Single cell chromatin accessibility reveals regulatory elements and developmental trajectories in the embryonic forebrain

Christopher T. Rhodes[1], Apratim Mitra[2], Dongjin R. Lee[1], Daniel J. Lee[3], Yajun Zhang[1], Joyce J. Thompson[3], Pedro P. Rocha[3,4], Ryan K. Dale[2], Timothy J. Petros[1]

[1]Unit on Cellular and Molecular Neurodevelopment

[2]Bioinformatics and Scientific Programming Core

[3]Unit on Genome Structure and Regulation

Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD 20892, USA.

[4]National Cancer Institute (NCI), NIH, Bethesda, MD 20982, USA.

***Correspondence: tim.petros@nih.gov**

Keywords: Interneuron, Neurodevelopment, Epigenetic, Chromatin Accessibility, Single cell sequencing

SUMMARY

While epigenetic modifications are critical for cell state changes throughout development, a detailed characterization of chromatin accessibility during neurogenesis has not been explored. We collected single-cell chromatin accessibility profiles from four distinct neurogenic regions of the embryonic mouse forebrain using single nuclei ATAC-Seq (snATAC-Seq). We identified thousands of differentially accessible peaks, many restricted to distinct progenitor cell types or brain regions. Integrating snATAC-Seq and transcriptome data, we characterized changes of chromatin accessibility at enhancers and promoters that were tightly coupled to transcript abundance during neurodevelopment. Integrating chromatin accessibility profiles from embryonic and adult interneurons with the iPSYCH2012 dataset revealed several disease-associated polymorphisms overlapped with accessible regions in embryonic cells. These findings highlight a diverse chromatin landscape in embryonic neural progenitors, extensive coordination between chromatin accessibility and gene expression during neuron fate determination, and open the door for future studies to define critical enhancer-promoter interactions that direct cell fate decisions.

INTRODUCTION

Inhibitory GABAergic interneurons are an extremely heterogeneous cell population that can be classified based on electrophysiological properties, morphologies, synaptic connectivity, neurochemical markers and more recently, distinct transcriptomes (DeFelipe et al., 2013; Gouwens et al., 2020; Lim et al., 2018; Pelkey et al., 2017; Tasic et al., 2016; Tasic et al., 2018; Tremblay et al., 2016). In the forebrain, GABAergic neurons are born from transient embryonic structures in the ventral telencephalon known as the medial, caudal and lateral ganglionic eminences (MGE, CGE, and LGE, respectively), whereas glutamatergic projection neurons arise from the dorsal telencephalon (Bandler et al., 2017; Greig et al., 2013; Marin and Muller, 2014; Parnavelas, 2000; Rudy et al., 2011; Turrero Garcia and Harwell, 2017). The MGE and CGE (and adjacent preoptic area) generate nearly all cortical and hippocampal interneurons, with each region generating almost entirely distinct, non-overlapping interneuron subtypes (Gelman et al., 2011; Gelman and Marin, 2010; Gelman et al., 2009; Miyoshi et al., 2007; Miyoshi et al., 2010; Pelkey et al., 2017; Sultan et al., 2013; Wonders and Anderson, 2006).

The embryonic brain contains two primary classes of neural progenitors: multipotent self-renewing apical progenitors (APs, also known as radial glia cells) located in the ventricular zone (VZ) and basal progenitors (BPs) that primarily undergo neurogenic divisions within the subventricular zone (SVZ) (Arai and Taverna, 2017; Florio and Huttner, 2014; Hu et al., 2017; Kriegstein and Alvarez-Buylla, 2009; Lui et al., 2011; Noctor et al., 2004; Petros et al., 2015). Both APs and BPs give rise to postmitotic neuronal precursors (Ns) within the GEs that undergo tangential migration to populate the telencephalon (Guo and Anton, 2014; Tanaka et al., 2006). Recent evidence demonstrates that initial interneuron subtype fate is specified within the GEs (Alvarez-Dolado et al., 2006; Fishell and Kepecs, 2020; Larimer et al., 2016; Larimer et al., 2017; Nery et al., 2002; Taniguchi et al., 2013; Wichterle et al., 1999; Wichterle et al., 2001) and that initial aspects of interneuron subtype identity are established as interneuron progenitors exit the cell cycle (Mayer et al., 2018; Mi et al., 2018). Several factors contribute to these initial fate

decisions, including spatial gradients of secreted cues and transcription factors (Brandao and Romcy-Pereira, 2015; Flames et al., 2007; McKenzie et al., 2019; Tyson et al., 2015; Wonders et al., 2008; Xu et al., 2010), temporal birthdates during embryogenesis (Bandler et al., 2017; Inan et al., 2012; Miyoshi and Fishell, 2011; Rymar and Sadikot, 2007), and the mode of neurogenesis (Petros et al., 2015). However, given the incredible diversity of interneuron subtypes, we still lack a comprehensive understanding of the mechanisms that regulate early stages of interneuron fate determination during embryogenesis.

One significantly understudied area is how epigenetic mechanisms, particularly chromatin modifications, sculpt interneuron fate and maturation. Ever since Waddington's initial proposal that epigenetic factors influence cell trajectories through the differentiation landscape (Goldberg et al., 2007; Waddington, 1957), ample evidence has demonstrated that changes in a cell's epigenetic landscape can alter cell fate decisions throughout development (Hirabayashi and Gotoh, 2010; Hsieh and Gage, 2004; Hu et al., 2012; Mohamed Ariff et al., 2012; Podobinska et al., 2017; Sokpor et al., 2018; Yao et al., 2016; Ye et al., 2017). However, the role of epigenetic mechanisms influencing early interneuron fate and diversity remains understudied.

Abnormal development and function of cortical interneurons has been implicated in the pathobiology of many neurological and psychiatric disorders such as schizophrenia, autism, epilepsy and intellectual disabilities (de Lanerolle et al., 1989; Kaar et al., 2019; Magloczky and Freund, 2005; Miri et al., 2018; Nakazawa et al., 2011; Rapanelli et al., 2017). Several recent studies performing meta-analysis of genome-wide association studies (GWAS) covering numerous psychiatric disorders found an enrichment of disease-associated gene expression during embryonic development (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019; Grove et al., 2019). Additional evidence indicates that disease-associated genes may be enriched in radial glia cells and immature interneurons (Schork et al., 2019; Trevino et al., 2020). This raises the possibility that a critical 'susceptibility' window during interneuron neurogenesis and maturation may represent a vulnerable period when genetic, gene regulatory or

environmental alterations can lead to multiple psychiatric diseases. A growing body of work highlights the involvement of chromatin modifications and chromatin remodeling in neurodevelopmental disorders and neurological disease (Dall'Aglio et al., 2018; Sokpor et al., 2017; Wang et al., 2018; Wickramasekara and Stessman, 2019; Zahir and Brown, 2011). Prevalent neurological disorders have been linked directly to polymorphisms in enhancer regions (Carullo and Day, 2019; Kikuchi et al., 2019) and GWAS have shown that more than 90% of disease-associated single nucleotide polymorphisms (SNPs) are located outside of coding genes, potentially disrupting transcriptional regulation (Maurano et al., 2012; Perenthaler et al., 2019). Thus, numerous research avenues are converging on the fact that (1) defects in interneuron fate, maturation and function, and (2) perturbation of chromatin modifications may be common etiologies of neurological and psychiatric disorders. Yet, the intersecting role of chromatin state and early interneuron development in such disorders is largely unexplored.

The dearth of studies examining chromatin accessibility and chromatin modifications during neural development *in vivo* is due in part to insufficient tools and techniques, as many traditional assays require large cell numbers that are suboptimal for heterogeneous cell populations. Recent groundbreaking advances now allow characterization of chromatin accessibility and transcriptional activity in individual cells during critical developmental timepoints. Using a single nuclei assay for transposase accessible chromatin followed by sequencing (snATAC-Seq) (Buenrostro et al., 2015a; Buenrostro et al., 2015b), we characterized the chromatin accessibility of cells during the transition from progenitors to lineage restricted neurons within the GEs and dorsal telencephalon. We identified differentially accessible peaks (DA peaks) that were enriched in specific brain regions and/or distinct progenitor cell types. Among chromatin accessibility profiles, individual loci smoothly transition from open to closed chromatin (or vice versa) as cells progress along their developmental timeline. Some DA peaks displayed increased chromatin accessibility in a lineage-specific manner, many of which were associated with known fate-determining genes. By integrating snATAC-Seq with single cell RNA sequencing (scRNA-

seq) and ENCODE ChIP-Seq data (Gorkin et al., 2020), we quantified changes to chromatin accessibility, transcriptional state and cis-regulatory elements of developmentally critical genes. Overall, we found that for a given gene, chromatin accessibility and gene expression patterns were similar as differentiation progressed, with many promoters and enhancers being accessible prior to and well after mRNA detection. Last, by comparing gene expression and accessibility profiles to the iPSYCH2012 dataset (Pedersen et al., 2018; Schork et al., 2019), we detected more hits to genes that are differentially expressed in progenitor cells compared to postmitotic neurons in the GEs or mature interneurons, raising the possibility that gene regulatory elements present during interneuron development could be involved in neurological and psychiatric diseases.

RESULTS

Differentially accessible peaks in the embryonic forebrain based on brain region and cell state

To ascertain the chromatin accessibility landscape of differentiating neurons, we microdissected the MGE, CGE, LGE and cortex from wild-type mice at embryonic day 12.5 (E12.5) when cells in the GEs are undergoing neurogenesis (Bandler et al., 2017) (Figure 1A). Since cortical neurogenesis occurs later, we also harvested E14.5 dorsal cortex to compare both temporally (E12.5 GEs vs. E12.5 cortex) and neurogenically (E12.5 GEs vs. E14.5 cortex) matched dorsal and ventral forebrain. Tissue was lysed and nuclei isolated for snATAC-Seq on the 10X Genomics platform. Sequencing libraries contained a total of 39,253 single nuclei from multiple biological replicates including 10,310 from MGE, 8,543 from CGE, 11,346 from LGE and 9,054 from cortex. Libraries were aggregated, downsampled to equal numbers of median fragments per nuclei and normalized by latent semantic analysis (LSA) before peak calling, construction of cell-by-peak count matrices and removal of batch effects (Figure S1).

Using uniform manifold approximation and projection (UMAP), nuclei segregated largely by tissue region, with distinct populations detected for the MGE and the cortex while CGE and LGE overlapped substantially (Figure 1B). The smart local moving (SLM) algorithm (Waltman and van Eck, 2013) detected 27 clusters, of which three non-neuronal clusters (i.e., endothelial cells) were removed based on gene markers to retain 96.8% of nuclei in 24 clusters (Figure 1C). Cell types were assigned by inspecting promoter accessibility (PA) (defined as the sum of reads mapping within -2000 bp of a TSS) of canonical cell type markers and were further refined by transferring cell type assignments from droplet-based scRNA-Seq data of E12.5 embryos (Figures 1D and S1) to the snATAC-seq dataset. PA for markers of interneuron and excitatory glutamatergic pyramidal cell maturation segregated clusters into mitotic apical progenitors (APs), basal progenitors (BPs) and postmitotic neuronal precursors (Ns) (Figures 1D and S1O-Q). MGE and most CGE and LGE nuclei display accessible chromatin at markers of GABAergic neurons, while virtually all cortical nuclei have accessible chromatin at markers of glutamatergic neurons (Figures S1O and S2). A group of nuclei collected in the LGE and CGE populations (clusters 9, 14, 16, 18, 23 and 24) displayed accessibility profiles more similar to glutamatergic neuron markers and were labeled as a 'mixed' neuron population (Figures S1O and S2). This was likely contamination at the pallial-subpallial boundary along the LGE and CGE because this population was not observed in the MGE population. To quantify temporal differentiation programs, a minimum spanning tree was constructed in Monocle3 (Cao et al., 2019; Trapnell et al., 2014) (Figure 1E). The progression of pseudotime largely recapitulated known maturation markers, from *Nes*+ and *Ccnd2*+ cycling progenitors to *Dcx*+ and *Rbfox3*+ postmitotic neurons (Figures 1F-I). Additionally, region restricted genes such as *Nkx2-1* and *Lhx6* in the MGE, and *Neurod6* in the cortex, displayed accessibility profiles that were constrained to these regions (Figures 1J-M). We examined pseudobulk ATAC read pileups within each cluster for regionally restricted genes for the MGE (*Nkx2-1*) and cortex (*Neurod6*) (Figures 1N and 1O) and for two pan-neuronal maturation markers for APs (*Nes*) and BPs/Ns (*Dcx*) populations (Figures 1P and 1Q). Read

pileups and signal strength for *Nkx2-1* and *Neurod6* reads were restricted to the MGE and cortex/mixed clusters, respectively. As expected, *Nes* reads were enriched in AP cell clusters with diminished signals in BP through N clusters, whereas *Dcx* displayed the inverse low-AP to high-N accessibility profile.

To differentiate between enhancer-like open chromatin and promoter-like open chromatin within early interneuron lineages, we performed Cut&Tag (Kaya-Okur et al., 2019) on pooled samples of MGE, LGE and CGE for H3K4me3, a histone post-translational modification associated with promoters of transcriptionally active and poised genes. Comparing cluster-specific snATAC-Seq signals to H4K3me3 enriched regions from Cut&Tag and ENCODE ChIP-Seq experiments demonstrated high concordance between H3K4me3 Cut&Tag and ChIP-Seq signal, both of which overlapped snATAC-Seq signal at promoters. Some regions had high snATAC-Seq signal but lacked strong H3K4me3 signal, indicating chromatin accessibility at a likely enhancer, as is the case for the second intron of *Nes* which contains a known enhancer (Lothian and Lendahl, 1997; Zimmerman et al., 1994). (Figures 1N-Q). These scATAC-Seq results revealed an extraordinary diversity of chromatin accessibility profiles within the embryonic forebrain that largely correlate with known cell type.

Differentially accessible peak profiles during initial differentiation

We detected differentially accessible (DA) peaks among all possible peaks (intergenic peaks and those in promoters/gene bodies) for each cluster compared to all other clusters. We detected a total of 30,046 DA peaks (FDR <= 0.05, average log(fold change) > 0) across all clusters, visualized in a dot plot (Figure 2A, Table S1) and a corresponding heatmap (Figure S3A). To characterize DA peak profiles across clusters, we asked whether the genomic coordinates bounding DA peaks of one cluster had reads in any other cluster. If there were reads in a DA peak from one cluster in another cluster, this peak was considered overlapping between the clusters. No minimum threshold for overlapping peak counts was used before calculating the percentage

of overlapping DA peaks from one cluster compared to all peaks from the same cluster (Figure 2A, “% Overlap”). We also counted the number of DA peaks from each cluster to assess if there were differences in the number of DA peaks called for each cluster (Figure 2A, “Peak counts”).

Unsupervised hierarchical clustering (HC) of DA peaks profiles created a dendrogram that segregated clusters first by maturation state (AP, BP and N), and secondarily by tissue origin. Since the overall profiles of LGE and CGE nuclei were very similar, these two regions were labeled ‘LGE/CGE’ for this analysis. The ‘mixed’ neuron population (Figure S1O) was also left as an individual group for this analysis. The dendrogram generated by HC is very similar to cluster relationships in LSA/UMAP space, which is encouraging since different features were used in each analysis (HC: DA peaks in cluster pairs, LSA: all peaks across libraries). As expected, dot plot positions containing both high DA peaks counts and high percent overlap were almost exclusively grouped along the diagonal, while positions with limited numbers of DA peaks or low between-cluster peak overlap populated the off-diagonals, indicating high specificity of DA peaks to specific clusters (Figure 2A). One exception are the four cortical AP clusters, which had greater similarity between each other compared to other clusters. The cell type- and brain region-specific DA peak profiles are consistent with previous models of chromatin reorganization during cellular differentiation (Hu et al., 2018; Maybury-Lewis et al., 2020; Preissl et al., 2018).

We also visualized binarized peak signals (i.e., “open” or “closed” regions) per cluster using a heatmap and again observed high peak signals primarily along the diagonal (Figure S3A). Intriguingly, the mean number of DA peaks decreased as maturation progressed, with a significant decrease in DA peaks in BP and N nuclei compared to APs in each tissue (Figures S3C-D). Despite this decrease, DA peak profiles became more distinct as maturation progressed, as indicated by the low between-cluster peak overlap along the dot plot off-diagonals (Figures 2A and S3A). The decrease in global accessibility over time is consistent with previous observations during cellular differentiation (Ludwig et al., 2019; Mo et al., 2016; Stergachis et al., 2013). There

were also a greater number of DA peaks in all maturation stages in the cortex compared to GEs (Figures S3E-F).

Regions of accessible chromatin are enriched for transcription factor (TF) binding motifs that often play essential roles in setting up regulatory mechanisms and drive cell specification. To define which TFs might drive differentiation in distinct brain regions, we performed motif analysis on DNA sequences within DA peaks using the JASPAR CORE (Khan et al., 2018) vertebrates collection. Motif analysis detected TF motifs that have lineage- and tissue-specific roles during interneuron neurogenesis (Figure S2B, Table S2). Further, among motifs enriched in specific clusters that also had a corresponding differentially accessible promoter, the top 5 matches contain one or more TF motif patterns with known spatial or temporal expression profiles correlating with the cell cluster (Figure 2B, Table S3). For example, UMAP visualization of motif enrichment by chromVAR, which measures the gain or loss of accessibility of a given genomic annotation relative to the average cell profile (Schep et al., 2017) for NR2F1 (MGE and CGE-specific), ASCL1 (GABAergic BP-enriched), EMX1 (cortical progenitor AP/BP-specific), NEUROD2 (cortical postmitotic N-specific) motifs demonstrated the expected neuronal lineage progression and/or region-restricted patterns (Figures 2B-E). We further quantified genome-wide DA peaks distributions within annotated gene regions and found the majority of DA peaks were constrained to promoter and intronic regions of gene bodies as well as distal intergenic loci (Figure S3G). These findings indicate that DA peaks are specific to brain region and cell-state, and importantly, that the genomic sequence of DA peaks include lineage specific TF motifs that can shine light on mechanisms regulating neuronal differentiation.

Candidate cis-regulatory elements are dynamic and cell state dependent

The global decrease of DA peak numbers from cycling neural progenitors to postmitotic immature neurons prompted us to examine changes to DA peaks in the context of maturation and/or lineage commitment. For this and all future analysis, we removed the ‘mixed’ population in

order to characterize true LGE and CGE-derived GABAergic cells. The Monocle3 extension Cicero (Pliner et al., 2018) detects interactions or “connections" between any 2 genomic loci and then assigns a co-accessibility score between the 2 sites analogous to a correlation, thereby calculating the proportion of nuclei containing a given loci-loci interaction within a population. We quantified such interactions within all nuclei and detected 92,414 connections that had a co-accessibility score equal to or greater than 0.25, meaning a given loci-loci interaction is detectable in $\geq$ 25% of all nuclei in the defined population (henceforth “Cicero connections”).

To explore co-accesibility changes during initial neuron differentiation, we used a heatmap to visualize Cicero connections in which at least one of the 2 interacting loci was a DA peak, representing potential interactions between DA peaks and putative cis-regulatory elements (cREs). Nuclei containing these Cicero connections were divided into 10 bins of equal pseudotime intervals along the Y-axis, with individual DA peaks grouped via hierarchical clustering along the X-axis. The heatmap color (Z-score) represents the proportion of nuclei with a given DA peak (interacting via a Cicero connection) with a co-accessibility score $\geq$ 0.25 (Figure 3A). Nearly half of all such Cicero connections were enriched in AP nuclei, which is consistent with an overall decrease in accessibility as development progresses (Figures S3B-C) and indicates a progressive change in the Cicero connection profiles as development proceeds. As chromatin regions with dynamic accessibility are associated with gene regulation during neural stem cell activation (Maybury-Lewis et al., 2020), we hypothesized that the Cicero connections which became more prevelant in immature neurons might encode lineage specific cREs that may play a role in neuronal lineage commitment.

To identify candidate cREs at specific genes, we examined Cicero connections within 0.5 Mb windows of gene TSSs after filtering for co-accessibility scores $\geq$ 0.25. We also identified Cicero connections in which only one of the 2 interacting loci overlapped any TSS of a gene, representing potential interactions between a TSS and putative cREs (henceforth “TSS-cRE connections”). All Cicero connections within this window are visualized as orange arcs whereas

TSS-cRE connections for a selected gene are highlighted in purple. We focused on several genes that have temporal and/or spatially restricted expression patterns (Figures 3B-C). To characterize Cicero connections that were spatially restricted, we downsampled tissues to equal nuclei numbers and detected 91,904, 76,858, 89,366, and 148,942 Cicero connections with co-accessibility scores $\geq$ 0.25 in MGE, LGE, CGE and cortex, respectively. *Lhx6* and *Neurod6* interactions are restricted to the MGE and cortex, respectively, and TSS-cRE connections for *Lhx6* and *Neurod6* are virtually exclusive to these regions. *Nr2f2* (*CoupTF-II*) is a marker for APs in the VZ of the CGE and caudal MGE (Kanatani et al., 2008); and this gene has its highest co-accessibility scores and TSS-cRE connections counts in these two regions. *Sp8* is a marker for LGE progenitors that is excluded from the MGE (Ma et al., 2012; Waclaw et al., 2006); this gene has its highest co-accessibility scores and TSS-cRE connections counts in the LGE and CGE compared to MGE and cortex. Overall, tissue specific TSS-cRE connectivity patterns were quite similar to regionally restricted gene expression patterns that are critical to neuronal development.

Numerous genes had TSS-cRE connections with a temporal component (Figure 3C). To characterize Cicero connections that varied between neurogenic cell types, we downsampled to equal nuclei numbers and detected 89,312, 88,980, and 110,362 Cicero connections with co-accessibility scores $\geq$ 0.25 in AP, BP and N nuclei, respectively. Amongst these Cicero connections, postmitotic genes *Lhx6* and *Neurod6* had their highest co-accessibility scores and TSS-cRE connections in BP and N nuclei in the MGE and cortex, respectively (Figure 3C). Progenitor enriched genes *Nr2f2* and *Sp8* (Kanatani et al., 2008; Ma et al., 2012; Waclaw et al., 2006) had their highest co-accessibility scores and TSS-cRE connections in AP nuclei, with decreased connections and co-accessibility scores in N nuclei. The pan-AP marker nestin (*Nes*) had its highest co-accessibility scores and TSS-cRE connections in AP nuclei throughout all regions compared to BP and N nuclei. Conversely, the postmitotic GABAergic marker *Gad1* had the its highest co-accessibility scores and connection counts in BP and N nuclei. As with regional specificity of TSS-cRE connections, the temporal connection patterns largely recapitulate known

in situ expression patterns as neurons mature. In sum, cREs likely interact with DA peaks and TSSs to regulate genes in regionally and temporally restricted patterns, and the co-accessibility patterns of TSS-cREs connections closely resemble known spatial and temporal expression patterns in the embryonic forebrain.

Transcription, gene accessibility and enhancer activation are highly coordinated in neural progenitors

To enhance our understanding of the relationship between chromatin accessibility and gene expression during cortical neurogenesis, we combined the chromatin accessibility profiles from snATAC-Seq data with scRNA-Seq data from matched embryonic brain regions. Integrating ATAC and RNA data involves quantifying ATAC reads in or near gene bodies by a Gene Activity Score (GAS) as a proxy to transcript abundance. Yet, embryonic brain tissue contains cell lineages that are undergoing multiple state changes in a short developmental time span. Events that occur during such dynamic maturation processes include enhancer activation, transcription initiation and enhancer decommissioning. We reasoned that this would have a significant impact on how accurately the GAS of a gene would reflect the same gene's transcriptional state.

Before quantifying GAS, we refined our cRE predictions with the goal of detecting presumptive enhancers (H3K27ac+ accessible chromatin) that may be critical to neuronal fate determination. We first combined our Cicero TSS-cRE analysis, as described above, with TSS-cRE predictions as determined by the SnapATAC algorithm (Fang et al., 2019) (Table S4). SnapATAC predictions aim to link distal regulatory elements to target genes based on the transcript count of a gene and chromatin accessibility at peaks flanking the gene. Specifically, SnapATAC performs logistic classification using gene expression as input variable to predict the binarized chromatin state at peaks. Our rationale for using multiple algorithms to characterize gene-peak interactions was that some TSS-cRE connections detected by Cicero (or SnapATAC) may not be detectable by the other method, but TSS-cREs connections common to both methods

may improve cRE predictions. After combining TSS-cRE connections from both methods, we retained common cREs to generate a list of all detectable presumptive enhancers (Figure 4A). These loci were then integrated with ENCODE H3K27ac ChIP-Seq peaks (Gorkin et al., 2020) from E12.5 and E14.5 forebrain, resulting in a set of 'high-confidence' enhancer candidates (henceforth "presumptive enhancers") (Table S5).

The VISTA Enhancer Browser is a resource for experimentally validated human and mouse noncoding fragments with gene enhancer activity, identified by extreme evolutionary sequence conservation or by ChIP-seq and then assessed in transgenic mice (https://enhancer.lbl.gov) (Visel et al., 2007). To validate our presumptive enhancers, we detected interactions between VISTA enhancers and genes with spatial and/or temporal restricted expression patterns (Figures 4B-E). The respective gene-enhancer interaction scores from both methods (i.e. Cicero: co-accessibility; SnapATAC: -log10(p-value)) were at or near the maximum possible values for several representative genes, indicating this strategy can predict enhancers with high-confidence. Among TSS-cRE connections, Cicero detected 9,541 connections, snapATAC detected 79,146, and 3,855 were common between both methods (Figure 4F). Interestingly, while Cicero detected a lower number of TSS-cRE connections than SnapATAC, a higher proportion of Cicero cREs overlapped ENCODE H3K27ac peaks, suggesting that Cicero may be more stringent in identifying presumptive enhancers (possibly due to how Cicero's algorithm weights proximal connections higher than distal ones) while SnapATAC may have a higher false positive rate. Last, the distance between genes and our enhancers varied considerably, but the majority of gene-enhancer interactions were dozens to hundreds of kilobases apart (Figure 4G).

After testing multiple Gene Activity Score metrics for snATAC-Seq/scRNA-Seq integration (Figure S4), we defined GAS as the sum of all ATAC reads mapping to the promoter, first exon of all transcripts and presumptive enhancers of a given gene. By incorporating presumptive enhancers and removing intronic reads which may contain enhancers associated with separate

genes, this GAS metric produced the highest concordance between ATAC and RNA assays, as well as the expected clustering profiles based on brain regions (Figure S4D). Following integration, the clustering of snATAC-Seq and scRNA-Seq cells were highly concordant (Figures 5A-B). As with nuclei from non-integrated snATAC data, the MGE and cortex integrated cells formed distinct clusters whereas the LGE and CGE cells were largely overlapping. Following integration, the Louvain algorithm detected 26 clusters (Figure 5C) and Monocle3 assigned pseudotime (Figure 5D) which largely recapitulated temporal and spatially restricted expression patterns expected in embryonic forebrain neurogenic regions (Figures 5E-J and S5). Characterizing developmentally critical genes and associated presumptive enhancers from snATAC-Seq/scRNA-Seq data, we detected previously validated VISTA enhancers (Visel et al., 2007) interacting with genes in the MGE (hs704 and hs1538 regulating *Nkx2-1*) and the cortex (hs627 regulating *Neurod2*), as well as genes expressed by GABAergic progenitors (hs967, hs998, hs1114, hs1354, and hs1540 regulating *Ascl1*) and glutamatergic progenitors (hs1025 regulating *Emx1*) (Figures 5K-N, Table S6). Overall, integration of snATAC-Seq and scRNA-Seq retains features from individual (non-integrated) datasets and allows exploration of the transcriptional regulatory mechanisms.

To characterize the temporal expression and chromatin accessibility profiles during neurogenesis, we utilized a hierarchical clustering based approach from the DEGReport package (Pantano, 2017) to group RNA, GAS and enhancer counts from our integrated scRNA-Seq/snATAC-Seq data (Figures 6A-R). We detected 3,330 differentially expressed genes (DEGs) among APs, BPs, and Ns from the E12.5 scRNA-Seq data by pairwise comparisons between groups. Then we selected RNA counts (from scRNA-Seq data) for which there were corresponding GAS and presumptive enhancer counts (from snATAC-Seq data) from our E12.5 integrated data, which reduced this set to 286 genes. DEGReports hierarchical clustering uncovered 5 clusters of genes each containing at least 5 genes, of which over 85% of the genes fell into two categories: one cluster is consistent with high expression and accessibility profiles within progenitors (APs and BPs) that

are downregulated in postmitotic neurons (156 DEGs, Figure 6A), and another cluster with the complimentary profile (90 DEGs, Figure 6B). Genes within these clusters displayed similar patterns of expression and chromatin accessibility over pseudotime. Visualization of representative early expressing genes *Hes1* (Figures 6C-F) and *Lmo1* (Figures 6G-J) and late expressing genes *Myt1l* (Figures 6K-N) and *Lhx6* (Figures 6O-R) demonstrated consistent trends for transcript, GAS and enhancer counts over pseudotime within each cluster. We quantified the number of high-confidence enhancers associated with upregulated DEGs as maturation progressed. We found 200 enhancers associated with DEGs that had a positive fold change from APs to BPs, 175 enhancers for BP to N transitions, and 269 enhancers from APs to Ns (Figure S6C). Likewise, we found 188 enhancers associated with DEGs that had a negative fold change from APs to BPs, 179 enhancers for BPs to Ns transition, and 320 enhancers from APs to Ns (Figure S6C). There is a gradual decrease in the ratio of the number of enhancers being activated versus the number being decommissioned as maturation progressed (Figure S6D), suggesting that a greater number of genes and associated enhancers are being repressed as interneuron progenitors exit the cell cycle. Taken together, DEGs and associated enhancers exhibit reorganization during the transition from progenitors to lineage committed interneuron precursors.

To further analyze expression and chromatin accessibility dynamics during neurogenesis, we sought to characterize continuous differentiation processes by visualizing matched heatmaps for RNA, GAS and presumptive enhancer counts of highly variable genes. (Figures 6S-U). We selected transcript counts (from scRNA-Seq data) and corresponding GAS and enhancer counts (from snATAC-Seq data) for the top 500 most variable genes from the E12.5 integrated dataset. Not all genes contained detectable GAS and/or enhancer counts (Figures S6E-G), so we retained only the 210 highly variable genes that had corresponding GAS and enhancer counts. RNA, GAS and enhancer count matrices for these genes were co-clustered using hierarchical clustering with a correlation distance metric and average linkage and visualized in matched heatmaps (Figures 6S-U). Partitioning early, transitional and late expressing gene profiles with respect to pseudotime

largely followed a continuous progression as cells matured from APs to BPs to Ns. Overall, there was high similarity between 1) early and late gene expression patterns detected by degPatterns (Figures 6A-B and S6A-B) and 2) early to late expressing genes visualized in heatmaps (Figures 6S-U), indicating distinct, dynamic expression and chromatin accessibility in AP, BP and N cells. By integrating multiple modalities, we have characterized relationships between gene expression, chromatin accessibility and presumptive enhancers during forebrain neurogenesis, and revealed a strong correlation between these features during initial neuron differentiation.

Progenitor restricted genes are implicated in psychiatric disorders

As many psychiatric disease-associated genes are enriched in immature interneurons during embryonic development (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019; Grove et al., 2019; Schork et al., 2019; Trevino et al., 2020), we were curious if these genes (and genomic loci) were expressed (or accessible) in our embryonic scRNA-seq and snATAC-seq dataset. To increase confidence that any disease-associated genes are enriched in embryonic tissue, we generated transcriptome and chromatin accessibility profiles of postnatal (P)30 mouse forebrain mature GABAergic interneurons. We utilized *Dlx6aCre*;*Sun1-GFP* mice combined with cell sorting to specifically harvest nuclei from all MGE- and CGE-derived interneurons in the cortex and hippocampus of P30 mice. We performed both scRNAseq and scATACseq on GFP+ nuclei from these mice and integrated this data with the embryonic MGE and CGE populations. Not surprisingly, the embryonic and P30 nuclei formed distinct clusters in RNA-only (Figure S7A), ATAC-only (Figure S7B) and integrated RNA+ATAC datasets (Figures 7A-B). APs, BPs, and Ns formed separate clusters in embryonic tissues while the P30 nuclei could be further divided into MGE and CGE-derived interneuron subtypes (Figures 7C and S7A-B). Pseudotime values assigned by Monocle3 recapitulated developmental trajectories and maturation stages (Figure 7D).

Using interneuron specific datasets, we could now compare differentially expressed genes (DEGs) from each age group (embryonic or P30) or cell type (progenitors, immature neurons, and mature interneuron subtypes) with disease-associated genes. The Integrative Psychiatric Research (iPSYCH) consortium established a large population-based Case–Cohort sample (iPSYCH2012) aimed at unravelling the genetic and environmental architecture of severe mental disorders (Pedersen et al., 2018). The iPSYCH2012 GWAS dataset includes candidate disease-associated single nucleotide polymorphisms (SNPs) which are associated to specific genes based on gene function, SNP loci and Hi-C. We compared DEGs from our embryonic and P30 scRNA-Seq data to a list of 160 candidate disease-associated genes from the iPSYCH2012 project (Pedersen et al., 2018; Schork et al., 2019). Using DEGs from pairwise comparisons between interneuron maturation stages, we found 73 hits present in mouse scRNA-Seq DEGs that showed maturation stage restricted expression patterns. (Fig 7E-F). Of these DEGs, many were restricted to progenitors (APs/BPs) or mature P30 interneurons (Figure 7G), consistent with previous findings (Grove et al., 2019; Schork et al., 2019). Candidate disease-associated genes from iPSYCH that were hits among DEGs had similar distributions of odds ratios (the chance of developing disease given a SNP) across maturation stage (Figure 7H). Focusing on specific embryonic (AP/BP/N) and mature interneuron cell types, we identified DEGs between one cell type versus all others and found 70 disease-associated hits that were expressed in specific cell types (Figure 7I). The number of cell type markers with iPSYCH hits was greatest in embryonic APs, followed by BPs, and then neurons (Figure 7J). In mature P30 interneurons, there were significantly more DEGs expressed in the MGE-derived *Pvalb*- and *Sst*-expressing populations compared to CGE-derived cells (Figure 7J). Individual SNPs with the highest odds ratios were largely restricted to progenitors, immature neurons and *Ndnf*-, *Pvalb*-, and *Sst*-expressing interneurons (Figure 7K).

One potential mechanism by which SNPs in non-coding regions disrupt gene function is altering transcription factor binding to cREs. To explore whether any SNPs from the iPSYCH2012

cohort mapped to cREs, we compared our embryonic and P30 ATAC data sets with these SNPs. We detected candidate gene-enhancer interactions by combining Cicero and SnapATAC TSS-cRE predictions and retained presumptive enhancers regulating the 70 cell type markers that had iPSYCH hits (Figure 7I). Coordinates for the presumptive enhancers were converted from the mouse genome (GRCm10/mm10) to the human genome build (GRC37/hg19) used for iPSYCH2012 SNPs using the UCSC Genome Browser's LiftOver utility, resulting in 2,338 presumptive enhancers that successfully mapped between species. These presumptive enhancers were intersected with SNP coordinates associated with the 70 cell type markers, resulting in 5 presumptive enhancers that contain disease-associated SNPs from iPSYCH: rs142351870 (48778362:C:T, p-value 6.323e-07), rs34559126 (49575913:G:GTAT, p-value 3.263e-06), rs12631989 (49273996:C:T, p-value 1.967e-05), rs13064780 (49274826:T:G, p-value 1.967e-05), and rs4850441 (198913413:A:G, p-value 1.495e-04). Genes regulated by these 5 disease-associated SNPs were restricted to embryonic cell types (APs, BPs, N) and *Pvalb* and *Sst* expressing MGE-derived mature interneurons (Figure 7L). Three of the disease associated SNPs were located on chromosome 3, with SNPs rs12631989 and rs13064780 separated by only a couple kb with identical Hi-C interactions. We plotted these three SNPs and associated Hi-C interactions from the human dataset with the corresponding accessibility profiles from our scATAC-seq data along the same genomic region. Of note, all three SNPs interact with numerous loci (based on human Hi-C interactions) that are also accessible in our mouse dataset, with many of the strongest peaks found in the embryonic tissue (Figure 7M). These insights reveal that our single cell transcriptome and accessibility profiles from mouse embryonic can be used to query disease-associated SNPs from GWAS hits in human studies and could open new avenues for understanding how cREs influence disease by perturbing interneuron fate and maturation.

DISCUSSION

We utilized single nuclei chromatin accessibility and transcriptomic analyses to characterize changes to the chromatin accessibility landscape associated with early interneuron differentiation stages. While several studies performed single cell sequencing experiments to characterize the chromatin accessibility landscape in the mouse (Markenscoff-Papadimitriou et al., 2020; Preissl et al., 2018; Ziffra et al., 2019) and human (Markenscoff-Papadimitriou et al., 2020; Preissl et al., 2018; Ziffra et al., 2019) forebrain, we present the first detailed characterization of chromatin accessibility profiles from distinct embryonic brain regions that generate diverse neuronal cell populations. This approach allowed us to compare both the spatial (between different regions) and temporally (between different progenitor cell types) changes in chromatin accessibility during the initial stages of neurogenesis. We identified presumptive enhancers for specific genes involved in well-characterized neuronal lineages and brain regions. We detected large reorganizations of the chromatin accessibility landscape as cells transitioned from neuronal progenitors with high fate capacity to lineage restricted postmitotic precursors and neurons. By utilizing pattern analysis with our integrated scRNA-seq and snATAC-seq data, we characterized how the mRNA, GAS and presumptive enhancers of individual genes change during initial interneuron differentiation and across different neurogenic cell types. The improved resolution of our chromatin accessibility landscape and accompanying unbiased gene pattern analysis provides insights into mechanisms that regulate early stages of interneuron fate determination during embryogenesis.

We developed a robust method for predicting presumptive enhancers (Figure 4), and several enhancers were substantiated using embryonic forebrain H3K27ac ChIP-Seq data from the ENCODE project (Gorkin et al., 2020). We found that global transcriptional state and chromatin accessibility at promoters/gene bodies were highly correlated, and that chromatin accessibility at promoters and H3K4me3 Cut&Tag peaks were also correlated albeit to a lesser extent. In most instances, there was a lag for GAS and enhancer peaks to be downregulated compared to RNA downregulation, both at the individual gene level (Figures 5E-J) and global

level (Figures 6S-U). This observation implies that at least some repressive mechanisms (i.e. repressor TFs, DNA methylation, etc.) likely precede repressive histone modifications and decreased chromatin accessibility at cREs. Focusing on individual genes, we also observed numerous instances where accessibility of promoters, gene body and/or enhancers precede transcript upregulation (see *Nkx2-1* in Figures 5G-H and *Myt1l* and *Lhx6* in Figures 6K-R), which is in agreement with several recent reports (Ma et al., 2020; Ziffra et al., 2019). Future multiomics studies that can simultaneously capture the epigenome and transcriptome within single cells during development should provide significant insight into this relationship. Additionally, using CRISPR-based tools to manipulate candidate fate-determining cREs (Adli, 2018) will be required to characterize enhancers and genomic loci in initial cell fate decisions.

We note several additional intriguing observations from our data. First, the chromatin accessibility profiles reveal significant diversity in AP clusters from the GEs, much more so than cortical APs (Figure 2). Much of the reported transcriptional diversity within the GEs has been restricted to postmitotic cells (Mayer et al., 2018; Mi et al., 2018), so our data suggests that there may be greater transcriptional and chromatin state diversity in APs within the GEs than previously appreciated.

Second, the number and score of global Cicero connections (orange arcs from Figures 3B-C) are only loosely correlated to TSS-cRE connections. In some instances, the number of global Cicero connections at certain genomic loci appear to be tissue specific (greater co-accessibility in regions flanking *Neurod6* restricted to cortex and *Lhx6* restricted to MGE), while global Cicero connections near other genes appear more similar between tissues (similar co-accessibility for all brain regions flanking *Sp8* and *Nr2f2*). This may suggest a role of Cis-Co-accessibility Networks (CCANs) (Pliner et al., 2018), modules of connection sites that are highly co-accessible with one another in specific brain regions during neuronal fate determination.

Third, the population of ‘mixed’ cells that were collected with the LGE and CGE tissue expressed markers for both GABAergic and cortical glutamatergic cells yet formed a distinct

cluster from these populations (Figures 2 and S1). These 'mixed' cells likely reside at the pallial/subpallial boundary as no comparable cells were detected from the MGE which is further away from the pallial-subpallial boundary. And they are unlikely to be doublets since nuclei with extreme read depths were removed during QC steps (Figure S1). Further work is needed to better characterize this cell population.

Using the iPsych database, we found several disease-associated SNPs that were accessible in both embryonic interneuron progenitors and mature *Pvalb*- and *Sst*-expressing interneurons. A recent report characterized genome-wide cross-species Hi-C contact patterns, including conserved and lineage-specific patterns in both local interactions and long-range interactions (Yang et al., 2019), which raises the possibility that insights gleaned from the current mouse-to-human comparison could open new avenues for studying mammalian regulatory elements. At the same time, approximately one third of mouse ATAC coordinates were unmapped during the liftover from mouse to human coordinates. These peaks could denote mouse specific regulatory features and highlight that distinct higher-order chromatin organizations likely exist between species. As the vast majority of disease-associated SNPs are found in non-coding regions, linking these SNPs with accessibility profiles in specific cell types during developmental time periods, as we have done here, will be critical for understanding their relationship to disease going forward. The abundance of large-scale GWAS for specific neurodevelopmental diseases such as autism (Autism Spectrum Disorders Working Group of The Psychiatric Genomics, 2017; Grove et al., 2019; Polimanti and Gelernter, 2017; Robinson et al., 2016) and schizophrenia (Pardinas et al., 2018; Schizophrenia Working Group of the Psychiatric Genomics, 2014), as well as other multi-psychiatric disease consortiums (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2019; Wang et al., 2018), will make this a fruitful endeavor in the future.

STAR METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
    - Lead Contact
    - Materials Availability
    - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
    - Mice
- METHOD DETAILS
    - Tissue dissection
    - Nuclei extraction for single nuclei ATAC-seq and Cut&Tag
    - Cell dissociation for single cell RNA-Seq
    - Nuclei extraction for single nuclei RNA-seq
    - snATAC library preparation and sequencing
    - scRNA/snRNA library preparation and sequencing
    - Cut&Tag library preparation and sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
    - scRNA/snRNA-Seq Analysis
    - snATAC-Seq Analysis
    - Cut&Tag Analysis

ACKNOWLEDGEMENTS

We thank C. J. McBain, K. Pelkey and J. A. Kassis for helpful discussions; S. Henikoff for pA-Tn5 transposome aliquot.

AUTHOR CONTRIBUTIONS

C.T.R and T.J.P. designed the study and wrote the paper. C.T.R., Y.Z. and T.J.P. extracted and purified nuclei. D.R.L. and T.J.P. extracted and purified embryonic cells. Y.Z. prepared single cell sequencing libraries. C.T.R., A.M., D.R.L. and R.K.D. analyzed single cell data. D.J.L. and J.J.T. performed Cut&Tag and D.J.L., J.J.T. and P.R. analyzed Cut&Tag data.

STAR METHODS

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Timothy J. Petros (tim.petros@nih.gov).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

The GEO accession numbers for the sequencing data reported in the paper are GSE167047 (embryonic snATAC-Seq), GSE167048 (P30 snATAC-Seq), GSE167013 (embryonic scRNA-Seq), GSE167016 (P30 snRNA-Seq).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

The following mouse lines were used in this study: C57BL/6J (JAX# 000664), *CAG-Sun1/sfGFP* (B6.129-*Gt(ROSA)26Sor*$^{tm5(CAG\text{-}Sun1/sfGFP)Nat}$/MmbeJ, JAX# 030952) (Mo et al., 2015), and *Dlx6a-Cre* (Tg(dlx5a-cre)1Mekk/J, JAX# 008199) (Monory et al., 2006). For timed matings, noon on the day a vaginal plug was observed was denoted E0.5. All experimental procedures were conducted in accordance with the National Institutes of Health guidelines and were approved by the NICHD Animal Care and Use Committee (ACUC). Both male and female embryonic and adult mice were used without bias for the single cell sequencing reactions.

METHOD DETAILS

Tissue dissection

To recover embryonic tissue, dams were anesthetized with isoflurane and then euthanized by cervical dislocation. Embryos were removed from the uterus and kept in chilled artificial cerebrospinal fluid (ACSF, in mM: 87 NaCl, 26 $NaHCO_3$, 2.5 KCl, 1.25 NaH_2PO_4, 0.5 $CaCl_2$, 7 $MgCl_2$, 10 glucose, 75 sucrose, saturated with 95% O_2, 5% CO_2, pH 7.4). Brains were extracted from E12.5 and E14.5 embryos, hemisected, microdissected to obtain the MGE, LGE, CGE and somatosensory cortex, and kept in ACSF. To recover adult tissue, P30 *Dlx6a-Cre;*CAG-Sun1/sfGFP mice were anesthetized with isoflurane and decapitated. Fresh hippocampus and primary somatosensory cortex regions were microdissected and kept in chilled ACSF.

Nuclei extraction for single nuclei ATAC-seq and Cut&Tag

Nuclei isolation followed the 10X Genomics ATAC nuclei isolation protocol with several modifications. All steps were performed on ice. For each brain region, tissue was transferred to a Dounce homogenizer containing 1mL Lysis buffer (10 mM Tris-HCL, pH 7.4, 10 mM NaCl, 3 mM

MgCl2, 0.1% Tween-20, 0.1% IGEPAL, 2% BSA). Samples were dounced with 10 strokes pestle A and 10 strokes pestle B. Lysate was strained through a 40 µm filter pre-wetted with wash buffer (10 mM Tris-HCL, pH 7.4, 10 mM NaCl, 3 mM MgCl2, 0.1% Tween-20, 2% BSA) and neutralized with 2 mL wash buffer.

For snATAC-Seq from embryonic tissue, lysates were centrifuged 500 g for 5 min at 4°C, supernatant removed, nuclei pellet washed once with 1 mL wash buffer, and centrifuged 500 g for 5 min. For snATAC-Seq from adult tissue, DAPI (1:10,000 from 1 mg/ml stock) was added to the nuclei suspensions which were passed through a 30 µm filter. GFP+ nuclei were collected using a SH800 Cell Sorter (Sony Biotechnology) using a 100 µm nozzle. DAPI+/GFP+ nuclei were selected using Sony Cell Sorter Software (v2.1.5) and collected in low-bind 1.5 mL tubes containing 50 µL wash buffer. After sorting, nuclei were centrifuged 1000 g for 10 min at 4°C in a swinging bucket centrifuge. For both embryonic and adult samples, a diluted aliquot of nuclei solution was mixed with Trypan Blue (1:1) and counted on a hemocytometer. For all samples, we prepared 3,000 nuclei/µL samples for snATACseq reactions, with 5 µL (~15,000 nuclei) used for each snATAC reaction.

For Cut&Tag, embryonic tissue was homogenized as described above. Nuclei suspensions were centrifuged 500 g for 5 min at 4°C, supernatant removed, and washed twice with 1 mL Cut&Tag wash buffer (1 mL of 1 M HEPES pH7.5 (Sigma #H3375-100G), 1.5 mL of 5 M NaCl, 12.5 µL of 2 M spermidine, 47.5 µL H_2O, 5X mini protease inhibitor tablets (Roche #11836170001)). A diluted aliquot of nuclei solution was mixed with Trypan Blue (1:1) and counted on a hemocytometer. Equal numbers of MGE, CGE, and LGE nuclei were pooled and diluted to a final concentration of 500 nuclei/µL in Cut&Tag wash buffer, with 100 µL (~50,000 nuclei) used for each Cut&Tag reaction.

Cell dissociation for single cell RNA-Seq

Embryonic tissue was dissected as described above. To collect whole cells, embryonic MGE, LGE, CGE and cortex tissue was incubated in 1 mg/mL Pronase (Roche #10165921001) in ACSF for 20 minutes at RT. Pronase solution was removed and 2 ml of reconstitution solution (1% fetal bovine serum + DNAse (1:10,000, Roche #47167280001)) in oxygenated ACSF was added. Cells were triturated sequentially with fire-polished large, medium and small bore Pasteur pipettes to mechanically dissociate tissue.

For scRNA-Seq, DRAQ5 (20 μM) and DAPI (1:10,000) was added to the single cell suspension, which was passed through a pre-wet 30 μm filter and then processed on an SH800 cell sorter to purify the sample. DRAQ5+/DAPI- live cells were collected in low-bind 1.5 mL tubes containing 100 μL ACSF. Cell solutions were centrifuged at 300 g for 5 min at 4 ^{o}C in a swinging bucket centrifuge and then counted on a hemocytometer. 15,000 cells (or highest amount recovered after sorting) was used for 10X Genomics scRNA-Seq experiments.

Nuclei extraction for single nuclei RNA-seq

Adult tissue from *Dlx6a-Cre;CAG-Sun1/sfGFP* mice was dissected, as described above. Nuclei was isolated following a previously reported method (Matson et al., 2018) and FACS enriched for Draq5+/GFP+ nuclei as described above.

snATAC library preparation and sequencing

snATAC reaction was carried out following 10X Genomics ATAC User Guide (revision C), libraries were prepared following 10X Genomics and Illumina guidelines, and sequenced on an Illumina HiSeq2500. Sequencing metrics were as follows: **CGE**: Replicate 1: Read pairs: 469,280,227; Estimated number of cells: 4,013; Median fragments per cell: 13,003; Fraction of fragments in peaks: 73.3%, Fraction of transposition events in peaks: 58.4%. Replicate 2: Read pairs: 115,917,541; Estimated number of cells: 4,530; Median fragments per cell: 11,082; Fraction of fragments in peaks: 80.3%, Fraction of transposition events in peaks: 72.1%. **MGE**: Replicate

1: Read pairs: 429,523,963; Estimated number of cells: 6,845; Median fragments per cell: 12,807; Fraction of fragments in peaks: 72.0%, Fraction of transposition events in peaks: 54.2%. Replicate 2: Read pairs: 107,989,853; Estimated number of cells: 3,465; Median fragments per cell: 12,278; Fraction of fragments in peaks: 77.4%, Fraction of transposition events in peaks: 67.3%. **LGE**: Replicate 1: Read pairs: 491,904,518; Estimated number of cells: 6,577; Median fragments per cell: 13,426; Fraction of fragments in peaks: 67.7%, Fraction of transposition events in peaks: 47.4%. Replicate 2: Read pairs: 115,545,951; Estimated number of cells: 4,769; Median fragments per cell: 10,660; Fraction of fragments in peaks: 77.7%, Fraction of transposition events in peaks: 65.8%. **Cortex (E12.5)**: Read pairs: 112,120,408; Estimated number of cells: 4,946; Median fragments per cell: 11,257; Fraction of fragments in peaks: 78.0%, Fraction of transposition events in peaks: 68.1%. **Cortex (E14.5)**: Read pairs: 433,510,039; Estimated number of cells: 4,108; Median fragments per cell: 16,810; Fraction of fragments in peaks: 74.2%, Fraction of transposition events in peaks: 63.6%. **Cortex (P30)**: Replicate 1: Read pairs: 437,124,392; Estimated number of cells: 2,355; Median fragments per cell: 63,184; Fraction of fragments in peaks: 54.5%, Fraction of transposition events in peaks: 35.0%; Replicate 2: Read pairs: 232,666,518; Estimated number of cells: 4,191; Median fragments per cell: 24,540; Fraction of fragments in peaks: 45.8%, Fraction of transposition events in peaks: 14.8%. **Hippocampus (P30)**: Replicate 1: Read pairs: 220,442,648; Estimated number of cells: 5,576; Median fragments per cell: 20,478; Fraction of fragments in peaks: 49.4%, Fraction of transposition events in peaks: 30.4%. Replicate 2: Read pairs: 219,203,262; Estimated number of cells: 3,184; Median fragments per cell: 28,519; Fraction of fragments in peaks: 51.1%, Fraction of transposition events in peaks: 25.3%.

scRNA/snRNA library preparation and sequencing

cDNA libraries were prepared using 10X Genomics 3’ RNA v3 chemistry. Library preparation was carried out following 10X Genomics RNA User Guide (rev C) and sequenced

following 10X Genomics and Illumina guidelines. Samples were sequenced to the following depths: **CGE**: Reads: 154,348,231; Estimated number of cells: 4,522; Median reads per cell: 34,133; Median genes per cell: 2,805. **MGE**: Reads: 192,833,398; Estimated number of cells: 6,331; Median reads per cell: 30,459; Median genes per cell: 2,877. **LGE**: Reads: 180,580,907; Estimated number of cells: 6,843; Median reads per cell: 26,389; Median genes per cell: 2,726. **Cortex (E12.5)**: Reads: 176,542,467; Estimated number of cells: 7,453; Median reads per cell: 23,687; Median genes per cell: 2,318. **Cortex (P30)**: Reads: 259,151,993; Estimated number of cells: 7,857; Median reads per cell: 32,984; Median genes per cell: 3,267. **Hippocampus (P30)**: Reads: 217,933,257; Estimated number of cells: 5,623; Median reads per cell: 38,757; Median genes per cell: 3,345.

Cut&Tag library preparation and sequencing

Cut&Tag was performed as previously described (Kaya-Okur et al., 2019) (https://www.protocols.io/view/bench-top-cut-amp-tag-bcuhiwt6). 10 µL of activated Concanavalin A coated magnetic beads was added to an aliquot of ~50,000 nuclei in Cut&Tag Wash Buffer (20 mM HEPES pH 7.5; 150 mM NaCl; 0.5 mM Spermidine; 1x Protease inhibitor cocktail) and incubated 15 min at room temperature (RT).

The unbound supernatant was removed and bead-bound nuclei were resuspended in 50 Dig-wash Buffer (20 mM HEPES pH 7.5; 150 mM NaCl; 0.5 mM Spermidine; 1x Protease inhibitor cocktail; 0.05% Digitonin) containing 2 mM EDTA and a 1:50 dilution of the appropriate primary antibody. Primary antibody incubation was performed on a rotating platform for 2 h at room temperature (RT). The primary antibody was removed by placing the tube on the magnet stand to clear and pulling off all of the liquid. To increase the number of Protein A binding sites for each bound antibody, an appropriate secondary antibody (such as Guinea Pig anti-Rabbit IgG antibody for a rabbit primary antibody) was diluted 1:50 in 50 µL of Dig-Wash buffer and cells were incubated at RT for 1 h. Cells were washed twice in 1 mL Dig-Wash buffer for 5 minutes and

placed on the magnet stand to remove unbound antibodies. A 1:100 dilution of pA-Tn5 adapter complex (~5 μM) was prepared in Dig-med Buffer (0.05% Digitonin, 20 mM HEPES, pH 7.5, 300 mM NaCl, 0.5 mM Spermidine, 1x Protease inhibitor cocktail). After removing the liquid on the magnet stand, 100 μL of pA-Tn5 dilution was added to the nuclei with gentle vortexing and incubated at RT for 1 h. Cells were washed twice in 1 mL Dig-Wash buffer for 5 minutes and placed on the magnet stand to remove unbound pA-Tn5 protein complex. Next, cells were resuspended in 300 μL Tagmentation buffer (10 mM MgCl2 in Dig-med Buffer) and incubated at 37 °C, 500 rpm, for 1 h. To stop tagmentation, 10 μL of 0.5 M EDTA, 3 μL of 10% SDS and 2.5 μL of 20 mg/mL Proteinase K was added to 300 μL of sample followed by incubation at 55 °C for 30 min or overnight at 37 °C, and then at 70 °C for 20 min to inactivate Proteinase K. To extract the DNA, 300 μL PCI was added to each sample and mixed by full-speed vortexing. Solutions were transferred to phase-lock tubes and centrifuged at 16,000 g for 5 min at RT. After centrifugation, 300 μL chloroform was then added to each tube and mixed via inversion. Tubes were centrifuged once more at 16,000 g for 5 min at RT. The aqueous phase was then transferred to tubes containing 750 μL chilled 100% ethanol and centrifuged at 16,000 g for 15 min at 4 ºC. Supernatant was decanted and 1 mL 100% ethanol was added before centrifuging a final time for 2 min at 4 ºC. Supernatant was decanted and pellets were dried before resuspending in 20-25 μL 10 mM Tris pH 8.

To amplify libraries, 20 μL DNA was mixed with 10 μL 5x Q5 reaction buffer, 0.5 μL Q5 polymerase, 1 μL dNTPs, 16 μL water, and 2.5 μL of a universal i5 and a uniquely barcoded i7 primer, using a different barcode for each sample. The sample was placed in a Thermocycler with a heated lid using the following cycling conditions: 72 °C for 5 min (gap filling); 98 °C for 30 sec; 14-18 cycles of 98 °C for 10 sec and 63 °C for 30 sec; final extension at 72 °C for 1 min and hold at 8 °C. Post-PCR clean-up was performed by performing a SPRIselect bead 0.6x left/1x right double size selection, then washed twice gently in 80% ethanol, and eluted in 10-12 μL 10 mM Tris pH 8.0. Multiplexed libraries were pooled and paired-end sequenced (25 x 25 bp) on an

Illumina HiSeq2500. Samples were sequenced to a following depths per library: **Ventral E12**: Replicate 1: 3 million; Replicate 2: 3 million.

QUANTIFICATION AND STATISTICAL ANALYSIS

scRNA/snRNA-Seq Analysis

Cellranger

The Cellranger (v3.0.0) pipeline was used to process single cell/nuclei RNA-Seq libraries, all steps used default parameters unless otherwise noted. Cellranger mkfastq converted BCL files generated from sequencing to demultiplexed FASTQ files. Reads were aligned to 10X Genomics's pre-built mouse (GRCm38/mm10) reference genome with Cellranger count. For single cell RNA-Seq libraries, 10X Genomics's pre-built mRNA model of GRCm38/mm10 annotation (refdata-cellranger-mm10-3.0.0) was used to map reads to genes, while single nuclei RNA-Seq libraries used a corresponding pre-mRNA model constructed per 10X Genomics guidelines. Aligned reads were de-duplicated, filtered for valid cell barcodes and used to construct gene-by-barcode matrices.

Seurat

Filtered gene-by-barcode matrices were used as input to Seurat (v3.0.0, https://satijalab.org/seurat/) in R (v.4.0.0, https://cran.r-project.org/). For each cell barcode, summary statistics were calculated for the metadata columns n_Features, n_genes, and percent reads mapping to mitochondrial genes (if present). Outliers within the metadata columns were detected using Tukey's fence method for far-out outliers ([Q1 - k(Q3 - Q1), Q3 + k(Q3 - Q1)], where k = 3) which is resistant to extreme values (https://datatest.readthedocs.io/en/stable/how-to/outliers.html) and barcodes with any metadata column that contained outliers were removed. The remaining barcodes were processed using the SCTransform workflow in Seurat.

snATAC-Seq Analysis

Cellranger-atac

The Cellranger-atac (v1.2.0) pipeline was used to process single nuclei ATAC-Seq data, all steps used default parameters unless otherwise noted. Cellranger-atac mkfastq converted BCL files generated from sequencing to demultiplexed FASTQ files. Reads were aligned to 10X Genomics's pre-built mouse (GRCm38/mm10) reference genome and genomic annotation (refdata-cellranger-atac-mm10-1.2.0) by Cellranger-atac count. Libraries were aggregated and downsampled to equal numbers of median fragments per nuclei by Cellranger-atac aggr. Aligned reads were de-duplicated, filtered for valid cell barcodes, and constructed peak-by-barcode and TF-by-barcode matrices and fragments file.

Signac

Filtered peak-by-barcode matrix and fragments file were analyzed using Signac (v1.0.0) (https://satijalab.org/signac/index.html), all steps used default parameters unless otherwise noted. For each cell, summary statistics were calculated for the metadata columns n_Features, n_genes, and percent reads mapping to mitochondrial genes (if present). Outliers within the metadata columns were detected using Tukey's fence method for far-out outliers, as previously described, and barcodes with any metadata column that contained outliers were removed. Following pre-processing, peak-by-barcode matrices were processed as instructed by Signac documentation: normalized by Term Frequency - Inverse Document Frequency (TF-IDF) method, datasets were integrated to remove batch effects, dimensional reduction with Singular Value Decomposition (SVD) and Uniform Manifold Approximation and Projection (UMAP), and cluster detection using smart local moving (SLM) algorithm. RNA cell type predictions from scRNA-Seq were transferred to snATAC-Seq nuclei using the label transfer method outlined by Signac. Following the standard Signac workflow, promoter accesbility (PA) for each gene was calculated

by summing all reads mapping to a gene body plus promoter (2 Kb upstream from gene TSSs). Differential testing we performed for peak counts and PA by Seurat::FindAllMarkers(min.pct = 0.2,test.use = 'LR',latent.vars = 'nCount_peaks') and Seurat::FindAllMarkers(), respectively. Motif analysis was performed using Signac by Seurat::FindMarkers(only.pos=TRUE,min.pct=0.2,test.use='LR',latent.vars='nCount_peaks') followed by Signac::FindMotifs() and using the JASPAR (http://jaspar.genereg.net/) CORE vertebrates collection as a reference database. Per cell motif deviation scores were computed using chromVAR (v1.12.0) and the UCSC mm10 genome sequences (BSgenome.Mmusculus.UCSC.mm10, v1.4.0), both from Bioconductor (v3.12), with the Signac::RunChromVAR() wrapper.

Cicero

Following outlier detection and dimensional reduction, the filtered peak-by-barcode counts matrix was transferred from the Seurat object to a Monocle3 (https://cole-trapnell-lab.github.io/monocle3/) cell_data_set (CDS) object. The counts object was processed with Cicero (v1.9.1) (https://cole-trapnell-lab.github.io/cicero-release/) following user documentation with defaults unless otherwise specified. Prior to calculating Cicero co-accessibility scores, a Cicero CDS object was created using the newly created Monocle3 cds object and UMAP coordinates from the Seurat object (i.e. cicero::make_cicero_cds(cds.object,reduced_coordinates=seurat.object@reductions$umap@cell.embeddings)).

SnapATAC

snATAC-Seq reads were processed with SnapATAC (v1.0.0) independently of Seurat/Signac/Cicero workflows following user documentation (https://github.com/r3fang/SnapATAC) with default parameters unless otherwise specified.

FASTQ reads were converted into snap files using Snaptools (v1.4.1) and Python 2.7 and ATAC accessible regions were binned into 5000 kB windows and counted to create cell-by-bin matrices. Snap files were analyzed by SnapATAC. Valid cell barcodes with log10(UMI) between 3 to 6, and promoter ratio between 0.05 to 0.6 were retained for further analysis. All other barcodes were discarded. cell-by-bin matrices were merged, binarized, filtered against ENCODE blacklist regions (http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/), and barcodes with bin counts less than 1000 and greater than the 95th percentile were removed.

Dimensional reduction was performed using Nyström landmark diffusion maps algorithm, batch reduction was performed using Harmony (https://github.com/immunogenomics/harmony) by SnapATAC::runHarmony(eigs.dim=1:40), a K Nearest Neighbor (KNN) graph constructed by SnapATAC::runKNN(eigs.dims=1:dims_use,k=15) and the Louvain algorithm used to detect clusters. Gene activity scores were calculated by counting snATAC-Seq reads mapping to known gene bodies. scRNA-Seq/snRNA-Seq based cell type predictions were transferred to corresponding snATAC-Seq barcodes by SnapATAC::snapToSeurat(eigs.dims=1:40,norm=TRUE,scale=TRUE), Seurat::FindTransferAnchors() and Seurat::TransferData() as described in SnapATAC documentation. Cluster specific peaks were detected using MACS2 by SnapATAC::runMACS(macs.options="--nomodel --shift 100 --ext 200 --qval 5e-2 -B --SPMR"). Gene-enhancer pairs were predicted by SnapATAC::predictGenePeakPair() for every TSS that had a non-zero peak count across all barcodes. The gene-enhancer pairs list was filtered by removing pairs where FDR > 0.05.

Enhancer Prediction

The greatest SnapATAC -log10(adjusted p-value) value was multiplied by 2 as an upper limit, and all -log10(adjusted p-values) were percentile ranked per gene. Similarly, Cicero co-accessibility scores were percentile ranked for each gene. Peaks detected by SnapATAC or

Cicero that did not overlap gene TSSs were extended 500 bp upstream and downstream before merging overlapping peak coordinates using the GenomicRanges (v1.42.0) Bioconductor package with GenomicRanges::reduce(). To generate a list of all presumptive enhancers, merged SnapATAC peaks and merged Cicero peaks were concatenated. To generate a more stringent list of enhancers for snATAC-Seq/scRNA-Seq integration, the intersection SnapATAC and Cicero peaks was taken by GenomicRanges::intersect().

To detect H3K27ac+ presumptive enhancers, concatenated peaks were further filtered by retaining only peaks that intersected forebrain H3K27ac ChIP-Seq peaks for age E12.5 (UCSC Database: mm10, Primary Table: encode3Ren_forebrain_H3K27ac_E12, Big Bed File: /gbdb/mm10/encode3/histones/ENCFF957YEE.bigBed) or E14.5 (UCSC Database: mm10, Primary Table: encode3Ren_forebrain_H3K27ac_E14, Big Bed File: /gbdb/mm10/encode3/histones/ENCFF088LWR.bigBed) by GenomicRanges::SubsetByOverlaps().

scRNA-Seq/snATAC-Seq Integration

Integration of single cell/nuclei RNA-Seq and single nuclei ATAC-Seq datasets was performed using Seurat following Signac documentation. For each data type, barcodes were filtered to remove far-out outliers, as previously described by Tukey's fence method. Prior to integrating data, for snATAC-Seq barcodes, gene activity score (GAS) was calculated by counting reads that mapped to the promoter regions (2kb upstream of TSSs), the first exon of each transcript, and (if detected) presumptive enhancer loci associated to each gene.

degPatterns

To obtain a list of differentially expressed genes (DEGs), Seurat::FindMarkers() was used to detect DEGs from embryonic and P30 scRNA-Seq data. Barcodes were ordered by increasing pseudotime, as assigned by Monocle3. The Seurat object containing integrated snATAC-

Seq/scRNA-Seq datasets was subset into 2 Seurat objects, one containing scRNA-Seq and one containing snATAC-Seq. degPatterns were quantified using the previously detected DEGs list and 3 separate matrices: gene-by-barcode counts (from scRNA-Seq Seurat object), gene-by-barcode counts (GAS slot from snATAC-Seq Seurat object) and gene-by-barcode counts (enhancer slot from snATAC-Seq Seurat object) by DEGReport:: degPatterns() using the DEGReport (https://github.com/lpantano/DEGreport) package.

RNA/GAS/Enhancers Heatmaps

The top 500 highly variable genes from embryonic scRNA-Seq data were detected by Seurat::FindVariableFeatures(nfeatures = 500). Barcodes in the Seurat object containing integrated snATAC-Seq/scRNA-Seq were ordered by increasing pseudotime, as assigned by Monocle3. This Seurat object was subset into 2 Seurat objects, one containing scRNA-Seq and one containing snATAC-Seq, which were further split into 3 matrices: gene-by-barcode counts (from scRNA-Seq Seurat object), gene-by-barcode counts (GAS slot from snATAC-Seq Seurat object) and gene-by-barcode counts (enhancer slot from snATAC-Seq Seurat object). RNA, GAS and enhancer gene-by-barcode matrices were concatenated together, and hierarchical clustering of genes was performed using correlation distance and average linkage metrics. Following clustering, the concatenated matrix was re-split into RNA, GAS and enhancer gene-by-barcode matrices. Matrices were centered and scaled for each gene prior to rendering heatmaps and the gene dendrograms were constructed using the distances calculated with the concatenated matrix. Manually annotated color bars were based on gene cohorts detected by degPatterns().

iPsych Comparison

Disease-associated genes for SNPs associated with psychiatric disorders from the iPSYCH2012 dataset (https://static-content.springer.com/esm/art%3A10.1038%2Fs41593-018-0320-0/MediaObjects/41593_2018_320_MOESM3_ESM.xlsx) were intersected with DEGs from

integrated embryonic scRNA-Seq and P30 snRNA-Seq datasets detected by Seurat. Genes were visualized using EnhancedVolcano (v1.8.0), venneuler (v1.1.0) and ggplot2 (v3.3.0).

To detect enhancers that overlap SNP coordinates, coordinates for presumptive enhancers from the integrated embryonic and P30 ATAC-Seq/RNA-Seq dataset were converted from the mouse (GRCm38/mm10 build) to human (hg19 build) using UCSC Genome Browser's LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). The intersection of SNP coordinates and presumptive enhancers was taken with GenomicRanges::SubsetByOverlaps().

Cut&Tag Analysis

Unless otherwise specified, libraries were processed with the ENCODE Histone ChIP-Seq Processing Pipeline (https://www.encodeproject.org/chip-seq/histone/). Reads were mapped to a reference genome following the **Pipeline for Mapping of FASTQs**: For each library, FASTQs from all technical replicates were concatenated and aligned to mouse genome (GRCm38/mm10 build) with the Bowtie2 aligner, PCR duplicates removed with PICARD. For visualization of histone marks, normalized bigwig files were created using DeepTools, and filtered to retain only fragments greater than 150 base-pairs. Peak calling followed the **Pipeline for Histone ChIP-seq peak calling (Replicated experiments):** Peak calling and signal generation for each biological replicate was done by MACS. Replicate concordance (within group) was performed using the 'overlap_peaks.py' script from the ENCODE histone ChIP-Seq pipeline, and peaks with an irreproducible discovery rate (IDR) < 0.05 were retained for further analysis. Between groups peak call comparisons were performed with DiffBind Bioconductor package.

FIGURE LEGENDS

Figure 1. Chromatin accessibility in the embryonic forebrain is cell type and state specific

(A) Schematic of snATAC-Seq workflow and neurogenic cell types: apical progenitors (APs), basal progenitors (BPs) and neurons (Ns).

(B-E) UMAP visualization of single nuclei clustered by brain region (B), SLM (C), neurogenic cell type (D) and pseudotime (E). In (D), PA = promoter accessibility, representing reads mapping within 2 kb upstream of TSSs.

(F-M) PA scores for genes enriched in specific neurogenic cell types (F-I) or distinct brain regions (J-M).

(N-Q) Aggregated reads per SLM cluster. *Nkx2-1* (MGE), *Neurod6* (cortex), *Nes* (APs) and *Dcx* (BPs/Ns), arranged by either brain region (N-O) or neurogenic cell type (P-Q). The y-axis range for chromatin accessibility tracks are normalized to the maximum reads per gene. Peaks: differentially accessible peak coordinates, Cut&Tag: H3K4me3 signal from E12.5 ventral forebrain using Cut&Tag, ENCODE: H3K4me3 signal from E12.5 forebrain from ENCODE ChIP-Seq data. H3K4me3 Cut&Tag signal tracks had range 0-825, H3K4me3 ENCODE signal had range 0-155.

Figure 2. Differentially accessible peaks are cluster and lineage specific

(A) Embryonic snATAC-Seq dot plot of differentially accessible peaks (DA peaks) for each cluster. Dot diameter indicates the percent of DA peaks from one cluster (column cluster labels) which are detectable in any other cluster (row cluster labels). Color intensity represents the total DA peak count per cluster. Hierarchical clustering was performed using correlation distance and average linkage.

(B-E) UMAP plots of nuclei colored by ChromVAR global motif deviations (Z-scores) for NR2F1, ASCL1, EMX1 and NEUROD2.

Figure 3. Detection of cis-regulatory elements within the developing mouse forebrain

(A) Heatmap depicting DA peaks binned along pseudotime from embryonic snATAC-Seq nuclei. DA peaks were filtered to retain only peaks that had at least one Cicero peak-peak connection. Stacked bar plots to left of heatmap depicts proportion of AP, BP and N nuclei per bin.

(B-C) Cicero connections within 0.5 Mbp window centered around TSS of *Lhx6* (postmitotic MGE marker), *Neurod6* (postmitotic cortex marker), *Nr2f2* (CGE progenitors), *Sp8* (LGE progenitors), *Nes* (pan-AP) and *Gad1* (pan-interneuron) broken down based on tissue (B) or neurogenic cell type (C). Cicero connections overlapping the TSS of selected genes shown as purple arcs, all Cicero connections in genomic region shown as orange arcs. Y-axis unit is co-accessibility score. Only connections with co-accessibility scores greater than 0.25 are depicted. Peaks: snATAC peaks used by Cicero to quantify peak-peak connections. Gene models are visualized at the bottom of each column with gene of interest highlighted.

Figure 4. Characterization of enhancers in embryonic neurogenic tissue.

(A) Gene-enhancer interactions originating at the *Sox2* locus as defined by Cicero (cyan arcs), SnapATAC (magenta arcs), the combination of Cicero and SnapATAC (black arcs) and the intersection of Cicero and SnapATAC (dark blue arcs, representing high confidence enhancers). Window is 0.5 Mb centered on TSS. Arc height represents interaction score.

(B-E) Visualization of intersection of Cicero and SnapATAC gene-peak interactions for *Nes* (B), *Neurog2* (C), *Gsx2* (D) and *Dlx1* (E). Window is 100 Kb, centered on the gene TSS. Peaks: displays co-accessible coordinates, Encode: displays H3K27ac signal from E12.5 forebrain from ENCODE ChIP-Seq data, VISTA: enhancers with positive activity in the forebrain and midbrain from the VISTA Enhancer Browser.

(F) Bar chart depicting the total number of gene-peak interactions detected by Cicero, SnapATAC and high-confidence interactions via the intersection of Cicero and SnapATAC (“merged” label).

(G) Bar chart of the number of gene-peak interactions binned by gene-peak distance for high-confidence gene-peak interactions.

Figure 5. Integrative analysis of embryonic snATAC-Seq and scRNA-Seq

(A) Workflow depicting integration of embryonic snATAC-Seq (top left) and scRNA-Seq (top right) data. Bottom, UMAP plot showing integrated snATAC-Seq nuclei and scRNA-Seq cells colored by tissue.

(B-D) UMAP visualization of integrated snATAC-Seq/scRNA-Seq data colored by assay (B), Louvain cluster (C), and pseudotime (D)

(E-J) UMAP visualization of integrated scRNA-Seq cells and snATAC-sec nuclei colored by transcript counts or GAS for *Nes* (E-F), *Nkx2-1* (G-H) and *Neurod2* (I-J). Grey dots in background represent cells/nuclei from other assay.

(K-M) Genome browser tracks displaying enhancer predictions regulating *Nkx2-1* (K), *Neurod2* (L), *Ascl1* (M) and *Emx1* (N). Gene-Peak interactions are visualized in blue arcs, arc heights indicates the relative interaction scores between gene TSS (red line) and peaks. snATAC-Seq Peaks: displays co-accessible coordinates, H3K27ac ENCODE: Forebrain E12.5 H3K27ac peaks from ENCODE project, VISTA: enhancers from the VISTA Genome Browser project.

Figure 6. Transcription, gene accessibility and active enhancer utilization are highly correlated and dynamic during neuron lineage commitment

(A-B) Line charts of 'early expressed' (A) and 'late expressed' (B) DEG clusters detected by degPatterns using embryonic integrated snATAC-Seq/scRNA-Seq data. Y-axis is Z-score for RNA, GAS, or enhancers counts per gene. X-axis is binned pseudotime periods. RNA, GAS, or enhancers for individual genes in these clusters are plotted (156 'early' and 90 'late' genes).

(C-R) Line chart and UMAP visualizations of RNA, GAS and enhancer read counts for representative 'early' genes *Hes1* and *Lmo1* (C-J) and 'late' genes *Myt1l* and *Lhx6* (K-R). For line

charts, Y-axis is Z-score for RNA, GAS, or enhancers counts per gene. X-axis is binned pseudotime periods. Grey dots in background of UMAP plots represent cells/nuclei from other assay.

(S-U) Heatmaps of RNA (S), GAS (T) and enhancers (U) counts for 210 variable genes. Heatmap columns were ordered by hierarchical clustering of 210 variable genes with correlation distance and average linkage. Rows were ordered by pseudotime (assigned by Moncole3). Color bars in (S) indicate genes grouped together based on 'early', 'transition' and 'late' profiles (as assigned by degPatterns and manually refined).

Figure 7. Interneuron development and psychiatric disorders share developmentally critical genes

(A-D) UMAP visualization of integrated embryonic and adult integrated scRNA-Seq and snATAC-Seq data colored by assay (A), tissue (B), cell type (C) and pseudotime (D).

(E) Volcano plot of DEGs from embryonic and adult scRNA-seq data per maturation stage that have hits in the iPSYCH2012 disease-associated genes.

(F) Venn diagram of DEGs that have hits in the iPSYCH2012 disease-associated genes.

(G) Number of DEGs with hits expressed in embryonic progenitors (AP/BP), immature neurons (N) and mature neurons (P30).

(H) Box plot of odds ratios (ORs) of iPSYCH2012 disease-associated genes that are DEG hits.

(I) Volcano plot of DEGs from embryonic and adult scRNA-seq data per cell type that have hits in the iPSYCH2012 disease-associated genes.

(J) Number of DEGs hits expressed in defined cell populations.

(K) Box plot of odds ratios (ORs) of iPSYCH2012 disease-associated genes that are DEG hits.

(L) Bar graph of DEGs that have hits in the iPSYCH2012 disease-associated genes and an associated enhancer from integrated embryonic and adult snATAC-Seq and scRNA-Seq data.

(M) Representative iPSYCH2012 candidate SNP that intersects an enhancer from embryonic and adult integrated snATAC-Seq/scRNA-Seq data. Interactions between SNP and disease-associated genes are visualized as either red arcs (rs34559126) or green arcs (rs126631989 and rs130664780), and height of connection is OR value for SNP. Gold boxes indicate genes with Hi-C contacts with all three SNPs, red and green boxes indicate genes contacting only rs34559126 or rs126631989/rs13064780, respectively. ATAC signal tracks were normalized by library depth and cell type abundance.

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY FIGURES

Figure S1 (Supplementary to Figure 1). snATAC-Seq and scRNA-Seq quality control and batch reduction, and scRNA-Seq based snATAC-Seq annotation

(A) Violin plots of snATAC-Seq QC metrics before outlier removal.

(B) Violin plots of snATAC-Seq QC metrics after outlier removal.

(C-E) UMAP visualization of snATAC-Seq nuclei prior to batch removal colored by replicate (C), tissue origin (D) and SLM cluster (E).

(F-H) UMAP visualization of batch corrected snATAC-Seq nuclei colored by replicate (F), tissue origin (G) and SLM cluster (H).

(I) Violin plots of scRNA-Seq QC metrics before outlier removal.

(J) Violin plots of scRNA -Seq QC metrics after outlier removal.

(K-N) UMAP visualization of scRNA-Seq cells colored by tissue (K), neuronal class (L), Louvain cluster (M) and cell type (N).

(O-Q) UMAP visualization of snATAC-Seq nuclei colored by neuronal class (O), scRNA-Seq cell type predictions (P) and assigned neurogenic cell type (Q).

Figure S2 (Supplementary to Figure 1). Visualization of GAS for critical interneuron markers

UMAP visualization of GAS from snATAC-Seq nuclei for specified genes. Legend depicts normalized read counts.

Figure S3 (Supplementary to Figure 2). DA Peaks per cluster and read distribution across gene features

(A) Embryonic snATAC-Seq signal plots of DA peaks for each cluster (same cluster order as Fig 2). Signal was binarized to indicate DA peaks from one cluster (column cluster labels) which are detectable in any other cluster (row cluster labels).

(B) For each cluster (rows of the dot plot), motifs representative of lineage and maturation stage, corresponding motif name and P-value are displayed to the right of the dot plot. All motifs are in the top five hits for each cluster, with top hits indicated with an asterisk.

(C) Bar chart of DA peak counts where -log(Fold Change) > 0 per SLM cluster colored by neurogenic cell type.

(D) Bar graph of mean DA peak counts where -log(Fold Change) > 0 per SLM cluster colored by neurogenic cell type and grouped by tissue.

(E) Bar chart of DA peak counts where -log(Fold Change) > 0 per SLM cluster colored by tissue.

(F) Bar chart of mean DA peak counts where -log(Fold Change) > 0 per SLM cluster colored by tissue and grouped by neurogenic cell type.

(G) Pie chart of relative enrichment of DA peaks within annotated gene features.

Figure S4 (Supplementary to Figure 5). Gene activity score comparisons

(A-D) UMAP visualizations of integrated snATAC-Seq and scRNA-Seq data colored by assay (left) and tissue samples (right). For integration, snATAC-Seq Gene Activity Score (GAS) was quantified using four different strategies: (A) quantified peaks within gene promoters (-2000 bp to

0 bp of 5' transcript coordinate) and gene bodies, (B) quantified peaks within gene promoters and the first exon of all transcripts, (C) quantified peaks within gene promoters only, and (D) quantified peaks within promoters, the first exon of all transcripts, and any associated putative enhancers (H3K27ac+ chromatin).

Figure S5 (Supplementary to Figure 5). Visualization of RNA counts and GAS following integration

UMAP visualization of transcript abundance from scRNA-Seq cells (left) and GAS from snATAC-seq nuclei (right) for genes enriched in specific brain regions (A-J) or in distinct neurogenic cell types (K-P).

Figure S6 (Supplementary to Figure 6). degPatterns analysis and heatmaps of RNA, GAS and enhancer counts for top 500 highly variable genes

(A-B) Line graphs of genes within each cluster detected by degPatterns that contained > 5 genes per cluster using embryonic integrated snATAC-Seq/scRNA-Seq data. Y-axis is Z-score for RNA, GAS, or enhancers counts per gene. X-axis is binned pseudotime periods with origin corresponding to early development and higher pseudotime values corresponding to more differentiated cells. RNA, GAS, or enhancers for each gene in this cluster are plotted individually.

(C) Bar graph of change of enhancer usage for all detected H3K27ac+ presumptive enhancers associated with DEGs per state transition: AP to BP, BP to N and AP to N. All detectable H3K27ac+ presumptive enhancers defined as genomic intersection of 1) the genomic union of (a) SnapATAC cREs and b) Cicero cREs with co-accessibility score >0.25) and 2) E12.5 ENCODE H3K27ac ChIP-Seq peaks.

(D) Bar graph of ratio of enhancer activation versus enhancer deactivation per state transition: AP to BP, BP to N and AP to N.

(E-G) Heatmaps of RNA (E), GAS (F) and enhancer (G) counts for top 500 highly variable genes from scRNA-Seq data and corresponding genes from snATAC-Seq GAS and enhancer matrices. Heatmap columns were ordered by hierarchical clustering with correlation distance and average linkage. Rows were ordered by pseudotime (assigned by Moncole3).

Figure S7 (Supplementary to Figure 7). Integration and visualization of embryonic and adult datasets

(A) UMAP visualization of embryonic and P30 scRNA-Seq cells colored by tissue (top) and cell type (bottom).

(B) UMAP visualization of embryonic and P30 snATAC-Seq nuclei colored by tissue (top) and cell type (bottom).

SUPPLEMENTARY TABLES

Table S1: DA peaks per cluster

Table S2: DA motifs per cluster

Table S3: DA motifs filtered by DEGs

Table S4: Embryonic enhancers

Table S5: High-Confidence Embryonic Enhancers

Table S6: High-Confidence Embryonic Enhancers Intersecting VISTA Enhancers

BIBLIOGRAPHY

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nat Commun *9*, 1911.

Alvarez-Dolado, M., Calcagnotto, M.E., Karkar, K.M., Southwell, D.G., Jones-Davis, D.M., Estrada, R.C., Rubenstein, J.L., Alvarez-Buylla, A., and Baraban, S.C. (2006). Cortical inhibition modified by embryonic neural precursors grafted into the postnatal brain. J Neurosci *26*, 7380-7389.

Arai, Y., and Taverna, E. (2017). Neural Progenitor Cell Polarity and Cortical Development. Front Cell Neurosci *11*, 384.

Autism Spectrum Disorders Working Group of The Psychiatric Genomics, C. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism *8*, 21.

Bandler, R.C., Mayer, C., and Fishell, G. (2017). Cortical interneuron specification: the juncture of genes, time and geometry. Curr Opin Neurobiol *42*, 17-24.

Brandao, J.A., and Romcy-Pereira, R.N. (2015). Interplay of environmental signals and progenitor diversity on fate specification of cortical GABAergic neurons. Front Cell Neurosci *9*, 149.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015a). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr Protoc Mol Biol *109*, 21 29 21-21 29 29.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015b). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486-490.

Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J.*, et al.* (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496-502.

Carullo, N.V.N., and Day, J.J. (2019). Genomic Enhancers in Brain Health and Disease. Genes (Basel) *10*.

Cross-Disorder Group of the Psychiatric Genomics Consortium, C. (2019). Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell *179*, 1469-1482 e1411.

Dall'Aglio, L., Muka, T., Cecil, C.A.M., Bramer, W.M., Verbiest, M., Nano, J., Hidalgo, A.C., Franco, O.H., and Tiemeier, H. (2018). The role of epigenetic modifications in neurodevelopmental disorders: A systematic review. Neurosci Biobehav Rev *94*, 17-30.

de Lanerolle, N.C., Kim, J.H., Robbins, R.J., and Spencer, D.D. (1989). Hippocampal interneuron loss and plasticity in human temporal lobe epilepsy. Brain Res *495*, 387-395.

DeFelipe, J., Lopez-Cruz, P.L., Benavides-Piccione, R., Bielza, C., Larranaga, P., Anderson, S., Burkhalter, A., Cauli, B., Fairen, A., Feldmeyer, D.*, et al.* (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. Nat Rev Neurosci *14*, 202-216.

Fang, R., Preissl, S., Hou, X., Lucero, J., Wang, X., Motamedi, A., Shiau, A.K., Mukamel, E.A., Zhang, Y., Behrens, M.M.*, et al.* (2019). Fast and Accurate Clustering of Single Cell Epigenomes Reveals Cis-Regulatory Elements in Rare Cell Types. bioRxiv.

Fishell, G., and Kepecs, A. (2020). Interneuron Types as Attractors and Controllers. Annu Rev Neurosci *43*, 1-30.

Flames, N., Pla, R., Gelman, D.M., Rubenstein, J.L., Puelles, L., and Marin, O. (2007). Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. J Neurosci *27*, 9682-9695.

Florio, M., and Huttner, W.B. (2014). Neural progenitors, neurogenesis and the evolution of the neocortex. Development *141*, 2182-2194.

Gelman, D., Griveau, A., Dehorter, N., Teissier, A., Varela, C., Pla, R., Pierani, A., and Marin, O. (2011). A wide diversity of cortical GABAergic interneurons derives from the embryonic preoptic area. J Neurosci *31*, 16570-16580.

Gelman, D.M., and Marin, O. (2010). Generation of interneuron diversity in the mouse cerebral cortex. Eur J Neurosci *31*, 2136-2141.

Gelman, D.M., Martini, F.J., Nobrega-Pereira, S., Pierani, A., Kessaris, N., and Marin, O. (2009). The embryonic preoptic area is a novel source of cortical GABAergic interneurons. J Neurosci *29*, 9380-9389.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Gorkin, D.U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A.Y., Li, B., Chiou, J., Wildberg, A., Ding, B.*, et al.* (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. Nature *583*, 744-751.

Gouwens, N.W., Sorensen, S.A., Baftizadeh, F., Budzillo, A., Lee, B.R., Jarsky, T., Alfiler, L., Baker, K., Barkan, E., Berry, K.*, et al.* (2020). Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. Cell *183*, 935-953 e919.

Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H., and Macklis, J.D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci *14*, 755-769.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R.*, et al.* (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat Genet *51*, 431-444.

Guo, J., and Anton, E.S. (2014). Decision making during interneuron migration in the developing cerebral cortex. Trends Cell Biol *24*, 342-351.

Hirabayashi, Y., and Gotoh, Y. (2010). Epigenetic control of neural precursor cell fate during development. Nat Rev Neurosci *11*, 377-388.

Hsieh, J., and Gage, F.H. (2004). Epigenetic control of neural stem cell fate. Curr Opin Genet Dev *14*, 461-469.

Hu, G., Cui, K., Fang, D., Hirose, S., Wang, X., Wangsa, D., Jin, W., Ried, T., Liu, P., Zhu, J.*, et al.* (2018). Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. Immunity *48*, 227-242 e228.

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

Hu, X.L., Wang, Y., and Shen, Q. (2012). Epigenetic control on cell fate choice in neural stem cells. Protein Cell *3*, 278-290.

Inan, M., Welagen, J., and Anderson, S.A. (2012). Spatial and temporal bias in the mitotic origins of somatostatin- and parvalbumin-expressing interneuron subgroups and the chandelier subtype in the medial ganglionic eminence. Cereb Cortex *22*, 820-827.

Kaar, S.J., Angelescu, I., Marques, T.R., and Howes, O.D. (2019). Pre-frontal parvalbumin interneurons in schizophrenia: a meta-analysis of post-mortem studies. J Neural Transm (Vienna) *126*, 1637-1651.

Kanatani, S., Yozu, M., Tabata, H., and Nakajima, K. (2008). COUP-TFII is preferentially expressed in the caudal ganglionic eminence and is involved in the caudal migratory stream. J Neurosci *28*, 13582-13591.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun *10*, 1930.

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

Kikuchi, M., Hara, N., Hasegawa, M., Miyashita, A., Kuwano, R., Ikeuchi, T., and Nakaya, A. (2019). Enhancer variants associated with Alzheimer's disease affect gene expression via chromatin looping. BMC Med Genomics *12*, 128.

Kriegstein, A., and Alvarez-Buylla, A. (2009). The glial nature of embryonic and adult neural stem cells. Annu Rev Neurosci *32*, 149-184.

Larimer, P., Spatazza, J., Espinosa, J.S., Tang, Y., Kaneko, M., Hasenstaub, A.R., Stryker, M.P., and Alvarez-Buylla, A. (2016). Caudal Ganglionic Eminence Precursor Transplants Disperse and Integrate as Lineage-Specific Interneurons but Do Not Induce Cortical Plasticity. Cell Rep *16*, 1391-1404.

Larimer, P., Spatazza, J., Stryker, M.P., Alvarez-Buylla, A., and Hasenstaub, A.R. (2017). Development and long-term integration of MGE-lineage cortical interneurons in the heterochronic environment. J Neurophysiol *118*, 131-139.

Lim, L., Mi, D., Llorca, A., and Marin, O. (2018). Development and Functional Diversification of Cortical Interneurons. Neuron *100*, 294-313.

Lothian, C., and Lendahl, U. (1997). An evolutionarily conserved region in the second intron of the human nestin gene directs gene expression to CNS progenitor cells and to early neural crest cells. Eur J Neurosci *9*, 452-462.

Ludwig, L.S., Lareau, C.A., Bao, E.L., Nandakumar, S.K., Muus, C., Ulirsch, J.C., Chowdhary, K., Buenrostro, J.D., Mohandas, N., An, X.*, et al.* (2019). Transcriptional States and Chromatin Accessibility Underlying Human Erythropoiesis. Cell Rep *27*, 3228-3240 e3227.

Lui, J.H., Hansen, D.V., and Kriegstein, A.R. (2011). Development and evolution of the human neocortex. Cell *146*, 18-36.

Ma, S., Zhang, B., LaFave, L.M., Earl, A.S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V.K., Tay, T.*, et al.* (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. Cell *183*, 1103-1116 e1120.

Ma, T., Zhang, Q., Cai, Y., You, Y., Rubenstein, J.L., and Yang, Z. (2012). A subpopulation of dorsal lateral/caudal ganglionic eminence-derived neocortical interneurons expresses the transcription factor Sp8. Cereb Cortex *22*, 2120-2130.

Magloczky, Z., and Freund, T.F. (2005). Impaired and repaired inhibitory circuits in the epileptic human hippocampus. Trends Neurosci *28*, 334-340.

Marin, O., and Muller, U. (2014). Lineage origins of GABAergic versus glutamatergic neurons in the neocortex. Curr Opin Neurobiol *26*, 132-141.

Markenscoff-Papadimitriou, E., Whalen, S., Przytycki, P., Thomas, R., Binyameen, F., Nowakowski, T.J., Kriegstein, A.R., Sanders, S.J., State, M.W., Pollard, K.S.*, et al.* (2020). A Chromatin Accessibility Atlas of the Developing Human Telencephalon. Cell *182*, 754-769 e718.

Matson, K.J.E., Sathyamurthy, A., Johnson, K.R., Kelly, M.C., Kelley, M.W., and Levine, A.J. (2018). Isolation of Adult Spinal Cord Nuclei for Massively Parallel Single-nucleus RNA Sequencing. J Vis Exp.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J.*, et al.* (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190-1195.

Maybury-Lewis, S.Y., Brown, A.K., Yeary, M., Sloutskin, A., Dhakal, S., McCarthy-Sinclair, B., Juven-Gershon, T., and Webb, A.E. (2020). Changing and stable chromatin accessibility supports transcriptional overhaul during neural stem cell activation. bioRxiv.

Mayer, C., Hafemeister, C., Bandler, R.C., Machold, R., Batista Brito, R., Jaglin, X., Allaway, K., Butler, A., Fishell, G., and Satija, R. (2018). Developmental diversification of cortical inhibitory interneurons. Nature *555*, 457-462.

McKenzie, M.G., Cobbs, L.V., Dummer, P.D., Petros, T.J., Halford, M.M., Stacker, S.A., Zou, Y., Fishell, G.J., and Au, E. (2019). Non-canonical Wnt Signaling through Ryk Regulates the Generation of Somatostatin- and Parvalbumin-Expressing Cortical Interneurons. Neuron *103*, 853-864 e854.

Mi, D., Li, Z., Lim, L., Li, M., Moissidis, M., Yang, Y., Gao, T., Hu, T.X., Pratt, T., Price, D.J.*, et al.* (2018). Early emergence of cortical interneuron diversity in the mouse embryo. Science *360*, 81-85.

Miri, M.L., Vinck, M., Pant, R., and Cardin, J.A. (2018). Altered hippocampal interneuron activity precedes ictal onset. Elife *7*.

Miyoshi, G., Butt, S.J., Takebayashi, H., and Fishell, G. (2007). Physiologically distinct temporal cohorts of cortical interneurons arise from telencephalic Olig2-expressing precursors. J Neurosci *27*, 7786-7798.

Miyoshi, G., and Fishell, G. (2011). GABAergic interneuron lineages selectively sort into specific cortical layers during early postnatal development. Cereb Cortex *21*, 845-852.

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

Mo, A., Luo, C., Davis, F.P., Mukamel, E.A., Henry, G.L., Nery, J.R., Urich, M.A., Picard, S., Lister, R., Eddy, S.R.*, et al.* (2016). Epigenomic landscapes of retinal rods and cones. Elife *5*, e11613.

Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., Henry, G.L., Picard, S., Urich, M.A., Nery, J.R., Sejnowski, T.J., Lister, R.*, et al.* (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron *86*, 1369-1384.

Mohamed Ariff, I., Mitra, A., and Basu, A. (2012). Epigenetic regulation of self-renewal and fate determination in neural stem cells. J Neurosci Res *90*, 529-539.

Monory, K., Massa, F., Egertova, M., Eder, M., Blaudzun, H., Westenbroek, R., Kelsch, W., Jacob, W., Marsch, R., Ekker, M.*, et al.* (2006). The endocannabinoid system controls key epileptogenic circuits in the hippocampus. Neuron *51*, 455-466.

Nakazawa, K., Zsiros, V., Jiang, Z., Nakao, K., Kolata, S., Zhang, S., and Belforte, J.E. (2011). GABAergic interneuron origin of schizophrenia pathophysiology. Neuropharmacology *62*, 1574-1583.

Nery, S., Fishell, G., and Corbin, J.G. (2002). The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations. Nat Neurosci *5*, 1279-1287.

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

Pantano, L. (2017). DEGreport: Report of DEG analysis. R package version 1.13.8.

Pardinas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L.*, et al.* (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet *50*, 381-389.

Parnavelas, J.G. (2000). The origin and migration of cortical neurones: new vistas. Trends Neurosci *23*, 126-131.

Pedersen, C.B., Bybjerg-Grauholm, J., Pedersen, M.G., Grove, J., Agerbo, E., Baekvad-Hansen, M., Poulsen, J.B., Hansen, C.S., McGrath, J.J., Als, T.D.*, et al.* (2018). The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. Mol Psychiatry *23*, 6-14.

Pelkey, K.A., Chittajallu, R., Craig, M.T., Tricoire, L., Wester, J.C., and McBain, C.J. (2017). Hippocampal GABAergic Inhibitory Interneurons. Physiol Rev *97*, 1619-1747.

Perenthaler, E., Yousefi, S., Niggl, E., and Barakat, T.S. (2019). Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. Front Cell Neurosci *13*, 352.

Petros, T.J., Bultje, R.S., Ross, M.E., Fishell, G., and Anderson, S.A. (2015). Apical versus Basal Neurogenesis Directs Cortical Interneuron Subclass Fate. Cell Rep *13*, 1090-1095.

Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A.*, et al.* (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol Cell *71*, 858-871 e858.

Podobinska, M., Szablowska-Gadomska, I., Augustyniak, J., Sandvig, I., Sandvig, A., and Buzanska, L. (2017). Epigenetic Modulation of Stem Cells in Neurodevelopment: The Role of Methylation and Acetylation. Front Cell Neurosci *11*, 23.

Polimanti, R., and Gelernter, J. (2017). Widespread signatures of positive selection in common risk alleles associated to autism spectrum disorder. PLoS Genet *13*, e1006618.

Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B.C., Afzal, V., Dickel, D.E.*, et al.* (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci *21*, 432-439.

Rapanelli, M., Frick, L.R., and Pittenger, C. (2017). The Role of Interneurons in Autism and Tourette Syndrome. Trends Neurosci *40*, 397-407.

Robinson, E.B., St Pourcain, B., Anttila, V., Kosmicki, J.A., Bulik-Sullivan, B., Grove, J., Maller, J., Samocha, K.E., Sanders, S.J., Ripke, S.*, et al.* (2016). Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nat Genet *48*, 552-555.

Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. Dev Neurobiol.

Rymar, V.V., and Sadikot, A.F. (2007). Laminar fate of cortical GABAergic interneurons is dependent on both birthdate and phenotype. J Comp Neurol *501*, 369-380.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods *14*, 975-978.

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427.

Schork, A.J., Won, H., Appadurai, V., Nudel, R., Gandal, M., Delaneau, O., Revsbech Christiansen, M., Hougaard, D.M., Baekved-Hansen, M., Bybjerg-Grauholm, J.*, et al.* (2019). A genome-wide association study of shared risk across psychiatric disorders implicates gene regulation during fetal neurodevelopment. Nat Neurosci *22*, 353-361.

Sokpor, G., Castro-Hernandez, R., Rosenbusch, J., Staiger, J.F., and Tuoc, T. (2018). ATP-Dependent Chromatin Remodeling During Cortical Neurogenesis. Front Neurosci *12*, 226.

Sokpor, G., Xie, Y., Rosenbusch, J., and Tuoc, T. (2017). Chromatin Remodeling BAF (SWI/SNF) Complexes in Neural Development and Disorders. Front Mol Neurosci *10*, 243.

Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R.*, et al.* (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell *154*, 888-903.

Sultan, K.T., Brown, K.N., and Shi, S.H. (2013). Production and organization of neocortical interneurons. Front Cell Neurosci *7*, 221.

Tanaka, D.H., Maekawa, K., Yanagawa, Y., Obata, K., and Murakami, F. (2006). Multidirectional and multizonal tangential migration of GABAergic interneurons in the developing cerebral cortex. Development *133*, 2167-2176.

Taniguchi, H., Lu, J., and Huang, Z.J. (2013). The spatial and temporal origin of chandelier cells in mouse neocortex. Science *339*, 70-74.

Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T.*, et al.* (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. Nat Neurosci.

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

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

Tremblay, R., Lee, S., and Rudy, B. (2016). GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. Neuron *91*, 260-292.

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

Turrero Garcia, M., and Harwell, C.C. (2017). Radial glia in the ventral telencephalon. FEBS Lett *591*, 3942-3959.

Tyson, J.A., Goldberg, E.M., Maroof, A.M., Xu, Q., Petros, T.J., and Anderson, S.A. (2015). Duration of culture and sonic hedgehog signaling differentially specify PV versus SST cortical interneuron fates from embryonic stem cells. Development *142*, 1267-1278.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res *35*, D88-92.

Waclaw, R.R., Allen, Z.J., 2nd, Bell, S.M., Erdelyi, F., Szabo, G., Potter, S.S., and Campbell, K. (2006). The zinc finger transcription factor Sp8 regulates the generation and diversity of olfactory bulb interneurons. Neuron *49*, 503-516.

Waddington, C.H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology (London,: Allen & Unwin).

Waltman, L., and van Eck, N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. Eur Phys J B *86*, 471.

Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F.C.P., Clarke, D., Gu, M., Emani, P., Yang, Y.T.*, et al.* (2018). Comprehensive functional genomic resource and integrative model for the human brain. Science *362*.

Wichterle, H., Garcia-Verdugo, J.M., Herrera, D.G., and Alvarez-Buylla, A. (1999). Young neurons from medial ganglionic eminence disperse in adult and embryonic brain. Nature Neuroscience *2*, 461-466.

Wichterle, H., Turnbull, D.H., Nery, S., Fishell, G., and Alvarez-Buylla, A. (2001). In utero fate mapping reveals distinct migratory pathways and fates of neurons born in the mammalian basal forebrain. Development *128*, 3759-3771.

Wickramasekara, R.N., and Stessman, H.A.F. (2019). Histone 4 Lysine 20 Methylation: A Case for Neurodevelopmental Disease. Biology (Basel) *8*.

Wonders, C.P., and Anderson, S.A. (2006). The origin and specification of cortical interneurons. Nat Rev Neurosci.

Wonders, C.P., Taylor, L., Welagen, J., Mbata, I.C., Xiang, J.Z., and Anderson, S.A. (2008). A spatial bias for the origins of interneuron subgroups within the medial ganglionic eminence. Dev Biol *314*, 127-136.

Xu, Q., Guo, L., Moore, H., Waclaw, R.R., Campbell, K., and Anderson, S.A. (2010). Sonic hedgehog signaling confers ventral telencephalic progenitors with distinct cortical interneuron fates. Neuron *65*, 328-340.

Yang, Y., Zhang, Y., Ren, B., Dixon, J.R., and Ma, J. (2019). Comparing 3D Genome Organization in Multiple Species Using Phylo-HMRF. Cell Syst *8*, 494-505 e414.

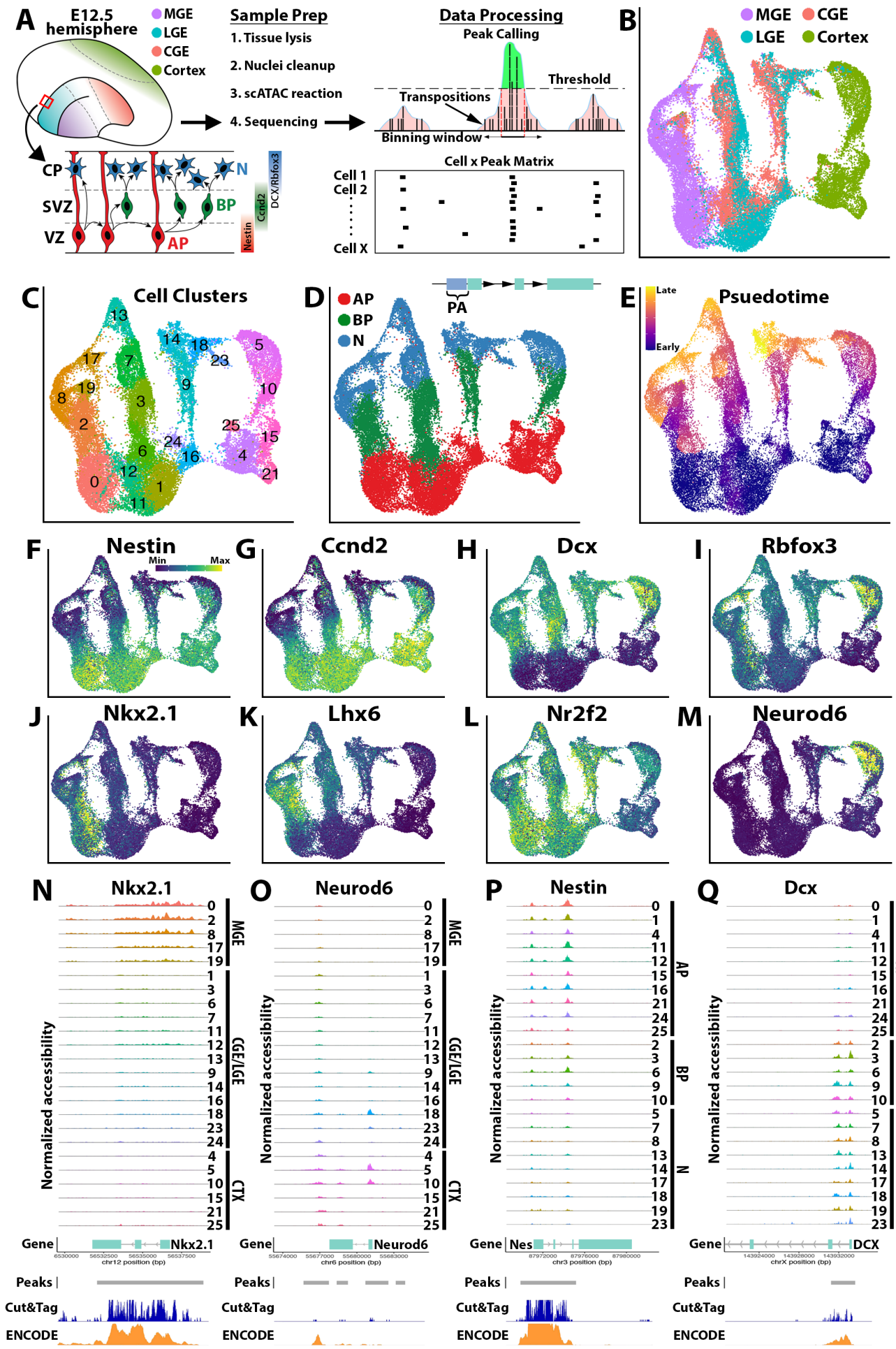
Yao, B., Christian, K.M., He, C., Jin, P., Ming, G.L., and Song, H. (2016). Epigenetic mechanisms in neurogenesis. Nat Rev Neurosci *17*, 537-549.

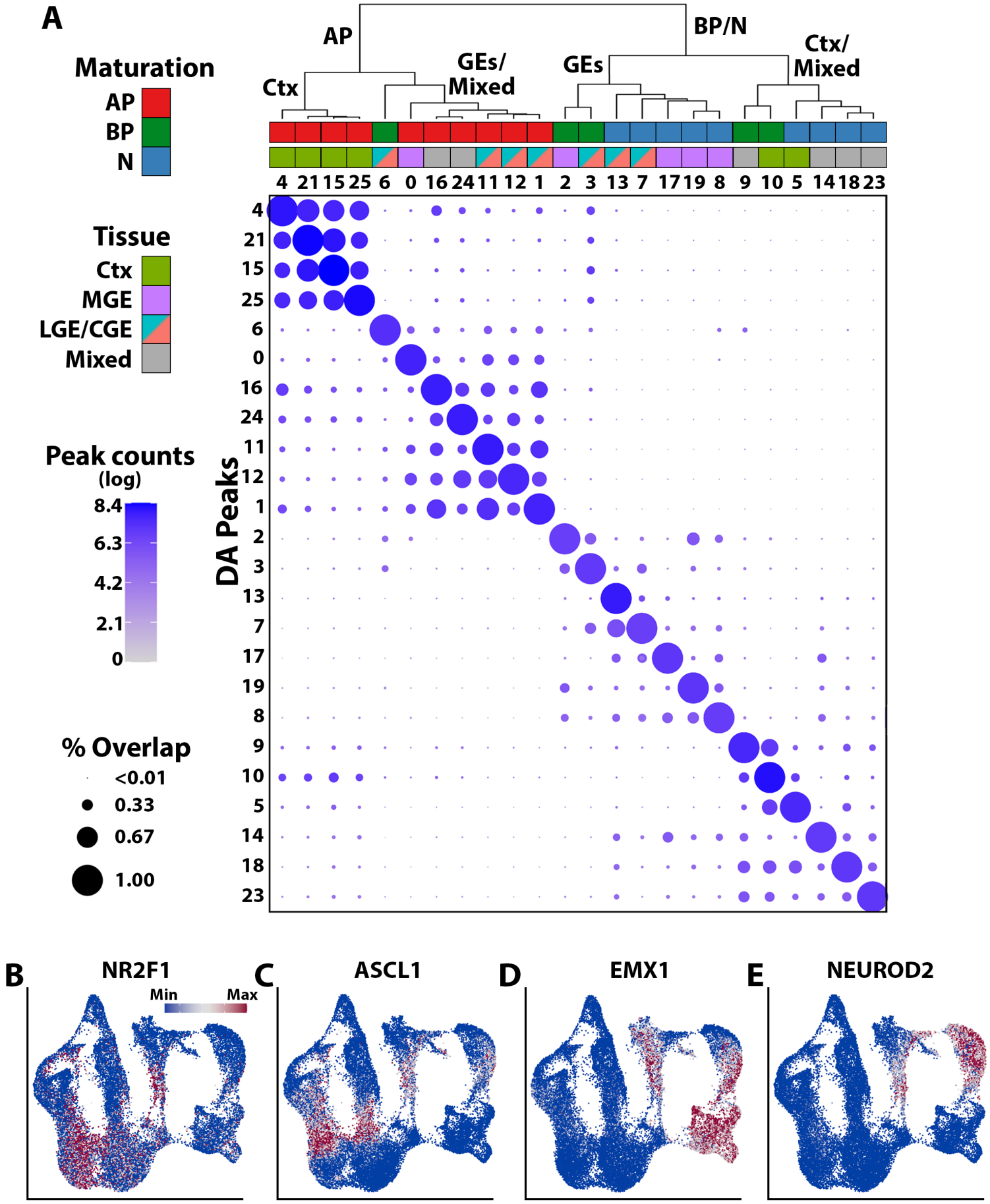
Ye, Y., Li, M., Gu, L., Chen, X., Shi, J., Zhang, X., and Jiang, C. (2017). Chromatin remodeling during in vivo neural stem cells differentiating to neurons in early Drosophila embryos. Cell Death Differ *24*, 409-420.

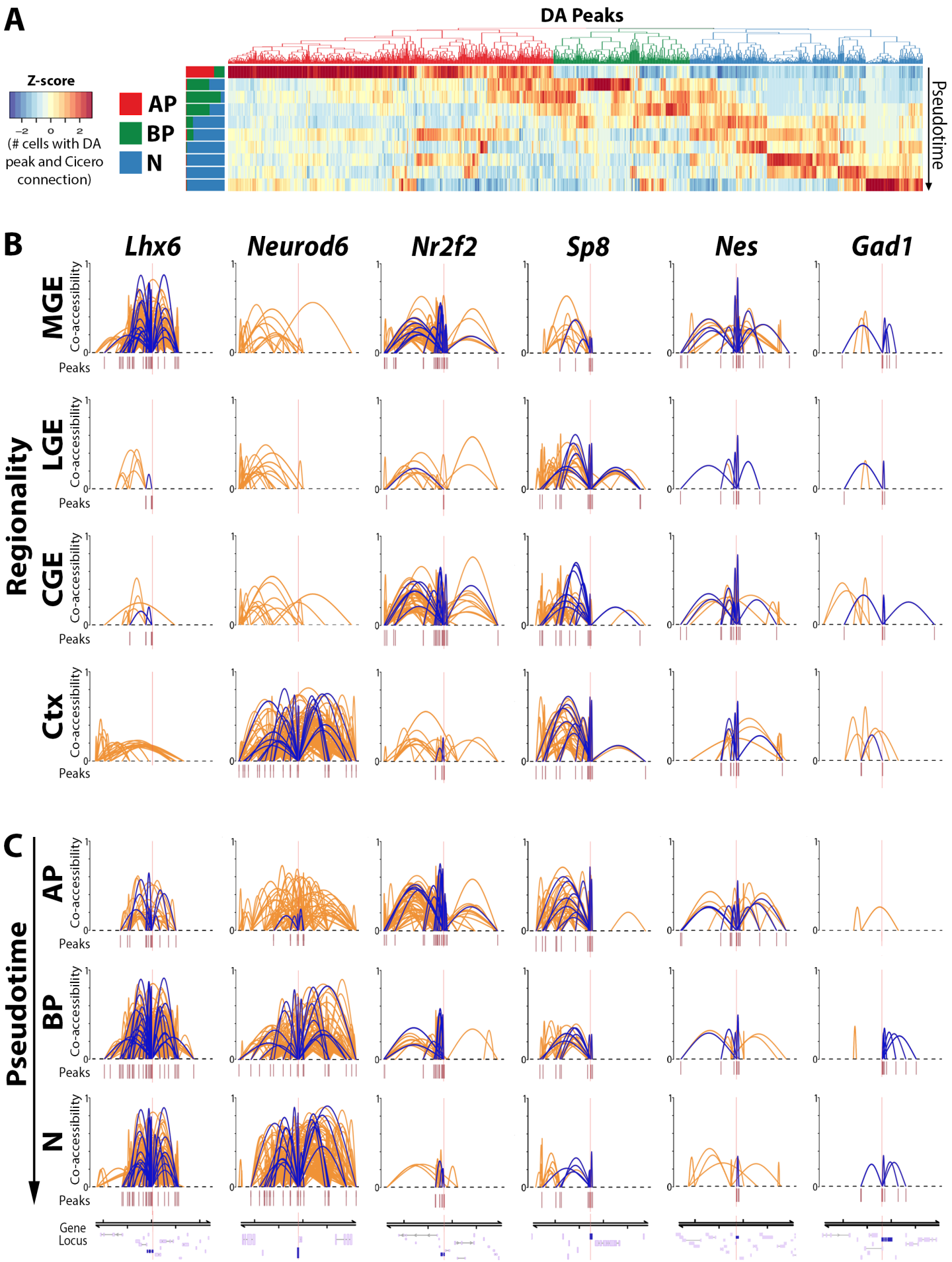
Zahir, F.R., and Brown, C.J. (2011). Epigenetic impacts on neurodevelopment: pathophysiological mechanisms and genetic modes of action. Pediatr Res *69*, 92R-100R.

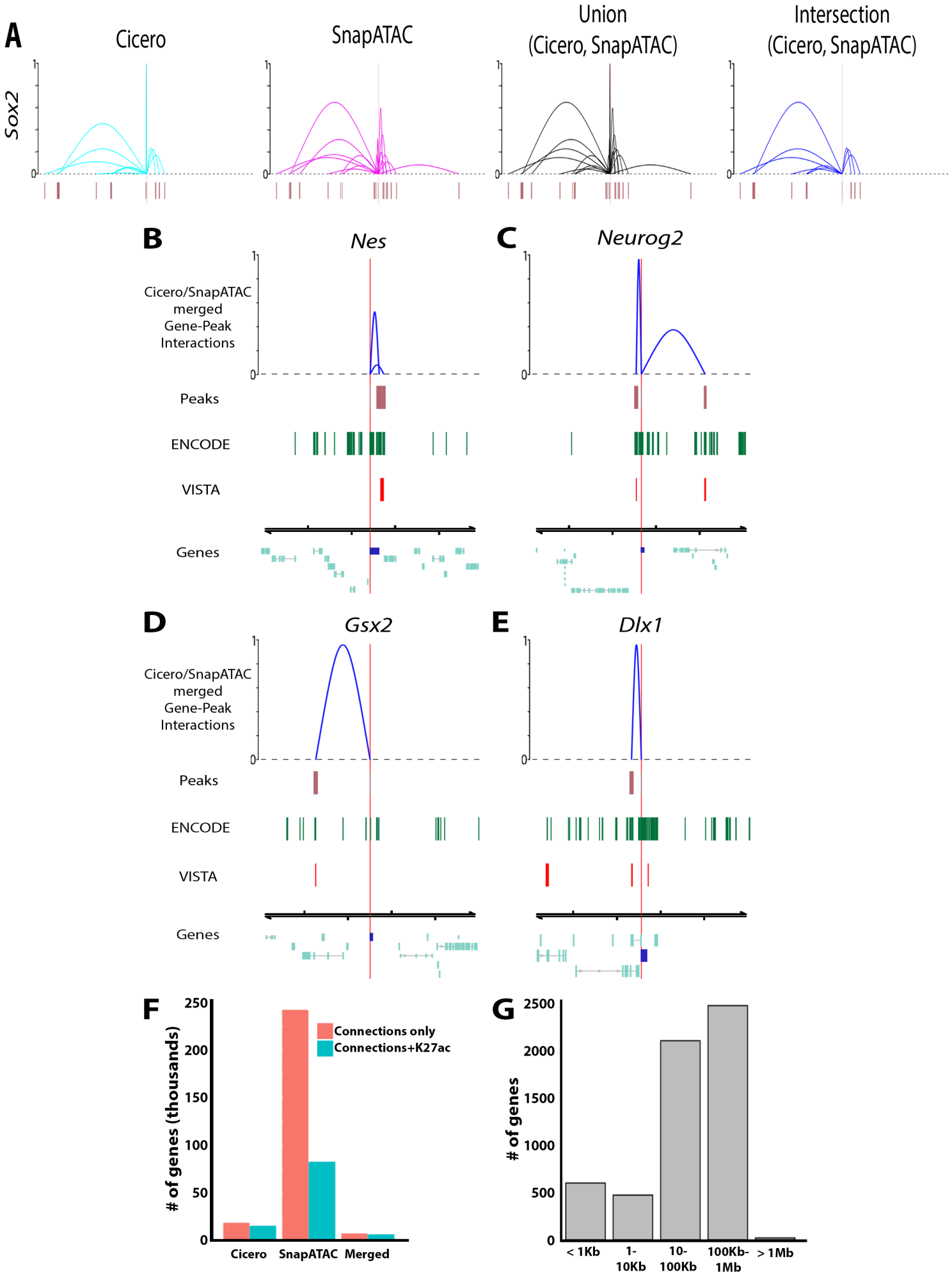
Ziffra, R.S., Kim, C.N., Wilfert, A., Haeussler, M., Casella, A.M., Przytycki, P.F., Kreimer, A., Pollard, K.S., Ament, S.A., Eichler, E.E.*, et al.* (2019). Single cell epigenomic atlas of the developing human brain and organoids. bioRxiv.

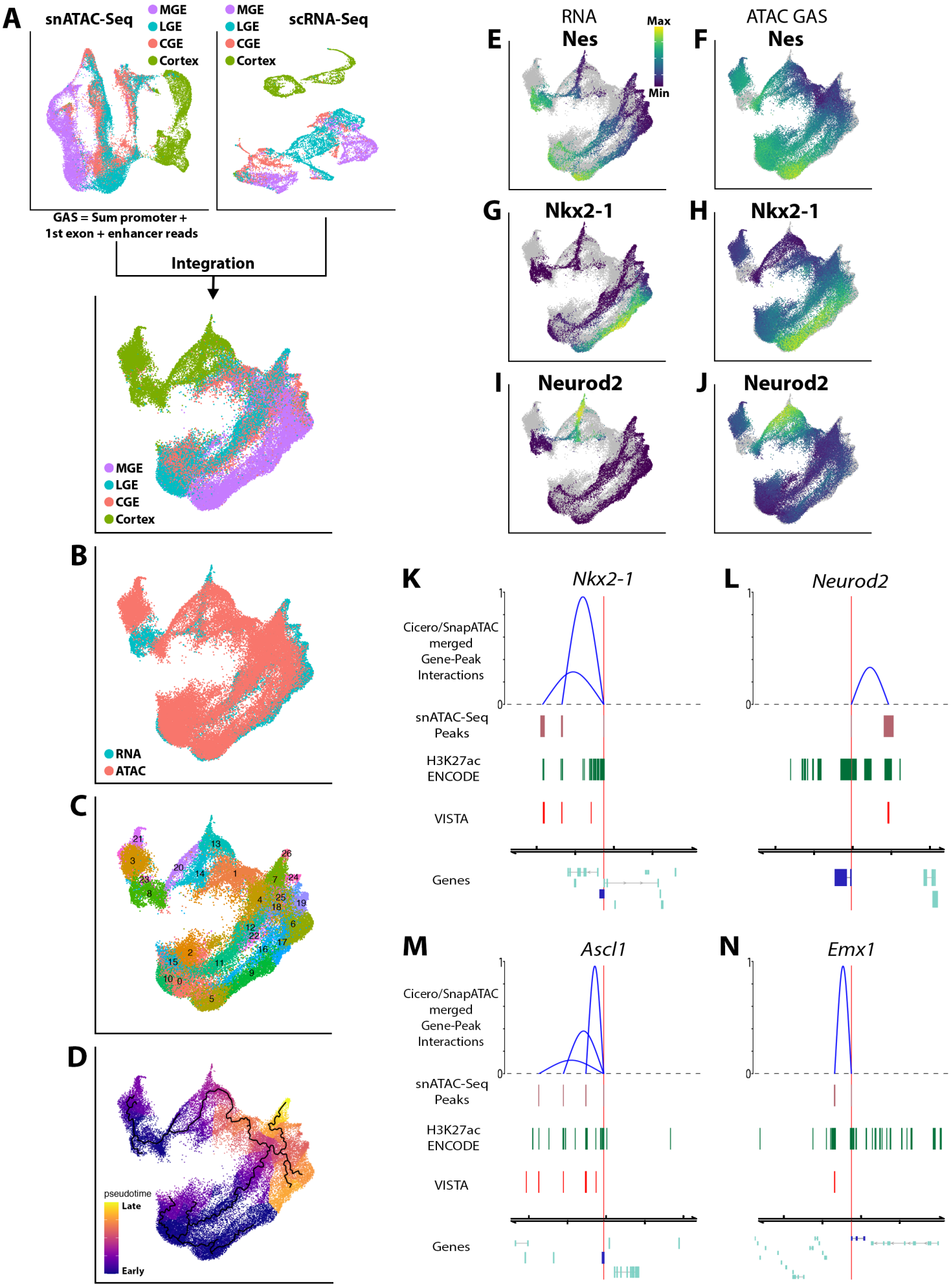
Zimmerman, L., Parr, B., Lendahl, U., Cunningham, M., McKay, R., Gavin, B., Mann, J., Vassileva, G., and McMahon, A. (1994). Independent regulatory elements in the nestin gene direct transgene expression to neural stem cells or muscle precursors. Neuron *12*, 11-24.

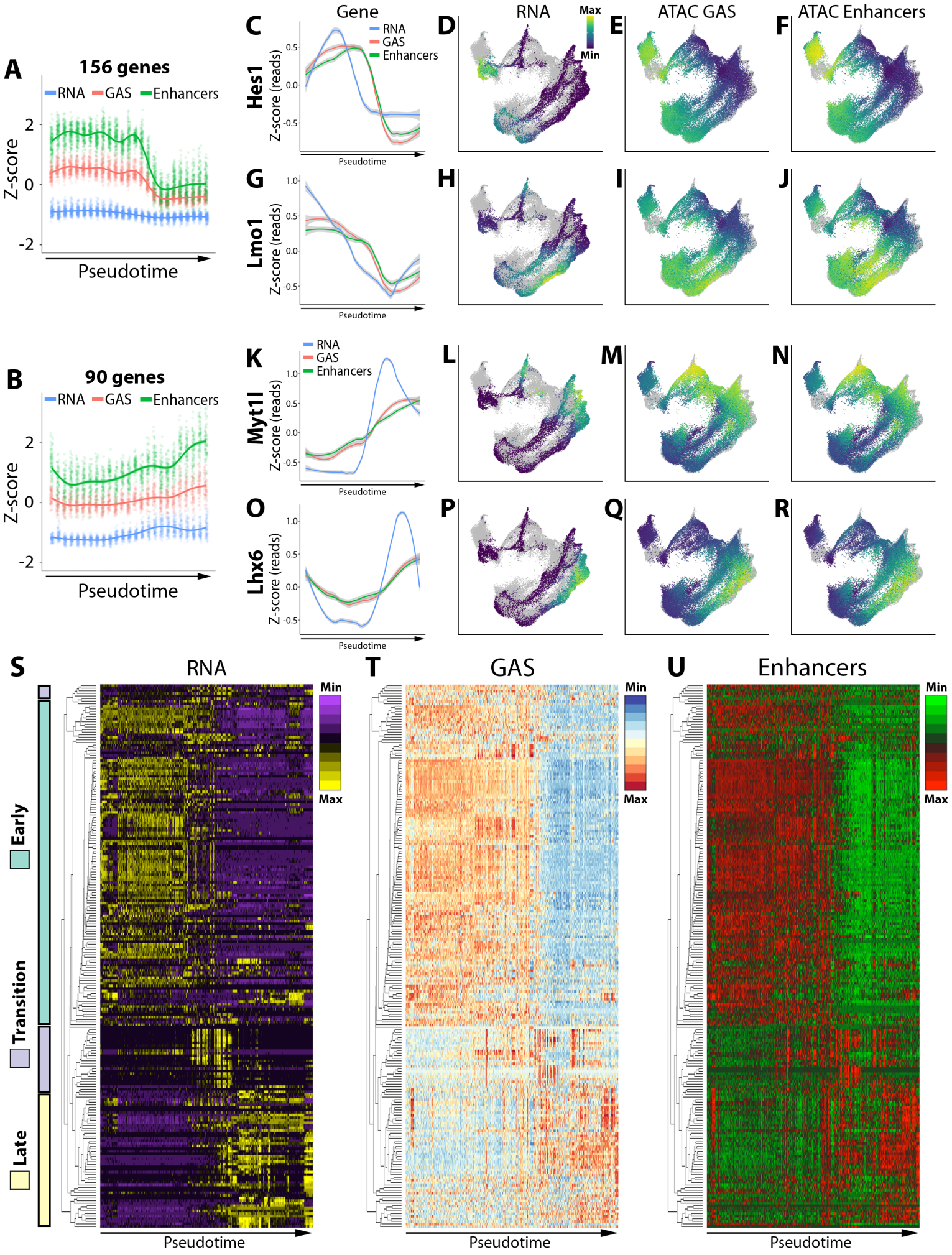

A
E12.5 hemisphere
MGE
LGE
CGE
Cortex
CP
SVZ
VZ
N
BP
AP
Nestin
Ccnd2
DCX/Rbfox3
Sample Prep
1. Tissue lysis
2. Nuclei cleanup
3. scATAC reaction
4. Sequencing
Data Processing
Peak Calling
Threshold
Transpositions
Binning window
Cell x Peak Matrix
Cell 1
Cell 2
Cell X
B
MGE
CGE
LGE
Cortex
C
Cell Clusters
D
AP
BP
N
PA
E
Psuedotime
Late
Early
F
Nestin
Min
Max
G
Ccnd2
H
Dcx
I
Rbfox3
J
Nkx2.1
K
Lhx6
L
Nr2f2
M
Neurod6
N
Nkx2.1
O
Neurod6
P
Nestin
Q
Dcx
Normalized accessibility
MGE
CGE/LGE
CTX
AP
BP
N
Gene
Nkx2.1
Neurod6
Nes
DCX
chr12 position (bp)
chr6 position (bp)
chr3 position (bp)
chrX position (bp)
Peaks
Cut&Tag
ENCODE

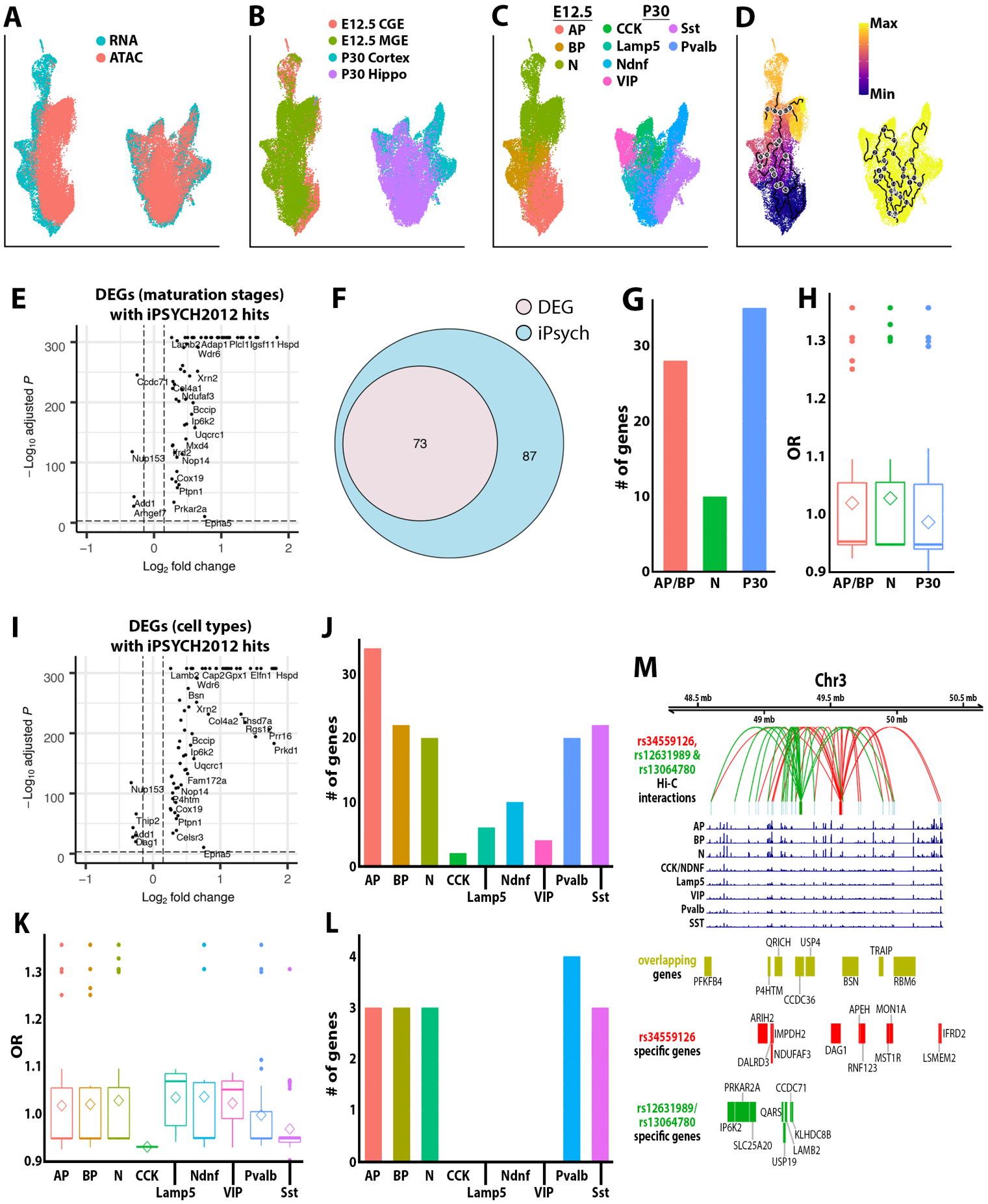

A
Maturation
AP
BP
N
Tissue
Ctx
MGE
LGE/CGE
Mixed
Peak counts
(log)
8.4
6.3
4.2
2.1
0
% Overlap
<0.01
0.33
0.67
1.00
AP
BP/N
Ctx
GEs/
Mixed
GEs
Ctx/
Mixed
4 21 15 25 6 0 16 24 11 12 1 2 3 13 7 17 19 8 9 10 5 14 18 23
DA Peaks
B
NR2F1
Min
Max
C
ASCL1
D
EMX1
E
NEUROD2


A
DA Peaks
Z-score
−2
0
2
(# cells with DA peak and Cicero connection)
AP
BP
N
Pseudotime
B
Lhx6
Neurod6
Nr2f2
Sp8
Nes
Gad1
Regionality
MGE
LGE
CGE
Ctx
Co-accessibility
Peaks
C
Pseudotime
AP
BP
N
Co-accessibility
Peaks
Gene
Locus


A
Cicero
SnapATAC
Union
(Cicero, SnapATAC)
Intersection
(Cicero, SnapATAC)
Sox2
B
Nes
C
Neurog2
Cicero/SnapATAC
merged
Gene-Peak
Interactions
Peaks
ENCODE
VISTA
Genes
D
Gsx2
E
Dlx1
F
# of genes (thousands)
Connections only
Connections+K27ac
Cicero
SnapATAC
Merged
G
# of genes
< 1Kb
1-
10Kb
10-
100Kb
100Kb-
1Mb
> 1Mb


A
snATAC-Seq
MGE
LGE
CGE
Cortex
scRNA-Seq
GAS = Sum promoter + 1st exon + enhancer reads
Integration
B
RNA
ATAC
C
D
pseudotime
Late
Early
RNA
ATAC GAS
E
Nes
F
Nes
Max
Min
G
Nkx2-1
H
Nkx2-1
I
Neurod2
J
Neurod2
K
Nkx2-1
L
Neurod2
Cicero/SnapATAC merged Gene-Peak Interactions
snATAC-Seq Peaks
H3K27ac ENCODE
VISTA
Genes
M
Ascl1
N
Emx1


A
156 genes
RNA GAS Enhancers
Z-score
Pseudotime
B
90 genes
RNA GAS Enhancers
Z-score
Pseudotime
Gene
C
Hes1
RNA
GAS
Enhancers
Z-score (reads)
Pseudotime
D
RNA
Max
Min
E
ATAC GAS
F
ATAC Enhancers
G
Lmo1
H
I
J
K
Myt1l
L
M
N
O
Lhx6
P
Q
R
S
RNA
Min
Max
Early
Transition
Late
Pseudotime
T
GAS
U
Enhancers


A
RNA
ATAC
B
E12.5 CGE
E12.5 MGE
P30 Cortex
P30 Hippo
C
E12.5
AP
BP
N
P30
CCK
Lamp5
Ndnf
VIP
Sst
Pvalb
D
Max
Min
E
DEGs (maturation stages) with iPSYCH2012 hits
−Log10 adjusted P
Log2 fold change
F
DEG
iPsych
73
87
G
# of genes
AP/BP
N
P30
H
OR
AP/BP
N
P30
I
DEGs (cell types) with iPSYCH2012 hits
−Log10 adjusted P
Log2 fold change
J
# of genes
AP
BP
N
CCK
Ndnf
Lamp5
VIP
Pvalb
Sst
K
OR
AP
BP
N
CCK
Ndnf
Lamp5
VIP
Pvalb
Sst
L
# of genes
AP
BP
N
CCK
Ndnf
Lamp5
VIP
Pvalb
Sst
M
Chr3
48.5 mb
49.5 mb
50.5 mb
49 mb
50 mb
rs34559126, rs12631989 & rs13064780 Hi-C interactions
AP
BP
N
CCK/NDNF
Lamp5
VIP
Pvalb
SST
overlapping genes
PFKFB4
QRICH
USP4
P4HTM
CCDC36
BSN
TRAIP
RBM6
rs34559126 specific genes
ARIH2
IMPDH2
NDUFAF3
DALRD3
DAG1
APEH
MON1A
RNF123
MST1R
IFRD2
LSMEM2
rs12631989/ rs13064780 specific genes
PRKAR2A
CCDC71
QARS
IP6K2
SLC25A20
KLHDC8B
LAMB2
USP19