, synaptic pruning, myelination, and plasticity during postnatal development give rise to the functional complexity and capabilities of the mature brain

(Silbereis et al., 2016). However, the molecular dynamics driving the cellular identities and functional changes that emerge during brain development, in particular postnatally, remain largely unknown.

Gene expression changes throughout brain development have been studied using bulk tissue samples (Hernandez et al., 2012; Kang et al., 2011), and recent studies have analyzed transcriptome and chromatin states at single-cell resolution at early fetal and adult stages (Morabito et al., 2021; Nowakowski et al., 2017; Figure S1A; Table S1A). However, there has been no systematic characterization of gene expression and chromatin dynamics throughout the entirety of postnatal human brain
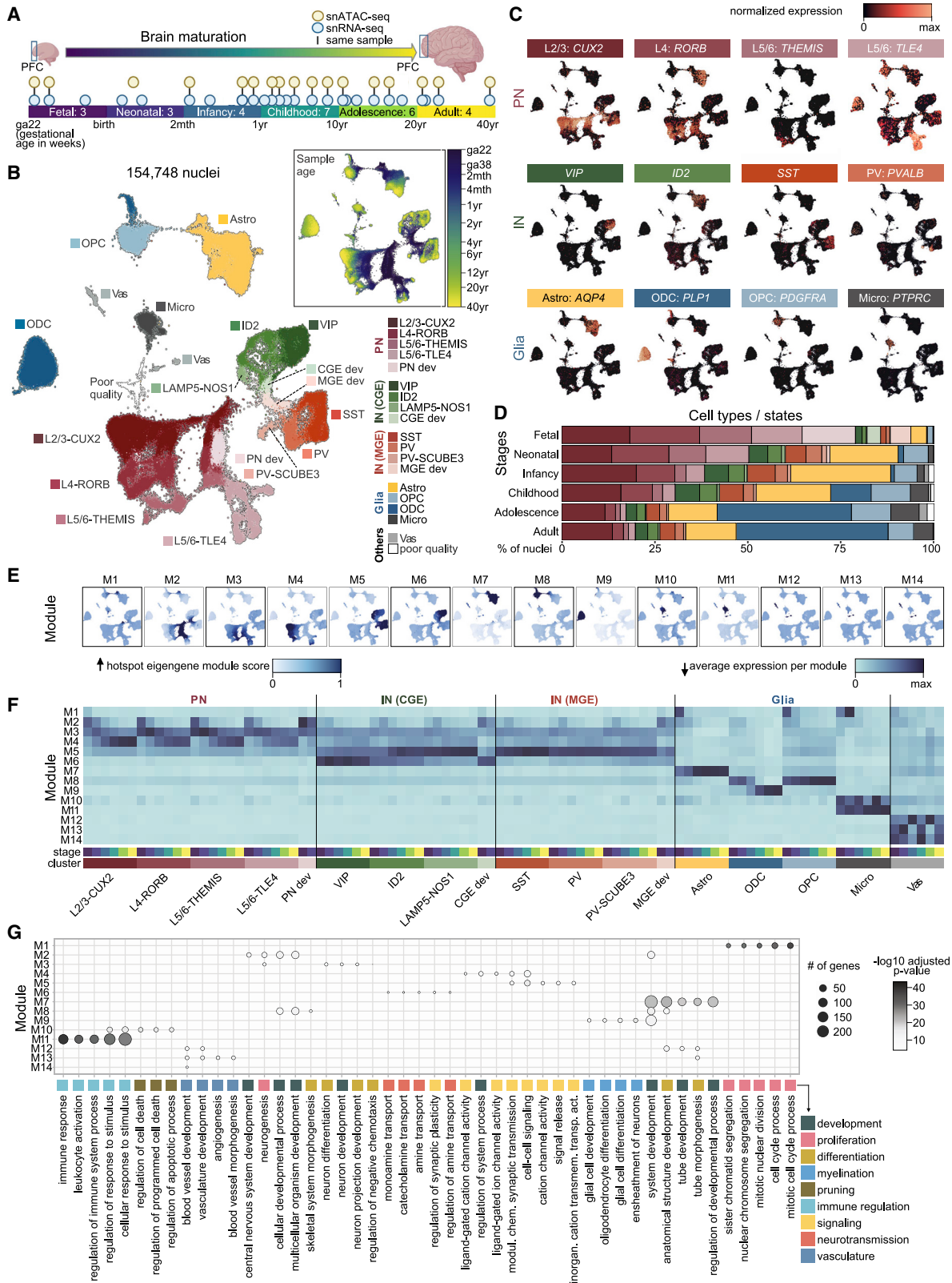
development. This gap spans major milestones, such as language acquisition and development of executive functions. Consequently, we lack a detailed understanding of the timing and nature of gene expression and chromatin dynamics that underpin these developmental processes. Here, we characterize at single-cell resolution the gene expression and chromatin accessibility changes that occur through human PFC development, from gestation to adulthood. This enables characterization of active pathways and their dynamics in diverse cortical cell types throughout development, cell-type-specific maturation timing, and prediction of the controlling *cis*-regulatory logic and associated factors, underpinning molecular dissection of neural cell developmental processes.

RESULTS

A single-nucleus resolution transcriptome reference of human brain development

We performed single-nucleus RNA-seq (snRNA-seq) profiling of 26 postmortem PFC samples from individuals spanning fetal, neonatal, infancy, childhood, adolescence, and adult stages of development (Figure 1A; Table S1B), providing extensive sampling of stages poorly represented by previous single-cell studies (neonatal, infant, child, and adolescent). We generated 154,748 single nuclei transcriptome profiles after quality filtering (Figures S1C–S1F; Table S1C) and assembled an integrated developmental reference with distinct chronological ordering of all major neural lineages (Figure 1B). We identified 86 distinct clusters across all developmental stages (Figure S1G) and annotated them by separating nuclei into either excitatory principal neurons (PNs), inhibitory interneurons (INs, from medial ganglionic eminence [MGE] or caudal ganglionic eminence [CGE]), or glia, followed by further separation of the major clusters using layer and subtype-specific marker genes (Figures 1C and S2A; Tables S2A–S2C). This developmental map revealed systematic expression changes through postnatal development (Figure 1B) and temporal changes in cell-type abundance (Figures 1D, S1H, and S1I), such as expansion of oligodendrocytes (ODCs) from ODC precursor cells (OPCs) beginning in infancy and progressing to adulthood (Figure 1D; Perlman et al., 2020; van Tilborg et al., 2018).

We further identified modules representing the organized coexpression of multiple genes in various subsets of nuclei using hotspot (DeTomaso and Yosef, 2021; Figure 1E; Table S2D). Each module reflects cell-type-specific transcriptional changes align-

ing with known cell functions and processes that occur during brain development (Figures 1E–1G). These coexpression modules represent the broad cell states and dynamics of the PFC, capturing major cell types and their functions, developmental processes, and cell-type-specific expression changes throughout brain development.

Cell-type-specific gene expression dynamics during PFC development

To characterize gene expression changes throughout development, we organized the 86 distinct subclusters into cell-type trajectories. Each trajectory contains all the nuclei corresponding to a particular cell type spanning all profiled stages. Subcluster assignment into trajectories was based on the known cell type and developmental markers, as well as cell neighborhoods in the uniform manifold approximation and projection (UMAP) representation over age progression (Figures S2A and S2B; Table S2B; STAR Methods). We applied a hierarchical annotation strategy by first separating glia, PNs, and INs, followed by separation into the major known cell types, and then into sub-trajectories. Our marker-based cell and cluster assignment to trajectories was also supported by RNA velocity vector fields (Bergen et al., 2020) and cellular trajectory reconstruction analysis using gene counts and expression (CytoTRACE) (Gulati et al., 2020) scores (Figure S3A; STAR Methods).

In total, 45 distinct sub-trajectories were defined that encompass development into terminal cell types (Figures S2B and S2C; Table S2B; STAR Methods). To facilitate interpretability, we aggregated related cell trajectories into 15 major cell-type trajectories for further analyses (4 PN, 6 IN, 4 glial, and 1 vasculature; Figures 2A and S2C; Table S2C). We performed differential expression testing over development for each major trajectory while adjusting for covariates (Law et al., 2016; Figure S4A; STAR Methods) and validated the result robustness (Figure S3C; STAR Methods). In total, we identified 14,984 unique development-associated differentially expressed genes (devDEGs, false discovery rate [FDR] <5%) within at least one major trajectory (Figures 2B and S3B; Table S3; see online browser). We identified robust markers of neural cell maturity and immaturity that show widespread, large, and extended expression dynamics over development (Figure S3E): *KCNH1* and *ARHGAP26*, which encode a voltage-gated potassium channel and a Rho GTPase activating protein, increase progressively to become highly expressed throughout neurons, and in some glia; and *BHLHE22*, encoding a transcription factor (TF) involved in embryonic neural

---

Figure 1. A single-nucleus resolution transcriptome reference of human PFC development
(A) Schematic of developmental stages, sample ages, brain region, profiling methods, and libraries generated in this study. Ages used to define the developmental stages: fetal (ga22 to <ga38; ga, gestational age, in weeks), neonatal (≥ga38 to <2 months), infancy (≥2 months to <1 year), childhood (≥1 to <10 years), adolescence (≥10 to <20 years), and adult (≥20 years), as guided by Kang et al. (2011).
(B) UMAP plot depicting 154,748 nuclei in 18 major clusters. Inset shows UMAP with each nucleus colored by the arcsinh transformation of its age. Astro, astrocytes; Micro, microglia; "dev" indicates early developing cells.
(C) Expression of marker genes for major cell-type annotations.
(D) Fraction of each cell type/state in each developmental stage.
(E) Hotspot eigengene module scores overlaid onto UMAP embeddings.
(F) Mean scaled and normalized expression for all genes in the module for each cluster and developmental stage. x axis depicts major clusters and developmental stages from fetal to adult in each cluster.
(G) Top 5 enriched biological process GO terms for each module.
See Tables S1 and S2.
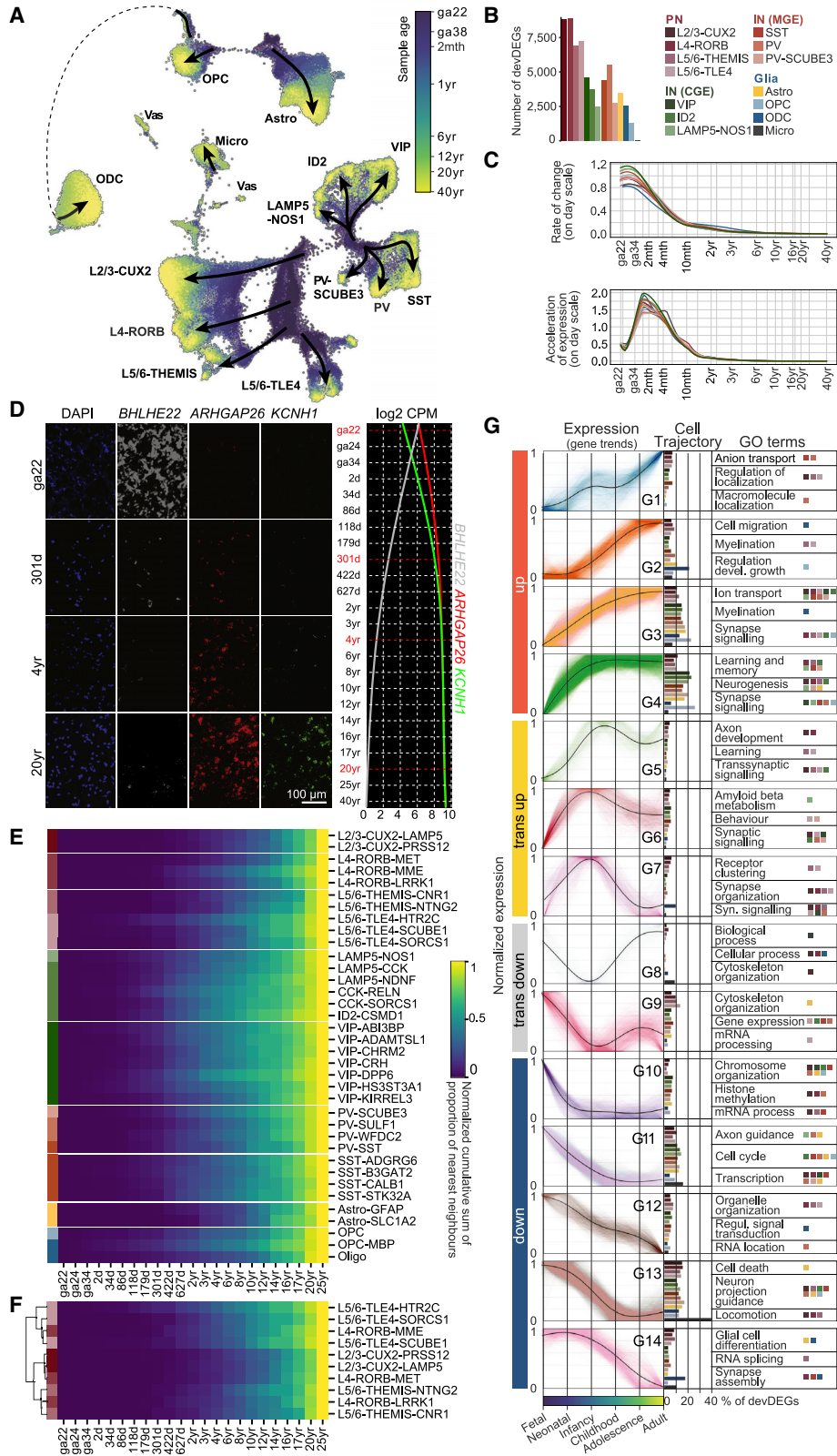
cell fate determination and differentiation (Dennis et al., 2019), is abundant in fetal PNs and decreases rapidly after birth. These genes more effectively represent dynamic markers of neural cell maturity and immaturity than the commonly used markers *SOX2* and *RBFOX3* (encoding NeuN) (Figure S3E). devDEG dynamics were consistent with developmental changes in expression detected for these markers by RNAscope *in situ* hybridization (Figures 2D, S3D, and S3E), and for larger sets of dynamic genes identified in a previous developmental RNA-seq study of bulk PFC tissue (Li et al., 2018; Figure S3F).

To examine the absolute rate of change in expression over development, we fit trends to the devDEGs using a generalized additive model (GAM) (Trapnell et al., 2014). This revealed rapid rate changes from the prenatal period to 1 year of age, before tapering off into childhood (Figure 2C), similar across all cell types and consistent with bulk PFC tissue analyses (Colantuoni et al., 2011; Kang et al., 2011). Acceleration of expression peaks at birth for all cell types (Figure 2C), indicating that an increase in rates of change in expression coincides with this major event. Regularization of spacing between ages, to account for disparate timescales (days versus years), revealed that ODCs exhibit an increase in the rate of change peaking in childhood, which tapers off in early adolescence (Figure S3G), corresponding well to established timelines for cortical synaptogenesis (Huttenlocher and Dabholkar, 1997).

The human brain continues to mature into the 3rd decade of life; however, the timing of when each cell type attains an adult-like transcriptional state is unknown. Therefore, for each cell-type developmental sub-trajectory, we determined the timing of the cumulative acquisition of an adult-like gene expression state by approximating the maturity of each nucleus based on the portion of its nearest neighbors that were from adult samples (Figure 2E; STAR Methods). This revealed surprising diversity in maturation timing across cell types. PN maturation mirrors the inside-out developmental pattern of the cortex, with the development starting with deep-layer neurons and ending with upper-layer neurons (Agirman et al., 2017). PN maturation timing clustered into three timeline categories, with waves of maturation starting in infancy, childhood, and adolescence (Figure 2F). Deep-layer L5/6-TLE4 subtypes and L4-RORB-MME show acquisition of adult-like features starting in infancy, potentially reflective of their roles in forming early sensory-related circuits with thalamic axons (Kostović and Judas, 2010). L5/6-THEMIS subtypes and L4-RORB-LRRK1 display an increase in maturation in childhood coinciding with a transition in neural networks from local intra-cortical connections to more distal inter-regional

connections (Supekar et al., 2009). Lastly, layer L2/3 and L4-RORB-MMT begin maturation in adolescence during a period of increased cortical myelination (Arain et al., 2013; Guillery, 2005). Unlike PNs, INs reach adult-like identities in an order that does not reflect their birthdate, where MGE neurogenesis precedes CGE (Miyoshi et al., 2010). Despite broad similarities among sub-trajectories within MGE and CGE cell types, differing maturation timescales are evident. For example, the parvalbumin-expressing (PV)-somatostatin-expressing (SST) sub-trajectory acquires adult-like features later than its MGE peers, and vasoactive intestinal polypeptide-expressing (VIP)-DPP6 neurons show accelerated maturation within the CGE sub-trajectories.

To further investigate the diversity and timing of expression dynamics over development, we clustered the devDEGs into 14 "gene trends" (Figure 2G; STAR Methods). We also categorized these into four general developmental gene trends: up, transiently up, transiently down, and down (Figure 2G; Table S3). These gene trends reveal complex and diverse expression dynamics from gestation to adulthood, associated with different cell types and processes critical for brain development and function (Figure 2G; Table S4).

Although all cell types showed similar rates of change (Figure 2C), we posited that divergent dynamics could exist within individual processes relevant to distinct neural cell functionalities. Thus, we calculated the weighted average of expression dynamics throughout development (the "eigentrend") of devDEGs enriched in manually curated gene ontologies for each major cell trajectory. This revealed diverse cell-type-specific developmental dynamics of genes involved in key neural function and development processes (Figure 3A).

The "neuron migration" eigentrend peaks during the third trimester before declining at birth (Figure 3A), concordant with the known timeline of IN migration (Arshad et al., 2016). ID2 INs show expression through to adulthood, longer than other cell types (Figure 3A), suggesting ID2 INs may undergo extended migration beyond infancy (p value $\leq 1.8 \times 10^{-11}$; Table S2E). This could represent a longer migration of these IN subtypes than was previously shown, where a subset of indeterminate INs migrate until at least 5 months postnatally (Paredes et al., 2016).

Two distinct programs of "neuron projection development" are evident for PNs (Figure 3A). L4-RORB and L5/6-TLE4 PN eigentrends peak in fetal development, coinciding with known timing of thalamic axons reaching the cortical plate of the neocortex and forming early sensory-related circuits (Kostović

Figure 2. Cell-type-specific gene expression and maturation dynamics across PFC development

(A) UMAP plot with each nucleus colored by the arcsinh transformation of its donor age. Arrows indicate the major trajectories.

(B) Number of devDEGs detected in each major trajectory.

(C) Absolute mean rates of change over development of devDEGs (top) and acceleration of expression of devDEGs (bottom), for each major trajectory.

(D) RNAscope in PFC tissue for four ages spanning maturation (left), detecting: mature cells marked by *KCNH1* (green) and *ARHGAP26* (red), and immature cells marked by *BHLHE22* (gray). DAPI counterstain (blue, nuclei). Pseudo-bulked snRNA-seq $\log_2$ counts per million (CPM) expression trend across ages for the markers (right).

(E) Heatmap of scaled (0–1) cumulative sum of proportion of nearest neighbors that are nuclei from the adult developmental stage across sample ages.

(F) As per (E), but for PN sub-trajectories only, with rows hierarchically clustered.

(G) Scaled (0–1) expression of the devDEGs clustered into 14 "gene trends." Faint lines show the individual devDEGs, and bold line is the average across all devDEGs in the trend. Gene trends are grouped into four general classes. Bar plots indicate the percentage of devDEGs from each major trajectory in each gene trend. Selected enriched GO terms and major trajectories they are enriched in are shown.
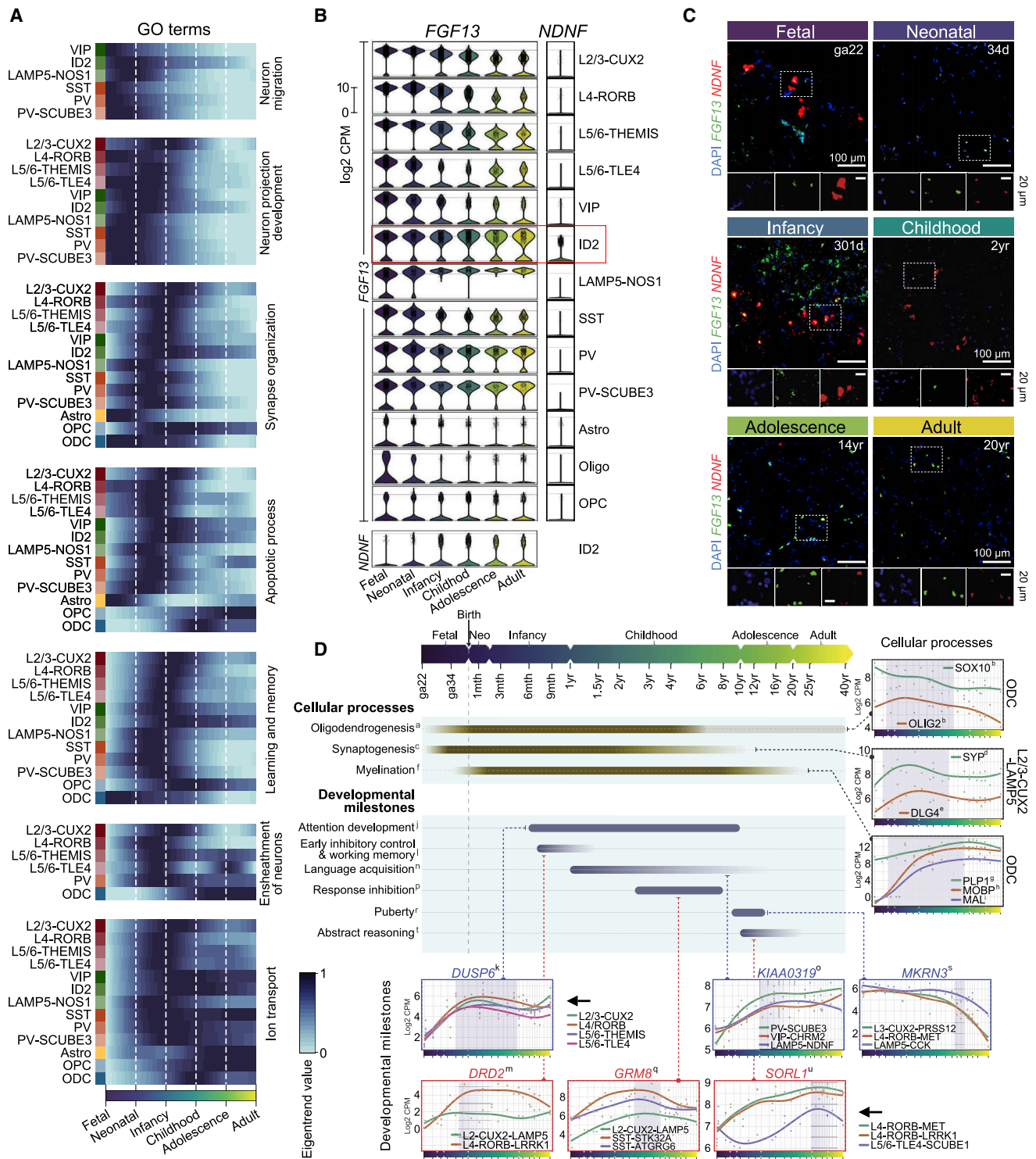
Figure 3. Neurodevelopmental processes and milestones with distinct cell-type-specific dynamics

(A) Eigentrend values for devDEGs in selected GO terms over development in different cell-type trajectories.

(B) Log$_2$ CPM expression of *FGF13* in major trajectories for each developmental stage, and log$_2$ CPM expression of *NDNF* in each major trajectory (right) and for each developmental stage for ID2 neurons (bottom).

(legend continued on next page)

and Judas, 2010; Kostović et al., 2019). In contrast, L2/3-CUX2 and L5/6-THEMIS show maximum eigentrend values after birth, reflecting the increasing concentration of local projections at this time (Kostović et al., 2019). The "synapse organization" eigentrends also suggest earlier activity in L4 and L5/6-TLE4 than in L2/3 and L5/6-THEMIS neurons (Figure 3A; Table S2E; p value $\leq 8.6 \times 10^{-7}$).

As ID2 INs show expression through later postnatal stages for "neuron migration," "neuron projection development," and "synapse organization," which share 35 genes. Of these, we identified *FGF13*, a factor critical for neuronal migration, polarization, and development of axons (Wu et al., 2012). *FGF13* is highly expressed in the early postnatal (P7) cerebral cortex in mice, significantly lower by P14, and markedly reduced by early adulthood (P60) (Wu et al., 2012). Surprisingly, we found that *FGF13* is highly expressed in a subset of human cortical neurons through to adulthood, in particular in several IN subtypes, including ID2, LAMP5-NOS1, PV-SCUBE3, and PV (Figure 3B). To validate this protracted *FGF13* expression, we focused on *NDNF+* ID2s, as these ID2 neurons can be unambiguously detected based on *NDNF* expression throughout development (Figure 3B). RNAscope confirmed that *NDNF+* ID2 neurons expressing *FGF13* are detectable in the human PFC throughout development, including into adulthood (Figure 3C), potentially reflecting ongoing plasticity in a subset of neurons.

Myelination is vital for brain function, allowing fast transmission of electrical signals along axons (Duncan et al., 2021; Williamson and Lyons, 2018). Analysis of the "ensheathment of neurons" eigentrend identified expression linked to myelination processes through development to adulthood, and more evident in deep-layer PNs as observed in neuroimaging studies (Whitaker et al., 2016) (Figure 3A). ODCs and multiple neuronal subtypes (L5/6-TLE, L5/6-THEMIS, and PV) exhibit increased ensheathment of neurons values spanning childhood to adulthood, suggesting coordinated gene expression in these cell types underpinning the increased myelination during adolescence (Figure 3A). Because myelination requires adaptive cell-to-cell communication between neurons and ODCs, we investigated cell-cell communication networks through development (Hou et al., 2020). We found that the expression of ligands for ODCs and respective receptors in PNs begins in the neonatal stage and increases through to adulthood, indicating reinforcement of communication through changes in regulatory networks (Figure S3H). Through development, we observe differential expression of genes encoding interacting cell adhesion receptors and their cognate ligands in ODCs and PNs, such as *PTPRD* and *NTRK3* (Figure S3I).

Eigentrends for "ion transport" show peak values during early infancy for most PNs, coinciding with synaptogenesis (Figure 3A; Huttenlocher and Dabholkar, 1997). Later onset during childhood is seen in L2/3 PNs and extends into adolescence in some CGE-derived INs (VIP, ID2), but not in MGE-derived INs (SST, PV), which resemble PNs from deep cortical layers. The difference between MGE- and CGE-derived INs could stem from known differences in developmental timing of migration patterns or their role in PN regulation (Molnár et al., 2019). Further analysis of ion transporter expression dynamics revealed enrichment of general gene trends for different types of ion transporters, with potassium ion transporters following an up gene trend for a majority of neurons starting from the neonatal period (Figure S3J).

We next explored the dynamics of genes linked to cellular and functional milestones of human brain development (Figure 3D). Inspecting genes linked to developmental and cognitive abilities, we identified neuronal subtypes in which gene upregulation temporally coincides with emergence of these capabilities (Figure 3D). For example, expression of *GRM8*, linked to response inhibition (Bauer and Covault, 2020), peaks in subtypes including some SSTs and L2/3s, coincident with the development of this capability from ~2 to 8 years.

Together, these analyses identify diverse expression dynamics associated with known cellular and developmental processes in all cell types, spanning gestation to adulthood.

Interneuron emergence, dynamics, and specialization through development

MGE- and CGE-derived INs exhibit diversity in transcriptional programs and function (Lim et al., 2018; Zecevic et al., 2011). Our profiling enabled detailed exploration of the origins, emergence, dynamics, and specialization of IN subtypes throughout human development. We identified 31 distinct IN subtypes/states (Figure 4A), with clusters expressing known IN subtype-specific markers (Figures 4B–4D). MGE- and CGE-derived subtypes exhibited distinct transcriptional profiles in adults, with MGE subtypes showing greater heterogeneity between each other than adult CGE subtypes (Figure 4E). We also identified rare GABAergic IN populations, including an *NPY*- and *NOS1*-expressing SST population (SST-NPY), representing only ~1% of INs, which we validated by RNAscope (Figures 4A and 4C). In mice, *NPY*-, *NOS1*-, and *SST*-expressing neurons project to other neocortical regions and are active during sleep (Lim et al., 2018). We also identified and independently validated another rare cell population that shares characteristics of both PV and SST subtypes (PV-SST) (Figures 4A and 4D). In mice, multipolar bursting interneurons with characteristics of PV and SST subtypes have been observed that exhibit narrow action potentials, such as fast spiking PV, but have a preference for dendritic postsynaptic targets, such as SST (Blatow et al., 2003; Thomson and Lamy, 2007).

To visualize the development of sub-trajectories, we overlaid the UMAP with scVelo velocity vector fields (Bergen et al., 2020). Although vectors largely agree with age progression, some disjunct patterns are visible (Figure S4B). To more explicitly compare transition vectors and development, we introduce UMAP of maturation (UMAT), which limits UMAP neighbor selection to cells from adjacent developmental stages (Figures 4F and S4C). UMAT vector field overlays agree with age progression

---

(C) RNAscope of *NDNF* (red) and *FGF13* (green) in PFC sections of representative ages of each developmental stage. DAPI counterstain (blue, nuclei).

(D) Expression dynamics of genes linked to milestones of human brain development and cognition. Horizontal timelines and graph shading indicate approximate age of milestone onset/development. Individual plots show expression of genes ($\log_2$ CPM; dot: individual sample; line: GAM fit) linked to specific milestones in select cell types, supported as per Table S1G.
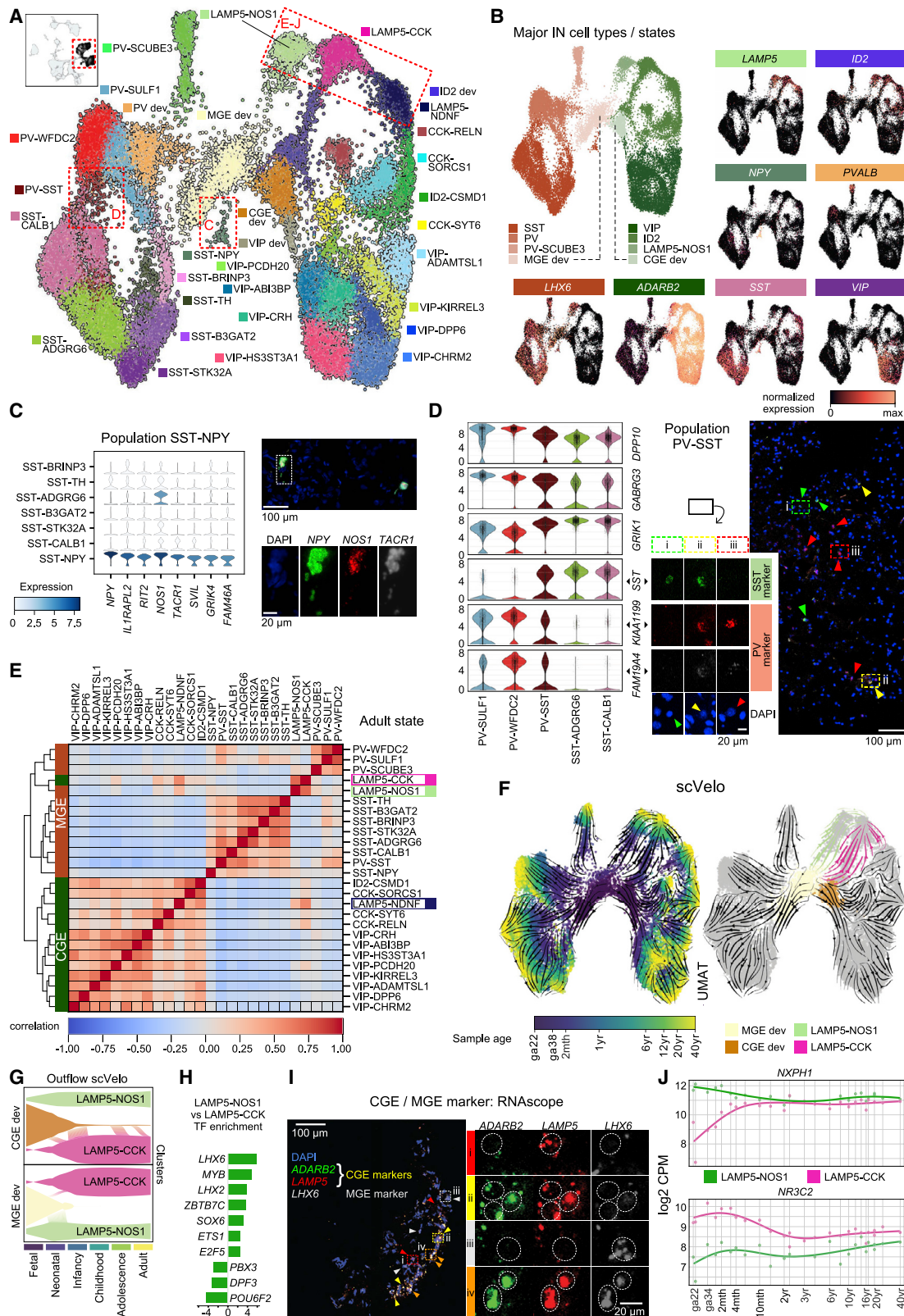
and allow visualization of developmental trajectories of the diverse IN subtypes.

Neurogliaform cells (NGFCs), marked by *LAMP5* expression, are found throughout the cortex and likely refine cortical circuits; however, the developmental origins of human cortical NGFCs are not well understood (Overstreet-Wadiche and McBain, 2015). UMAT-based sub-trajectory analysis suggests that an MGE-derived LAMP5+ population (NOS1) exists and differs in its origin from the other two LAMP5+ populations (CGE-derived CCK and NDNF; Figures 4A and 4F). Moreover, transition probabilities calculated from CellRank scVelo and CytoTRACE kernels indicate a high transition probability from developing MGE cells (MGE dev) to LAMP5-NOS1, and from developing CGE cells (CGE dev) to LAMP5-CCK, supporting distinct developmental origins and trajectories (Figures 4G, S4D, and S4E). Differential expression analysis of TFs between NOS1 and CCK sub-trajectory populations revealed TFs known to be specifically expressed in MGE- versus CGE-derived neurons (Figure 4H; Table S2F; Long et al., 2009), supporting their independent developmental origins. Finally, RNAscope demonstrated the existence of LAMP5+ cells co-staining for classical MGE (*LHX6*) and CGE (*ADARB2*) markers (Figures 4B and 4I).

Vector field projections also suggest convergence of the NOS1 and CCK subtypes at the adult stages (Figures 4F and S4D), potentially due to shared environmental cues and function. Molecular convergence of these subtypes is illustrated in several ways. Initially disparate expression of the MGE-derived cortical IN marker *NXPH1* and mineralocorticoid receptor *NR3C2* becomes similar postnatally (Figure 4J). Furthermore, given the ion transport eigentrend differences between MGE- and CGE-derived INs (Figure 3A), we further analyzed these in the NGFC sub-trajectories. In NOS1 INs, the ion transport eigentrend begins earlier than in other subtypes (Figure S4F), likely due to earlier migration of MGE- than CGE-derived INs (Lim et al., 2018). However, from infancy onward the eigentrends are similar, further supporting subtype convergence. Finally, based on gene expression, adult NOS1 and CCK populations are more similar to MGE-derived IN populations, whereas NDNF are more similar to other CGE-derived popula-tions (Figure 4E). MGE-derived NGFCs are implicated in memory consolidation during non-REM sleep (Valero et al., 2021). Based on devDEG similarities, including ion channel expression and sub-trajectory analysis, we speculated that these subtypes could share a similar function. DEGs between the adult cells of each subtype (NOS1 and CCK) compared with all other IN subtypes are enriched in memory-related processes in the adult NOS1 and CCK subtypes (Table S2G), suggesting that the CGE-derived population may have a similar function in memory consolidation. In contrast, the NDNF subtype likely corresponds to canopy cells located in L1 based on marker expression (Figures S4H and S4I; Hodge et al., 2019; Schuman et al., 2019) and shows localization at the edge of cortical sections (Figure S4G).

Together, this provides a detailed molecular characterization of human postnatal cortical IN development, revealing subtype features and possible functional convergence of NGFC subtypes with different origins.

A single-nucleus resolution chromatin accessibility map of human PFC development

To investigate the *cis*-regulatory elements (CREs) that govern the expression dynamics underpinning PFC development, we performed single-nucleus chromatin accessibility profiling (snA-TAC-seq) of 17 postmortem PFC samples (16 matching the snRNA-seq samples) from the same developmental stages (Figure 1A; Tables S1B and S1E). We generated chromatin accessibility profiles for 87,339 individual nuclei after quality filtering (Table S1E), which clustered into 12 major cell states over development using label transfer from the snRNA-seq (Figures 5A and S5A–S5F). UMAP representation revealed chronological ordering within neuronal cell types, representing progression of regulatory programs across development (Figure 5A inset), and age-related cell-type proportion changes (Figures 5B and S5E).

As described for the snRNA-seq, we grouped cell types/states into trajectories spanning development and identified 252,606 open chromatin regions (OCRs) across all trajectories, revealing extensive cell-type- and developmental-stage-specific OCRs.

Figure 4. Interneuron emergence, dynamics, and specialization through development

(A) UMAP plot of 27,561 IN nuclei annotated with 31 cell subtype labels, based on gene expression. Inset shows the location of IN nuclei in UMAP representation of entire dataset. Red boxes indicate cell populations featured in (C)–(J).

(B) UMAP plot annotated by major IN cell states/types. Normalized expression of well-established IN subtype markers projected onto IN UMAP embedding.

(C) Expression (ln CPM) of SST-NPY population markers in SST subtypes (left). RNAscope validation (right) of the SST-NPY population in 2-day-old PFC, detecting: *NPY* (green), *NOS1* (red), and *TACR1* (gray). DAPI counterstain (blue, nuclei). Dotted box indicates magnified example shown below.

(D) Expression (ln CPM) of marker genes for the PV-SST population (left). RNAscope validation of the PV-SST population in 6-year-old PFC (right): *SST* (green) staining SST INs, and *KIAA1199* (red) and *FAM19A4* (gray) staining PV INs. PV-SST neurons show all three markers. DAPI (blue, nuclei). Arrowheads: yellow, PV-SST neurons (*SST*+ *KIAA1199*+ *FAM19A4*+); green, SST (*SST*+); red, PV (*KIA1199*+ *FAM19A4*+). Middle panel shows magnification of colored boxes on right.

(E) Pearson correlation between gene expression of IN subtypes in the adult stage.

(F) UMAP of maturation (UMAT) overlaid by the arcsinh age of each nucleus (left), and UMAT overlaid by select subcluster colorations (CGE-dev and MGE-dev, right). scVelo vector fields projected onto UMATs with colorations of LAMP5-NOS1 and LAMP5-CKK subclusters.

(G) Probability mass flow plots from a CellRank velocity kernel for outflows to LAMP5-NOS1 or LAMP5-CCK clusters across developmental stages from combined CGE-dev and ID2-dev clusters (top) and MGE-dev cluster (bottom).

(H) Expression differences (log$_2$ CPM fold change) between TFs in LAMP5-NOS1 and LAMP5-CCK populations within the infancy to adult stages. Only significant TFs are displayed.

(I) RNAscope validation of a *LAMP5*+ cell population expressing both MGE- and CGE-population markers in 8-year-old PFC, detecting: CGE markers *ADARB2* (green) and *LAMP5* (red); MGE marker *LHX6* (gray). DAPI (blue, nuclei). Arrowheads: orange, nuclei co-staining for all 3 markers; white, *LHX6*+ cells; red, *LAMP5*+ cells; yellow, *ADARB2*+ *LAMP5*+ cells. Right: selected magnified examples of the section locations indicated by dashed boxes and roman numerals on the left.

(J) Expression (log$_2$ CPM) of *NXPH1* and *NR3C2* across development. Dot: sample; line: GAM fit.
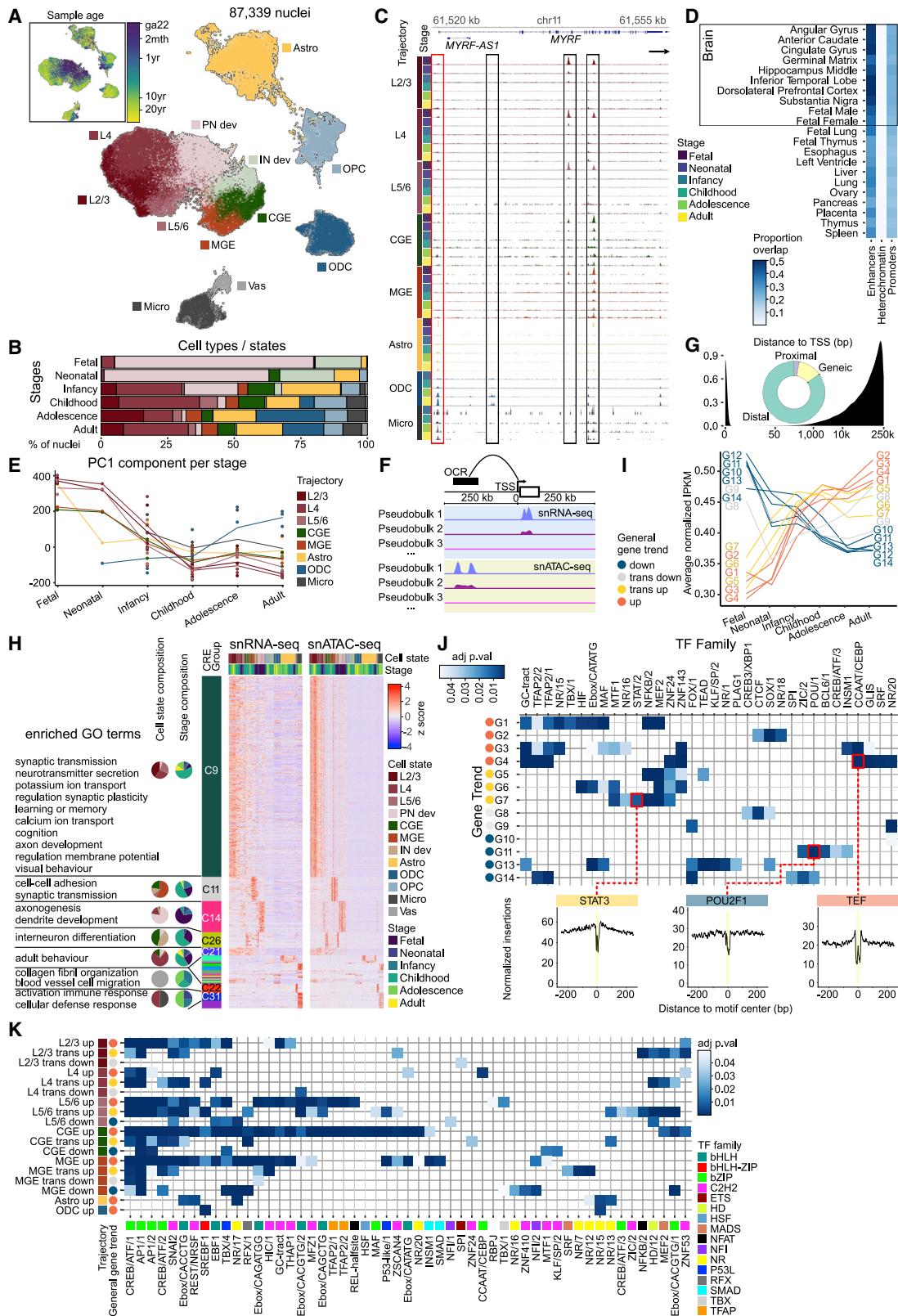
For example, *MYRF*, encoding a TF required for myelination, has a promoter OCR with high accessibility in ODCs starting in childhood, coinciding with myelination and *MYRF* expression (Figures 5C and S5G; Bujalka et al., 2013). OCRs overlap most strongly with predicted ENCODE brain enhancers (39%–50%) (Ernst and Kellis, 2017) but show negligible intersection (<0.1%) with predicted heterochromatin (Figures 5D and S5H) and are enriched in promoters (10%), introns (55%), and exons (10%; Figure S5I), supporting the OCR validity.

OCR accessibility trends over development for all major cell trajectories are similar to those in the snRNA-seq (Figure 5E). Neuronal trajectories show the greatest change in accessibility between neonatal and infancy stages, and to a lesser extent between infancy and childhood. In contrast, accessibility changes in ODCs are most pronounced between childhood and adolescence (Figure 5E), supporting the protracted myelination processes observed by snRNA-seq (Figure 3).

We matched snATAC-seq nuclei to their nearest neighbor in the snRNA-seq data, enabling identification of putative CREs where the change in accessibility and expression of a nearby gene (<250-kb distance; Figure 5F) are correlated (Granja et al., 2019). Across the trajectories, we identified 304,741 CRE-gene pairs that represent potential regulatory interactions, with a median of 5 unique CREs linked to each gene, a long-tailed distribution of link number (Figure S5K), and 84% of CREs located >1 kb from their respective gene (Figure 5G). Clustering CRE-gene pairs into groups by CRE accessibility similarity revealed seven groups specific to neuronal cell types at different developmental stages (Figure 5H). For example, a mature PN group (C9) is linked to processes including learning and memory, cognition, and potassium ion transport, supporting the importance of such transporter up-regulation for mature PNs. CREs in this group are enriched for motifs of TFs involved in cell survival, such as SP1 and SP3 (Figure S5L; Gao et al., 2009; Liang et al., 2013).

As CREs are identified mainly by variation across cell types, we assessed whether changes in CRE accessibility reflected gene expression dynamics. For all snRNA-seq gene trends, we observed that their linked CRE accessibility recapitulated their expression trend over development (Figure 5I). To predict transcriptional regulators of these expression dynamics, we tested for TF motif enrichment in CREs linked to particular gene trends (Figure 5J; Table S5C). Genes with slowly decreasing trends (G11–13, "down") are enriched in developmental TF motifs, such as POU2F1. Further trajectory-specific analysis showed that POU2F1 enrichment is restricted to neurons (Figure S6A; Table S5D), mirroring findings in early human fetal development (Domcke et al., 2020). STAT3, a TF involved in neurite outgrowth (Ihara et al., 1997), is enriched in transiently up gene trends in all major trajectories (Figure 5J). Gene trends with rapidly increasing expression through development (G4, "up") are enriched for TFs in the CCAAT/CEBP family, which plays an essential role in memory (re)consolidation (Arguello et al., 2013), whereas other members, such as TEF, have been associated with major depressive disorder (Hua et al., 2014). TF footprinting analyses further supported the presence of these TFs at the putative CREs (Figure 5J). We used the predicted TF binding sites to construct cell type- and stage-specific regulatory networks (Figure S5M), revealing a central role for *NTRK3* in PNs for ensheathment of neurons during infancy and adolescence.

We leveraged our devDEG gene trends (Figure 2) to investigate trajectory-specific TF enrichment in CREs for devDEGs linked to ion transport (Figure 5K). This identified bZIP family TFs, in particular FOS and JUN that dimerize to form the AP-1 complex (Shaulian and Karin, 2002), as potential regulators of transcriptional programs with increasing expression over development in all neuronal trajectories (Figure 5K). FOS and JUN function in activity-dependent transcriptional control of neural circuitry form and function and support higher cognitive functions (Yap and Greenberg, 2018). Although the predicted regulators of developmentally upregulated ion transporter expression programs showed substantial commonality in MGE and upper-layer PN trajectories, the CGE trajectory appears to be governed by TF regulatory networks shared with deeper layer PN trajectories (Figure 5K), supporting differences observed by snRNA-seq. Collectively, this validates our

Figure 5. A single-nucleus resolution chromatin accessibility map of human PFC development
(A) UMAP of 87,339 nuclei based on chromatin accessibility (snATAC-seq), annotated by cell type/state. Inset UMAP depicts nuclei colored by their sample donor age.
(B) Percentages of cell types/states for each developmental stage.
(C) snATAC-seq signal at *MYRF* for each trajectory and developmental stage. Peaks in boxes exemplify stage- and trajectory-specific dynamics. Red box: promoter.
(D) Proportion of peaks overlapping chromHMM-predicted enhancers, heterochromatin, and promoters across fetal and adult human tissues.
(E) First principal component of accessibility in OCRs for each trajectory across developmental stages.
(F) CRE identification strategy: pseudo-bulks of matching snRNA-seq and snATAC-seq nuclei used to correlate expression and accessibility in peaks near transcriptional start sites (TSSs, <250 kb) of respective gene.
(G) Distribution of the distance between CREs to the nearest TSS of their linked genes, and proportion classified as genic (0 bp), proximal (1–1,000 bp), and distal (>1,000 bp).
(H) Heatmap of CRE accessibility (right) and expression (left) of linked CRE-gene pairs (rows) across 400 pseudo-bulks of snRNA-seq and snATAC-seq nuclei, clustered into groups. CRE-gene pairs assigned to only one group are shown (n = 707). Enriched GO terms and cell-type and stage composition shown in pie charts for selected groups.
(I) Average of normalized snATAC-seq Tn5 insertions per kilobase per million reads mapped (IPKM) across developmental stages for all CREs linked to devDEGs in each gene trend.
(J) Enrichment of TF motif similarity modules in CREs linked to devDEGs in each snRNA-seq gene trend. Color indicates summarized adjusted p value (top). Only TF motif similarity modules enriched in <5 gene trends are displayed. Bias corrected ATAC-footprints for SOX10, POUF21, and TEF centered around the motif-binding site (bottom).
(K) Enrichment of TF motifs in CREs linked to devDEGs in the ion transporter GO term for general gene trends and major trajectories. TF motifs are summarized similarity modules (motif family indicated on x axis). Color represents adjusted p value.

Figure 6. Cell-type- and dynamics-resolved brain-related disease associations and their regulatory drivers

(A) Enrichment of known disease genes for neurological and psychiatric disorders and three control diseases (DisGeNet) in general gene trends across cell type trajectories. Color represents the adjusted p value.

(B) Enrichment of TF motifs in linked CREs for significant combinations of general gene trend and disease. TF motifs are summarized to the TF motif similarity modules. Color represents adjusted p value. TF motif family indicated on the x axis.

(C) TF binding regions: average of normalized snATAC-seq Tn5 IPKM (snATAC-seq reads) in L5/6 trajectory nuclei over development in CREs with a FOS/JUND motif linked to genes from up gene trends, and that are associated with bipolar disorder. Expression: log₂ CPM of *MCTP1* over development in L5/6 neurons (dot: sample; line: GAM fit). CRE region: snATAC-seq signal at a CRE with FOS/JUND motifs (indicated by vertical gray bar) and linked to *MCTP1*. Browser: snATAC-seq signal over developmental stages at *MCTP1*, with linked CRE highlighted.

Cell-type- and dynamics-resolved disease associations and their regulatory drivers

Symptoms for many neurological and psychiatric diseases manifest or worsen at distinct ages, suggesting the involvement

observations from the snRNA-seq and reveals cell-type-specific regulatory programs of postnatal brain development.

of dynamically regulated cellular processes in development. To test for cell type and developmental dynamics contribution to disease, we used our atlas to explore developmental disease-gene associations. Assessing enrichment of brain-disease-associated genes with devDEGs enabled us to identify potentially causative cell types and predict expression dynamics that may indicate when their dysregulation is relevant to disease.

(legend on next page)

Of the 23 diseases analyzed, all show associations with cell-type-specific gene expression trends, except for three blood disorders included as negative controls (Figure 6A). The majority (80%) of associated expression trends have up or transiently up dynamics, demonstrating that developmental upregulation of these disease-gene functions is important. The dynamics of individual genes potentially indicates the developmental timing of their dysfunction relevant to disease manifestation (Figure 6A). Enrichment for the analyzed neurodevelopmental diseases covers all general gene expression trends and almost all cell types. For example, autism spectrum disorder (ASD) shows enrichment for most cell types and dynamics, reflecting the disease complexity (Figure S6B). Multiple disorders that typically occur later in life (dementia, mild cognitive disorder, amnesia, and Huntington's disease) are enriched predominantly for genes transiently upregulated during postnatal development in PN and IN subtypes (Figure 6A) and highly enriched for roles in synaptic signaling (Table S5G). This suggests that disruption of synaptic functions refined during postnatal brain maturation is linked to cognitive disorders of later life.

TFs may play an important role in mediating the temporally defined onsets of these diseases through their regulation of disease-linked genes. For the significant disease associations, we predicted TFs regulating the developmental expression changes by motif enrichment analysis in the CREs linked to the known disease genes of the relevant gene trend (Figure 6B; Table S5F). Perturbation of regulatory activities of such TFs may cause cell-type- and developmental stage-specific dysregulation leading to disease. This revealed a potential role of the AP1/1 and CREB/AFT/1 TF motif families, which includes FOS and JUN, in many brain diseases (Figure 6B). Especially interesting is the association of FOS and JUN with impaired cognition across multiple PNs, as these TFs have been linked to learning and memory mechanisms (Gass et al., 2004). Further analysis revealed that genes associated with bipolar disorder are upregulated during development in L5/6 PNs and linked to CREs enriched for the binding motif of FOS and JUND (Figures 6A and 6B), which are known to be regulated by common bipolar medications (Gao et al., 2021). For example, the bipolar-associated gene *MCTP1* (Scott et al., 2009) is upregulated after birth in L5/6 PNs, has an intronic linked CRE that gains accessibility from infancy onward, and contains a FOS/JUND-binding motif (Figure 6C). This exemplifies how our atlas of PFC development can yield insights into regulatory drivers and implicated cell types in brain disorders.

A reference for cell developmental analysis in health, disease, and organoid models of maturation

We also leveraged our atlas to enhance interpretation of additional snRNA-seq analyses of the PFC and its *in vitro* models, modifying a transfer learning strategy (Lotfollahi et al., 2022) to stably integrate additional snRNA-seq datasets into our reference. We first integrated neurotypical control samples from snRNA-seq characterization of the PFC in ASD (Velmeshev et al., 2019), correctly predicting the original cell-type assignment (>86% accuracy; Figures 7A, S7A, and S7B). We next investigated whether this strategy would enable prediction of developmental state, restricting predictions to astrocytes or L2/3 PNs as they represent the most well-captured developmental trajectories (Figure 1D; Table S3O). For the 4–22 years Velmeshev et al. (2019) samples, predicted ages and actual ages correlate well (Spearman correlation = 0.77 for astrocytes, 0.59 for L2/3 PNs; Figure 7B). We also generated and integrated three additional PFC samples (Table S1B), from donors 261 and 443 days of age and gestational age (ga, in weeks) 39. We correctly estimated their developmental stages using predicted astrocyte nuclei ages (Figure 7B), whereas using L2/3 PNs we predicted that the ga39 sample was neonatal (Figures 7B and 7C). Our strategy shows improved age prediction accuracy and technical variability robustness (Figures S7C and S7D), allowing confident assessment of maturity.

We also explored prediction of cell type and developmental age for contextualizing disease states by integrating an adult glioblastoma snRNA-seq dataset, identifying mesenchymal, neuronal, astrocyte-like, and ODC-like populations (Figure 7D). Around 18% of nuclei align with fetal astrocytes and OPCs (Figure S7E; Table S6A) and likely correspond transcriptomically to progenitor cells and functionally to pluripotent apical glioma stem cells (Couturier et al., 2020), based on higher expression of embryonic stem cell network and cell-cycle pathway genes (Figure S7G).

Figure 7. A reference for cell developmental analysis in health, disease, and organoid models of maturation
(A) UMAP plot overlaid with sample nuclei age (left) and cell type labels (right).
(B) Distributions of age prediction for query astrocyte and L2/3 PN nuclei from control samples (Velmeshev et al., 2019) and our PFC samples (ga39, and 261 and 443 days).
(C) Projection onto reference of nuclei from snRNA-seq of ga39, and 261- and 443-day samples. Density of points indicated by orange contour lines (as also in D and F).
(D) Projection of glioblastoma snRNA-seq nuclei.
(E) Microscopic images of human cerebral organoids cultured for 9 months, showing astrocytes (GFAP+, white), neuronal nuclei (NeuN+, green), and neuronal dendrites (MAP2+, red) on top left, and detailed views on right. RNAscope in 3- and 9-month-old organoids, detecting *BHLHE22* (gray) and *ARHGAP26* (red). DAPI counterstain (blue, nuclei). Scale bars, 100 μm.
(F) Projection onto reference of nuclei from snRNA-seq of human cerebral organoids cultured for 5, 9, or 12 months.
(G) Projection of all organoid snRNA-seq indicating location of PNs reaching postnatal maturity versus those that do not (top left). Volcano plot: differentially expressed gene (DEG) analysis between postnatally mature (purple) and immature (green) PNs. Enriched GO terms for upregulated and downregulated DEGs (bottom). See also Table S6C.
(H) Projection of all organoid snRNA-seq indicating location of postnatally mature PNs versus matched nearest neighbors in snRNA-seq data of normal PFC cells (top left). Volcano plot: DEG analysis between postnatally mature PNs in normal brain (blue) and organoids (purple). Enriched GO terms for upregulated DEGs in brain PNs (bottom left). Ranking of TFs upregulated in brain PNs by z score of network connectivity with other upregulated non-TF genes. Left side of plot indicates if TF is a devDEG of an indicated general gene trend. TF family specified in parentheses.
See also Table S6E.

We next sought to identify the distribution of cell states and their developmental ages in an *in vitro* model of cortical development, by generating cerebral organoids cultured for 5, 9, or 12 months (Iefremova et al., 2017) and profiling them by snRNA-seq (Figures 7E, 7F, and S7E). These organoids contained the expected major cell types (astrocytes, OPCs, PNs, and INs; Figure 7F). With 12-month organoids, we identified non-neuronal nuclei mapping to mature ODCs and vasculature-associated cells, confirming the emergence of these cells with increased culture time (Kanton et al., 2019; Ormel et al., 2018). Organoid neurons mapped to most cortical layers for PNs and both CGE- and MGE-clusters for INs (Figures 7F and 7G). Strikingly, most organoid nuclei align with early fetal populations; however, with increasing organoid age more nuclei show signatures of early postnatal neurons, in particular for PNs and astrocytes, with 2% of PNs at 5 months aligned to early postnatal stages, compared with 15% for the 9- and 12-month organoids (Figures 7F and S7E). Based on this evaluation of molecular equivalence using a real brain development reference, most organoid PNs remain immature. Notably, the less numerous mature PNs could not be discerned using traditional analysis methods in the absence of our reference map (Figure S7H).

To identify genes associated with increased maturity in *in vitro* neuronal development, we conducted differential expression analysis between organoid PNs classified as postnatally mature versus those that are not. Consistent with the temporal changes uncovered in our developmental reference, we found several potassium ion channels (e.g., *KCNB2, KCNMA1*) and neurotransmitter receptors (e.g., *GABRB1*, *GRIA4*) that are also upregulated during normal postnatal development (Figure 7G; Table S6B). Furthermore, upregulated DEGs show a high overlap with devDEGs in up gene trends and are enriched in pathways that are hallmarks of neuronal and brain functional development, whereas downregulated DEGs are enriched in early differentiation and developmental processes (e.g., neuronal precursor cell proliferation, stem cell population maintenance; Figures 7G and S7I; Table S6C).

By matching the mature PNs in organoids to their nearest neighbor nuclei in our developmental reference, we identified genes and their programs that are dysregulated in the postnatal maturity organoid PNs compared with their closest brain equivalents (Figure 7H; Table S6D). We detected downregulation of many genes (n = 7,497) enriched in canonical pathways for synaptic organization, structure, activity, and regulation (Table S6E). Similar to the previous analysis, this included several potassium ion channels (e.g., *KCND2*, *KCNQ5*) and neurotransmitter receptors (e.g., *GRIN2B*, *GRIA1*). Using network approaches based on estimated CREs in PNs and TF motif-binding sites, we ranked TFs by their likelihood of regulating these differences. Many of the highly ranked TFs (n = 31) correspond to devDEGs upregulated in normal brain development (Figure 7H) and were identified as regulators of maturation-induced changes in ion transport in PNs (Figure 5K). Together, this suggests that organoids lack the synaptic networks to achieve PN maturation (Figure 7H) and highlights regulators that may be manipulated in the future to overcome these *in vitro* deficiencies. Overall, these integration analyses exemplify the value of our developmental map as a reference for gauging cellular development in normal, disease, and *in vitro* contexts.

DISCUSSION

Here, we provide a thorough characterization of gene expression and chromatin accessibility dynamics throughout human PFC development at single-cell resolution, allowing the molecular investigation of critical developmental events, some previously only described using neuroimaging and microscopy studies (Sydnor et al., 2021). Although past studies of gene expression in bulk brain tissue showed major expression changes at the prenatal-to-postnatal transition, it was not possible to determine whether they reflected differences in cell-type-specific transcriptional regulation or changes in cellular composition. Our atlas reveals that individual brain cell types respond with substantial and specific transcriptional changes to the transition from fetal to postnatal environment, unique to this early stage. Characterizing the timing of when each neural cell subtype reaches an adult-like transcriptional state reveals surprising diversity in maturation timing across cell types, which will inform future efforts to link subtype-specific functions to the emergence of brain activities during development. By linking molecular dynamics of key genes to the known timing of functional milestones they are associated with, we demonstrate the value of these maps for exploring molecular processes of brain development.

Our analyses provide detailed characterization of postnatal cortical IN development and maturation, unveiling the complexity of the neural subtypes and their development. By charting IN emergence, dynamics, and specialization, detecting poorly described neuronal subtypes, and resolving previously unknown developmental origins, these findings significantly extend our knowledge of human cortical IN maturation and composition.

This developmental map provides a powerful resource for improving our understanding of brain-related disorders, including causative cell types, developmental windows of dysregulation, and regulators for multiple disorders. Our reference and stable integration technique enable investigation of cell-type-specific acceleration of maturation. Applying this approach to cerebral organoids, which are increasingly used to model brain development and disorders (Sidhaye and Knoblich, 2021; Tian et al., 2020), suggests that in their current state these may be unsuitable to model postnatal onset disorders. Prediction of TFs that regulate neuronal maturation processes that are dysregulated in brain organoids indicates regulatory pathways that may be pivotal in advancing neuronal maturity. Collectively, this reference constitutes an important advance in the understanding of brain cell states and dynamics in health and disease.

Limitations of the study

Although our study spans gestation to adulthood, the scarcity of biobanked samples, particularly in the third trimester, makes complete temporal coverage and controlling for inter-sample variability challenging, limiting interpretation of cell-type abundance changes. Although sex differences play a role in PFC maturation, with males typically developing executive functions later than females, we were unable to account for this due to sample availability limitations. We acknowledge that our

application of velocity-based trajectory methods is outside the implied theory, where the robustness of velocity is untested.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell lines
  - Human subjects
- METHOD DETAILS
  - Data reporting
  - Generation of cerebral organoids
  - Histology and imaging
  - RNAscope *in-situ* hybridization
  - Sample preparation
  - Library preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - snRNA-seq analysis
  - snATAC-seq analysis
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2022.09.039.

AUTHOR CONTRIBUTIONS

R.L. conceived the study. R.L., S.F., C.A.H., R.K.S., and D.P. designed experiments and analyses. Sample preparation and data collection, R.K.S., D.P., D.B.V.-L., L.M., C.P., O.C., and J.M.P.; sequencing, D.P. and J.P.; data analysis and interpretation, C.A.H., S.F., R.K.S., D.P., S.B., O.C., C.P., A.A.-F., and R.L.; algorithm implementation, C.A.H., S.F., and J.M.P. Manuscript was writ-

ten by R.K.S., S.F., C.A.H., D.P., S.B., and R.L. All authors approved of and contributed to the final manuscript. C.A.H., R.K.S., S.F., and D.P. contributed equally and have the right to list their name first in their CV.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Agirman, G., Broix, L., and Nguyen, L. (2017). Cerebral cortex development: an outside-in perspective. FEBS Lett. *591*, 3978–3992.

Aiken, J., Buscaglia, G., Bates, E.A., and Moore, J.K. (2017). The α-tubulin gene TUBA1A in Brain Development: A Key Ingredient in the Neuronal Isotype Blend. J. Dev. Biol. *5*, 8.

Alcamo, E.A., Chirivella, L., Dautzenberg, M., Dobreva, G., Fariñas, I., Grosschedl, R., and McConnell, S.K. (2008). Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. Neuron *57*, 364–377.

Andrews, W., Barber, M., Hernandez-Miranda, L.R., Xian, J., Rakic, S., Sundaresan, V., Rabbitts, T.H., Pannell, R., Rabbitts, P., Thompson, H., et al. (2008). The role of Slit-Robo signaling in the generation, migration and morphological differentiation of cortical interneurons. Dev. Biol. *313*, 648–658.

Ang, Y.-S., Tsai, S.-Y., Lee, D.-F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. Cell *145*, 183–197.

Arain, M., Haque, M., Johal, L., Mathur, P., Nel, W., Rais, A., Sandhu, R., and Sharma, S. (2013). Maturation of the adolescent brain. Neuropsychiatr. Dis. Treat. *9*, 449–461.

Arguello, A.A., Ye, X., Bozdagi, O., Pollonini, G., Tronel, S., Bambah-Mukku, D., Huntley, G.W., Platano, D., and Alberini, C.M. (2013). CCAAT enhancer binding protein δ plays an essential role in memory consolidation and reconsolidation. J. Neurosci. *33*, 3646–3658.

Arshad, A., Vose, L.R., Vinukonda, G., Hu, F., Yoshikawa, K., Csiszar, A., Brumberg, J.C., and Ballabh, P. (2016). Extended production of cortical interneurons into the third trimester of human gestation. Cereb. Cortex *26*, 2242–2256.

Baker, A., Kalmbach, B., Morishima, M., Kim, J., Juavinett, A., Li, N., and Dembrow, N. (2018). Specialized subpopulations of deep-layer pyramidal neurons in the neocortex: bridging cellular properties to functional consequences. J. Neurosci. *38*, 5441–5455.

Batiuk, M.Y., Martirosyan, A., Wahis, J., de Vin, F., Marneffe, C., Kusserow, C., Koeppen, J., Viana, J.F., Oliveira, J.F., Voet, T., et al. (2020). Identification of region-specific astrocyte subtypes at single cell resolution. Nat. Commun. *11*, 1220.

Bauer, L.O., and Covault, J.M. (2020). GRM8 genotype is associated with externalizing disorders and greater inter-trial variability in brain activation during a response inhibition task. Clin. Neurophysiol. *131*, 1180–1186.

Bayatti, N., Moss, J.A., Sun, L., Ambrose, P., Ward, J.F.H., Lindsay, S., and Clowry, G.J. (2008). A molecular neuroanatomical study of the developing human neocortex from 8 to 17 postconceptional weeks revealing the early differentiation of the subplate and subventricular zone. Cereb. Cortex *18*, 1536–1548.

Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. *38*, 1408–1414.

Blatow, M., Rozov, A., Katona, I., Hormuzdi, S.G., Meyer, A.H., Whittington, M.A., Caputi, A., and Monyer, H. (2003). A novel network of multipolar bursting

interneurons generates theta frequency oscillations in neocortex. Neuron 38, 805–817.

Boldog, E., Bakken, T.E., Hodge, R.D., Novotny, M., Aevermann, B.D., Baka, J., Bordé, S., Close, J.L., Diez-Fuertes, F., Ding, S.-L., et al. (2018). Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. Nat. Neurosci. 21, 1185–1195.

Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D., et al. (2008). Satb2 is a postmitotic determinant for upper-layer neuron specification in the neocortex. Neuron 57, 378–392.

Brunjes, P.C., and Osterberg, S.K. (2015). Developmental markers expressed in neocortical layers are differentially exhibited in olfactory cortex. PLoS One 10, e0138541.

Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., et al. (2013). API design for machine learning software: experiences from the scikit-learn project. Preprint at arXiv. arXiv:1309.0238.

Bujalka, H., Koenning, M., Jackson, S., Perreau, V.M., Pope, B., Hay, C.M., Mitew, S., Hill, A.F., Lu, Q.R., Wegner, M., et al. (2013). MYRF is a membrane-associated transcription factor that autoproteolytically cleaves to directly activate myelin genes. PLoS Biol. 11, e1001625.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420.

Cahill, K.M., Huo, Z., Tseng, G.C., Logan, R.W., and Seney, M.L. (2018). Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. Sci. Rep. 8, 9588.

Colantuoni, C., Lipska, B.K., Ye, T., Hyde, T.M., Tao, R., Leek, J.T., Colantuoni, E.A., Elkahloun, A.G., Herman, M.M., Weinberger, D.R., et al. (2011). Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 478, 519–523.

Cooper, J.A. (2014). Molecules and mechanisms that regulate multipolar migration in the intermediate zone. Front. Cell. Neurosci. 8, 386.

Couturier, C.P., Ayyadhury, S., Le, P.U., Nadaf, J., Monlong, J., Riva, G., Allache, R., Baig, S., Yan, X., Bourgey, M., et al. (2020). Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. Nat. Commun. 11, 3406.

Cubelos, B., Briz, C.G., Esteban-Ortega, G.M., and Nieto, M. (2015). Cux1 and Cux2 selectively target basal and apical dendritic compartments of layer II–III cortical neurons. Dev. Neurobiol. 75, 163–172.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. USA 112, 7285–7290.

Denisenko, E., Guo, B.B., Jones, M., Hou, R., de Kock, L., Lassmann, T., Poppe, D., Clément, O., Simmons, R.K., Lister, R., et al. (2020). Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. Genome Biol. 21, 130.

Dennis, D.J., Han, S., and Schuurmans, C. (2019). bHLH transcription factors in neural development, disease, and reprogramming. Brain Res. 1705, 48–65.

DeTomaso, D., and Yosef, N. (2021). Hotspot identifies informative gene modules across modalities of single-cell genomics. Cell Syst. 12, 446–456.e9.

Di Bella, D.J., Habibi, E., Stickels, R.R., Scalia, G., Brown, J., Yadollahpour, P., Yang, S.M., Abbate, C., Biancalani, T., Macosko, E.Z., et al. (2021). Molecular logic of cellular diversification in the mouse cerebral cortex. Nature 595, 554–559.

Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., Pliner, H.A., Aldinger, K.A., Pokholok, D., Zhang, F., Milbank, J.H., et al. (2020). A human cell atlas of fetal chromatin accessibility. Science 370, eaba7612.

Duncan, G.J., Simkins, T.J., and Emery, B. (2021). Neuron-oligodendrocyte interactions in the structure and integrity of axons. Front. Cell Dev. Biol. 9, 653101.

ENCODE Project Consortium, Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., et al. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583, 699–710.

Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. Nat. Protoc. 12, 2478–2492.

Fan, X., Dong, J., Zhong, S., Wei, Y., Wu, Q., Yan, L., Yong, J., Sun, L., Wang, X., Zhao, Y., et al. (2018). Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. Cell Res. 28, 730–745.

Ferguson, B.R., and Gao, W.-J. (2018). PV interneurons: critical regulators of E/I balance for prefrontal cortex-dependent behavior and psychiatric disorders. Front. Neural Circuits 12, 37.

Ferland, R.J., Cherry, T.J., Preware, P.O., Morrisey, E.E., and Walsh, C.A. (2003). Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. J. Comp. Neurol. 460, 266–279.

Fonseca, M.I., Chu, S.-H., Hernandez, M.X., Fang, M.J., Modarresi, L., Selvan, P., MacGregor, G.R., and Tenner, A.J. (2017). Cell-specific deletion of C1qa identifies microglia as the dominant source of C1q in mouse brain. J. Neuroinflammation 14, 48.

Galazo, M.J., Emsley, J.G., and Macklis, J.D. (2016). Corticothalamic projection neuron development beyond subtype specification: Fog2 and intersectional controls regulate intraclass neuronal diversity. Neuron 91, 90–106.

Gao, T.-H., Ni, R.-J., Liu, S., Tian, Y., Wei, J., Zhao, L., Wang, Q., Ni, P., Ma, X., and Li, T. (2021). Chronic lithium exposure attenuates ketamine-induced mania-like behavior and c-Fos expression in the forebrain of mice. Pharmacol. Biochem. Behav. 202, 173108.

Gao, Z., Ure, K., Ables, J.L., Lagace, D.C., Nave, K.-A., Goebbels, S., Eisch, A.J., and Hsieh, J. (2009). Neurod1 is essential for the survival and maturation of adult-born neurons. Nat. Neurosci. 12, 1090–1092.

Gass, P., Fleischmann, A., Hvalby, O., Jensen, V., Zacher, C., Strekalova, T., Kvello, A., Wagner, E.F., and Sprengel, R. (2004). Mice with a fra-1 knock-in into the c-fos locus show impaired spatial but regular contextual learning and normal LTP. Brain Res. Mol. Brain Res. 130, 16–22.

Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. 53, 403–411.

Granja, J.M., Klemm, S., McGinnis, L.M., Kathiria, A.S., Mezger, A., Corces, M.R., Parks, B., Gars, E., Liedtke, M., Zheng, G.X.Y., et al. (2019). Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat. Biotechnol. 37, 1458–1465.

Guillery, R.W. (2005). Is postnatal neocortical maturation hierarchical? Trends Neurosci. 28, 512–517.

Gulati, G.S., Sikandar, S.S., Wesche, D.J., Manjunath, A., Bharadwaj, A., Berger, M.J., Ilagan, F., Kuo, A.H., Hsieh, R.W., Cai, S., et al. (2020). Single-cell transcriptional diversity is a hallmark of developmental potential. Science 367, 405–411.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for the ENCODE Project. Genome Res. 22, 1760–1774.

Heimberg, G., Bhatnagar, R., El-Samad, H., and Thomson, M. (2016). Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing. Cell Syst. 2, 239–250.

Hernandez, D.G., Nalls, M.A., Moore, M., Chong, S., Dillman, A., Trabzuni, D., Gibbs, J.R., Ryten, M., Arepalli, S., Weale, M.E., et al. (2012). Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. Neurobiol. Dis. 47, 20–28.

Herrero-Navarro, Á., Puche-Aroca, L., Moreno-Juan, V., Sempere-Ferràndez, A., Espinosa, A., Susín, R., Torres-Masjoan, L., Leyva-Díaz, E., Karow, M.,

Figueres-Oñate, M., et al. (2021). Astrocytes and neurons share region-specific transcriptional signatures that confer regional identity to neuronal reprogramming. Sci. Adv. 7, eabe8978.

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.

Hoerder-Suabedissen, A., and Molnár, Z. (2015). Development, evolution and pathology of neocortical subplate neurons. Nat. Rev. Neurosci. 16, 133–146.

Hou, R., Denisenko, E., Ong, H.T., Ramilowski, J.A., and Forrest, A.R.R. (2020). Predicting cell-to-cell communication networks using NATMI. Nat. Commun. 11, 5011.

Hu, J.S., Vogt, D., Sandberg, M., and Rubenstein, J.L. (2017). Cortical interneuron development: a tale of time and space. Development 144, 3867–3878.

Hua, P., Liu, W., Chen, D., Zhao, Y., Chen, L., Zhang, N., Wang, C., Guo, S., Wang, L., Xiao, H., et al. (2014). Cry1 and Tef gene polymorphisms are associated with major depressive disorder in the Chinese population. J. Affect. Disord. 157, 100–103.

Huttenlocher, P.R., and Dabholkar, A.S. (1997). Regional differences in synaptogenesis in human cerebral cortex. J. Comp. Neurol. 387, 167–178.

Iefremova, V., Manikakis, G., Krefft, O., Jabali, A., Weynans, K., Wilkens, R., Marsoner, F., Brändl, B., Müller, F.-J., Koch, P., et al. (2017). An organoid-based model of cortical development identifies non-cell-autonomous defects in Wnt signaling contributing to Miller-Dieker syndrome. Cell Rep. 19, 50–59.

Ihara, S., Nakajima, K., Fukada, T., Hibi, M., Nagata, S., Hirano, T., and Fukui, Y. (1997). Dual control of neurite outgrowth by STAT3 and MAP kinase in PC12 cells stimulated with interleukin-6. EMBO J. 16, 5345–5352.

Jolly, S., Lang, V., Koelzer, V.H., Sala Frigerio, C., Magno, L., Salinas, P.C., Whiting, P., and Palomer, E. (2019). Single-cell quantification of mRNA expression in the human brain. Sci. Rep. 9, 12353.

Kamimoto, K., Hoffmann, C.M., and Morris, S.A. (2020). CellOracle: dissecting cell identity via network inference and in silico gene perturbation. Preprint at bioRxiv. https://doi.org/10.1101/2020.02.17.947416.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature 478, 483–489.

Kang, J.B., Nathan, A., Weinand, K., Zhang, F., Millard, N., Rumker, L., Moody, D.B., Korsunsky, I., and Raychaudhuri, S. (2021). Efficient and precise single-cell reference atlas mapping with Symphony. Nat. Commun. 12, 5890.

Kanton, S., Boyle, M.J., He, Z., Santel, M., Weigert, A., Sanchís-Calleja, F., Guijarro, P., Sidow, L., Fleck, J.S., Han, D., et al. (2019). Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature 574, 418–422.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 46, D260–D266.

Kim, H., and Park, H. (2007). Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis. Bioinformatics 23, 1495–1502.

Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2016). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296.

Kostović, I., and Judas, M. (2010). The development of the subplate and thalamocortical connections in the human foetal brain. Acta Paediatr. 99, 1119–1127.

Kostović, I., Sedmak, G., and Judaš, M. (2019). Neural histology and neurogenesis of the human fetal and infant brain. Neuroimage 188, 743–773.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature 560, 494–498.

Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., et al. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat. Biotechnol. 36, 70–80.

Lange, M., Bergen, V., Klein, M., Setty, M., Reuter, B., Bakhti, M., Lickert, H., Ansari, M., Schniering, J., Schiller, H.B., et al. (2020). CellRank for directed single-cell fate mapping. Preprint at bioRxiv. https://doi.org/10.1101/2020.10.19. 345983.

Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G.K., and Ritchie, M.E. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res. 5, 1408.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29.

Leifer, D., Li, Y.L., and Wehr, K. (1997). Myocyte-specific enhancer binding factor 2C expression in fetal mouse brain development. J. Mol. Neurosci. 8, 131–143.

Li, M., Santpere, G., Imamura Kawasawa, Y., Evgrafov, O.V., Gulden, F.O., Pochareddy, S., Sunkin, S.M., Li, Z., Shin, Y., Zhu, Y., et al. (2018). Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science 362.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1779.

Li, Y.E., Preissl, S., Hou, X., Zhang, Z., Zhang, K., Fang, R., Qiu, Y., Poirion, O., Li, B., Liu, H., et al. (2020). An atlas of gene regulatory elements in adult mouse cerebrum. Preprint at bioRxiv. https://doi.org/10.1101/2020.05.10.087585.

Liang, H., Xiao, G., Yin, H., Hippenmeyer, S., Horowitz, J.M., and Ghashghaei, H.T. (2013). Neural development is dependent on the function of specificity protein 2 in cell cycle progression. Development 140, 552–561.

Lim, L., Mi, D., Llorca, A., and Marín, O. (2018). Development and functional diversification of cortical interneurons. Neuron 100, 294–313.

Long, J.E., Cobos, I., Potter, G.B., and Rubenstein, J.L.R. (2009). Dlx1&2 and Mash1 transcription factors control MGE and CGE patterning and differentiation through parallel and overlapping pathways. Cereb. Cortex 19 (Suppl 1), i96–i106.

Lopez, R., Regier, J., Cole, M.B., Jordan, M.I., and Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nat. Methods 15, 1053–1058.

Lotfollahi, M., Naghipourfar, M., Luecken, M.D., Khajavi, M., Büttner, M., Wagenstetter, M., Avsec, Ž., Gayoso, A., Yosef, N., Interlandi, M., et al. (2022). Mapping single-cell data to reference atlases by transfer learning. Nat. Biotechnol. 40, 121–130.

Lun, A.T.L., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res 5, 2122.

Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippani, M., et al. (2021). Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. 24, 425–436.

McInnes, L., Healy, J., Saul, N., and Großberger, L. (2018). UMAP: uniform manifold approximation and projection. J. Open Source Software 3, 861.

Mickelsen, L.E., Bolisetty, M., Chimileski, B.R., Fujita, A., Beltrami, E.J., Costanzo, J.T., Naparstek, J.R., Robson, P., and Jackson, A.C. (2019). Single-cell transcriptomic analysis of the lateral hypothalamic area reveals molecularly distinct populations of inhibitory and excitatory neurons. Nat. Neurosci. 22, 642–656.

Miller, D.J., Duka, T., Stimpson, C.D., Schapiro, S.J., Baze, W.B., McArthur, M.J., Fobbs, A.J., Sousa, A.M.M., Sestan, N., Wildman, D.E., et al. (2012). Prolonged myelination in human neocortical evolution. Proc. Natl. Acad. Sci. USA 109, 16480–16485.

Mittnenzweig, M., Mayshar, Y., Cheng, S., Ben-Yair, R., Hadas, R., Rais, Y., Chomsky, E., Reines, N., Uzonyi, A., Lumerman, L., et al. (2021). A single-embryo, single-cell time-resolved model for mouse gastrulation. Cell 184, 2825–2842.e22.

Miyoshi, G., and Fishell, G. (2012). Dynamic FoxG1 expression coordinates the integration of multipolar pyramidal neuron precursors into the cortical plate. Neuron 74, 1045–1058.

Miyoshi, G., Hjerling-Leffler, J., Karayannis, T., Sousa, V.H., Butt, S.J.B., Battiste, J., Johnson, J.E., Machold, R.P., and Fishell, G. (2010). Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. J. Neurosci. 30, 1582–1594.

Molnár, Z., Clowry, G.J., Šestan, N., Alzu'bi, A., Bakken, T., Hevner, R.F., Hüppi, P.S., Kostović, I., Rakic, P., Anton, E.S., et al. (2019). New insights into the development of the human cerebral cortex. J. Anat. 235, 432–451.

Morabito, S., Miyoshi, E., Michael, N., Shahin, S., Martini, A.C., Head, E., Silva, J., Leavy, K., Perez-Rosendahl, M., and Swarup, V. (2021). Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. Nat. Genet. 53, 1143–1155.

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

Oishi, K., Nakagawa, N., Tachikawa, K., Sasaki, S., Aramaki, M., Hirano, S., Yamamoto, N., Yoshimura, Y., and Nakajima, K. (2016). Identity of neocortical layer 4 neurons is specified through correct positioning into the cortex. eLife 5, e10907.

Ormel, P.R., Vieira de Sá, R., van Bodegraven, E.J., Karst, H., Harschnitz, O., Sneeboer, M.A.M., Johansen, L.E., van Dijk, R.E., Scheefhals, N., Berdenis van Berlekom, A., et al. (2018). Microglia innately develop within cerebral organoids. Nat. Commun. 9, 4167.

Overstreet-Wadiche, L., and McBain, C.J. (2015). Neurogliaform cells in cortical circuits. Nat. Rev. Neurosci. 16, 458–468.

Paredes, M.F., James, D., Gil-Perotin, S., Kim, H., Cotter, J.A., Ng, C., Sandoval, K., Rowitch, D.H., Xu, D., McQuillen, P.S., et al. (2016). Extensive migration of young neurons into the infant human frontal lobe. Science 354, aaf7073.

Perlman, K., Couturier, C.P., Yaqubi, M., Tanti, A., Cui, Q.-L., Pernin, F., Stratton, J.A., Ragoussis, J., Healy, L., Petrecca, K., et al. (2020). Developmental trajectory of oligodendrocyte progenitor cells in the human brain revealed by single cell RNA sequencing. Glia 68, 1291–1303.

Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., and Furlong, L.I. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. 48, D845–D855.

Polański, K., Young, M.D., Miao, Z., Meyer, K.B., Teichmann, S.A., and Park, J.-E. (2020). BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics 36, 964–965.

Puranam, R.S., He, X.P., Yao, L., Le, T., Jang, W., Rehder, C.W., Lewis, D.V., and McNamara, J.O. (2015). Disruption of Fgf13 causes synaptic excitatory-inhibitory imbalance and genetic epilepsy and febrile seizures plus. J. Neurosci. 35, 8866–8881.

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198.

Ren, Q., Chen, K., and Paulsen, I.T. (2007). TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res. 35, D274–D279.

Rivers, L.E., Young, K.M., Rizzi, M., Jamen, F., Psachoulia, K., Wade, A., Kessaris, N., and Richardson, W.D. (2008). PDGFRA/NG2 glia generate myelinating oligodendrocytes and piriform projection neurons in adult mice. Nat. Neurosci. 11, 1392–1401.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25.

Ruzicka, W.B., Brad Ruzicka, W., Mohammadi, S., Davila-Velderrain, J., Subburaju, S., Tso, D.R., Hourihan, M., and Kellis, M. (2020). Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. Preprint at medRxiv. https://doi.org/10.1101/2020.11.06.20225342.

Sarropoulos, I., Sepp, M., Frömel, R., Leiss, K., Trost, N., Leushkin, E., Okonechnikov, K., Joshi, P., Kutscher, L.M., Cardoso-Moreira, M., et al. (2020). The regulatory landscape of cells in the developing mouse cerebellum. Preprint at bioRxiv. https://doi.org/10.1101/2021.01.29.428632.

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

Schuman, B., Machold, R.P., Hashikawa, Y., Fuzik, J., Fishell, G.J., and Rudy, B. (2019). Four unique interneuron populations reside in neocortical Layer 1. J. Neurosci. 39, 125–139.

Scott, L.J., Muglia, P., Kong, X.Q., Guan, W., Flickinger, M., Upmanyu, R., Tozzi, F., Li, J.Z., Burmeister, M., Absher, D., et al. (2009). Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. Proc. Natl. Acad. Sci. USA 106, 7501–7506.

Servén, D., Brummitt, C., and Abedi, H. (2018). hlink. Dswah/Pygam: V0.8.0 (Zenodo). https://doi.org/10.5281/zenodo.1476122.

Shaulian, E., and Karin, M. (2002). AP-1 as a regulator of cell life and death. Nat. Cell Biol. 4, E131–E136.

Sidhaye, J., and Knoblich, J.A. (2021). Brain organoids: an ensemble of bioassays to investigate human neurodevelopment and disease. Cell Death Differ. 28, 52–67.

Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., and Sestan, N. (2016). The cellular and molecular landscapes of the developing human central nervous system. Neuron 89, 248–268.

Su, S., Law, C.W., Ah-Cann, C., Asselin-Labat, M.-L., Blewitt, M.E., and Ritchie, M.E. (2017). Glimma: interactive graphics for gene expression analysis. Bioinformatics 33, 2050–2052.

Supekar, K., Musen, M., and Menon, V. (2009). Development of large-scale functional brain networks in children. PLoS Biol. 7, e1000157.

Sydnor, V.J., Larsen, B., Bassett, D.S., Alexander-Bloch, A., Fair, D.A., Liston, C., Mackey, A.P., Milham, M.P., Pines, A., Roalf, D.R., et al. (2021). Neurodevelopment of the association cortices: patterns, mechanisms, and implications for psychopathology. Neuron 109, 2820–2846.

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al. (2018). Shared and distinct transcriptomic cell types across neocortical areas. Nature 563, 72–78.

Thomson, A.M., and Lamy, C. (2007). Functional maps of neocortical local circuitry. Front. Neurosci. 1, 19–42.

Tian, A., Muffat, J., and Li, Y. (2020). Studying human neurodevelopment and diseases using 3D brain organoids. J. Neurosci. 40, 1186–1193.

Tochigi, M., Iwamoto, K., Bundo, M., Sasaki, T., Kato, N., and Kato, T. (2008). Gene expression profiling of major depression and suicide in the prefrontal cortex of postmortem brains. Neurosci. Res. 60, 184–191.

Tomita, K., Gotoh, H., Tomita, K., Yamauchi, N., and Sanbo, M. (2012). Multiple patterns of spatiotemporal changes in layer-specific gene expression in the developing visual cortex of higher mammals. Neurosci. Res. 73, 207–217.

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

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. 32, 381–386.

Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic interneurons in the neocortex: From cellular properties to circuits. Neuron 91, 260–292.

Trevino, A.E., Müller, F., Andersen, J., Sundaram, L., Kathiria, A., Shcherbina, A., Farh, K., Chang, H.Y., Paşca, A.M., Kundaje, A., et al. (2021). Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell 184, 5053–5069.e23.

Trevino, A.E., Sinnott-Armstrong, N., Andersen, J., Yoon, S.J., Huber, N., Pritchard, J.K., Chang, H.Y., Greenleaf, W.J., and Paşca, S.P. (2020). Chromatin accessibility dynamics in a model of human forebrain development. Science 367, eaay1645.

Tricoire, L., Pelkey, K.A., Daw, M.I., Sousa, V.H., Miyoshi, G., Jeffries, B., Cauli, B., Fishell, G., and McBain, C.J. (2010). Common origins of hippocampal Ivy and nitric oxide synthase expressing neurogliaform cells. J. Neurosci. 30, 2165–2176.

Turnescu, T., Arter, J., Reiprich, S., Tamm, E.R., Waisman, A., and Wegner, M. (2018). Sox8 and Sox10 jointly maintain myelin gene expression in oligodendrocytes. Glia 66, 279–294.

Uzquiano, A., Kedaigle, A.J., Pigoni, M., Paulsen, B., Adiconis, X., Kim, K., Faits, T., Nagaraja, S., Antón-Bolaños, N., Gerhardinger, C., et al. (2022). Single-cell multiomics atlas of organoid development uncovers longitudinal molecular programs of cellular diversification of the human cerebral cortex. Preprint at bioRxiv. https://doi.org/10.1101/2022.03.17.484798.

Valério-Gomes, B., Guimarães, D.M., Szczupak, D., and Lent, R. (2018). The absolute number of oligodendrocytes in the adult mouse brain. Front. Neuroanat. 12, 90.

Valero, M., Viney, T.J., Machold, R., Mederos, S., Zutshi, I., Schuman, B., Senzai, Y., Rudy, B., and Buzsáki, G. (2021). Sleep down state-active ID2/Nkx2.1 interneurons in the neocortex. Nat. Neurosci. 24, 401–411.

van Tilborg, E., de Theije, C.G.M., van Hal, M., Wagenaar, N., de Vries, L.S., Benders, M.J., Rowitch, D.H., and Nijboer, C.H. (2018). Origin and dynamics of oligodendrocytes in the developing brain: implications for perinatal white matter injury. Glia 66, 221–238.

Vanlandewijck, M., He, L., Mäe, M.A., Andrae, J., Ando, K., Del Gaudio, F., Nahar, K., Lebouvier, T., Laviña, B., Gouveia, L., et al. (2018). A molecular atlas of cell types and zonation in the brain vasculature. Nature 554, 475–480.

Velmeshev, D., Schirmer, L., Jung, D., Haeussler, M., Perez, Y., Mayer, S., Bhaduri, A., Goyal, N., Rowitch, D.H., and Kriegstein, A.R. (2019). Single-cell genomics identifies cell type-specific molecular changes in autism. Science 364, 685–689.

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., et al. (2020). Global reference mapping of human transcription factor footprints. Nature 583, 729–736.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272.

Vullhorst, D., Neddens, J., Karavanova, I., Tricoire, L., Petralia, R.S., McBain, C.J., and Buonanno, A. (2009). Selective expression of ErbB4 in interneurons, but not pyramidal cells, of the rodent hippocampus. J. Neurosci. 29, 12255–12264.

Whitaker, K.J., Vértes, P.E., Romero-Garcia, R., Váša, F., Moutoussis, M., Prabhu, G., Weiskopf, N., Callaghan, M.F., Wagstyl, K., Rittman, T., et al. (2016). Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. Proc. Natl. Acad. Sci. USA 113, 9105–9110.

Williamson, J.M., and Lyons, D.A. (2018). Myelin dynamics Throughout life: an ever-changing landscape? Front. Cell. Neurosci. 12, 424.

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

Wolf, F.A., Hamey, F.K., Plass, M., Solana, J., Dahlin, J.S., Göttgens, B., Rajewsky, N., Simon, L., and Theis, F.J. (2019). PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 20, 59.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: computational identification of cell doublets in single-cell transcriptomic data. Cell Syst. 8, 281–291.e9.

Wu, Q.-F., Yang, L., Li, S., Wang, Q., Yuan, X.-B., Gao, X., Bao, L., and Zhang, X. (2012). Fibroblast growth factor 13 is a microtubule-stabilizing protein regulating neuronal polarization and migration. Cell 149, 1549–1564.

Yap, E.-L., and Greenberg, M.E. (2018). Activity-regulated transcription: bridging the gap between neural activity and behavior. Neuron 100, 330–348.

Zecevic, N., Hu, F., and Jakovcevski, I. (2011). Interneurons in the developing human neocortex. Dev. Neurobiol. 71, 18–33.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zhou, Q., and Anderson, D.J. (2002). The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. Cell 109, 61–73.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
mouse anti-MAP-2	Millipore	Cat# AB5622; RRID: AB_91939
rabbit anti-GFAP	Dako	Cat# ZO33429-2; RRID: AB_10013482
APC-linked mouse anti-NeuN	Novus Biologicals	Cat# NBP1-92693APC; RRID: AB_2894834
AlexaFluor 555, goat anti-mouse IgG (H&L)	Abcam	Cat# ab150114; RRID: AB_2687594
AlexaDluor 647, donkey anti-rabbit IgG (H&L)	Abcam	Cat# ab150063; RRID: AB_2687541
Chemicals, peptides, and recombinant proteins		
Essential 8 medium	Thermo Fisher	Cat# A1517001
Y-27632 2HCl	Selleckchem	Cat# S1049
DMEM/F12 medium	Thermo Fisher	Cat# 11320033
Neurobasal medium	Thermo Fisher	Cat# 21103049
N2 MAX Media Supplement (100x)	R&D Systems	Cat# AR009
MACS NeuroBrew-21	Miltenyi Biotec	Cat# 130-093-566
MEM Non-Essential Amino Acids Solution (100x)	Thermo Fisher	Cat# 11140050
GlutaMAX Supplement	Thermo Fisher	Cat# 35050061
Penicillin-Streptomycin (10,000 U/ml)	Thermo Fisher	Cat# 15140122
Heparin	Sigma	Cat# H3149
LDN-193189	Cayman Chemical	Cat# 19396
A83-01	Cayman Chemical	Cat# 9001799
IWR-1-endo	Adooq Bioscience	Cat# A12737
Cultrex 3D RGF BME	R&D Systems	Cat# 3445-005-001
MACS NeuroBrew-21 w/o Vitamin A	Miltenyi Biotec	Cat# 130-097-263
Insulin solution human	Sigma	Cat# I9278
DAPI	Cayman Chemical	Cat# 14285
RNAsin Plus RNAse inhibitor	Promega	Cat# N2615
Critical commercial assays		
Chromium Single cell 3′ GEM, Library & Gel Bead Kit v3	10x Genomics	PN: 1000075
ATAC-seq NextGEM kit	10x Genomics	PN: 1000175
Experimental models: Cell lines		
Human: Passage 40 H9 ES cells (WA09 cell line)	WiCell	RRID: CVCL_9773
Deposited data		
snRNA-seq, snATAC-seq	This study	GEO: GSE168408
Bulk RNA-seq	Li et al., 2018	http://development.psychencode.org
scRNA-seq data	Velmeshev et al., 2019	https://autism.cells.ucsc.edu
Software and algorithms		
Original code used for data analysis	This paper	https://doi.org/10.5281/zenodo.7113422
ImageJ (Fiji)	Schindelin et al., 2012	https://imagej.net/software/fiji
Genome assembly	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/human (hg19/GrChr37.p13)
Gene annotation	Harrow et al., 2012	https://www.gencodegenes.org/human/release_19.html

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell Ranger	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/downloads/3.1
Cell Ranger-ATAC	10x Genomics	https://support.10xgenomics.com/single-cell-atac/software/downloads/1.2
scanpy	Wolf et al., 2018	https://pypi.org/project/scanpy/1.6.0
anndata	Wolf et al., 2018	https://pypi.org/project/anndata/0.7.4
scrublet	Wolock et al., 2019	https://pypi.org/project/scrublet/0.2.2
Harmony	Korsunsky et al., 2019	https://pypi.org/project/harmonyTS/0.1.4
BBKNN	Polański et al., 2020	https://pypi.org/project/bbknn/1.5.1
Hotspot	DeTomaso and Yosef, 2021	https://github.com/yoseflab/Hotspot/tree/0.9.1_release
loompy	Linnarsson Lab	https://pypi.org/project/loompy/3.0.6
scVelo	Bergen et al., 2020	https://pypi.org/project/scvelo/0.2.2
CytoTRACE via CellRank	Gulati et al., 2020	https://pypi.org/project/cellrank/1.5.1
edgeR	Bioconductor; Robinson et al., 2010	https://bioconductor.org/packages/3.12/bioc/html/edgeR.html (v3.31.4)
Glimma	Bioconductor; Su et al., 2017	https://bioconductor.org/packages/3.13/bioc/html/Glimma.html (v2.2.0)
pyGAM	Servén et al., 2018	https://pypi.org/project/pygam/0.8.0
scipy	Virtanen et al., 2020	https://pypi.org/project/scipy/1.7.1
connectomeDB2020	Human Connectome Project	https://www.humanconnectome.org/software/connectomedb
gProfiler	Raudvere et al., 2019	https://biit.cs.ut.ee/gprofiler_archive3/e102_eg49_p15/gost
TransportDB	Ren et al., 2007	http://www.membranetransport.org/transportDB2/index.html
fgsea	Bioconductor; Korotkevich et al., 2016	https://bioconductor.org/packages/3.13/bioc/html/fgsea.html (v1.17.1)
SpatialLIBD	Maynard et al., 2021	https://bioconductor.org/packages/release/data/experiment/html/spatialLIBD.html (v1.0)
Symphony	Kang et al., 2021)	https://CRAN.R-project.org/package=symphony (v1.0)
DisGeNET	Piñero et al., 2020	https://www.disgenet.org/ (v7)
disgenet2r	Piñero et al., 2020	https://bitbucket.org/ibi_group/disgenet2r/src/master/r (v0.99.2)
scvi-tools	Lopez et al., 2018	https://pypi.org/project/scvi-tools/0.11.0
reticulate	RStudio, Tomasz Kalinowski	https://github.com/rstudio/reticulate/tree/release/1.20
UMAP via uwot	McInnes et al., 2018	https://cran.r-project.org/src/contrib/Archive/uwot (v0.1.10)
Seurat	Hao et al., 2021	https://github.com/satijalab/seurat/releases/tag/v4.0.3
FNN	CRAN	https://CRAN.R-project.org/package=FNN (v1.1.3)
scran	Bioconductor; Lun et al., 2016	https://bioconductor.org/packages/3.12/bioc/html/scran.html (v1.18.5)
ArchR	Granja et al., 2021	https://github.com/GreenleafLab/ArchR/tree/release_1.0.1

(Continued on next page)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
MACS2	Zhang et al., 2008	https://pypi.org/project/MACS2/2.2.7.1
IDR	Li et al., 2011	https://anaconda.org/bioconda/idr (v2.0.4)
sklearn	Buitinck et al., 2013	https://pypi.org/project/scikit-learn/0.24.0
JASPAR TF motifs	Khan et al., 2018	https://jaspar2018.genereg.net/
TF motif cluster	Vierstra et al., 2020	https://www.vierstra.org/resources/motif_clustering (v1.0)
igraph	CRAN	https://CRAN.R-project.org/package=igraph (v1.2.6)
CellRank	Lange et al., 2020	https://pypi.org/project/cellrank/1.5.1
RRHO2	Bioconductor; Cahill et al., 2018	https://www.bioconductor.org/packages/2.13/bioc/html/RRHO.html

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Ryan Lister (ryan.lister@uwa.edu.au).

Materials availability
This study did not generate new unique reagents.

Data and code availability
- Single-nuclei RNA-seq and single-nuclei ATAC-seq data have been deposited at GEO under accession number GEO: GSE168408 and are publicly available as of the date of publication. Accompanying interactive browsers for this study are available at http://brain.listerlab.org. All data for the Velmeshev et al. study is publicly available through the Sequence Read Archive, accession number PRJNA434002, and all data for Li et al. (2018) is publicly available through http://development.psychencode.org.
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact (RL) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines
Commercial human embryonic stem cell line H9 was purchased from WiCell and authenticated by the provider. H9 cells were cultivated in E8 media (Gibco) and passaged using EDTA (Sigma).

Human subjects
De-identified human prefrontal cortex (PFC) samples from neurotypical individuals of various ages from mid-gestation through to adulthood and one individual with ASD were obtained through collaboration with NIH NeuroBioBank in the United States and analyzed under a protocol approved by the University of Western Australia Human Research Ethics Committee (RA/4/20/6394). Samples were standardized as best as possible through brain region and age. Brain regions include Brodmann area 8 (BA8), Brodmann area 9 (BA9), Brodmann area 10 (BA10), and Brodmann area 46 (BA46), corresponding regions of the dorsolateral and medial PFC in the left and right hemispheres (see Table S1B). Age, developmental stage, and sex of each sample is detailed in Table S1B.

Note, it was not possible to source samples between gestational ages (ga) ga24 and ga34, largely corresponding to the 3rd trimester where samples could not be obtained due to major ethical, practical, and regulatory challenges. Nonetheless, we were able to infer broad trends across development, which includes this period. The primary glioblastoma sample was resected from an adult female and was obtained through the Australian Genomics and Clinical Outcomes of High Grade Glioma (AGOG) biobank under protocols approved by the Northern Sydney Local Health District Human Research Ethics Committee (0809-198M) and Sir Charles Gairdner and Osborne Park Health Care Group Human Research Ethics Committee (2008-094), and analyses approved by the University of Western Australia Human Research Ethics Committee (RA/4/20/5615). Organoid experiments and derivation

from established embryonic stem cell lines were approved by the University of Western Australia Human Research Ethics Committee (2021/ET000182).

METHOD DETAILS

Data reporting

This study complies with the MINSEQE guidelines for reporting high throughput sequencing experiments. No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment. No replication study was performed, but key results were validated using complementary techniques. Subjects were excluded when records indicated that the individual suffered from brain disorders or diseases.

Generation of cerebral organoids

Cerebral organoids were generated from H9 embryonic stem (ES) cells following an established protocol (Iefremova et al., 2017). In brief, ES cells were dissociated into single cells and plated in U-shaped ultra low attachment 96-well plates (Corning) at a density of 9,000 cells per well in E8 media (Gibco) with 50 μM Y-27632 (Selleckchem). Media was changed to only E8 2 days later and to a neural induction media another 2 days later when embryoid bodies were transferred to ultra low attachment 24 well plates (Corning). Neural induction media contained DMEM/F12 and Neurobasal (both Gibco) at a 50:50 ratio with 1:200 N2 supplement (R&D Systems), 1:100 Neurobrew supplement (Miltenyi), 1:200 non essential amino acids (Gibco), 1:100 Glutamax (Gibco), 1:100 Pen/Strep (Gibco), 10 μg/ml Heparin (Sigma), 200 nM LDN-193189 (Cayman Chemical), 500 nM A83-01 (Cayman Chemical), and 2 μM IWR-1 (Adooq Bioscience). After 4 days, LDN-193189, A83-01 and IWR-1 were removed from the media, and organoids were embedded in a 20 μl volume of Cultrex 3D RGF BME (R&D Systems) on day 12. Media was switched to Maintenance media contained of DMEM/F12 and Neurobasal 50:50 ratio with 1:200 N2 supplement, 1:100 Neurobrew supplement without vitamin A (Miltenyi), 1:100 Glutamax, 1:100 Pen/Strep, 1:2,500 human Insulin solution (Sigma), and 1:350,000 beta-Mercaptoethanol (Sigma). After 6 days, Neurobrew in the media was switched to the vitamin A containing variant and organoids were moved to an orbital shaker (Instrument Choice) for constant agitation at 75 rpm with media changes every 3 days until analysis.

Histology and imaging

Organoids were fixed in 4% PFA in PBS for 1 h at room temperature (RT) and incubated in 30% sucrose (Sigma) in PBS at 4°C overnight before being frozen at -80°C and embedded in OCT compound (VWR) for cryosectioning at 20 μm. Sections were recovered on Superfrost glass slides (Menzel) and stored at -80°C until processed for immunostaining. Sections were blocked in PBS with 2% FCS (Gibco), 1% BSA, 0.1% Triton X-100, and 0.05% Tween-20 (all Sigma) for 2 h at RT and incubated with primary antibodies overnight at 4°C. After two washes with PBS, secondary antibodies were incubated for 2 h at RT before being counterstained with DAPI (1 μg/ml, Cayman Chemical) after 3 additional washes. Sections were mounted with fluorescence mounting medium (Agilent) and imaged on a Nikon C2+ confocal microscope. Primary antibodies: GFAP (Dako ZO33429-2) used 1:750, MAP-2 (Millipore AB5622) used 1:400. Conjugated antibody: NeuN (Novus Biologicals NBP1-92693APC) used 1:500. Secondary antibodies: anti-mouse (Abcam 150114) used 1:1,000, anti-rabbit (Abcam ab150063) used 1:1,000.

RNAscope *in-situ* hybridization

Flash-frozen brain tissue or organoids were sectioned at 20 μm using a Leica CM3050S cryostat and frozen overnight on SuperFrost Plus glass slides (Thermo Scientific). Sections were fixated with 4% PFA for 15 minutes and further processed following the standard protocol for RNAscope Multiplex Fluorescent Reagent Kit v2 Assay (Advanced Cell Diagnostics), using Opal Polaris 480, Opal 570, and Opal 650 dyes (Akoya Biosciences) with the following RNAscope 2.5 probes (all Advanced Cell Diagnostics): Hs-ADARB2-C3, Hs-ARHGAP26, Hs-BHLHE22-C3, Hs-CEMIP-C2, Hs-FGF13, Hs-KCNH1-C2, Hs-LAMP5-C2, Hs-LHX6, Hs-NDNF-C3, Hs-NOS1-C2, Hs-NPY-O1, Hs-SST-C3, Hs-TACR1-C3, Hs-TAFA4-C1. Sections were counterstained with DAPI, mounted with ProLong Gold Antifade (Thermo Scientific) and imaged on a Nikon C2+ confocal microscope.

Sample preparation

Nuclei isolation for single nucleus RNA-seq (snRNA-seq) using fluorescence-activated nuclei sorting (FANS)

A single nuclei suspension was obtained following an established protocol for nuclei isolation from small amounts of tissue (Denisenko et al., 2020). Briefly, tissue samples were lysed in a 2.5 ml lysis buffer (10 mM Tris-HCl, 1% Nonident P40, 3 mM MgCl₂, and 10 mM NaCl) with 0.2 U/μl RNasin Plus RNase Inhibitor (Promega) for 17 min on ice. After lysis, 2.5 ml of ice-cold PBS was added to the lysis buffer and sample for a final volume of 5 ml. Tissue was homogenized using a Pasteur pipette until no large pieces were visible. The homogenate was then filtered through a 30 μm filter (Miltenyi) and centrifuged at 2,000 x g for 5 minutes at 4°C. Supernatant was removed and the pellet was resuspended in 400 μl wash buffer (1x PBS with 1% BSA) containing DAPI at a 1:10,000 dilution and with 0.2 U/μl RNasin Plus RNase Inhibitor (Promega). DAPI-positive nuclei were isolated by sorting using a BD Influx to obtain a final concentration of 500 - 800 nuclei/μl in wash buffer with 0.2 U/μl RNasin Plus RNase Inhibitor (Promega) (Figure S1B).

Nuclei isolation for snRNA-seq using sucrose gradient isolation

Nuclei from two samples, ga39 and the 14 year old ASD sample, were isolated with a sucrose gradient method. Briefly, tissue samples were lysed in 400 µl of lysis buffer (10 mM Tris-HCl, 1% Nonident P40, 3 mM MgCl2, and 10 mM NaCl) with 0.2 U/µl RNasin Plus RNase Inhibitor (Promega) in 1.5 ml tubes and broken down with a pellet pestle. Tissue was dissociated by passing through a polished silanized Pasteur pipette 3-4 times then incubated on ice for 10 min. Dissociation with a Pasteur pipette was repeated at 5 min and 10 min. After incubation, the dissociated tissue was added to 2.5 ml of wash buffer (1x PBS with 1% BSA) in a 15 ml falcon tube. The sample was then passed through a 30 µm cell strainer (Miltenyi) into a 50 ml falcon tube and centrifuged for 5 minutes at 500 x g at 4°C in a swinging bucket centrifuge. Following centrifugation, the supernatant was removed and the tissue was resuspended in 100 µl of wash buffer for every 30 mg of tissue used. Using only 100 µl of the resuspended sample, 180 µl of 1.8 M sucrose solution (Nuclei PURE kit, Sigma) was added to the sample and homogenized using a P1000 micropipette. 1 ml of 1.3 M sucrose solution (Nuclei PURE kit, Sigma) was placed at the bottom of a 2ml Eppendorf tube and the 280 µl of the nuclei suspension mixed with 1.8M sucrose was slowly layered on top of the 1.3 M sucrose solution. The sucrose gradient was centrifuged for 10 minutes at 3000 x g at 4°C in a swinging bucket centrifuge. After centrifugation, the debris from the top of the sucrose gradient was removed with a Kimwipe wrapped around a spatula. The remaining supernatant was removed with a P1000 micropipette. Nuclei were resuspended in 50 µl of wash buffer. 10 µl of the suspension was stained with DAPI (1:10,000) and used to count the nuclei concentration.

Nuclei isolation for single nucleus assay for Transposase-Accessible chromatin using sequencing (snATAC-seq)

Nuclei were isolated following the sucrose gradient isolation described above for snRNA-seq, with the addition of 0.01% digitonin added to the lysis buffer and incubated for 3 minutes after the final dissociation step.

Library preparation

snRNA-seq

Single nuclei suspensions were loaded on 10X Genomics Chromium Chip B (10x Genomics) to generate single-cell GEMs. Single-nuclei RNA-Seq libraries were prepared using the Chromium Single Cell 3' Reagents kit as per the manufacturer's instructions. The protocol was followed as outlined in the user guide, with the exception of performing 18 cycles for cDNA amplification and a capture target of ~10,000 nuclei per sample. Library size distribution and abundance was assessed with aD1000 ScreenTape (Agilent) and their concentration quantitated on a C1000/CFX96 qPCR system (Bio-Rad) using Luna Universal qPCR mix (NEB) and Truseq PCR primers (Illumina). Libraries were sequenced on either a NextSeq 550 or a NovaSeq 6000 (Illumina) in paired-end mode (read1: 28 cycles, read 2: 91 cycles, index1: 8 cycles) to generate approximately 200 M reads per sample.

snATAC-seq

Single nuclei suspensions were loaded on 10X Genomics Chromium Chip E (10x Genomics) to generate single-cell GEMS. Single-nuclei ATAC-seq libraries were prepared using the Chromium Single Cell ATAC v1 Reagents kit as per the manufacturer's instructions, with a capture target of ~10,000 nuclei per sample. Library size distribution and abundance was assessed with a D5000 ScreenTape (Agilent) and their concentration quantitated on a C1000/CFX96 qPCR system (Bio-Rad) using Luna Universal qPCR mix (NEB) and Truseq PCR primers (Illumina). Libraries were sequenced on a NovaSeq 6000 (Illumina) in paired-end mode (read1: 50 cycles, read2: 49 cycles, index1: 8 cycles, index2: 16 cycles) to generate approximately 250 M reads per sample.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed in R v4.0.5 or Python v3.8.

snRNA-seq analysis

Sequence Alignment and UMI counting

Cellranger (v3.1.0) analysis pipeline, with default parameters, was used to process raw reads. Briefly, Cellranger mkfastq was used to convert raw base call (BCL) files of the snRNA-seq libraries into FASTQ files. A pre-mRNA transcriptome was built using the cellranger mkref command starting with the refdata-cellranger-hg19-3.0.0 transcriptome as per the instructions provided by 10x Genomics. The Cellranger count command was used for alignment, filtering, barcode counting, and unique molecular identifier counting (Tables S1C and S1D).

Nuclei and gene quality control

Filtered count matrices from each sample were combined into a scanpy (v1.6.0) anndata (v0.7.4) object for processing and filtering, and unless noted otherwise, default parameters were used (Wolf et al., 2018). Genes not observed in ≥5 nuclei across all batches were removed and potential nuclei doublets were removed using scrublet (v0.2.2) with 10 principal components for each batch individually (Wolock et al., 2019). On a per sample basis, nuclei with gene counts <3 median absolute deviations (MADs) and <300 gene counts were removed. Nuclei with ribosomal or mitochondrial gene count percentages >20 percent or >3 per sample MADs were removed (Table S1C). Highly expressed *MALAT1* gene and ribosomal and mitochondrial genes were removed. Finally, we investigated the relationship between sample quality and PMI and found them to be uncorrelated (Figure S1F).

Data integration for dimension reduction and clustering

A developmental reference was integrated by downsampling to 1,000 Unique Molecular Identifier counts (UMIs) per nucleus to remove sampling bias from differing sequencing depths between batches. Downsampling was carried out by randomly sampling,

without replacement, 1,000 UMIs from the total UMIs for a given nucleus, and nuclei with less than 1,000 total UMIs were removed. The 1,000 UMIs cutoff was selected to limit the number of nuclei removed and to include enough counts per nuclei to be informative, which was previously shown to be at 1,000 UMIs when the accuracy of cell type classification plateaus (Heimberg et al., 2016). Analysis was also performed with nuclei with less than 1,000 UMIs included. When included, these nuclei cluster together, exhibit no apparent age-related progression, and resemble a spectrum of cell states from a mixture of cell types (Figure S1E), making them difficult to categorize for further analysis and warranting removal. Downsampled data was used for dimension reduction and clustering, while the full count matrix was used for all other analyses unless noted otherwise.

Post downsampling, the count matrix was scaled to Counts Per Million (CPM), natural log plus one transformed, and matrices from all batches were combined. Five thousand Highly Variable Genes (HVGs) were selected, and data dimensions were reduced via principal component analysis (PCA) to components explaining 50% of the variance. A neighborhood graph was constructed using scanpy preprocessing neighbors function with 25 neighbors on the reduced components, and a 2-dimensional Uniform Manifold Approximation and Projection (UMAP) embedding was generated (McInnes et al., 2018). Subclusters were generated via scanpy Leiden clustering at a resolution of 7.0 with the neighborhood graph serving as the basis (Traag et al., 2019). At the whole-tissue level, 63 Leiden clusters were identified. Inhibitory interneuron (IN) subclustering was performed using the same workflow but with 4,000 HVGs and a Leiden clustering resolution of 5.0. There were 36 IN Leiden clusters identified, replacing 13 IN whole-tissue clusters and bringing the total Leiden cluster count to 86.

Alternative batch correction methods
Harmony (v0.1.0) (Korsunsky et al., 2019) and BBKNN (v1.5.1) (Polański et al., 2020) were also investigated as possible batch corrections. While these methods resulted in integrated batches, the integrations lacked a clear progression of age-related development (Figure S6C). These corrections also fully integrated the poor quality nuclei that our downsampling approach readily differentiated. Harmony and BBKNN were run using default parameters within the scanpy external application programming interface. Except for the downsampling of UMIs and removal of nuclei under 1,000 UMIs, the data was processed using the same pipeline outlined in nuclei and gene quality control and data integration for dimension reduction and clustering.

Cluster annotation
PN were classified based on the expression of marker genes *CUX2, RORB, THEMIS*, and *TLE4*, while IN were classified based on the expression of marker genes *PVALB, SST, VIP*, and *ID2*. Using this approach, we identified 18 major clusters, including 5 PN major clusters (immature principal neurons [*PN-dev*], layer 2 and 3 [*L2/3-CUX2*], layer 4 [*L4-RORB*], and layer 5 and 6 [*L5/6-THEMIS, L5/6-TLE4*]), 6 IN major clusters (early developing MGE [*MGE-dev*], early developing CGE [*CGE-dev*], inhibitor of DNA binding 2-expressing [*ID2*], vasoactive intestinal polypeptide-expressing [*VIP*], somatostatin-expressing [*SST*], and parvalbumin-expressing [*PV*]), astrocytes [*Astro*], MBP+ oligodendrocyte precursor cells and mature oligodendrocytes [*Oligo*], oligodendrocyte precursor cells [*OPCs*], microglia [*Micro*], and vasculature [*Vas*] (which encompasses endothelial, pericytes, fibroblast-like, and pericytes/smooth muscle cells) (Figure 1B). In cases where several subclusters within a major cluster could be grouped based on additional common markers, these were given a common subcluster name. For example, using this strategy, L5/6-THEMIS PNs could be further subclustered into L5/6-THEMIS-NTNG2 and L5/6-THEMIS-CNR1 (Figure S2A). Membership of Leiden subclusters in each major cluster is listed in Tables S2A–S2C.

Principal neuron annotation
Immature principal neurons (PNs). PN-developing (PN dev) (subcluster 13). Marker Genes - *MEIS2* (Bayatti et al., 2008; Hoerder-Suabedissen and Molnár, 2015), *UNC5D* (Cooper, 2014; Miyoshi and Fishell, 2012), *FOXG1* (Cooper, 2014; Miyoshi and Fishell, 2012), *LINC01158* (Fan et al., 2018), *EIF1B* (Bayatti et al., 2008; Hoerder-Suabedissen and Molnár, 2015), *MEF2C* (Cooper, 2014), *DCX* (Herrero-Navarro et al., 2021).

Principal neurons L2-3. Immature. L2/3_CUX2_dev-1 (subcluster 11), L2/3_CUX2_dev-2 (subcluster 15), L2/3_CUX2_dev-3 (subcluster 22), L2/3_CUX2_dev-4 (subcluster 32), L2-3_CUX2_dev-5 (subcluster 36), L2/3_CUX2_dev-6 (subcluster 43), L2/3_CUX2_dev-fetal (subcluster 46). Marker Genes - *SATB2* (Alcamo et al., 2008; Britanova et al., 2008), *FAM19A2* (Fan et al., 2018), *CUX2* (Fan et al., 2018), *FGF13* (Puranam et al., 2015; Wu et al., 2012), *MEF2C* (Leifer et al., 1997), *STMN2* (Colantuoni et al., 2011; Tomita et al., 2012). *Mature.* L2_CUX2_LAMP5 (subcluster 8), L2_CUX2_LAMP5_dev (subcluster 34), L3_CUX2_PRSS12 (subcluster 31). Marker Genes - *CUX2* (Cubelos et al., 2015), *CAMK2A* (Tochigi et al., 2008).

Principal neurons L4. Immature. L4_RORB_dev-1 (subcluster 57), L4_RORB_dev-2 (subcluster 10), L4_RORB-fetal (subcluster 18). Marker Genes – *RORB* (Oishi et al., 2016), *SATB2* (Britanova et al., 2008), *FOXP2* (Ferland et al., 2003), *FOXP1* (Ferland et al., 2003). Mature. L4_RORB_LRRK1 (subcluster 17), L4_RORB_MET (subcluster 28), L4_RORB_MME (subcluster 44). Marker Genes - *FOXP2* (Baker et al., 2018; Brunjes and Osterberg, 2015), *RORB* (Lake et al., 2018).

Principal neurons L5/6 THEMIS. Immature. L5/6_THEMIS_dev-1 (subcluster 21), L5/6_THEMIS_dev-2 (subcluster 25). Marker Genes – *THEMIS* (Hodge et al., 2019), *SATB2* (Britanova et al., 2008), *FOXP1* (Ferland et al., 2003). *Mature. L5/6_THEMIS_CNR1 (subcluster 40), L5/6_THEMIS_NTNG2 (subcluster 55).* Marker Genes - *THEMIS* (Hodge et al., 2019).

Principal neurons L5/6 TLE4. Immature. L5/6_TLE4_dev (subcluster 24). Marker Genes - *SOX5* (Galazo et al., 2016), *TLE4* (Galazo et al., 2016), *FOXP2* (Ferland et al., 2003), *NFIB* (Galazo et al., 2016), *BCL11B* (Britanova et al., 2008), *HS3ST4* (Hodge et al., 2019), *TUBA1A* (Aiken et al., 2017).

Mature. L5/6_TLE4_SCUBE1 (subcluster 26), L5/6_TLE4_SORCS1 (subcluster 27), L5/6_TLE4_HTR2C (subcluster 45).

Marker Genes - *TLE4* (Galazo et al., 2016), *FOXP2* (Galazo et al., 2016), *NFIB* (Galazo et al., 2016), *BCL11B* (Aiken et al., 2017), *HS3ST4* (Lake et al., 2018).

Inhibitory interneuron annotation

IN prefix to Leiden cluster identifiers indicates cluster was identified with GABAergic interneuron clustering described in data integration for dimension reduction and clustering.

Immature GABA interneurons. MGE_dev-1 (subcluster IN6), MGE_dev-2 (subcluster IN23), CGE_dev (subclusters IN8). Marker Genes - *LHX6* (Fan et al., 2018; Hu et al., 2017), *SOX6* (Hu et al., 2017), *DLX1* (Fan et al., 2018; Hu et al., 2017), *DLX2* (Fan et al., 2018; Hu et al., 2017), *DCX* (Herrero-Navarro et al., 2021), *GAD1* (Lake et al., 2018).

GABA PVALB. PV_WFDC2 (subcluster IN0), PV_dev (subcluster IN7), PV_SULF1 (subcluster IN18), PV_SULF1_dev (subcluster IN29). Marker Genes - *LHX6* (Hu et al., 2017), *SOX6* (Hu et al., 2017), *PVALB* (Ferguson and Gao, 2018), *SLIT2* (Andrews et al., 2008).

GABA PVALB SCUBE3. PV_SCUBE3 (subcluster IN21), PV_SCUBE_dev (subcluster IN30). Marker Genes - *SCUBE3* (Hodge et al., 2019).

GABA SST. SST_TH (subcluster IN31), SST_B3GAT2 (subcluster IN10), SST_STK32A (subcluster IN12), SST_BRINP3 (subcluster IN28), SST_NPY (subcluster IN35), SST_ADGRG6 (subcluster IN5), SST_ADGRG6_dev (subcluster IN15), SST_CALB1 (subcluster IN14), SST_CALB1_dev (subcluster IN12), PV_SST (subcluster IN27). Marker Genes - *SST* (Tasic et al., 2018), *LHX6*, *SOX6* (Hu et al., 2017), *ERBB4* (Vullhorst et al., 2009), *RELN* (Lake et al., 2018).

GABA ID2. LAMP5_CCK (subcluster IN4), LAMP5_NDNF (subcluster IN11), CCK_SORCS1 (subcluster IN17), ID2_CSMD1 (subcluster IN22), CCK_RELN (subcluster IN26), CCK_SYT6 (subcluster IN32), ID2_dev (subcluster IN25). Marker Genes - *RELN* (Overstreet-Wadiche and McBain, 2015), *SV2C* (Boldog et al., 2018), *KIT* (Mickelsen et al., 2019), *CXCL14* (Boldog et al., 2018), *LAMP5* (Mickelsen et al., 2019), *CCK* (Lake et al., 2018).

LAMP5_NOS1. LAMP5_NOS1 (subcluster IN13). Marker Genes - *NOS1* (Tricoire et al., 2010), *CA1* (Hodge et al., 2019).

GABA VIP. VIP_HS3ST3A1 (subcluster IN2), VIP_CHRM2 (subcluster IN3), VIP_KIRREL3 (subcluster IN19), VIP_CRH (subcluster IN9), VIP_ADAMTSL1 (subcluster IN20), VIP_ABI3BP (subcluster IN16), VIP_DPP6 (subcluster IN24), VIP_PCDH20 (subcluster 34), VIP_dev (subcluster IN33). Marker Genes - *GAD1* (Lake et al., 2018), *VIP* (Tremblay et al., 2016), *CALB2* (Darmanis et al., 2015).

Glial cell annotation

Astrocytes. *Immature.* Astro_dev-1 (subcluster 2), Astro_SLC1A2_dev (subcluster 14), Astro_dev-2 (subcluster 19), Astro_dev-3 (subcluster 54), Astro_dev-4 (subcluster 56), Astro_dev-5 (subcluster 60). Marker Genes - *MEIS2* (Fan et al., 2018), *LINC01158* (Fan et al., 2018), *EIF1B* (Fan et al., 2018), *SLC1A3* (Batiuk et al., 2020), *SLC1A2* (Jolly et al., 2019).

Mature. Astro_SLC1A2 (subcluster 1), Astro_GFAP (subcluster 16). Marker Genes - *AQP4* (Batiuk et al., 2020), *SLC1A2* (Jolly et al., 2019), *SLC1A3* (Batiuk et al., 2020).

Oligodendrocyte precursor cells (OPCs). OPC (subcluster 0), OPC_dev (subcluster 38). Marker Genes - *PDGFRA* (Rivers et al., 2008), *OLIG1* (Zhou and Anderson, 2002), *OLIG2* (Zhou and Anderson, 2002).

Oligodendrocytes (ODCs). OPC_MBP (cluster 49), Oligo_mat (subcluster 3), Oligo-1 (subcluster 4), Oligo-2 (subcluster 6), Oligo-3 (subcluster 9), Oligo-4 (subcluster 35), Oligo-5 (subcluster 50), Oligo-6 (subcluster 61), Oligo-7 (subcluster 62). Marker Genes - *MAG* (Valério-Gomes et al., 2018), *MOG* (Valério-Gomes et al., 2018), *PLP1* (Turnescu et al., 2018).

Microglia. Micro (subcluster 5), Micro_out (subcluster 53). Marker Genes - *C1QA* (Fonseca et al., 2017).

Vasculature. Vas_TBX18 (cluster 51), Vas_PDGFRB (subcluster 58), Vas_CLDN5 (subcluster 59). Marker Genes – Pericytes and smooth muscle cells: *PDGFRB* (Vanlandewijck et al., 2018), Endothelial cells: *CLDN5* (Vanlandewijck et al., 2018), Fibroblast-like cells: *TBX18* (Vanlandewijck et al., 2018).

Identification of poor quality cells (subcluster 42)

As has been previously noted, we identified a subcluster with high expression of *NRNG* and *THY1* (Velmeshev et al., 2019; Figure S6F). We additionally observed high mitochondrial percentages per nuclei within this subcluster, potentially indicating the cluster contained nuclei of poor quality (Figure S4J). In order to assess the quality of the nuclei in this cluster versus potential contamination from background reads, pseudo-cells were generated by sub-sampling from a pseudo-bulked count matrix. More specifically, each batch's unfiltered count matrix (containing both called and uncalled cells) from 10x Genomics Cell Ranger (v3.1.0) was summed along the gene axis. Pseudo-bulked genes were removed to match genes post preprocessing (see nuclei and gene quality control), and pseudo-bulked counts were randomly sub-sampled, without replacement, to 1,000 UMIs. Quantity and batch origin were selected to replicate cells in the labeled poor-quality subcluster. Pseudo-cells were then incorporated into the pipeline used for whole tissue clustering (see data integration for dimension reduction and clustering), and resulting clusters were checked for overlapping memberships of previously called cell types (Figures S6D and S6E). Using this method, we observed that the previously identified subcluster high in *NRGN* and *THY1* now contained most of the pseudo-cells, indicating the quality of these nuclei is similar to background or ambient RNA. Previous studies have reported this population of nuclei as a new PN population relevant to autism (Velmeshev et al., 2019) and schizophrenia (Ruzicka et al., 2020). However, these studies also reported high expression of mitochondrial genes and similar gene expression profiles as our study, suggesting these nuclei may also represent mostly ambient RNA.

Hotspot analysis of gene coexpression

Hotspot (v0.9.1) is a tool for identifying highly correlated gene modules in a single-cell dataset (DeTomaso and Yosef, 2021). Hotspot computes gene modules by finding informative genes with high local autocorrelation, evaluating the pairwise correlation between genes, and clustering the results in a gene-gene affinity matrix. The Hotspot depth-adjusted negative binomial model was run using

the full count matrix with the 5,000 downsampled HVGs and downsampled principal components explaining 50% of the variation (see data integration for dimension reduction and clustering). The Hotspot computed a weighted K-nearest-neighbors (KNN) graph with 25 neighbors, and 187 genes with an autocorrelation false discovery rate (FDR) greater than 0.05 were removed. Gene modules were created by agglomerative clustering with the minimum number of genes per module set to 65. 14 modules were identified, and 922 genes were not assigned to a module (Table S2D). Hotspot eigengene module scores were calculated by first centering the UMIs using the depth-adjusted negative binomial model. The centered values were then smoothed using the weighted average of their 25 nearest neighbors. These smoothed values were then modeled with PCA using the first principal component, and the cell-loadings reported as the model scores.

Trajectories
Subclusters containing cells from a majority of adult samples were classified as mature, and major and sub-trajectories inferred to give rise to these mature subclusters over development were manually assembled based on age progression of adjacent subclusters and shared expression of markers. For example, distinct layers 2/3 and layer 4 PN sub-trajectories (L2/3-CUX2, L4-RORB) were defined that share cells of early developmental stages but subsequently differentiate into distinct terminal cell types (Figures S2A and S2C; Table S2B). 45 sub-trajectories were collapsed into 15 major trajectories for further analysis based on major clusters.

scVelo velocity analysis
Spliced and unspliced read counts were generated from FASTQ files (see sequence alignment and UMI counting) using loompy (v3.0.6) mapped to hg19-3.0.0 (La Manno et al., 2018). As above (see data integration for dimension reduction and clustering), a neighborhood graph *[n_neighbors=50]* was constructed from the normalized downsampled count matrix on the reduced components. Spliced and unspliced reads were filtered and normalized using scVelo (v0.2.2) (Bergen et al., 2020) filter_and_normalize function *[min_shared_counts=5, n_top_genes=1500 (500 for INs)]*. Velocity scores were calculated using the scVelo stochastic model of transcriptional dynamics. Single nucleus velocities were then projected into UMAP and UMAT embeddings (see UMAP of maturation and data integration for dimension reduction and clustering). Nucleus-to-nucleus transition probabilities were calculated using a CellRank velocity kernel.

CytoTRACE
Cellular Trajectory Reconstruction Analysis using gene Counts and Expression (CytoTRACE, v0.3.3) is a computation framework based on the observation that the number of genes expressed in a cell decreases during differentiation (Gulati et al., 2020). More specifically, the geometric mean of the 200 most correlated genes with gene counts has been shown to predict a nucleus' differentiation state accurately. To calculate the CytoTRACE values, we first KNN imputed our log normalized downsampled data (see data integration for dimension reduction and clustering) with 25 neighbors. We then utilized the CellRank (Lange et al., 2020) implementation of CytoTRACE to calculate the value for each nucleus, where a low CytoTRACE value corresponds to an earlier developmental state and a higher value to a later state. Finally, within a CellRank kernel, CytoTRACE scores and the KNN graph were used to compute directed nucleus-to-nucleus transition probabilities.

Mass flow analysis
A transition matrix representing nucleus-to-nucleus transition probabilities was generated from each CellRank kernel (Lange et al., 2020) (see scVelo velocity analysis and *CytoTRACE*). A transition matrix was used to calculate the probability of mass flows from one cluster to the others at each developmental stage. These mass flows were summarized in vein plots using a CellRank re-implementation of Mittnenzweig et al. (2021), where total flows between clusters at subsequent developmental stages is one. The total frequency of each cluster per stage is represented by veins with changing width proportional to the frequency of that cluster. Outflows between nuclei types are visualized by edges connecting clusters whose width is proportional to flow magnitude. Flows between clusters were simplified by eliminating low-magnitude flows with a threshold of 0.05.

Pseudo-bulked trajectory data
Batches with at least ten nuclei in each major trajectory were pseudo-bulked by summing the UMI counts of all genes. Pseudo-bulked counts were then normalized by the trimmed mean of M-values (TMM) (Robinson and Oshlack, 2010) and scaled to $\log_2$ CPM, and lowly expressed genes were removed with edgeR (v3.31.4) (Robinson et al., 2010) *filterByExpr* function.

Differential expression analysis of genes over development
Differential expression analysis was carried out with the limma-voom pipeline (Law et al., 2014) from the edgeR (Robinson et al., 2010) package following a published workflow (Law et al., 2016). Briefly, using pseudo-bulked trajectory data, potential confounding factors were screened using multi-dimensional scaling (MDS) scatter plots (Figure S4A), where samples that cluster by a given factor suggests the factor contributes to expression differences and is worth including in the analysis and factors that show little or no effect may be left out of the downstream analysis. Of those screened, donor sex, 10x chemistry, and library prep lot number were deemed worth including (Figure S4A). These factors were used along with developmental stages to establish a design matrix, and contrasts were set up for pairwise comparisons between stages. Linear modeling was carried out using limma *lmFit* and *contrasts.fit* functions, and empirical Bayes moderation was carried out to obtain more precise estimates of gene-wise variability. Assumptions for this model were checked using exploratory visualizations, such as biological coefficient of variation plots. Differentially expressed genes (DEGs) were defined as those with an FDR <5% (Table S3). Interactive mean-difference plots (http://brain.listerlab.org) were produced via Glimma (v2.2.0) (Su et al., 2017) *glMDPlot* function. Note that results pertaining to major clusters are presented here to maximize statistical power and facilitate interpretability, but results on all sub-trajectories are also available at http://brain.listerlab.org.

Rank-rank hypergeometric overlap analysis

We conducted a leave-one-out analysis for differential gene expression analysis between different developmental stages for the L2/3-CUX2 major trajectory. We implemented our devDEG strategy, including pseudo-bulking and correction for confounding factors, but removed each sample once. Rank-rank hypergeometric overlap (RRHO2) (Cahill et al., 2018) was used to compare ranked DEGs lists between the two independent gene profiling experiments. Each pairwise comparison resulted in a RRHO2 plot showing the strength and pattern of the correlation between the DEGs found when a sample was left out and the DEGs found by the original analysis.

Fitting and clustering gene trends

Trend curves were fit to all development-associated differentially expressed genes (devDEGs, see differential expression analysis of genes over development), allowing for analysis of continuous expression changes across development, even where age gaps may exist. Rather than the raw data, clustering trend curves produce more coherent clusters and sharper kinetic trends, allowing analysis of a more assorted set of patterns (Trapnell et al., 2014). To fit the trends, we fit a Generalized Additive Model (GAM) to the pseudo-bulked trajectory data using the pyGAM python package (v0.8.0) (Servén et al., 2018). More specifically, GAM was fit with pyGAM's LinearGAM function, which uses an identity link function, a Normal error distribution, and a cubic smoothing function. Once a GAM was fit to a gene, a response curve, or trend fit, was produced using 100 evenly spaced grid points. Gene trends on a stage scale were fit to pseudo-bulked trajectory gene expression binned by developmental stages and set at six even intervals, while trends on an age scale were fit to pseudo-bulked trajectory gene expression scaled to regularize the spacing between ages while maintaining a grounding in time. Age scaling was performed by first dividing ages in years by a cofactor of 0.55 and then arcsinh transformed. The cofactor was chosen to optimally linearize the spacing between samples in our dataset.

Gene trends set on the stage scale were scaled zero to one and hierarchical clustered using scipy (v1.7.1) python package (Virtanen et al., 2020) with a Ward linkage and Euclidean distance metric. The hierarchical tree was cut at a height of 90 using the scipy cut_tree function, resulting in 14 clusters or gene trends. The height of the tree cut was selected to maximum distinctiveness and to limit the redundancy of the gene trends.

snRNA-seq pseudo-bulked trend comparison to Li et al. bulk study

We analyzed an independent bulk RNA-seq dataset consisting of 40 samples from the dorsolateral prefrontal cortex (DFC) spanning 8 PCW to 40 yr, published by Li et al. (2018), restricting the analysis to 28 samples matching our age range, 19 PCW (21 ga) to 40 yr. We limited the gene comparisons to those with either similar or dissimilar expression profiles across the major cell trajectories of our study. We defined genes with similar expression profiles as those occupying the same trend class (i.e. up or down) in enough major trajectories to account for 75% of all nuclei in the same trend class, thereby selecting a set of genes that show similar developmental expression dynamics in a large fraction of all analyzed cell types. Genes with dissimilar profiles were defined as those with >35% of nuclei in both up and down trend classes, thereby selecting a set of genes that have opposite developmental dynamics between groups of cell types. Selected genes were GAM fit to both our pseudo-bulked data and the selected published DFC data (see manuscript methods fitting and clustering gene trends). Both trend sets were normalized zero to one, and the difference measured using euclidean distance.

Genes with similar expression dynamics in our major trajectories have similar expression changes over development when comparing our pseudo-bulked snRNA-seq data to the bulk RNA-seq data (Figure S3F). Genes that demonstrate very different dynamic trends between major trajectories in our snRNA-seq data tend to disagree with the bulk data, as would be expected when comparing pseudo-bulked data to bulk data for genes that show divergent dynamics in different cell types of the analyzed tissue.

Rates of change analysis

The rate of change analysis was performed on a per major trajectory basis by taking the mean of the absolute rate of change devDEG trend fits. Whilst the rate of change analyses on a day scale suggest only small scale changes beyond infancy, disparate time scales (days for early development versus years for older ages) could mask rate changes occurring over larger time intervals. Therefore, we calculated the difference between time points on two scales to highlight changes occurring at different time intervals. In both cases, the grid points from the GAM fit to the arcsinh transformed ages were used. The GAM grid points were left evenly distributed on the arcsinh transformed ages to highlight protracted differences between ages (Figure S3G), and the rate of change was calculated by taking the difference between gene trend values at subsequent grid points. The rate of change on a year scale was calculated by first transforming the arcsinh scaled grid points back to a year scale by transforming with the sinh function and multiplying by the 0.55 cofactor. The rate of change was then calculated by taking the difference between trend expression values at subsequent grid points and dividing by the difference of grid points on the year scale. Similarly, acceleration of expression was calculated by taking the difference in the rate of change values at subsequent grid points and dividing by the difference of grid points on the year scale.

Nearest neighbor acquisition of adult-like identities

KNN graphs were constructed for all nuclei and separately for INs using the scanpy preprocessing neighbors function with 25 neighbors as described in data integration for dimension reduction and clustering. For each sub-trajectory age, the sum of the number of nuclei whose nearest neighbor is an adult nucleus and the number of adult nuclei whose nearest neighbor is within the query age is calculated. The sum of adult neighbors is normalized by sample size, the product of the number of nuclei in the query age and the number of adult nuclei within a sub-trajectory. The cumulative sum of the adult neighbor proportions is taken across the ages of development and scaled from zero to one. PN sub-trajectories in Figure 1F were hierarchically clustered using a Ward linkage. The 40yr sample was omitted from this analysis to avoid any potential detection of aging effects, and sub-trajectories with <30 adult

nuclei were not included. This analysis was robust to the number of nearest neighbors, showing little change in maturity patterns with as few as 15 and as many as 100 nearest neighbors.

Eigentrends

To inspect the expression dynamics of a set of genes linked to brain development and function processes, we calculated the eigentrend of devDEGs enriched in Gene Ontologies (GO) for each major trajectory (see Enrichment). In order to calculate the eigentrends, we first quantile normalized the devDEG stage fit trends (see Fitting and clustering gene trends) from all major trajectories as a combined group. Next, the intersection of devDEGs for each major trajectory and the enriched GO term gene set was taken. For each major trajectory, PCA, without zero-centering, was performed on the quantile normalized trends of the intersected genes with respect to the stage-scaled GAM grid points, and the eigentrend was taken as the first principal component scaled from zero to one. Statistical testing of comparisons between eigentrend values was performed with a one-sided Wilcoxon signed-rank test.

Cell-to-cell communication of ensheathment genes

Starting with the connectomeDB2020 database of 2293 manually curated ligand-receptor pairs with literature support (Hou et al., 2020), we filtered ligand-receptor pairs to those contained in the gene ontology term "neuron ensheathment" (Raudvere et al., 2019). For each major trajectory, only pairs where both the ligand and receptor are DEGs over development were considered. From pseudo-bulked trajectory data, the mean gene expression of batches within a developmental stage was used to calculate the product of expression for each ligand-receptor pair.

Umap of MATuration (UMAT)

To visualize developmental sub-trajectories we overlaid the UMAP with velocity vector fields derived from nucleus-to-nucleus transition probabilities (Bergen et al., 2020). While vector fields largely agree with age progression, some disjunct developmental patterns are visible (Figure S4B), potentially indicating a paucity of some intermediate cell states. To more explicitly compare transition probabilities and development, we introduce Umap of MATuration (UMAT), which limits UMAP neighbor selection to cells from adjacent developmental stages (Figures 4F and S4C).

Scanpy *neighbors* function was used to iteratively populate the connectivity and distance matrices so that the nearest neighbors of a nucleus were limited to nuclei within the same developmental stage or adjacent stages. For example, nuclei in the neonatal stage could only select neighbors from the fetal or infancy stages. Once fully populated, the connectivity and distance matrices were used as the basis for UMAP embedding. The same preprocessing procedures and parameters described in Data integration for dimension reduction and clustering were used within the Umap of MATuration (UMAT) workflow.

Enrichment analyses

Gene set enrichment analysis of GO terms was performed with gProfiler (Raudvere et al., 2019) with default parameters (adjusted p-values <0.05), and results were limited to enrichment in GO molecular function and GO biological processes. Enrichment for Hotspot gene modules (Table S2D) and devDEGs (Table S3) for each major trajectory was done with a background consisting of genes with a least one UMI in the downsampled count matrix, while, for each major trajectory, enrichment of devDEGs within a gene trend was run with a background consisting of all genes expressed within the major trajectory.

For the enrichment analysis across different types of ion transporters, the TransportDB database (Ren et al., 2007) was subsetted to transporters relevant in the brain. Next, for each major gene trend and major trajectory using the function fora in the fgsea package (v1.17.1) we determined whether there was enrichment for a particular ion transporter type (Korotkevich et al., 2016). The background was set to be all devDEGs in the respective trajectory. P-values were then adjusted for multiple testing using FDR and enrichment was deemed present when adjusted p-value <0.05.

Disease enrichment analysis

To investigate enrichment of known neurological and psychiatric disease genes in major trajectory-specific gene trends, gene-disease associations defined in the DisGeNET database (Piñero et al., 2020) were used (v7, accessed via R package disgenet2r v0.99.2). We focused on the DisGeNet database, as it represents one of the largest publicly available and most regularly updated collections of genes associated with human diseases (Piñero et al., 2020). The function fora in the fgsea package was used to calculate a p-value regarding the statistical significance of enrichment over a background set of genes (Korotkevich et al., 2016). The background set of genes was set to include all genes that were expressed in ≥5% of the cells of the relevant cell type. p-values were then adjusted for multiple testing using FDR. Diseases were deemed enriched where the adjusted p-value was smaller than 0.05.

Stable integration into the reference atlas

Integrating query datasets and organoid samples onto the reference dataset was performed using the scArches approach from scVI (scvi-tools v0.11.0) (Lopez et al., 2018), via the R package reticulate (v1.20). Raw count data from the reference was downsampled to a maximum of 1,000 UMIs per cell, as well as for query datasets. Features were then reduced to the highly variable genes in the reference that are also present in the query dataset to be integrated. A model is trained using a variational autoencoder to produce a latent space, distributing cells among 30 latent variables. This trained model informs the distribution of cells in the query dataset into the same low-dimensional space as the reference. Parameters were initialized in accordance with the scArches optimization (use_layer_norm = 'both', use_batch_norm = 'none', encodec_covariates = TRUE, dropout_rate = 0.2, n_layers =2).

Once both reference and query are represented in the same latent space, we first verify the validity of their integration in a screening check called a partial initialisation. The partial initialisation employs the uwot package (v0.1.10) to create a UMAP visualization of the query dataset aligned to the reference, but allows the reference structure to change in shape (McInnes et al., 2018). The latent representations of the query and reference are combined together, and a nearest neighbor graph using 25 neighbors is constructed.

Then, point (nuclei) locations are initialized as follows: reference nuclei are initialized to the UMAP coordinates of the original reference embedding, and query nuclei are randomly distributed over this coordinate range (only the reference nuclei receive a specific initialisation, hence the name 'partial initialisation'). The UMAP is then optimized over 50 epochs to produce a final visualization of the partial initialisation of the query and reference.

The partial initialisation is applied to determine whether query datasets are aligning to the reference, or largely forming their own isolated clusters. Because the later stable integration forces query nuclei onto the reference clusters and can thus result in misclassification of cell types, this preliminary testing is important to confirm that integrations are valid and sensible. A qualitative approach can be taken, where query datasets observed to form their own clusters in the partial initialisation are not considered for stable integration, as their expression is too different to align to any reference cluster. However, understanding the context of the query samples is also important, as disease and cerebral organoid models may not completely recapitulate neurotypical brain nuclei expression and thus not align well, but will elicit meaningful results from a subsequent stable integration. Furthermore, the Partition-based graphical abstraction algorithm (PAGA) from the scanpy package can be used to calculate the strength of connections between partitions of a dataset, comparing observed interedges between expected number of interedges assuming random connections (Wolf et al., 2019). Leiden clusters are created to form partitions of the partial initialisation. Few strong connections (weight >0.15) between query-dominated and reference-dominated clusters suggest poor integration that may be inappropriately forced onto the reference, whereas abundant and diverse connections suggest a well-aligned integration. This method demonstrated that our organoid samples have abundant connections to astrocyte and neuronal clusters in the reference, regions they integrate onto heavily in the reference. In contrast, very few connections were inferred between non-neuroectoderm snRNA-seq datasets and our reference map, and corresponded to poor-quality nuclei, microglia, and vasculature, in line with expectations of minimal similarity between these datasets (Figure S7F). The glioblastoma sample shows a connectivity level between non-neuroectoderm samples and organoids, with few connections to vasculature and astrocyte populations (data not shown). As glioblastoma represents a non-neurotypical landscape, its lack of alignment during the partial initialisation step is understandable; and further, it reinforces our assignment of nuclei to be astrocyte-like, rather than astrocytic cells.

Once we have confirmed the validity of the partial initialisation, the uwot package can be utilized to regain the UMAP visualization of our reference (McInnes et al., 2018). A nearest neighbor graph using 25 neighbors is constructed from the latent variable representation of the reference. Point (cell) locations are initialized to the UMAP coordinates of the original reference embedding, with zero epochs to prevent changes to the structure. The query data points are then fitted to the reference using the umap_transform function. Nearest neighbors of the latent representations of query cells are located in the reference, which informs initial positioning of the query points onto the UMAP axes. Points are then fitted to the reference embedding through 50 epochs to approach their final location on the visualization. We also applied two similar methods. This included a method by Hao et al. (2021) utilizing canonical correlation analysis (CCA) as available in Seurat (v4.0.3), and Symphony (v1.0) by Kang et al. (2021), which incorporates soft cluster assignment. These algorithms receive the initial PCA of the reference dataset to inform their integration, and thus create dimensional embeddings equal to the number of principal components leveraged (365). For direct comparison, the Velmeshev dataset was integrated with the scArches parameterisation, but with 365 latent variables instead of 30. Comparison to both methods in terms of cell type prediction accuracy showed comparable performance (Figures S7B and S7C), but substantial improvement in age prediction, primarily using astrocytes (average Spearman correlation improvement = 0.52 for Velmeshev et al. samples). Note that one of the additional samples used to test the integration method (ga39) used a sucrose nuclei isolation method.

Age and cell type prediction

Predicting the cell type from integrated datasets involves KNN cluster prediction with 10 nearest neighbors within the integrated UMAP space. The KNN function of the FNN package (v1.1.3) facilitated this. Query cells were assigned the cell type most common among their 10 nearest neighbors in the reference dataset. For samples that supplied original cell annotations, these were used to measure the accuracy of the cluster prediction, indicating the accuracy of the integration. Accuracy was given as the percentage of cells in the dataset that had their cell type correctly predicted, relative to their original annotation. For age prediction, a KNN regression was performed using FNN's knn.reg function, with 10 nearest neighbors. Cell ages were estimated from the arcsinh transformation of the reference ages, taking the mean of its neighbors. An estimation of sample age was calculated by taking the largest mode of the density of age distributions for that sample. If the density had two similarly large modes, both were considered. Alternative predictions using k=50, 70, and 200 had little effect on age and cell type predictions. Finally, we also tested whether organoids were exhibiting higher levels of metabolic stress than postmortem brain samples that could interfere with our ability to predict cell types. Inspection of expression levels of genes involved in metabolic stress response and glycolysis (data not shown) demonstrated this was not the case, and is in line with recently reported results (Uzquiano et al., 2022).

Differential expression comparing immature and mature PNs in organoids

Using the R package scran (v1.18.5) (Lun et al., 2016), we identified DEGs between PNs predicted to be of fetal age (immature PNs) and PNs predicted to be non-fetal (mature PNs) using a t-test with a blocking effect indicating the nucleus organoid of origin. P-values were corrected for multiple testing using FDR. Additionally, a permutation analysis was performed to confirm that the results could not be produced when randomly shuffling the mature and immature labeling.

snATAC-seq analysis

Sequence alignment and UMI counting

Reads were demultiplexed by sample index using the cellranger-atac (v1.2.0) mkfastq command. Fastq files were aligned to the human (hg19) genome, cell barcodes were demultiplexed, and UMIs corresponding to genes were counted using the cellranger-atac count command and default parameters. Some samples were excluded based on indicators of low quality, such as high doublet score or low fraction of fragments overlapping called peaks (Table S1F).

ArchR preprocessing

ArchR (v1.0.1) is an R package for processing and analyzing snATAC-seq data (Granja et al., 2021). Using the function createArrow-Files the fragments files produced by cellranger-atac were loaded into R, excluding sex chromosomes and restricting the output to contain only barcodes for valid cells as defined by cellranger-atac. ArchR then was used for doublet removal, resulting in the removal of 6.69% of nuclei on average (Table S1E). Next, we excluded any nuclei that did not have sufficient TSS enrichment in a sample adaptive fashion by excluding any nuclei with TSS enrichment lower than median TSS enrichment in the sample minus 1 standard deviation. This resulted in the removal of 6,011 nuclei across all samples (Table S1E).

Dimension reduction and clustering

In order to cluster the data, features were transformed using term frequency that has been depth normalized to a constant (10,000) followed by normalization with the inverse document frequency and then log transformation (log(TF-IDF)). Next, iterative latent semantic indexing available in ArchR using the 50,000 most variable features and 20 dimensions was applied. Four iterations were performed, each time sampling 10,000 cells with 10 random starts at a resolution of 4. To combat batch effects Harmony was applied to the dataset. Using the Harmony-corrected dimension reduction, the data were clustered into 73 subclusters (Korsunsky et al., 2019). This constitutes overclustering, which was addressed during the annotation step.

Cluster annotation

The annotation of subclusters was driven by the snRNA-seq data. To this end, canonical correlation analysis (CCA) was used to match each nuclei in the snATAC-seq data to its closest neighbor in the snRNA-seq data (Butler et al., 2018). Note that for this analysis the union of the 2,000 most variable genes in each modality. Subcluster labels were transferred from the snRNA-seq nuclei to their matched snATAC-seq nuclei. Subclusters were then annotated according to their most prevalent major cluster label as well as by manual inspection of known marker genes (as previously described in *Cluster Annotation for snRNA-seq*). At this stage, several clusters were removed due to high doublet score (Figure S5A). Further clusters that could not be confidently assigned to one cell type and probably constitute doublets or poor quality nuclei were identified. All these subclusters were labeled "Poor Quality" (constituting 4,281 nuclei) and removed from further analysis.

Using the reduced dataset, dimension reduction, clustering and annotation were repeated. This resulted in 67 subclusters, which were labeled as 12 major cell types: Astrocytes, Oligodendrocytes, Oligodendrocyte Precursor Cells, Microglia, Vasculature (which encompass endothelial, fibroblast-like, and pericytes/smooth muscle cells), L2/3, L4, L5/6, PN developing (PN dev), IN developing (IN dev), MGE derived (MGE der, includes SST and PV populations), CGE derived (CGE der, includes VIP and ID2 populations).

Peak calling

Major cell types were first combined into the following trajectories, which match major trajectories used in the snRNA-seq analysis: Oligodendrocytes (Oligodendrocyte Precursor Cells, Oligodendrocytes), Astrocytes, Microglia, Vasculature, L2/3 (PN dev, L2/3), L4 (PN dev, L4), L5/6 (PN dev L5/6), MGE der (IN dev, MGE der), CGE der (IN dev, CGE der). Peaks were called using a modified ENCODE ATAC-seq pipeline with default parameters (Trevino et al., 2021). Fragments were first combined from nuclei according to their sample of origin and trajectory assignment. Additionally, fragments from nuclei were combined according to their trajectory assignment, ignoring their sample of origin. For each of these newly created fragment files, two pseudo replicate fragment files were created, one containing fragments from randomly sampled half the nuclei and the second containing fragments from the other half. Note that fragment files with <40 nuclei were removed from this analysis. Furthermore, sex chromosomes, the mitochondrial chromosome, and blacklisted regions were removed. MACS2 (v2.2.7.1) was then used to call peaks in both the pseudo replicate fragment files as well as the complete fragment files (Zhang et al., 2008). Peaks were filtered out that did not meet the significance threshold of 0.05 as well as an Irreproducible Discovery Rate (IDR, v2.0.4) of 0.05, which was calculated with the help of the pseudo replicate fragment files (Li et al., 2011). This was done in order to select for a more consistent and confident peak set. Next, the function nonOverlappingGR in ArchR was used to obtain peak sets that were non-overlapping and prevent daisy-chaining.

Next, to obtain even more reliable peak sets that ensure decent replication across samples, the following procedure was applied. For each sample and cell type peak set, all peaks were removed that did not show a 50% overlap with peaks in the peak set called for the corresponding trajectory. Additionally, peaks that were smaller than 300 bp or larger than 30 kb were filtered out. All peaks had to show in at least two samples Insertions Per Kilobase per Million reads mapped (IPKM) values above 2. This resulted in finding 152,329 peaks for L2/3, 181,948 peaks for L4, 128,135 peaks for L5/6, 74,202 for MGE-der, 80,411 for CGE-der, 72,263 for Oligodendrocytes, 45,453 for Microglia, 75,411 for Astrocytes, and 16,385 peaks for Vasculature.

For each trajectory, the mean of IPKM values in each peak in a stage was also used to find the first PC. This analysis is robust with regards to the different number of nuclei across trajectories due to pseudo-bulking.

ChromHMM overlap and peak annotation

Identified peaks were intersected with Encode chromHMM annotations specific for humans (ENCODE Project Consortium et al., 2020) using a variety of different tissues and developmental stages (Sarropoulos et al., 2021). The fraction of identified peaks that overlapped Encode-annotated enhancers ("E"), promoters ("P"), and heterochromatin ("H") elements by at least one base pair was determined.

Identification of Cis Regulatory Elements (CREs)

Using a correlation-based approach, peak-to-gene links were identified by applying pseudo-bulking of counts from nuclei of matched scATAC-seq and snRNA-seq (Di Bella et al., 2021; Trevino et al., 2020). This approach has recently been validated using single multiome technology that measures ATAC and RNA from the same cell (Trevino et al., 2021). To this end, 400 nuclei from the entire scATAC-seq dataset were randomly sampled using geosketch to preserve rarer cell types. Using the R package FNN, these 400 seed nuclei were combined with their respective 50 nearest neighbor nuclei (PCA space), such that each pseudo-bulk sample comprised 51 cells in total. Pseudobulk ATAC insertion counts for peaks were obtained by summing peak insertion counts across the respective single-nuclei members. Matching RNA nuclei were obtained by selecting the 50 nearest neighbors of scRNA nuclei matched to seed nuclei as determined by Seurat's FindIntegrationAnchors and IntegrateData (Butler et al., 2018). Similar to the ATAC nuclei groups, the counts for each group of 51 scRNA nuclei were summed. Each matched pseudo-bulk sample was annotated with the majority cluster and age assignments of its contingent ATAC nuclei respectively.

Candidate peak-gene pairs were obtained by associating peaks with a genomic distance between 1 and 250 kb to the TSS of the respective gene. For each candidate peak-gene pair the Pearson correlation coefficient of CPM-normalized counts of accessibility and gene expression data as well as adjusted p-values for these coefficients based on their t-statistic were computed. To further ensure selection of real CREs, we additionally determined a cut-off for the Pearson correlation coefficient value. For this, a previously described method (Sarropoulos et al., 2021) based on computing interchromosomal correlations to obtain an empirical null distribution and identify a biologically meaningful correlation cut-off was adapted. For each gene, 10 randomly chosen regions located on chromosomes different to the gene were considered. We then proceeded as described above to find the Pearson correlation coefficient. Using this data, we then identified the 5% threshold (two-tailed) as the cut-off (Figure S5J). Regions that passed both FDR and Pearson correlation coefficient cut-off were considered CRE. This resulted in the identification of 304,741 trajectory-specific CREs (Table S5A).

Clustering CREs

Nonnegative Matrix Factorization (NMF) implemented via Python package sklearn (v0.24) was used to group CREs into groups based on their CPM value across pseudo-bulk samples (Li et al., 2020). Integral to a successful decomposition of a matrix into a basis and coefficient matrix is the choice of the rank. The rank was optimized with the help of two measurements, sparseness and entropy, with the idea that an optimal solution would show sparsity in both resulting matrices and low entropy in the basis matrix. Average values were calculated from 10 times for NMF runs at each given rank with random seed, which will ensure the measurements are stable. Rank 31 produced the best combination of sparseness and low entropy.

Next, the normalized coefficient matrix was used to associate groups with distinct pseudo-bulk samples. Since the values in the normalized matrix indicate weight of pseudo-bulk samples in a group, we assigned each pseudo-bulk sample to its respective maximally weighted group. In addition, each group was associated with CREs using the basis matrix. For each CRE and each group, the basis coefficient score and feature score via the "Kim method" (Kim and Park, 2007), which give an indication of distinctness of association between a specific CRE and cluster, were derived. CREs that had a feature score smaller than the median feature score plus 3 standard deviations or a basis coefficient score smaller than its median over the whole matrix were filtered. This resulted in 1,968 CREs that were distinctly associated with a small number of groups.

For plotting, the CPM value of CREs associated with the same gene were combined. Both CRE accessibility and associated gene expression were normalized by their row mean.

Plotting CREs trends

To identify CRE accessibility trends over development, gene trends identified from the snRNA-seq data were linked to the CRE via the devDEG. The mean IPKM of the number of Tn5 insertions across all samples for each stage was then determined. To visualize whether CREs show similar gene trends, the IPKM value across stages of each CRE was normalized to a scale from 0 to 1 and then averaged over all CREs belonging to the same gene trend.

Motif enrichment

Motif enrichment was performed using the JASPAR2018 catalog (Khan et al., 2018). Promoter regions were removed from this analysis in order to focus on enhancer regulation. Utilizing the ArchR addMotifAnnotations function, peaks were annotated with known motifs. Appropriate GC matched background peaks for each peak were next determined using getBgdPeak. Using the internal ArchR computeEnrichment function then allowed determination of enriched motifs in chosen sets of regions over their respective background peaks. TFs not expressed as determined via the snRNA-seq data were removed. If multiple sets of regions were tested, FDR corrections to account for multiple testing were used. Finally, motifs for the gene trend analysis were grouped into larger motif families based on their sequence similarity as described in Vierstra et al. (2020). P-values were combined using the Fisher method.

Footprinting analysis

To obtain Tn5 bias-corrected footprints, the ArchR function getFootprints is used. Pseudo-bulk footprints are created for the CREs regions with enriched TF motifs associated with particular gene trends across major trajectories.

Construction of networks

To construct networks we adapted an approach from Kamimoto et al. (2020). The approach has the following steps: 1) For every trajectory, motif binding sites located in the CREs were identified and used to establish potential connections between TFs and the genes associated with the CREs. These connections between target genes and TFs serve as the base networks, which are the input for the machine learning process conducted with the snRNA-seq data in the next step. 2) A relatively simple machine learning model was fitted that predicts a target gene expression based on the gene expression of the potential regulatory TFs as determined by the base network. Note that only cells of a particular cell type of interest are used in order to ensure that the described process represents indeed a linear relationship and is in fact cell type-specific. The linear models were fitted with a Bayesian Ridge model using a non-informative prior via sklearn. 3) Utilizing the ability to easily calculate p-values for the coefficients associated with the TFs with the help of the posterior, we were able to establish connections between TFs and their targets by removing non-significant connections. 4) The resulting network could then be interrogated for the relative importance of each member via the hub score and betweenness measure (igraph v1.2.6).

Ranking of TFs

To rank TFs that were upregulated in a differential expression analysis between mature PNs in organoids and their closest matched nuclei in the reference brain atlas, we employed a strategy based on the inferred base network described in *Construction of networks*. For each TF, we counted the number of potential connections with other upregulated non-TF genes. We then transformed this number into a z-score to rank the TFs according to their likelihood of regulating genes that are differentially expressed.

ADDITIONAL RESOURCES

An accompanying website for this study provides all code as well as interactive browser tools for interrogation of multiple datasets, available at http://brain.listerlab.org.

Supplemental figures

A Previous human single cell PFC samples (as of May 16th, 2022)

B Nuclei/debris separation — Singlet/multiplet separation

C Nuclei counts per sample

D Number of genes detected per nucleus — Number of UMI counts per nucleus

E Leiden clustering before removal of low UMI nuclei — UMI counts per nuclei

F

G UMAP 86 subclusters

H Nuclei counts per stage

I Cell types / states per sample
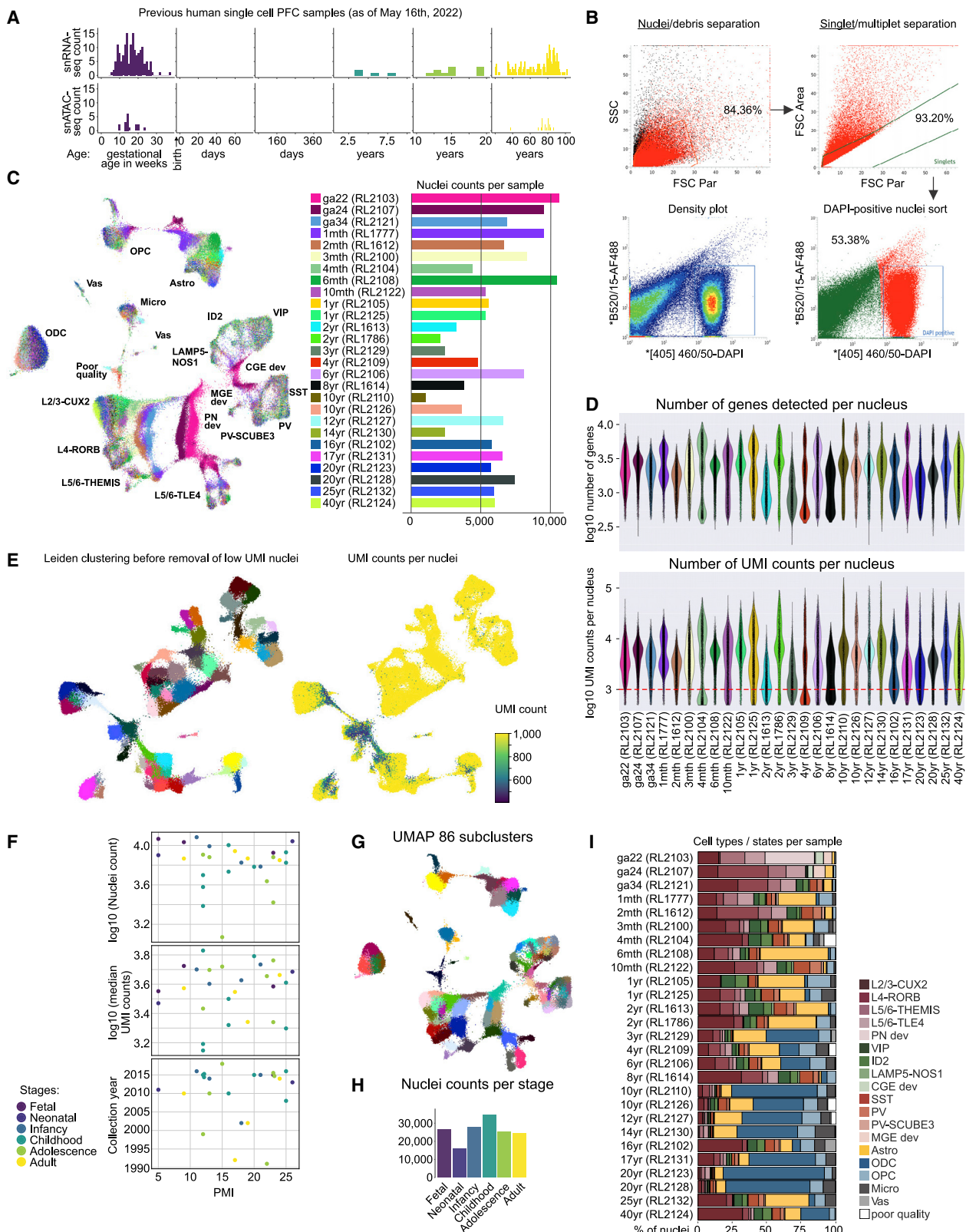
(legend on next page)

Figure S1. Construction of a snRNA-seq map of the human PFC during development, related to Figure 1

(A) Number of human PFC sc/snRNA-seq and sc/snATAC-seq samples generated in previous studies and the sample ages. Brain samples not from the PFC were included for fetal stages. ga, gestational age, in weeks.

(B) Nuclei were stained with DAPI and gated to discriminate intact single nuclei using area plots of forward scatter (FSC Par) versus side scatter (SSC). Nuclei were sorted for DAPI+ events.

(C) UMAP of snRNA-seq data overlaid with donor origin for each nucleus (left). Bar plot of number of single nuclei transcriptome profiles for each sample after stringent quality filtering (right).

(D) Distribution of number of genes detected per nucleus (log_{10} transformed, top) and number of UMI counts per nucleus (log_{10} transformed, bottom), for each sample prior to stringent quality filtering. Dotted line on plot of UMI counts per nucleus indicates threshold of the minimum number of UMIs for a nucleus to be included in subsequent analyses.

(E) UMAPs, Leiden clustering (left) and UMI counts (right), of downsampled snRNA-seq prior to removal of nuclei with <1,000 UMIs (STAR Methods).

(F) PMI and common quality measures (nuclei counts, UMI counts, and collection year) for all snRNA-seq samples, colored by stage.

(G) UMAP of snRNA-seq data with each of the 86 clusters represented by a distinct color.

(H) Number of nuclei per stage.

(I) Proportion of each major cell type/state in each sample.

A

C

Developmental major trajectories

B

Developmental sub-trajectories

(legend on next page)
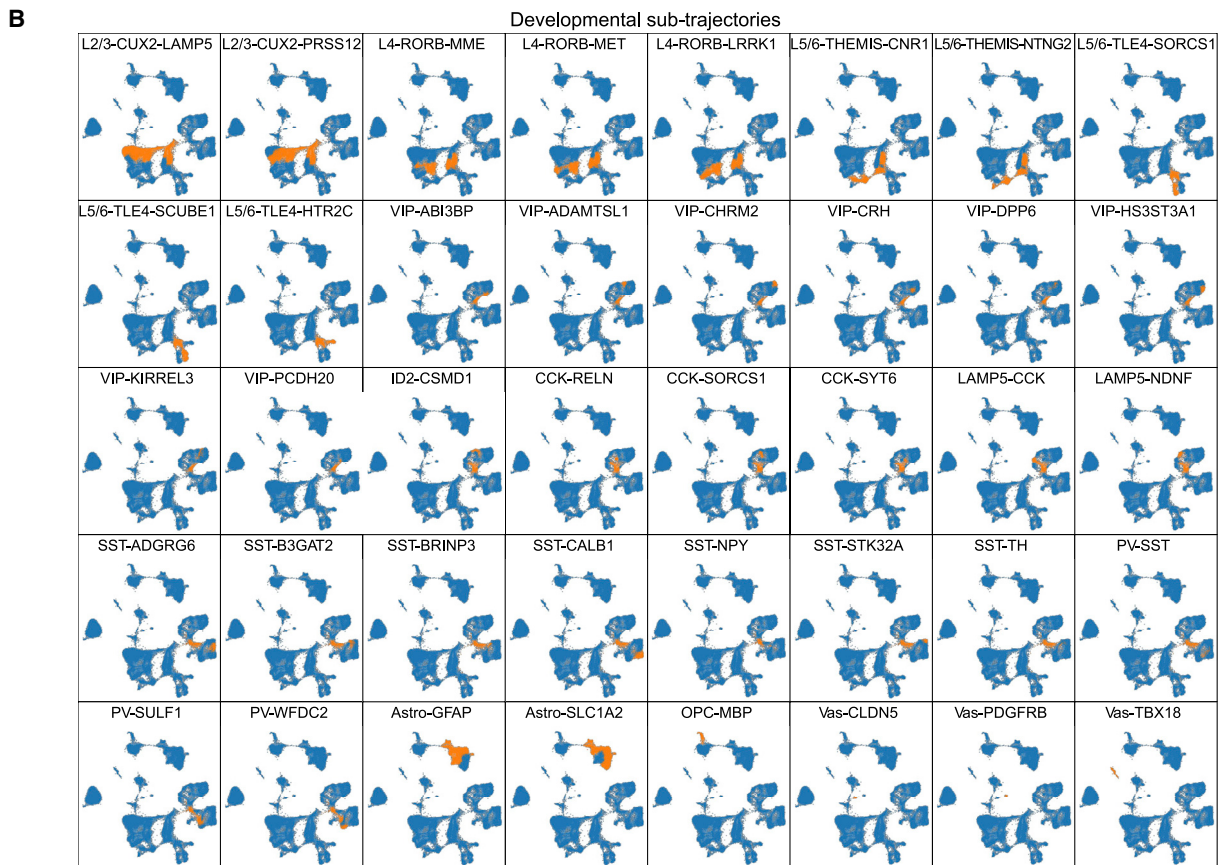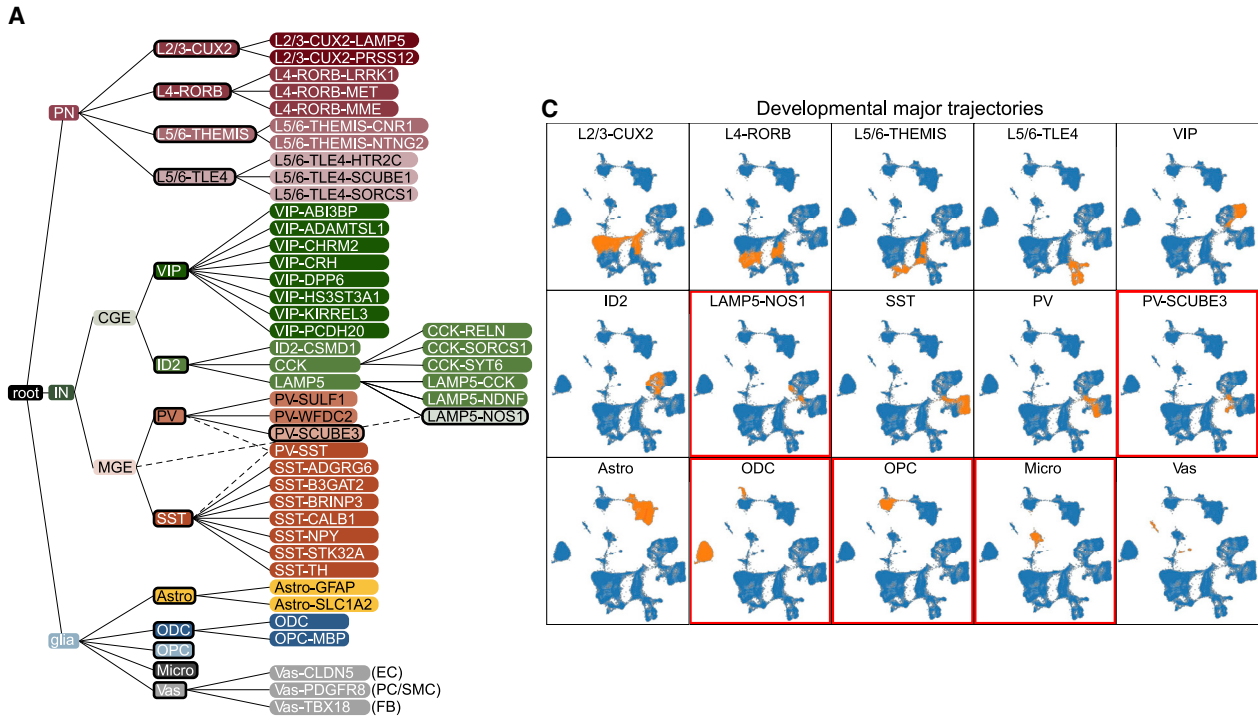
Figure S2. Identification of developmental trajectories in the snRNA-seq human PFC map, related to Figures 1 and 2

(A) Hierarchical cluster naming strategy for all nuclei. Bold outlines indicate major trajectories.

(B) UMAP representation of 40 of the 45 identified sub-trajectories, with nuclei that are members of each trajectory indicated in orange. The 5 remaining major trajectories are indicated by a red outline in (B).

(C) UMAP representation of 15 major trajectories, including vasculature-associated cells, with nuclei that are members of each trajectory indicated in orange.

Figure S3. Robustness of trajectories and analyses, related to Figures 1, 2, and 3

(A) UMAP overlaid with CytoTRACE scores (left) and vector fields derived from scVelo stochastic model (right).

(B) Number of nuclei versus number of devDEGs in each trajectory.

(C) RRHO2 plots comparing L2/3-CUX2 trajectory devDEGs for every stage contrast from the original analysis with the one where the indicated sample in the empty square was removed. Each such pairwise comparison results in a plot showing the strength and pattern of the correlation between the DEGs found when a sample was left out and the DEGs found by the original analysis. Negative $\log_{10}$ p values represent the correlation strength, and low p values in the upper right and lower left quadrants represent concordant up and down gene regulation, respectively.

(D) RNAscope in PFC tissue for six further representative ages (for additional ages see Figure 2D), detecting: postnatally mature cells marked by *KCNH1* (green) and *ARHGAP26* (red), and immature cells marked by *BHLHE22* (gray). DAPI counterstain (blue, nuclei).

(E) Bar plot of proportion of PNs with non-zero expression with a line plot of average PN $\log_2$ CPM expression (left) and UMAP overlaid with $\log_2$ CPM expression (right) of new (*BHLHE22*, *ARHGAP26*, *KCNH1*) and conventional (*SOX2*, *RBFOX3*/NeuN) markers.

(F) Euclidean distances between scaled gene expression trends of pseudo-bulked snRNA-seq data from this study and dorsolateral PFC (DFC) bulk RNA-seq from Li et al. (2018), for devDEGS identified in this study that exhibit either similar or dissimilar expression profiles across different PFC cell types (see STAR Methods).

(G) Absolute mean rates of change of devDEGs calculated on arcsinh age scale.

(H) Selected ligand-receptor pair expression (product of mean ligand expression × mean receptor expression) between PNs and ODCs across developmental stages.

(I) Expression ($\log_2$ CPM; dot: individual sample; line: GAM fit) of receptor *PTPRD* and corresponding ligand *NTRK3* across development. Points represent expression levels in individuals with ≥ 10 nuclei in the trajectory.

(J) Enrichment, indicated by p value and number of intersecting genes, of investigated ion transporter types (calcium [Ca], potassium [K], sodium [Na], neurotransmitter [NT]) in general gene trends in cell types where significant enrichments were detected.

See also Tables S2 and S4.

A

Brain region

Cause of death

Chemistry

Collection year

Donor sex

Ethnicity

Library prep lot number

Post-mortem interval (PMI)

B scVelo

C

D cytoTRACE

E Outflow cytoTRACE

F Ion transport

UMAP

UMAT

Clusters

Sample age

ga22 ga38 2mth 1yr 6yr 12yr 20yr 40yr

MGE dev   LAMP5-NOS1
CGE dev   LAMP5-CCK

CGE dev
MGE dev

Fetal Neonatal Infancy Childhood Adolescence Adult

Fetal Neonatal Infancy Childhood Adolescence Adult

LAMP5-NOS1
LAMP5-CCK
LAMP5-NDNF

Eigentrend value
0   1

H *LAMP5* *SST* *NMBR* *NDNF* *NPY*

log2 CPM

G

Population LAMP5-NDNF

8yr   DAPI   2yr   DAPI

NPY   NPY

NOS1   NOS1

NDNF   NDNF

100 µm   20 µm   100 µm   20 µm

I

LAMP5
SST
NMBR
NDNF
NPY

MGE dev
CGE dev
VIP dev
PV-SCUBE3
PV dev
PV-SULF1
PV-WFDC2
SST-NPY2
PV-SST
SST-CALB1
SST-BRINP3
SST-TH
SST-ADGRG6
SST-B3GAT2
SST-STK32A
ID2-CSMD1
LAMP5-CCK
LAMP5-NOS1
CCK-RELN
CCK-SYT6
LAMP5-NDNF
CCK-SORCS1
ID2-CSMD1
VIP-ADAMTSL1
VIP-KIRREL3
VIP-DPP6
VIP-PCDH20
VIP-CHRM2
VIP-ABI3BP
VIP-CRH

Mean scaled
log2 CPM
0   1

J Percentage mitochondrial reads per cell state

0.2%

0.1%

0.0%

Astro
CGE dev
ID2
L2/3-CUX2
L4-RORB
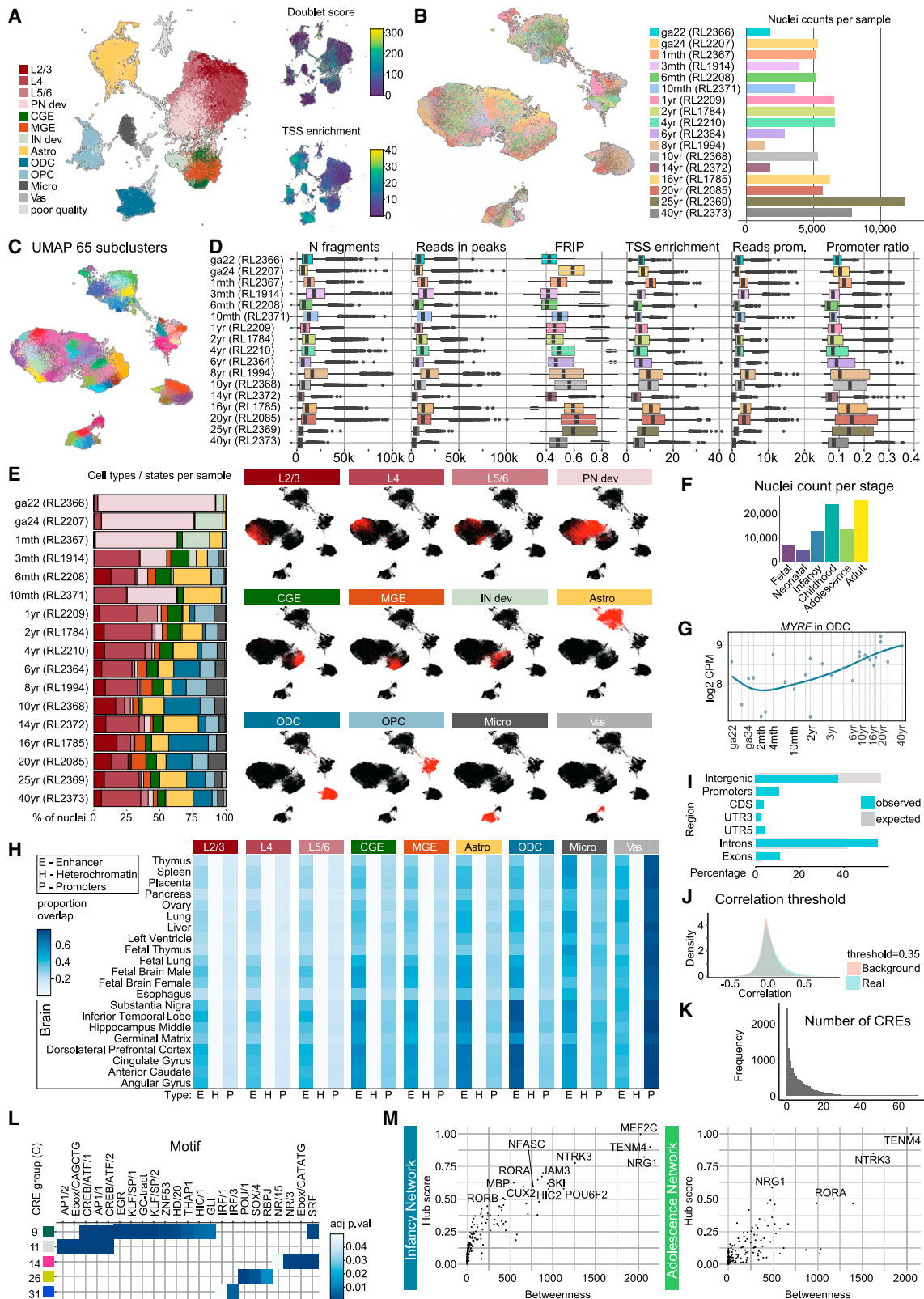L5/6-THEMIS
L5/6-TLE4
LAMP5-NOS1
MGE dev
Micro
OPC
ODC
PN dev
PV
PV-SCUBE3
poor quality
SST
VIP
Vas

Figure S4. Emergence and maturation of LAMP5+ subtypes, related to Figures 1, 3, and 4

(A) Multi-dimensional scaling $\log_2$ CPM scatterplots investigating correlation with potential confounders, including brain region, cause of death, 10x chemistry version, collection year, donor sex, ethnicity, library prep lot number, and post-mortem interval (PMI), in the pseudo-bulked snRNA-seq dataset.

(B) UMAP overlaid by the arcsinh age of each nucleus and vector fields derived from scVelo stochastic model.

(C) UMAP of maturation (UMAT) overlaid by the arcsinh age of each nucleus.

(D) UMAT overlaid by select subcluster colorations (CGE-dev and MGE-dev), and vector fields derived from scVelo projected onto a UMAT embedding with colorations of select subclusters (LAMP5-NOS1 and LAMP5-CKK).

(E) Probability mass flow plots from a CellRank CytoTRACE kernel for outflows to LAMP5-NOS1 or LAMP5-CCK clusters across developmental stages from combined CGE-dev and ID2-dev clusters (top) and MGE-dev cluster (bottom). The width of horizontal veins is proportional to nuclei type frequency at each developmental stage and the width of edges between nuclei types is proportional to flow probability.

(F) Eigentrend values of devDEGs in the ion transport GO-term across development for different LAMP5 subtypes.

(G) RNAscope in PFC tissue sections from 8-year-old (left) and 2-year-old (right) individuals, showing the location of LAMP5-NDNF nuclei positive for *NOS1* (red) and *NDNF* (gray), but negative for *NPY* (green) at the edge of the cortical sections. Sections counterstained with DAPI (blue).

(H and I) (H) $\log_2$ CPM expression in nuclei of LAMP5+ populations from this study, and (I) mean scaled $\log_2$ CPM expression (scaled per gene) in all IN sub-clusters from this study, for previously identified marker genes for layer 1 canopy cells from Schuman et al. (2019) (*NDNF+ NPY−*) and the layer 1 LAMP5-NDNF population from Hodge et al. (2019) (*LAMP5+ SST+ NMBR+*).

(J) Percentage of mitochondrial reads per cell state.

Figure S5. Construction of a chromatin accessibility reference of the human PFC over development, related to Figure 5

(A) UMAP of unfiltered snATAC-seq dataset colored by annotation of different cell types/states and poor quality nuclei (left). UMAPs colored by TSS enrichment and doublet score for each nucleus (right).

(B) UMAP of snATAC-seq nuclei colored by donor origin (left). Number of high quality nuclei per sample (right).

(C) UMAP overlaid with all 65 identified subclusters.

(D) Number of fragments, reads in peaks, fraction of reads in peaks (FRIPs), TSS enrichment, reads in promoters, and fraction of reads in promoters per sample.

(E) Proportion of cell types/states in each sample (left), using the (A) cell type color key. UMAPs overlaid with matched cell type/state labels for each nucleus, derived from matched snRNA-seq data (right).

(F) Number of nuclei in each stage.

(G) Expression ($\log_2$ CPM; dot: individual sample; line: GAM fit) of *MYRF* across development in ODCs.

(H) Proportion of peaks in different cell types from this study overlapping chromHMM predicted enhancers, heterochromatin, and promoters identified across a series of human tissues and developmental stages (Ernst and Kellis, 2017).

(I) Proportion of peaks of particular types of genomic regions, and the observed versus expected frequency.

(J) Density plot of Pearson correlation coefficients between CRE accessibility and gene activity in ±250 kb windows considered for the real target assignment (teal) and in different chromosomes used as background (red).

(K) Number of CREs associated with each gene.

(L) Enrichment of TF motifs in various CRE modules; color represents enrichment. Only significant enrichments found in <3 CRE modules are displayed. TF motifs are summarized to TF motif similarity modules.

(M) Scatterplots of hub score versus betweenness of networks of interactions between genes in ensheathment of neurons GO term and TFs in L4 PNs in infancy and adolescence. Networks were constructed using motif binding sites in CRE elements, as estimated from the snATAC-seq, and co-expression of genes and TFs, as estimated from snRNA-seq (STAR Methods). Scatterplots for other PNs showed similar results.
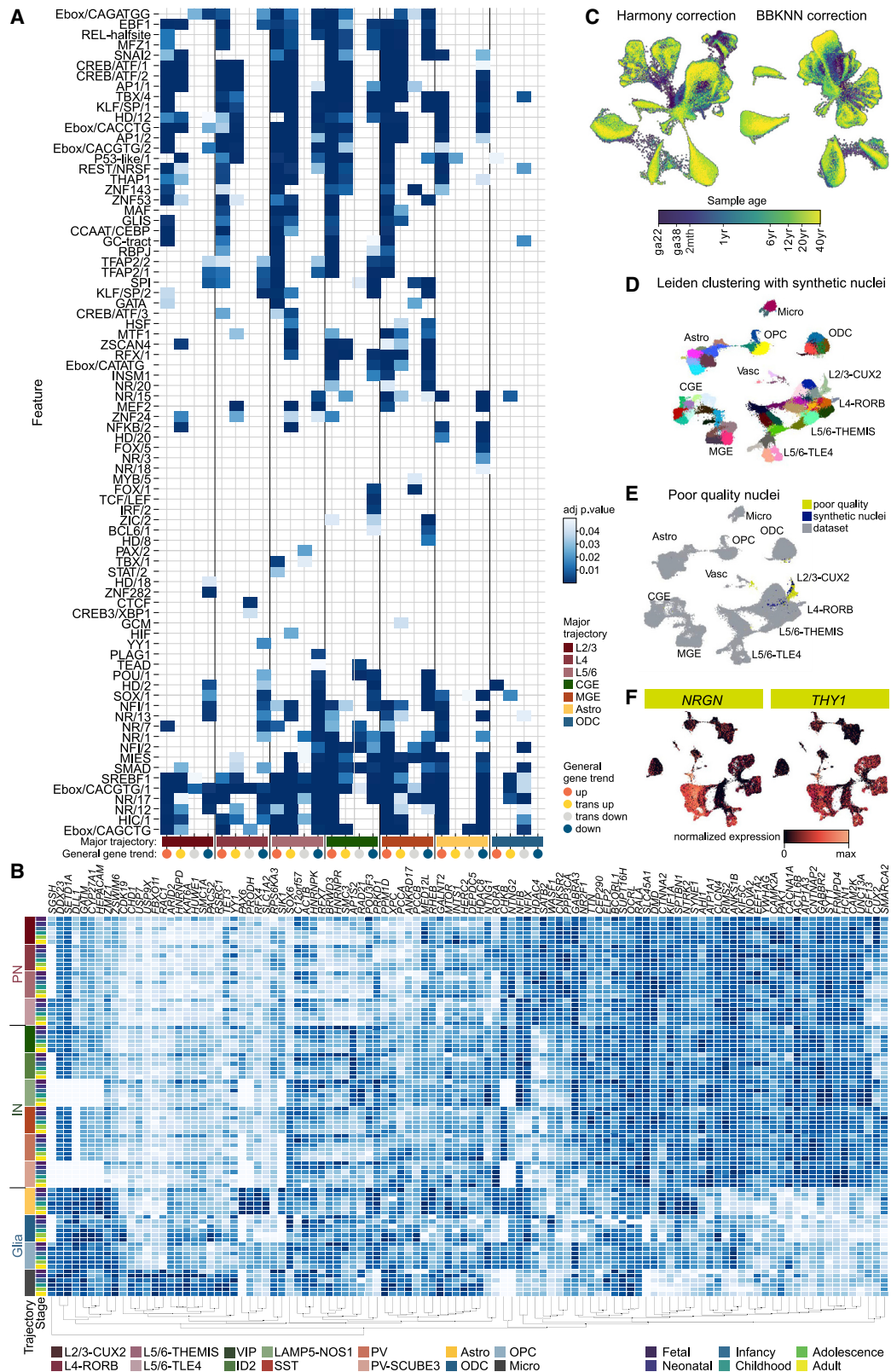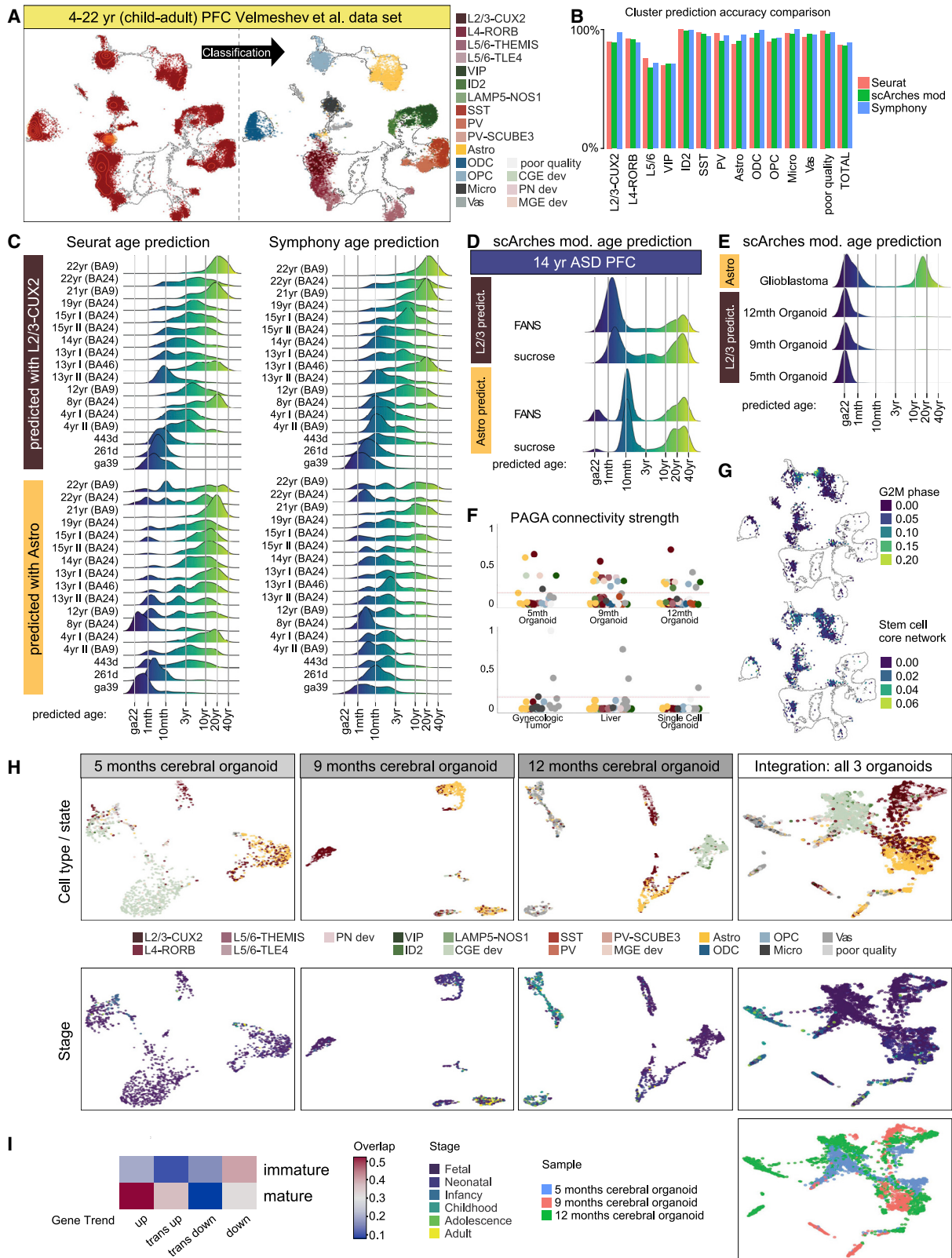
(A) CREs linked to devDEGs in general gene trends across different major trajectories were tested for enrichment of TF binding motifs. TF motifs are summarized to TF motif similarity modules. Color represents adjusted p value for enrichment.

(B) Heatmap of expression (snRNA-seq) of syndromic ASD genes (Simons Foundation Autism Research Initiative) across development in 10 major trajectories.

(C) UMAP of Harmony (left) and BBKNN (right) corrected snRNA-seq data, with nuclei colored by the donor age.

(D) UMAP of Leiden clustering when synthetic low-quality nuclei are included in snRNA-seq data.

(E) UMAP of snRNA-seq data including synthetic low-quality nuclei; also marked is the population of poor quality nuclei identified in the dataset.

(F) UMAPs overlaid with normalized expression of *NRGN* (left) and *THY1* (right).

A 4-22 yr (child-adult) PFC Velmeshev et al. data set

L2/3-CUX2
L4-RORB
L5/6-THEMIS
L5/6-TLE4
VIP
ID2
LAMP5-NOS1
SST
PV
PV-SCUBE3
Astro
ODC   poor quality
OPC   CGE dev
Micro   PN dev
Vas   MGE dev

B Cluster prediction accuracy comparison

Seurat
scArches mod
Symphony

C Seurat age prediction   Symphony age prediction

D scArches mod. age prediction

14 yr ASD PFC

E scArches mod. age prediction

F PAGA connectivity strength

G
G2M phase
Stem cell core network

H
5 months cerebral organoid | 9 months cerebral organoid | 12 months cerebral organoid | Integration: all 3 organoids

L2/3-CUX2   L5/6-THEMIS   PN dev   VIP   LAMP5-NOS1   SST   PV-SCUBE3   Astro   OPC   Vas
L4-RORB   L5/6-TLE4   CGE dev   ID2   PV   MGE dev   ODC   Micro   poor quality

I
Overlap
Stage
Sample

(legend on next page)

Figure S7. Establishing stable integration and age prediction, related to Figure 7

(A) Projection of stably integrated nuclei from snRNA-seq data of PFC from 14 control individuals from Velmeshev et al. (2019), ranging from 4 to 22 years of age. Density of points is indicated by orange contour lines.

(B) Cluster prediction accuracy for each cell type for Seurat (Hao et al., 2021), Symphony (Kang et al., 2021), and modified scArches.

(C) Distributions of age prediction for nuclei from snRNA-seq data from neurotypical control PFC samples (Velmeshev et al., 2019), and our ga39, and 261- and 443-day PFC samples, using only astrocytes of L2-3 PNs for Seurat (left) and Symphony (right).

(D) Distributions of age prediction for nuclei from our snRNA-seq of the ASD PFC sample prepared by two different nuclei isolation methods (FANS versus sucrose gradient nuclei purification).

(E) Distributions of age prediction for nuclei from snRNA-seq profiling of the glioblastoma and cerebral organoid samples.

(F) Cell type identities of clusters connected to query-dominated clusters. For all partition-based graphical abstraction algorithm (PAGA) connections between query-dominated and reference-dominated Leiden clusters, the weight or strength of these connections (y axis) is plotted for each sample, colored by any cell type with over 100 nuclei present in the reference-dominated cluster. Where multiple cell types are prevalent within a single cluster, multiple points are plotted at the same weight. The red dashed line represents the filtering weight of 0.15.

(G) snRNA-seq UMAP outline overlaid with projection of nuclei from the glioblastoma snRNA-seq in hexbin format colored by average G2M Phase Score (top) and Stem Cell Core Network Score (Ang et al., 2011) (bottom).

(H) UMAP of snRNA-seq of each cerebral organoid sample, annotated by their cell type labels (top row) and stage prediction (bottom row) as determined by developmental snRNA-seq reference integration. Note that the points are transparent to allow overlaid points to be visible.

(I) Proportion overlap between gene trends and the DEGs from the analysis of immature organoid PNs versus mature organoid PNs of brain organoids.