Transcriptome analysis of early stage of neurogenesis reveals regulatory gene network for preplate neuron differentiation and CR cell specification

Jia Li[1,3†], Lei Sun[2†], Xue-Liang Peng[1], Shao-Jun Qi[1,3], Qin Shen[1*].

†Equal contributors

[1] Brain and Spinal Cord Innovative Research Center of Tongji Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200065, People's Republic of China

[2]School of Life Sciences, Tsinghua University, Beijing 100084, People's Republic of China

[3]School of Medicine, Tsinghua University, Beijing, 100084, People's Republic of China

*Correspondence: shenqin@tongji.edu.cn

Abstract

Background: Preplate is a transient layer critical for cortex development. Cajal-Retzius (CR) cells are the major cell type that later restrict to layer 1 and play important role in organizing cortical layering architecture. However, the molecular mechanisms underlying the generation of preplate and CR cell differentiation remain poorly understood.

Results: We used Ebf2-EGFP transgenic mouse that labeled preplate first then persisted to CR, to purify Ebf2 positive cells at different early embryonic stages. Bioinformatic analysis of high throughput RNA-seq revealed differentially expressed genes profile in Ebf2+ and Ebf2- cell population and temporal gene expression dynamics during early neurogenesis and CR neuron specification. Furthermore, our single cell RNA-seq unraveled E15.5 CR cells could be classified into three subpopulations based on their mature states. We also found that the upstream regulatory regions of Ebf2+ cell-enriched genes and lncRNAs were associated with active histone modifications by ChIP-seq studies, indicating epigenetic involvements in early differentiating neuron gene expression. Combined with functional prediction analysis, we identified a number of coding genes and lncRNAs that might regulate different aspects of CR neuron differentiation via primary cell culture, wholemount culture and *In Utero* Electroporation layer 1 assays.

Conclusions: Our work has identified genes and provided histone modification atlas that are related to early neurogenesis, particularly CR neuron identity, and will help understand the molecular mechanisms governing the developmental function of CR neurons.

Keywords: Early Neurogenesis, Preplate, Cajal-Retzius neuron development, RNA-seq, Single-cell RNA-seq, Histone modification, ChIP-seq, LncRNAs

Background

The mammalian cortical neurogenesis occurs on a precise time schedule during development. Neural stem cells (NSCs) in the ventricular zone (VZ) including neuroepithelial cells and radial glial cells first undergo symmetric cell divisions to expand, then begin to produce neurons [1]. The first-born neurons are generated as early as E10 in mice and form the preplate or primordial plexiform layer, which situates above the ventricular zone and beneath the pial surface [1-3]. The preplate is then split into the marginal zone (MZ, layer 1) and the subplate by immigrating cortical plate neurons around E13.5 in mouse and the seventh to eighth week of gestation in human [4-6]. The marginal zone is the most superficial cortical layer containing Cajal-Retzius (CR) cells, whereas the subplate neurons line the lower border of the cortical plate. The formation and splitting of preplate, setting up of the upper and lower borders of the cortical plate, are the first steps in the establishment of the highly organized cortical layers.

The preplate and its derivatives play a critical role in regulating subsequent inside-out neuronal migration, cortical lamination and the establishment of topographic connections [7]. Subplate neurons contribute to the guidance of corticofugal and thalamocortical axons during early development and are important for the functional maturation and plasticity of the cortical circuitry [8, 9]. CR neurons, the major cells type in layer 1, have been studied extensively since they were discovered more than a century ago by Retzius and Ramon y Cajal [3, 10]. Best known for their expression of Reln, CR neurons regulate the inside-out migration of cortical neurons, correct layer formation, maintenance of radial glia and cortical patterning [11]. Loss of Reln or mislocalization of CR cells contributes to defective migration and disordered cortical layer formation that are associated with human brain malformation such as lissencephaly [12] and

dystroglycanopathy [13]. Although functional importance of these early differentiating neurons is well recognized, characterization of cell-intrinsic, dynamic gene expression patterns in these cells remains incomplete. Recent studies have identified the genes enriched in subplate neurons by microarray and RNA-seq analysis [14-18], providing a comprehensive gene expression profile of subplate at different developmental stages, which offers valuable insight on molecular mechanisms underlying neurodevelopmental disorders [17]. However, how gene expression is dynamically regulated during the transition from the preplate to CR neurons remains unclear.

We previously have shown that in the Ebf2-EGFP transgenic mice (GENSAT project) [19] , the GFP signal is expressed specifically in early differentiating preplate neurons and persistent in CR cells [20]. After preplate splitting, virtually all of the EGFP-positive cells exhibit morphological and molecular features of CR neurons in the embryonic and postnatal brain [20], consistent with the pattern of the $Ebf2^{GFPiCre}$ [21], making this line suitable for tracing the development and profiling genetic and epigenetic pattern of early born neurons including CR neurons. In addition, single-cell gene expression profiling allows further dissection of the heterogeneity and developmental trajectory within a cell population [22, 23]. While multiple types of neurons in hippocampus and adult cortex have been identified by single-cell RNA-sequencing [23, 24], a full characterization of early differentiating neurons at single-cell level has not been done before. In this study, we used FACS-purified Ebf2-EGFP-expressing cells to probe the dynamic molecular features of preplate and CR neurons during the early stages of forebrain development. Whole genome transcriptome analysis identified the initial difference between the preplate and progenitor cells at E11.5, reflecting the molecular choice during the first cell fate decision in neural differentiation. We also recognized clusters of genes that displayed temporal

changes corresponding to different developmental trajectories. Furthermore, single-cell sequencing of E15.5 CR neurons identified its subpopulations that not reported before. Genome-wide overview of active (H3K27ac) and repressive (H3K27me3) histone modification in preplate neurons, CR neurons and NPCs revealed that their occupancies around TSS and distal region were associated with expression of early differentiating neurons-enriched genes and lncRNAs. Accompanied by bioinformatic functional enrichment analysis, we also identified and tested the function of CR-specific genes and lncRNAs that were enriched in CR neurons in neuronal differentiation. Hence our study provides unique molecular markers to extend the repertoire of preplate and CR neuron labels and reveals remarkable dynamism in the pattern of gene expression during early cortical development.

Results

Characterization of Ebf2-EGFP-expressing cells in the early cerebral cortex

We purified Ebf2-EGFP expressing cells by FACS from the embryonic mouse forebrain at E11.5, E13.5 and E15.5, to compare the developmental gene expression dynamics during preplate formation and differentiation of early neurons including CR neurons. At E11.5, Ebf2-EGFP+ cells are mostly preplate neurons including both subplate neurons and CR neurons, and at E13.5, Ebf2-EGFP expression is diminishing in subplate neurons but maintain in CR neurons, while by E15.5 Ebf2-EGFP is specifically expressed in CR neurons [20]. At E11.5, Ebf2-expressing cells consisted of 17.1% ± 1.07% of the total cells. At E13.5 and E15.5, the percentage of Ebf2-expressing cells gradually decreased to 8.37 ± 2.29% and 3.9 ± 1.59%, respectively (Figure 1A). 100% of FACS-sorted Ebf2-EGFP+ cells were β-III tubulin positive at all stages, indicating

these cells have committed neuronal fate decision, consistent with our previous in vivo analysis [20]. Accordingly, after culturing sorted and unsorted cells for 4 days, almost all of the Ebf2-EGFP-positive cells remained single and extended processes, while the Ebf2-EGFP-negative cells proliferated and generated large clones similar to the unsorted cell group, indicating a good separation of differentiating neurons from progenitor cells (Figure S1A).

Identification of preplate enriched genes in the developing mouse forebrain

Next, we collected FACS-selected Ebf2-EGFP positive cells and Ebf2-EGFP negative cells from the three early neurogenic stages and isolated total RNA for whole-genome transcriptome sequencing. Two biological replicates for each cell type (Ebf2-EGFP+ and Ebf2-EGFP-) were sequenced and expressed transcripts were identified as described in method. To verify the reproducibility of our samples, we analyzed the Pearson's correlation coefficient (Pearson's r) for the replicates and there was a strong association between different preparation of FACS-sorted cells from the same stage (Figure S1B).

To validate the RNA-seq results, we selected several protein-coding genes, including CR neurons markers (Ebf2 and Calb2) and genes associated with CR neurons fate regulation (Foxg1, Tbr1, Ebf1, and Ebf3) [10, 20, 25-28], and verified by qPCR that the temporal dynamic expression levels of these genes were in accordance with that detected by RNA-seq (Figure S1C). Furthermore, the temporal expression trends were consistent with the mRNA expression pattern revealed by in situ hybridization (ISH) from the data available in Allen Brain Atlas data (Figure S1C).

We used DE-Seq to analyze the dynamic expression profiles during the early cortical

differentiation period. 4859 genes exhibited significant difference in expression level between the two cell populations during the time course (FDR<0.05, p-values<0.05 by Fisher's exact test) and were selected as differential expressed genes (DEGs) for further analysis (Figure 1B; Additional file 1).

We analyzed DEGs expressed in any developmental stage as E11.5, E13.5 and E15.5. Genes with expression level (FPKM) more than 1, and 2-fold higher in Ebf2-EGFP positive population compared to that in Ebf2-EGFP negative population were defined as Ebf2-EGFP positive cells enriched genes, Ebf2+ genes in short. We found that there were 1044 genes at E11.5, 643 genes at E13.5, and 541 at E15.5 enriched in Ebf2-EGFP positive cells, respectively. Considerable overlap was found between the different stages (Figure 1C). In total, 1513 genes displayed enriched expression in Ebf2-EGFP positive population at least at one stage, and 1894 such genes in Ebf2-EGFP negative population. Of these, 166 genes were consistently enriched in Ebf2-EGFP positive population throughout all three embryonic stages, and 420 genes were differentially expressed in Ebf2-EGFP negative population (Figure 1C). Of the 166 DEGs included widely-used CR neuron markers such as Reln and Trp73, and genes known to be expressed in the CR neurons such as Calb2 (Calretinin), Tbr1, Ebf3, Cdkn1a [10, 25, 28, 29]. We further defined the DEGs with FPKM $\geq$ 5 in Ebf2-EGFP+ cells, and fold change $\geq$ 2 at all three stages as CR-specific genes. After excluding 11 subplate genes reported in previous studies [14, 16], 14 CR-specific genes reported in previous studies [25], 76 are novel genes with no reports association with CR neurons. Among the 76 CR-specific genes, 55 genes have valid in situ data from public resources. We also noticed that there were genes showing high expression levels in both Ebf2-EGFP+ cells and Ebf2-EGFP- cells although their expression levels meet the fold change criteria. To obtain the

genes specifically expressed in Ebf2-EGFP+ cells but not in Ebf2-EGFP- cells, we added a cutoff as CR-specific genes should be with FPKM ≤ 15 in Ebf2-EGFP- cells at these stages. We found 33 genes and confirmed that 8 with specific expression in layer 1 cells (Figure 1D; Figure S1D-S1E; Additional file 1). Hence, these genes could serve as CR-specific markers for future study.

Remarkably, the expression of these 8 genes were regulated by overexpression or knockdown of Ebf2, and Foxg1 knockdown in E11.5 cortical cell culture (Figure 1E; Figure S1F-S1G). These genes included transcription factors, synapse molecules, and axon guidance molecules, such as Nhlh2, Sncb, Slit1. For example, Nhlh2 was upregulated or inhibited by overexpression and knockdown of Ebf2, respectively. While Sncb and Ppp2r2c responded to Ebf2 overexpression but not knockdown, the expression of Slit1 and Brinp1 were decreased by Ebf2 knockdown, suggesting they might be also regulated by other molecules involved in developmental pathways rather than Ebf2 (Figure 1E; Figure S1F-S1G). Cortical CR neurons are considered to have multiple extra-cortex origins [30]. We and others have shown previously that Foxg1 suppressed CR neuron fate [27], therefore we hypothesized that CR-specific genes and lncRNAs identified by our RNA-seq analysis could be regulated by Foxg1 expression. Indeed, silencing of Foxg1 using shRNA up-regulated Ebf2 and CR-specific genes (Nhlh2) and lncRNA candidates (ln-CR5), as well as the CR marker gene Reln in the mRNA level (Figure 3F; Figure S2B).

Biological processes related to Ebf2-EGFP+ neuron-enriched (FC>2, FPKM>1) genes

We performed Gene Ontology analysis using the DAVID software to predict the potential function of the DEGs that were enriched in the Ebf2-EGFP+ or Ebf2-EGFP- cell population, respectively

(Figure 2A-2D; Additional file 2). The GO terms associated with neuronal differentiation and functions were highly enriched in DEGs expressed in the Ebf2-EGFP+ cell population, including “synapse”, “neuron projection”, “neuron differentiation”, “cell morphogenesis involved in neuron”, “axonogenesis” and “neuron migration” (Figure 2A). In contrast, the GO terms enriched in the Ebf2-EGFP- cell population were mostly correlated with progenitor cell properties, such as “forebrain development”, “regulation of cell proliferation”, “growth factor binding”, “regulation of neurogenesis” and “dorsal/ventral pattern formation” (Figure 2B).

In addition, we utilized the KEGG pathway repository to help determine the relative representation of specific pathways in our global RNA-seq expression data. Significantly enriched KEGG pathways of the global DEGs pointed out many signaling pathways implicated in normal neural development (Figure 2E; Additional file 2). For example, genes falling under the MAPK Signaling Pathway, Wnt Signaling Pathway and DNA replication were enriched, which played important roles in cell proliferation and differentiation. The Focal Adhesion and Axon Guidance pathway were also enriched, indicating regulation of cell morphology and neuronal differentiation were involved. Moreover, we performed the KEGG pathway analysis focused on DEGs enriched in Ebf2-EGFP+ cell population (Figure 2F; Additional file 2). As expected, the major groups of KEGG categories enriched in this data set were Calcium Signaling Pathway, Neuroactive Ligand-receptor interaction, Gap Junction, GnRH Signaling Pathway, signifying differentiation of neuronal properties. Interestingly, as Ebf2 was an important factor in GnRH neuron differentiation [31], the enrichments of GnRH Signaling pathway suggested that it would regulate similar pathway genes in cortical neurons.

Temporal transcriptome analysis of Ebf2-EGFP+ cells revealed early neuronal specification and CR differentiation dynamics

The high enrichment of these GO analysis terms and KEGG pathways imply that the transcriptional difference between the two cell populations is consistent with their different cell type identities. The components of Ebf2-EGFP+ and Ebf2-EGFP- cell population change as development proceeds. At E11.5, Ebf2-EGFP+ cells are mostly preplate neurons including both subplate neurons and CR neurons, while by E15.5 Ebf2-EGFP is specifically expressed in CR neurons [20]. In contrast, at E11.5 most if not all of the Ebf2-EGFP- cells are progenitor cells while at E13.5 and E15.5 the Ebf2-EGFP- population also includes cortical plate neurons. Therefore, the genes that are highly enriched in Ebf2-EGFP+ cells at all three stages, specifically those with increased enrichment, could be specifically involved in CR neuron differentiation, while the genes that are highly enriched in Ebf2-EGFP+ cells at E11.5 but not at E13.5 or E15.5 could be involved in regulating general neuronal differentiation processes. Hence, we further compared temporal changes in global DEGs at the three different stages. Genes with expression level (FPKM) in Ebf2-EGFP positive population 2 folds of that in Ebf2-EGFP negative population were defined as “up-regulated”. If the ratio of FPKM (Ebf2-EGFP+/ Ebf2-EGFP-) was between 2 and 1, these genes were defined as “unchanged”. If the ratio was between 0 and 1, these genes were defined as “down-regulated”. Theoretically, we expected 27 possible combination of expression change patterns, while finally there were 26 classes excepted for “E11up - E13down - E15up”. We found 26 classes could be classified into 6 groups represented different aspects of early neurogenesis, and we selected 6 typical classes for example as follows: “E11up - E13up - E15up”, “E11down - E13down - E15down”, “E11unchange - E13unchange - E15unchange”,

“E11up - E13unchange - E15down”, “E11down - E13unchange - E15up”, “E11down - E13unchange - E15down” (Figure 2G; Figure S2A; Additional file 3). We hypothesize that CR-specific genes are consistently higher at the three stages, while the difference in genes involved in global neuronal differentiation would be down after E13.5 as the Ebf2-EGFP negative population includes newly generated cortical plate neurons. Indeed, as we expect, the genes in the group “E11up - E13up - E15up” include the well-known CR marker genes, such as Reln, Lhx5, P73 and Ebf2, also CR-specific genes, and mature neuron genes related to axon regeneration like Cacna2d2 [32]. In contrast, genes in the “E11down - E13down - E15down” keep higher expression in Ebf2-EGFP- cells, showing that they may function on regulating neuronal proliferation, such as Sox2, Pax6 and Hes5. There are also some genes showing decreased expression levels from E11.5 to E15.5, that they may be related with promoting neuronal differentiation, such as Tle1 [33]. “E11unchange - E13unchange - E15unchange” group enriched genes associative with basic cellular organelles or activities that required for various eukaryotic cells. For example, Atp6v0c and Atp6v0e2 in this group are vacuolar ATPase and play essential roles in maintaining PH in the cellular compartments, protein degradation and endocytic pathway [34, 35]. Rab11fip5 is reported to participate in protein trafficking between different organelle membranes [36]. “E11up - E13unchange - E15down” group enrich subplate genes, such as Syt4, Igfbpl1, Myt1 [14, 16], reflect subplate genes are highly expressed in the preplate on E11.5, and with decreased expression level in layer1 as they split to subplate later on E13.5 and E15.5. Genes in “E11down - E13unchange - E15up” group imply that they may be involved in migrating from various CR origins to cortical surface. Take Zic1, Zic2 for example, they are expressed in the progenitor cells in the septum and cortical hem, and Bmp7 is expressed in the cortical hem,

meninges, and choroid plexus, these are typical CR neuron origins and location [37, 38]. We also found that “E11down - E13unchange - E15down” group enriched genes for cortical neuron development like Foxg1, Satb2, Bcl6, which are known to be involved in neurogenesis and neuronal differentiation. This data indicates, genes exhibiting different temporal dynamics are involved in distinct developmental processes, thus allowing us to identify novel factors that play roles in early cortical neurogenesis and CR development.

Differentially expressed lncRNAs in the Ebf2-EGFP+ cell and Ebf2-EGFP- cell

From our RNA-seq analysis, we identified 2417 long non-coding RNA transcripts using the Cufflinks method. 200 lncRNAs were differentially expressed between the Ebf2-EGFP+ and the Ebf2-EGFP- population (FDR<0.05) and termed DElncRNAs (Figure 3A). These DElncRNAs formed two main clusters according to their expression level at the three stages tested (Figure 3B). One cluster included lncRNAs highly expressed in Ebf2-EGFP+ cells compared to Ebf2-EGFP- cells, while the other cluster of lncRNAs showed the opposite situation. As previous reports suggested that expression of lncRNAs tended to be cell type-specific [39], we reasoned that DElncRNAs enriched in Ebf2-EGFP+ cells, especially increased expression levels as development, might be specifically associated with early neuron development. Hence, we selected DElncRNAs with these criterions: 1) specifically expressed in Ebf2-EGFP+ cells with FPKM at least 2-fold of that in Ebf2-EGFP negative cells; 2) highly expressed in Ebf2-EGFP+ cells, especially on E15.5 with FPKM $\geq$ 5, and expressed in Ebf2-EGFP- cells with FPKM $\leq$ 15 during all three stages, similar cutoff standards in CR-specific coding genes above. The in vivo in situ verification of CR-specific coding genes at layer1 (Figure 1D) made it as more confidential prediction of

CR-specific lncRNAs involvements in early neurons differentiation and to be specific, CR neuron development. We identified 9 lncRNAs that meet the criterions and had never been studied in the context of neural development (Additional file 4). The qPCR results to validate the mRNA expression level were generally in agreement with the RNA-seq data (Figure 3C; Figure S3A). Notably, we detected strong ln-CR2 signals enriched in E11.5 Ebf2-EGFP+ cells, and comparably lower or even no signals in Ebf2-EGFP- cells, indicating Ebf2-EGFP+ cell-specific expression. In situ hybridization on E11.5, E13.5 and E15.5 mouse embryonic brain sections revealed that ln-CR2 was selectively expressed on the cortical layer next to the pial surface, also expressed on pallium, TE (thalamus eminence), from which cortical CR neurons had been found to be originated, indicating cell-type specificity (Figure 3D). Further, we performed GO analysis (see method) of DElncRNAs to predict their potential functions (Figure 3E). The analysis results shown significant enrichments of neuronal development process related GO terms, such as “nervous system development”, “neuron migration”, “negative regulation of neuron apoptotic process” and “neuron differentiation”, neuron identities related GO terms such as “neuron projection”, “axon” and “dendrite”, also transcription factors regulation related GO terms as “transcription factor complex”, “transcription factor binding”. The functional annotation of DElncRNAs identified from our sequencing data revealed these lncRNAs’ involvements in the neurogenesis progression, neuron maturation and transcription factor regulation in these development events. While overexpressed Ebf2 did not significantly increase expression levels of CR-specific DElncRNAs except for ln-CR9, such might imply that most of them were not directly regulated by Ebf2 (Figure 3F).

Overview of subtypes and differentiation states of Cajal-Retzius neurons at the single-cell level

To gain comprehensive understanding of early differentiating neurons located on the cerebral cortex layer1 and to characterize their components in depth with high resolution, we performed single-cell RNA-seq (10X genomic Chromium) based on 3000 Ebf2-EGFP positive cells from E15.5 separated from Ebf2-EGFP negative cells by FACS. After filtering out potential noises as multiplet cells and mitochondria genes, then normalization and PCA analysis, the resulting acquired gene expression data sets were selected for the following analysis (Figure S4A-S4C). In this stage, Ebf2-EGFP positive cells were molecularly clustered into eight major subpopulations by K-Means clustering, each shown distinct gene expression pattern and identified by the most specific markers. The expression profiles of cluster 2 genes were quite different from other clusters, while cluster 3 genes shown similar high expression levels also in cluster 0 and 4 (Figure 4A-4C).

We undertook analysis of Ebf2+ genes highly expressed in scRNA-seq, since they were also specifically enriched in Cajal-Retzius neurons (E15.5 Ebf2-EGFP positive cells here). We performed pseudotime reconstruction by using Monocle for cells derived from Ebf2+ genes pool, and chose for trajectory analysis that expressed across more than 10 cells and their average expression levels were more than 1. Monocle is an algorithm that cells with similar Ebf2+ genes expression pattern can be put together to form sequential transcriptional changes response to various cell states along pseudotime points (Figure 4D; Figure S4D). We found that how the eight major cell subpopulations clustered from t-SNE plot corresponded to the temporal cortex neuron development progress from early to late stages (Figure 4E). Genes in cluster 5 could be the firstly

activated population during the pseudotemporal ordering, sequentially followed by cluster 3 and 7 genes, they displayed high expression levels at the beginning of time period, later dropped to low levels. These genes were most likely associated with a cascade of neocortical neuron differentiation events. For example, Cluster 5 gene Neurod1 expressed in progenitor cells which committed to neurons at the upper SVZ and in newborn neurons at the lower IZ, Neurod1 also had a prominent function of cell fate specification [40, 41]. Cluster 3 gene Rorb was a marker of layer 4 neurons and also labeled a subset of layer 5 neurons [42] (Figure 4E-4F). Thus, it was suggested that cluster 5, 3 and 7 genes might involve in immature neuron formation. In contrast, genes in cluster 2 were with low expression levels at the start side, but high expression levels at the end of pseudotime. The top differential expressed genes in this cluster like Reln, Nhlh2, Lhx5 were expressed on the CR neurons, and were also representing a more mature neuron state. Genes in cluster 7, 0, 6, 1, 4 displayed higher expression levels during the middle period of pseudotemporal ordering than the start or the terminal. Genes in these clusters were correlated with regulation or negative regulation of neuron differentiation (Tle4, Rnd3, Pou3f3) [43, 44] (Figure 4F; Figure S4E). Therefore, it implied that they might function on the transition stage helping immature neuron differentiate to more mature state. Above all, our scRNA-seq results unravel a previously unreported information that E15.5 CR neurons could be divided into three subtypes as immature, mature and much more mature neuron population.

Epigenetic regulation of early differentiating neuron-related genes and lncRNAs

Histone modification is a critical determinant of transcriptionally active or silent chromatin states [45]. To gain a comprehensive view of histone modification in the early differentiating neurons,

we performed H3K27ac and H3K27me3 chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) in Ebf2-EGFP+ cells during the early embryonic development (E11.5, E13.5, E15.5) and Ebf2-EGFP- cells at E11.5 (mostly neural stem cells and progenitor cells at this stage). We analyzed the occupancy of these histone modifications combined with RNA-seq data on these cell types in the followed analysis. First, our data shown strong concordance in H3K27ac and H3K27me3 bindings across two biological replicates (Figures S5A-S5B). According to a general strategy used to classify REs (regulatory elements) [46-48], we classified H3K27ac and H3K27me3 bound TSS (± 5kb) by the pattern of histone modification as TSS activation (H3K27ac), repression (H3K27me3) and bivalent (H3K27ac, H3K27me3). Interestingly, we observed being bivalent with both H3K27ac and H3K27me3 bindings at TSS sites in Ebf2-EGFP- cells. While the majority of genes in Ebf2-EGFP+ cells carried TSS activating modifications with strong H3K27ac enrichments and weak H3K27me3 enrichments (Figure 5A). Such were consistent with previous studies that bivalent histone modification patterns formed a poised state to be ready for initiation of further developmental events in the ES cells or progenitor cells, while monovalent states occurred in the differentiated cells meant more organized states in the established specific cell identity [48, 49]. As development proceeded from E11.5 to E15.5, H3K27ac enrichments around TSS in Ebf2-EGFP+ cells become stronger. In contrast, H3K27ac enrichments around TSS in E11.5 neural progenitor cells were lowest compared to others. However, H3K27me3 bindings did not show temporal changes in Ebf2-EGFP+ cells (Figure 5A; Figure S5C-S5D). These patterns suggested that the epigenetic modifications were different between early differentiating neurons and neural progenitor cells and dynamically changed as Ebf2-EGFP+ cells differentiate.

On Ebf2-EGFP+ cells, H3K27ac bindings were located in intergenic (61.3%), gene regions (62.2%, including 3'UTRs, 5'UTRs, Exons, Introns) and promoter regions (11.8%), H3K27me3 bindings were located in intergenic (65.6%), gene regions (43.3%, including 3'UTRs, 5'UTRs, Exons, Introns) and promoter regions (3.9%) (Figure 5B; Figure S5E-S5F). Intergenic regions contained the highest percentage of both H3K27ac and H3K27me3 peaks, likely implying potential distal regulatory elements. In addition, more peaks of H3K27ac were located in promoter (11.8%) and 5'UTRs (11.0%) than H3K27me3 (promoter 3.9% and 5'UTRs 2.8%) in Ebf2-EGFP+ cells (Figure S5E-S5F), consistent with H3K27ac's role in transcription regulation. On E11.5, genomic distribution patterns of H3K27ac and H3K27me3 peaks in Ebf2-EGFP- cells were similar to that in Ebf2-EGFP+ cells (Figure S5G-S5J).

On Ebf2-EGFP+ cells, almost all (97.4%, 1470 genes) of Ebf2+ genes had H3K27ac bindings. 64% (966 genes) of them had H3K27me3 bindings, 63.8% (964 genes) of them had both histone marks, 33.5% had only H3K27ac bindings while 0.1% had only H3K27me3 bindings (Figure 5C). Further, we investigated TSS and distal region distribution of these two kind of histone bindings around Ebf2+ genes. We found that for the H3K27ac bindings, 88% (1334) of Ebf2+ genes in Ebf2-EGFP+ cells had peaks around TSS, 42% (639) had peaks around distal region (within 100kb). And 33% (503) of Ebf2+ genes had peaks on both TSS and distal region, like Calb2, Ebf1, Zic5 and Sncb. Furthermore, 55% (831) of Ebf2+ genes had H3K27ac peaks only on TSS, such as most CR marker genes and CR-specific expressed genes like Reln, Ebf2, Ebf3, Trp73, Lhx1, Lhx5, Nhlh2, B3gat1, klf7, Ctsf. 9% (136) had enrichments only on distal region (Figure 5D), showing that H3K27ac bindings occurred both TSS and distal region of Ebf2+ genes. For H3K27me3 bindings, about 56.3% (851) of Ebf2+ genes had H3K27me3 peaks on

distal regions, 18.1% (274) on TSS, 10.5% (159) on both TSS and distal region like Ebf2, Ebf3, Trp73. 45.8% (692) of them had H3K27me3 bindings only on distal regions, as most CR marker genes and CR-specific expressed genes like Calb2, Tbr1, Ebf1, Lhx1, Lhx5, Zic5, Klf7 and Sncb, 7.6% (115) of Ebf2+ genes had bindings only on TSS like Zic2 (Figure 5E), showing that more H3K27me3 enrichments on distal region of Ebf2+ genes than on TSS. Notably, H3K27ac and H3K27me3 bindings on Ebf2+ genes distributed on TSS and distal region in E11.5 Ebf2-EGFP- cells were quite different from that in Ebf2-EGFP+ cells. In E11.5 Ebf2-EGFP- cells, only 6% (90) and 6.4% (96) Ebf2+ genes had H3K27ac peaks around TSS and distal region, respectively. However, there were 65.5% (989) and 25.8% (390) of Ebf2+ genes with H3K27me3 peaks around TSS and distal region, respectively (Figure 5D-5E). Low H3K27ac and high H3K27me3 bindings around Ebf2+ genes in NPCs implied that these genes were repressed, while they were activated in the Ebf2-EGFP+ cells. Therefore, H3K27ac enrichments on TSS and distal sites, along with H3K27me3 enrichments on distal region in Ebf2-EGFP+ cells might contribute to maintain or promote Ebf2+ genes highly expressed in CR neurons and distinguished CR neurons identities during early embryonic stages.

The epigenomic data shown that chromatin state round TSS was a strong indicator of RE function [48]. The data we obtained revealed that chromatin state at H3K27ac or H3K27me3-bound REs were associated with differential expression of the nearest CR-specific genes. To directly test for H3K27ac-mediated activation or H3K27me3-mediated repression via genomic binding to gene TSS regions, we examined the CR-specific genes and lncRNAs that were specifically enriched in Ebf2-EGFP+ cells identified in this paper. As expected, we observed unique epigenomic signatures mentioned above in the region around the genes and lncRNAs

reflected their potential roles in Ebf2-EGFP+ cells or NPC cells. For example, B3gat1, Ctsf and ln-CR5 had increased expression in Ebf2-EGFP+ cells from E11.5 to E15.5, accompanied by increasingly stronger H3K27ac enrichments at TSS in Ebf2-EGFP+ cells during these stages. Meanwhile, these three genes had low expression levels in Ebf2-EGFP- cells on E11.5, corresponding to H3K27me3 enrichments at TSS of these genes in Ebf2-EGFP- cells (Figure 5F). Moreover, ln-CR2, which showed reduced expression in Ebf2-EGFP+ cells on E13.5 and followed by increased expression level on E15.5, exhibited a histone mark with no H3K27ac peaks on the TSS region of ln-CR2 on E13.5 but high H3K27ac peaks on E15.5 (Figure 5F). The chromatin states at H3K27ac and H3K27me3-bound TSS associated with CR-specific genes and lncRNAs, suggesting that the expression of CR-specific genes were regulated by histone modification.

CR-specific LncRNAs In Vitro and In Vivo Functional Analysis

To further investigate what roles these lncRNA candidates may play during early neuronal differentiation, we performed in vitro and in vivo functional assays using lentivirus-mediated shRNA silencing and overexpression. We cultured single E11.5 cortical progenitor cells in adherent conditions for 5 days (DIV5), then fixed the culture and stained for neuronal markers β-III tubulin (immature neurons) and Reln (CR neurons) (Figure 6A). In the control culture there were 29.89% Tuj1+ neurons, of which 3.16% were Reln+ (Figure 6B). Reln+ neurons were increased to 11.56% by ln-CR1 overexpression, 6.09% and 6.83% by ln-CR2 and ln-CR3 overexpression, separately (Figure 6B). We also found that overexpression of ln-CR1, ln-CR2 and ln-CR3 increased β-III tubulin+ cells and reduced total cell number (Figure 6C-6D). Interestingly, the number of small clones with less than 4 cells and all Tuj1+ increased from CAG-control to

CAG-ln-CR1, CAG-ln-CR2 and CAG-ln-CR3, respectively, while the number of clones with more than 8 cells was much less than that in control (Figure 6C and 6E). These results suggested that overexpression of CR-specific lncRNAs promoted neuronal differentiation. Moreover, we also observed that overexpression of ln-CR1 altered the morphology of Reln+ neurons, increased whole processes length, branch depth and branch levels, while no significant differences were about neuron soma size (Figure 6F).

In addition, we performed wholemount culture and in utero electroporation cortex layer1 to investigate whether these CR-specific lncRNAs also had influences on CR neurons in vivo. Since CR neurons become more mature and have more complex morphology in the postnatal stage, we cultured wholemount cortical hemispheres isolated from E15.5 Ebf2-EGFP+ embryos and cultured to P5.5 (DIV10), we also electroporated cortex on E15.5 and collected embryos at P3.5. We found that overexpression ln-CR1 increased the morphology complexity of CR neurons, leading to more filopodia, spine-like structures (arrows) and bouton-like structures (arrowheads) (Figure 6H; Figure S6A-S6B).

To uncover how overexpression of these CR-specific lncRNAs molecularly affected CR neuron development, we assessed the mRNA level changes of several known CR neuron related genes via qPCR (Figure 6G). We found that CR marker genes including Reln, Ebf2, Calb2 and P73, and genes that were expressed in CR neurons such as Ebf1, Ebf3, Tbr1, Nhlh2 and Lhx1, were selectively altered when overexpressed CR-specific lncRNAs candidates ln-CR1, ln-CR2 and ln-CR3 (Figure 6G), implying these lncRNAs might be involved in the molecular pathways for CR neuronal differentiation.

Discussion

In this study, we characterized the dynamic transcriptomic profile of early neurogenesis and identified gene expression signatures associated with preplate neuronal differentiation and subsequent establishment of CR cell properties. Further, single cell RNA-seq deepen our understandings that E15.5 CR cells could be classified into three subpopulations based on their neuron mature states, which was unreported before. We also provided an overall understanding of active, repressive or bivalent histone modification states in the early differentiating neurons and neural progenitor cells, and indicated epigenetic regulation involved in these cell identities. Novel lncRNAs expressed in CR cells were also identified to be important for CR cell differentiation. Our findings provide a comprehensive molecular depiction of the early neuronal differentiation events, shedding new light on the mechanism underlying preplate formation and differentiation of CR cells, the first step of cortical neurogenesis.

Molecular transition from preplate to CR neurons

As cortical neurogenesis proceeds, the preplate is physically separated into the subplate and the layer 1 by migrating cortical plate neurons. We first identified over 1044 coding genes and 40 lncRNAs with higher expression in the newly differentiated preplate neurons (Ebf2-EGFP+) at E11.5 (Figure 1C; Figure 3B; Additional file 1 and Additional file 4). At E13.5, 370 genes and continue to show enrichment in preplate neurons, while additional 20 genes also have higher expression in subplate neurons. A subset of these genes continue to be differentially expressed in the Ebf2-EGFP+ cells at E15.5, while the total DEGs at E15.5 are much less than earlier stages, suggesting further specification of the CR cell fate. Indeed, we did not find any overlap between

the confirmed subplate neuron-specific genes at this age [17] (28 genes overlap, 2.3% of total DEGs at E15.5), confirming the molecular segregation of preplate neurons into the subplate and layer 1 by E15.5. The fact that the number of DEGs is the highest at E11.5 reflects the degree of the difference between the two populations: At E11.5, all Ebf2-EGFP+ cells are early differentiating neurons, while the Ebf2-EGFP- cells are mostly cortical progenitor cells. As the Ebf2-negative population includes both cortical plate neurons and neural progenitor cells at E13.5 and E15.5, genes common to neurons might not have significant changes.

This dynamic transcriptional profile may underlie the progression of preplate differentiation to the acquisition of unique molecular identity of CR neurons and subplate. For example, 166 genes are enriched in E11.5 and highly expressed in preplate, then later become restricted in CR neurons, while 851 genes which are also enriched in E11.5 but no longer specifically expressed in Layer 1 at E15.5.

Our transcriptome analysis also revealed a surprising enrichment of genes categorized in the GO terms such as “neuron projection”, “neuron differentiation”, “axonogenesis”, even “synapse” and “synaptic transmission” as early as E11.5. Terms such as “gate channel activity” and “ion channel activity” also appear before E15.5. Interestingly, whole-cell patch-clamp recording showed an inward sodium current in putative CR cells as early as E12 [50]. In addition, dynamic expression of K+ channels and Ca2+ channels were also described in pre- and postnatal CR cells [11, 50-52]. We found that Ebf2-EGFP-positive cells expressed neurotransmitter receptors including ionotropic and metabotropic glutamate receptors, AMPA glutamate receptors, at E11.5. Interestingly, it has been reported that at very early stages of neural development, subplate neurons form transient synapses with axons coming from the subcortical regions to guide later formation

of cortico-subcortical connectivity [53, 54]. Our results suggest that populations of early differentiating neurons, including Cajal-Retzius cells, are equipped with machineries suited for electric or chemical synaptic connections shortly after they are born. It's possible that functional connections are established earlier than currently recognized. Alternatively, expression of these synapse-related proteins may not necessarily confer electrophysiological properties, but rather prime early differentiating neurons to be ready for synaptic connection once they encounter with other cortical plate neurons.

Disease-related genes are highly expressed during early differentiation period [17, 55]. Abnormality in the process of preplate formation and splitting are associated with neurological disorders including mental retardation, epilepsy, and autism [7]. Interestingly, schizophrenia has a highly susceptible stage between 6-8 weeks gestation, the time when preplate differentiation occurs [56]. These findings suggest the importance of proper regulation during the early differentiation process.

Single-cell RNA-seq Analysis Reveal subpopulation of CR neurons

CR neurons were reported to be generated from many areas of developing pallium and were redistributed on the cortical surface via tangential migration. While did previous studies unravel all sites of CR origins? How many origins did CR population have, and where were they? These mysteries still long remained elusive. At first, we planned to undertake E15.5 Ebf2-EGFP+ cells single-cell RNA-seq to answer these questions. After sequencing, we selected 13 genes as Trp73, Calb2, Cxcr4, Cxcr7, Cd200, Reln, B3gat1, Snca, Pbx3, Ebf2, Tbr1, Lhx1, Lhx5 that reported to be specifically expressed in five well-known CR generation regions as hem, septum, PSB, MGE,

TE [57-62], and picked cell population that highly expressed (expression level ≥ 5) any of these 13 genes to perform K-means clustering. Finally, we analyzed whether different CR cell subclasses clustered could reflect their diverse origins.

The results shown eight cell subpopulations clustered (Figure 4A-4B) and we expected that there would be eight CR neuron sources. While we found that some CR-origin specific genes were found in various cell subpopulations. For example, previous study demonstrated Trp73-expressing CR cells were hem-derived [57], while our scRNA-seq data shown Trp73 widely expressed in all eight cell subpopulations. Also, previous study shown Lhx1 was one of Thalamic eminence (TE)-specific expressing gene, TE was thought to be a putative region to generate CR cells [63, 64], however, our scRNA-seq results shown Lhx1 expressed in two distinct cell clusters. On the other hand, each cell clusters contained genes from different CR origins. For example, Cluster 5 contained many CR-specific genes that were from various CR cell origin territories, such as Reln (hem), Trp73 (hem), Ebf2 (PSB), Snca (septum), Lhx1 (TE) [25, 57, 58, 60, 62]. Therefore, single-cell RNA-seq results of E15.5 CR cells provided limited associative information with CR origins mystery.

Here, we reported that CR neurons could be classified into 3 subgroups as immature, mature and much mature states, respectively. The immature subtype of CR consisted of activated cluster 5, 3 and 7 genes, they were upregulated at the beginning of pseudotemporal period, corresponding to genes function in these three clusters had association with characteristics of upper VZ located neural progenitor cells and lower cortical plate newborn neurons [40-42]. The following mature state subtype enriched cluster 7, 6, 0, 1, 4 genes that were activated during the middle of pseudotime trajectory, these genes were quite relevant to specify late born neuron to the correct

upper layer destination and neurite outgrowth that produced new neuron projection [43, 44, 65]. The more mature state consisted of cluster 2 genes that exhibited upregulation at later period. Many CR marker genes, synapse and axonogenesis related genes in this cluster made it more promising as a mature neuron class [66, 67]. We report the mature state differences of CR population here, the biological meaning of these differences whether contribute to sequential cortical laminar formation, or cortical neuron maturation and circuitry network establishment is waiting to be resolved. Overall, our study provides a single-cell RNA-seq dataset and molecular dynamics for each gene over continuous progression to identify new mechanism underlying CR neurons biology.

Methods

Mouse

All animal protocols used in this study were approved by the IACUC (Institutional Animal Care and Use Committee) of Tsinghua University and performed in accordance with guidelines of the IACUC. The laboratory animal facility has been accredited by AAALAC (Association for Assessment and Accreditation of Laboratory Animal Care International). For timed mice mating, the day of the vaginal plug was considered E0.5.

Mapping and Transcripts Assembling

Sequencing reads were mapped to mouse genome using bowtie with parameter "-v 2 -X 350 -fr -m 10 -S -k 1". After mapping sequencing reads to the mouse genome, we assembled the transcripts using Cufflinks [68]. We assembled transcripts into known gene models or novel gene models

guided by known gene annotations. Known gene annotations were downloaded from Ensembl (Mus_musculus.GRCm38.74). The assembled transcripts were then merged by the Cuffmerge utility provided by the Cufflinks package. Then the merged transcripts were concurrently annotated by Cuffcompare program in Cufflinks suite, of which known transcripts were used as reference to screen for novel lncRNAs. Only transcripts with tag “intergenic” and “intronic” were retained. Transcripts shorter than 200nt or with FPKM lower than 0.3 were filtered out from the transcripts set.

Clustering analysis and functional enrichment analysis

K-means clustering algorithm based on Euclidean distance was used to distinguish different gene expression patterns, which was visualized by Java TreeView software (http://bioinformatics.oxfordjournals.org/content/20/17/3246.abstract?keytype=ref&ijkey=31ed76d5EFKWk). Then, we used findGO.pl program in Homer to perform functional enrichment (GO annotation, KEGG and Wiki Pathway) analysis for each gene set with different expression patterns [69]. GO analysis was also performed by DAVID website (https://david.ncifcrf.gov/tools.jsp; version 6.8).

LncRNA Differential Expression

We used DE-seq to identify differential expressed lncRNAs. As input, the DESeq package expected count data as obtained from RNA-Seq in the form of a rectangular table of integer values. The table cell in the i-throw and the j-th column of the table told how many reads had been mapped to gene i in sample j. The count values must be raw counts of sequencing reads. As a first

processing step, we needed to estimate the effective library size. This step was sometimes also called normalization. The effective library size information was called the size factors vector, since the package only needed to know the relative library sizes. If the counts of non-differentially expressed genes in one sample were, on average, twice as high as in another (because the library was sequenced twice as deeply), the size factor for the first sample should be twice that of the other sample. The function estimateSizeFactors estimated the size factors from the count data. The inference in DESeq relied on an estimation of the typical relationship between the data's variance and their mean, or, equivalently, between the data's dispersion and their mean. The dispersion could be understood as the square of the coeffcient of biological variation. The variance seen between counts was the sum of two components: the sample-to-sample variation just mentioned, and the uncertainty in measuring a concentration by counting reads. The latter, known as shot noise or Poisson noise, was the dominating noise source for lowly expressed genes. The former dominated for highly expressed genes. The sum of both, shot noise and dispersion, was considered in the differential expression inference. Having estimated the dispersion for each gene, it was straight-forward to lookford differentially expressed genes. To contrast two conditions, e.g., to see whether there was differential expression between conditions “positive" and “negative", we simply called the function nbinom Test. It returned a data frame with the p-values and other useful information. Genes and lncRNAs with q-value lower than 0.1 were regarded as differential expressed.

Functional Analysis of lncRNAs

To identify the function of lncRNAs, we calculated the expression correlation of lncRNAs with

coding genes. We assumed that co-expressed genes were more likely to be correlated in function. Based on expression profiles in lncRNAs and protein coding genes, we calculated the correlation between them. FPKM of genes and lncRNAs from RNA-seq data sets were used to calculate the correlation matrix. Before calculating the correlation matrix, we filtered some low-expressed genes and lncRNAs for the accuracy of further characterization.

Known protein coding genes with FPKM ranked in top 75% in at least one data set were retained. For lncRNAs, the filtering processes were as follows: (1) expressed (FPKM>0) in at least 1 conditions, (2) FPKM ranked in top 75% in at least one datasets, (3) co-efficient of variation (cv) of FPKM ranked in top 75% in all ncRNA candidates. Cv represented for coefficient of variance, the standard deviation divided by mean. Cv showed the extent of variability in relation to mean of the population.

Correlations between gene-gene and gene-lncRNA were calculated. We chose to calculate Pearson correlation in the following analysis. The Pearson correlation coefficient was a measure of the linear correlation between two variables X and Y, giving a value between +1 and −1 inclusive, where 1 was total positive correlation, 0 was no correlation, and −1 was negative correlation. It was widely used in the sciences as a measure of the degree of linear dependence between two variables. Pearson's correlation coefficient between two variables was defined as the covariance of the two variables divided by the product of their standard deviations.

$$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 (Y - \overline{Y})^2}}$$

The calculation of correlation was done by SciPy, a fundamental python library for scientific computing. The two-sided p-value for a hypothetical test whose null hypothesis was that two sets of data were uncorrelated. The p-value roughly indicated the probability of an uncorrelated system

producing datasets that had a Pearson correlation at least as extreme as the one computed from these datasets.

Based on the assumption that functionally correlated genes were likely to co-express in various conditions, the functional prediction or annotation of lncRNAs could be performed. One of the most direct way was to look at the functional annotations of genes in its immediate neighborhood. For each lncRNAs, we chose top 1000 correlated genes to perform Gene Ontology analysis. We used a R package GOstats to perform GO analysis. The enriched GO term could be used to illustrate the possible function of lncRNAs.

Quantitative RT-PCR

Total RNA was isolated from purified Ebf2-EGFP positive or negative cells, extracted by using Trizol reagent (Cat#15596018, Life Technologies), and reverse transcribed, performed qPCR using PrimeScript RT regent Kit (Perfect Real Time, Cat#DRR047A, Takara). Gene expression level were detected on a BioRad CFX384 Touch™ real-time PCR detection system (Bio-Rad Laboratories) and determined with delta delta Ct method, normalized to an internal cotrol GAPDH mRNA for quantification. All reactions were performed 3 times replication. qRT-PCR primers for mouse gene sequence were see the Supplemental Experimental Procedures.

Tissue collection and RNA In Situ hybridization

Mice brains at various developmental stages were collected and fixed by 4h immersion in 4% paraformaldehyde (PFA) at 4°C. After briefly rinsed in PBS, brains were dehydrated in 30% sucrose in PBS at 4°C until the time they sinked, and embedded in OCT (Sakura Finetek, Torrance,

CA), cryosectioned to 16 μm tissue slides. In situ hybridization was performed according to instructions as previously described (Wigard et al., 2006; Gregor et al., 2007; Robert et al., 2007). Digoxigenin-labeled RNA probes were designed based on lncRNA sequences determined by Cufflinks and Ensembl. Images were taken on a Zeiss Slide Scanner (Axio Scan. Z1) fluorescent microscopy.

Embryonic Cortex Layer 1 Electroporation

Embryonic cortex layer 1 electroporation were carried out as described previously [70]. Briefly, induced anesthesia on embryonic day 15.5 (E15.5) mice and placed them on a heating pad. Then carefully pulled accessible embryos chain out and injected proper volume of plasmid mix (0.5-1ul/mice, 2-3 ug/ul) into the gap under the skull and above the pail surface with a glass capillary. Set the Electroporator (CUY21VIVO-SQ; BEX) to five 50ms, 36V pulses with an interval of 999ms, and electroporated embryos' head with a 5-mm electrode (LF650P5; BEX). After that, placed the embryos back into their original location of abdominal cavity, and closed the abdomen wall and skin. Then waited for a while until the mice recover from surgery. For phenotype observation, sacrificed the mice progenies on postnatal day 3.5 or 5.5 (P3.5 or P5.5) and selected both Ebf2-EGFP and RFP signals positive pups for the following PFA fixation dealing and immunohistochemistry. Confocal images were taken using LSM780 (Zeiss) Laser Scanning System on a Zeiss microscope. At least three independent experiments, three animals each experiment were analyzed for each condition.

Wholemount Culture Neuronal Processes Analysis

For neuronal processes and spine analysis, Ebf2-EGFP+ wholemounts on E15.5 were infected by lenti-virus CAG-RFP, H1-RFP, or CAG-ln-CR1, CAG-ln-CR2, CAG-ln-CR3. Next, wholemounts were cultured 12 days on 35℃ bioincubator. Then media was replaced every 3 days. After culture, wholemounts were fixed and stained with chicken anti-GFP (GFP-1010, aves lab lnc.) and rat anti-chicken-Alexa488 (Jackson Immuno Research lab). Confocal microscopy was performed with a LSM780 (Zeiss) Laser Scanning System on a Zeiss microscope, and wholemount tissues were analyzed by taking 1um serial optical sections with ZEN_2011_Lite_x64 software (Zeiss). Images were acquired using a 1024x1024 scan format with a 63x objective.

Immunocytochemistry

Cell culture were removed media and fixed in 4% (w/v) PFA for 30min at RT, rinsed by 1×PBS three times. Fixed cells were blocked in 5% (v/v) BSA with 0.3% (v/v) Triton X-100 in 1×PBS for 1h and rinsed by 1×PBS three times. Immunostaining were performed for 4h at RT or overnight at 4°C with rotation by incubating with primary antibodies, then washed by 1×PBS three times. Cells were incubated with secondary antibodies and DAPI diluted in 1×PBS for 1h in the dark, followed by washing 1×PBS three times. Images were taken on Zeiss fluorescent microscopy (AxioVision). Primary antibodies used: GFP (GFP-1010, aves lab lnc.), Reln (MAB5364, Millipore), Tuj1 (T8660, Sigma).

Chromatin-immunoprecipitation and sequencing (ChIP-Seq)

ChIP assays were performed in mice primary cells Ebf2-EGFP positive cells at E11.5, E13.5, E15.5 and Ebf2-EGFP negative cells at E11.5, with antibody against H3K27ac (Ab4729, Abcam)

and H3K27me3 (Ab6002, Abcam), experimental procedures were as previously described [71]. Briefly, Ebf2-EGFP positive and negative cells (~1-5 million cells) were separated and collected from FACS, and fixed by 0.4% (v/v) formaldehyde at room temperature (RT) for 10 min. Then stopped by adding 0.125 M glycine for 5 min. After rinsing cells three times with cold PBS, the nuclei were resuspended in Nuclei Lysis Buffer (50 mM Tri-Cl, pH8.1, 10 mM EDTA, 1% SDS, 1 mM DTT, 1 mM PMSF and Sigma protease inhibitor cocktail, RT) and incubated on ice for 10 min. Nuclear suspension were sonicated 6" on, 15" off, 20 cycles (power set at 25%) on ice. The fragmented chromatin was diluted for 5 folds in ChIP Dilution Buffer (16.7 mM Tri-Cl, pH8.1, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 1 mM PMSF, and Sigma protease inhibitor cocktail). Then added 5 ug IP antibody and were overnight incubated at 4℃. Chromatin-protein complex were pulled down with Dynabeads protein G (Cat#10004D, Invitrogen) for 2h at 4℃ with rotation and followed by 5 sequential washes. Once with low salt wash buffer (20 mM Tri-Cl, pH8.1, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate), once with high salt wash buffer (20 mM Tri-Cl, pH8.1, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate), once with LiCl wash buffer (10 mM Tri-Cl, pH8.1, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Deoxycholic acid (Na salt)), and twice with TE buffer (10 mM Tri-Cl, pH8.1, 1 mM EDTA). Immunoprecipitated complex were eluted from the antibody in elution buffer (1% SDS, 50 mM TrisCl (pH 8.1), 1 mM EDTA) once with votexing at 65℃ for 20min and repeat elution once at RT. Eluted chromatin were incubated 65℃ overnight for reverse crosslinks and chromatin fragmented purification with PCR column purification. ChIP-Seq libraries were prepared using NEB Next Ultra™ II ChIP-Seq Library Prep Kit (Cat#E7645S & Cat#7335S, NEB) according to manufacturer's instructions. ChIP-seq data were aligned to mouse

genome mm10, performing Bowtie v. 1.1.2 on the Galaxy platform with the following options: -v 3 –M 1 -5 15. Resulting output SAM files were converted to BAM format, and finally converted to BED files. Calling peaks were used MACS (Model-Based Analysis of ChIP-Seq), with a p-value cutting off of 1e-5.

Figure Legends

Figure 1. Transcriptome sequencing of Ebf2-EGFP+ and Ebf2-EGFP- cells in the early embryonic stage

(A) FACS purification of Ebf2-EGFP+ and Ebf2-EGFP- cells from early embryonic stages, respectively. 17.1%, 8.37% and 3.9% Ebf2-EGFP+ cells on E11.5, E13.5 and E15.5 were collected, respectively. Combined with Ebf2-EGFP- cells, after RNA isolation and reverse transcription, these samples were sent to RNA-seq. On E11.5, Ebf2-EGFP+ cells are early-bron neurons, Ebf2-EGFP- cells are neural progenitor cells. On E13.5 and E15.5, Ebf2-EGFP+ cells are mainly CR cells, and Ebf2-EGFP- cells are mixed pool of neural progenitor cells and immature neurons.

(B) Expression levels of all differential expressed genes (FDR<0.05) in Ebf2-EGFP+ & Ebf2-EGFP- samples during the three early embryonic developmental stages are shown.

(C) Venn diagrams displaying the extent of overlapping between DEGs at the three early embryonic periods.

(D) Mouse brain expression of selected CR-specific coding genes are found to be specifically expressed in layer 1 on E14.5 (In Situ from Eurexpress Anatomy Atlas and GenePaint.org).

(E) RT-PCR analysis of CR-specific genes in In Vitro NSCs culture assay after virus infection with control (H1) or Ebf2-overexpression. Data represent mean ± SEM (n=3 independent experiments, *P<0.05, T test).

Figure 2. Functional prediction of differentially expressed DEGs in the Ebf2-EGFP+ cell and Ebf2-EGFP- cells

(A-D) Enriched functional assessments (including biological processes, celluar components and molecular function) of highly expressed in Ebf2-EGFP+ or Ebf2-EGFP- cell population DEGs (fold change ≥ 2). And to be specific, enriched functional assessments of highly expressed on E11.5, E13.5 or E15.5 Ebf2-EGFP+ cell population DEGs, respectively.

(E-F) Significantly (p<0.001, FDR<0.001) enriched KEGG pathways within all RNA-seq sample DEGs and Ebf2-EGFP+ sample DEGs.

(G) Stage-specific expression level analysis of all RNA-seq DEGs reveal their temporal dynamic expression pattern (y axis = FPKM). Noted in the presence of several markers of neurogenic commitment.

Figure 3. Functional prediction of differentially expressed lncRNAs in the Ebf2-EGFP+ cell and Ebf2-EGFP- cells

(A) Expression levels of all differential expressed lncRNAs (FDR<0.05, p-value<0.05) in Ebf2-EGFP+ & Ebf2-EGFP- samples during the three early embryonic developmental stages are shown.

(B) Venn diagram displaying the extent of overlapping between DElncRNAs at the three early embryonic periods.

(C) RT-PCR to validate the consistency of RNA-seq results via comparing expression trends of DElncRNAs identified by RNA-seq and enriched in Ebf2-EGFP+ or Ebf2-EGFP- cells at different embryonic stages.

(D) In Situ showing CR-specific lncRNA ln-CR2 endougenous expression levels in early embryonic stages as E11.5, E13.5 and E15.5. Higher expression regions are labled, they are E11.5 on medial pallium（MP），ventral pallium（VP）；E13.5 on medial pallium（MP），thalamus（Th），lateral pallium（LP），ventral pallium（VP），posteromedial cortical amygdalar area (part of pallial amygdala)（PMCo）； E15.5 on dorsal pallium（DP），medial pallium（MP），striatum（Str），pallial septum（PSe）(TE, thalamus eminence).

(E) Enriched functional assessments (including biological processes, celluar components and molecular function) of highly expressed in Ebf2-EGFP+ or Ebf2-EGFP- cell population DElncRNAs (fold change ≥ 2).

(F) RT-PCR analysis of CR-specific lncRNAs in In Vitro NSCs culture assay after virus infection with control (H1) or Ebf2-overexpression. Data represent mean ± SEM (n=4 independent experiments, *P<0.05, T test).

Figure 4. Single cell overview of early differentiating neuron subtypes

(A) Heatmap of E15.5 Ebf2-EGFP+ cell sub-clustering deferentially expressed genes.

(B) The K-means algorithm-based plots showing eight major distinct populations.

(C) Visualization of the eight major sub-classes of cells using t-SNE clustering. Each dot represents a single cell, dot color represents expression level, color from blue to red, expression levels from low to high.

(D) Pseudotime trajectories of E15.5 Ebf2-EGFP+ cells reconstructed by Monocle. Cell states are indicated by colors and numbers in the black circle.

(E) Heatmap shows gene dynamics along pseudotemporal cortical neuron differentiation. x-axis shows pseudotime ordering from early to late, y-axis shows a single gene expression level dynamic per row. Gene expression level from high to low is indicated by colors from red to blue.

(F) Expression level dynamics across pseudotime ordering of Lhx5 (in cluster2) and Neurod1 (in cluster5) are shown.

Figure 5. Genomic characterization of histone modification in the early differentiating neurons during embryonic stage

(A) Heatmap representation of ChIP-seq signal density for H3K27ac and H3K27me3 for 10 kb centered on predicted Gene TSSs in Ebf2-EGFP+ (E11.5, E13.5 and E15.5) and Ebf2-EGFP- (E11.5) cells.

(B) Genomic features of H3k27ac and H3k27me3-bound REs in Ebf2-EGFP+ cells.

(C) Comparison of Ebf2+ genes number with H3k27ac and H3k27me3 bindings in Ebf2-EGFP+ cells.

(D) Proportion of Ebf2+ genes with TSS and distal H3k27ac bindings in Ebf2-EGFP+ cells (upper) and NPCs (lower).

(E) Proportion of Ebf2+ genes with TSS and distal H3k27me3 bindings in Ebf2-EGFP+ cells (upper) and NPCs (lower).

(F) The CR-specific genes B3gat1, Ctsf, and CR-specific lncRNAs ln-CR2, ln-CR5 TSS region demonstrate a similar pattern of histone modification during early embryonic development (bivalent in NPCs and become more and more activated in Ebf2-EGFP+ cells from E11.5 to E15.5).

Figure 6. Functional validation of CR-specific LncRNAs ln-CR1, ln-CR2 and ln-CR3 *In Vitro* and *In Vivo*

(A) Immunostaining of NSCs treated with control (CAG-control) and overexpressed CR-specific lncRNAs ln-CR1 (CAG-ln-CR1), ln-CR2 (CAG-ln-CR2) and ln-CR3 (CAG-ln-CR3) after In Vitro culture 5 days. Cells were stained for Reln (red), Tuj1 (cyan), or DAPI (blue) for CR neurons, neurons, and nucleus, respectively, and analyzed by fluorescent microscopy. Overexpression of ln-CR1, ln-CR2 and ln-CR3 showed more CR neurons and more neurite-like processes, 5 days after virus transfection. Scale bar, left 3 rows 150um, right 50um.

(B-E) Overexpression of CR-specific lncRNAs significantly increased CR neuron number, promoted NSC differentiation and generated more neuron-like small cell clones. All the experiments were performed in 4 times and 8 randomly selected microscopy fields were counted

each time. Data represented mean ± SEM. p values indicated were calculated by Student's t test (unpaired). **p<0.01, *p<0.05.

(F) Overexpression of CR-specific lncRNAs also increased CR neuron processes branch depth, branch levels and total dendrites length, no significant increation in soma size. All the experiments were performed in 3 times and 8 randomly selected microscopy fields were counted each time. Data represented mean ± SEM. p values indicated were calculated by Student's t test (unpaired). **p<0.01, *p<0.05.

(G) Overexpression of CR-specific lncRNAs resulted in upregulation of CR neuron molecular markers as Reln, Ebf2, and Calb2 72h after virus infection in the NSC In Vitro culture assay. Data represented mean ± SEM (**p<0.01, *p<0.05, n>3, Student's T test).

(H) Comparison of dendritic spine, filopodia and processes cortical wholemounts overexpressing ln-CR1 (CAG-ln-CR1) to wholemount expressing CAG-control. Scale bar, 20 um. ln-CR1 overexpression resulted in increased in spine and filopodia (arrows) as compared to CAG-control transfected CR neurons. Cortical wholemounts were collected on P3 from *in utero* electroporation to cortex layer 1 on E15.5.

Additional files

Figure S1. RNA-seq quality control, related to Figure 1

(A) Ebf2-EGFP+ cells and Ebf2-EGFP- cells purity after FACS, their cells state and growth abilities after 4 days in vitro culture, compared with unsorting cells and viewed with microscope (scale bar, 25um).

(B) Sample to sample Pearson's correlation (top-right quadrants) and pairwise comparison (bottom-left quadrants) of normalized gene expression between two biological replicates and cell types.

(C) RT-PCR and endogenous expression (In Situ from Allen Brain Atlas) to validate the consistency of RNA-seq results via comparing expression trends of protein-coding genes at different embryonic stages. CR neuron molecular markers as Calb2, Ebf2, Tbr1, and CR neuron related genes as Ebf1, Ebf3 and Foxg1 were selected for validation.

(D) RT-PCR and endogenous expression (In Situ from Allen Brain Atlas) to validate the consistency of RNA-seq results via comparing expression trends of Nhlh2 at three embryonic stages.

(E) Mouse brain expression of selected CR-specific coding genes Dnm3 is found to be specifically expressed in layer 1 at E14.5 (In Situ from Eurexpress Anatomy Atlas).

(F-G) RT-PCR analysis of CR-specific genes in In Vitro NSCs culture assay after virus infection with control (H1) or Ebf2-knockdown, or Foxg1-knockdown. Data represented mean ± SEM (n=3 independent experiments, *P<0.05, T test).

Figure S2. Annotation of temporal dynamic genes, related to Figure 2

(A) Several markers of neurogenic commitment corresponding to temporal dynamic expression pattern of all DEGs are selected with gene annotation.

Figure S3. Functional prediction of CR-specific DElncRNAs, related to Figure 3

(A) RT-PCR analysis to validate the consistency of RNA-seq results via comparing expression trends of lncRNAs at different embryonic stages. CR-specific lncRNAs ln-CR6 and Meg3 were selected for test.

(B) RT-PCR analysis of CR-specific lncRNAs in In Vitro NSCs culture assay after virus infection with control (H1) or Foxg1-knockdown. Data represent mean ± SEM (n=4 independent experiments, *P<0.05, T test).

Figure S4. Quality control and pseudotime reconstruction analysis of single cell RNA-seq, related to Figure 4.

(A) Violin plots shown genes, UMIs and mitochondrion ratio during scRNA-seq. Left, violin plot shown detected gene number in all sample cells, x-axis is sample name, y-axis is gene numbers detected in a single cell; Middle, x-axis is the same with left, y-axis is UMIs detected in a single cell; Right, y-axis is mitochondrion numbers detected in a single cell.

(B) Scatter diagram of mitochondrion ratio, gene numbers compared to UMI numbers, respectively. x-axis is UMI numbers, y-axis is mitochondrion ratio detected in a single cell in the left, y-axis is gene numbers detected in a single cell in the right.

(C) Scatter diagram of all detected genes dispersion. x-axis is average expression of a gene in all cells, y-axis is dispersion value of that gene.

(D) Pseudotime trajectories of E15.5 Ebf2-EGFP+ cells reconstructed by Monocle. Pseudotime

points are indicated by colors and cell clusters are indicated by numbers in the black circle.

(E) Expression level dynamics across pseudotime ordering of Reln (in cluster2) and Tle4 (in cluster4) are shown.

Figure S5. Chromatin signatures in Ebf2-EGFP+ cells and NPCs, related to Figure 5

(A-B) Average profile of H3k27ac(A) or H3k27me3(B) binding at called ChIP-seq peaks compared the two biological replicates.

(C-D) A profile plot shown normalized H3k27ac(C) or H3k27me3(D) enrichment around TSS of all genes during the early embryonic stages.

(E-J) Quantitative data shown genomic features of H3k27ac and H3k27me3-bound REs in Ebf2-EGFP+ cells (E, F), E11.5 NPCs (G, H) and E11.5 Ebf2-EGFP+ cells (I, J).

Figure S6. ex vivo wholemount culture assay show LncRNAs function in CR neurons, related to Figure 6

(A) Comparison of dendritic spine, filopodia and processes cortical cultured wholemounts overexpressing ln-CR1 (CAG-ln-CR1), ln-CR2 (CAG-ln-CR2), ln-CR3 (CAG-ln-CR3) and knockdown ln-CR2 (shRNA-ln-CR2) to wholemounts expressing CAG-control. Scale bar, 15um. ln-CR1 overexpression result in increased in spine and filopodia as compared to CAG-control transfected CR neurons, with no differences are observed when overexpressing ln-CR2 and ln-CR3 in CR neurons. ln-CR2 knockdown result in significant increase in processes density

around soma when compared to H1 transfected CR neurons, in cultured cortical wholemounts from E15.5-P7.5.

(B) Quantification of CR-specific lncRNAs overexpression alter CR neuron processes branch depth, branch levels and total dendrites length. All the experiments were performed in 3 times and 4 randomly selected microscopy fields were counted each time.Data represents mean ± SEM. p values indicated were calculated by Student's t test (unpaired). **p<0.01, *p<0.05.

References

1. C. VS. Neocortical Histogenesis in Normal and Reeler Mice: A Developmental Study Based Upon [3H]Thymidine Autoradiography. Developmental Brain Research. 1982;4:293-302.
2. M.-P. M. Ontogenesis of the Pyramidal Cell of the Mammalian Neocortex and Developmental Cytoarchitectonics: A Unifying Theory. J Comp Neurol. 1992;321:321-3.
3. Hevner RF, Neogi T, Englund C, Daza RM,F. A. Cajal–Retzius cells in the mouse: transcription factors, neurotransmitters, and birthdays suggest a pallial origin. Developmental Brain Research. 2003;141:39-53.
4. Bayer SA,A. J. Development of the Endopiriform Nucleus and the Claustrum in the Rat Brain. Neuroscience. 1991;45:391-412.
5. Takahashi T, Goto T, Miyama S, Nowakowski RS,C. V. Jr. Sequence of Neuron Origin and Neocortical Laminar Fate: Relation to Cell Cycle of Origin in the Developing Murine Cerebral Wall. J Neurosci. 1999;19:10357-71.
6. Bystron I, Blakemore C,R. P. Development of the human cerebral cortex: Boulder Committee revisited. Nat Rev Neurosci. 2008;9:110-22.
7. O. EC. Analysis of preplate splitting and early cortical development illuminates the biology of neurological disease. Front Pediatr. 2014;2:1-9.
8. Ghosh A, Antonini A, McConnell SK,S. CJ. Requirement of Subplate Neurons in the Formation of Thalamocortical Connections. Nature. 1990;347:179-181.
9. Molnar Z, Adams R,B. C. Mechanisms Underlying the Early Establishment of Thalamocortical Connections in the Rat. J Neurosci. 1998;18:5723-45.
10. Del Rio JA, Martinez A, Fonseca M, Auladell C,S. E. Glutamate-like Immunoreactivity and Fate of Cajal-Retzius Cells in the Murine Cortex as Identified with Calretinin Antibody. Cereb Cortex. 1995;5:13-21.
11. Kirischuk S, Luhmann HJ,K. W. Cajal-Retzius cells: update on structural and functional properties of these mystic neurons that bridged the 20th century. Neuroscience. 2014;275:33-46.
12. Hong SE, Shugart YY, Huang DT, Shahwan SA, Grant PE, Hourihane JOB, Martin NDT,W. CA. Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human

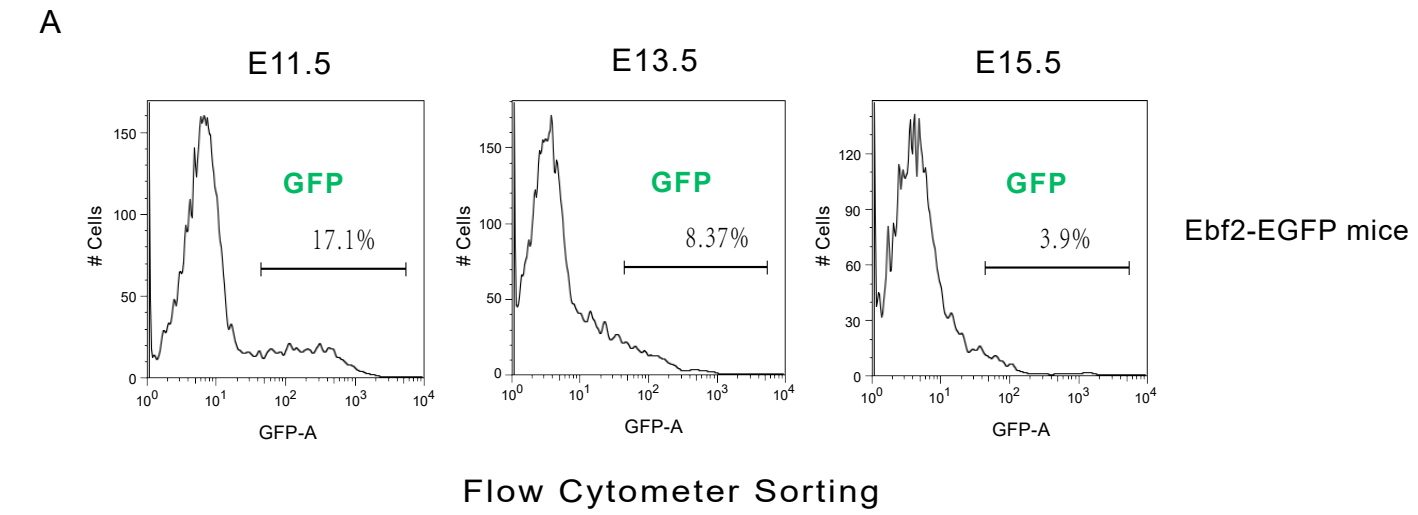
RELN mutations. Nature Genetics. 2000;26:93-6.
13. Nakagawa N, Yagi H, Kato K, Takematsu H,O. S. Ectopic clustering of Cajal-Retzius and subplate cells is an initial pathological feature in Pomgnt2-knockout mice, a model of dystroglycanopathy. Sci Rep. 2015;5:11163.
14. Oeschger FM, Wang WZ, Lee S, Garcia-Moreno F, Goffinet AM, Arbones ML, Rakic S,M. Z. Gene expression analysis of the embryonic subplate. Cereb Cortex. 2012;22:1343-59.
15. Osheroff H,H. ME. Gene expression profiling of preplate neurons destined for the subplate: genes involved in transcription, axon extension, neurotransmitter regulation, steroid hormone signaling, and neuronal survival. Cereb Cortex. 2009;19 Suppl 1:i126-34.
16. Hoerder-Suabedissen A, Wang WZ, Lee S, Davies KE, Goffinet AM, Rakic S, Parnavelas J, Reim K, Nicolic M, Paulsen O,M. Z. Novel markers reveal subpopulations of subplate neurons in the murine cerebral cortex. Cereb Cortex. 2009;19:1738-50.
17. Hoerder-Suabedissen A, Oeschger FM, K. ML, Belgard TG, Wang WZ, Lee S, Webber C, Petretto E, Edwards AD,M. Z. Expression profiling of mouse subplate reveals a dynamic gene network and disease association with autism and schizophrenia. Proc Natl Acad Sci U S A. 2013;110:3555-60.
18. Belgard TG, Marques AC, Oliver PL, Abaan HO, Sirey TM, Hoerder-Suabedissen A, Garcia-Moreno F, Molnar Z, Margulies EH,P. CP. A transcriptomic atlas of mouse neocortical layers. Neuron. 2011;71:605-16.
19. Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME,H. N. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature. 2003;425:917-925.
20. Chuang SM, Wang Y, Wang Q, Liu KM,S. Q. Ebf2 marks early cortical neurogenesis and regulates the generation of cajal-retzius neurons in the developing cerebral cortex. Dev Neurosci. 2011;33:479-93.
21. Chiara F, Badaloni A, Croci L, Yeh ML, Cariboni A, Hoerder-Suabedissen A, Consalez GG, Eickholt B, Shimogori T, Parnavelas JG,R. S. Early B-cell factors 2 and 3 (EBF2/3) regulate early migration of Cajal-Retzius cells from the cortical hem. Dev Biol. 2012;365:277-89.
22. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;161:1202-1214.
23. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He LQ, Hjerling-Leffler J,L. S. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science. 2015;347:1138-42.
24. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, Enikolopov G, Nauen DW, Christian KM, Ming GL,S. H. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. Cell Stem Cell. 2015;17:360-72.
25. Yamazaki H, Sekiguchi M, Takamatsu M, Tanabe Y,N. S. Distinct ontogenic and regional expressions of newly identified Cajal–Retzius cell-specific genes during neocorticogenesis. PNAS. 2004;101:14509-14.
26. Hanashima C, Li SC, Shen L, Lai E,F. G. Foxg1 Suppresses Early Cortical Cell Fate. Science. 2004;303:56-59.
27. Shen Q, Wang Y, Dimos JT, Fasano CA, Phoenix TN, Lemischka IR, Ivanova NB, Stifani S, Morrisey EE,T. S. The timing of cortical neurogenesis is encoded within lineages of individual

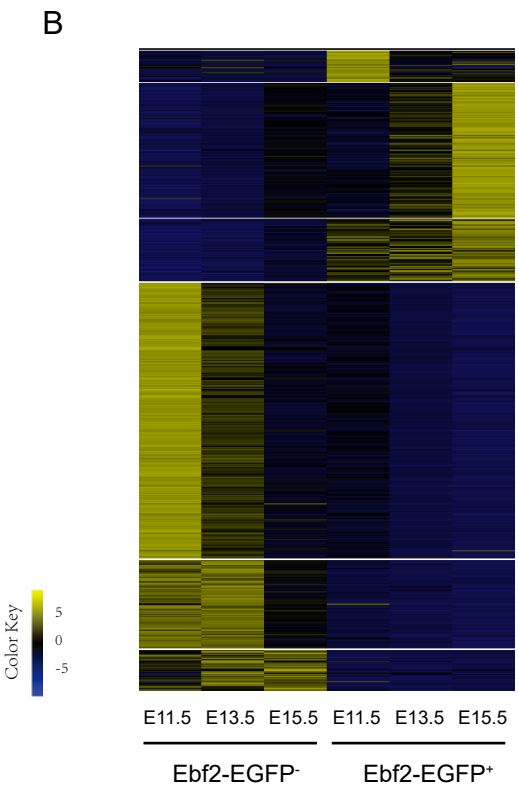
progenitor cells. Nat Neurosci. 2006;9:743-51.
28. Hevner RF, Shi L, Justice N, Hsueh YP, Sheng M, Smiga S, Bulfone A, Goffinet AM, Campagnoni AT,R. JL. Tbr1 Regulates Differentiation of the Preplate and Layer 6. Neuron. 2001;29:353-366.
29. S. JA.,M. MW. Generation of Cajal-Retzius neurons in mouse forebrain is regulated by transforming growth factor beta-Fox signaling pathways. Dev Biol. 2008;313:35-46.
30. M. G. Building a human cortex: the evolutionary differentiation of Cajal-Retzius cells and the cortical hem. J Anat. 2010;217:334-43.
31. Corradi A, Croci L, Broccoli V, Zecchini S, Previtali S, Wurst W, Amadio S, Maggi R, Quattrini A,C. GG. Hypogonadotropic hypogonadism and peripheral neuropathy in Ebf2-null mice. Development. 2003;130:401-410.
32. Tedeschi A, Dupraz S, Laskowski CJ, Xue J, Ulas T, Beyer M, Schultze JL,B. F. The Calcium Channel Subunit Alpha2delta2 Suppresses Axon Regeneration in the Adult CNS. Neuron. 2016;92:419-434.
33. Buscarlet M, Perin A, Laing A, Brickman JM,Stifani S. Inhibition of Cortical Neuron Differentiation by Groucho/TLE1 Requires Interaction with WRPW, but Not Eh1, Repressor Peptides. Journal of Biological Chemistry. 2008;283:24881-8.
34. Nishi T,F. M. The vacuolar (H+)-ATPases--nature's most versatile proton pumps. Nat Rev Mol Cell Biol. 2002;3:94-103.
35. Ueda T, Ugawa S,S. S. A novel putative M9.2 isoform of V-ATPase expressed in the nervous system. Neuroreport. 2003;14:25-30.
36. Boal F, Hodgson LR, Reed SE, Yarwood SE, Just VJ, Stephens DJ, McCaffrey MW,T. JM. Insulin promotes Rip11 accumulation at the plasma membrane by inhibiting a dynamin- and PI3-kinase-dependent, but Akt-independent, internalisation event. Cell Signal. 2016;28:74-82.
37. Inoue T, Ogawa M, Mikoshiba K,A. J. Zic deficiency in the cortical marginal zone and meninges results in cortical lamination defects resembling those in type II lissencephaly. J Neurosci. 2008;28:4712-25.
38. Segklia A, Seuntjens E, Elkouris M, Tsalavos S, Stappers E, Mitsiadis TA, Huylebroeck D, Remboutsika E,G. D. Bmp7 regulates the survival, proliferation, and neurogenic properties of neural progenitor cells during corticogenesis in the mouse. PLoS One. 2012;7:e34088.
39. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 2012;22:1775-89.
40. Hevner RF, Hodge RD, Daza RA,E. C. Transcription factors in glutamatergic neurogenesis: conserved programs in neocortex, cerebellum, and adult hippocampus. Neurosci Res. 2006;55:223-33.
41. Lee JE, Hollenberg SM, Snider L, Turner DL, Lipnick N,W. H. Conversion of Xenopus Ectoderm into Neurons by NeuroD, a Basic Helix-Loop-Helix Protein. Science. 1995;268:836-44.
42. Schaeren-Wiemers N, Andre E, Kapfhammer JP,B.-A. M. The Expression Pattern of the Orphan Nuclear Receptor RORb in the Developing and Adult Rat Nervous System Suggests a Role in the Processing of Sensory Information and in Circadian Rhythm. Eur J Neurosci. 1997;9:2687-701.
43. Pacary E, Heng J, Azzarelli R, Riou P, Castro D, Lebel-Potter M, Parras C, Bell DM, Ridley AJ, Parsons M,G. F. Proneural transcription factors regulate different steps of cortical neuron

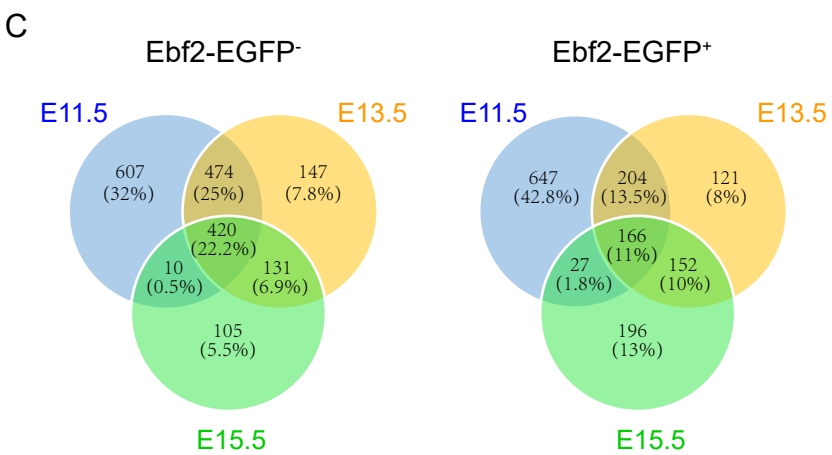
migration through Rnd-mediated inhibition of RhoA signaling. Neuron. 2011;69:1069-84.
44. Oishi K, Aramaki M,N. K. Mutually repressive interaction between Brn1/2 and Rorb contributes to the establishment of neocortical layer 2/3 and layer 4. Proc Natl Acad Sci U S A. 2016;113:3371-6.
45. Jenuwein T,A. CD. Translating the Histone Code. Science. 2001;293:1074-1080.
46. Attanasio C, Nord AS, Zhu Y, Blow MJ, Biddie SC, Mendenhall EM, Dixon J, Wright C, Hosseini R, Akiyama JA, et al. Tissue-specific SMARCA4 binding at active and repressed regulatory elements during embryogenesis. Genome Res. 2014;24:920-9.
47. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473:43-9.
48. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125:315-26.
49. Albert M, Kalebic N, Florio M, Lakshmanaperumal N, Haffner C, Brandl H, Henry I,H. WB. Epigenome profiling and editing of neocortical progenitor cells during development. EMBO J. 2017;36:2642-58.
50. Albrieux M, Platel JC, Dupuis A, Villaz M,M. WJ. Early expression of sodium channel transcripts and sodium current by cajal-retzius cells in the preplate of the embryonic mouse neocortex. J Neurosci. 2004;24:1719-25.
51. Mienville JM,B. JL. Potassium Current Expression during Prenatal Corticogenesis in the Rat. Neuroscience. 1997;81:163-172.
52. Kirmse K, Grantyn R,K. S. Developmental downregulation of low-voltage-activated Ca2+ channels in Cajal-Retzius cells of the mouse visual cortex. Eur J Neurosci. 2005;21:3269-76.
53. M. AR., Filipovic R, Mo Z, Rasband MN, Zecevic N,A. SD. Electrical Excitability of Early Neurons in the Human Cerebral Cortex during the Second Trimester of Gestation. Cerebral Cortex. 2009;19:1795-1805.
54. Selemon LD,Z. N. Schizophrenia: a tale of two critical periods for prefrontal cortical development. Transl Psychiatry. 2015;5:e623.
55. Heather LC, Wang X, West JA,G. JL. A practical guide to metabolomic profiling as a discovery tool for human heart disease. J Mol Cell Cardiol. 2013;55:2-11.
56. K. I. Reelin-mediated signaling in neuropsychiatric and neurodegenerative diseases. Progress in Neurobiology. 2010;91:257-274.
57. Meyer G, Perez-Garcia CG, Abraham H,C. D. Expression of p73 and Reelin in the Developing Human Cortex. J Neurosci. 2002;22:4973–86.
58. Yoshida M, Assimacopoulos S, Jones KR,G. EA. Massive loss of Cajal-Retzius cells does not disrupt neocortical layer order. Development. 2006;133:537-45.
59. Bielle F, Griveau A, Narboux-Neme N, Vigneau S, Sigrist M, Arber S, Wassef M,P. A. Multiple origins of Cajal-Retzius cells at the borders of the developing pallium. Nat Neurosci. 2005;8:1002-12.
60. Hanashima C, Fernandes M, Hebert JM,F. G. The role of Foxg1 and dorsal midline signaling in the generation of Cajal-Retzius subtypes. J Neurosci. 2007;27:11103-11.
61. Lavdas AA, Grigoriou M, Pachnis V,P. JG. The Medial Ganglionic Eminence Gives Rise to a Population of Early Neurons in the Developing Cerebral Cortex. J Neurosci. 1999;99:7881-8.

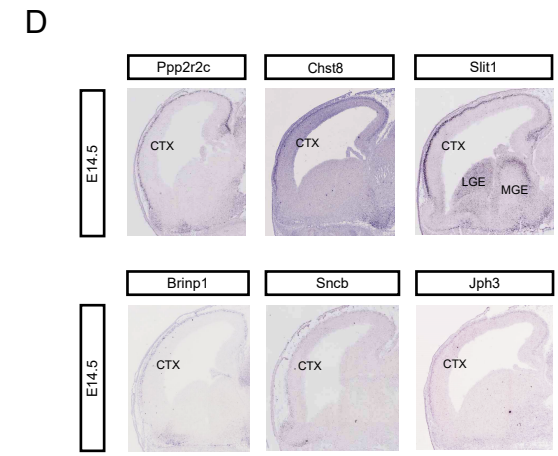
62. Takiguchi-Hayashi K, Sekiguchi M, Ashigaki S, Takamatsu M, Hasegawa H, Suzuki-Migishima R, Yokoyama M, Nakanishi S,T. Y. Generation of reelin-positive marginal zone cells from the caudomedial wall of telencephalic vesicles. J Neurosci. 2004;24:2286-95.
63. Cabrera-Socorro A, Hernandez-Acosta NC, Gonzalez-Gomez M,M. G. Comparative aspects of p73 and Reelin expression in Cajal-Retzius cells and the cortical hem in lizard, mouse and human. Brain Res. 2007;1132:59-70.
64. Tissir F, Ravni A, Achouri Y, Riethmacher D, Meyer G,G. AM. DeltaNp73 regulates neuronal survival in vivo. PNAS. 2009;106:16871–16876.
65. J. Rong, J. R. McGuire, Z. H. Fang, G. Sheng, J. Y. Shin, S. H. Li,X. J. Li. Regulation of intracellular trafficking of huntingtin-associated protein-1 is critical for TrkA protein levels and neurite outgrowth. J Neurosci. 2006;26:6019-30.
66. Fujino T, Leslie JH, Eavri R, Chen JL, Lin WC, Flanders GH, Borok E, Horvath TL,N. E. CPG15 regulates synapse stability in the developing and adult brain. Genes Dev. 2011;25:2674-85.
67. Mascaro ALA, Cesare P, Sacconi L, Grasselli G, Mandolesi G, Maco B, Knott GW, Huang L, De Paola V, Strata P,P. FS. In vivo single branch axotomy induces GAP-43–dependent sprouting and synaptic remodeling in cerebellar cortex. PNAS. 2013;110:10824-9.
68. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ,P. L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511-5.
69. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H,G. CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38:576-89.
70. Levy R, Molina J, Danielpour M,B. JJ. Neonatal pial surface electroporation. J Vis Exp. 2014:1-7.
71. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y,O. SH. Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell. 2009;139:1303-14.

Figure 1


A
E11.5
E13.5
E15.5
GFP
17.1%
8.37%
3.9%
Cells
GFP-A
Ebf2-EGFP mice
Flow Cytometer Sorting


B
Color Key
5
0
-5
E11.5 E13.5 E15.5 E11.5 E13.5 E15.5
Ebf2-EGFP-
Ebf2-EGFP+


C
Ebf2-EGFP-
Ebf2-EGFP+
E11.5
E13.5
E15.5
607 (32%)
474 (25%)
147 (7.8%)
420 (22.2%)
10 (0.5%)
131 (6.9%)
105 (5.5%)
647 (42.8%)
204 (13.5%)
121 (8%)
166 (11%)
27 (1.8%)
152 (10%)
196 (13%)


D
Ppp2r2c
Chst8
Slit1
Brinp1
Sncb
Jph3
E14.5
CTX
LGE
MGE

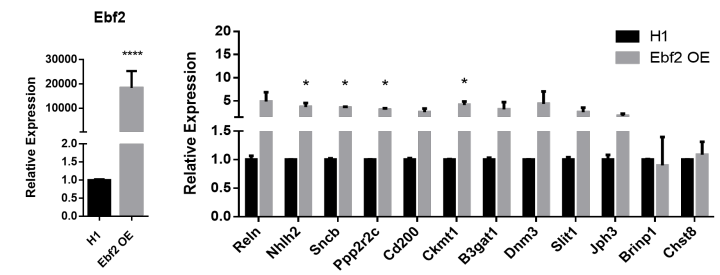

E
Ebf2
Relative Expression
H1
Ebf2 OE
****
*
Reln
Nhlh2
Sncb
Ppp2r2c
Cd200
Ckmt1
B3gat1
Dnm3
Slit1
Jph3
Brinp1
Chst8

Figure 2

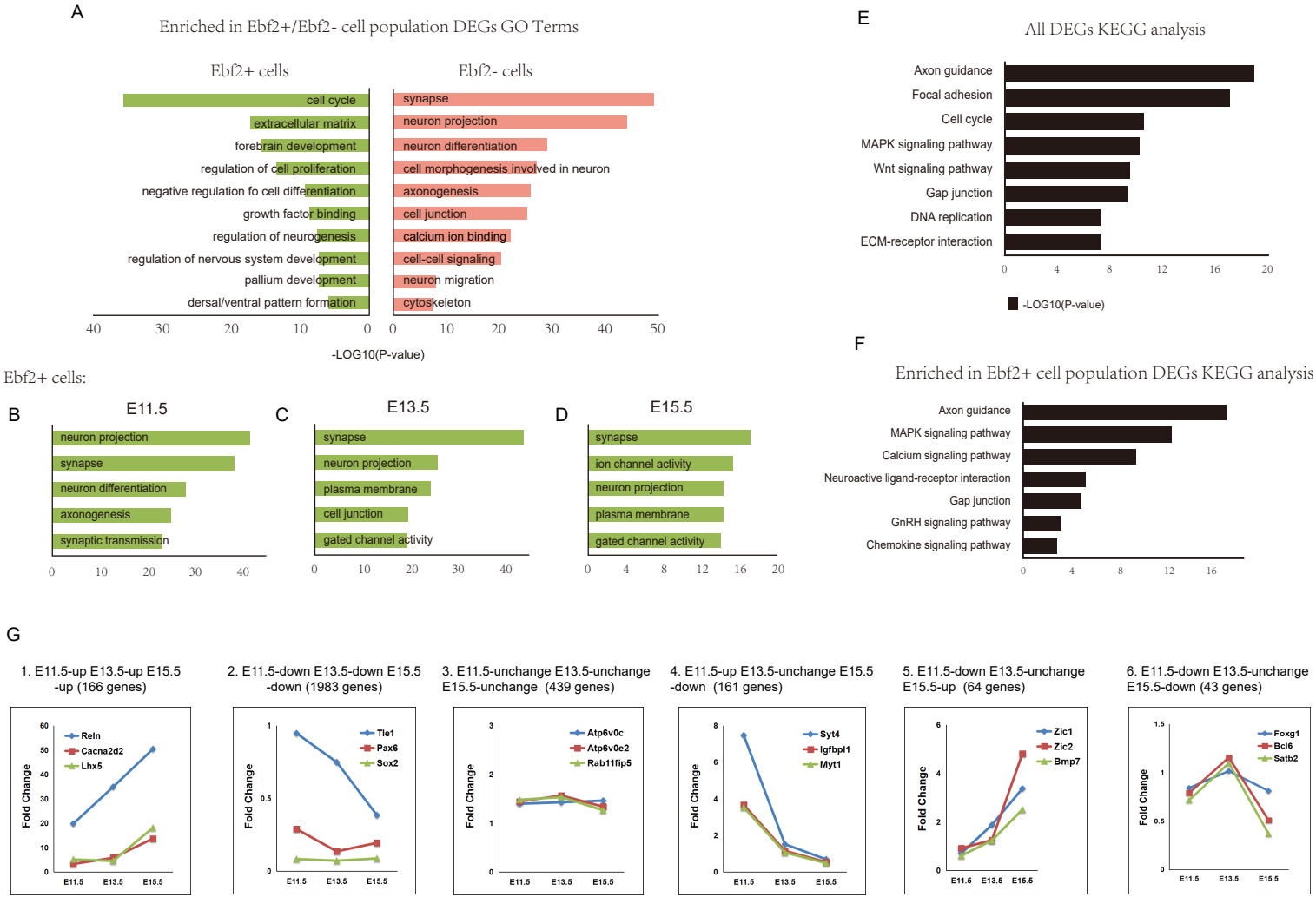

A
Enriched in Ebf2+/Ebf2- cell population DEGs GO Terms
Ebf2+ cells
Ebf2- cells
cell cycle
extracellular matrix
forebrain development
regulation of cell proliferation
negative regulation fo cell differentiation
growth factor binding
regulation of neurogenesis
regulation of nervous system development
pallium development
dersal/ventral pattern formation
synapse
neuron projection
neuron differentiation
cell morphogenesis involved in neuron
axonogenesis
cell junction
calcium ion binding
cell-cell signaling
neuron migration
cytoskeleton
-LOG10(P-value)
E
All DEGs KEGG analysis
Axon guidance
Focal adhesion
Cell cycle
MAPK signaling pathway
Wnt signaling pathway
Gap junction
DNA replication
ECM-receptor interaction
-LOG10(P-value)
Ebf2+ cells:
B
E11.5
neuron projection
synapse
neuron differentiation
axonogenesis
synaptic transmission
C
E13.5
synapse
neuron projection
plasma membrane
cell junction
gated channel activity
D
E15.5
synapse
ion channel activity
neuron projection
plasma membrane
gated channel activity
F
Enriched in Ebf2+ cell population DEGs KEGG analysis
Axon guidance
MAPK signaling pathway
Calcium signaling pathway
Neuroactive ligand-receptor interaction
Gap junction
GnRH signaling pathway
Chemokine signaling pathway
G
1. E11.5-up E13.5-up E15.5 -up (166 genes)
Reln
Cacna2d2
Lhx5
2. E11.5-down E13.5-down E15.5 -down (1983 genes)
Tle1
Pax6
Sox2
3. E11.5-unchange E13.5-unchange E15.5-unchange (439 genes)
Atp6v0c
Atp6v0e2
Rab11fip5
4. E11.5-up E13.5-unchange E15.5 -down (161 genes)
Syt4
Igfbpl1
Myt1
5. E11.5-down E13.5-unchange E15.5-up (64 genes)
Zic1
Zic2
Bmp7
6. E11.5-down E13.5-unchange E15.5-down (43 genes)
Foxg1
Bcl6
Satb2
Fold Change
E11.5
E13.5
E15.5

Figure 3

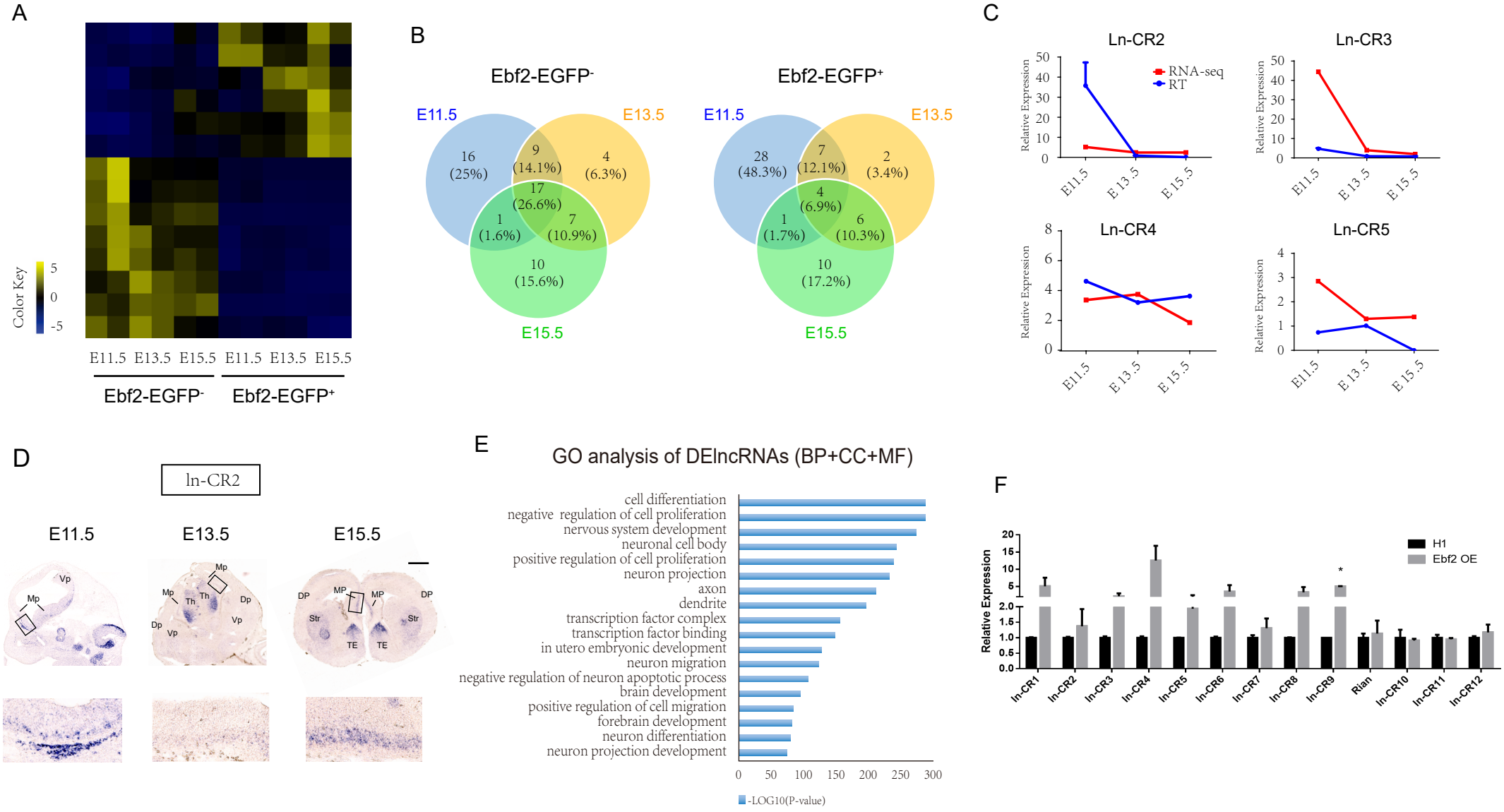

A
Color Key
5
0
-5
E11.5 E13.5 E15.5 E11.5 E13.5 E15.5
Ebf2-EGFP-
Ebf2-EGFP+
B
Ebf2-EGFP-
E11.5
E13.5
E15.5
16 (25%)
9 (14.1%)
4 (6.3%)
17 (26.6%)
1 (1.6%)
7 (10.9%)
10 (15.6%)
Ebf2-EGFP+
E11.5
E13.5
E15.5
28 (48.3%)
7 (12.1%)
2 (3.4%)
4 (6.9%)
1 (1.7%)
6 (10.3%)
10 (17.2%)
C
Ln-CR2
Ln-CR3
Ln-CR4
Ln-CR5
RNA-seq
RT
Relative Expression
E11.5
E13.5
E15.5
D
ln-CR2
E11.5
E13.5
E15.5
Vp
Mp
Th
Dp
DP
Str
TE
MP
E
GO analysis of DElncRNAs (BP+CC+MF)
cell differentiation
negative regulation of cell proliferation
nervous system development
neuronal cell body
positive regulation of cell proliferation
neuron projection
axon
dendrite
transcription factor complex
transcription factor binding
in utero embryonic development
neuron migration
negative regulation of neuron apoptotic process
brain development
positive regulation of cell migration
forebrain development
neuron differentiation
neuron projection development
0 50 100 150 200 250 300
-LOG10(P-value)
F
H1
Ebf2 OE
Relative Expression
In-CR1
In-CR2
In-CR3
In-CR4
In-CR5
In-CR6
In-CR7
In-CR8
In-CR9
Rian
In-CR10
In-CR11
In-CR12

Figure 4

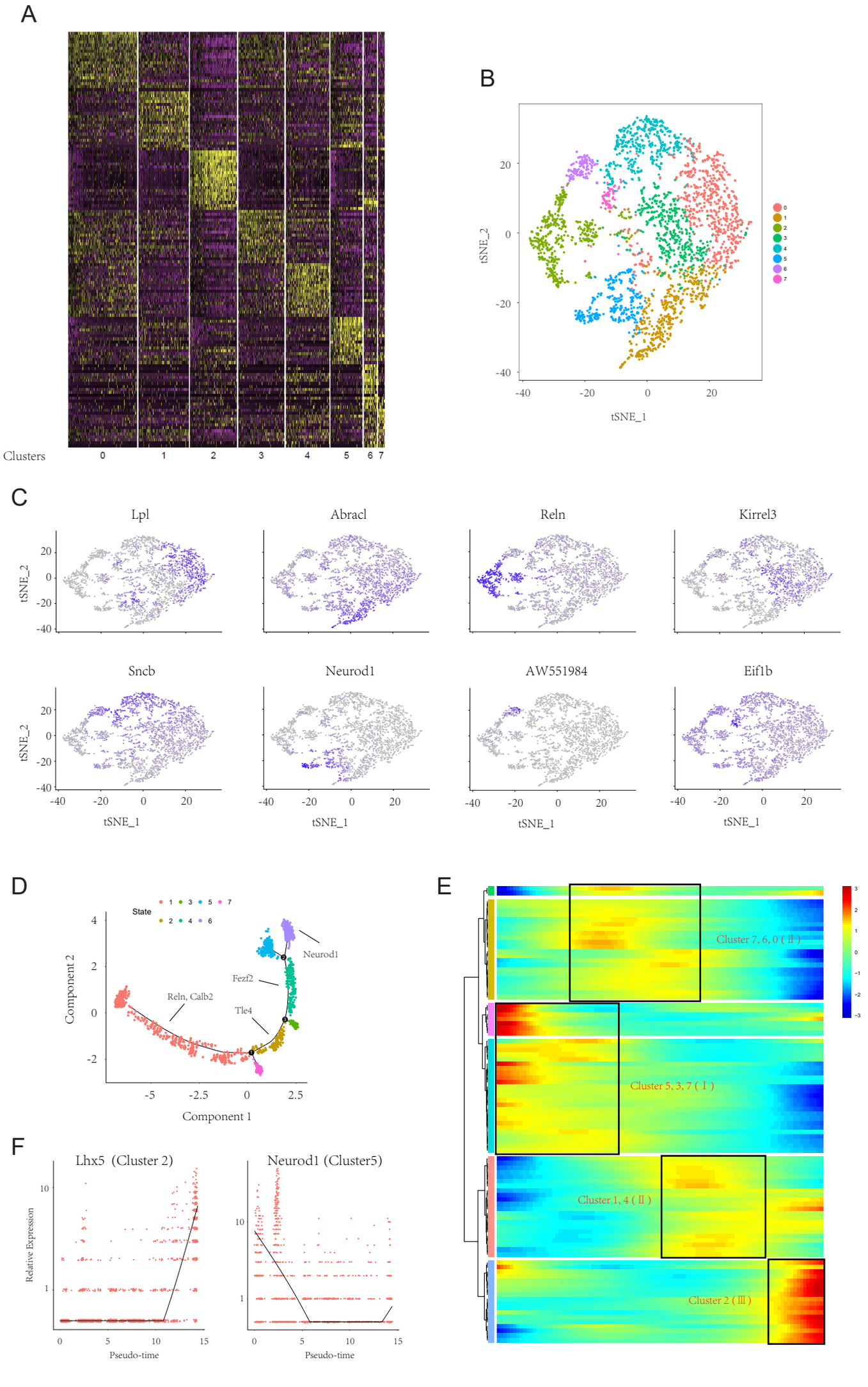
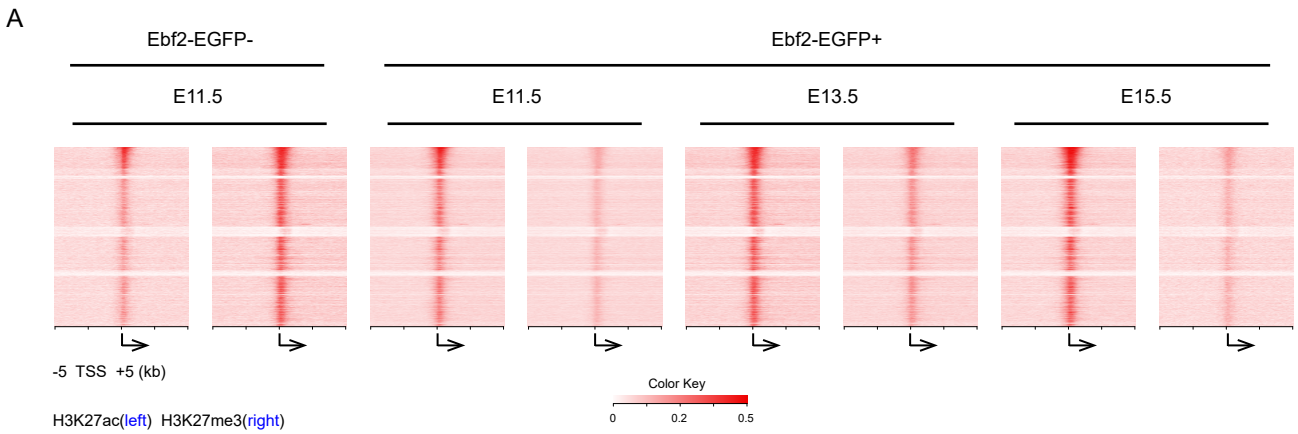
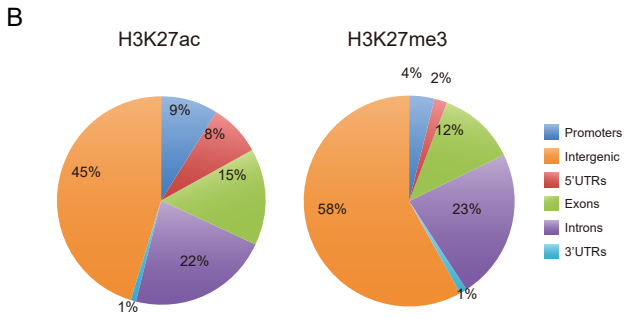
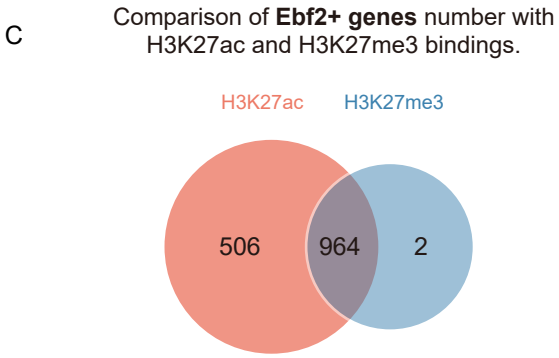
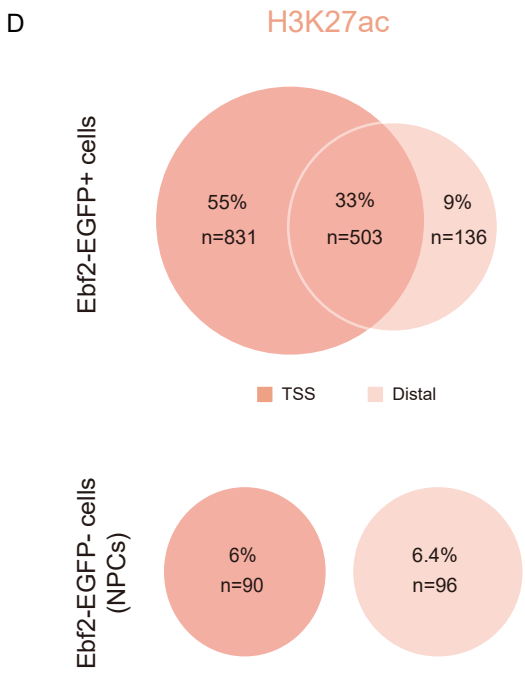

A
Clusters
0 1 2 3 4 5 6 7
B
tSNE_2
tSNE_1
C
Lpl
Abracl
Reln
Kirrel3
Sncb
Neurod1
AW551984
Eif1b
tSNE_2
tSNE_1
D
State
Component 2
Component 1
Neurod1
Fezf2
Reln, Calb2
Tle4
E
Cluster 7, 6, 0 (Ⅱ)
Cluster 5, 3, 7 (Ⅰ)
Cluster 1, 4 (Ⅱ)
Cluster 2 (Ⅲ)
F
Lhx5 (Cluster 2)
Neurod1 (Cluster5)
Relative Expression
Pseudo-time

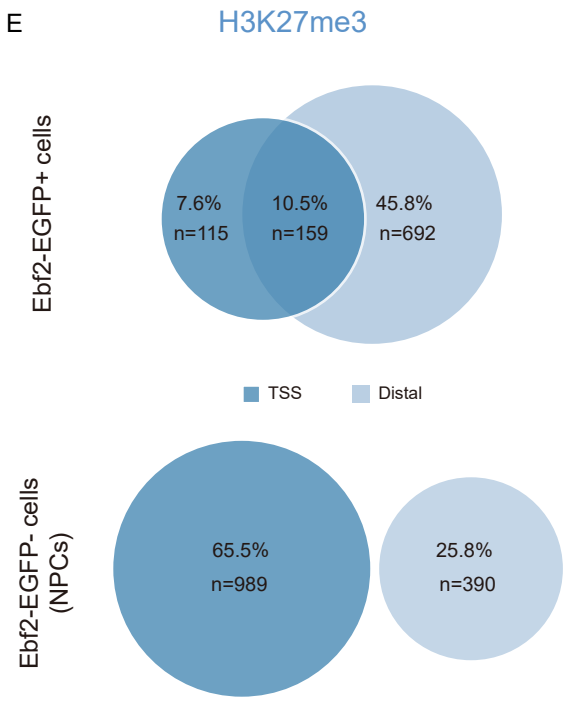
Figure 5


A
Ebf2-EGFP-
Ebf2-EGFP+
E11.5
E11.5
E13.5
E15.5
-5 TSS +5 (kb)
Color Key
0
0.2
0.5
H3K27ac(left) H3K27me3(right)


B
H3K27ac
H3K27me3
9%
8%
15%
22%
1%
45%
4%
2%
12%
23%
1%
58%
Promoters
Intergenic
5'UTRs
Exons
Introns
3'UTRs


C
Comparison of Ebf2+ genes number with H3K27ac and H3K27me3 bindings.
H3K27ac
H3K27me3
506
964
2


D
H3K27ac
Ebf2-EGFP+ cells
55%
n=831
33%
n=503
9%
n=136
TSS
Distal
Ebf2-EGFP- cells (NPCs)
6%
n=90
6.4%
n=96


E
H3K27me3
Ebf2-EGFP+ cells
7.6%
n=115
10.5%
n=159
45.8%
n=692
TSS
Distal
Ebf2-EGFP- cells (NPCs)
65.5%
n=989
25.8%
n=390

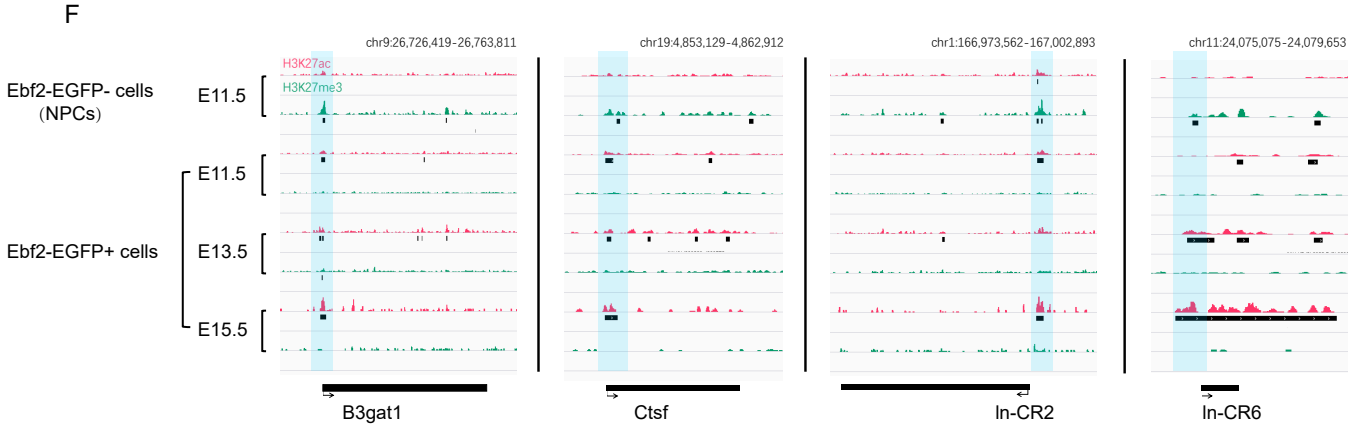

F
chr9:26,726,419-26,763,811
chr19:4,853,129-4,862,912
chr1:166,973,562-167,002,893
chr11:24,075,075-24,079,653
H3K27ac
H3K27me3
Ebf2-EGFP- cells (NPCs)
E11.5
Ebf2-EGFP+ cells
E11.5
E13.5
E15.5
B3gat1
Ctsf
In-CR2
In-CR6

Figure 6

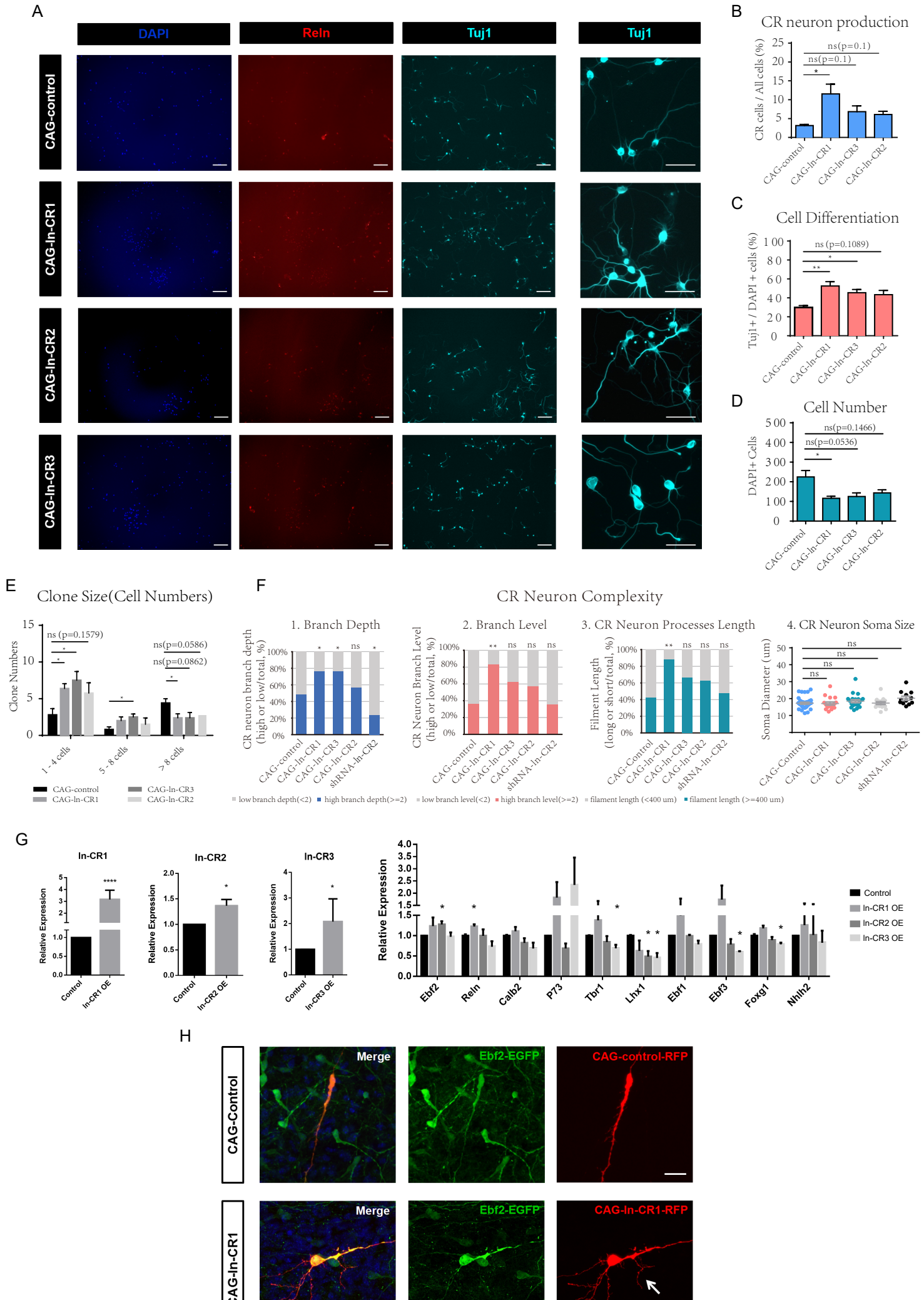

A
DAPI
Reln
Tuj1
Tuj1
CAG-control
CAG-ln-CR1
CAG-ln-CR2
CAG-ln-CR3
B
CR neuron production
CR cells / All cells (%)
ns(p=0.1)
ns(p=0.1)
*
CAG-control
CAG-ln-CR1
CAG-ln-CR3
CAG-ln-CR2
C
Cell Differentiation
Tuj1+ / DAPI + cells (%)
ns (p=0.1089)
*
**
D
Cell Number
DAPI+ Cells
ns(p=0.1466)
ns(p=0.0536)
*
E
Clone Size(Cell Numbers)
Clone Numbers
ns (p=0.1579)
ns(p=0.0586)
ns(p=0.0862)
1 - 4 cells
5 - 8 cells
> 8 cells
CAG-control
CAG-ln-CR1
CAG-ln-CR3
CAG-ln-CR2
F
CR Neuron Complexity
1. Branch Depth
CR neuron branch depth (high or low/total, %)
2. Branch Level
CR Neuron Branch Level (high or low/total, %)
3. CR Neuron Processes Length
Filment Length (long or short/total, %)
4. CR Neuron Soma Size
Soma Diameter (um)
CAG-Control
CAG-ln-CR1
CAG-ln-CR3
CAG-ln-CR2
shRNA-ln-CR2
low branch depth(<2)
high branch depth(>=2)
low branch level(<2)
high branch level(>=2)
filament length (<400 um)
filament length (>=400 um)
G
In-CR1
In-CR2
In-CR3
Relative Expression
Control
In-CR1 OE
In-CR2 OE
In-CR3 OE
Ebf2
Reln
Calb2
P73
Tbr1
Lhx1
Ebf1
Ebf3
Foxg1
Nhlh2
H
CAG-Control
CAG-ln-CR1
Merge
Ebf2-EGFP
CAG-control-RFP
CAG-ln-CR1-RFP

Fig S1, related to Figure 1

A

Ebf2-EGFP mice, E11.5

Sorted: Ebf2-EGFP+ cell

Day 0

Day 4

Sorted: Ebf2-EGFP- cell

Day 0

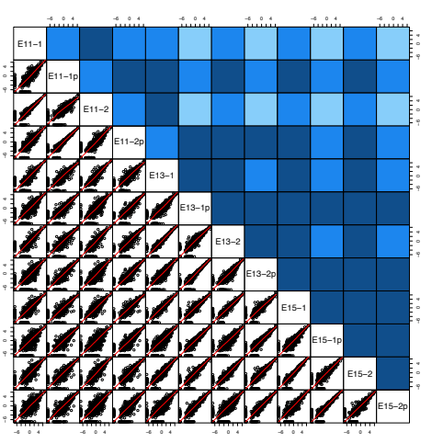
Day 4

Sorted: Ebf2-EGFP+ cell &
Ebf2-EGFP- cell

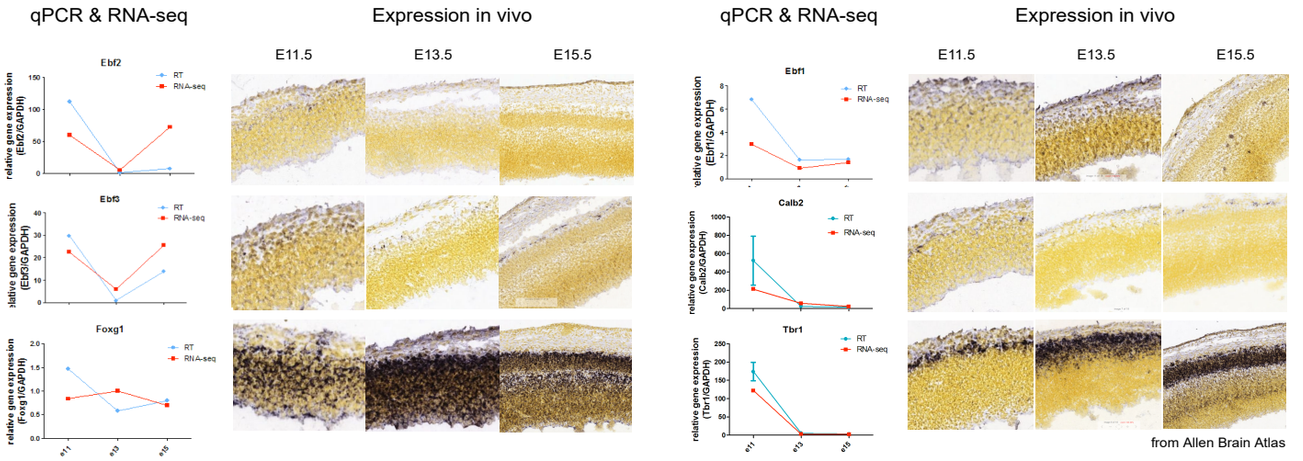
Day 0

Day 4

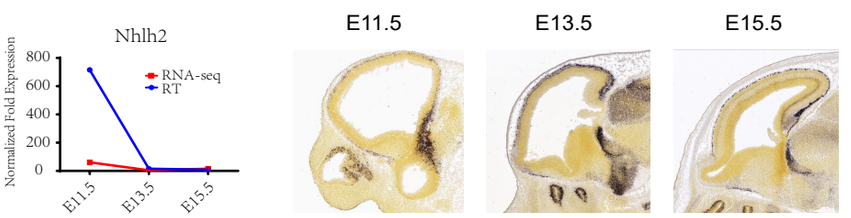
B



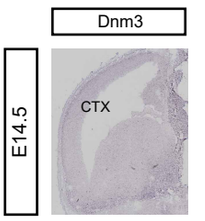
C



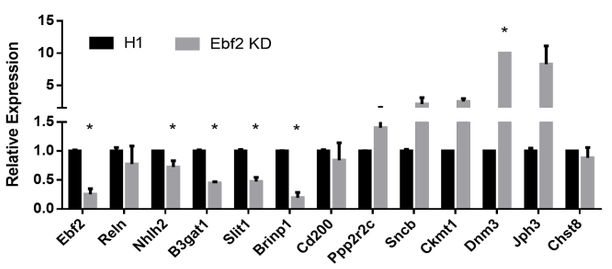
D



E



F



G

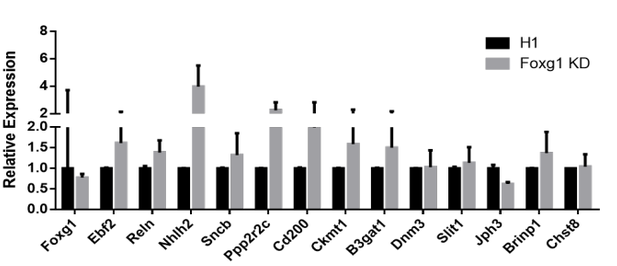


Fig S2, related to Fig2

A

Gene	Gene Ontology
Reln	Cell migration, Generation of neurons, Neuron differentiation, Neurite development
Cacna2d2	Axon growth and regeneration, Epilepsy, Ataxia
Lhx5	Forebrain neuron differentiation, Cajal-Retzius cells normal distribution, cerebellar Purkinje cell differentiation

Gene	Gene Ontology
Tle1	Cortical cell differentiation, Wnt signaling pathway, Transcription factor binding
Pax6	Neural precursor cell proliferation; Interkinetic nuclear migration; Cell fate commitment;
Sox2	Stem cell maintenance; determination of cell fate; forebrain development

Gene	Gene Ontology
Atp6v0c	ATP hydrolysis coupled proton transport; Vacuolar acidification; Ion transport
Atp6v0e2	ATP hydrolysis coupled proton transport; Proton transmembrane transport; V-type ATPase
Rab11fip5	Cellular response to acidic pH; protein transport; regulated exocytosis

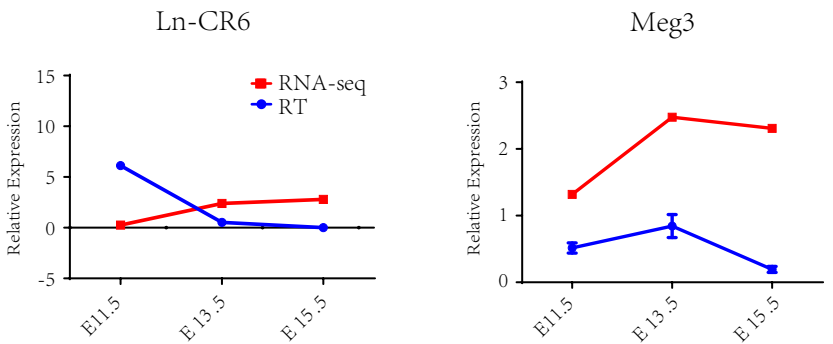
Gene	Gene Ontology
Syt4	Subplate gene Synaptic vesicle exocytosis; Neurotransmitter secretion
Igfbpl1	Subplate gene; Regulation of cell growth
Myt1	Subplate gene; Regulation of oligodendrocyte differentiation and proliferation

Gene	Gene Ontology
Zic1	Medial forebrain development; Cerebellum development; Cell differentiation
Zic2	Cortical lamination; Medial forebrain development; Holoprosencephaly
Bmp7	Epithelial cell morphogenesis; Premature radial gial differentiation; Neuronal migration

Gene	Gene Ontology
Foxg1	Cortical neuron development; Neuron fate determination; Telencephalic patterning
Bcl6	Neuronal differentiation; Regulation of Notch signaling; Cell morphogenesis
Satb2	Upper layer neuron specification; Chromatin remodeling; Transcription regulation

Fig S3, related to Fig3

A



B

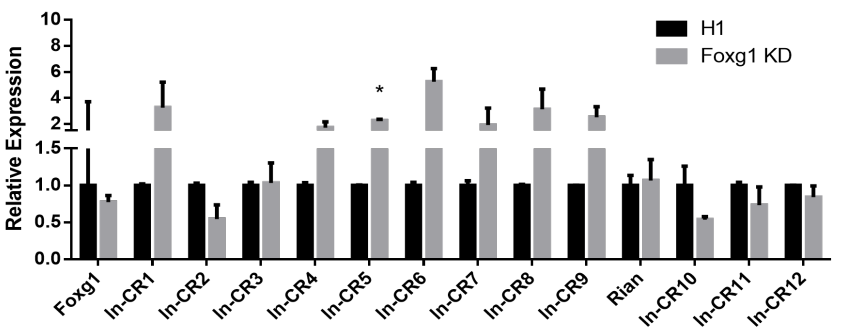


Fig S4, related to Figure 4

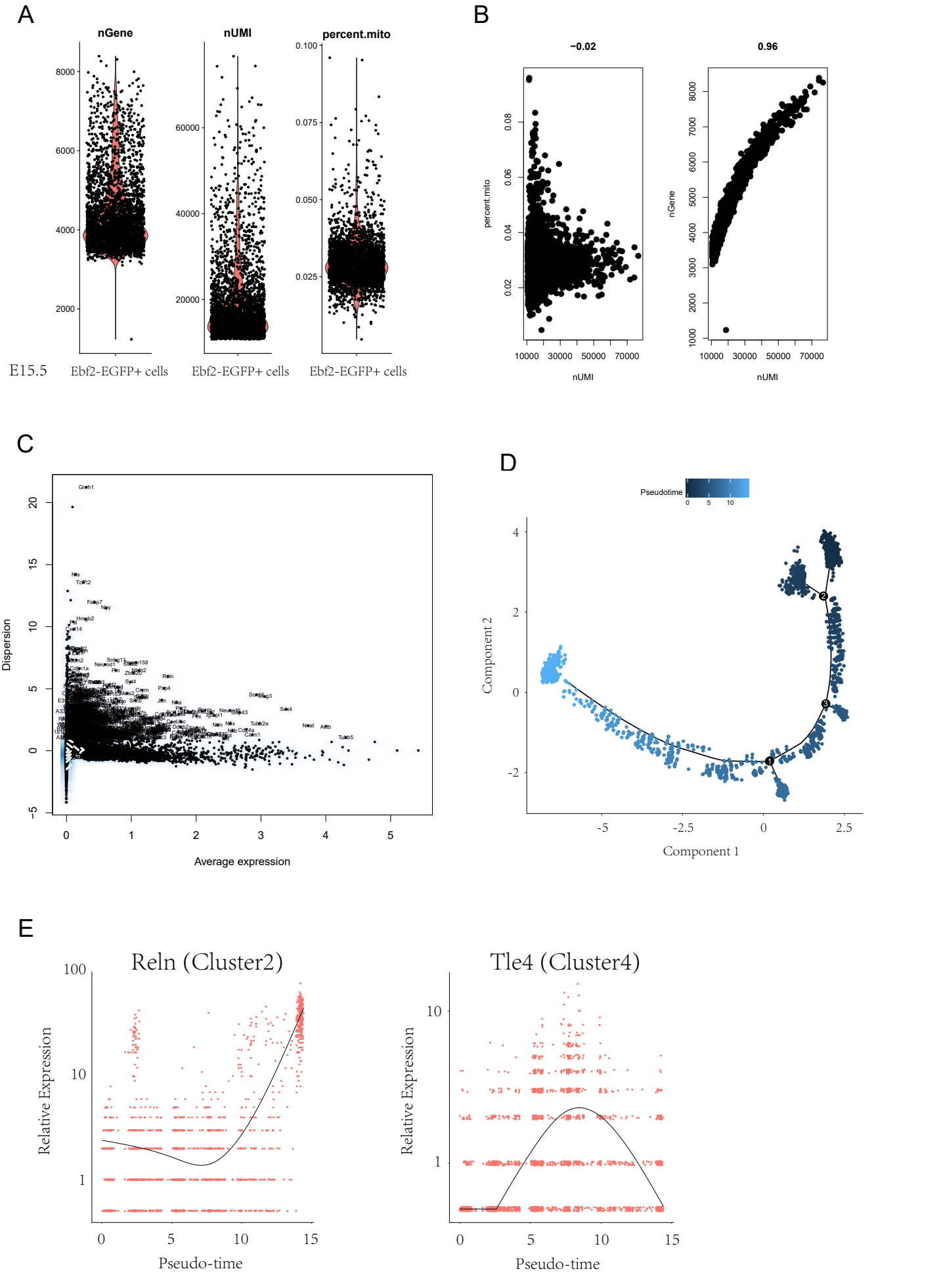


Fig S5, related to Figure5

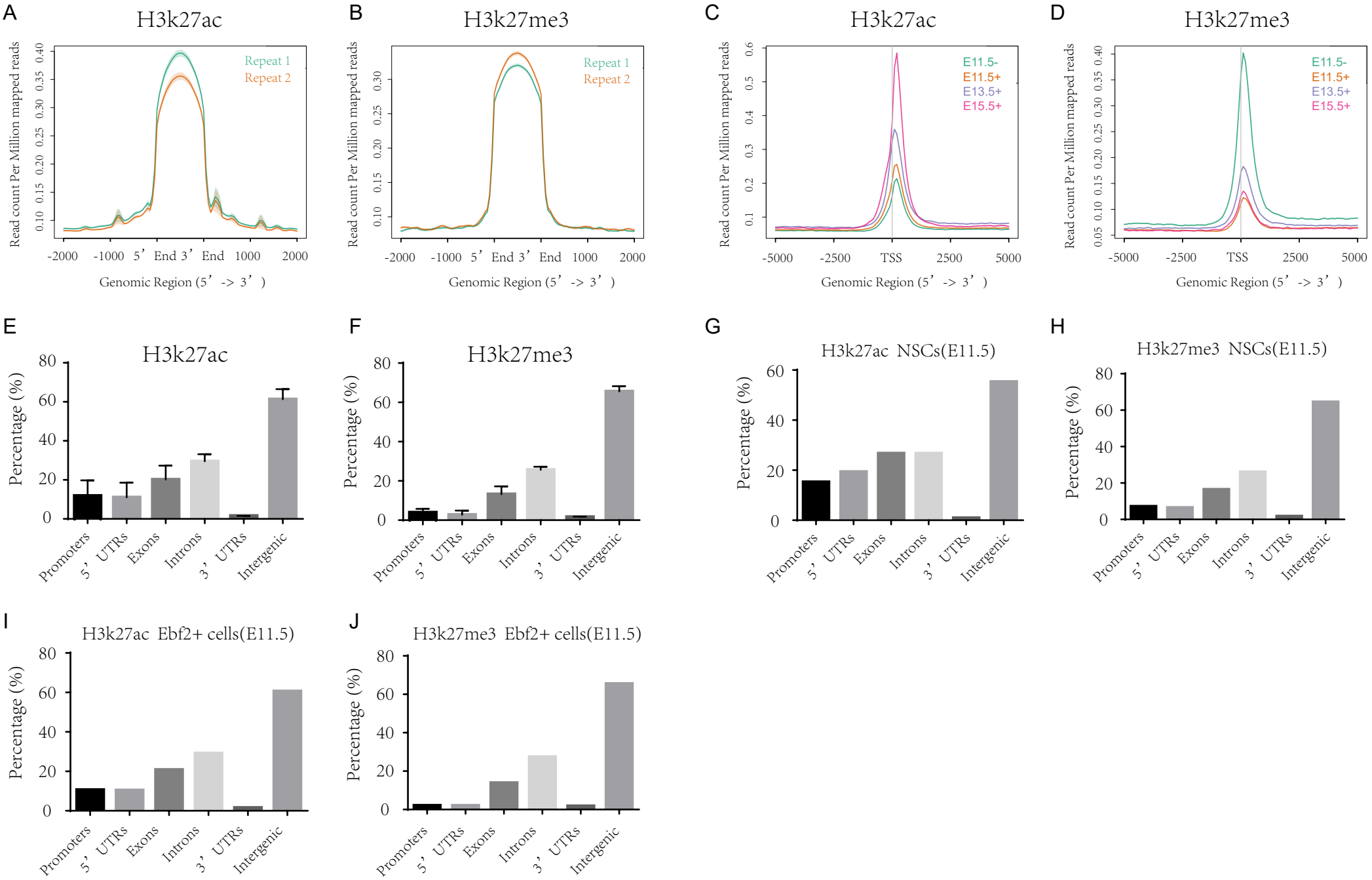
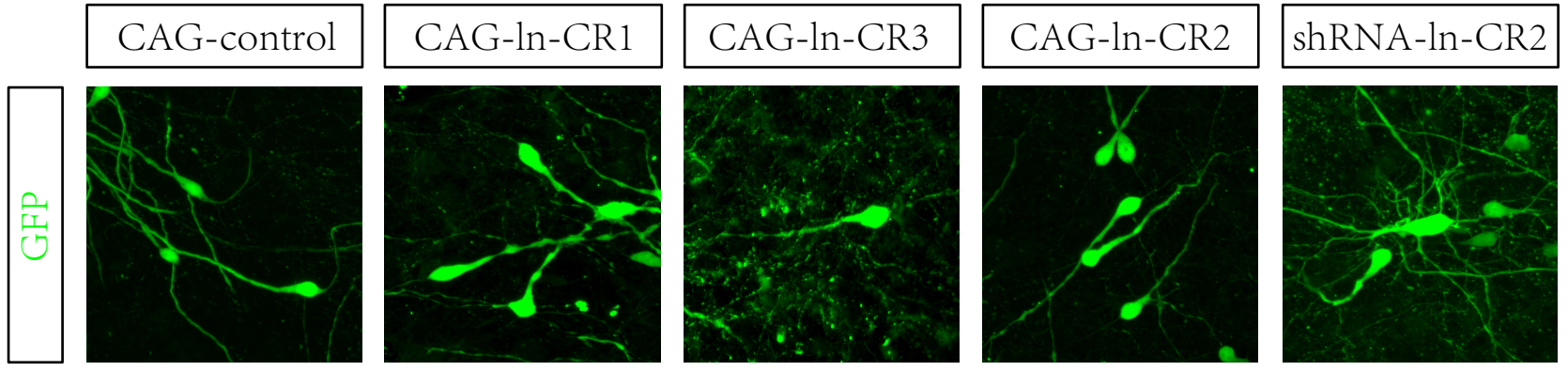
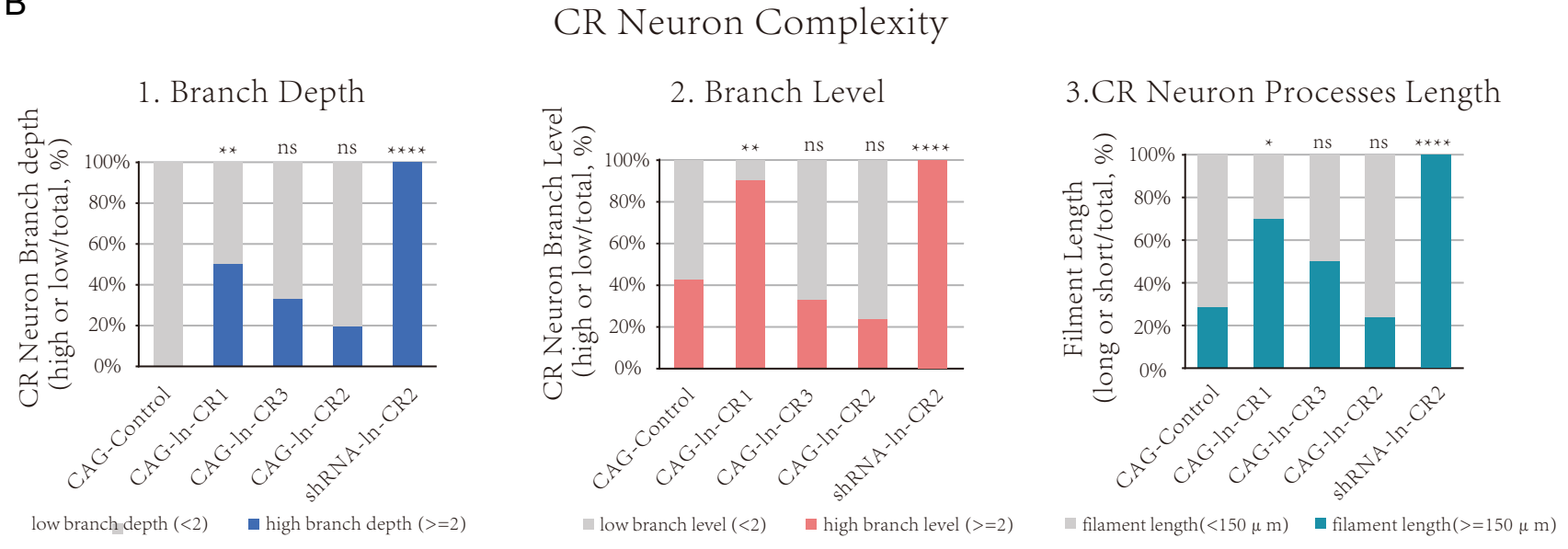


Fig S6, related to Figure6

A



B