Published in final edited form as:
Neuron. 2019 September 04; 103(5): 785–801.e8. doi:10.1016/j.neuron.2019.06.011.

A single cell transcriptomic atlas of human neocortical development during mid-gestation.

Damon Polioudakis[1,13], **Luis de la Torre-Ubieta**[1,2,13], **Justin Langerman**[3], **Andrew G. Elkins**[1], **Xu Shi**[4,5], **Jason L. Stein**[6], **Celine K. Vuong**[7], **Susanne Nichterwitz**[2], **Melinda Gevorgian**[2,8], **Carli K. Opland**[1], **Daning Lu**[1], **William Connell**[1], **Elizabeth K. Ruzzo**[1], **Jennifer K. Lowe**[1], **Tarik Hadzic**[1,2], **Flora I. Hinz**[1], **Shan Sabri**[3], **William E. Lowry**[9], **Mark B. Gerstein**[4,5,10,11], **Kathrin Plath**[3], **Daniel H. Geschwind**[1,12,14,*]

[1]Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, UCLA.

[2]Department of Psychiatry and Biobehavioral Sciences, Semel Institute, David Geffen School of Medicine, UCLA.

[3]Department of Biological Chemistry, David Geffen School of Medicine, UCLA.

[4]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA.

[5]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA.

[6]Department of Genetics & UNC Neuroscience Center, University of North Carolina, Chapel Hill.

[7]Department of Microbiology, Immunology and Molecular Genetics, UCLA.

[8]Department of Biology, CSUN

[9]Department of Molecular, Cell and Developmental Biology, UCLA.

[10]Department of Computer Science, Yale University, New Haven, CT 06520, USA.

[11]Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA.

[12]Department of Human Genetics, David Geffen School of Medicine, UCLA.

[13]These authors contributed equally.

[*]**Corresponding author:** dhg@mednet.ucla.edu.

Author Contributions:
L.T.U, J.S., and D.H.G. designed the study. D.H.G. oversaw all of the analyses and supervised the study, with help from K.P., M.B.G., and B.L. L.T.U, J.L., S.N., M.G., C.K.O., and D.L., performed the single-cell experiments. D.P. and J.L.S. processed and managed the data. D.P., L.T.U., A.G.E., X.S., and W.C., analyzed the data, with input from S.S. L.T.U., D.L, and C.K.V. performed the validation experiments. T.H. and J.K.L performed CNV analyses. D.P., L.T.U., and D.H.G. wrote and prepared the manuscript, with feedback from all authors.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

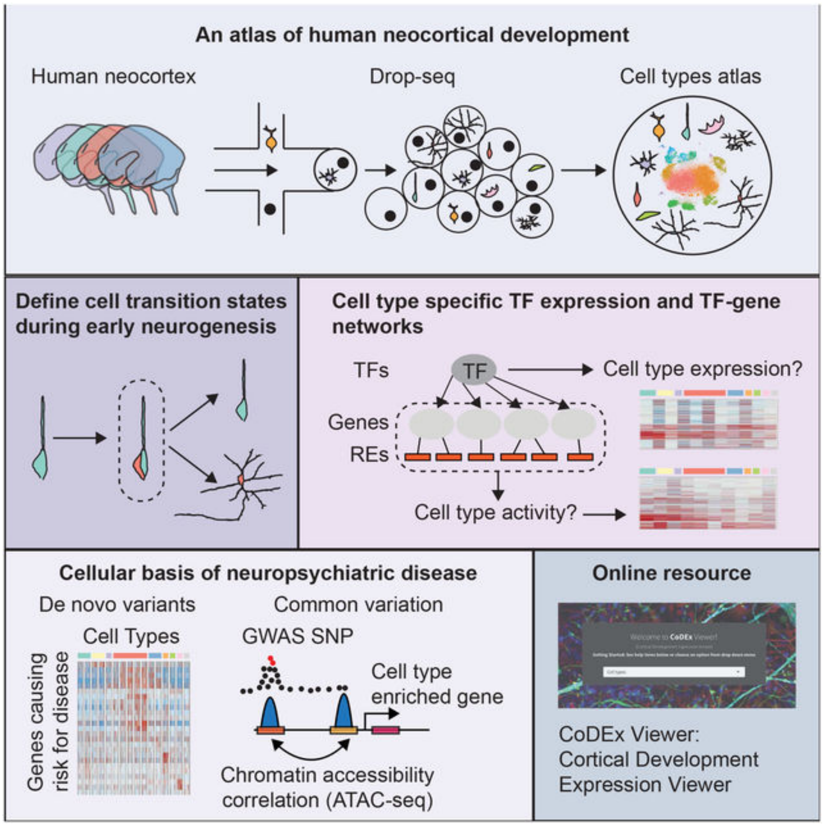
Declaration of Interests:
The authors declare no competing interests.

[14]Lead contact.

Summary:

We performed RNA sequencing on 40,000 cells to create a high-resolution single-cell gene expression atlas of developing human cortex, providing the first single-cell characterization of previously uncharacterized cell types, including human sub-plate neurons, comparisons with bulk tissue, and systematic analyses of technical factors. These data permit deconvolution of regulatory networks connecting regulatory elements and transcriptional drivers to single-cell gene expression programs, significantly extending our understanding of human neurogenesis, cortical evolution, and the cellular basis of neuropsychiatric disease. We tie cell-cycle progression with early cell fate decisions during neurogenesis, demonstrating that differentiation occurs on a transcriptomic continuum; rather than only expressing a few transcription factors that drive cell fates, differentiating cells express broad, mixed cell-type transcriptomes before telophase. By mapping neuropsychiatric disease genes to cell types, we implicate dysregulation of specific cell types in ASD, ID, and epilepsy. We developed CoDEx, an online portal to facilitate data access and browsing.

Graphical Abstract


An atlas of human neocortical development
Human neocortex
Drop-seq
Cell types atlas
Define cell transition states during early neurogenesis
Cell type specific TF expression and TF-gene networks
TFs
TF
Cell type expression?
Genes
REs
Cell type activity?
Cellular basis of neuropsychiatric disease
De novo variants
Common variation
Cell Types
Genes causing risk for disease
GWAS SNP
Cell type enriched gene
Chromatin accessibility correlation (ATAC-seq)
Online resource
Welcome to CoDEx Viewer!
CoDEx Viewer: Cortical Development Expression Viewer

ETOC:

An extensive single-cell catalog of cell types in the mid-gestation human neocortex extends our understanding of early cortical development, including subplate neuron transcriptomes, cell type specific regulatory networks, brain evolution and the cellular basis of neuropsychiatric disease.

Keywords

cortical development; neurogenesis; human; subplate; autism; schizophrenia; intellectual disability; epilepsy; evolution; differentiation

Introduction:

The human cortex is composed of billions of cells estimated to encompass hundreds or thousands of distinct cell types each with unique functions (Silbereis et al., 2016). Groundbreaking work in mouse revealed the power of single-cell transcriptomics to provide a framework for understanding the complexity and heterogeneity of cell types in the brain (Hrvatin et al., 2018; Macosko et al., 2015; Saunders et al., 2018; Shekhar et al., 2016; Tasic et al., 2016; Zeisel et al., 2018). The availability of high-quality tissue and advances in single-cell transcriptomic technologies now permit us to catalog the cell type diversity of the human cortex in a comprehensive and unbiased manner (Ecker et al., 2017).

Despite the enormous progress that has been made in characterizing early cortical development (Geschwind and Rakic, 2013; Lui et al., 2011; Silbereis et al., 2016), many of the molecular mechanisms underpinning the generation, differentiation, and development of the diverse types of cells remain largely unknown (Molnar, 2011). Molecular taxonomies of cortical cell types from developing human brains enable us to understand the mechanisms of neurogenesis and how the remarkable cellular diversity found in the human cortex is achieved (Camp et al., 2015; Fan et al., 2018; Liu et al., 2016; Nowakowski et al., 2017; Pollen et al., 2015; Zhong et al., 2018). Several recent studies have taken a first step in this direction, analyzing several hundred, or a few thousand cells from developing human brain (Fan et al., 2018; Liu et al., 2016; Nowakowski et al., 2017; Pollen et al., 2015; Zhong et al., 2018). However, advances in technology and throughput [*e.g.* Drop-seq (Macosko et al., 2015)] now allow us to analyze an order of magnitude more cells to complement and extend these studies, providing a deeper picture of human cortical development and its perturbation in disease.

Results:

A catalog of cell types in developing human neocortex identifies major cell types, progenitor states, and subtypes of excitatory and inhibitory cells.

Here we use single-cell RNA sequencing (scRNA-seq) to define cell types and compile cell type transcriptomes in the developing human neocortex. We focus on the cortical anlage at mid-gestation (gestation week (GW) 17 to 18) (Figure 1A) because this period contains the major germinal zones and the developing cortical laminae containing migrating and newly born neurons, and neurodevelopmental processes occurring during this epoch are implicated in neuropsychiatric disease (de la Torre-Ubieta et al., 2016; Gandal et al., 2016). To optimize detection of distinct cell types, we separated the cortex into the germinal zones [ventricular zone (VZ) and subventricular zone (SVZ)] and developing cortex [subplate (SP) and cortical plate (CP)] prior to single-cell isolation. Using Drop-seq (Macosko et al., 2015), we obtained and compared high quality profiles for ~40,000 cells from human cortex (Figures 1A, S1A–B, and Tables S1–S3), and a small subset with microfluidics approaches (Fluidigm) for technical comparisons.

We first applied unbiased clustering based on stochastic nearest neighbor embedding (tSNE; see materials and methods) and spectral K-nearest neighbor graph based clustering (Butler et al., 2018), identifying 16 transcriptionally distinct cell groups. Cell types originated from the expected anatomical source, and clustered by biological cell type rather than batch or technical artifacts (Figures 1B–G and S1C–D). We identified multiple groups of cells at different stages of neuronal differentiation and maturation, corresponding to all known major cell types at this developmental time period (Figures 1B–F, S1E, and Table S4). Clusters contained between 50 and 2,000 cells. The smallest cluster captured, which belonged to microglia, was comprised of ~50 cells. Other small clusters for oligodendrocyte precursors, endothelia, and pericytes were comprised of 306, 237, 114 cells, respectively (Figure 1G and Table S4). Clusters were reproducible and robust as ascertained by bootstrapping (Figure S1F). Ordering of cells by pseudo-time in an unbiased manner using Monocle 2, a computational method that performs lineage trajectory reconstruction based on single-cell transcriptomics data, (Qiu et al., 2017; Trapnell et al., 2014) confirmed the predicted developmental trajectory (Figures 1H–I). For example, it is possible to observe the ordered transitions between different neural progenitor types and maturing glutamatergic neurons, with radial glia (RG) transitioning to intermediate progenitors (IPs), and IPs transitioning to newborn migrating neurons (Figure 1I).

We observed that cell type detection appears to be more sensitive to the number of cells profiled than sequencing depth (Figures S2A–D). Further, while each individual cell profile is an incomplete representation of that cell type (Lun et al., 2016) (Figures S2E–F), pooling transcriptomes within cells of a given type provides more complete cell type transcriptome representations. We iteratively subsampled cells from clusters to empirically assess the completeness of cell type signatures with different sample sizes (Figure S2G). At a depth of 40,000 cells, we obtain stable transcriptomes representing 3,000–5,000 genes for most of the cell types present (Table S4). Comparison of these data with a lower throughput, higher sequencing depth method (Fluidigm C1, (Nowakowski et al., 2017), Figure S3), revealed that the ability to leverage an order of magnitude more cells yielded more stable mRNA transcript profiles for a given cell type (Figures S3A–C). Integration of our dataset with the largest previous study (4,000 cells; (Nowakowski et al., 2017)) using canonical correlation analysis (Butler et al., 2018) showed substantial alignment of cells between the two datasets (Figures S3D–E). This is the first direct comparison of different human fetal single cell data sets and it demonstrates the reproducibility of these expression profiles, and significantly extends them by providing more stable gene expression rankings. We provide these cell-type specific expression profiles with annotated gene expression ranking confidence measures for each cell type (Table S4), and a web interface for browsing these data (http://geschwindlab.dgsom.ucla.edu/pages/codexviewer).

Comparison of scRNA-seq datasets to bulk RNA-seq expression profiles (de la Torre-Ubieta et al., 2018) showed consistently that gene expression profiles generated using different scRNA-seq methodologies across different laboratories strongly correlated with bulk RNA-seq gene expression profiles (Spearman 0.69–0.83) (Figures S4A–B). However, we did observe that approximately 400 protein coding genes representing longer, brain-enriched, cell adhesion molecules involved in neuronal development (Figures S4E–H) were consistently under-represented in single-cell datasets compared to bulk tissue RNA-seq (Figures S4C–D). Overall comparison of expression of canonical cell type marker genes showed similar expression levels compared to bulk tissue RNA-seq (Figure S4I) indicating that despite small biases in gene detection shared across scRNA-seq methods, the relative frequencies of major cell types were not over- or under-represented, further demonstrating the robustness of the scRNA-seq dataset (Figure S4).

We next reasoned we could use the depth of our dataset to identify cell states and cell sub-types not identified in previous studies with smaller cell numbers. We performed an additional round of clustering on each major cell cluster (Figure 2, Table S5, see materials and methods), which finely resolved maturation states during neurogenesis, and identified multiple cell sub-types not previously characterized in single-cell datasets in humans: SP neurons (Figure 2D), distinct subtypes of glutamatergic neurons (Figure 2D), early parvalbumin (PV) interneurons, and NPY expressing SST interneurons (Figure 2C). A previous study profiling ~2,300 cells from developing human cortex did not detect PV interneurons, and suggested PV interneurons may not develop until after GW26 (Zhong et al., 2018). Here, we find that at GW17–18 PV interneurons comprise ~0.1% of the total population, underscoring how necessary larger datasets are to identify rare cell types.

The provenance of human neocortical interneurons has been disputed (Hansen et al., 2013; Ma et al., 2013; Radonjic et al., 2014; Zhong et al., 2018). We observed no clusters of progenitors expressing markers of interneurons, and no clusters of interneurons expressing mitotic or progenitor markers (Figure S1E). In addition, sub- clustering of interneurons did not identify a cell population displaying characteristics of interneuron progenitors. OLIG2 is a marker of both medial ganglionic eminence progenitors and oligodendrocyte precursors (OPCs) (Miyoshi et al., 2007). We observed OLIG2+ cells only in the OPC cluster, which express other OPC markers, but do not express interneuron marker genes (Figure S1E). Thus, even with the order of magnitude greater cell depth and increased ability to detect low abundance cell types (*e.g.* 0.1%), we do not find evidence of a neocortical interneuron progenitor during mid-gestation in humans.

Cell type enrichment of TFs and co-factors.

We next sought to gain insight into cell-type specific regulatory programs by comparing TF expression across major cell types. We find previously characterized TFs and co-factors enriched in their corresponding cell types (Figures 1F and 3A) and multiple TFs and co-factors that have not been associated with specific neocortical cell types (Figure 3). These TFs also displayed laminae-specific expression in a bulk tissue laser capture micro-dissected (LCM) expression dataset (Miller et al., 2014), and temporal trajectories similar to canonical cell type markers (Figures 3B and S5A).

To validate predictions for these putative novel cell type markers, we performed RNA FISH, which confirmed laminae-specific expression of each of the TFs and co-factors tested: *ZFHX4* and *CARHSP1* in neural progenitors, and *CSRP2* in excitatory neurons (Figures 3C–H). Of particular interest was *ZFHX4*, which has been previously associated with 8q21.11 microdeletion syndrome (Palomares et al., 2011). Our data localizes ZFHX4 specifically to neural progenitors in the developing human neocortex for the first time (Figures 3A–E), implicating specific dysregulation of neural progenitors as the mechanism underlying this syndrome.

The TF ST18 appeared to partially cluster with SP markers (Figures 2D and 4A–B) (Hoerder-Suabedissen and Molnar, 2015; Oeschger et al., 2012). We found that SP markers previously defined in other species were not uniquely expressed in the SP in another fetal gene expression atlas (Miller et al., 2014) (Figure 4B, C). We identified SP enriched genes in this cortical laminar atlas (Figure 4C, see materials and methods), which showed strong overlap with ST18 (Figure 4B). Sub-clustering separated deep layer neurons from the ST18-expressing SP neurons (Figure 4D). Genes enriched in the SP neuron cluster, or highly correlated with ST18 display strong SP enrichment in the cortical laminae dataset, verifying our capture of SP neurons and identification of many additional SP neuron markers (Figures 4E–H). Additionally, we performed RNA FISH to confirm SP specific expression of ST18 (Figure 4I). This represents the first transcriptomic characterization of human SP neurons at single-cell resolution.

We next reasoned that we could begin to leverage these single cell data to uncover some of the cellular and molecular mechanisms driving human cortical evolution by determining whether specific cell types were enriched with genes showing human specific expression trajectories (hSET) in bulk tissue (Bakken et al., 2016) (see materials and methods). We observed the strongest enrichment of hSET genes in outer RG (oRG) and the excitatory upper layer enriched cluster (Figure S5B). This is notable, since both of these cell types represent processes central to both neocortical expansion (Lui et al., 2011) and the elaboration of cortical connectivity in humans (Fame et al., 2011). Among the approximately 600 genes with oRG-enriched expression, we identified *LYN*, a Src tyrosine kinase previously implicated in neuronal polarization and AMPA signaling (Hayashi et al., 1999; Namba et al., 2014), which had not been previously associated with this cell type We used a fetal LCM atlas (Miller et al., 2014) and RNA FISH to further validate these observations, showing that LYN localized to the germinal zones and was specifically expressed in the VZ and oSVZ (Figure S5C–D).

Mapping of cell-type specific gene regulatory networks in the developing human neocortex.

To deconvolute the cell type specificity of regulatory elements, we leveraged a recently generated map of regulatory elements active in developing fetal cortex and their putative target genes (de la Torre-Ubieta et al., 2018) to identify promoters and enhancers regulating the expression of genes enriched in cells defined in this study (see materials and methods) (Table S6). Enhancers associated with specific cell types were characterized by remarkable consistency in mean enhancer size, number associated with each gene, and distance to the target gene for each cell type (Figures 5A–F). In addition, there was no correlation between target gene length or GC content and number of associated enhancers (Figures 5G–H). We extended this map by computationally reconstructing gene regulatory networks using the SCENIC pipeline (Aibar et al., 2017) (Figure 5I) with empirically determined regulatory elements (de la Torre-Ubieta et al., 2018), rather than standard promoter annotations. This produced 124 regulons, each representing a TF along with a set of co-expressed and motif enriched target genes, and the regulon activity scores for each cell (Table S7). Multiple TFs previously associated with specific cell types showed enriched regulon activity in the expected cell types (Figures 5J–K and Table S7). We also identified TFs with previously uncharacterized cell-type or cell subtype specific activity, including *NFE2L2* in RG, *NHLH1* in post-mitotic IPs, *ZNF354C* in excitatory neurons, and *BACH2* in maturing excitatory neurons (Figure 5L and Table S7). This represents a first- generation map of cell-type specific gene regulatory networks in the developing human neocortex.

Dissecting the acquisition of a neuronal program.

Neurons are generated from the controlled asymmetric division of neural progenitors, which prompted us to analyze the distinct transcriptional states of cycling cells during this process (Lui et al., 2011). Neural progenitors clustered by cell cycle state in addition to cell type (Figures 1E, 6A, and S6A–C), with about 30% of progenitors cycling, roughly consistent with previous observations (37% based on immunostaining) (Hansen et al., 2010). Remarkably, we also observed that many of the cycling progenitors individually expressed markers of several distinct major cell types, including RGs, IPs, and neurons (Figures 6A–C). Doublets were an insufficient explanation for the co-expression of distinct cell type makers for multiple reasons, including that the number of cells expressing multiple major cell type markers is twice the empirically assessed doublet rate (Table S2 and Figure S1B—also see materials and methods) and the highly non-random distribution of the cell types expressing markers of two cell types (Figure 6C).

Therefore, as an alternative explanation, we hypothesized that we were identifying an intermediate or transition state: mitotically active cells in the early stages of neurogenesis, *i.e*. RG producing IP, RG producing neurons, and IP producing neurons. Consistent with this hypothesis, mixed marker cells progressing through different stages of the cell cycle consistently displayed transcriptomes comprised of multiple major cell types (Figures 6D and S7A–B). By S-phase, RG+IP+ and IP+Neuron+ cells more closely resembled their presumed endpoint cell type, IP and neuron, respectively (Figures 6D and S7B). The transcriptomic signature of RG+Neuron+ S-phase and G2/M phase cells was closer to RG, potentially reflecting the greater dissimilarity between RG and neurons (Figures 6D and S7A–B). In addition, the mixed marker cells share a high percentage of the end point cell type signature, but the magnitude of expression of the cell type relevant signature genes is smaller than in cells in the fully differentiated cell clusters (Figures 6E and S7C–F). Mixed marker cells not in S, G2, or M phase may represent cells starting to cycle and differentiate, consistent with findings in mice that some RG precursors also express neuronal marker genes of both deep and superficial layers, representing transcriptionally primed cells (Zahr et al., 2018). Alternatively, these mixed marker cells may be newborn cells that still retain some transcripts of the mother cell type, as has been previously suggested in mice (Zahr et al., 2018; Zhong et al., 2018).

To independently validate the existence of cells in these transition states, we performed RNA FISH, observing S-phase neural progenitors in the VZ expressing both PAX6 and STMN2, indicating an induction of a neuronal program in a cell before its neurogenic division (Figure 6F). Indeed, 8.9% (VZ), 6.7% (iSVZ) and 7.5% (oSVZ) of these cells co-express markers of RG and neurons (Figure 6G), confirming our scRNA-seq data (see materials and methods). We were able to quantify the relative proportions of progenitors undergoing distinct differentiation divisions (Figure 6H), finding that RG produce roughly equal numbers of RG, IP, and neurons, but that IP produce approximately two times as many neuronal progeny as IP progeny. Taken together, these results indicate that during early neurogenesis: 1) Cell fate decisions occur prior to S-phase; 2) Differentiating “parent” cells not only express the few key TFs that drive cell fates, but express broad, mixed cell type transcriptomes; 3) Neural cell-type differentiation occurs on a continuum and involves transcriptomic transitions tied to cell cycle progression (Figure 6I).

Cellular determinants of disease.

We next reasoned that we could use this atlas of developing human brain cell types to identify the developmental stages and cell types where mutations causing high risk for neuropsychiatric disease act, so as to provide a reference for understanding disease mechanisms and circuits (Figure 7). We first examined enrichment of high confidence autism spectrum disorder (ASD) risk genes, defined by harboring high risk likely protein-disrupting mutations (Sanders et al., 2015) (Figures 7A, 7D, and S8A). The majority of ASD-risk genes were expressed in developing glutamatergic neurons, both deep and upper layer (Figures 7A and 7D), consistent with previous studies (Amiri et al., 2018; Parikshak et al., 2013). However, at the individual gene level, there is substantial variability, and several genes are expressed in inhibitory neurons as well as excitatory neurons or progenitors (Figures 7A and S8A). For example, MYT1L and AKAP9 display pan-neuronal expression, whereas GRIN2B is glutamatergic subtype specific, and ILF2 is expressed in cycling progenitors (Figures 7A and S8A). In adult, expression again concentrated in glutamatergic neurons, with some genes exhibiting more pan-neuronal expression patterns (Figure S8A).

In addition, our expanded atlas of cell types identified several genes that showed remarkably distinct patterns of extra-neuronal expression, including SLC6A1, which was enriched in pericytes, and TRIO, SETD5, TCF7L2, and KAT2B, which were enriched in oligodendrocyte precursors (Figures 7A and S8A). For the first time these data suggest that cell types involved in maintenance of the blood brain barrier and the peri-neural environment may also mediate ASD risk. It should be noted that several of these genes are expressed in different cell types in adult, such as SLC6A1 in interneurons, highlighting the importance of broader single-cell catalogs (Figure S8A). Expanding this analysis to high confidence intellectual disability (ID) and epilepsy risk genes, (Figures 7B–D and S8B–C) showed that most epilepsy risk genes are expressed in glutamatergic neurons (Figures 7B, 7D, and S8B). ID risk genes were also enriched in glutamatergic neurons, but also showed enrichment in RG, which was not observed with ASD or epilepsy (Figures 7C–D, and S8C). The impact on early progenitor types in ID relative to ASD and epilepsy is consistent with the more severe disease phenotype in ID. Although the results for ID were highly significant, It should be noted that the ID risk gene list is smaller, making the comparisons less powered. Taken together, these results demonstrate cell-type specific expression of ASD, epilepsy, and ID risk genes by mid fetal development and provide a framework for the cellular and developmental context in which individual ASD, epilepsy, and ID genes should optimally be studied.

The majority of neuropsychiatric disease risk loci are found in the non-coding genome, where functional interpretation is hampered by limited knowledge of the genomic location and spatiotemporal activity of regulatory elements. Leveraging our cell-type specific map of regulatory elements active in the human neocortex (Figure 8A and Table S6, see materials and methods), (de la Torre-Ubieta et al., 2018) we used a partitioned heritability approach based on LD score regression (Finucane et al., 2015) to identify cell types enriched for variants influencing brain volume, cognition, or causing risk for neuropsychiatric disease. We found that variants influencing adult intracranial volume (Adams et al., 2016) were specifically enriched in the regulatory elements of cycling progenitors (PgS, PgG2M), pinpointing a specific cell type and state likely associated with neural progenitor expansion (Figures 8B–C). Importantly, by connecting causal genetic drivers to specific genes within a specific cell type, this not only identifies putative cell-type specific mechanisms involved in cortical expansion, but provides further support for the radial unit hypothesis of cortical expansion on the human lineage (Lui et al., 2011; Rakic, 1995).

In contrast, common genetic variants influencing educational attainment (Edu) (Okbay et al., 2016) were enriched in cycling neural progenitors, cortical plate glutamatergic neurons, MGE derived interneurons, and intriguingly, in pericytes (Figures 8B–C). A less powered IQ genome-wide association study (GWAS) (Sniekers et al., 2017) also found enrichment in maturing cortical plate glutamatergic neurons, but not in other cell types (Figures 8B). Unfortunately, most of the psychiatric disease GWAS remain underpowered (n = ~46,000 and ~34,000 for ASD and epilepsy respectively). However, variants causing risk for schizophrenia (n = ~105,000) (Pardinas et al., 2018) were enriched in multiple cell types, including neural progenitors, glutamatergic neurons, interneurons, oligodendrocyte precursors and microglia (Figures 8B–C). A recent study, using a partitioned heritability approach, found enrichment for schizophrenia variants in adult cortical glutamatergic neurons and cortical interneurons, consistent with bulk tissue analysis (Horvath and Mirnics, 2015), but was unable to assess enrichment in human fetal cortical cell types given a lack of available data (Skene et al., 2018). Our results implicate neural progenitors, oligodendrocyte precursors and fetal microglia in schizophrenia, highlighting the importance of generating single-cell resources from multiple time periods and brain regions. Given the complex etiology and phenotypic diversity of schizophrenia it may be expected that multiple cells are affected. These results highlight how combining DNA accessibility profiling and single-cell sequencing can facilitate interpretation of the function of variants influencing brain structure and function.

Discussion:

This resource of transcriptomic profiles of 40,000 single cells in human fetal cortex demonstrates the utility of single-cell analysis for characterizing human neurogenesis, identifying novel cell type regulatory mechanisms, and for understanding the cellular basis of brain phenotypes with neurodevelopmental origins. By expanding the publicly available number of human fetal brain single-cell transcriptomes by an order of magnitude, these data provide a high-resolution map of expression profiles for all known major cell types from mid-gestation human brains with more complete cell type specific mRNA transcript profiles than previously available. To facilitate sharing, exploration, and use of this unique and valuable resource we developed a powerful and easy to use online browser that allows rapid queries to ascertain cell type specific expression patterns presented in an intuitive graphical interface. We leverage the breadth and robustness this high depth catalog of neocortical cell types to characterize rare cell types and states, including PV interneurons, NPY expressing SST interneurons, and subplate neurons (Fan et al., 2018; Liu et al., 2016; Nowakowski et al., 2017; Pollen et al., 2015; Zhong et al., 2018). We show that most markers for subplate cells identified in other species are not specific to subplate in human, and provide a new cadre of marker genes for this important cell class that has expanded substantially on the primate lineage (Hoerder-Suabedissen and Molnar, 2015).

Some of the rare cell states identified include transitional forms that we validate via in situ hybridization. Characterization of these rare cell states provides novel insight into neurodevelopmental cell dynamics. Specifically, our data implicate early decision points in cell fate trajectories that are pre S-phase, leading to transcriptomically mixed cell states prior to their division into two distinct cell types. An early cell fate decision point tied to cell cycle is consistent with previous work indicating cell fate decisions in neurogenesis are made in G1 (Lange et al., 2009; Pilaz et al., 2009). However, previous models of asymmetric neurogenic divisions suggest that only a few key TFs of the “daughter” lineage are expressed in the asymmetrically dividing cell, whereas we observe early induction of more extensive cell type transcriptional programs (Bertrand and Hobert, 2010; Pfeuty, 2015). This is particularly surprising in that cells are expressing transcriptomes of two distinct cell types, prior to telophase. In addition, the transition state dynamics during early neurogenesis show that cell type differentiation is on a gradual continuum and involves transcriptomic transitions tied to cell cycle progression, rather than off or on expression of a small group of TFs.

We also perform a systematic exploration of the impact of multiple technical factors including comparisons across methods and to bulk tissue RNA-seq, enabling us to thoroughly evaluate gene and cell type detection and coverage. Surprisingly, these types of analyses have not been performed in most published papers to date. By direct comparison with bulk tissue data, we show that there is a high level of correspondence between the transcriptomes identified in both single cell and bulk tissue data. However, we do find that single cell transcriptomes do miss a small number of certain genes, biased towards long neuronal enriched transcripts such as cell adhesion molecules, an observation that has not been noted in previous studies (Nowakowski et al., 2017; Pollen et al., 2015; Zhong et al., 2018). In addition, we evaluate and provide relatively high confidence cell type transcriptomes relative to previously published lower-throughput, higher sequencing depth methods. Finally, we provide the first direct comparison and integration of different human fetal brain single cell data sets, demonstrating the reproducibility of these methods for identifying cell clusters.

By integrating these data with tissue specific regulatory information we provide a map of TF gene regulatory networks for specific cell types in developing human brain. We highlight how this can be used to identify critical cell types in monogenic disorders (e.g. ZFHX4 and 8q21.11 deletion), as well as in ASD, expanding the implicated cell landscape in this disorder to include inhibitory neurons and in a few cases, non-neural cells in addition to glutamatergic neurons. These results emphasize the importance of expanding single-cell taxonomies to include single-cell epigenetic analysis (Luo et al., 2017). Lastly, we show that genes with human specific expression patterns act preferentially in oRG and upper cortical layer neurons, which is consistent with the expansion of these zones during brain evolution. These data provide a molecular context for cortical expansion and increased cortical-cortical connectivity in humans, and extend our understanding of developmental dynamics and the origin of neuropsychiatric disease risk in human neocortex.

STAR methods:

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents should be directed to and will be fulfilled by the Lead Contact, Daniel H. Geschwind (dhg@mednet.ucla.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Developing human brain tissue samples—De-identified fetal tissue samples were obtained from the UCLA Gene and Cell Therapy Core according to IRB guidelines or from the University of Maryland Brain and Tissue Bank (RNA FISH). For all donors profiled for single-cell RNA-seq (Drop-seq), DNA was acquired from fetal brain tissue and donors were genotyped with Illumina HumanOmni2.5 chips. Sex was determined based on homozygosity in X chromosome SNPs (3 male, 1 female) and expression of XIST and Y chromosome genes. High confidence CNVs were called with plumbCNV (Cooper et al., 2015). No known major pathogenic CNVs implicated in neuropsychiatric disorders were found in these donors. The largest CNV called and confirmed visually was 380kb. Samples processed for Dropseq were obtained from 4 donors [female 17,17,18 gestation weeks (GW); male: 18 GW]. Samples processed for Fluidigm were obtained from 2 donors [female 17, 17.5 GW]. Samples processed for RNA fluorescent in situ hybridization (RNA FISH) were obtained from the UCLA Gene and Cell Therapy Core or from the University of Maryland Brain and Tissue Bank according to IRB guidelines from five donors aged GW15.5–18. This study was performed according to the legal and institutional ethical regulations of the UCLA Office of Human Research Protection. Full informed consent was obtained from all of the parent donors.

METHOD DETAILS

Tissue dissection and single-cell isolation—Coronal sections were prepared from fetal cortices using a razor blade under a dissection microscope in ice-cold Hank's Balanced Salt Solution (HBSS). The coronal sections were then further dissected at the intermediate zone (IZ) to divide them into two regions: 1) consisting of the germinal zones (GZ) [ventricular zone (VZ), and subventricular zone (SVZ], and 2) consisting of the developing cortex (CP) [subplate (SP), cortical plate (CP), and marginal zone (MZ)]. The majority of the IZ was included as part of the "CP" dissection, but there is likely a small amount included in the GZ. Following dissection, GZ and CP sections were separately gently dissociated via enzymatic digestion with papain (Worthington) and filtered into a pure homogeneous cell suspension through dual filtering with a 40μm strainer followed by an ovomucoid gradient (Worthington). Cell survival (90–95%) and yield were quantified with Trypan blue staining, before immediately proceeding with Drop-seq or Fluidigm single-cell isolation. To assess doublet rates by human-mouse cell mixing experiments, mouse E15 cortical cultures were prepared in parallel. Briefly, mouse cortices were dissected in ice-cold HBSS and enzymatically dissociated with trypsin into a homogeneous cell suspension. Survival (90–95%) and yield were quantified with Trypan blue staining. Mouse and human cells were mixed in a 1:10 ratio immediately prior to single-cell isolation by Drop-seq.

Single-cell RNA-seq—Drop-seq was run on single cells according to the online Drop-seq protocol v.3.1 (http://mccarrolllab.com/download/905/) and the methods published in Macosko *et al.* (Macosko et al., 2015). Cells were maintained on ice and diluted to 125,000/mL in PBS + 0.01% BSA immediately prior to isolation. Barcoded beads were obtained from Chemgenes and cells were isolated in a Polydimethylsiloxane (PDMS) microfluidics device. Libraries were prepared with the Nextera XT DNA Library Preparation Kit (Illumina) according to the manufacturer's instructions. Libraries were then sequenced to an average of 57,814 reads/cell in an Illumina HiSeq2500 instrument with a modified 100bp paired-end protocol where R1 =25bp and R2 =75bp to maximize mapping. This read depth was empirically determined to yield the best per cell gene detection versus sequencing depth.

Fluidigm C1 scRNA-seq was run using the Fluidigm low-throughput small IFC (96 cells) or the high-throughput small IFC (800 cells) according to the manufacturer's instructions. Libraries were then prepared with the Nextera XT DNA Library Preparation Kit (Illumina) according to the manufacturer's instructions and sequenced to an average of 414,411 (high-throughput) and 682,569 (low-throughput) reads/cell in an Illumina HiSeq2500 instrument with a modified 100bp paired-end protocol where R1 =25bp and R2 =75bp to maximize mapping. The low-throughput libraries were sequenced in an Illumina HiSeq2500 where R1 =50bp and R2 =50bp. The high-throughput libraries were sequenced in an Illumina HiSeq3000 where R1 =12bp and R2 =126bp.

RNA fluorescent in situ hybridization (RNA FISH)—In order to independently validate single-cell expression profiles, we used RNAscope (Wang et al., 2012), an RNA FISH technique capable of single-molecule RNA detection with minimal off-target signal. Fetal tissue samples were obtained from the UCLA Gene and Cell Therapy Core or from the

University of Maryland Brain and Tissue Bank according to IRB guidelines from five donors aged GW15.5–18. Developing human cortices were flash-frozen, embedded in OCT and cryosectioned in the coronal plane (15μm section thickness). Sections were then subjected to RNA FISH following the manufacturer's protocol for fresh frozen tissues using the Fluorescent Multiplex Assay Kit v1 (Advanced Cell Diagnostics Cat# 320850) with the following probes: CARHSP1-C1, CSRP2-C2, CRYAB (Cat# 426271-C2), EOMES (Cat# 429691-C3), LYN-C3, PAX6 (Cat# 588881-C1), PAX6 (Cat# 588881-C2), PCNA (Cat# 553071-C1), SATB2 (Cat# 420981-C1), ST18-C2, STMN2-C3, and ZFHX4-C2.

To determine the expression pattern across cortical layers for cell-enriched TFs (Figures 3C–G and 4I), tiled images of multiple coronal sections from three independent donors spanning the entire cortex were acquired using a Leica DMi8 epifluorescence microscope at 40X magnification. Fluorescence analyses were performed in ImageJ version 2.0.0. For each image, regions of interest (ROIs) outlining individual cortical layers were manually created based on nuclear packing as visualized by DAPI staining. Background fluorescence was subtracted using a mask based on an empirically determined thresholded value for each image. To account for changes in cell density across cortical layers, the background-corrected signal for each channel was normalized to the DAPI intensity. The normalized fluorescence intensity is the mean gray value of each RNAFISH channel for each cortical layer divided by the mean DAPI gray value for the corresponding layer. DAPI-corrected values outside of three standard deviations of the mean were removed.

To quantify the co-expression of RG and neuronal markers in S-phase cells, images of coronal sections probed with the RG marker PAX6, the S-phase marker PCNA and the neuronal marker STMN2 were acquired on a Zeiss LSM780 confocal using a 63X magnification objective. Confocal tiled images of developing human cortex encompassing the VZ and SVZ were acquired for the entire thickness of the section at a 0.29μm Z-step size. For each cell quantified, the presence of puncta for each marker gene was determined across the entire Z-plane. Only cells expressing all three markers in the same Z-plane overlapping the DAPI staining were considered to be an RG-Neuron transition state. A total of 5,692 cells were quantified from two donors. Similar image acquisition and analysis was performed to quantify cells co-expressing PAX6 or EOMES, and ZFHX4. A total of 4,723 cells were quantified from three donors.

QUANTIFICATION AND STATISTICAL ANALYSIS

Alignment and processing—The raw Drop-seq data was processed using the Drop-seq tools v1.12 pipeline from the McCarroll Laboratory (http://mccarrolllab.com/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf). Reads were aligned to the Ensembl release 87 *Homo sapiens*genome. We calculated unique molecular identifier (UMI) counts for each gene of each cell by collapsing UMI reads using Drop-seq tools.

The raw Fluidigm C1 data was processed using a custom pipeline. Cell barcode demultiplexing and initial processing was performed with Fluidigm mRNASeqHT_demultiplex.pl v1.0.2. Raw reads were aligned to the Ensembl release 75 *Homo sapiens* genome with RNA STAR (Dobin et al., 2013). Aligned reads were sorted and alignments mapping to different chromosomes were removed from the BAM file using samtools (Li et al., 2009). Gene expression levels were quantified using HTSeq with a union exon model (Anders et al., 2015).

Quality control statistics were collected using RNA STAR statistics, Drop-seq tools metrics, and PicardTools (commands ReorderSam, CollectAlignmentSummaryMetrics, CollectRnaSeqMetrics, CollectGcBiasMetrics) and samtools (duplication metrics).

Assessment of doublet rate—Doublet rate, *i.e*. the frequency of which more than one cell was captured in a single Drop-seq droplet, was assessed by species mixing experiments (Figure S1 and Table S2). Overall doublet rate (mouse+mouse, human+human, human +mouse) is derived based on the frequency of beads associating to both mouse and human cells in a single drop or well. For the species mixing experiments, cells from mouse E15 cortical cultures were added to human cells at a concentration of 1:10 as described above. The raw Drop-seq data was processed using the Drop-seq tools v1.12 pipeline from the McCarroll Laboratory (http://mccarrolllab.com/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf). Reads were aligned to a mixed species reference genome (*Homo sapiens* and *Mus musculus*) obtained from GEO GSE63269. The BAMs were then filtered into two organism specific BAMs using the Drop-seq tools command 'FilterBAM'. For each species-specific BAM, UMI counts were then calculated for each gene of each cell by collapsing UMI reads using Drop-seq tools.

Filtering and normalization—To select Drop-seq cells for downstream analysis: 1) Cells were selected for downstream analysis using the cell barcodes associated with the most UMIs. We estimated the number of cells captured as 5% of the input beads and retained this many cell barcodes for downstream analysis. 2) For samples with mouse cells spiked in, mouse cells were removed by filtering all cells with > 250 UMIs mapping to the mouse genome. 3) Removed cells with <200 unique genes detected (gene detection: ³1 count). 4) Removed cells with >3 standard deviations above the mean number of genes detected (3152). 5) Removed cells with >5% of their counts mapping to MT genes. 6) Removed genes detected in <3 cells.

Normalization was performed using Seurat (v2.3.4 (Butler et al., 2018). Briefly, raw counts are read depth normalized by dividing by the total number of UMIs per cell, then multiplying by 10,000, adding a value of 1, and log transforming (ln (transcripts-per-0,000 + 1)) using the Seurat function 'CreateSeuratObject'. Raw UMI counts data were assessed for the effects from biological covariates (anatomical region, donor, age, sex), and technical covariates (library batch, sequencing batch, number of UMI, number of genes detected, CDS length, GC content) (Figure S1). The effects of number of UMI (sequencing depth), donor, and library preparation batch were removed using a linear model from the read depth normalized expression values (custom R scripts, lm(expression ~ number_of_UMI + donor + lab_batch), and Seurat function 'ScaleData').

Single-cell clustering and visualization—Clustering was performed using Seurat (v2.3.4) (Butler et al., 2018). Read depth normalized expression values were mean centered and variance scaled for each gene, and the effects of number of UMI (sequencing depth), donor, and library preparation batch were removed using a linear model with Seurat

(‘ScaleData’ function). Highly variable genes were then identified and used for the subsequent analysis (Seurat ‘MeanVarPlot’ function). Briefly, average expression and dispersion are calculated for each gene, genes are placed into bins, and then a z-score for dispersion within each bin is determined. Principal component analysis (PCA) was then used to reduce dimensionality of the dataset to the top 40 PCs (Seurat ‘RunPCA’ function). Clustering was then performed using graph based clustering implemented by Seurat (‘FindClusters’ function). Briefly, a K-nearest neighbor graph based on Euclidean distance in PCA space is constructed from the PC scores for each cell. Edges between cells are weighted based on shared overlap in neighborhoods determined by Jaccard distance. Cells are then iteratively grouped together with the goal of optimizing the density of links inside communities as compared to links between communities. Cell clusters with fewer than 30 cells were omitted from further analysis.

For visualization, t-distributed stochastic neighbor embedding (tSNE) coordinates were calculated in PCA space, independent of the clustering, using Seurat (‘RunTSNE’ function). tSNE plots were then colored by the cluster assignments derived above, gene expression values, or other features of interest. Gene expression values are mean centered and variance scaled unless otherwise noted.

For sub-clustering analysis an iterative approach was used (Figure 2), cells from each initial cluster were re-processed, clustered, and analyzed from the raw counts matrix using Seurat as described above. For larger clusters (>1,000 cells) the top 10 PCs were used, and for smaller clusters (<1,000 cells) the top 5 PCs were used.

Cluster stability—To assess cluster stability, we adapted the approach from Hennig *et al.*using bootstrapping and the Jaccard index (Hennig, 2007). Briefly, a bootstrap sample of cells is drawn with replacement from the original dataset, then re-quality filtered, normalized, analyzed, and clustered, then this process is repeated over 100 iterations. For each iteration, the maximum Jaccard index is computed between the new clustering and the original clustering. The mean Jaccard index of 100 iterations of bootstrapping is reported (Figure S1F). Jaccard Index values range from 0–1, with >0.5 indicating stable clustering.

Differential gene expression analysis and cell type enrichment—In general, differentially expressed genes between different cell groups were determined using a linear model implemented in R as follows: lm(expression ~ number_of_UMI + donor + lab_batch). P-values were then Benjamini-Hochberg corrected. To identify cell type enriched genes, differential expression analysis was performed for each cluster individually versus all other cells in the dataset for genes detected in at least 10% of cells in the cluster. Genes were considered enriched if they were detected in at least 10% of cells in the cluster, 0.2 $\log_2$fold enriched, and Benjamini-Hochberg corrected p-value < 0.05 (Table S6).

Pseudo-time analysis—Monocle 2.0 was used to construct single-cell pseudo-time trajectories (Qiu et al., 2017; Trapnell et al., 2014). First, the dataset was subset to cells in Seurat clusters inferred to be part of the neurogenesis differentiation axis (progenitor and excitatory neuron clusters) (Figure 1I). The subset dataset was then run through the Monocle 2.0 pipeline beginning with raw counts. Dispersed genes to use for pseudo-time ordering were calculated using the 'estimateDispersions' function and required to be expressed in at least 10 cells. DDRTree was used to reduce dimensions and the effects of number of UMI (sequencing depth), donor, and library preparation batch were corrected for (Monocle function: reduceDimension(mo_filtered, max_components = 30, residualModelFormulaStr = "~number_of_UMI+donor+lab_batch"). The visualization function 'plot_cell_trajectory' was used to plot the minimum spanning tree on cells.

Stability of cluster gene expression signatures—Cluster gene expression signatures were evaluated by the stability in mean gene expression level ranking. Gene expression ranking was determined by mean expression level for each gene across all cells in the cluster. A bootstrapping approach was then used to evaluate stability of the gene expression rankings. Cells from each respective cluster were sampled with replacement over 1,000 iterations. At each iteration, a sample population was drawn pseudo-randomly and the mean expression level across the population was calculated and then genes were ranked by mean expression level, with a ranking of 1 being the most highly expressed. The mean of the rankings over all iterations and the standard deviation were plotted (Figures S2G and S5B–C).

Alignment of single-cell datasets—Datasets were aligned using Seurat canonical correlation analysis (Butler et al., 2018). First, read depth normalized expression values were mean centered and variance scaled for each gene. Then the 2,000 most highly variable genes were identified for each dataset using the Seurat 'FindVariableGenes' function. Next, canonical correlation analysis was run using the union of the variable gene sets as described in (Butler et al., 2018). The analysis returns canonical correlation vectors (CCV) across both datasets, which are then used to align the datasets. We used the top 20 CCVs to run the alignment procedure. The datasets were then aligned using the Seurat 'AlignSubspace' function, which utilizes CCVs and a nonlinear time warping algorithm to align metagenes between datasets. We then applied t-SNE to reduce dimensionality to plot the aligned datasets.

Comparison to bulk tissue RNA-seq—Bulk tissue RNA-seq samples from human fetal neocortex (GW17–19) GZ (n=9) and CP (n=9) were obtained from de la Torre-Ubieta *et al.* (de la Torre-Ubieta et al., 2018). GZ and CP dissections were carried out as described above. Bulk tissue RNA-seq samples were read depth normalized to counts per million (CPM). For comparison to single-cell RNA-seq data from a different laboratory, we obtained Fluidigm C1 generated raw sequencing data from human fetal brain from Pollen *et al.* (Pollen et al., 2015). For all single-cell datasets, single-cell expression profiles were pooled by aggregating gene expression counts across groups of cells to simulate bulk tissue RNA-seq samples. To aggregate or pool gene expression counts, groups of cells were randomly drawn from the single-cell dataset, and raw counts were summed across each group of cells for each gene to pool the expression profiles. Each pooled expression profile was then read depth normalized to CPM.

To compare pooled samples of different sizes to bulk tissue RNA-seq by correlation in gene expression values, 1) Samples of different numbers of cells were drawn, and counts were summed across the cells for each gene to make pooled samples of single-cells. 2) Gene expression levels (summed counts) of the pooled single-cell datasets were then correlated to the mean CPM of the bulk tissue RNA-seq dataset (Figure S4B).

To identify genes under-represented in single-cell RNA-seq compared to bulk tissue RNA-seq, we identified genes with higher or lower relative expression in the pooled single-cell expression profiles compared to bulk tissue RNA-seq. Genes greater than two standard deviations from the mean relative expression level of pooled versus bulk were labeled as under or over-represented in the respective single-cell RNA-seq dataset. To assess biases in capture of different cell types with Drop-seq, the expression of groups of cell type marker genes in the pooled Drop-seq dataset were compared to expression in the bulk tissue RNA-seq dataset. The expression ratios of pooled Drop-seq versus bulk tissue RNA-seq were converted to a z-score for plotting (Figure S4I).

Gene Ontology enrichment analysis was performed using g:Profiler (Reimand et al., 2016).

Cell type enrichment of TFs and co-factors—TFs, co-factors, and chromatin remodelers were obtained from AnimalTFDB 2.0 (Zhang et al., 2015). Genes were considered enriched in a major cell type if they were >0.4 $\log_2$fold enriched for any cluster corresponding to cells of that type, and were <0.25 $\log_2$fold enriched for any other cluster (Figure 3A). For example, radial glia (RG) enriched genes are >0.4 $\log_2$fold enriched in either or both the ventricular radial glia (vRG) and outer radial glia (oRG) cluster, and <0.25 $\log_2$fold enriched in any other cluster.

Subplate markers—To derive a human SP set of markers across mid-gestation, we used the fetal LCM laminae dataset (Miller et al., 2014) to identify SP enriched genes (Figures 4B–D). Genes were sorted by fold change of SP versus the VZ, IZ, CP, and MZ, using the Brainspan online tool (http://www.brainspan.org/lcm/search/index.html), and then manually curated for SP specificity.

Cell cycle analysis—Cell cycle state was determined by mean expression of groups of cell cycle stage marker genes obtained from Macosko *et al.* (Macosko et al., 2015) (Figures S6A–C). Two methods for cell cycle normalization were tested: 1) All cell cycle stage marker genes were excluded from the highly variable genes used for PCA, tSNE, and Seurat clustering described above. 2) Cell cycle correction by removing the effects of cell cycle state using a linear model via Seurat. First, each cell is assigned a S-phase and G2/M phase score using Seurat's 'CellCycleScoring' function. Then the cell cycle score is regressed out along with number of UMI (sequencing depth), donor, and library preparation batch using a linear model as described above.

Transition state analysis—Cells were considered positive for markers of a cell type if the mean expression of a group of cell type marker genes was >0.5 log normalized expression, *e.g.* RG+Neuron+cells express RG marker genes at a mean expression level >0.5 log normalized expression and neuronal marker genes at a mean expression level >0.5 log normalized expression.

Cell type gene signatures were determined using two methods: 1) Differential expression of cells in a type versus cells in another type, *e.g.* to determine an RG signature and newborn neuron signature differentiating the two cell types, differential expression of RG cells from RG clusters (oRG and vRG cluster) versus newborn migrating excitatory neurons (ExN cluster). 2) Enrichment of genes in a cell type. For each cluster, genes enriched in the cluster were determined as described above. Then the union of enriched genes for all clusters in a cell type was taken. The RG signature is the union of genes enriched in vRG and oRG clusters, the intermediate progenitor (IP) signature is genes enriched in the IP cluster, and the Neuron signature is genes enriched in the migrating excitatory neuron cluster.

Transcriptomic analysis of cycling mixed marker cells used cells dual positive for RG, IP, or neuronal markers specifically in the S-phase or G2/M phase clusters, *e.g.* RG+Neuron+ S-phase cells are cells from the S-phase cluster and dual positive for RG and neuronal markers. Dual positive cells were then compared to single marker type positive cells, *e.g.* RG+Neuron + S-phase cells were compared to RG+ Neuron- IP- and RG- Neuron+ IP- cells to ensure comparison to cells of a clear transcriptomic type. The eigengene of the cell type signatures was then calculated for each cell positive for cell type markers in an expected differentiation trajectory, *e.g.* to explore the RG to neuronal transition the Neuron eigengene was calculated using the neuron signature across RG+ Neuron- IP- negative cells, RG+ Neuron+ cells, and RG- Neuron+ IP- cells. Cell type signatures were determined as described above.

The amount of overlap and the magnitude of expression of the cell type gene signature of cycling mixed marker cells was compared to the end point cell type. The gene signatures of the cell types involved in a differentiation trajectory were compared by fold change of the gene signature in the beginning state cell type to the endpoint cell type. For example, for the RG to neuron transition, the RG and neuronal gene signatures are determined by differential expression of cells from the RG to the newborn neuronal cluster. The expression of the neuronal gene signature genes is then compared in RG+ cells to RG+Neuron+ cells. The percent of the neuronal genes that are more highly expressed in RG+Neuron+ cells versus RG+ cells is ascertained, this is the percent of shared genes. The mean fold change of the neuronal genes in RG+Neuron+ cells versus RG+ cells is also determined and compared to the magnitude of the mean fold change in RG+ versus Neuron+ cells, this is the percent of fold change.

Cell-type specific regulatory elements—A map of regulatory elements active in developing fetal cortex generated from chromatin accessibility data (ATAC-seq) (Buenrostro et al., 2013) was obtained from de la Torre-Ubieta *et al.* (de la Torre-Ubieta et al., 2018). Promoter elements were identified as accessible chromatin peaks within annotated gene promoters (within 2kb upstream and 1kb downstream of the transcription start site). Distal regulatory elements were then linked to genes by correlation between the promoter accessible peak and distal ATAC-seq peaks as described (de la Torre-Ubieta et al., 2018).. Cell-type specific regulatory elements comprise the union of accessible chromatin within the promoter of a given gene and associated distal regulatory elements (enhancers) for each set of genes enriched within specific cell types. Regulatory elements associated with genes enriched in specific cell types were then used for regulatory element metrics (Figures 5A–H) and partitioned heritability analyses (Figures 8).

Gene regulatory networks—Cell-type specific regulatory networks were identified based on genes enriched in cell clusters using the SCENIC pipeline (Aibar et al., 2017). First, co-expressed modules between TFs and genes were identified from the single-cell expression data using GRNBoost. The set of TFs used for the co-expression analysis consisted of 782 well known TFs from RcisTarget database. In total, 2,862,989 TF and gene linkages were identified. To obtain more reliable TF and gene linkages, the top 10% of the linkages with highest scores were kept for further analysis. Then the regulatory elements for each module were extracted from the active regulatory elements obtained from (de la Torre-Ubieta et al., 2018). A motif enrichment analysis was done for the regulatory elements using Homer. If the co-expressed TF has a motif enriched in the regulatory elements (p-value <0.01), it will be the regulon for this gene module. The motif database consists of known motifs from Homer (Heinz et al., 2010) and novel motifs from JASPAR (Khan et al., 2018). The JASPAR motifs were formatted by MEME (Bailey et al., 2009). Finally, the regulon activity (AUC score) for each module in each cell was scored by AUCell (Aibar et al., 2017). Given the distribution of activity scores, the cells that have the regulon enriched were identified. A Fisher's exact test was performed to evaluate if a regulon is significantly enriched (p-value < 0.05) in specific cell types based on the enriched cell clusters (Table S7).

Partitioned heritability analysis—Partitioned heritability was assessed using LD score regression (v1.0.0) (Finucane et al., 2015). Heritability was calculated by comparing the association statistics for common genetic variants falling within regulatory elements associated with specific cell types, with the LD-score, a measure of the extent of the LD block. First, an annotation file was created, which marked all HapMap3 SNPs that fell within the regulatory elements for each cell type. LD-scores were calculated for these SNPs within 1 cM windows using the 1000 Genomes EUR data. These LD-scores were included simultaneously with the baseline distributed annotation file from (Finucane et al., 2015). Subsequently, the heritability explained by these annotated regions of the genome was assessed from phenotypes for 18 GWAS (see Table S8 for references and sample sizes). The enrichment was calculated as the heritability explained for each phenotype within a given annotation divided by the proportion of SNPs in the genome and FDR correction within each GWAS was used to correct for multiple comparisons.

Comparisons to adult brain single-nuclei expression profiles—For comparison to single-nuclei RNA-seq data from human adult brain we obtained the gene expression counts matrix and cluster enrichment scores from (Lake et al., 2018) (GEO: GSE92942). The single-nuclei RNA-seq raw counts were read depth normalized by dividing by the total number of UMIs per cell, then multiplying by 10,000, adding a value of 1, and log transforming (ln(transcripts-per-10,000 +1).

Gene list enrichment analysis—Genes with human specific expression patterns across cortical development were obtained from (Bakken et al., 2016). High confidence ASD risk genes, defined by harboring high risk likely protein-disrupting mutations, were obtained from Sanders *et al.* combined *de novo* and TADA analysis (66 genes FDR < 0.1) (Sanders et al., 2015). Intellectual disability (ID) risk genes were obtained from previous exome sequencing of patients with idiopathic and non-syndromic (de Ligt et al., 2012; Rauch et al.,

2012) and subset to mutations presenting *de novo* in patients and likely to be gene-disrupting (frameshift, nonsense or splicing sites).

In order to obtain a list of high-confidence epilepsy risk genes, we curated ClinVar and OMIM (access date: May 2018). We first searched ClinVar for “epilepsy” associated variants which resulted in 7,011 variant-phenotype entries. Variants identified as having a clinical significance of uncertain, benign, likely benign, not provided, or drug response were removed while variants with a clinical significance of risk factor, likely pathogenic, pathogenic, or conflicting interpretations of pathogenicity were retained. Large structural variants disrupting more than one gene were also removed. This resulted in 1,227 entries with an identical gene-condition-clinical significance (>=1 variant listed per gene). Genes lacking a single variant with a clinical significance of pathogenic were removed and the remaining entries were then manually curated using OMIM. OMIM genes for which the molecular basis was known were then evaluated for the strength of evidence from the reported clinical features and the molecular genetics. If the clinical features revealed the patients never met the criteria for epilepsy (*e.g.*, two or more unprovoked seizures) or if seizures were a variable clinical feature rather than a defining feature of the syndrome these genes were excluded from the high-confidence list. Similarly, if the molecular genetic evidence was not sufficient (*e.g.*, not a large enough sample size) these genes were excluded from the high-confidence list. Finally, we cross checked that all high-confidence genes from OMIM’s Phenotypic Series for Epileptic encephalopathy early infantile were included on our list, resulting in the addition of 14 genes. This resulted in a final list of 109 high-confidence epilepsy risk genes.

Enrichment $\log_2$odds ratios were calculated using a general linear model (binomial distribution).

DATA AND SOFTWARE AVAILABILITY

The accession number for the transcriptomic dataset reported in this paper is dbGaP: phs001836.

ADDITIONAL RESOURCES

An online interface to facilitate sharing, exploration, and use of the dataset: http://geschwindlab.dgsom.ucla.edu/pages/codexviewer

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Drop-seq libraries were sequenced by the UCLA BSCRC, RNA-seq libraries were prepared and sequenced by the UCLA Neuroscience Genomics Core, and fetal tissue was collected from the UCLA CFAR (5P30 AI028697). Finger whorl GWAS data were kindly provided by Sarah Medland. We thank members of the Geschwind lab for helpful discussions and critical reading of the manuscript. Funding: This work was supported by NIH grants to D.H.G. (1U01 MH105991, 5R01 MH081754, 5R01 MH100027, 1R01MH110927 and 1U01MH116489) and the Allen Distinguished Investigator Program to DHG, the California Institute for Regenerative Medicine (CIRM)-BSCRC Training Grant (TG2–01169) to L.T.U., K.P. was supported by NIH (P01GM099134) and the Allen

Distinguished Investigator Program and the Paul G. Allen Frontiers Group, J. L. by the UCLA Tumor Cell Biology Training Program (USHHS Ruth L. Kirschstein Institutional National Research Service Award # T32 CA009056), and S. S. by the UCLA Broad Stem Cell Research Center – Rose Hills Foundation Training Award.

References:

Adams HH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Renteria ME, Trompet S, Arias-Vasquez A, Seshadri S, Desrivieres S, et al. (2016). Novel genetic loci underlying human intracranial volume identified through genome-wide association. Nat Neurosci 19, 1569–1582. [PubMed: 27694991]

Aibar S, Gonzalez-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, Rambow F, Marine JC, Geurts P, Aerts J, et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14, 1083–1086. [PubMed: 28991892]

Amiri A, Coppola G, Scuderi S, Wu F, Roychowdhury T, Liu F, Pochareddy S, Shin Y, Safi A, Song L, et al. (2018). Transcriptome and epigenome landscape of human cortical development modeled in organoids. Science 362.

Anders S, Pyl PT, and Huber W (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. [PubMed: 25260700]

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res 37, W202–208. [PubMed: 19458158]

Bakken TE, Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, Dalley RA, Royall JJ, Lemon T, et al. (2016). A comprehensive transcriptional map of primate brain development. Nature 535, 367–375. [PubMed: 27409810]

Bertrand V, and Hobert O (2010). Lineage programming: navigating through transient regulatory states via binary decisions. Curr Opin Genet Dev 20, 362–368. [PubMed: 20537527]

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, and Greenleaf WJ (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10, 1213–1218. [PubMed: 24097267]

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. [PubMed: 29608179]

Camp JG, Badsha F, Florio M, Kanton S, Gerber T, Wilsch-Brauninger M, Lewitus E, Sykes A, Hevers W, Lancaster M, et al. (2015). Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. Proc Natl Acad Sci U S A 112, 15672–15677. [PubMed: 26644564]

Cooper NJ, Shtir CJ, Smyth DJ, Guo H, Swafford AD, Zanda M, Hurles ME, Walker NM, Plagnol V, Cooper JD, et al. (2015). Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes. Hum Mol Genet 24, 1774–1790. [PubMed: 25424174]

de la Torre-Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, and Geschwind DH (2018). The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. Cell 172, 289–304 e218. [PubMed: 29307494]

de la Torre-Ubieta L, Won H, Stein JL, and Geschwind DH (2016). Advancing the understanding of autism disease mechanisms through genetics. Nat Med 22, 345–361. [PubMed: 27050589]

de Ligt J, Willemsen MH, van Bon BW, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, de Vries P, Gilissen C, et al. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 367, 1921–1929. [PubMed: 23033978]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Ecker JR, Geschwind DH, Kriegstein AR, Ngai J, Osten P, Polioudakis D, Regev A, Sestan N, Wickersham IR, and Zeng H (2017). The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. Neuron 96, 542–557. [PubMed: 29096072]

Fame RM, MacDonald JL, and Macklis JD (2011). Development, specification, and diversity of callosal projection neurons. Trends Neurosci 34, 41–50. [PubMed: 21129791]

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. [PubMed: 29867213]

Ferland RJ, Cherry TJ, Preware PO, Morrisey EE, and Walsh CA (2003). Characterization of Foxp2 and Foxp1 mRNA and protein in the developing and mature brain. J Comp Neurol 460, 266–279. [PubMed: 12687690]

Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat Genet 47, 1228–1235. [PubMed: 26414678]

Gandal MJ, Leppa V, Won H, Parikshak NN, and Geschwind DH (2016). The road to precision psychiatry: translating genetics into disease mechanisms. Nat Neurosci 19, 1397–1407. [PubMed: 27786179]

Gawad C, Koh W, and Quake SR (2016). Single-cell genome sequencing: current state of the science. Nat Rev Genet 17, 175–188. [PubMed: 26806412]

Geschwind DH, and Rakic P (2013). Cortical evolution: judge the brain by its cover. Neuron 80, 633–647. [PubMed: 24183016]

Hansen DV, Lui JH, Flandin P, Yoshikawa K, Rubenstein JL, Alvarez-Buylla A, and Kriegstein AR (2013). Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. Nat Neurosci 16, 1576–1587. [PubMed: 24097039]

Hansen DV, Lui JH, Parker PR, and Kriegstein AR (2010). Neurogenic radial glia in the outer subventricular zone of human neocortex. Nature 464, 554–561. [PubMed: 20154730]

Hayashi T, Umemori H, Mishina M, and Yamamoto T (1999). The AMPA receptor interacts with and signals through the protein tyrosine kinase Lyn. Nature 397, 72–76. [PubMed: 9892356]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576–589. [PubMed: 20513432]

Hennig C (2007). Cluster-wise assessment of cluster stability. Comput Stat Data An 52, 258–271.

Hoerder-Suabedissen A, and Molnar Z (2015). Development, evolution and pathology of neocortical subplate neurons. Nat Rev Neurosci 16, 133–146. [PubMed: 25697157]

Horvath S, and Mirnics K (2015). Schizophrenia as a disorder of molecular pathways. Biol Psychiatry 77, 22–28. [PubMed: 24507510]

Hrvatin S, Hochbaum DR, Nagy MA, Cicconet M, Robertson K, Cheadle L, Zilionis R, Ratner A, Borges-Monroy R, Klein AM, et al. (2018). Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. Nat Neurosci 21, 120–129. [PubMed: 29230054]

Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, Bessy A, Cheneby J, Kulkarni SR, 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. [PubMed: 29140473]

Lake BB, Chen S, Sos BC, Fan J, Kaeser GE, Yung YC, Duong TE, Gao D, Chun J, Kharchenko PV, et al. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol 36, 70–80. [PubMed: 29227469]

Lange C, Huttner WB, and Calegari F (2009). Cdk4/cyclinD1 overexpression in neural stem cells shortens G1, delays neurogenesis, and promotes the generation and expansion of basal progenitors. Cell Stem Cell 5, 320–331. [PubMed: 19733543]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ, He D, Weissman JS, Kriegstein AR, Diaz AA, et al. (2016). Single-cell analysis of long non-coding RNAs in the developing human neocortex. Genome Biol 17, 67. [PubMed: 27081004]

Lui JH, Hansen DV, and Kriegstein AR (2011). Development and evolution of the human neocortex. Cell 146, 18–36. [PubMed: 21729779]

Lun AT, Bach K, and Marioni JC (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol 17, 75. [PubMed: 27122128]

Luo C, Keown CL, Kurihara L, Zhou J, He Y, Li J, Castanon R, Lucero J, Nery JR, Sandoval JP, et al. (2017). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science 357, 600–604. [PubMed: 28798132]

Ma T, Wang C, Wang L, Zhou X, Tian M, Zhang Q, Zhang Y, Li J, Liu Z, Cai Y, et al. (2013). Subcortical origins of human and monkey neocortical interneurons. Nat Neurosci 16, 1588–1597. [PubMed: 24097041]

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161, 1202–1214. [PubMed: 26000488]

Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, et al. (2014). Transcriptional landscape of the prenatal human brain. Nature 508, 199–206. [PubMed: 24695229]

Miyoshi G, Butt SJ, Takebayashi H, and Fishell G (2007). Physiologically distinct temporal cohorts of cortical interneurons arise from telencephalic Olig2-expressing precursors. J Neurosci 27, 7786–7798. [PubMed: 17634372]

Molnar Z (2011). Evolution of cerebral cortical development. Brain Behav Evol 78, 94–107. [PubMed: 21691047]

Molyneaux BJ, Arlotta P, Menezes JR, and Macklis JD (2007). Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci 8, 427–437. [PubMed: 17514196]

Namba T, Kibe Y, Funahashi Y, Nakamuta S, Takano T, Ueno T, Shimada A, Kozawa S, Okamoto M, Shimoda Y, et al. (2014). Pioneering axons regulate neuronal polarization in the developing cerebral cortex. Neuron 81, 814–829. [PubMed: 24559674]

Nowakowski TJ, Bhaduri A, Pollen AA, Alvarado B, Mostajo-Radji MA, Di Lullo E, Haeussler M, Sandoval-Espinosa C, Liu SJ, Velmeshev D, et al. (2017). Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. Science 358, 1318–1323. [PubMed: 29217575]

Oeschger FM, Wang WZ, Lee S, Garcia-Moreno F, Goffinet AM, Arbones ML, Rakic S, and Molnar Z (2012). Gene expression analysis of the embryonic subplate. Cereb Cortex 22, 1343–1359. [PubMed: 21862448]

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen GB, Emilsson V, Meddens SF, et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533, 539–542. [PubMed: 27225129]

Palomares M, Delicado A, Mansilla E, de Torres ML, Vallespin E, Fernandez L, Martinez-Glez V, Garcia-Minaur S, Nevado J, Simarro FS, et al. (2011). Characterization of a 8q21.11 microdeletion syndrome associated with intellectual disability and a recognizable phenotype. Am J Hum Genet 89, 295–301. [PubMed: 21802062]

Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, Legge SE, Bishop S, Cameron D, Hamshere ML, et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet 50, 381–389. [PubMed: 29483656]

Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, Horvath S, and Geschwind DH (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 1008–1021. [PubMed: 24267887]

Pfeffer CK, Xue M, He M, Huang ZJ, and Scanziani M (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. Nat Neurosci 16, 1068–1076. [PubMed: 23817549]

Pfeuty B (2015). A computational model for the coordination of neural progenitor selfrenewal and differentiation through Hes1 dynamics. Development 142, 477–485. [PubMed: 25605780]

Pilaz LJ, Patti D, Marcy G, Ollier E, Pfister S, Douglas RJ, Betizeau M, Gautier E, Cortay V, Doerflinger N, et al. (2009). Forced G1-phase reduction alters mode of division, neuron number, and laminar phenotype in the cerebral cortex. Proc Natl Acad Sci U S A 106, 21924–21929. [PubMed: 19959663]

Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, Shuga J, Liu SJ, Oldham MC, Diaz A, et al. (2015). Molecular identity of human outer radial glia during cortical development. Cell 163, 55–67. [PubMed: 26406371]

Qiu X, Hill A, Packer J, Lin D, Ma YA, and Trapnell C (2017). Single-cell mRNA quantification and differential analysis with Census. Nat Methods 14, 309–315. [PubMed: 28114287]

Radonjic NV, Ayoub AE, Memi F, Yu X, Maroof A, Jakovcevski I, Anderson SA, Rakic P, and Zecevic N (2014). Diversity of cortical interneurons in primates: the role of the dorsal proliferative niche. Cell Rep 9, 2139–2151. [PubMed: 25497090]

Rakic P (1995). A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. Trends Neurosci 18, 383–388. [PubMed: 7482803]

Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, Albrecht B, Bartholdi D, Beygo J, Di Donato N, et al. (2012). Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet 380, 1674–1682. [PubMed: 23020937]

Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, and Vilo J (2016). g:Profiler-a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res 44, W83–89. [PubMed: 27098042]

Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, et al. (2015). Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 87, 1215–1233. [PubMed: 26402605]

Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, de Rivera H, Bien E, Baum M, Bortolin L, Wang S, et al. (2018). Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. Cell 174, 1015–1030 e1016. [PubMed: 30096299]

Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, Adiconis X, Levin JZ, Nemesh J, Goldman M, et al. (2016). Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell 166, 1308–1323 e1330. [PubMed: 27565351]

Silbereis JC, 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. [PubMed: 26796689]

Skene NG, Bryois J, Bakken TE, Breen G, Crowley JJ, Gaspar HA, Giusti- Rodriguez P, Hodge RD, Miller JA, Munoz-Manchado AB, et al. (2018). Genetic identification of brain cell types underlying schizophrenia. Nat Genet 50, 825–833. [PubMed: 29785013]

Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E, Taskesen E, Hammerschlag AR, Okbay A, Zabaneh D, et al. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. Nat Genet 49, 1107–1112. [PubMed: 28530673]

Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci 19, 335–346. [PubMed: 26727548]

Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, and Rinn JL (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32, 381–386. [PubMed: 24658644]

Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, Wu X, Vo HT, Ma XJ, and Luo Y (2012). RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. J Mol Diagn 14, 22–29. [PubMed: 22166544]

Zahr SK, Yang G, Kazan H, Borrett MJ, Yuzwa SA, Voronova A, Kaplan DR, and Miller FD (2018). A Translational Repression Complex in Developing Mammalian Neural Stem Cells that Regulates Neuronal Specification. Neuron 97, 520–537 e526. [PubMed: 29395907]

Zeisel A, Hochgerner H, Lonnerberg P, Johnsson A, Memic F, van der Zwan J, Haring M, Braun E, Borm LE, La Manno G, et al. (2018). Molecular Architecture of the Mouse Nervous System. Cell 174, 999–1014 e1022. [PubMed: 30096314]

Zhang HM, Liu T, Liu CJ, Song S, Zhang X, Liu W, Jia H, Xue Y, and Guo AY (2015). AnimalTFDB 2.0: a resource for expression, prediction and functional study of animal transcription factors. Nucleic Acids Res 43, D76–81. [PubMed: 25262351]

Zhong S, Zhang S, Fan X, Wu Q, Yan L, Dong J, Zhang H, Li L, Sun L, Pan N, et al. (2018). A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. Nature 555, 524–528. [PubMed: 29539641]

Highlights:

- High resolution transcriptome map of 40,000 cells from developing human brain
- Comparisons to other published single cell data and bulk transcriptomes
- Defines intermediate cell transition states during early neurogenesis
- Implicates specific cell types in neuropsychiatric disorders

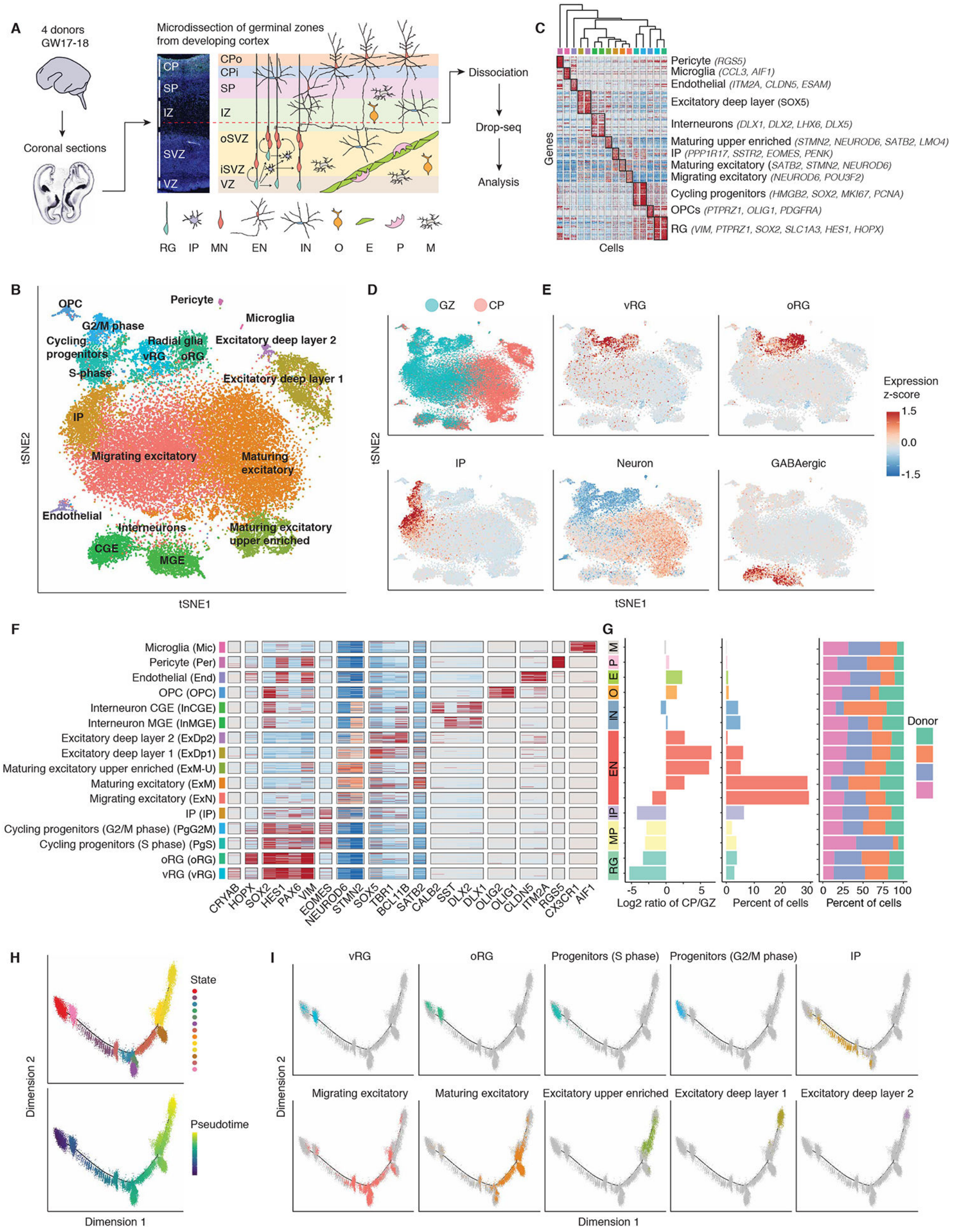


Figure 1. A catalog of cell types in developing human neocortex.

(A) Schematic illustrating experimental design and anatomical dissections. VZ: ventricular zone; iSVZ: inner subventricular zone; oSVZ: outer subventricular zone; IZ: intermediate zone; SP: subplate; CPi: inner cortical plate; CPo: outer cortical plate; RG: radial glia; IP: intermediate progenitor; MN: newborn migrating excitatory neuron; EN: excitatory neuron; IN: interneuron; O: oligodendrocyte precursor; E: endothelial cell; P: pericyte; M: microglia. (B) Scatter plot visualization of cells after principal components analysis and t-stochastic neighbor embedding (tSNE), colored by Seurat clustering, and annotated by major cell types. (C) Heatmap of gene expression for each cell. Cells are grouped by Seurat clustering, and the mean expression profile of enriched genes for each cluster was used to hierarchically cluster the Seurat clusters. The top 20 most enriched genes are shown per cluster, and anatomical marker genes in the top 20 are noted. Color bar matches Seurat clusters in B. (D and E) tSNE of cells colored by anatomical source (D), or mean expression of groups of canonical marker genes of major cell types (E). (F) Heatmap of expression profiles of canonical cell type marker genes. Cells are grouped by Seurat clustering. Color bar matches Seurat clusters in B. (G) Cluster metrics. Ratio of cells derived from GZ or CP. Percent of total cell population. Percent of cells derived from each donor. Bar colors indicate grouping of cells by major cell type, *e.g*. CGE and MGE derived interneurons are both blue. MP: mitotic progenitor. (H) Pseudo-time analysis using Monocle 2.0 of cells expected to be part of the neurogenesis-differentiation axis, colored by Monocle state or pseudo-time. Each point represents a cell. Pseudo-time represents an ordering of cells based upon the inferred trajectory, predicting the lineage trajectory. (I) Pseudo-time trajectory colored by Seurat clusters.

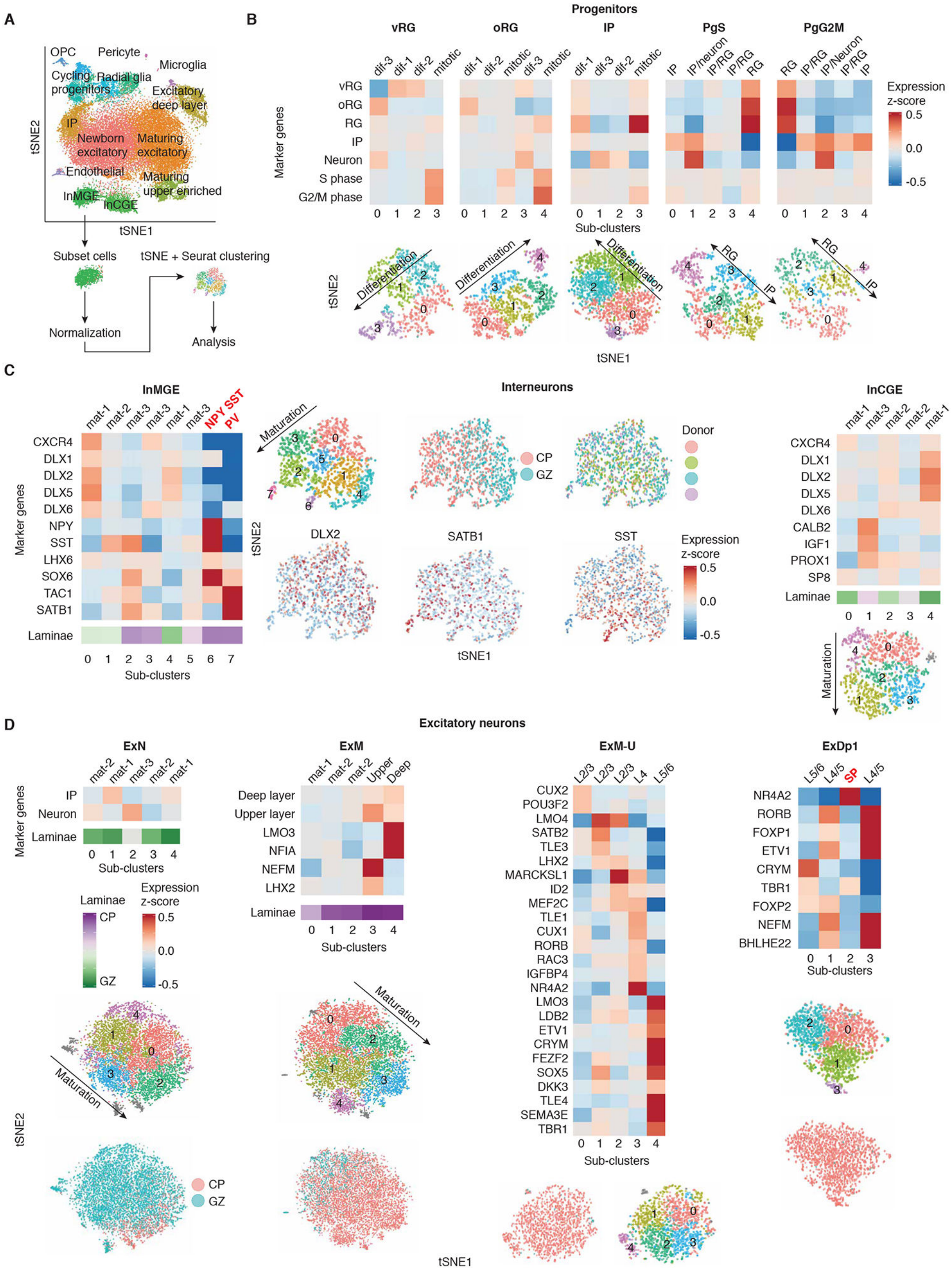


Figure 2. Sub-clustering analysis identifies progenitor states and subtypes of excitatory and inhibitory cells.

(A) Diagram of sub-clustering analysis workflow. An iterative approach was used, cells from each initial cluster were re-processed, clustered, and analyzed from the raw counts matrix using Seurat. tSNE is colored by Seurat clustering, and annotated by major cell types. (B) Sub-clustering of progenitors. Progenitors separate by cell type and cell cycle state. (C) Sub-clustering of interneurons. InMGE sub-clusters by maturity and cell subtype. InMGE-7 displays enrichment of TAC1, a marker of PV interneurons, and does not express SST (Pfeffer et al., 2013). InMGE-6 shows strong enrichment of NPY and SST. InCGE sub-clusters by maturity. All clusters are CALB2+, with differing levels of expression likely reflecting maturity. (D) Sub-clustering of excitatory neurons. New born (ExN) and maturing excitatory neurons (ExM) sub-cluster by maturity. ExM begin to display separation of laminae markers. The excitatory upper layer enriched cluster (ExM-U) shows enrichment of laminae markers for different sub-clusters, and expression of the callosal marker LMO4 (Molyneaux et al., 2007). The deep layer cluster (ExDp1) separates by layer. ExDp1–2 is enriched for the subplate marker NR4A2 (Hoerder-Suabedissen and Molnar, 2015), ExDp1–0 is enriched for lower L5 and L6 markers (CRYM, TBR1, FOXP2) (Molyneaux et al., 2007), and ExDp1–1 and 3 are enriched for L4 and upper L5 markers (RORB, FOXP1, ETV1) (Ferland et al., 2003; Molyneaux et al., 2007). Heatmaps: Heatmaps of expression profiles by sub-cluster of groups or individual marker genes (Y-axis). The laminae bar indicates the percent of cells derived from the CP. Purple: 100% of cells derive from the CP, 0% GZ; Green: 0% of cells derive from the CP, 100% GZ. Upper layer and deep layer gene groups are the top 50 most enriched genes from the excitatory upper enriched cluster and the deep layer cluster, respectively. tSNES: tSNEs of cells are colored by features of interest: sub-cluster, anatomical source, donor, or gene expression. Grey indicates cells with an undefined transcriptional signature. For heatmaps and tSNE, gene expression is plotted as a z-score for the population of cells in the plot, therefore some cell types display differences in relative expression of cell type markers between sub-clusters of the same major cell type, but all express the marker at some level (*e.g.* all RG express markers of RG, but some sub-clusters of RG have higher relative expression than other sub-clusters of RG). Labels "mat" and "dif" indicate inferred order of differentiation or maturation.

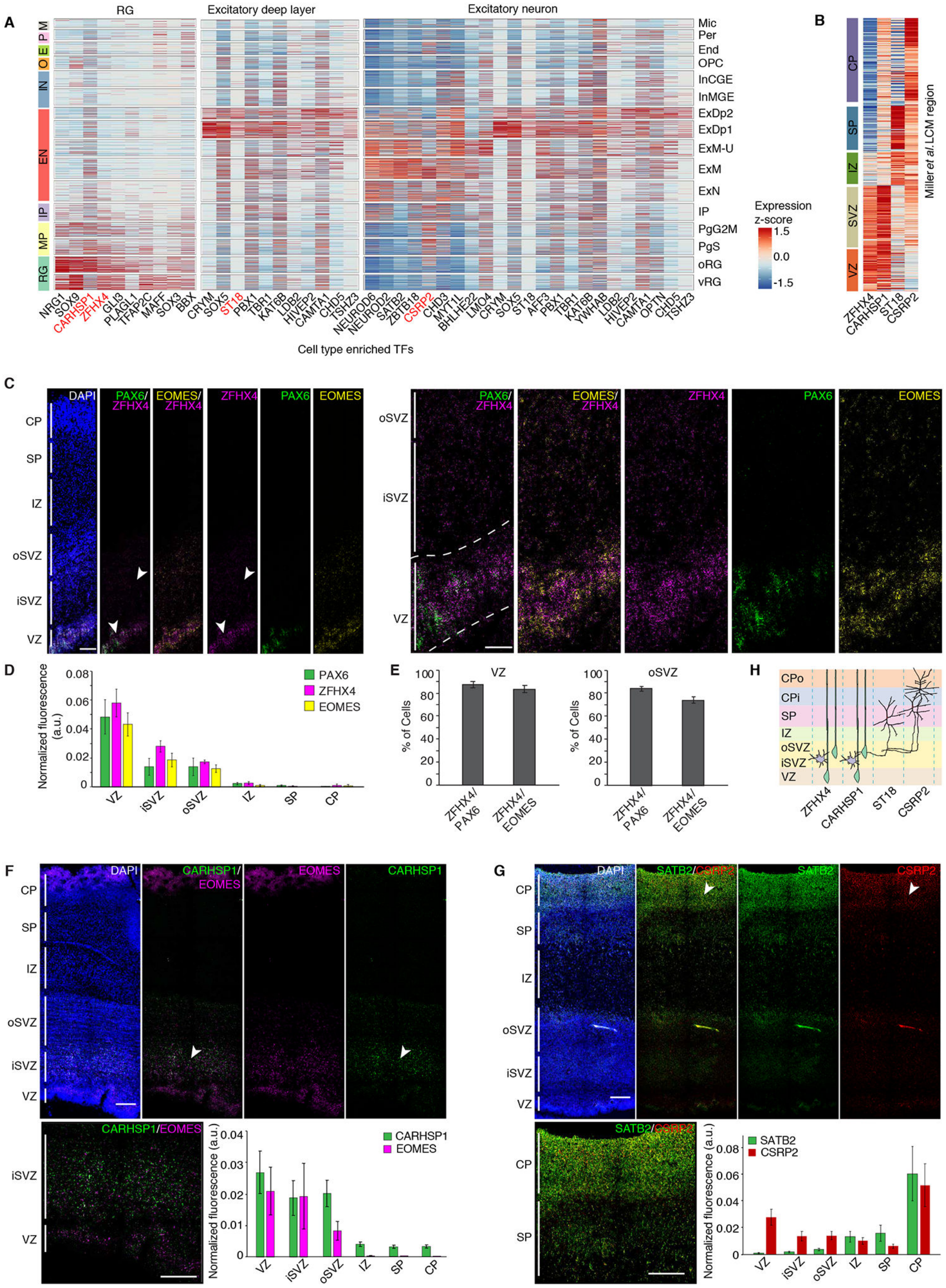


Figure 3. Cell type enrichment of TFs and co-factors.

(A) Heatmap of expression of TFs, co-factors, and chromatin remodelers enriched in RG, excitatory neurons, and deep layer excitatory neurons. Cells are grouped by cluster. Red indicates factors previously unknown to be enriched in the neocortical cell types of interest. (B) Expression of factors of interest in bulk tissue LCM laminae from developing cortex. (C) RNA FISH of fetal cortex probed with the newly identified cell-enriched TF ZFHX4 (neural progenitors in the VZ and SVZ), and known markers PAX6 (RG marker) and EOMES (IP marker). Insets show higher magnification of the VZ and SVZ. (E) Quantification of the percentage of PAX6+ or EOMES+ cells co-expressing ZFHX4. ZFHX4 is expressed in both

RG and IPs. (F and G) RNA FISH of fetal cortex probed with the newly identified cell-enriched TFs CARHSP1 (neural progenitors in the VZ and SVZ), and CSRP2 (glutamatergic neurons in the CP). (C, D, F, G) Quantification of normalized fluorescence intensity per layer for each set of probes (see materials and methods). Scale bar = 250μm (left) or 100μm (inset). (H) Schematic of cell-type specific expression of factors of interest. Color indicates -$\log_{10}$ p-value from Fisher’s test.

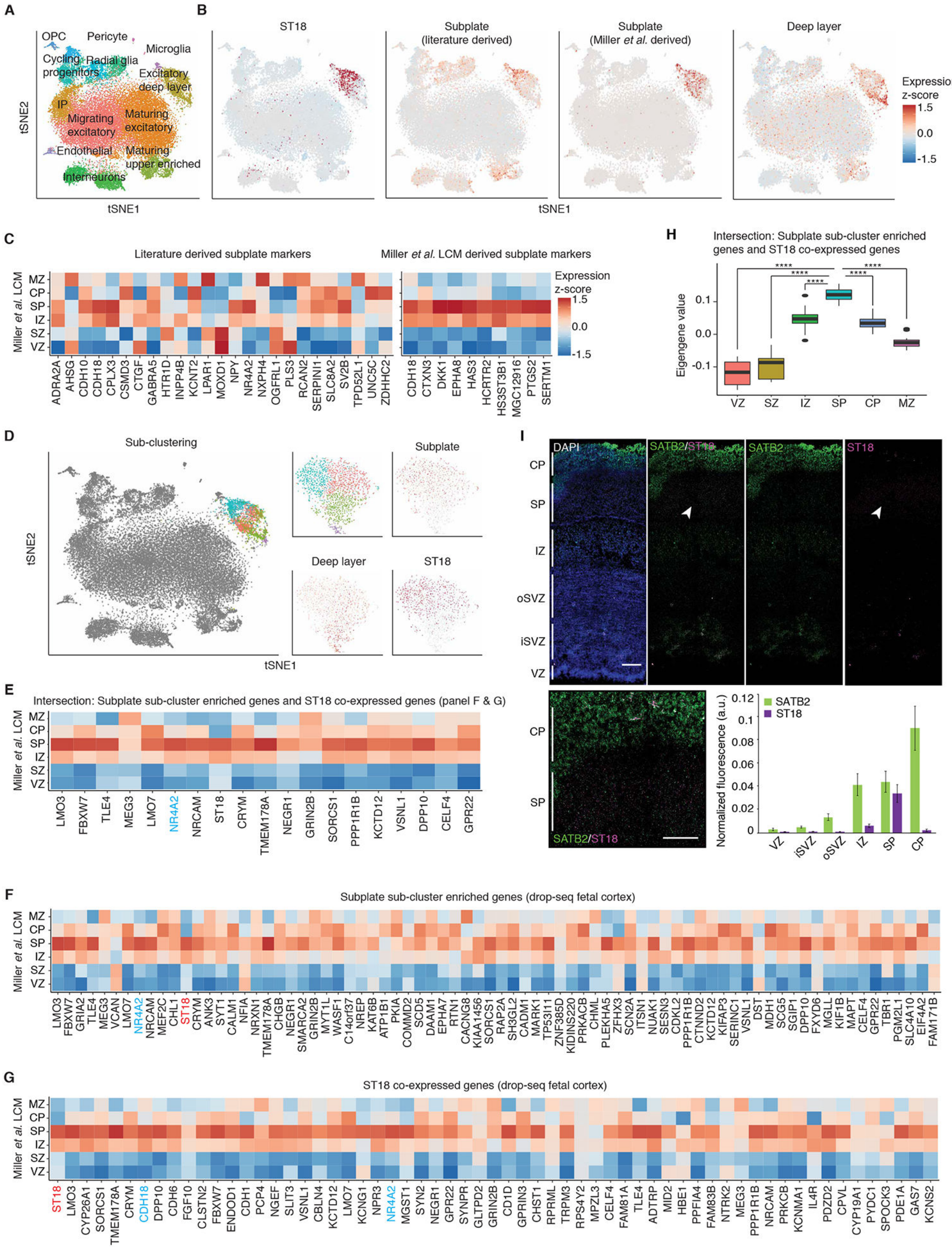


Figure 4. Characterization of subplate neuron expression profiles.

(A) tSNE colored by Seurat clustering, and annotated by major cell types. (B) tSNE of cells colored by mean expression of groups of marker genes or expression of specific genes. (C) Expression of SP markers in bulk tissue LCM laminae from developing cortex. SP markers were derived from literature sources (left), or by differential expression of the SP versus the VZ, SZ, CP, and MZ and visual confirmation of SP specificity (right). (D) Sub-clustering of the deep layer excitatory cluster 1. tSNE for the full dataset colored by sub-clustering (left). tSNE of cells belonging to the deep layer excitatory cluster (right), colored by sub-clustering, mean expression of groups of marker genes, or expression of specific genes. (E to G) Expression of SP cluster enriched genes (F), ST18 co-expressed genes (G), and the intersection of both F and G (E) in bulk tissue LCM laminae from developing cortex. Genes are ordered left to right by enrichment or correlation (highest left). Light blue text indicates SP markers previously identified. (H) Eigengene of intersected ST18 co-expressed and SP cluster enriched genes (E) plotted in bulk tissue LCM laminae from developing cortex. P-values: * <0.05, ** <0.01, *** <0.001, **** <0.0001. (I) RNA FISH of fetal cortex probed with the newly identified subplate enriched TF ST18. Quantification of normalized fluorescence intensity per layer for each set of probes (see materials and methods). Scale bar = 250μm (left) or 100μm (inset).

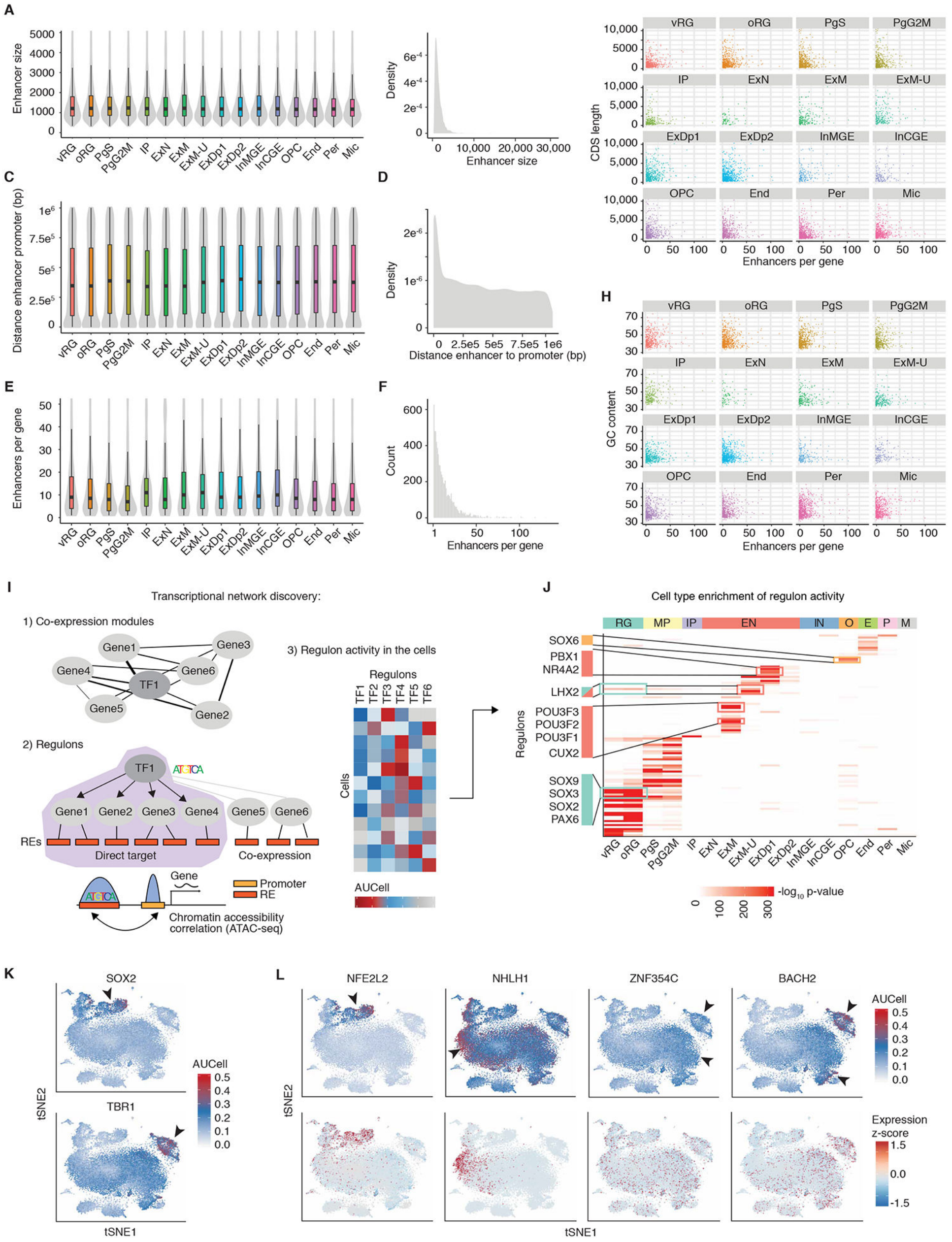


Figure 5. Transcriptional network discovery.

(A to H) Regulatory elements for cell-type specific genes. (A) Enhancer size by cell type. Enhancers are assigned to cell types by cell type enriched genes. (B) Density plot of enhancer sizes that are assigned to specific cell types. (C) Distance (base pairs) from enhancer to promoter by cell type. (D) Density plot of distance (base pairs) from enhancer end to promoter start that are assigned to specific cell types. (E) Enhancers per gene by cell type. (F) Histogram of enhancers per gene that are assigned to specific cell types. (G) Number of enhancers per gene versus CDS length of the gene by cell type. (H) Number of enhancers per gene versus GC content of the gene by cell type. (I) Schematic showing the computational approach used for transcriptional network discovery with the SCENIC pipeline (see materials and methods). 1) Co-expression modules between transcription factors and candidate genes are constructed. 2) Genes in co-expression modules are then pruned to genes which are inferred to be direct targets of the transcription factor, making a regulon. Direct targets are determined by the presence of the transcription factors binding motif in the regulatory elements associated with that gene. 3) The activity of each regulon is then assessed in each cell. (J) Cell type enrichment of regulon activity. Each regulon was scored as active or inactive for each cell, and cluster enrichment was then determined by Fisher’s test. Color indicates FDR-corrected -$\log_{10}$ p-value. (K) SCENIC regulon activity in each cell (AUCell) for the indicated TF plotted on tSNE. (L) TFs with previously uncharacterized cell type or cell subtype specific activity. Regulon activity in each cell (AUCell) for the indicated TF (top panels) or expression of the TF plotted on tSNE (bottom panels).

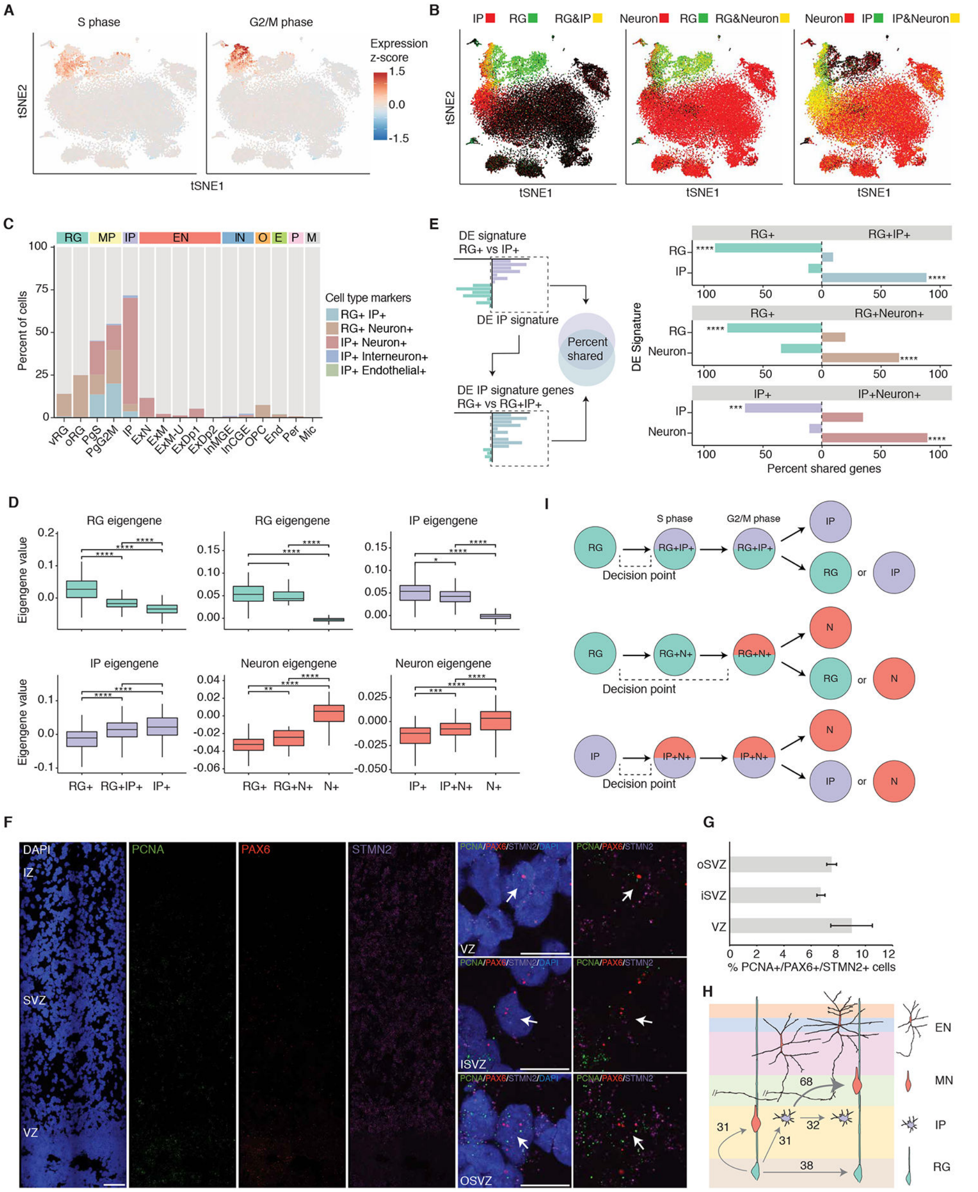


Figure 6. Dissecting the acquisition of a neuronal program.

(A) tSNE colored by mean expression of cell cycle phase markers. (B) tSNE colored by co-expression of groups of canonical cell type markers. Yellow indicates co-expression. (C) Percent of cells in each Seurat cluster displaying co-expression of major cell type markers. (D) Mixed transcriptomic signatures of mixed marker cells in S-phase corresponding to the expression of markers from multiple cell types. For the RG to IP comparison, RG and IP eigengenes were derived from differentially expressed genes between RG and IP cells, and similarly for the RG to Neuron comparison and IP to Neuron comparison. Boxplots: box indicates first and third quartiles; the whiskers extend from the box to the highest or lowest value that is within 1.5 * inter-quartile range of the box; and the line is the median. (E) Shared gene signatures between major cell types and mixed cell types. Overlap of gene signatures from major cell types (y-axis), and genes differentially expressed between major cell types and mixed marker cells (labeled on the grey bar). X-axis: percentage of genes differentially expressed between major cell types that are also differentially expressed between the corresponding major cell type and mixed marker cells. For example, ~85% of IP signature genes are more highly expressed in RG+IP+ cells than RG+ cells. (F) RNA FISH of fetal cortex probed with the S-phase marker PCNA (green), the RG marker PAX6 (red), the neuron marker STMN2 (magenta), and stained with DAPI (blue). Panels on the right show high magnification single-plane confocal images of individual cells expressing all three markers. Scale bar = 100μm (left) or 10μm (right). (G) Quantification of the percentage of cells co-expressing the S-phase marker PCNA, the RG marker PAX6 and the neuron marker STMN2. (H) Quantification of relative amounts of mitotic RG and relative amounts of IPs undergoing different differentiation events. (I) Diagram of mixed cell type transcriptomic states that is characteristic of neurogenic differentiation trajectories in human neocortex. P-values: * <0.05, ** <0.01, *** <0.001, **** <0.0001.

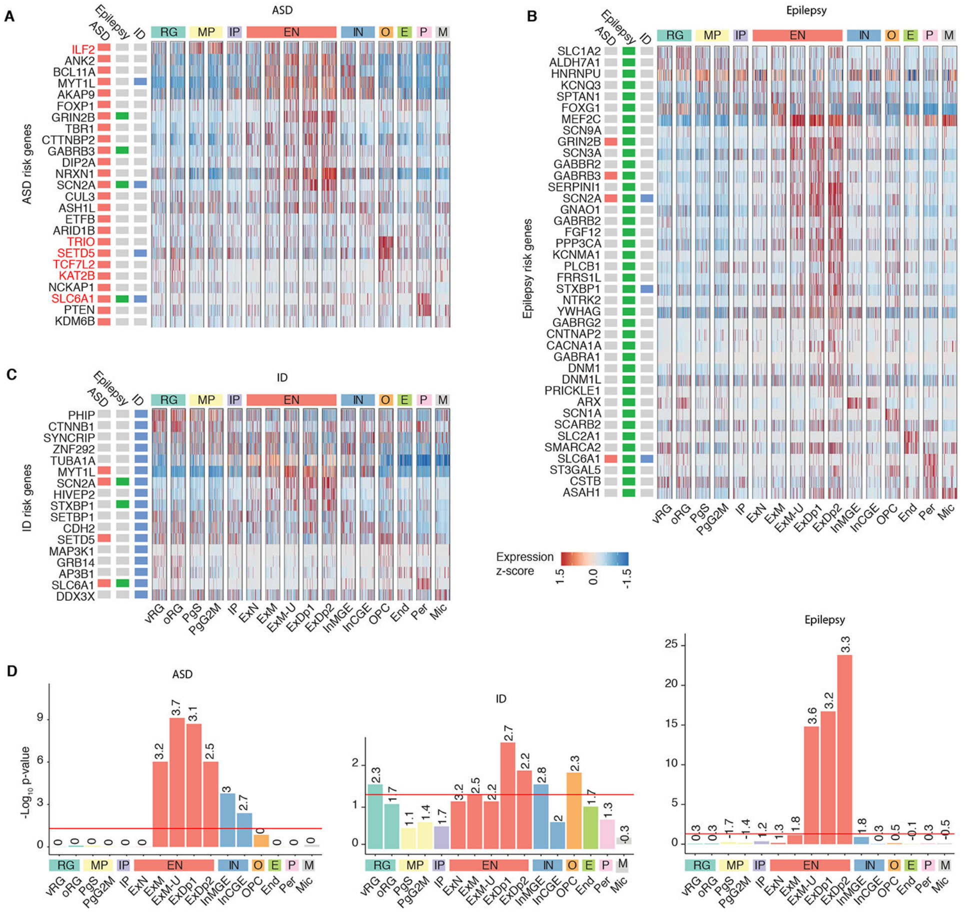


Figure 7. Cellular determinants of disease.

(A to C) Cell type expression of ASD, epilepsy, or ID risk genes respectively. Expression of ASD risk genes is enriched in fetal glutamatergic neurons with some genes specifically expressed in other cell types. Red: gene is discussed in text. Cells are ordered by cluster. (D) Cell type enrichment of ASD, epilepsy, or ID risk genes. Numbers indicate $\log_2$ odds ratio, the red line indicates FDR-significance threshold (p-value 0.05).

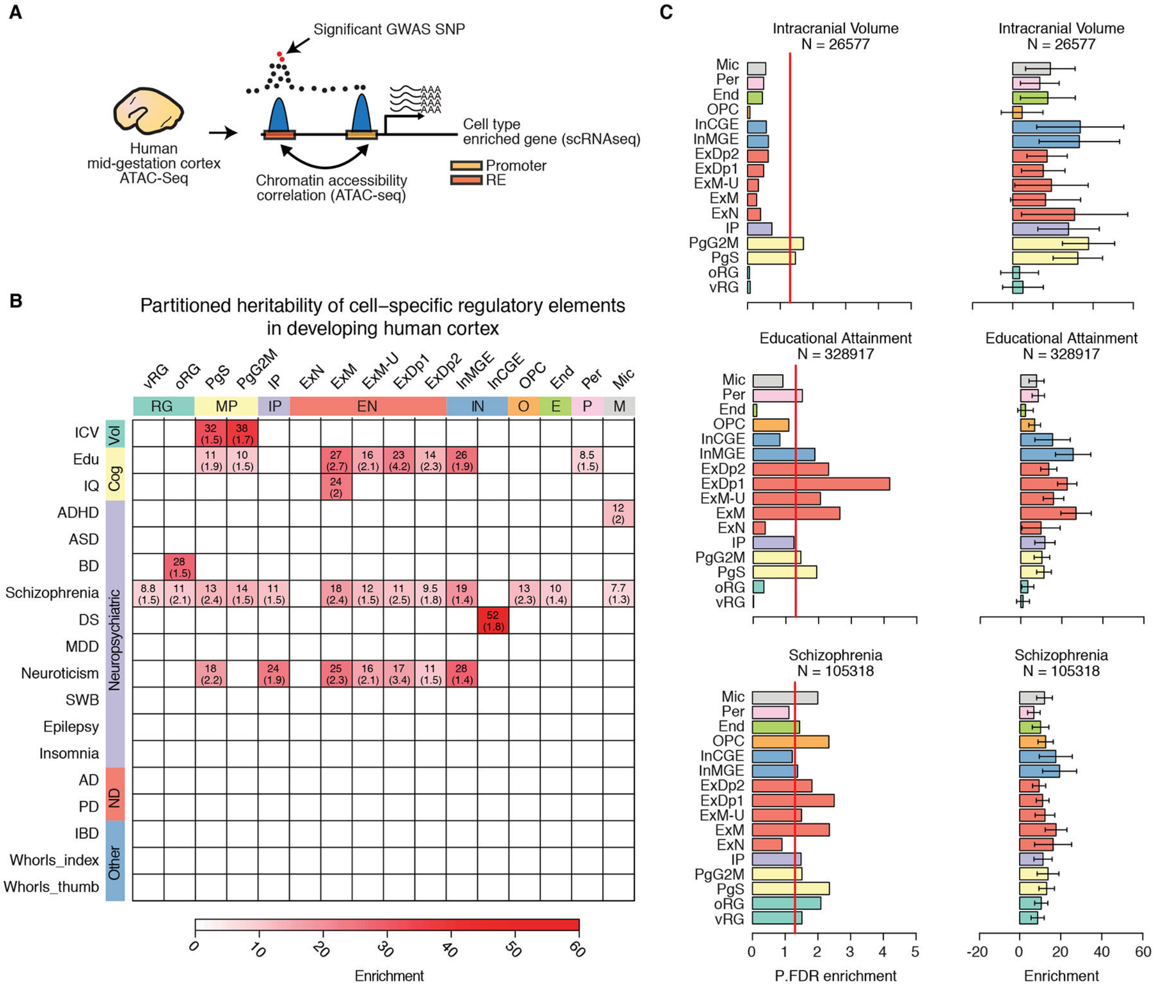


Figure 8. Partitioned heritability analysis demonstrates enrichment of heritability in specific brain traits and neuropsychiatric diseases in diverse cell types.

(A) Schematic showing the approach to identify regulatory elements (RE) for specific cell types and assess enrichment for specific brain traits. REs of genes enriched in specific cell types are identified by chromatin accessibility correlation between the promoter of the gene and other accessible peaks within 1Mb. The set of promoter and distal RE peaks are then tested for enrichment in SNPs associated with brain traits and neuropsychiatric disease using partitioned heritability by LD score regression. (B) Heatmap showing significant partitioned heritability enrichment for specific brain traits and neuropsychiatric disorders in different cell populations. Color indicates the partitioned heritability enrichment. Numbers are the FDR-corrected p-values. References for each GWAS are in Table S8. We did not observe enrichment of IBD or finger whorl variants in the regulatory elements of any of the cortical derived cell types, supporting the cell type specificity of gene regulation. (C) For selected GWAS, barplots indicate the FDR-corrected significance or the enrichment (right) of partitioned heritability. Red vertical line indicates FDR-significance threshold (p-value 0.05). Error bars represent standard error. N: GWAS sample size.