Single-cell analysis of prenatal and postnatal human cortical development

Dmitry Velmeshev*[#1,2,3], Yonatan Perez*[1,2], Zihan Yan[3], Jonathan E. Valencia[4], David R. Castaneda-Castellanos[4], Lucas Schirmer[5,6,7], Simone Mayer[8], Brittney Wick[9], Shaohui Wang[1,2], Tomasz Jan Nowakowski[10], Mercedes Paredes[2], Eric Huang[11], Arnold Kriegstein[1,2#]

[1]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco; [2]Department of Neurology, University of California, San Francisco, [3]Department of Neurobiology, Duke University School of Medicine; [4]Rebus Biosystems, 2255 Martin Ave. Suite D, Santa Clara, CA 95050, [5]Division of Neuroimmunology, Department of Neurology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [6]Mannheim Center for Translational Neuroscience and Institute for Innate Immunoscience, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [7]Interdisciplinary Center for Neurosciences, Heidelberg University, Heidelberg, Germany; [8]Hertie Institute for Clinical Brain Research, University of Tübingen, Germany; [9]UC Santa Cruz Genomics Institute, [10]Department of Neurological Surgery, University of California, San Francisco; [11]Department of Pathology, University of California, San Francisco.

[#]Correspondence: arnold.kriegstein@ucsf.edu and dmitry.velmeshev@duke.edu

*These authors contributed equally to this work.

Abstract:

Human cortical development involves birth, differentiation and maturation of a multitude of cellular lineages over the course of prenatal and postnatal life. Single-cell genomics provided insight into the molecular changes that underlie the development of early cortical lineages but has mostly been applied to the second trimester of prenatal human cortical development. Here, we utilize single-nucleus RNA sequencing to capture the development of the main human cortical lineages across prenatal and postnatal stages until adulthood and generate a comprehensive single-cell transcriptomic atlas of human cortical development. We collected more than 350,000 single-nucleus RNA-seq profiles from 108 tissue samples from 60 donors representing the second and third trimester of gestation, as well as neonatal, early and late postnatal stages of development. We identify lineage-specific developmental programs that underlie the development of specific subtypes of excitatory cortical neurons, interneurons, glial cell types and brain vasculature. Moreover, we discover sex- and region-specific developmental changes in each of these lineages. By intersecting our results with genetic risk factors for autism, we identify the cortical cell types and lineages most vulnerable to genetic insults that cause autism. We show that lineage-specific gene expression programs upregulated in female cells are especially enriched for the genetic risk factors of autism. Since these genes are expressed at lower levels in males during development, they may be more susceptible to heterozygous loss of function in ASD, providing a basis for the increased male to female ratio of autism incidence.

Main Text:

Development of the human cerebral cortex spans months during prenatal stages and years after birth, generating tens to hundreds of cell types across multiple cortical areas. This complex process is orchestrated by lineage-specific gene expression programs that guide production, migration, cell type differentiation and maturation of neuronal and glia cell types, as well as formation of projections and neuronal circuits. Changes in these regulatory gene programs during development lead to the pathogenesis of neurodevelopmental and psychiatric disorders, including autism spectrum disorder and schizophrenia. Previous studies have focused on investigating the molecular processes that underly human cortical development during the

second trimester of gestation (*1-4*), the peak of cortical neurogenesis and neuronal migration. These studies revealed the molecular signatures of progenitor cells and neuronal and glial cell types, as well as early specification of neurons into broad subtypes and their arealization. However, the later stages of human cortical development, including the third trimester of gestation, birth, and neonatal and early postnatal development, have been largely studied using bulk genomic approaches. In order to gain a comprehensive view of human cortical development across prenatal and postnatal stages, we utilized single-nucleus RNA sequencing (snRNA-seq) (*5*) to profile 413,682 nuclei from 108 tissue samples derived from 60 individuals that were not diagnosed with any brain disorders or had evidence for neuropathological abnormalities (**Tables S1**). We sampled nuclei from ages spanning the range from the second trimester of gestation to adulthood (**Fig 1a, Table S2**), including samples from the third trimester and early postnatal stages that are often excluded or significantly underrepresented in genomic studies of the human brain, and from both female (N = 25) and male (N = 35) subjects. We acquired data from the cortex, as well as the ganglionic eminences, the major source of cortical interneurons (*6, 7*). We used Seurat (*8*) to perform unbiased clustering and UMAP embedding. After removing a cluster of cell debris (**Fig S1a**), we retained 358,663 nuclei and identified clusters corresponding to neural progenitors, as well as all the major subtypes of excitatory and inhibitory neurons and glial cells in the cortex (**Fig 1b-c, Fig S1b**), as well as brain vascular cells, indicating that we were able to capture the transcriptomic changes underlying differentiation and maturation of cortical cell types across development. We detected a median of 1106 genes and 1609 transcripts per nucleus, which is comparable with published single-cell genomics data collected from the human brain (*9*), with mature neuron cell types expressing higher number of genes and transcripts (**Fig S1c**). We did not observe significant batch effects, with nuclei from different samples well intermixed, and none of the clusters were formed by nuclei from a single sample (**Fig S1d**). Nuclei were captured from the prefrontal, cingulate, temporal and insular cortices (**Fig S1e**). For prenatal samples that were not sex-identified, we determined their sex using sex-specific gene expression (**Fig S1f**). For one individual, we found that its sex was documented incorrectly, as was evident from expression of sex genes, and this was corrected in the final sex annotation. We observed that nuclei clustered according to their developmental age (**Fig 1d**), suggesting that transcriptomics changes associated with development are a major driver of cell identity.

We next aimed to examine the developmental trajectories of excitatory neurons and astrocytes that are generated by dorsal cortical neural progenitors. We selected clusters corresponding to dorsal forebrain progenitors (including radial glia and intermediate precursor cells) as well as clusters containing excitatory neurons and astrocytes. By re-clustering these data and referencing molecularly defined cell types annotated in the Allen Brain Atlas (*10*), we identified clusters corresponding to known subtypes of excitatory neurons, including layer 5 corticofugal (L5), layer 6 corticothalamic (L6-CT), and upper (L2-3) and deep-layer intertelencephalic (L5-6-IT, L6-IT) neurons (**Fig 2a, Fig S2a**). We next used monocle 3 (*11*), as well as custom scripts (**Methods**) to construct trajectories based on snRNA-seq data (**Fig 2b**) and calculated pseudotime for each nucleus. We observed several branching points in the trajectory, including between excitatory neuron and astrocyte lineages, as well as between subcortical projection neurons (L5, L6-CT and L6b) and upper-layer cortico-cortical projection neurons (L2-3) and deep-layer IT and layer 4 neurons. Pseudotime generally corresponded well to the developmental age of nuclei in each

lineage (**Fig 2c**). However, we observed that the relationship between pseudotime and age for intratelencephalic projecting neurons, layer 4 neurons, and astrocytes, was different from excitatory neurons projecting to subcortical regions (**Fig 2d**). The former cell types demonstrated changes extending into postnatal life, while there were almost no pseudotime changes in cortico-subcortical projection neurons after birth. This suggests that unlike these corticofugal projection neurons, cortico-cortical, layer 4 neurons, and astrocytes continue to mature months to years after birth. Next, we aimed to identify lineage-specific gene expression programs for each of the excitatory neuron and astrocyte lineages that we identified. First, we selected nuclei along branches corresponding to each lineage (**Fig S2b, Methods**). We then performed Moran's I test for each lineage to identify dynamically expressed genes while controlling for confounding factors, such as sex, brain region and postmortem interval (**Methods**). We compressed gene expression along pseudotime in each lineage using a sliding window approach (**Methods**), fit gene expression curves, and compared them between lineages to find lineage-specific genes. We next classified these genes based on the pattern and age of onset of their expression (**Fig 2e, Table S3**). We identified genes that were expressed transiently during progression of specific lineages, such as the L2-3-specific gene, *RERGL*, as well as genes that were upregulated more than once during lineage progression, including a deep-layer corticofugal-specific gene, *DIP2A*. This suggests that certain genes play roles at multiple stages of differentiation and maturation of cortical cell types. Gene ontology (GO) analysis of lineage-specific genes suggests that they are associated with neurodevelopmental processes, especially synaptic development and signaling, as well as ion transport and membrane potential homeostasis (**Fig S2c**). Remarkably, expression of most of the genes associated with these processes peaked either during the second trimester of gestation or in the adult, suggesting ongoing maturation of excitatory neurons in the human cortex through adolescence and young adulthood. Using a multiplexed single-molecule RNA *in situ* hybridization (smFISH) platform, we performed analysis of 24 dynamically expressed genes in GW20 human cortex (**Fig 2f, Fig S2d**). All but one of these genes were specific to the cortical plate, and 7 genes (*HS6ST3*, *PTCHD1-AS*, *UNC5C*, *ST18*, *EPHA6*, *RBM20, HS3ST4*) demonstrated layer-restricted expression in the GW20 cortical plate, suggesting early commitment of specific excitatory neuron subtypes during the second trimester, including L2-3, L4 and L5-6-IT subtypes (**Fig 2f**, **Fig S2e)**. Interestingly, *SP18*, marker specific to the L5 lineage, localized to the subplate, suggesting that subplate neurons contribute to layer 5 excitatory neurons in the adult cortex. By comparing nuclei from male and female individuals in each lineage, we identified lineage-specific genes that are enriched in either male or female cells during development (**Table S3**). We observed more male-enriched genes in astrocytes; however, all excitatory neuronal subtypes had more female-enriched genes (**Fig 2g**). Interestingly, while relatively more male-enriched genes were expressed from the Y chromosome, many of which were detected in a previous bulk gene expression study of the developing human brain (*12*), we did not observe an increase in X-linked genes in female-specific genes (**Fig S3a**). While several top sex-enriched genes were known sex-specific genes, such as *XIST*, others were associated with brain development, such as *CNTN3* (contactin 3) and *SYT6* (synaptotagmin 6) (**Fig 2h, Fig S3b**). GO analysis demonstrated that male-enriched genes were solely associated with protein translation, while female-enriched genes were linked to neurodevelopmental processes (**Fig S3c**). These results point to the existence of sex-specific gene expression programs during specific stages of development of cortical excitatory neurons and astrocytes. In addition, by comparing nuclei from the prefrontal and

cingulate cortex, we characterized genes that are dynamically regulated in each lineage in a region-specific manner, including antisense transcription factor *ID3* and connective tissue growth factor (*CTGF*) expressed at higher levels in the frontal cortex astrocytes and L6b neurons (**Fig S3d-f, Table S3**).

Next, we aimed to investigate developmental gene expression changes during differentiation and maturation of GABAergic interneuron (IN) lineages. We selected nuclei from ventral forebrain progenitors, as well as cortical interneurons, re-clustered the data and identified known classes of cortical interneurons (**Fig 3a, Fig S4a**), including interneurons expressing VIP, calretinin, reelin, nitric acid synthase (NOS), two subtypes of parvalbumin (PV)-expressing, and two types of somatostatin (SST)-expressing interneurons. Most IN types expressed either markers of the medial (MGE) or caudal (CGE) ganglionic eminences, with the exception of NOS interneurons, that co-expressed MGE marker genes, *MAF* and *LHX6*, along with the CGE marker, *PROX1*. We captured points at which IN trajectories diverged, including a divergence between MGE and CGE-derived INs. We then reconstructed lineage trajectories corresponding to each interneuron subtype (**Fig 3b**) and calculated pseudotime for each nucleus. Pseudotime correlated well with the developmental age of interneurons (**Fig 3c**). The transcriptional profiles of different IN types changed in a similar fashion across development and displayed most profound changes before birth (**Fig S4b**), which is in contrast with intertelencephalic and layer 4 excitatory neurons, as well as astrocytes. We then identified genes specific to each branch of the lineage trajectory, as well as every IN subtype (**Table S4**). A number of genes ranked as highly lineage and subtype-specific are known interneuron markers, such as the transcription factors *MAF*, *LHX6* (MGE), *PROX1* and *NR2F2* (CGE). In addition, we identified a number of genes that are highly IN subtype-specific but that had not been previously described as playing a role in interneuron specification, including the cytoskeletal protein fidgetin (*FIGN*), restricted to VIP INs, and the transcription factor *AFF1*, specific to PV interneurons (**Fig 3d**). Gene ontology analysis showed that top GO terms enriched for IN lineage-specific genes were associated with synaptic signaling (**Fig S4c**). Interestingly, compared to excitatory neurons, few IN genes peaked during the second trimester, with most of them reaching maximum expression in the adult. Next, we investigated sex-specific programs underlying cortical interneuron development. Since we only had access to a few ganglionic eminence samples, all of which were from male subjects, we focused on analyzing sex-divergent gene expression in INs after migration into the cortex (**Fig 3e**). We identified multiple genes that were enriched in male and female INs as they migrated and matured in the cortex. These included the syndecan proteoglycan *SDC3* enriched in male SST interneurons and *CBLN2* (cerebellin 2 precursor), which is enriched in female SST interneurons and is involved in synapse structure maintenance and synaptic transmission. Both male and female-enriched IN genes were associated with neurodevelopmental pathways (**Fig S4d**). We also performed region-specific analysis of IN developmental programs by comparing cells that migrated into the frontal or cingulate cortex (**Fig 3f**). The number of region-specific genes was much smaller in INs than in excitatory neurons and astrocytes (**Fig 3g, Fig S4e**), consistent with observations in adult mouse cortical cell types (*13*). Among the few frontal cortex-enriched genes, we observed *EGFR* in reelin and *NCAM* in PV INs (**Fig 3h**), suggesting some degree of regional specification of certain subtypes of cortical interneurons.

We further investigated lineage progression in human cortical oligodendrocytes. We selected cells from clusters containing oligodendrocyte precursor cells (OPC) and mature oligodendrocytes (OL), as well as ventral and dorsal forebrain neural precursors, and re-clustered the data (**Fig 4a-b, Fig S5a**). Ventral forebrain radial glia are known to produce an initial wave of oligodendrocytes that populate the cortex starting as early as gestational week 10 (*14*), while at later developmental stages oligodendrocytes are produced by dorsal forebrain radial glia. Trajectory analysis (**Fig 4c**) revealed a linear progression of progenitor cells through OPCs to mature OLs. We did not observe two separate trajectories originating in ventral and dorsal progenitor clusters, but rather observed a single trajectory starting in the ventral progenitor cluster that passed through dorsal progenitors. Considering that the earliest developmental time point we sampled was GW16, we may not have captured an earlier ventral wave of OL production. We therefore focused on the trajectory starting from dorsal progenitors. We calculated pseudotime based on this trajectory (**Fig 4d**) and identified dynamically regulated lineage-specific genes (**Table S5**) by comparing genes expressed in the OL lineage to genes expressed in subtypes of excitatory neurons, astrocytes and interneurons. We observed known genes involved in oligodendrocyte lineage commitment and maturation (**Fig 4e**), such as *PDGFRA*, *OLIG1* and *MAG*, as well as genes previously not known to be involved in OL development, including the membrane protein coding gene, *LRRN1*, the presenilin 1 (*PSEN1*), gene involved in Alzheimer's disease (AD) and detected in aging and AD OLs, as well as several long noncoding intergenic RNAs (lincRNAs) (**Fig 4e**). Gene ontology analysis revealed that genes involved in processes such as glia differentiation and extracellular matrix organization started to peak during early postnatal life, while genes associated with myelination and axonal ensheathment reached maximum expression in the adult (**Fig S5b**). By comparing male and female OL lineage cells, we detected sex-enriched genes expressed during OL development (**Table S5**), including genes known to regulate neural development, such as *CTNNA2* and *NEUROD6* (**Fig 4f**). However, all of the sex-enriched OL genes were expressed at higher levels in radial glia and reduced their expression during OL differentiation, suggesting that sex-specific transcriptomic programs do not play a major role in oligodendrocyte differentiation. Similar to the sex-enriched genes of excitatory neurons and astrocytes, we found that genes enriched in female OL cells were associated with neurodevelopmental processes, while male-enriched genes were associated with translation regulation. The number of region-enriched genes expressed during OL lineage development was smaller than in excitatory neuron and astrocyte lineages but larger than in interneurons (**Fig 4g, Fig S5c**). Among region-enriched genes, we observed several cell adhesion molecules, such as *NRXN3* and *NRCAM*, suggesting a degree of regional specification of cortical oligodendrocytes.

Finally, we sought to elucidate gene expression changes that underlie maturation of mesenchymal brain cell types, including microglia and brain vascular cells. By performing trajectory analysis of microglial cells from different developmental stages, we identified three trajectories that correspond to microglia maturation in the cortex (**Fig 5a**). Two of these trajectories passed through all developmental stages and converge on the same population of microglia in the adult cortex, while the third trajectory culminates in a microglial population mostly from aged individuals. We therefore concluded that the third trajectory is associated with microglia aging and activation and focused on the first two developmental trajectories. These

two microglia trajectories differed in the expression of some microglial signatures, such as integrin subunit beta 2 (*ITGB2*), but shared most developmentally upregulated genes (**Fig 5b, Table S6**). In addition to classic microglia markers, such as *P2RY12*, we observed upregulation of a long noncoding RNA, *LINC01141,* and a transcription factor, *ZNF710,* during human cortical microglia development. In addition to microglia, we analyzed a cluster containing brain vascular cells which contained endothelial cells, pericytes, smooth muscle cells, and fibroblasts (**Fig 5c**). Because fibroblasts and smooth muscle cells were relatively scarce, we focused our analysis on the endothelial cells and pericytes. We observed three different trajectories (**Fig 5d**), one of which corresponded to pericytes (**Fig 5e**) and two to endothelial cells. One of the endothelial cell trajectories originated from a population of young (2-10 years of age) endothelial cells and expressed markers of endothelial cells activated as a result of inflammation and ischemia, such as *APOL1 (15)* and E-selectin (*SELE*) (*16*). We therefore focused on the other endothelial lineage and identified both known and novel developmentally regulated endothelial cell markers (**Fig 5f, Table S6**).

Once we defined lineage-specific developmental gene programs in human cortical cell types, we sought to investigate how these transcriptional programs may be affected in autism spectrum disorder (ASD), one of the most common neurodevelopmental diseases. We compiled all lineage- and cell type-specific gene signatures for excitatory neurons, astrocytes, oligodendrocytes, interneurons, microglia, endothelial cells and pericytes, in total obtaining 3404 unique genes. We then overlapped this gene list with lists of genes associated with ASD risk. We used a broad list of ASD genes from the SFARI database (N = 1045) (*17, 18*), as well as a list of 102 high-confidence ASD genes (ASD-HC) (*19*). We observed a strong enrichment of SFARI genes in our compendium of lineage-specific genes (**Fig 6a**) (N = 326, hypergeometric P value = 1.6×10^{-160}). Of these 326 SFARI genes, 27 were among high-confidence ASD genes (P = 1.7×10^{-12}). We next asked which specific cortical lineages and cell types are most enriched in these genes. We calculated the proportion of lineage-specific genes that are among the ASD risk genes in the SFARI database, as well as the hypergeometric p value (**Table S7**). We performed this analysis for each branch point in the excitatory neuron and interneuron lineages, as well as the specific neuronal and glial cell types (**Fig 6b**) and colored each branch point and terminal cell type according to the proportion of SFARI genes and the p value. We only considered cell types with at least 5 SFARI genes among their developmental signatures; of all such cell types and branches had hypergeometric p values < 0.05. We observed the strongest enrichment for ASD risk genes among earliest-emerging lineages, including pan-excitatory neurons (Ex-1), CGE (IN-1) and MGE-derived (IN-2) interneurons, as well as astrocytes. Among specific subtypes of neurons, genes specific to somatostatin interneurons and layer 5 excitatory neurons were most enriched for ASD risk genes, followed by parvalbumin interneurons, deep-layer intertelencephalic (L5-6-IT) projection neurons, VIP interneurons, and upper-layer excitatory neurons. We also observed significant enrichment in developmental genes specific to oligodendrocytes, as well as in microglial developmental programs. By analyzing the developmental stage at which these developmentally regulated ASD risk genes reached peak expression, we observed two enriched age periods: 2nd trimester to 1 year of age and the adult. We observed almost no ASD risk genes that peaked between 1-2 years and fewer genes that reached maximum expression between 2 and 20 years. This suggested that a while a large number of ASD risk genes is expressed at the highest level

during prenatal and neonatal development, a portion of these genes does not reach peak expression until adulthood. We next explored enrichment of ASD risk genes in sex-specific developmental programs. We observed strong enrichment of female-specific developmental genes in both SFARI and HC-ASD gene lists (**Fig 6d-e**). Male-specific genes were much less frequently found among SFARI genes and, strikingly, we did not find a single male-enriched gene among high-confidence ASD genes. This was in contrast to 25 female-enriched high-confidence ASD genes, pointing to strong enrichment of the genetic risk of ASD among developmental genes that are more highly expressed in female cells. Female-specific SFARI genes were expressed across all neuronal cell types, as well as astrocytes and oligodendrocytes, while we saw fewer of these genes in microglia and vascular cell types (**Fig 6e**). Female-specific HC-ASD genes were mostly observed in excitatory neuron lineages and astrocytes. Interestingly, among the 37 male-specific SFARI genes, almost a third were detected in microglia (**Fig 6f**). Among female-specific high-confidence ASD risks genes were the transcription factor *ARID1B* and cytoskeletal scaffolding protein ankyrin 2 (*ANK2*) that were specifically upregulated in female upper-layer cortical excitatory neurons during their development, *CACNA1E* encoding calcium voltage-gated channel subunit expressed at higher levels in female L5 neurons, as well as transcription factor FOXP1 expressed in a sex-specific manner in L5 and L6b neurons (**Fig 6g**). This provides strong evidence to support the ASD female protective effect (FPE) hypothesis (*20*), and suggest that fine-tuning of developmental trajectories of cortical lineages by sex-specific developmental programs can contribute to the male bias in the pathogenesis of ASD. These findings suggest that fine-tuning of developmental trajectories of cortical lineages by sex-specific developmental programs can underlie the male bias in the pathogenesis of ASD.

We performed a large-scale unbiased single-nucleus transcriptomic analysis of human cortical development throughout the second and third trimester of prenatal development, as well as the period from birth to adulthood. We captured the major cell types and lineages encompassing classes of excitatory neurons, inhibitory interneurons, astrocytes, oligodendrocytes, microglia and brain vasculature cells. By reconstructing single-cell trajectories and identifying genes that are expressed in a lineage-specific manner as cells differentiate and mature, we were able to create a compendium of developmental programs for all the major cortical cell types across prenatal and postnatal life. In addition, we characterized sex and brain region-specific gene programs that are used by specific lineages of cortical cell types. We observed that female-enriched developmental genes are associated with neurodevelopmental processes, while male-enriched genes are involved in protein translation control, suggesting sex-specific fine-tuning of developmental trajectories. We demonstrated that the developmental gene programs utilized by cortical excitatory neurons are the most region-specific, followed by astrocytes and oligodendrocytes. Interneurons, in contrast, express few region-specific genes during development, consistent with data on regional signatures of cortical cell types in the mature human brain (*13*). By utilizing our insights into cortical lineage-specific developmental programs, we investigated their enrichment among genetic risk factors for autism spectrum disorder. We found that early pan-excitatory neuron programs, as well as pan-MGE and pan-CGE programs are most enriched for ASD risk genes, suggesting that a number of mutations underlying ASD can affect genes expressed in broad cortical neuronal lineages. In addition, we found enrichment of ASD genes in more specific lineages, including layer 5 excitatory neurons, somatostatin

interneurons, and astrocytes. These data are in agreement with previous reports of enrichment of ASD genes in deep-layer cortical neurons during mid-gestation (*21, 22*). We previously reported that upper-layer cortical excitatory neurons are most dysregulated in the cortex of idiopathic ASD patients. It would be an important future direction to elucidate how changes in pan-excitatory neuron programs during development can culminate in dysfunction of specific cortical neuronal populations, such as L2-3 neurons, and whether molecular pathology of upper-layer neurons is specific to idiopathic ASD and is driven by common gene variants, rather than rare variants with strong effect sizes (*23*). Finally, we observed a striking enrichment of ASD genetic risk factors among female-specific developmental genes. This finding might explain the 4:1 male to female ratio of individuals affected by ASD and suggests the importance of sexual dimorphism in human brain development. It is possible that higher baseline expression of female-enriched ASD risk genes during development renders female brain more resistant to genetic insults causing autism, especially to haploinsufficiency that can reduce transcript or protein levels by affecting one of the two alleles. The data generated here may help enable fine-grained understanding of human brain development and provide insight into mechanisms of neurodevelopmental disorders.

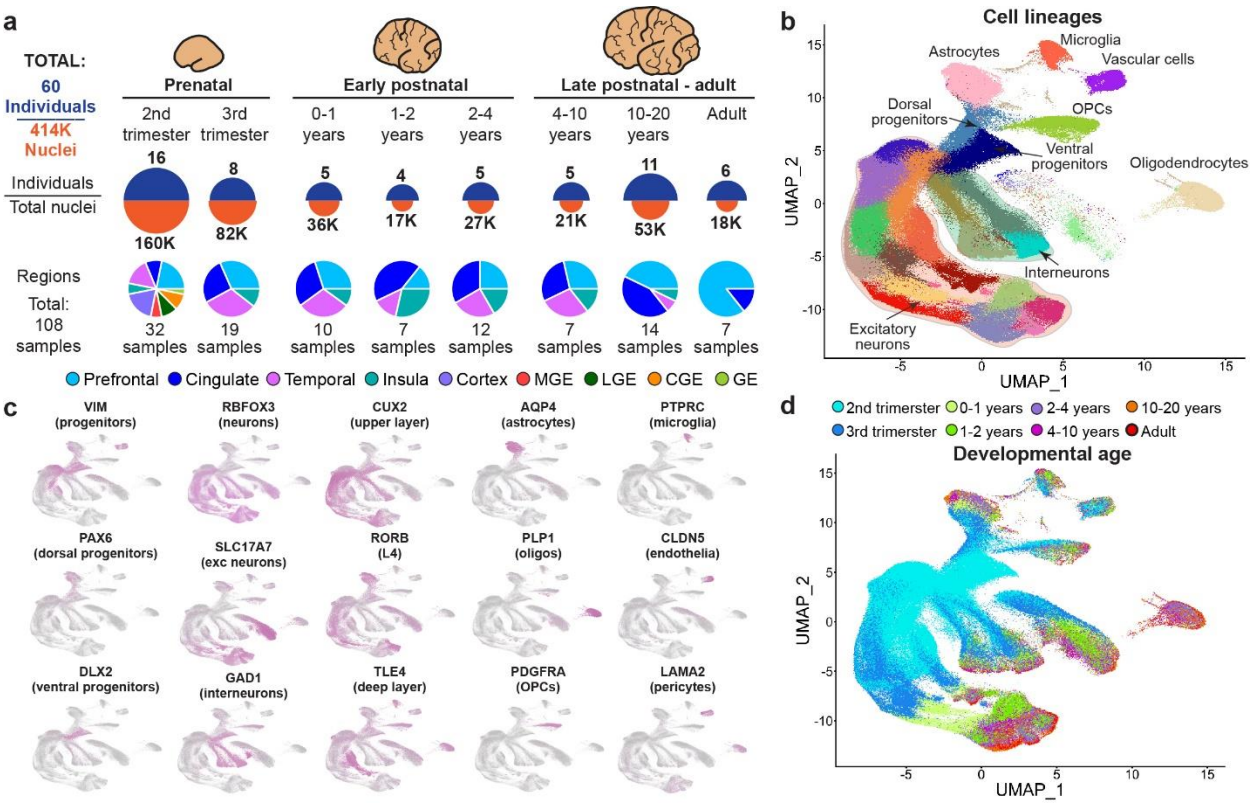

Figure 1. Brain tissue samples used for data collection and initial clustering of snRNA-seq data. a) Overview of the tissue samples used in the current study, including the number of individuals, as well as ages and brain regions captured in the snRNA-seq dataset. MGE-medial ganglionic eminence, LGE-lateral ganglionic eminence, CGE-caudal ganglionic eminence, GE- ganglionic eminence. **b)** Initial clustering of the entire dataset, with the major lineages labeled. **c)** Expression of cell type-specific marker used to determine cardinal lineages. **d)** Nuclei labeled by their developmental age.

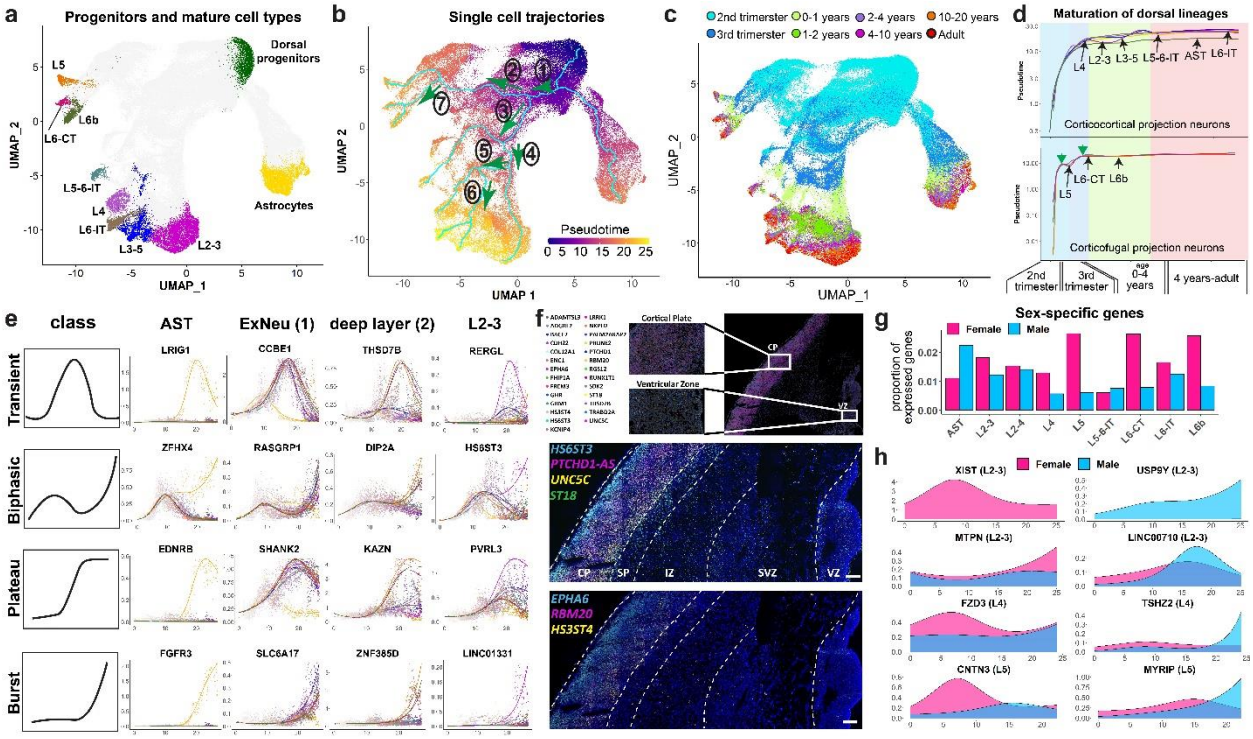

Figure 2. Analysis of excitatory cortical neuron and astrocyte lineages. a) Clusters of cortical progenitors, as well as mature subtypes of excitatory neurons and astrocytes. L2-3 – upper-layer cortico-cortical projection neurons, L3-5 – layer 3-5 excitatory neurons, L4 – layer 4 neurons, L6-IT – layer 6 intratelencephalic projection neurons, L5-6-IT – deep-layer intratelencephalic projection neurons, L6b – layer 6b neurons, L6-CT – layer 6 corticothalamic neurons, L5 – layer 5 neurons. **b)** Reconstructed single-cell trajectories, with labeled trajectory branch points. **c)** Nuclei labeled by developmental age. **d)** Pseudotime/developmental age curves for excitatory neuron and astrocyte lineages. **e)** Classification of genes based on dynamic expression pattern and examples of lineage-specific genes. **f)** Spatial transcriptomics analysis of lamina-restricted expression of lineage-specific genes in GW20 cortex. SP – subplate, SVZ – subventricular zone, VZ – ventricular zone. Scale bar = 250 µm. **g)** Proportion of female and male-enriched genes in each lineage relative to all expressed genes. **h)** Examples of sex-enriched genes in specific lineages.

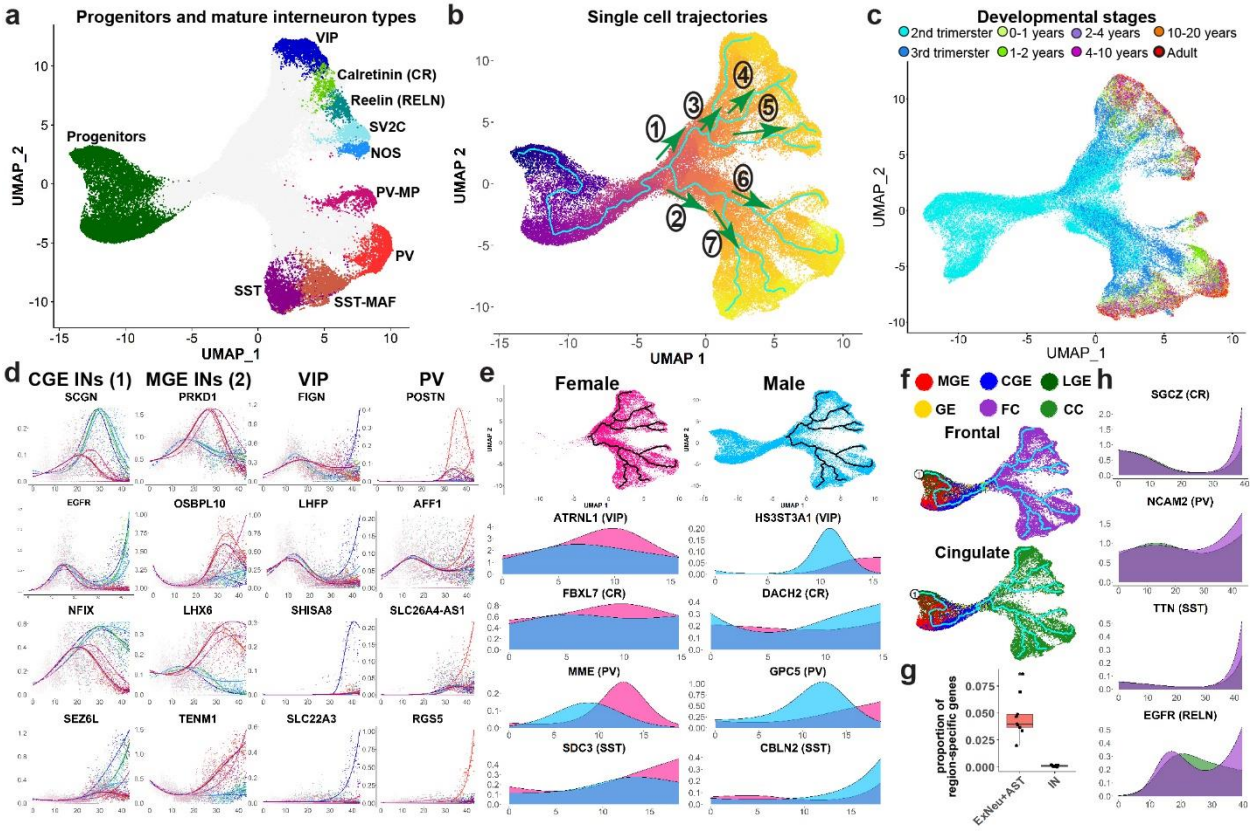

Figure 3. Analysis of inhibitory cortical interneuron lineages. a) Clusters of ventral forebrain progenitors and subtypes of cortical interneurons. **b)** Single-cell trajectories and trajectory branch points. **c)** Distribution of nuclei sampled at different developmental stages. **d)** Examples of dynamically expressed genes specific to the MGE and CGE lineages, as well as specific classes of cortical interneurons. **e)** Analysis of sex-specific gene expression and examples of developmental expression of sex-enriched genes in IN lineages. **f)** Identification of region-specific developmentally regulated genes. FC – frontal cortex, CC – cingulate cortex. **g)** Number of region-specific genes in dorsal forebrain lineages (excitatory neurons and astrocytes) compared to interneurons. **h)** Examples of region-enriched genes expressed during development of specific IN lineages.

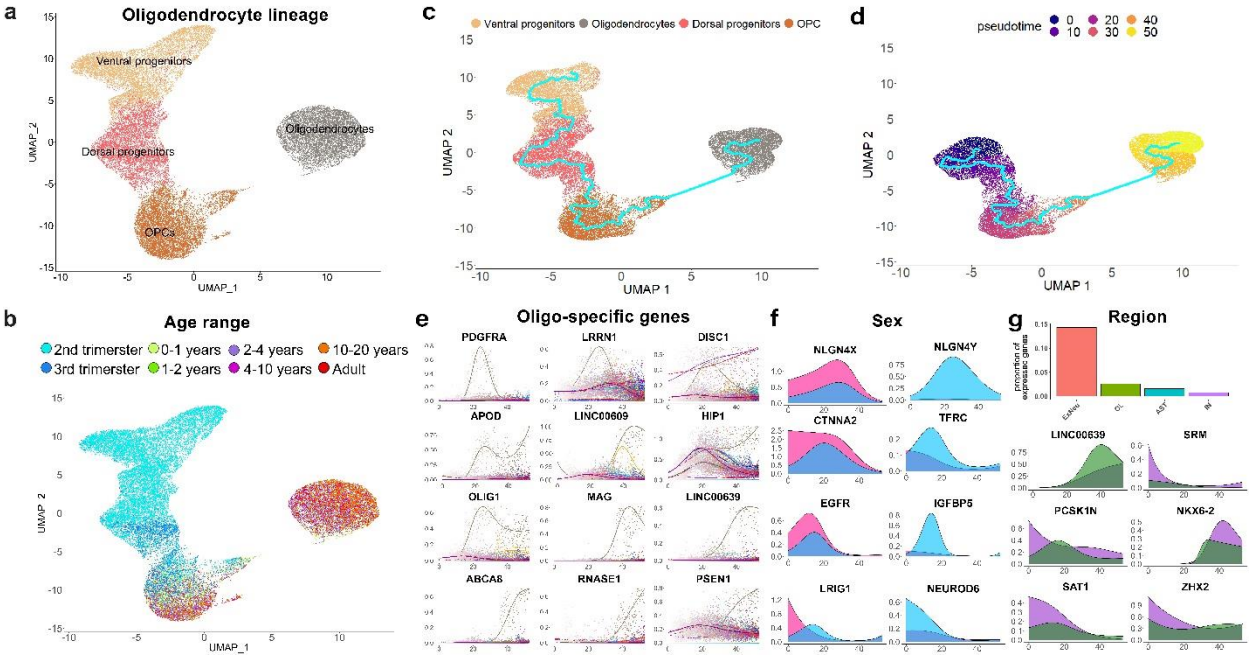

Figure 4. Elucidation of development of cortical oligodendrocytes (OL). a) Clusters of dorsal and ventral forebrain progenitors, oligodendrocyte precursor cells (OPC) and mature oligodendrocytes. **b)** Distribution of OL captured at different developmental stages. **c)** Original single-cell trajectory including ventral forebrain progenitors. **d)** Pseudotime calculated based on the revised trajectory including only dorsal forebrain progenitors. **e)** Genes expressed specifically during development of human oligodendrocytes. **f)** Examples of sex-specific OL genes. **g)** Comparison of the number of region-enriched OL genes with region-specific genes expressed during development of excitatory neurons, astrocytes and interneurons, as well as examples of region-enriched OL genes.

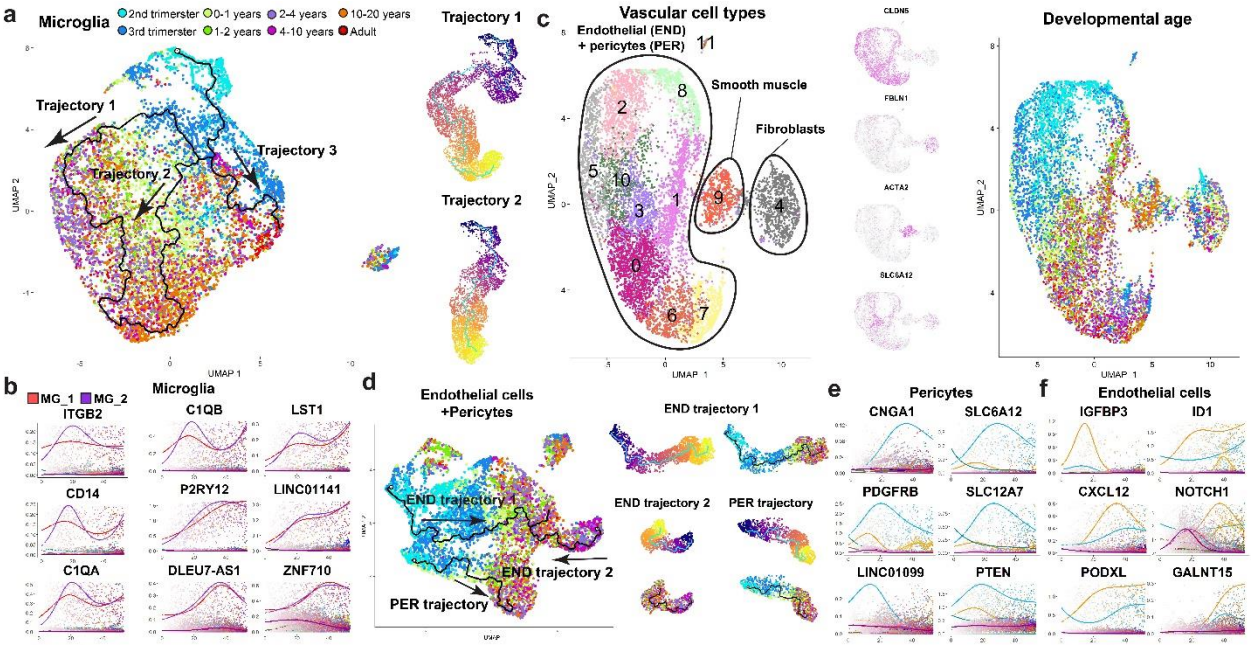

Figure 5. Analysis of microglia and brain vasculature developmental programs in human cortex. **a)** Developmental trajectories of microglia. **b)** Genes expressed specifically by the microglial lineage during development. **c)** Subclustering of brain vasculature cell types, as well as expression of subtype markers and correspondence to developmental age. **d)** Trajectories of developing endothelial cells and pericytes. **e)** Developmental gene signatures of brain vascular pericytes. **f)** Genes upregulated during development of human brain endothelial cells.

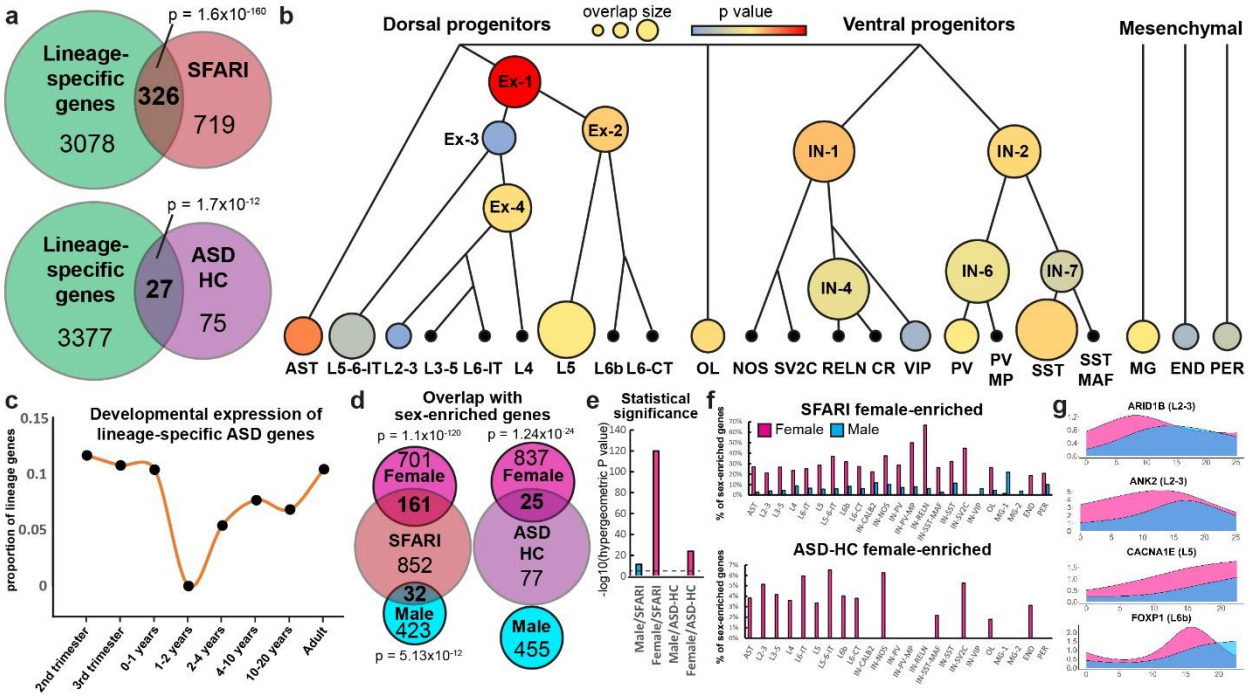

Fig 6. Lineage enrichment of ASD risk genes. a) Overlap of lineage-specific genes with genes in the SFARI database and high-confidence ASD risk genes from Satterstrom et al. **b)** Enrichment of SFARI genes in specific cortical lineages and cell types. Circle size is proportional to size of overlap, while color corresponds to the hypergeometic p value. Black circles are cell types with fewer than five SFARI genes. **c)** Classification of SFARI genes based on the developmental stage at which their expression peaked. **d)** Overlap of SFARI and ASD-HC genes with male and female-enriched lineage genes. **e)** Statistical significance of the overlap between sex-enriched genes and ASD risk genes. **f)** Number of female-enriched ASD genes across cortical cell types. **g)** Examples of high-confidence ASD risk genes that are upregulated during development of female cortical lineages.

Acknowledgments: Funding: This study was funded by the 1U01MH114825-01 grant from the NIMH awarded to Arnold Kriegstein. BW was funded by NHGRI 2U24HG002371-23. **Authors contributions:** DV designed the project, acquired tissue collection, performed nuclei isolation and 10x Genomics capture, analyzed the data and wrote the manuscript. YP performed spatial transcriptomic experiments, nuclei isolation and 10x Genomics capture and edited the manuscript. ZY performed analysis of the oligodendrocyte lineage. JEV and DRC performed probe design and analysis of spatial transcriptomics data. LS acquired adult samples, performed nuclei isolation and 10x Genomics capture. SM performed nuclei isolation and 10x Genomics capture. BW developed Cell Browser visualization tools. SW acquired second trimester tissue samples. TJN performed regional tissue dissections. MP and EH acquired third trimester and early postnatal samples. AK designed and supervised the project and edited the manuscript. All authors read the manuscript. **Competing interests:** Authors have no competing interests. **Data and materials availability:** Raw data can be accessed at the NeMo Archive, accession number nemo:dat-3ah9h9x (https://assets.nemoarchive.org/dat-3ah9h9x). Analyzed data (cell-count matrix and metadata) can be accessed through the UCSC Cell Browser, collection human-cortical-dev (https://human-cortical-dev.cells.ucsc.edu). Code is available at https://doi.org/10.5281/zenodo.7245297.

Supplementary Materials:
Materials and Methods
Table S1 – S7
Fig S1 – S5
References (24 – 27)

Supplementary Materials

Materials and Methods

Sample acquisition and selection

Samples were acquired from three different sources. 1) De-identified second-trimester tissue samples were collected at the Zuckerberg San Francisco General Hospital with previous patient consent in strict observance of the legal and institutional ethical regulations. Protocols were approved by the Human Gamete, Embryo, and Stem Cell Research Committee (institutional review board) at the University of California, San Francisco. These fresh tissue samples were dissected and snap-frozen in isopentane on dry ice. 2) De-identified second-trimester, third trimester and early postnatal tissue samples were obtained at the UCSF Pediatric Neuropathology Research Laboratory led by Dr. Eric Huang. These samples were acquired with patient consent in strict observance of the legal and institutional ethical regulations and in accordance to research protocols approved by the UCSF IRB committee. These samples were dissected and snap-frozen either on a cold plate placed on a slab of dry ice or in isopentane on dry ice. 3) Banked de-identified second-trimester, third trimester, early postnatal and adult tissue samples were obtained from the University of Maryland Brain and Tissue Bank through the NIH NeuroBioBank.

For postnatal ages, samples from individuals with known history of brain disorders or brain trauma were excluded from downstream analyses. For prenatal samples, samples with unusual neuropathology following pathological examination, as well as samples positive for commonly tested chromosomal aberrations, were excluded. Prior to performing nuclei isolation and single-nucleus RNA sequencing, samples were screened for RNA quality by collecting 100um-thick cryosections, isolating total RNA and measuring RNA Integrity Number (RIN) using the Agilent 2100 Bioanalyzer instrument. Only samples with RIN >= 6.5 were included in the study.

Nuclei isolation and generation of single-nucleus RNA-seq data using 10x Genomics platform

40 mg of sectioned brain tissue was homogenized in 5 mL of RNAase-free lysis buffer (0.32M sucrose, 5 mM $CaCl^2$, 3 mM $MgAc^2$, 0.1 mM EDTA, 10 mM Tris-HCl, 1 mM DTT, 0.1% Triton X-100 in DEPC-treated water) using glass dounce homogenizer (Thomas Scientific, Cat # 3431D76) on ice. The homogenate was loaded into a 30 mL thick polycarbonate ultracentrifuge tube (Beckman Coulter, Cat # 355631). 9 mL of sucrose solution (1.8 M sucrose, 3 mM $MgAc^2$, 1 mM DTT, 10 mM Tris-HCl in DEPC-treated water) was added to the bottom of the tube with the homogenate and centrifuged at 107,000 g for 2.5 hours at 4°C. Supernatant was aspirated, and the nuclei containing pellet was incubated in 250 uL of DEPC-treated water-based PBS for 20 min on ice before resuspending the pellet. The nuclear suspension was filtered twice through a 30 um cell strainer. Nuclei were counted using a hemocytometer and diluted to 2,000 nuclei/uL before performing single-nuclei capture on the 10X Genomics Single-Cell 3' system. Usually, the target capture of 3,000 nuclei per sample was used; the 10x capture and library preparation protocol was used without modification. Single-nucleus libraries from individual samples were pooled and sequenced on the NovaSeq 6000 machine (average depth 60,000 reads/nucleus).

snRNA-seq data processing with 10X Genomics CellRanger software and data filtering

For library demultiplexing, fastq file generation and read alignment and UMI quantification, CellRanger software v 1.3.1 was used. CellRanger was used with default parameters, except for using pre-mRNA reference file (ENSEMBL GRCh38) to insure capturing intronic reads originating from pre-mRNA transcripts abundant in the nuclear fraction.

Individual expression matrices containing numbers of Unique molecular identifiers (UMIs) per nucleus per gene were filtered to retain nuclei with at least 400 genes expressed and less than 10% of total UMIs originating from mitochondrial and ribosomal RNAs. Individual matrices were combined prior to pre-processing and clustering with Seurat.

Dimensionality reduction, UMAP embedding, clustering and cell type identification

All of the following bioinformatics analysis steps are documented in an R script available at https://doi.org/10.5281/zenodo.7245297.

Data preprocessing, normalization, variable feature selection and PCA was performed using the standard Seurat pipeline (*24*). Selection of significant principal components was done using the elbow method. The selected components were used to perform UMAP embedding and clustering using the Louvain method. The identity of specific lineages and cell types was determined based on expression of known marker genes, as is shown in Figure 1 and Figure S1.

Sex determination

To determine the sex of individuals for which sex information was not available, we aggregated gene expression of all nuclei by individual and plotted individual-wise expression of the following genes: *XIST*, *TSIX*, *DDX3Y*, *KDM5D*, *USP9Y*, *ZFY*, *EIF1AY*, *UTY*.

Trajectory reconstruction and Isolation of individual lineages

Seurat UMAP coordinates were imported into monocle3 (*25*) for trajectory reconstruction. learn_graph function with custom graph_control options was used to construct the trajectory graph. We noticed that while the original trajectory graph generated by monocle3 corresponded to the major cell lineages, it failed to connect some nodes that passed through populations of cells expressing shared lineage markers. Moreover, some trajectory branches did not correspond to biologically interpretable lineage progression, specifically the branches connecting two mature neuronal cell types containing only adult cells. We corrected these issues by modifying the trajectory according to the following principles: 1) if two terminal nodes failed to be connected but were passing through populations of cells expressing known lineage-specific markers (such as *RORB* for layer 4, *TLE4*/*SEMA4A* for layer 6b, *CUX2* for layer 2-3 and *CUX1* for layer 5-6-IT), we connected these nodes 2) if a branch connected nodes located in two mature cell types, we omitted this branch and 3) based on the first two principles, we isolated the shortest path between the node in the neural progenitor/radial glia cluster and the node in the mature cell type cluster.

Identification of lineage-specific dynamically expressed genes

First, we selected trajectory branches corresponding to specific lineages, as well as the cells along the branches. Then, monocle3's Moran's test (graph_test function) was used to identify genes that are dynamically expressed in each lineage. We modified graph_test function to utilize Moran's test with covariates to ensure that our results are not affected by uneven contribution

of cells from male and female subjects, as well as cells from different brain regions. We selected genes with adjusted p value < 0.05 as statistically significant dynamically expressed genes. To identify lineage-specific genes, we first compressed the single-cell expression data along each lineage by using a sliding window along pseudotime and averaging expression of neighboring cells for each gene. We generated 500 meta-cells in each lineage using this approach. Then, we fit the expression of each gene using a generalized linear model and the following formula: expression ~ splines::ns(pseudotime, df=3). Then, we calculated the area under the curve for the smoothed expression/pseudotime plot for each gene in each lineage. The difference of under the curve between the lineage of interest and all other lineages was used to rank genes according to their lineage specificity. This lineage specificity score of at least 20, as well as Moran's p value < 0.05 was used to define lineage-specific genes.

Identification of sex and region-enriched dynamically expressed genes

To identify male and female-enriched genes in each lineage, we selected cells from only males or females within each lineage and first performed Moran's I test separately for male and female data. Then, we compressed the data and calculated area under the curve for male and female gene expression. Genes with Moran's I statistic >= 0.1, adjusted Moran's p value<0.05 and the area under curve difference between male and female expression >= 50 were considered sex-specific in each given lineage.

Gene ontology analysis

We used ShinyGO (*26*) to perform gene ontology analysis using genes expressed in each lineage as the background gene list. In order to reduce redundancy of the identified GO terms, all significant (adjusted p value < 0.05) terms were used as input to Revigo (*27*). The value of the resulting gene list of 0.4 was used. The -log10(p value) and fold enrichment for the resulting non-redundant GO processes were reported.

Analysis of enrichment of ASD risk genes

We utilized two gene lists: the list of all SFARI genes and the list of 102 high-confidence ASD risk genes identified using TADA (Transmission And De novo Association) test. We intersected these gene lists with our list of lineage-specific genes, as well as genes enriched in male and female developmental lineages. We calculated hypergeometric p values for each lineage, displayed each lineage as a circle with the -log10(p value) proportional to the color intensity and the ratio of ASD genes among lineage-specific genes as the size of the circle.

Data visualization

Cell type, gene expression and lineage trajectories for each lineage can be visualized at https://human-cortical-dev.cells.ucsc.edu.

Spatial transcriptomic gene selection

Marker genes for validation using the spatial transcriptomic platform were chosen by exploring excitatory lineage specific genes showing marked expression changes during the 2nd trimester (Table S2). Narrower gene list was selected following the Esper™ probe design tool, to create a high-fidelity panel with a minimum of 22 probes per target, while considering the length of the transcript, sequence composition and transcript coverage. We designed probes to 30 genes, 27

of which were detected in the GW20 cortex and for the further 24 images could be obtained without imaging artifacts.

Rebus Esper spatial transcriptomics platform

Spatial analysis of the cortical sample was done as previously described (*1*). Briefly, De-identified tissue sample was collected after obtaining consent form, in alignment with legal and institutional ethical regulations. Protocols were approved by the Human Gamete, Embryo, and Stem Cell Research Committee (institutional review board) at the University of California, San Francisco. Cortical sample from GW20 primary tissue was dissected and flash frozen in OCT following the Rebus Esper protocol. Sample was mounted on functionalized coverslips and fixed in 4% PFA for 12 minutes at RT, washed twice with PBSx1 for 5 minutes, then twice with 70% ethanol before preserved in fresh 70% ethanol at 4°C for up to a week prior to processing. The spatially resolved, multiplexed in situ RNA detection and analysis was performed using the automated Rebus Esper spatial omics platform (Rebus Biosystems). The system is used in conjunction with Rebus' proprietary smFISH chemistry based on osmFISH, while integrating synthetic aperture optics (SAO) microscopy (*2*), fluidics and image processing software. Mounted fixed tissue section was assembled in a flow cell and connected to the Rebus machine. Hybridization cycles and imaging were done automatically under the instrumental control software. In brief, primary probes for all target genes were initially hybridized for 6 h and probes not specifically bound were washed away. Readout probes labelled with Atto532, Atto594 and Atto647N dyes for the first 3 genes were then hybridized, washed, counterstained with DAPI and then imaged with an Andor sCMOS camera (Zyla 4.2 Plus, Oxford Instruments) through a 20×, 0.45 NA dry lens (CFI S Plan Fluor ELWD, Nikon) with a 365-nm LED for DAPI and 532-nm, 595-nm and 647-nm lasers configured for SAO imaging. Multiple fields of view (FOVs) were imaged for each channel within the region of interest (ROI). Single z-planes with 2.8 µm depth of field were acquired for each FOV. After imaging, the first three readout probes were stripped and the readout probes for the next three genes were then hybridized, imaged and stripped. This process was repeated for 10 cycles, until readout was completed for all genes. Using the Rebus Esper image processing software, the images were analyzed to identify specific RNA signals. Only RNA spots with signal-to-background ratio higher than 1.5 were selected for quantification. In addition, all SAO images were processed using the SAO-reconstruction algorithm to generate high-resolution images and increase the resolution of single RNA transcript detection. RNA spots were automatically detected to generate high fidelity RNA spot tables containing xy positions and signal intensities. Nuclei segmentation software identified individual cells by finding nuclear boundaries from DAPI images. Using this approach, individual transcripts were automatically detected, assigned to specific cells, and counted. Cell by feature matrix, containing gene-expression and spatial location data for each individual cell, as well as registered imaging data was generated (Table S9).

Table S1. Sample information.

Table S2. Nuclei information and initial clustering.

Table S3. Analysis of excitatory neurons and astrocytes.

Table S4. Analysis of the interneuron lineages.

Table S5. Oligodendrocyte lineage analysis

Table S6. Analysis of microglia and brain vascular cells.

Table S7. Analysis of lineage-specific developmental and sex-specific expression of ASD risk genes.

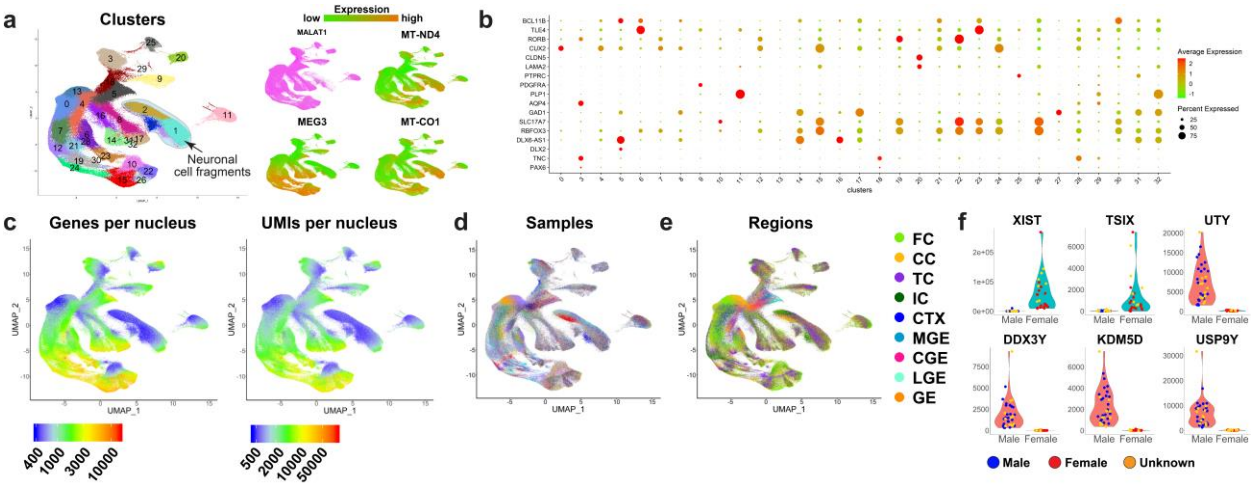

Figure S1. Technical and biological characteristics of the combined snRNA-seq dataset. a) Identification of the clusters containing neuronal debris. **b)** Marker gene expression used to identify specific cell lineages. **c)** Gene and UMI counts per nucleus across all cell types. **d-e)** Distribution of nuclei from different samples and regions. **f)** Expression of sex-specific genes used to determine sex of samples with unknown status.

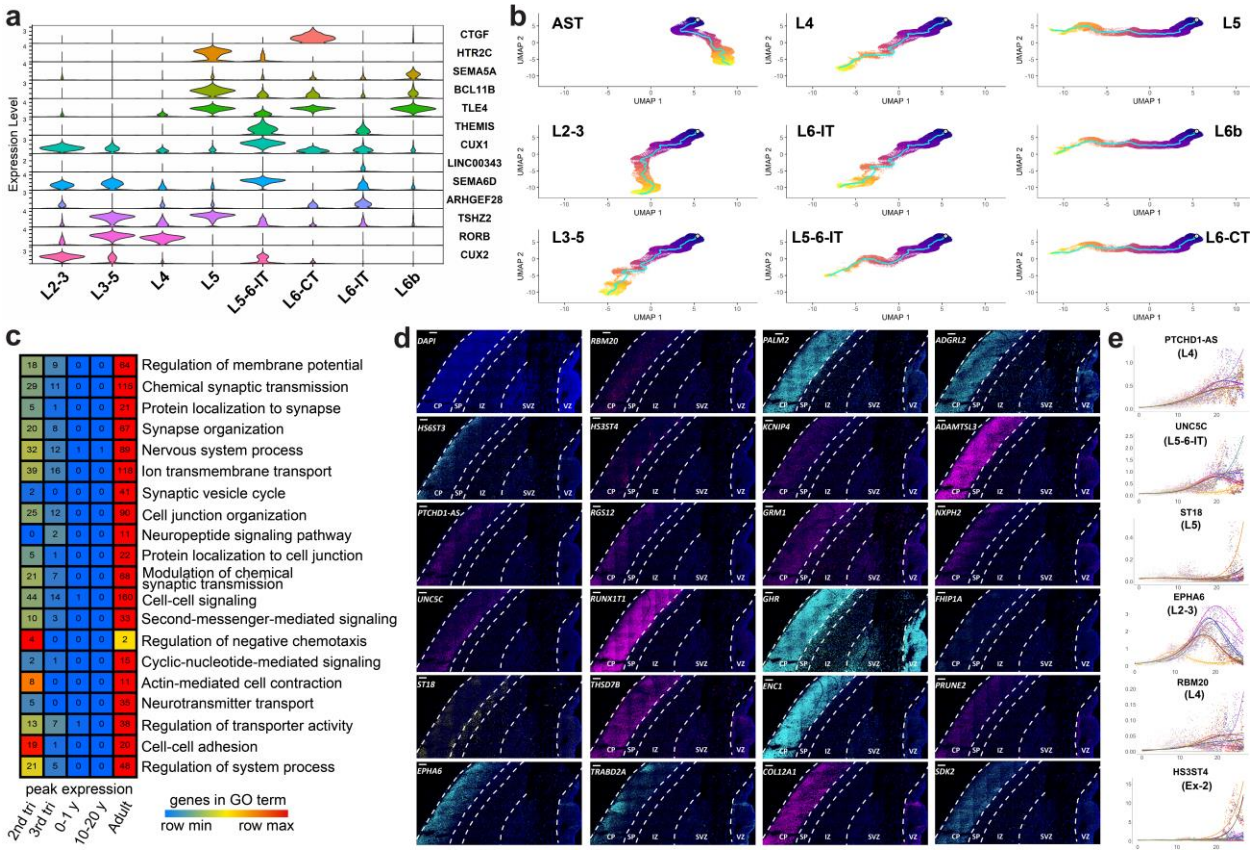

Figure S2. Excitatory neuron and astrocyte lineage analysis and results of spatial transcriptomics experiments. a) Expression of cortical excitatory neuron marker genes used to determine excitatory neuron lineages. **b)** Isolated lineages trajectories for excitatory neuron subtypes. **c)** Gene ontology (GO) processes enriched for excitatory lineage-specific genes with peak expression at different developmental stages. **d)** Fluorescent images for genes detected in GW20 cortical tissue using multiplexed spatial transcriptomics. Scale bar = 250 µm. **e)** Expression of lineage-specific genes validated using spatial transcriptomics.

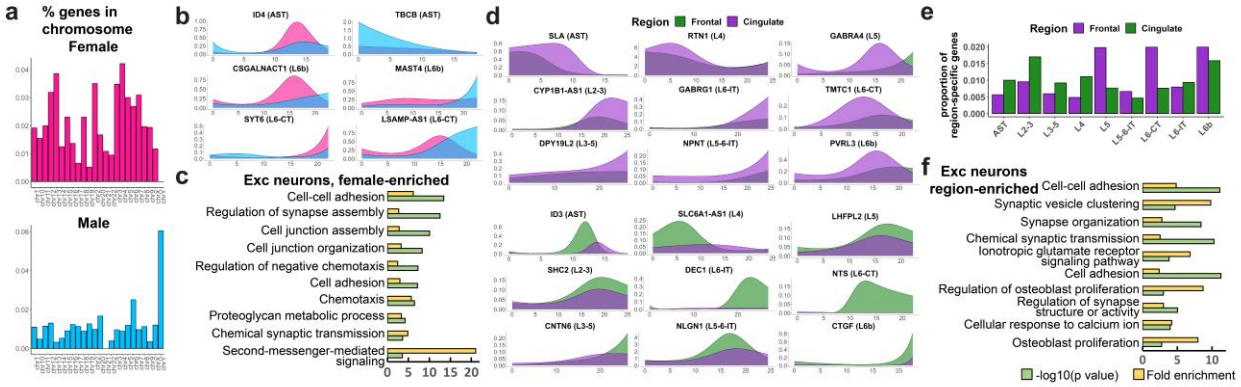

Figure S3. Analysis of sex and region-enriched genes during excitatory neuron and astrocyte development. a) Percent of sex-enriched genes across chromosomes. **b)** Expression of sex-enriched genes in astrocyte and excitatory neuron lineages **c)** Gene ontology analysis for female-enriched genes. **d)** Expression of region-specific developmental genes. **e)** Proportion of region-specific genes relative to all expressed genes in astrocytes and excitatory neurons. **f)** Gene ontology terms associated with region-enriched genes.

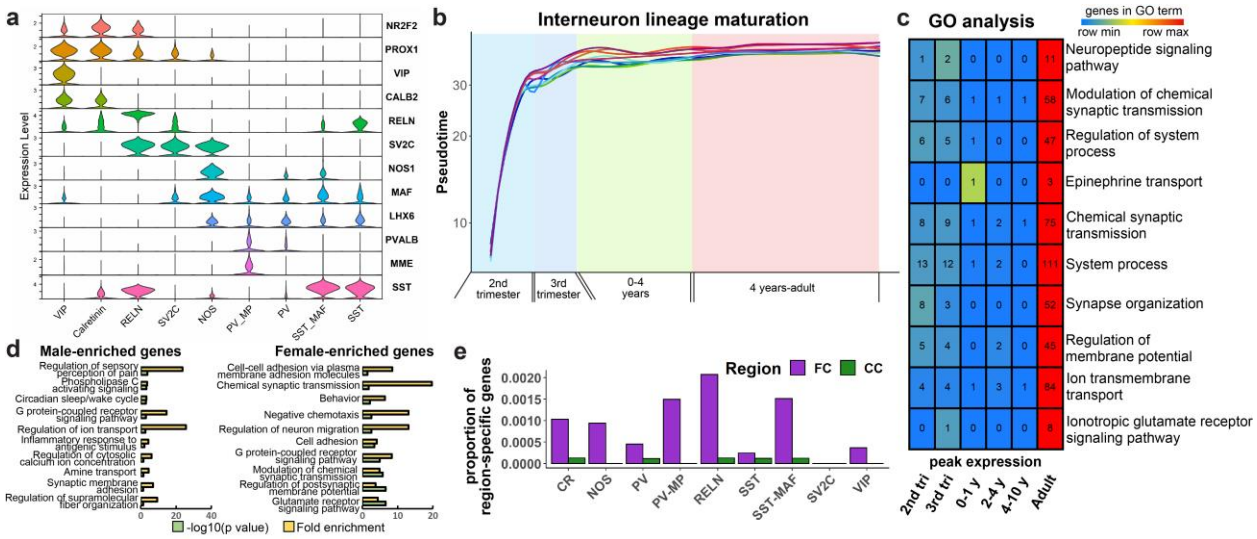

Figure S4. Analysis of interneuron lineages. a) Markers of interneuron lineages and cell types. **b)** Dynamics of development of interneuron lineages during prenatal and postnatal stages. **c)** Gene ontology processes enriched for IN lineage-specific genes expressed at different stages of development. **d)** GO analysis for region and sex-enriched genes. **e)** Proportion of region-specific genes in interneuron lineages.

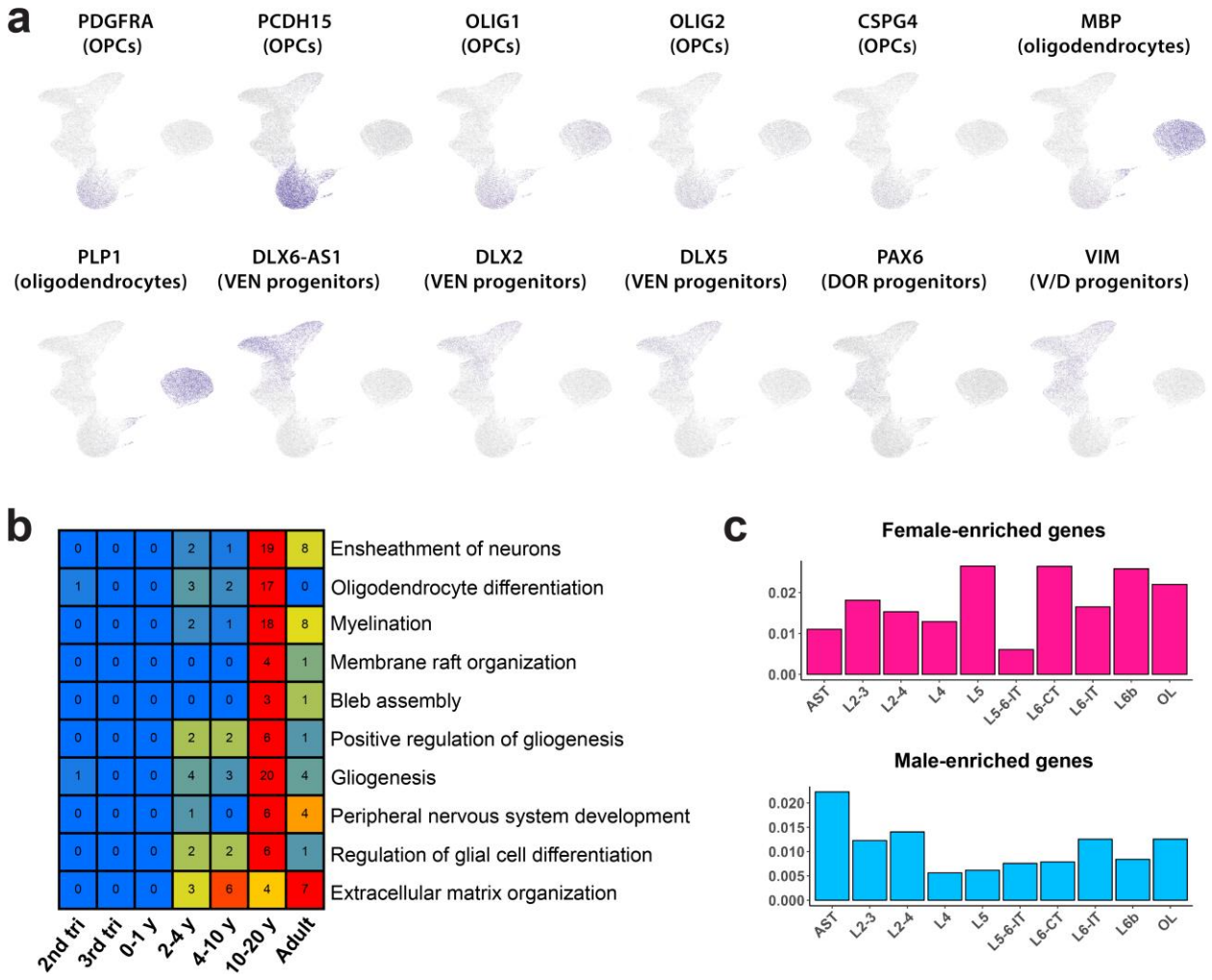

Figure S5. Results of oligodendrocyte lineage analysis. a) Expression of oligodendrocyte lineage markers expressed at different stages of OL development. **b)** GO enrichment analysis of OL lineage-specific genes expressed at different stages of development. **c)** Proportion of sex-enriched genes in the OL lineage compared to neuronal and astrocyte lineages.

References

1. A. Bhaduri *et al.*, An atlas of cortical arealization identifies dynamic molecular signatures. *Nature* **598**, 200-204 (2021).
2. U. C. Eze, A. Bhaduri, M. Haeussler, T. J. Nowakowski, A. R. Kriegstein, Single-cell atlas of early human brain development highlights heterogeneity of human neuroepithelial cells and early radial glia. *Nature Neuroscience* **24**, 584-594 (2021).
3. T. J. Nowakowski *et al.*, Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318-1323 (2017).
4. R. S. Ziffra *et al.*, Single cell epigenomic atlas of the developing human brain and organoids. *bioRxiv*, 2019.2012.2030.891549 (2020).
5. B. B. Lake *et al.*, Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586-1590 (2016).
6. D. V. Hansen *et al.*, Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat Neurosci* **16**, 1576-1587 (2013).
7. T. Ma *et al.*, Subcortical origins of human and monkey neocortical interneurons. *Nat Neurosci* **16**, 1588-1597 (2013).
8. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411-420 (2018).
9. N. Habib *et al.*, Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955-958 (2017).
10. R. D. Hodge *et al.*, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61-68 (2019).
11. J. Cao *et al.*, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
12. H. J. Kang *et al.*, Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483-489 (2011).
13. B. Tasic *et al.*, Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72-78 (2018).
14. I. Jakovcevski, R. Filipovic, Z. Mo, S. Rakic, N. Zecevic, Oligodendrocyte development and the onset of myelination in the human fetal brain. *Front Neuroanat* **3**, 5 (2009).
15. J. Wu *et al.*, APOL1 risk variants in individuals of African genetic ancestry drive endothelial cell defects that exacerbate sepsis. *Immunity* **54**, 2632-2649.e2636 (2021).
16. R. L. Zhang *et al.*, E-selectin in focal cerebral ischemia and reperfusion in the rat. *J Cereb Blood Flow Metab* **16**, 1126-1136 (1996).
17. B. S. Abrahams *et al.*, SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* **4**, 36 (2013).
18. S. Banerjee-Basu, A. Packer, SFARI Gene: an evolving database for the autism research community. *Disease Models & Mechanisms* **3**, 133-135 (2010).
19. F. K. Satterstrom *et al.*, Novel genes for autism implicate both excitatory and inhibitory cell lineages in risk. *bioRxiv*, 484113 (2018).

20. E. B. Robinson, P. Lichtenstein, H. Anckarsäter, F. Happé, A. Ronald, Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences* **110**, 5258-5262 (2013).
21. N. N. Parikshak *et al.*, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008-1021 (2013).
22. A. J. Willsey *et al.*, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997-1007 (2013).
23. T. Gaugler *et al.*, Most genetic risk for autism resides with common variation. *Nat Genet* **46**, 881-885 (2014).
24. E. Z. Macosko *et al.*, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).
25. X. Qiu *et al.*, Reversed graph embedding resolves complex single-cell trajectories. *Nature Methods* **14**, 979 (2017).
26. S. X. Ge, D. Jung, R. Yao, ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**, 2628-2629 (2019).
27. F. Supek, M. Bošnjak, N. Škunca, T. Šmuc, REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).