Transcriptome analysis of early stage of neurogenesis reveals temporal-specific gene network for preplate neuron differentiation and Cajal-Retzius cell specification

Jia Li[1,2,3†], Lei Sun[4†], Xue-Liang Peng[1], Xiao-Ming Yu[2], Shao-Jun Qi[1,2], Zhi John Lu[5], Jing-Dong J. Han[6], Qin Shen[1*].

†Equal contributors

[1] Current address: 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 Medicine, Tsinghua University, Beijing 100084, People's Republic of China

[3] PTN graduate program, School of Life Sciences, Peking University, Beijing 100871, People's Republic of China

[4] PTN graduate program, School of Life Sciences, Tsinghua University, Beijing 100084, People's Republic of China

[5] MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, School of Life Sciences, Tsinghua University, Beijing 100084, People's Republic of China

[6] Key Laboratory of Computational Biology, CAS Center for Excellence in Molecular Cell Science, Collaborative Innovation Center for Genetics and Developmental Biology, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, People's Republic of China

*Correspondence: shenqin@tongji.edu.cn

Abstract

Background: Neurogenesis in the developing cerebral cortex begins with the generation of the preplate, which consists of early born neurons including Cajal-Retzius (CR) cells and subplate neurons. CR cells later segregate to layer 1, the most superficial layer, and play an important role in organizing cortical lamination. However, the molecular mechanisms underlying the generation of preplate and CR cell differentiation remain poorly understood.

Results: Using the Ebf2-EGFP transgenic mouse line in which EGFP labels the preplate first then persists in CR cells, we compared differentially expressed gene profile in Ebf2-EGFP-positive and Ebf2-EGFP-negative cell population by high-throughput RNA-seq at different stages during early cortical development. Bioinformatic analysis revealed the temporal gene expression dynamics of Ebf2-EGFP+ cells and identified a number of coding genes and lncRNAs that were involved in early neuronal differentiation. Furthermore, single cell RNA-seq analysis unraveled eight subpopulations of Ebf2-EGFP+ cells at E15.5, which were classified into three developing states using pseudotiming analysis. We also found that the upstream regulatory regions of Ebf2+ cell-enriched genes including lncRNAs were associated with active histone modifications, indicating epigenetic involvements in early differentiating neuron gene expression. Furthermore, we characterized the functional role of Ebf2+ cell-enriched lncRNAs in early neuron differentiation using assays including primary cell culture, wholemount culture and *In Utero* Electroporation of layer 1.

Conclusions: We have provided a genetic and epigenetic atlas of early neurogenesis and identified genes associated with CR neuron identity. Our work will help understand the molecular mechanisms governing the first differentiation step of cortical development.

Keywords: neurogenesis, preplate, Cajal-Retzius neuron, cerebral cortex, neural 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), the neuroepithelial cells and radial glial cells in the ventricular zone (VZ) 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

58 plexiform layer, which situates above the ventricular zone and beneath the pial surface [1-3]. The
59 preplate is then split into the marginal zone (MZ, layer 1) and the subplate by immigrating cortical
60 plate neurons around E13.5 in mouse and the seventh to eighth week of gestation in human [4-6],
61 setting up of the upper and lower borders of the cortical plate. The formation and splitting of
62 preplate are the first steps in the establishment of the highly organized cortical layers.

63 The preplate and its derivatives play a critical role in regulating subsequent inside-out
64 neuronal migration, cortical lamination and the establishment of topographic connections [7].
65 Subplate neurons contribute to the guidance of corticofugal and thalamocortical axons during
66 early development and are important for the functional maturation and plasticity of the cortical
67 circuitry [8, 9]. Cajal-Retzius neurons are the major cell type in layer 1, the most superficial
68 cortical layer. Discovered more than a century ago by Retzius and Ramon y Cajal, CR neurons
69 have been studied extensively [3, 10]. Best known for their expression of Reln, CR neurons
70 regulate the inside-out migration of cortical neurons, correct layer formation, maintenance of
71 radial glia and cortical patterning [11]. Loss of Reln or mislocalization of CR cells contributes to
72 defective migration and disordered cortical layer formation that are associated with human brain
73 malformation such as lissencephaly [12] and dystroglycanopathy [13]. Although functional
74 importance of these early differentiating neurons is well recognized, characterization of
75 cell-intrinsic, dynamic gene expression patterns in these cells remains incomplete. Recent studies
76 have identified the genes enriched in subplate neurons by microarray and RNA-seq analysis
77 [14-18], providing a comprehensive gene expression profile of subplate at different developmental
78 stages, which offers valuable insight on molecular mechanisms underlying neurodevelopmental
79 disorders [17]. However, how gene expression is dynamically regulated during the transition from
80 the preplate to CR neurons remains unclear.

81 We previously have shown that in the Ebf2-EGFP transgenic mice (GENSAT project) [19],
82 the GFP signal is expressed specifically in early differentiating preplate neurons and persistent in
83 CR cells [20]. After preplate splitting, virtually all of the EGFP-positive cells exhibit
84 morphological and molecular features of CR neurons in the embryonic and postnatal brain [20],
85 consistent with the pattern of the Ebf2^{GFPiCre} [21], making this line suitable for tracing the
86 development and profiling genetic and epigenetic pattern of early born neurons including CR
87 neurons. In addition, single-cell gene expression profiling allows further dissection of the

88 heterogeneity and developmental trajectory within a cell population [22, 23]. While multiple types
89 of neurons in hippocampus and adult cortex have been identified by single-cell RNA-sequencing
90 [23, 24], a full characterization of early differentiating neurons at single-cell level has not been
91 done before. In this study, we used FACS-purified Ebf2-EGFP-expressing cells to probe the
92 dynamic molecular features of preplate and CR neurons during the early stages of forebrain
93 development. Whole genome transcriptome analysis identified the initial difference between the
94 preplate and progenitor cells at E11.5, reflecting the molecular choice during the first cell fate
95 decision in neural differentiation. We also recognized clusters of genes that displayed temporal
96 changes corresponding to different developmental trajectories. Furthermore, single-cell
97 sequencing of E15.5 CR neurons identified novel subpopulations. Genome-wide overview of
98 active (H3K27ac) and repressive (H3K27me3) histone modifications in preplate neurons, CR
99 neurons and NPCs revealed that their occupancies around TSS and distal region were associated
100 with expression of early differentiating neurons-enriched genes and lncRNAs. By bioinformatic
101 functional enrichment analysis, we also identified and tested the function of genes and lncRNAs
102 that were enriched in CR neurons in neuronal differentiation. Hence, our study provides unique
103 molecular markers to extend the repertoire of preplate and CR neuron labels and reveals
104 remarkable dynamics in gene expression during early cortical development.
105
106 **Results**
107 **Characterization of Ebf2-EGFP-expressing cells in the early cerebral cortex**
108 We purified Ebf2-EGFP expressing cells by FACS from the embryonic mouse forebrain at E11.5,
109 E13.5 and E15.5, to compare the developmental gene expression dynamics during preplate
110 formation and differentiation of early neurons. At E11.5, Ebf2-EGFP+ cells are mostly preplate
111 neurons including both subplate neurons and CR neurons, and at E13.5, Ebf2-EGFP expression is
112 diminishing in subplate neurons but maintained in CR neurons. By E15.5 Ebf2-EGFP is
113 specifically expressed in CR neurons [20]. At E11.5, Ebf2-expressing cells consisted of 17.1% ±
114 1.07% of the total cells. At E13.5 and E15.5, the percentage of Ebf2-expressing cells gradually
115 decreased to 8.37 ± 2.29% and 3.9 ± 1.59%, respectively (Figure 1A). 100% of FACS-sorted
116 Ebf2-EGFP+ cells were β-III tubulin positive at all stages, indicating these cells have committed
117 neuronal fate decision, consistent with our previous in vivo analysis [20]. Accordingly, after

118 culturing sorted and unsorted cells for 4 days, almost all of the Ebf2-EGFP-positive cells remained
119 single and extended processes, while the Ebf2-EGFP-negative cells proliferated and generated
120 large clones similar to the unsorted cell group, indicating a good separation of differentiating
121 neurons from progenitor cells (Figure S1A).
122
123 **Identification of Ebf2-EGFP+ cell-enriched genes in the developing mouse forebrain**
124 Next, we collected FACS-selected Ebf2-EGFP positive cells and Ebf2-EGFP negative cells from
125 the three early neurogenic stages and isolated total RNA for whole-genome transcriptome
126 sequencing. Two biological replicates for each cell type (Ebf2-EGFP+ and Ebf2-EGFP-) were
127 sequenced and expressed transcripts were identified as described in method. To verify the
128 reproducibility of our samples, we analyzed the Pearson's correlation coefficient (Pearson's r) for
129 the replicates and there was a strong association between different preparations of FACS-sorted
130 cells from the same stage (Figure S1B).
131 To validate the RNA-seq results, we selected several protein-coding genes, including CR
132 neurons markers (Ebf2 and Calb2) and genes associated with CR neurons fate regulation (Foxg1,
133 Tbr1, Ebf1, and Ebf3) [10, 20, 25-28], and verified by qPCR that the temporal dynamic expression
134 levels of these genes were in accordance with that detected by RNA-seq (Figure S1C).
135 Furthermore, the temporal expression trends were consistent with the mRNA expression pattern
136 revealed by in situ hybridization (ISH) from the data available in Allen Brain Atlas data (Figure
137 S1C).
138 We used DE-Seq to analyze the dynamic expression profiles during the early cortical
139 differentiation period. 4859 genes exhibited significant difference in expression level between the
140 two cell populations during the time course (FDR<0.05, p-values<0.05 by Fisher's exact test) and
141 were selected as differential expressed genes (DEGs) for further analysis (Figure 1B; Additional
142 file 1).
143 We analyzed DEGs expressed in any of the developmental stages sequenced (E11.5, E13.5
144 and E15.5). Genes with expression level (FPKM) more than 1, and 2-fold higher in Ebf2-EGFP
145 positive population compared to that in Ebf2-EGFP negative population were defined as
146 Ebf2-EGFP positive cells enriched genes, Ebf2+ genes in short. We found that there were 1044
147 genes at E11.5, 643 genes at E13.5, and 541 at E15.5 enriched in Ebf2-EGFP positive cells,

148 respectively. Considerable overlap was found between the different stages (Figure 1C). In total,
149 1513 genes displayed enriched expression in Ebf2-EGFP positive population at least at one stage,
150 and 1894 such genes in Ebf2-EGFP negative population. Of these, 166 genes were consistently
151 enriched in Ebf2-EGFP positive population throughout all three embryonic stages, and 420 genes
152 were differentially expressed in Ebf2-EGFP negative population (Figure 1C). Of the 166 DEGs
153 included widely-used CR neuron markers such as Reln and Trp73, and genes known to be
154 expressed in the CR neurons such as Calb2 (Calretinin), Tbr1, Ebf3, Cdkn1a [10, 25, 28, 29]. We
155 further defined the DEGs with FPKM ≥ 5 in Ebf2-EGFP+ cells, and fold change ≥ 2 at all three
156 stages as CR-specific genes. After excluding 10 subplate genes reported in previous studies [14,
157 16] and 14 CR-specific genes reported in previous studies [25], 77 are novel genes with no
158 reported association with CR neurons. Among the 77 CR-specific genes, 56 genes have valid in
159 situ data from public resources. We also noticed that there were genes showing high expression
160 levels in both Ebf2-EGFP+ cells and Ebf2-EGFP- cells although their expression levels meet the
161 fold change criteria. To obtain the genes specifically expressed in Ebf2-EGFP+ cells but not in
162 Ebf2-EGFP- cells, we added a cutoff of FPKM ≤ 15 as CR-specific genes should be expressed at
163 low level in Ebf2-EGFP- cells at these stages. We found 33 genes and confirmed that 8 with
164 specific expression in layer 1 cells (Figure 1D; Figure S1D-S1E; Additional file 1). Hence, these
165 genes could serve as CR-specific markers for future study.

166 Remarkably, the expressions of these 8 genes were regulated by overexpression or
167 knockdown of Ebf2, and Foxg1 knockdown in E11.5 cortical cell culture (Figure 1E; Figure
168 S1F-S1G). These genes included transcription factors, synapse molecules, and axon guidance
169 molecules, such as Nhlh2, Sncb, Slit1. For example, Nhlh2 was upregulated or inhibited by
170 overexpression and knockdown of Ebf2, respectively. While Sncb and Ppp2r2c responded to Ebf2
171 overexpression but not knockdown, the expression of Slit1 and Brinp1 were decreased by Ebf2
172 knockdown, suggesting they might also be regulated by molecules involved in developmental
173 pathways other than Ebf2 (Figure 1E; Figure S1F-S1G).

174 Cortical CR neurons are considered to have multiple extra-cortex origins [30]. We and others
175 have shown previously that Foxg1 suppressed CR neuron fate [27], therefore we hypothesized that
176 CR-specific genes identified by our RNA-seq analysis could be regulated by Foxg1 expression.
177 Indeed, silencing of Foxg1 using shRNA up-regulated Ebf2 and CR-specific genes (Nhlh2,

178 Ppp2r2c) in the mRNA level (Figure 1G).

179

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

181 We performed Gene Ontology analysis using the DAVID software to predict the potential function
182 of the DEGs that were enriched in the Ebf2-EGFP+ or Ebf2-EGFP- cell population, respectively
183 (Figure 2A-2D; Additional file 2). The GO terms associated with neuronal differentiation and
184 functions were highly enriched in DEGs expressed in the Ebf2-EGFP+ cell population, including
185 "synapse", "neuron projection", "neuron differentiation", "cell morphogenesis involved in
186 neuron", "axonogenesis" and "neuron migration" (Figure 2A). In contrast, the GO terms enriched
187 in the Ebf2-EGFP- cell population were mostly correlated with progenitor cell properties, such as
188 "forebrain development", "regulation of cell proliferation", "growth factor binding", "regulation
189 of neurogenesis" and "dorsal/ventral pattern formation" (Figure 2B).

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

203

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

206 The high enrichment of these GO analysis terms and KEGG pathways imply that the
207 transcriptional difference between the two cell populations is consistent with their different cell

208 type identities. The components of Ebf2-EGFP+ and Ebf2-EGFP- cell population change as
209 development proceeds. At E11.5, Ebf2-EGFP+ cells are mostly preplate neurons including both
210 subplate neurons and CR neurons, while by E15.5 Ebf2-EGFP is specifically expressed in CR
211 neurons [20]. In contrast, at E11.5 most if not all of the Ebf2-EGFP- cells are progenitor cells
212 while at E13.5 and E15.5 the Ebf2-EGFP- population also includes cortical plate neurons.
213 Therefore, the genes that are highly enriched in Ebf2-EGFP+ cells at all three stages, particularly
214 those with increased enrichment, could be specifically involved in CR neuron differentiation,
215 while the genes that are highly enriched in Ebf2-EGFP+ cells at E11.5 but not at E13.5 or E15.5
216 could be involved in regulating general neuronal differentiation processes. Hence, we further
217 compared temporal changes in global DEGs at the three different stages. Genes with expression
218 level (FPKM) in Ebf2-EGFP positive population 2 folds of that in Ebf2-EGFP negative population
219 were defined as "up-regulated". If the ratio of FPKM (Ebf2-EGFP+/ Ebf2-EGFP-) was between 2
220 and 1, these genes were defined as "unchanged". If the ratio was between 0 and 1, these genes
221 were defined as "down-regulated". Theoretically, we expected 27 possible combination of
222 expression change patterns, while our analysis showed 26 classes without the pattern of "E11up -
223 E13down - E15up". These 26 classes could be classified into 6 groups representing various
224 aspects of early neurogenesis, and 6 typical classes were shown in Figure 2G and Figure S2A
225 (Additional file 3). We hypothesized that CR-specific genes are consistently higher at the three
226 stages, while the difference in genes involved in global neuronal differentiation would be down
227 after E13.5 as the Ebf2-EGFP negative population also includes newly generated cortical plate
228 neurons. Indeed, as we expect, the genes in the group "E11up - E13up - E15up" include the
229 well-known CR marker genes, such as Reln, Lhx5, Trp73 and Ebf2, CR-specific genes, and
230 mature neuron genes related to axon regeneration like Cacna2d2 [32]. In contrast, genes in the
231 "E11down - E13down - E15down" keep higher expression in Ebf2-EGFP- cells, showing that they
232 may function on regulating NSC proliferation, such as Sox2, Pax6 and Hes5. Genes that are first
233 expressed in progenitor cells then in cortical neurons other than preplate and CR neurons may also
234 be included in this group, such as Tle1 [33]. The "E11up - E13unchange - E15down" group
235 includes subplate genes identified previously, such as Syt4, Igfbpl1, Myt1 [14, 16], reflecting that
236 subplate genes are highly expressed in the preplate on E11.5, and decreased in layer1 when the
237 preplate splits later on E13.5 and E15.5. Interestingly, we found that most genes in the "E11down

238 - E13unchange - E15up" group initially have higher expression in the cortical hem or MGE at
239 E11.5, then appear in layer 1 at E13.5 and E15.5, implying that these genes may be involved in
240 migration of CR neurons from various CR origins. For example, Zic1 and Zic2 are expressed in
241 the progenitor cells in the septum and cortical hem, and Bmp7 is expressed in the cortical hem,
242 meninges, and choroid plexus, typical CR neuron origins and location [34, 35]. The genes in the
243 "E11down - E13unchange - E15down" group include Foxg1, Bcl6 and Miat, which are expressed
244 along the lineage of neural progenitor cells [36, 37]. The "E11unchange - E13unchange -
245 E15unchange" group is enriched with genes associated with basic cellular organelles or activities
246 that required for various eukaryotic cells. For example, Atp6v0c and Atp6v0e2 in this group are
247 vacuolar ATPase and play essential roles in maintaining PH in the cellular compartments, protein
248 degradation and endocytic pathway [38, 39], while Rab11fip5 is reported to participate in protein
249 trafficking between different organelle membranes [40]. Taken together, our analysis revealed that
250 temporal gene expression dynamics reflects distinct developmental processes, allowing us to
251 identify novel factors that play specific roles in early cortical neurogenesis and CR development.
252

253 **Differentially expressed lncRNAs in the Ebf2-EGFP+ cells and Ebf2-EGFP- cells**

254 From our RNA-seq analysis, we identified 2417 long non-coding RNA transcripts in cortical cells
255 using the Cufflinks method. 200 lncRNAs were differentially expressed between the Ebf2-EGFP+
256 and the Ebf2-EGFP- population (FDR<0.05) and termed DElncRNAs (Figure 3A). 58 lncRNAs
257 were highly expressed in Ebf2-EGFP+ cells compared to Ebf2-EGFP- cells, while 64 of lncRNAs
258 showed the opposite pattern. We further filtered DElncRNAs with these criterions: 1) specifically
259 expressed in Ebf2-EGFP+ cells with FPKM at least 2-fold of that in Ebf2-EGFP negative cells
260 (Figure 3B); these lncRNAs also showed temporal specific expression patterns. 2) lncRNAs from
261 1) with FPKM ≤ 15 in Ebf2-EGFP- cells during all three stages, and FPKM ≥ 5 in Ebf2-EGFP+
262 cells at E15.5. We identified 9 lncRNAs that meet the criterions and these had never been studied
263 before in the context of neural development (Additional file 4). We called these lncRNAs as
264 ln-CR# for simplicity. The qPCR results to validate the mRNA expression level were generally in
265 agreement with the RNA-seq data (Figure 3C; Figure S3A). Notably, we detected strong
266 expression of ln-CR2 in E11.5 Ebf2-EGFP+ cells, and comparably lower or even no signals in
267 Ebf2-EGFP- cells, indicating Ebf2-EGFP+ cell-specific expression. In situ hybridization on E11.5,

268 E13.5 and E15.5 mouse embryonic brain sections revealed that ln-CR2 was selectively expressed
269 in the cortical layer next to the pial surface, as well as the TE (thalamus eminence), one of the
270 extra-cortical origins of CR neurons, indicating cell-type specificity (Figure 3D). Furthermore, we
271 performed GO analysis (see method) of DElncRNAs to predict their potential functions (Figure
272 3E). We found significant enrichments of neuronal development process-related GO terms, such
273 as "nervous system development", "neuron migration", "negative regulation of neuron apoptotic
274 process" and "neuron differentiation", neuron identities-related GO terms such as "neuron
275 projection", "axon" and "dendrite", also transcription factors regulation-related GO terms such as
276 "transcription factor complex", "transcription factor binding". The functional annotation of
277 DElncRNAs identified from our sequencing data suggests that these lncRNAs are potential
278 players in the processes of neurogenic lineage progression, neuronal maturation and transcription
279 factor regulation. Overexpression of Ebf2 altered the expression levels of some CR-specific
280 DElncRNAs but not others (Figure 3F), suggesting other regulatory factors might control their
281 expression. We also performed knockdown Foxg1, which only significantly up-regulated ln-CR5.
282 Some lncRNAs like ln-CR1 also had the trends of up regulation, while others had not, suggesting
283 that CR-specific lncRNAs might be involved in different aspects of early neurogenesis (Figure
284 S3B).
285
286 **Overview of subtypes and differentiation states of CR neurons at the single-cell level**
287 As early differentiating neurons located in the layer 1 of the embryonic cerebral cortex consist of
288 multiple cell types including CR neurons from different origins, we performed single-cell
289 RNA-seq (10X genomic Chromium) of 3000 E15.5 Ebf2-EGFP+ cells purified by FACS to
290 characterize individual gene expression profile and developmental states of early differentiating
291 neurons. After filtering out potential noises as multiplet cells and mitochondria genes, then
292 normalization and PCA analysis, the acquired gene expression data sets were selected for the
293 following analysis (Figure S4A-S4C). By K-Means clustering, Ebf2-EGFP positive cells at this
294 stage were molecularly clustered into eight major subpopulations, each showing distinct gene
295 expression pattern and with highly differentially expressed markers (Figure 4A). These eight
296 clusters can be visualized using the t-SNE plot (Figure 4B and 4C). We noticed that while Cluster
297 0, 3 and 4 had highly overlapped gene expression, Cluster 2 was the most segregated group from

298 the others. Interestingly, we found that 46% of E13.5 CR genes and 50.6% of P2 CR genes
299 identified from a previous study [25] were included in Cluster 2 (29 genes, and 44 genes,
300 respectively), indicating high enrichment of CR genes in this cluster.

301 To characterize the developmental state of the Ebf2+ neurons, we performed pseudo-time
302 reconstruction using Monocle analysis [41], which is an algorithm that cells with similar genes
303 expression pattern can be put together to identify sequential transcriptional changes in response to
304 various cell states along pseudo-time points. We chose genes that were expressed across more than
305 10 cells and their average expression levels were more than 1 for this analysis (Figure 4D; Figure
306 S4D). We found that the eight major cell subpopulations clustered from t-SNE plot corresponded
307 to the temporal cortical neuron development progress from early to late stages (Figure 4E). Genes
308 in Cluster 5 could be the first activated population during the pseudotemporal ordering,
309 sequentially followed by Cluster 3 and 7 genes, which displayed high expression levels at the
310 beginning of time period, later dropped to low levels. For example, Cluster 5 gene Neurod1 is
311 known to be expressed in committed neuronal progenitor cells at the upper SVZ and in newborn
312 neurons at the lower IZ, which plays a prominent role in cell fate specification [42, 43]. Cluster 3
313 gene Rorb is a marker of layer 4 neurons and also labels a subset of layer 5 neurons [44] (Figure
314 4E-4F). Thus, Cluster 5, 3 and 7 genes might be involved in initial neuronal differentiation. In
315 contrast, genes in Cluster 2 were expressed at low expression levels at the start side, but high
316 expression levels at the end of pseudotime. The top differential expressed genes in this cluster
317 such as Reln, Nhlh2, Lhx5 may represent a more committed CR neuron state. Genes in Cluster 7,
318 0, 6, 1, 4 displayed higher expression levels during the middle period of pseudotemporal ordering
319 than the start or the terminal. Genes in these clusters were correlated with regulation or negative
320 regulation of neuron differentiation (Tle4, Rnd3, Pou3f3) [45, 46] (Figure 4F; Figure S4E),
321 implying a function in progressive specification of neuronal cell fate. Taken together, our
322 scRNA-seq results unravel a previously unreported molecular trajectory of E15.5 CR neurons,
323 developing from general neuronal fate to committed CR neurons through suppressing other
324 cortical layer-associated genes.

325

326 **Epigenetic regulation of Ebf2-EGFP+ cell-enriched genes and lncRNAs**

327 Histone modification is a critical determinant of transcriptionally active or silent chromatin states

328 [47]. To gain a comprehensive view of histone modification in the early differentiating neurons,
329 we performed H3K27ac and H3K27me3 chromatin immunoprecipitation followed by
330 high-throughput sequencing (ChIP-seq) in Ebf2-EGFP+ cells during the early embryonic
331 development (E11.5, E13.5, E15.5) and Ebf2-EGFP- cells at E11.5 (mostly neural stem cells and
332 progenitor cells at this stage). We analyzed the occupancy of these histone modifications
333 combined with RNA-seq data on these cell types in the followed analysis. First, our data show
334 strong concordance in H3K27ac and H3K27me3 bindings across two biological replicates
335 (Figures S5A-S5B). According to a general strategy used to classify REs (regulatory elements)
336 [48-50], we classified H3K27ac and H3K27me3 bound TSS (± 5kb) by the pattern of histone
337 modification as TSS activation (H3K27ac), repression (H3K27me3) and bivalent (H3K27ac,
338 H3K27me3). Interestingly, we observed bivalent H3K27ac and H3K27me3 bindings at 58.3% of
339 TSS sites of all detected DEGs in Ebf2-EGFP- cells (0.9% had only H3K27ac bindings, 4.5% had
340 only H3K27me3 bindings). While the majority of genes (72.4% had only H3K27ac bindings,
341 0.5% had only H3K27me3 bindings, 16.4% had both bindings) in Ebf2-EGFP+ cells carried TSS
342 activating modifications with strong H3K27ac enrichments and weak H3K27me3 enrichments
343 (Figure 5A). Such observations were consistent with previous studies that bivalent histone
344 modification patterns form a poised state to be ready for initiation of further developmental events
345 in the ES cells or progenitor cells, while monovalent states occurred in the differentiated cells
346 mean more organized states in the established specific cell identity [50, 51]. As development
347 proceeded from E11.5 to E15.5, H3K27ac enrichments around TSS in Ebf2-EGFP+ cells became
348 stronger. In contrast, H3K27ac enrichments around TSS in E11.5 neural progenitor cells were
349 lowest compared to others. However, H3K27me3 bindings did not show temporal changes in
350 Ebf2-EGFP+ cells (Figure 5A; Figure S5C-S5D). These patterns suggested that the epigenetic
351 modifications were different between early differentiating neurons and neural progenitor cells and
352 dynamically changed as Ebf2-EGFP+ cells differentiate.

353 We found that in Ebf2-EGFP+ cells, H3K27ac bindings were located in intergenic (61.3%),
354 gene regions (62.2%, including 3'UTRs, 5'UTRs, Exons, Introns) and promoter regions (11.8%),
355 while H3K27me3 bindings were located in intergenic (65.6%), gene regions (43.3%, including
356 3'UTRs, 5'UTRs, Exons, Introns) and promoter regions (3.9%) (Figure 5B; Figure S5E-S5F).
357 Intergenic regions contained the highest percentage of both H3K27ac and H3K27me3 peaks,

358 likely implying potential distal regulatory elements. In addition, more peaks of H3K27ac were
359 located in promoter (11.8%) and 5'UTRs (11.0%) than H3K27me3 (promoter 3.9% and 5'UTRs
360 2.8%) in Ebf2-EGFP+ cells (Figure S5E-S5F), consistent with H3K27ac's role in transcription
361 regulation. In E11.5, genomic distribution patterns of H3K27ac and H3K27me3 peaks in
362 Ebf2-EGFP- cells were similar to that in Ebf2-EGFP+ cells (Figure S5G-S5J).

363 In the 1513 Ebf2+ genes enriched in Ebf2-EGFP+ cells, 97.2% had H3K27ac bindings and
364 63.8% had H3K27me3 bindings. 63.7% (964 genes) of them had both histone marks, 33.4% had
365 only H3K27ac bindings while 0.1% had only H3K27me3 bindings (Figure 5C). We further
366 investigated the distribution of these histone marks around the TSS and distal regions of the Ebf2+
367 genes. We found that for the H3K27ac bindings, 88% (1334) of Ebf2+ genes in Ebf2-EGFP+ cells
368 had peaks around TSS, 42% (639) had peaks around distal region (within 100kb). Interestingly,
369 the majority of Ebf2+ genes either had H3K27ac peaks only on TSS (55%, 831 genes) or on both
370 TSS and distal region (33%, 503 genes), while 9% (136) had enrichments only on distal region
371 (Figure 5D). Most CR marker genes and CR-specific expressed genes were occupied by H3K27ac
372 modification on TSS regions alone or together with distal sites. Conversely, more than half (56.3%,
373 851) of Ebf2+ genes had H3K27me3 peaks on distal regions, 18.1% (274) on TSS, 10.5% (159)
374 on both TSS and distal regions. We found that most CR marker genes and CR-specific expressed
375 genes such as Calb2, Tbr1, Ebf1, Lhx1, Lhx5, Zic5, Klf7 and Sncb, fall in the 45.8% (692) of
376 Ebf2+ genes that have H3K27me3 bindings only on distal regions. 7.6% (115) of Ebf2+ genes had
377 bindings only on TSS (Figure 5E), showing that more H3K27me3 enrichments on distal region of
378 Ebf2+ genes than on TSS. Notably, H3K27ac and H3K27me3 bindings on Ebf2+ genes
379 distributed on TSS and distal regions in E11.5 Ebf2-EGFP- cells were quite different from that in
380 Ebf2-EGFP+ cells. In E11.5 Ebf2-EGFP- cells, only 6% (90) and 6.4% (96) Ebf2+ genes had
381 H3K27ac peaks around TSS and distal region, respectively. However, there were 65.5% (989) and
382 25.8% (390) of Ebf2+ genes with H3K27me3 peaks around TSS and distal region, respectively
383 (Figure 5D-5E). Low H3K27ac and high H3K27me3 bindings around Ebf2+ genes in NPCs
384 implied that these genes were repressed, while they were activated in the Ebf2-EGFP+ cells.
385 Therefore, H3K27ac enrichments on TSS and distal sites, along with H3K27me3 enrichments on
386 distal region in Ebf2-EGFP+ cells might contribute to maintain or promote Ebf2+ genes to be
387 highly expressed in CR neurons.

388 The epigenomic data demonstrate that chromatin state round TSS is a strong indicator of RE
389 function [50]. Our results revealed that chromatin state at H3K27ac or H3K27me3-bound REs
390 were associated with differential expression of the nearest CR-specific genes. The dynamic
391 epigenomic signatures mentioned above in the regions around the genes and lncRNAs also
392 correlate with the changes in their expression levels in Ebf2-EGFP+ cells or NPC cells. For
393 example, B3gat1 and ln-CR2 expression increased in Ebf2-EGFP+ cells from E11.5 to E15.5,
394 accompanied by increasingly stronger H3K27ac enrichments at TSS in Ebf2-EGFP+ cells during
395 these stages. Meanwhile, these two genes are expressed at low levels in Ebf2-EGFP- cells on
396 E11.5, corresponding to H3K27me3 enrichments at TSS of these genes in Ebf2-EGFP- cells
397 (Figure 5F). Moreover, ln-CR2, which is increased in Ebf2-EGFP+ cells from E13.5 to E15.5,
398 exhibits a change in histone modification with no H3K27ac peaks on the TSS region on E13.5 but
399 high H3K27ac peaks on E15.5 (Figure 5F).
400
401 **Functional analysis of CR-specific lncRNAs from *In Vitro* and *In Vivo* aspects**
402 To further investigate what roles these lncRNA candidates may play during early neuronal
403 differentiation, we performed in vitro and in vivo functional assays using lentivirus-mediated
404 shRNA silencing and overexpression. We cultured single E11.5 cortical progenitor cells in
405 adherent conditions for 5 days (DIV5), then fixed the culture and stained for neuronal markers
406 β-III tubulin (immature neurons) and Reln (CR neurons) (Figure 6A). In the control culture there
407 were 29.89% Tuj1+ neurons, of which 3.16% were Reln+ (Figure 6B). Overexpression of ln-CR1,
408 ln-CR2 and ln-CR3 increased β-III tubulin+ cells and reduced total cell number (Figure 6C-6D).
409 Notably, Reln+ neurons were increased to 11.56% by ln-CR1 overexpression, 6.09% and 6.83%
410 by ln-CR2 and ln-CR3 overexpression, respectively (Figure 6B). Interestingly, the number of
411 small clones with less than 4 cells and all Tuj1+ increased in CAG-ln-CR1, CAG-ln-CR2 and
412 CAG-ln-CR3 compared to CAG-control, while the number of clones with more than 8 cells was
413 much less than that in control (Figure 6C and 6E). These results suggest that overexpression of
414 CR-specific lncRNAs promotes neuronal differentiation. Moreover, we also observed that
415 overexpression of ln-CR1 altered the morphology of Reln+ neurons, increased whole processes
416 length, branch depth and branch levels, while no significant differences were detected in neuron
417 soma size (Figure 6F).

418 During the postnatal stage, CR neurons develop more complex morphology. We further
419 analyzed the effect of ln-CR1 on cell morphology in an ex-vivo culture system using the
420 wholemount cortical hemispheres isolated from E15.5 Ebf2-EGFP+ embryos and cultured to P5.5
421 (DIV10). Compared to control, we found that overexpression of ln-CR1 increased the
422 morphological complexity of CR neurons, leading to more filopodia, spine-like structures (arrows)
423 and bouton-like structures (arrowheads) (Figure 6H; Figure S6A-S6B). We observed similar
424 results in P3.5 embryos that were ln-CR1 overexpressed by in utero electroporation into the layer
425 1 of the E15.5 cerebral cortex.

426 To uncover how overexpression of these CR-specific lncRNAs molecularly affected CR
427 neuron development, we assessed the mRNA level of several known CR neuron-related genes via
428 qPCR (Figure 6G). We found that CR marker genes including Reln, Ebf2, Calb2 and Trp73, and
429 genes that are expressed in CR neurons such as Ebf1, Ebf3, Tbr1, Nhlh2 and Lhx1, were
430 selectively altered when overexpressing CR-specific lncRNAs candidates ln-CR1, ln-CR2 and
431 ln-CR3 (Figure 6G), implying that these lncRNAs might be involved in the molecular pathways
432 for CR neuronal differentiation.

433

434 **Discussion**

435 In this study, we characterized the dynamic transcriptomic profile of early neurogenesis and
436 identified gene expression signatures associated with preplate neuronal differentiation and
437 subsequent establishment of CR cell properties. We also demonstrate that E15.5 CR cells could be
438 classified into eight subpopulations with three distinct maturation stages by single cell RNA-seq
439 analysis, which has not been reported before. Furthermore, we provide the first neuron subtype-
440 specific profile of histone modifications, revealing active, repressive or bivalent histone
441 modification states in the early differentiating neurons and neural progenitor cells that are
442 correlated with gene expression level. We have identified novel lncRNAs expressed in CR cells
443 and some of them are functionally important for CR cell differentiation. Our findings provide a
444 comprehensive molecular depiction of the early neuronal differentiation events, shedding new
445 light on the mechanism underlying preplate formation and differentiation of CR cells, the first step
446 of cortical neurogenesis.

447

448 **Molecular transition from preplate to CR neurons**

449 As cortical neurogenesis proceeds, the preplate is physically separated into the subplate and
450 the layer 1 by migrating cortical plate neurons. We first identified over 1044 coding genes and 40
451 lncRNAs with higher expression in the newly differentiated preplate neurons (Ebf2-EGFP+) at
452 E11.5 (Figure 1C; Figure 3B; Additional file 1 and Additional file 4). At E13.5, 370 genes and
453 continue to show enrichment in preplate neurons, while additional 20 genes also have higher
454 expression in subplate neurons. A subset of these genes continue to be differentially expressed in
455 the Ebf2-EGFP+ cells at E15.5, while the total DEGs at E15.5 are much less than earlier stages,
456 suggesting further specification of the CR cell fate. Indeed, we found fewer overlaps between the
457 confirmed subplate neuron-specific genes at this age [17] (28 genes overlap, 2.3% of total DEGs
458 at E15.5), confirming the molecular segregation of preplate neurons into the subplate and layer 1
459 by E15.5. The fact that the number of DEGs is the highest at E11.5 reflects the degree of the
460 difference between the two populations: At E11.5, all Ebf2-EGFP+ cells are early differentiating
461 neurons, while the Ebf2-EGFP- cells are mostly cortical progenitor cells. As the Ebf2-negative
462 population includes both cortical plate neurons and neural progenitor cells at E13.5 and E15.5,
463 genes common to neurons might not have significant changes.

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

469 Our transcriptome analysis also revealed a surprising enrichment of genes categorized in the
470 GO terms such as “neuron projection”, “neuron differentiation”, “axonogenesis”, even “synapse”
471 and “synaptic transmission” as early as E11.5. Terms such as “gate channel activity” and “ion
472 channel activity” also appear before E15.5. Interestingly, whole-cell patch-clamp recording
473 showed an inward sodium current in putative CR cells as early as E12 [52]. In addition, dynamic
474 expression of K+ channels and Ca2+ channels were also described in pre- and postnatal CR cells
475 [11, 52-54]. We found that Ebf2-EGFP-positive cells expressed neurotransmitter receptors
476 including ionotropic and metabotropic glutamate receptors, AMPA glutamate receptors, at E11.5.
477 Interestingly, it has been reported that at very early stages of neural development, subplate neurons

478 form transient synapses with axons coming from the subcortical regions to guide later formation
479 of cortico-subcortical connectivity [55, 56]. Our results suggest that populations of early
480 differentiating neurons, including Cajal-Retzius cells, are equipped with machineries suited for
481 electric or chemical synaptic connections shortly after they are born. It's possible that functional
482 connections are established earlier than currently recognized. Alternatively, expression of these
483 synapse-related proteins may not necessarily confer electrophysiological properties, but rather
484 prime early differentiating neurons to be ready for synaptic connection once they encounter with
485 other cortical plate neurons.

486 Disease-related genes are highly expressed during early differentiation period [17, 57].
487 Abnormality in the process of preplate formation and splitting are associated with neurological
488 disorders including mental retardation, epilepsy, and autism [7]. Interestingly, schizophrenia has a
489 highly susceptible stage between 6-8 weeks gestation, the time when preplate differentiation
490 occurs [58]. These findings suggest the importance of proper regulation during the early
491 differentiation process.

492

493 **Single-cell RNA-seq analysis reveals subpopulations of CR neurons**

494 CR neurons were reported to be generated from many areas of developing pallium and were
495 redistributed on the cortical surface via tangential migration. However, it is not known whether
496 CR neurons from different origins maintain their original identity or develop to the same type of
497 cells with identical gene expression profile. In the first case, we expect to identify subpopulations
498 with distinct expression level of regional genes. For example, Trp73, a gene highly expressed in
499 cortical hem-derived and TE-derived CR cells [59], would select the cells originated from the
500 cortical hem or TE. Interestingly, we found that Trp73 is only expressed in one cell subpopulation,
501 with pattern similar to other CR genes such as Ebf2, CXCR4 (Additional file 6). We also found
502 that genes that are expressed in various CR cell origin territories were enriched in the same cluster.
503 For example, Trp73 (hem), Lhx1 (Septum), Emx2 (VZ) were all included Cluster 2 (Additional
504 file 6) [59-61]. Therefore, our results revealed loss of the original regional identity of CR neurons
505 once they migrate into the cerebral cortex, suggesting that molecular and functional maturation of
506 CR neurons may follow a general path regardless of their origins.

507 Pseudo-time analysis of single cell RNA-seq data has offered a way to further dissect the

508 transitional difference in gene expression within a heterogeneous population. Here, we reported
509 that by pseudo-time analysis, CR neurons could be classified into 3 subgroups. The first group
510 consisted of activated cluster 5, 3 and 7 genes, which were highly expressed at the beginning of
511 pseudotemporal period. We found that genes in this group are associated with characteristics of
512 cells located in the upper VZ and lower cortical plate such as NeuroD1, representing transitioning
513 from neural progenitor cells to newborn neurons [42-44]. The second group is enriched with
514 cluster 7, 6, 0, 1, 4 genes that were activated during the middle of pseudotime trajectory. These
515 genes were categorized under "neural differentiation" "neurite outgrowth" and "migration" terms
516 using DAVID analysis (data not shown). The third group along the pseudo-time line consisted of
517 mainly cluster 2 genes, which includes many CR marker genes, synapse and axonogenesis-related
518 genes. Interestingly, we found that genes first expressed in neural progenitor cells then in cortical
519 plate neurons, such as Tle4 and Pou3f3, were expressed higher first, then down-regulated in the
520 third group, suggesting when cells undergo fate decisions to become CR neurons, genes related to
521 other neuronal identity would need to be suppressed at the later differentiation stage. We found
522 that Zic2, a CR-enriched gene, has binding site in the promoter of Tle4 and neuroD1, raising the
523 possibility that genes in the third group might regulate non-CR genes to allow CR-specific traits to
524 be established. Further studies will help resolve the biological relevance of these analysis-acquired
525 developmental trajectory of gene expression. Overall, our study provides a comprehensive profile
526 of gene expression for the early born neurons in the cerebral cortex at population and single cell
527 level, unravelling the molecular dynamics during early neurogenesis.
528

529 **Methods**

530 **Mouse**

531 All animal protocols used in this study were approved by the IACUC (Institutional Animal Care
532 and Use Committee) of Tsinghua University and performed in accordance with guidelines of the
533 IACUC. The laboratory animal facility has been accredited by AAALAC (Association for
534 Assessment and Accreditation of Laboratory Animal Care International). For timed mice mating,
535 the day of the vaginal plug was considered E0.5.
536

537 **Mapping and transcripts assembling**

538 Sequencing reads were mapped to mouse genome using bowtie with parameter "-v 2 -X 350 -fr -m
539 10 -S -k 1". After mapping sequencing reads to the mouse genome, we assembled the transcripts
540 using Cufflinks [62]. We assembled transcripts into known gene models or novel gene models
541 guided by known gene annotations. Known gene annotations were downloaded from Ensembl
542 (Mus_musculus.GRCm38.74). The assembled transcripts were then merged by the Cuffmerge
543 utility provided by the Cufflinks package. Then the merged transcripts were concurrently
544 annotated by Cuffcompare program in Cufflinks suite, of which known transcripts were used as
545 reference to screen for novel lncRNAs. Only transcripts with tag "intergenic" and "intronic" were
546 retained. Transcripts shorter than 200nt or with FPKM lower than 0.3 were filtered out from the
547 transcripts set.
548
549 **Clustering analysis and functional enrichment analysis**
550 K-means clustering algorithm based on Euclidean distance was used to distinguish different gene
551 expression patterns, which was visualized by Java TreeView software
552 (http://bioinformatics.oxfordjournals.org/content/20/17/3246.abstract?keytype=ref&ijkey=31ed76
553 d5EFKWk). Then, we used findGO.pl program in Homer to perform functional enrichment (GO
554 annotation, KEGG and Wiki Pathway) analysis for each gene set with different expression patterns
555 [63]. GO analysis was also performed by DAVID website (https://david.ncifcrf.gov/tools.jsp;
556 version 6.8).
557
558 **LncRNA differential expression**
559 We used DE-seq to identify differential expressed lncRNAs. As input, the DESeq package
560 expected count data as obtained from RNA-Seq in the form of a rectangular table of integer values.
561 The table cell in the i-throw and the j-th column of the table told how many reads had been
562 mapped to gene i in sample j. The count values must be raw counts of sequencing reads. As a first
563 processing step, we needed to estimate the effective library size. This step was sometimes also
564 called normalization. The effective library size information was called the size factors vector,
565 since the package only needed to know the relative library sizes. If the counts of non-differentially
566 expressed genes in one sample were, on average, twice as high as in another (because the library
567 was sequenced twice as deeply), the size factor for the first sample should be twice that of the

568 other sample. The function estimateSizeFactors estimated the size factors from the count data. The
569 inference in DESeq relied on an estimation of the typical relationship between the data's variance
570 and their mean, or, equivalently, between the data's dispersion and their mean. The dispersion
571 could be understood as the square of the coeffcient of biological variation. The variance seen
572 between counts was the sum of two components: the sample-to-sample variation just mentioned,
573 and the uncertainty in measuring a concentration by counting reads. The latter, known as shot
574 noise or Poisson noise, was the dominating noise source for lowly expressed genes. The former
575 dominated for highly expressed genes. The sum of both, shot noise and dispersion, was considered
576 in the differential expression inference. Having estimated the dispersion for each gene, it was
577 straight-forward to look for differentially expressed genes. To contrast two conditions, e.g., to see
578 whether there was differential expression between conditions “positive" and “negative", we
579 simply called the function nbinom Test. It returned a data frame with the p-values and other useful
580 information. Genes and lncRNAs with q-value lower than 0.05 were regarded as differential
581 expressed.
582
583 **Functional analysis of lncRNAs**
584 To identify the function of lncRNAs, we calculated the expression correlation of lncRNAs
585 with coding genes. We assumed that co-expressed genes were more likely to be correlated in
586 function. Based on expression profiles in lncRNAs and protein coding genes, we calculated the
587 correlation between them. FPKM of genes and lncRNAs from RNA-seq data sets were used to
588 calculate the correlation matrix. Before calculating the correlation matrix, we filtered out some
589 low-expressed genes and lncRNAs for the accuracy of further characterization. For protein coding
590 genes with FPKM ranked in top 75% in at least one data set were retained. For lncRNAs, the
591 filtering processes were as follows: (1) expressed (FPKM>0) in at least 1 condition, (2) FPKM
592 ranked in top 75% in at least one datasets, (3) co-efficient of variation (cv) of FPKM ranked in top
593 75% in all ncRNA candidates. Cv represented for coefficient of variance, the standard deviation
594 divided by mean. Cv showed the extent of variability in relation to mean of the population.
595 Correlations between gene-gene and gene-lncRNA were calculated. We chose to calculate
596 Pearson correlation in the following analysis. The Pearson correlation coefficient was a measure
597 of the linear correlation between two variables X and Y, giving a value between +1 and −1

598 inclusive, where 1 was total positive correlation, 0 was no correlation, and −1 was negative
599 correlation. It was widely used in the sciences as a measure of the degree of linear dependence
600 between two variables. Pearson's correlation coefficient between two variables was defined as the
601 covariance of the two variables divided by the product of their standard deviations.

602 $$r = \frac{\sum (X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum (X - \overline{X})^2 (Y - \overline{Y})^2}}$$

603 The calculation of correlation was done by SciPy, a fundamental python library for scientific
604 computing. The two-sided p-value for a hypothetical test whose null hypothesis was that two sets
605 of data were uncorrelated. The p-value roughly indicated the probability of an uncorrelated system
606 producing datasets that had a Pearson correlation at least as extreme as the one computed from
607 these datasets.

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

614

615 **Quantitative RT-PCR**

616 Total RNA was isolated from purified Ebf2-EGFP positive or negative cells, extracted using Trizol
617 reagent (Cat#15596018, Life Technologies), and reverse transcribed. qPCR was performed using
618 PrimeScript RT regent Kit (Perfect Real Time, Cat#DRR047A, Takara). Gene expression levels
619 were detected using a BioRad CFX384 Touch™ real-time PCR detection system (Bio-Rad
620 Laboratories) and determined with ΔΔCt method, normalized to an internal control GAPDH
621 mRNA for quantification. All reactions were replicated 3 times. qRT-PCR primers for mouse gene
622 sequence were included in the Additional file 7.

623

624 **Tissue collection and RNA In Situ hybridization**

625 Mouse brains at various developmental stages were collected and fixed in 4% paraformaldehyde

626 (PFA) at 4°C for 4h. After briefly rinsing in PBS, brains were dehydrated in 30% sucrose in PBS
627 at 4°C until the time they sank, and embedded in OCT (Sakura Finetek, Torrance, CA). Frozen
628 brains were cryosectioned to 16 μm and mounted to coverslip. In situ hybridization was performed
629 according to instructions as previously described (Wigard et al., 2006; Gregor et al., 2007; Robert
630 et al., 2007). Digoxigenin-labeled RNA probes were designed based on lncRNA sequences
631 determined by Cufflinks and Ensembl. Images were taken on a Zeiss Slide Scanner (Axio Scan.
632 Z1) fluorescent microscope.
633
634 **In utero electroporation of cortical layer 1**
635 The E15.5 embryonic cortical layer 1 were electroporated as described previously [64]. Briefly,
636 embryonic day 15.5 (E15.5) mice were anesthetized and placed on a heating pad. Then the uterus
637 were carefully pulled out and injected with proper volume of plasmid mix (0.5-1μl/embryo, 2-3
638 μg/μl) into the gap under the skull and above the pial surface with a glass capillary. The
639 Electroporator (CUY21VIVO-SQ; BEX) was set to five 50ms, 36V pulses with an interval of
640 999ms, and embryos' heads were electroporated with a 5-mm electrode (LF650P5; BEX). After
641 that, the embryos were placed back into their original location of abdominal cavity, and the
642 abdomen wall and skin were closed. Then waited for a while until the mice recovered from
643 surgery. For phenotype observation, the mice progenies were sacrificed on postnatal day 3.5 or 5.5
644 (P3.5 or P5.5), and both Ebf2-EGFP and RFP signals positive pups were selected for the following
645 PFA fixation dealing and immunohistochemistry. Confocal images were taken using LSM780
646 (Zeiss) Laser Scanning System on a Zeiss microscope. At least three independent experiments,
647 three animals each experiment were analyzed for each condition.
648
649 **Wholemount culture neuronal processes analysis**
650 For neuronal processes and spine analysis, Ebf2-EGFP+ wholemounts of E15.5 cerebral cortex
651 were infected by lenti-virus CAG-RFP, H1-RFP, or CAG-ln-CR1, CAG-ln-CR2, CAG-ln-CR3.
652 Next, wholemounts were cultured for 12 days in a 35℃ bioincubator. The culture medium was
653 replaced every 3 days. After culture, wholemounts were fixed and stained with chicken anti-GFP
654 (GFP-1010, Aves lab lnc.) and rat anti-chicken-Alexa488 (Jackson Immuno Research lab).
655 Confocal microscopy was performed with a LSM780 (Zeiss) Laser Scanning System on a Zeiss

656 microscope, and wholemount tissues were analyzed by taking 1μm serial optical sections with
657 ZEN_2011_Lite_x64 software (Zeiss). Images were acquired using a 1024x1024 scan format with
658 a 63x objective.
659
660 **Immunocytochemistry**
661 Cell culture media were removed and cells were fixed in 4% (w/v) PFA for 30min at RT, rinsed by
662 1×PBS three times. Fixed cells were blocked in 5% (v/v) BSA with 0.3% (v/v) Triton X-100 in
663 1×PBS for 1h and rinsed by 1×PBS three times. Immunostainings were performed for 4h at RT or
664 overnight at 4°C with rotation by incubating with primary antibodies, then washed by 1×PBS
665 three times. Cells were incubated with secondary antibodies and DAPI diluted in 1×PBS for 1h in
666 the dark, followed by washing with 1×PBS three times. Images were taken on Zeiss fluorescent
667 microscopy (AxioVision). Primary antibodies used: GFP (GFP-1010, Aves lab lnc.), Reln
668 (MAB5364, Millipore), Tuj1 (T8660, Sigma).
669
670 **Chromatin-immunoprecipitation and sequencing (ChIP-Seq)**
671 ChIP assays were performed using mice primary Ebf2-EGFP positive cells at E11.5, E13.5, E15.5
672 and Ebf2-EGFP negative cells at E11.5, with antibody against H3K27ac (Ab4729, Abcam) and
673 H3K27me3 (Ab6002, Abcam), experimental procedures were as previously described [65]. Briefly,
674 Ebf2-EGFP positive and negative cells (~1-5 million cells) were sorted and collected from FACS
675 (BD Influx^{TM}), and fixed by 0.4% (v/v) formaldehyde at room temperature (RT) for 10 min. The
676 reaction was stopped by adding 0.125 M glycine for 5 min. After rinsing cells three times with
677 cold PBS, the nuclei were resuspended in Nuclei Lysis Buffer (50 mM Tri-Cl, pH8.1, 10 mM
678 EDTA, 1% SDS, 1 mM DTT, 1 mM PMSF and Sigma protease inhibitor cocktail, RT) and
679 incubated on ice for 10 min. Nuclear suspension were sonicated 6” on, 15” off, 20 cycles (power
680 set at 25%) on ice. The fragmented chromatin was diluted for 5 folds in ChIP Dilution Buffer
681 (16.7 mM Tri-Cl, pH8.1, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 1 mM PMSF, and
682 Sigma protease inhibitor cocktail). Then 5 ug IP antibody was added and incubated overnight at
683 4 °C. Chromatin-protein complex was pulled down with Dynabeads protein G (Cat#10004D,
684 Invitrogen) for 2h at 4°C with rotation and followed by 5 sequential washes as the following: 1.
685 low salt wash buffer (20 mM Tri-Cl, pH8.1, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1%

686 SDS, 0.1% Na-deoxycholate); 2. high salt wash buffer (20 mM Tri-Cl, pH8.1, 500 mM NaCl, 1
687 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate); 3. LiCl wash buffer (10 mM
688 Tri-Cl, pH8.1, 250 mM LiCl, 1 mM EDTA, 0.5% NP-40, 0.5% Deoxycholic acid (Na salt)); and
689 the last two washes with TE buffer (10 mM Tri-Cl, pH8.1, 1 mM EDTA). Immunoprecipitated
690 complex was eluted from the antibody in elution buffer (1% SDS, 50 mM TrisCl (pH 8.1), 1 mM
691 EDTA) by votexing once at 65°C for 20min and once at RT. Eluted chromatins were incubated at
692 65°C overnight for reverse crosslinks and chromatin fragmented purification with PCR column
693 purification. ChIP-Seq libraries were prepared using NEB Next Ultra™ II ChIP-Seq Library Prep
694 Kit (Cat#E7645S & Cat#7335S, NEB) according to the manufacturer's instructions. ChIP-seq data
695 were aligned to mouse genome mm10, performing Bowtie v. 1.1.2 on the Galaxy platform with
696 the following options: -v 3 –M 1 -5 15. Resulting output SAM files were converted to BAM
697 format, and finally converted to BED files. Peaks were called using MACS (Model-Based
698 Analysis of ChIP-Seq), with a p-value cutting off of 1e-5.
699

700 **Acknowledgments**

701 We thank Dr. Zhengang Yang and Dr. Chunyang Wang at Fudan University Collaborative
702 Innovation Center for Brain Science for their assistance of In Situ hybridization experiments in
703 this study. We thank Yuchuan Wang, Chao Di and Xinqiang Ding for their help of bioinformatic
704 analysis. We thank for all members of Shen lab for their helpful discussion during the preparation
705 of manuscript. We are also grateful to Center of Biomedical Analysis at Tsinghua University.
706

707 **Funding**

708 This work was supported by grants from the National Natural Science Foundation of China
709 (31170247 to Q.S.).
710

711 **Availability of data and materials**

712 The sequencing data generated in this study will be deposited in GEO repository before
713 manuscript accepted.
714

715 **Author's contribution**

716 JL and QS designed the experiments. JL, LS, X-L P and S-J Q performed the experiments, JL and
717 X-M Y analyzed sequencing data. JL and QS wrote the manuscript. QS supervised the project.

718

719 **Ethics approval and consent to participate**

720 The study was approved by the IACUC (Institutional Animal Care and Use Committee) of
721 Tsinghua University. All animal experiments were performed in accordance with their guidelines.

722

723 **Consent for publication**

724 The authors declare that they have no competing interests.

725

726 **Competing interests**

727 Not applicable.

728

729 **Figure Legends**

730 **Fig 1. Transcriptome sequencing of Ebf2-EGFP+ and Ebf2-EGFP- cells in the early**
731 **embryonic stage**

732 (A) FACS purification of Ebf2-EGFP+ and Ebf2-EGFP- cells from early embryonic stages, E11.5,
733 E13.5 and E15.5 respectively. The percentage of Ebf2-EGFP+ cells at each stage was calculated
734 and shown in the graph.

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

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

739 (D) Expression of selected CR-specific coding genes are found to be specifically expressed in
740 cortical layer 1 of E14.5 mouse embryos (In situ hybridization images from Eurexpress Anatomy
741 Atlas and GenePaint.org).

742 (E) qRT-PCR analysis comparing the expression levels of CR-specific genes in NSC culture assay
743 after virus infection with control (H1) or Ebf2-overexpression. Data represent mean ± SEM (n=3
744 independent experiments, *P<0.05, T test).

745 **Fig 2. Functional prediction of DEGs in the Ebf2-EGFP+ cells and Ebf2-EGFP- cells**

746 (A) Enriched functional assessments (including biological processes, cellular components and
747 molecular function) of highly expressed DEGs in Ebf2-EGFP+ or Ebf2-EGFP- cell population
748 (fold change ⩾ 2).

749 (B-D) Enriched functional assessments of highly expressed DEGs in E11.5, E13.5 or E15.5
750 Ebf2-EGFP+ cell population, respectively.

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

753 (G) Temporal dynamic expression pattern of DEGs at the three different stages (y axis = Fold
754 Change between Ebf2-EGFP+ cells and Ebf2-EGFP- cells). Noted the dynamic pattern is related
755 to functions of associated genes.

756 **Fig 3. Functional prediction of differentially expressed lncRNAs in the Ebf2-EGFP+ and**
757 **Ebf2-EGFP- cells**

758 (A) Heatmap representing all differential expressed lncRNAs (FDR<0.05, p-value<0.05) in
759 Ebf2-EGFP+ & Ebf2-EGFP- cell samples at the three embryonic stages.

760 (B) Venn diagram displaying the extent of overlapping between DElncRNAs at the three
761 embryonic stages.

762 (C) The expression pattern of DElncRNAs between Ebf2-EGFP+ or Ebf2-EGFP- cells is
763 consistent revealed by RNA-seq and qRT-PCR.

764 (D) In situ hybridization showing the endogenous expression level of CR-specific lncRNA ln-CR2
765 at E11.5, E13.5 and E15.5. Regions with specific expression are labeled and higher magnification
766 images of boxed areas are shown below. Abbreviations: DP, dorsal pallium; VP, ventral pallium;
767 MP, medial pallium; Th, thalamus; Str, striatum; TE, thalamus eminence.

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

771 (F) qRT-PCR analysis comparing the expression levels of lncRNAs in NSC culture assay after
772 virus infection with control (H1) or Ebf2-overexpression. Data represent mean ± SEM (n=4
773 independent experiments, *P<0.05, T test).

774 **Fig 4. Single cell overview of Cajal-Retzius neuron subtypes**

775 (A) Heatmap of differentially expressed genes in E15.5 Ebf2-EGFP+ cells sub-clusters.

776 (B) The K-means algorithm-based plot showing eight major populations.

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

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

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

785 **(F)** Expression level dynamics across pseudotime ordering of Lhx5 (in Cluster 2) and Neurod1
786 (in Cluster 5) are shown.

787 **Fig 5. Genomic characterization of histone modification in the early differentiating neurons**
788 **and NPCs**

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

792 (B) Genomic features of H3K27ac and H3K27me3-bound REs in Ebf2-EGFP+ cells.
793 (C) Comparison of genes number with H3K27ac and H3K27me3 bindings in Ebf2-EGFP+ cells.
794 (D) Proportion of Ebf2-EGFP+ cell-enriched genes with TSS and distal H3K27ac bindings in
795 Ebf2-EGFP+ cells (upper) and NPCs (lower).

796 (E) Proportion of Ebf2+ genes with TSS and distal H3K27me3 bindings in Ebf2-EGFP+ cells

797 (upper) and NPCs (lower).

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

801 **Fig 6. Functional validation of CR-specific lncRNAs ln-CR1, ln-CR2 and ln-CR3 *In Vitro***
802 **and *In Vivo***

803 (A) Immunostaining of NSCs treated with control (CAG-control) or overexpression lentiviruses
804 for CR-specific lncRNAs ln-CR1 (CAG-ln-CR1), ln-CR2 (CAG-ln-CR2) and ln-CR3
805 (CAG-ln-CR3) after in vitro culture for 5 days. Reln (red), Tuj1 (cyan), or DAPI (blue) for CR
806 neurons, neurons, and nucleus, respectively. Scale bar, left 3 rows 150 μm, right 50 μm.

807 (B-E) Overexpression of CR-specific lncRNAs significantly increased CR neuron number,
808 promoted NSC differentiation and generated more neuron-like small cell clones. All the
809 experiments were performed 4 times and 8 randomly selected microscopy fields were counted
810 each time. Data represented mean ± SEM. p values indicated were calculated by Student's t test
811 (unpaired). **$p<0.01$, *$p<0.05$.

812 (F) Quantification of process branch depth, branch levels and total dendrites length and soma size
813 of CR neurons after overexpression of CR-specific lncRNAs compared to control. All the
814 experiments were performed 3 times and 8 randomly selected microscopy fields were counted
815 each time. Data represented mean ± SEM. p values indicated were calculated by Student's t test
816 (unpaired). **$p<0.01$, *$p<0.05$.

817 (G) Quantification of CR neuron molecular markers as Reln, Ebf2, and Calb2 at 72h after
818 lentiviral transduction in the NSC culture assay comparing the effect of overexpression of
819 CR-specific lncRNAs to control. Data represented mean ± SEM (**$p<0.01$, *$p<0.05$, $n>3$,
820 Student's T test).

821 (H) Comparison of dendritic spine, filopodia and processes in cortical wholemounts
822 overexpressing ln-CR1 (CAG-ln-CR1) to CAG-control. Scale bar, 20 μm. ln-CR1 overexpression

823 resulted in increased in spine and filopodia (arrows) as compared to CAG-control transfected CR
824 neurons. Cortical wholemounts were collected from P3 pups after *in utero* electroporation to
825 cortex layer 1 on E15.5.

826

827 **Additional files**

828 **Fig S1. RNA-seq quality control, related to Fig 1**

829 (A) Phase and fluorescent images of Ebf2-EGFP+ cells and Ebf2-EGFP- cells after FACS at Day
830 0 and Day 4 in vitro culture showing the purity and growth of sorted and unsorted cells (scale bar,
831 25 μm).

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

835 (C) qRT-PCR and endogenous expression patterns (In Situ Hybridization data from Allen Brain
836 Atlas) to validate the RNA-seq results at different embryonic stages. CR neuron molecular
837 markers as Calb2, Ebf2, Tbr1, and CR neuron related genes as Ebf1, Ebf3 and Foxg1 were
838 selected for validation.

839 (D) qRT-PCR and endogenous expression patterns (In Situ Hybridization data from Allen Brain
840 Atlas) to validate the RNA-seq results of Nhlh2 at three embryonic stages.

841 (E) Expression of the CR-specific coding gene Dnm3 is found to be specifically expressed in layer
842 1 of E14.5 mouse brain (In Situ Hybridization data from Eurexpress Anatomy Atlas).

843 (F-G) qRT-PCR analysis of CR-specific genes in NSCs culture assay after lentiviral transduction
844 of control (H1) or Ebf2-knockdown, or Foxg1-knockdown. Data represent mean ± SEM (n=3
845 independent experiments, *P<0.05, T test).

846 **Fig S2. Annotation of temporal dynamic genes, related to Fig 2**

847 (A) Gene Ontology analysis of selected genes with distinct temporal dynamic expression pattern.

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

849 (A) qRT-PCR analysis to validate the RNA-seq results at different embryonic stages. CR-specific
850 lncRNAs ln-CR6 and Meg3 were selected for test.

851 (B) qRT-PCR analysis of CR-specific lncRNAs in NSCs culture assay after lentiviral transduction
852 of control (H1) or Foxg1-knockdown. Data represent mean ± SEM (n=4 independent experiments,
853 *P<0.05, T test).

854 **Fig S4. Quality control and pseudotime reconstruction analysis of single cell RNA-seq,**
855 **related to Fig 4**

856 (A) Violin plots show genes, UMIs and mitochondrion ratio in scRNA-seq. Left, violin plot
857 shows detected gene number in all sample cells. X-axis is sample name and Y-axis is gene
858 numbers detected in a single cell; Middle, X-axis is the same with left and Y-axis is UMIs detected
859 in a single cell; Right, Y-axis is mitochondrion numbers detected in a single cell.

860 (B) Scatter diagram of mitochondrion ratio, gene numbers compared to UMI numbers,
861 respectively. X-axis is UMI numbers, Y-axis is mitochondrion ratio detected in a single cell in the
862 left, Y-axis is gene numbers detected in a single cell in the right.

863 (C) Scatter diagram of all detected genes dispersion. X-axis is average expression of a gene in all
864 cells and Y-axis is dispersion value of that gene.

865 (D) Pseudotime trajectories of E15.5 Ebf2-EGFP+ cells reconstructed by Monocle. Pseudotime
866 points are indicated by colors and cell clusters are indicated by numbers in the black circle.

867 (E) Expression level dynamics of Reln (in Cluster 2) and Tle4 (in Cluster 4) across pseudotime
868 ordering.

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

870 (A-B) Average profile of H3K27ac (A) or H3K27me3 (B) binding at called ChIP-seq peaks
871 comparing the two biological replicates.

872 (C-D) A profile plot showing normalized H3K27ac (C) or H3K27me3 (D) enrichment around TSS
873 of all genes at the three embryonic stages.

874 (E-J) Quantitative data showing genomic features of H3K27ac and H3K27me3-bound REs in

875 Ebf2-EGFP+ cells (E, F), E11.5 NPCs (G, H) and E11.5 Ebf2-EGFP+ cells (I, J).

876 **Fig S6. LncRNAs function in CR neurons in ex vivo wholemount culture assay, related to Fig**
877 **6**

878 (A) Comparison of dendritic spine, filopodia and processes of cultured cortical wholemounts
879 overexpressing ln-CR1 (CAG-ln-CR1), ln-CR2 (CAG-ln-CR2), ln-CR3 (CAG-ln-CR3) and
880 knockdown ln-CR2 (shRNA-ln-CR2) to wholemounts expressing CAG-control. Scale bar, 15 μm.

881 (B) Quantification of CR neuron processes branch depth, branch levels and total dendrites length
882 after overexpression of CR-specific lncRNAs. All the experiments were performed 3 times and 4
883 randomly selected microscopy fields were counted each time. Data represent mean ± SEM. p
884 values indicated were calculated by Student's t test (unpaired). **p<0.01, *p<0.05.
885
886
887 **References**
888

889 1. Caviness VS. Neocortical Histogenesis in Normal and Reeler Mice: A Developmental Study
890 Based Upon [3H]Thymidine Autoradiography. Developmental Brain Research. 1982;4:293-302.
891 2. Marin-Padilla M. Ontogenesis of the Pyramidal Cell of the Mammalian Neocortex and
892 Developmental Cytoarchitectonics: A Unifying Theory. J Comp Neurol. 1992;321:321-3.
893 3. Hevner RF, Neogi T, Englund C, Daza RM, Fink A. Cajal–Retzius cells in the mouse:
894 transcription factors, neurotransmitters, and birthdays suggest a pallial origin. Developmental Brain
895 Research. 2003;141:39-53.
896 4. Bayer SA, Altman J. Development of the Endopiriform Nucleus and the Claustrum in the Rat
897 Brain. Neuroscience. 1991;45:391-412.
898 5. Takahashi T, Goto T, Miyama S, Nowakowski RS, Caviness VS Jr. Sequence of Neuron Origin
899 and Neocortical Laminar Fate: Relation to Cell Cycle of Origin in the Developing Murine Cerebral
900 Wall. J Neurosci. 1999;19:10357-71.
901 6. Bystron I, Blakemore C, Rakic P. Development of the human cerebral cortex: Boulder Committee
902 revisited. Nat Rev Neurosci. 2008;9:110-22.
903 7. Olson EC. Analysis of preplate splitting and early cortical development illuminates the biology of
904 neurological disease. Front Pediatr. 2014;2:1-9.
905 8. Ghosh A, Antonini A, McConnell SK, Shatz CJ. Requirement of Subplate Neurons in the
906 Formation of Thalamocortical Connections. Nature. 1990;347:179-181.
907 9. Molnar Z, Adams R, Blakemore C. Mechanisms Underlying the Early Establishment of
908 Thalamocortical Connections in the Rat. J Neurosci. 1998;18:5723-45.
909 10. Del Rio JA, Martinez A, Fonseca M, Auladell C, Soriano E. Glutamate-like Immunoreactivity and
910 Fate of Cajal-Retzius Cells in the Murine Cortex as Identified with Calretinin Antibody. Cereb Cortex.
911 1995;5:13-21.
912 11. Kirischuk S, Luhmann HJ, Kilb W. Cajal-Retzius cells: update on structural and functional

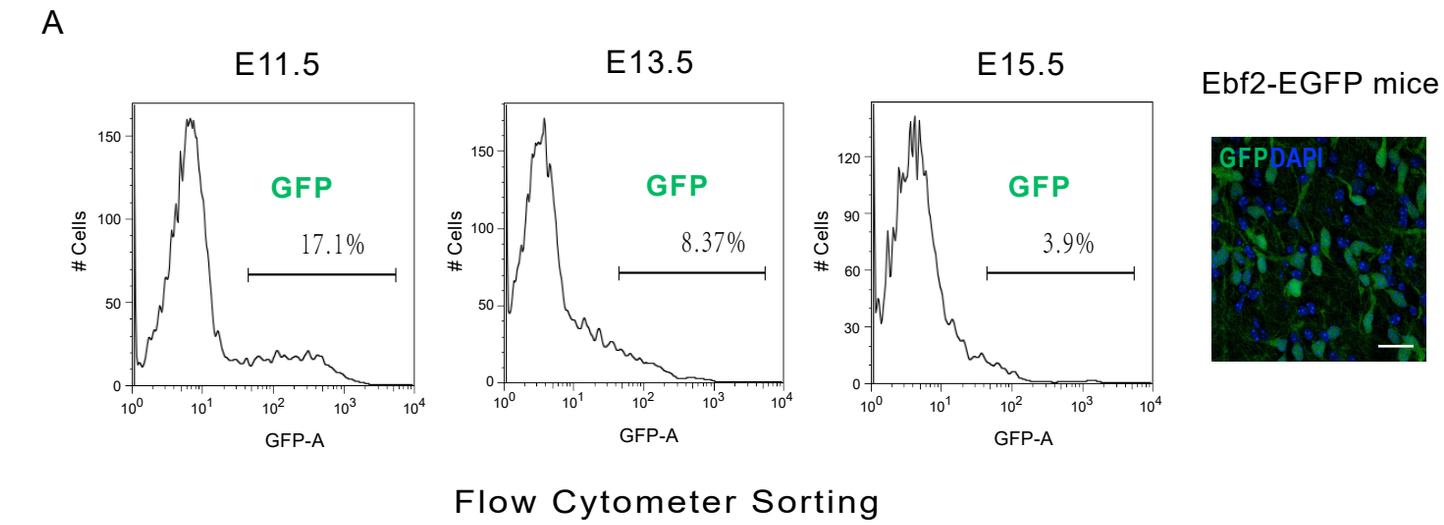
913 properties of these mystic neurons that bridged the 20th century. Neuroscience. 2014;275:33-46.
914 12. Hong SE, Shugart YY, Huang DT, Shahwan SA, Grant PE, Hourihane JOB, Martin NDT, Walsh
915 CA. Autosomal recessive lissencephaly with cerebellar hypoplasia is associated with human RELN
916 mutations. Nature Genetics. 2000;26:93-6.
917 13. Nakagawa N, Yagi H, Kato K, Takematsu H, Oka S. Ectopic clustering of Cajal-Retzius and
918 subplate cells is an initial pathological feature in Pomgnt2-knockout mice, a model of
919 dystroglycanopathy. Sci Rep. 2015;5:11163.
920 14. Oeschger FM, Wang WZ, Lee S, Garcia-Moreno F, Goffinet AM, Arbones ML, Rakic S, Molnar
921 Z. Gene expression analysis of the embryonic subplate. Cereb Cortex. 2012;22:1343-59.
922 15. Osheroff H, Hatten ME. Gene expression profiling of preplate neurons destined for the subplate:
923 genes involved in transcription, axon extension, neurotransmitter regulation, steroid hormone signaling,
924 and neuronal survival. Cereb Cortex. 2009;19 Suppl 1:i126-34.
925 16. Hoerder-Suabedissen A, Wang WZ, Lee S, Davies KE, Goffinet AM, Rakic S, Parnavelas J, Reim
926 K, Nicolic M, Paulsen O, Molnar Z. Novel markers reveal subpopulations of subplate neurons in the
927 murine cerebral cortex. Cereb Cortex. 2009;19:1738-50.
928 17. Hoerder-Suabedissen A, Oeschger FM, Krishnan ML, Belgard TG, Wang WZ, Lee S, Webber C,
929 Petretto E, Edwards AD, Molnar Z. Expression profiling of mouse subplate reveals a dynamic gene
930 network and disease association with autism and schizophrenia. Proc Natl Acad Sci U S A.
931 2013;110:3555-60.
932 18. Belgard TG, Marques AC, Oliver PL, Abaan HO, Sirey TM, Hoerder-Suabedissen A,
933 Garcia-Moreno F, Molnar Z, Margulies EH, Ponting CP. A transcriptomic atlas of mouse neocortical
934 layers. Neuron. 2011;71:605-16.
935 19. Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A,
936 Leblanc G, Hatten ME, Heintz N. A gene expression atlas of the central nervous system based on
937 bacterial artificial chromosomes. Nature. 2003;425:917-925.
938 20. Chuang SM, Wang Y, Wang Q, Liu KM, Shen Q. Ebf2 marks early cortical neurogenesis and
939 regulates the generation of cajal-retzius neurons in the developing cerebral cortex. Dev Neurosci.
940 2011;33:479-93.
941 21. Chiara F, Badaloni A, Croci L, Yeh ML, Cariboni A, Hoerder-Suabedissen A, Consalez GG,
942 Eickholt B, Shimogori T, Parnavelas JG, Rakic S. Early B-cell factors 2 and 3 (EBF2/3) regulate early
943 migration of Cajal-Retzius cells from the cortical hem. Dev Biol. 2012;365:277-89.
944 22. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki
945 N, Martersteck EM, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using
946 Nanoliter Droplets. Cell. 2015;161:1202-1214.
947 23. Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lönnerberg P, La Manno G, Juréus A, Marques S,
948 Munguba H, He LQ, Hjerling-Leffler J, Linnarsson S. Cell types in the mouse cortex and hippocampus
949 revealed by single-cell RNA-seq. Science. 2015;347:1138-42.
950 24. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, Enikolopov G, Nauen DW, Christian
951 KM, Ming GL, Song H. Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying
952 Adult Neurogenesis. Cell Stem Cell. 2015;17:360-72.
953 25. Yamazaki H, Sekiguchi M, Takamatsu M, Tanabe Y, Nakanishi S. Distinct ontogenic and regional
954 expressions of newly identified Cajal–Retzius cell-specific genes during neocorticogenesis. PNAS.
955 2004;101:14509-14.
956 26. Hanashima C, Li SC, Shen L, Lai E, Fishell G. Foxg1 Suppresses Early Cortical Cell Fate.

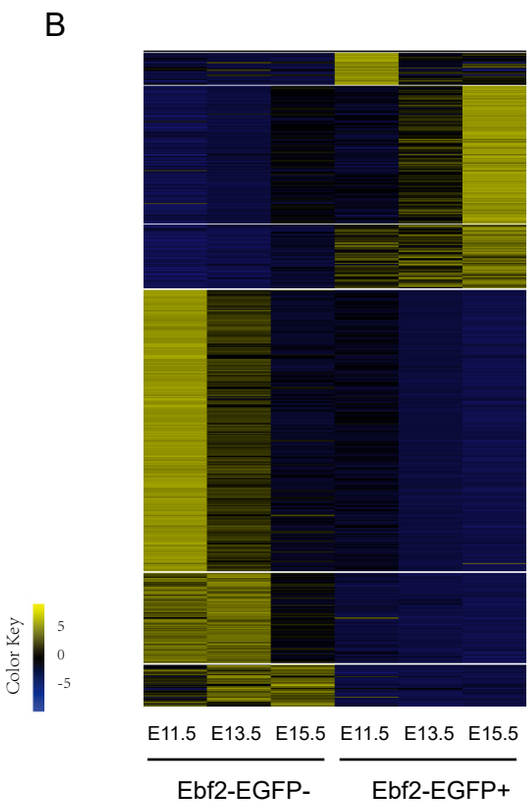
957 Science. 2004;303:56-59.
958 27. Shen Q, Wang Y, Dimos JT, Fasano CA, Phoenix TN, Lemischka IR, Ivanova NB, Stifani S,
959 Morrisey EE, Temple S. The timing of cortical neurogenesis is encoded within lineages of individual
960 progenitor cells. Nat Neurosci. 2006;9:743-51.
961 28. Hevner RF, Shi L, Justice N, Hsueh YP, Sheng M, Smiga S, Bulfone A, Goffinet AM,
962 Campagnoni AT, Rubenstein JL. Tbr1 Regulates Differentiation of the Preplate and Layer 6. Neuron.
963 2001;29:353-366.
964 29. Siegenthaler JA, Miller MW. Generation of Cajal-Retzius neurons in mouse forebrain is regulated
965 by transforming growth factor beta-Fox signaling pathways. Dev Biol. 2008;313:35-46.
966 30. Meyer G. Building a human cortex: the evolutionary differentiation of Cajal-Retzius cells and the
967 cortical hem. J Anat. 2010;217:334-43.
968 31. Corradi A, Croci L, Broccoli V, Zecchini S, Previtali S, Wurst W, Amadio S, Maggi R, Quattrini A,
969 Consalez GG. Hypogonadotropic hypogonadism and peripheral neuropathy in Ebf2-null mice.
970 Development. 2003;130:401-410.
971 32. Tedeschi A, Dupraz S, Laskowski CJ, Xue J, Ulas T, Beyer M, Schultze JL, Bradke F. The
972 Calcium Channel Subunit Alpha2delta2 Suppresses Axon Regeneration in the Adult CNS. Neuron.
973 2016;92:419-434.
974 33. Buscarlet M, Perin A, Laing A, Brickman JM, Stifani S. Inhibition of Cortical Neuron
975 Differentiation by Groucho/TLE1 Requires Interaction with WRPW, but Not Eh1, Repressor Peptides.
976 Journal of Biological Chemistry. 2008;283:24881-8.
977 34. Inoue T, Ogawa M, Mikoshiba K, Aruga J. Zic deficiency in the cortical marginal zone and
978 meninges results in cortical lamination defects resembling those in type II lissencephaly. J Neurosci.
979 2008;28:4712-25.
980 35. Segklia A, Seuntjens E, Elkouris M, Tsalavos S, Stappers E, Mitsiadis TA, Huylebroeck D,
981 Remboutsika E, Graf D. Bmp7 regulates the survival, proliferation, and neurogenic properties of neural
982 progenitor cells during corticogenesis in the mouse. PLoS One. 2012;7:e34088.
983 36. Tiberi L, van den Ameele J, Dimidschstein J, Piccirilli J, Gall D, Herpoel A, Bilheu A, Bonnefont
984 J, Iacovino M, Kyba M, et al. BCL6 controls neurogenesis through Sirt1-dependent epigenetic
985 repression of selective Notch targets. Nat Neurosci. 2012;15:1627-35.
986 37. Aprea J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, Zocher S, Massalini S,
987 Alexopoulou D, Lesche M, et al. Transcriptome sequencing during mouse brain development identifies
988 long non-coding RNAs functionally involved in neurogenic commitment. EMBO J. 2013;32:3145-60.
989 38. Nishi T, Forgac M. The vacuolar (H+)-ATPases--nature's most versatile proton pumps. Nat Rev
990 Mol Cell Biol. 2002;3:94-103.
991 39. Ueda T, Ugawa S, Shimada S. A novel putative M9.2 isoform of V-ATPase expressed in the
992 nervous system. Neuroreport. 2003;14:25-30.
993 40. Boal F, Hodgson LR, Reed SE, Yarwood SE, Just VJ, Stephens DJ, McCaffrey MW, Tavare JM.
994 Insulin promotes Rip11 accumulation at the plasma membrane by inhibiting a dynamin- and
995 PI3-kinase-dependent, but Akt-independent, internalisation event. Cell Signal. 2016;28:74-82.
996 41. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ,
997 Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by
998 pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32:381-386.
999 42. Hevner RF, Hodge RD, Daza RA, Englund C. Transcription factors in glutamatergic neurogenesis:
1000 conserved programs in neocortex, cerebellum, and adult hippocampus. Neurosci Res. 2006;55:223-33.

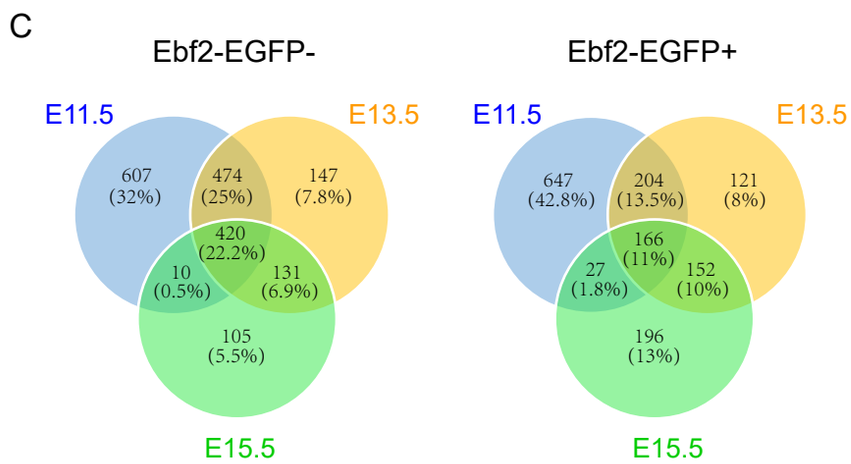
1001 43. Lee JE, Hollenberg SM, Snider L, Turner DL, Lipnick N, Weintraub H. Conversion of Xenopus
1002 Ectoderm into Neurons by NeuroD, a Basic Helix-Loop-Helix Protein. Science. 1995;268:836-44.
1003 44. Schaeren-Wiemers N, Andre E, Kapfhammer JP, Becker-Andre M. The Expression Pattern of the
1004 Orphan Nuclear Receptor RORb in the Developing and Adult Rat Nervous System Suggests a Role in
1005 the Processing of Sensory Information and in Circadian Rhythm. Eur J Neurosci. 1997;9:2687-701.
1006 45. Pacary E, Heng J, Azzarelli R, Riou P, Castro D, Lebel-Potter M, Parras C, Bell DM, Ridley AJ,
1007 Parsons M, Guillemot F. Proneural transcription factors regulate different steps of cortical neuron
1008 migration through Rnd-mediated inhibition of RhoA signaling. Neuron. 2011;69:1069-84.
1009 46. Oishi K, Aramaki M, Nakajima K. Mutually repressive interaction between Brn1/2 and Rorb
1010 contributes to the establishment of neocortical layer 2/3 and layer 4. Proc Natl Acad Sci U S A.
1011 2016;113:3371-6.
1012 47. Jenuwein T, Allis CD. Translating the Histone Code. Science. 2001;293:1074-1080.
1013 48. Attanasio C, Nord AS, Zhu Y, Blow MJ, Biddie SC, Mendenhall EM, Dixon J, Wright C, Hosseini
1014 R, Akiyama JA, et al. Tissue-specific SMARCA4 binding at active and repressed regulatory elements
1015 during embryogenesis. Genome Res. 2014;24:920-9.
1016 49. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner
1017 R, Coyne M, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature.
1018 2011;473:43-9.
1019 50. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M,
1020 Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells.
1021 Cell. 2006;125:315-26.
1022 51. Albert M, Kalebic N, Florio M, Lakshmanaperumal N, Haffner C, Brandl H, Henry I, Huttner WB.
1023 Epigenome profiling and editing of neocortical progenitor cells during development. EMBO J.
1024 2017;36:2642-58.
1025 52. Albrieux M, Platel JC, Dupuis A, Villaz M, Moody WJ. Early expression of sodium channel
1026 transcripts and sodium current by cajal-retzius cells in the preplate of the embryonic mouse neocortex.
1027 J Neurosci. 2004;24:1719-25.
1028 53. Mienville JM, Barker JL. Potassium Current Expression during Prenatal Corticogenesis in the Rat.
1029 Neuroscience. 1997;81:163-172.
1030 54. Kirmse K, Grantyn R, Kirischuk S. Developmental downregulation of low-voltage-activated
1031 Ca2+ channels in Cajal-Retzius cells of the mouse visual cortex. Eur J Neurosci. 2005;21:3269-76.
1032 55. Moore AR, Filipovic R, Mo Z, Rasband MN, Zecevic N, Antic SD. Electrical Excitability of Early
1033 Neurons in the Human Cerebral Cortex during the Second Trimester of Gestation. Cerebral Cortex.
1034 2009;19:1795-1805.
1035 56. Selemon LD, Zecevic N. Schizophrenia: a tale of two critical periods for prefrontal cortical
1036 development. Transl Psychiatry. 2015;5:e623.
1037 57. Heather LC, Wang X, West JA, Griffin JL. A practical guide to metabolomic profiling as a
1038 discovery tool for human heart disease. J Mol Cell Cardiol. 2013;55:2-11.
1039 58. Knuesel I. Reelin-mediated signaling in neuropsychiatric and neurodegenerative diseases.
1040 Progress in Neurobiology. 2010;91:257-274.
1041 59. Meyer G, Perez-Garcia CG, Abraham H, Caput D. Expression of p73 and Reelin in the
1042 Developing Human Cortex. J Neurosci. 2002;22:4973–86.
1043 60. Miquelajauregui A, Varela-Echavarria A, Ceci ML, Garcia-Moreno F, Ricano I, Hoang K,
1044 Frade-Perez D, Portera-Cailliau C, Tamariz E, De Carlos JA, et al. LIM-homeobox gene Lhx5 is

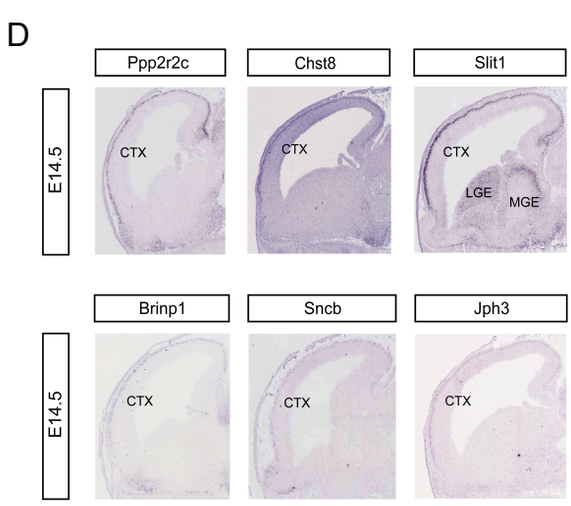
1045 required for normal development of Cajal-Retzius cells. J Neurosci. 2010;30:10551-62.
1046 61. Shinozaki K, Miyagi T, Yoshida M, Miyata T, Ogawa M, Aizawa S, Suda Y. Absence of
1047 Cajal-Retzius cells and subplate neurons associated with defects of tangential cell migration from
1048 ganglionic eminence in Emx1/2double mutant cerebral cortex. Development. 2002;129:3479-92.
1049 62. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ,
1050 Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and
1051 isoform switching during cell differentiation. Nat Biotechnol. 2010;28:511-5.
1052 63. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass
1053 CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements
1054 required for macrophage and B cell identities. Mol Cell. 2010;38:576-89.
1055 64. Levy R, Molina J, Danielpour M, Breunig JJ. Neonatal pial surface electroporation. J Vis Exp.
1056 2014:1-7.
1057 65. Shen X, Kim W, Fujiwara Y, Simon MD, Liu Y, Mysliwiec MR, Yuan GC, Lee Y, Orkin SH.
1058 Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell.
1059 2009;139:1303-14.

Figure 1


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

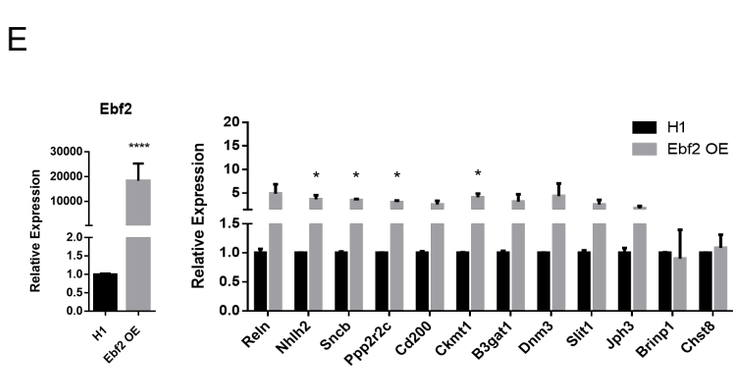

E
Ebf2
Relative Expression
H1
Ebf2 OE
****
Reln
Nhlh2
Sncb
Ppp2r2c
Cd200
Ckmt1
B3gat1
Dnm3
Slit1
Jph3
Brinp1
Chst8

Figure 2

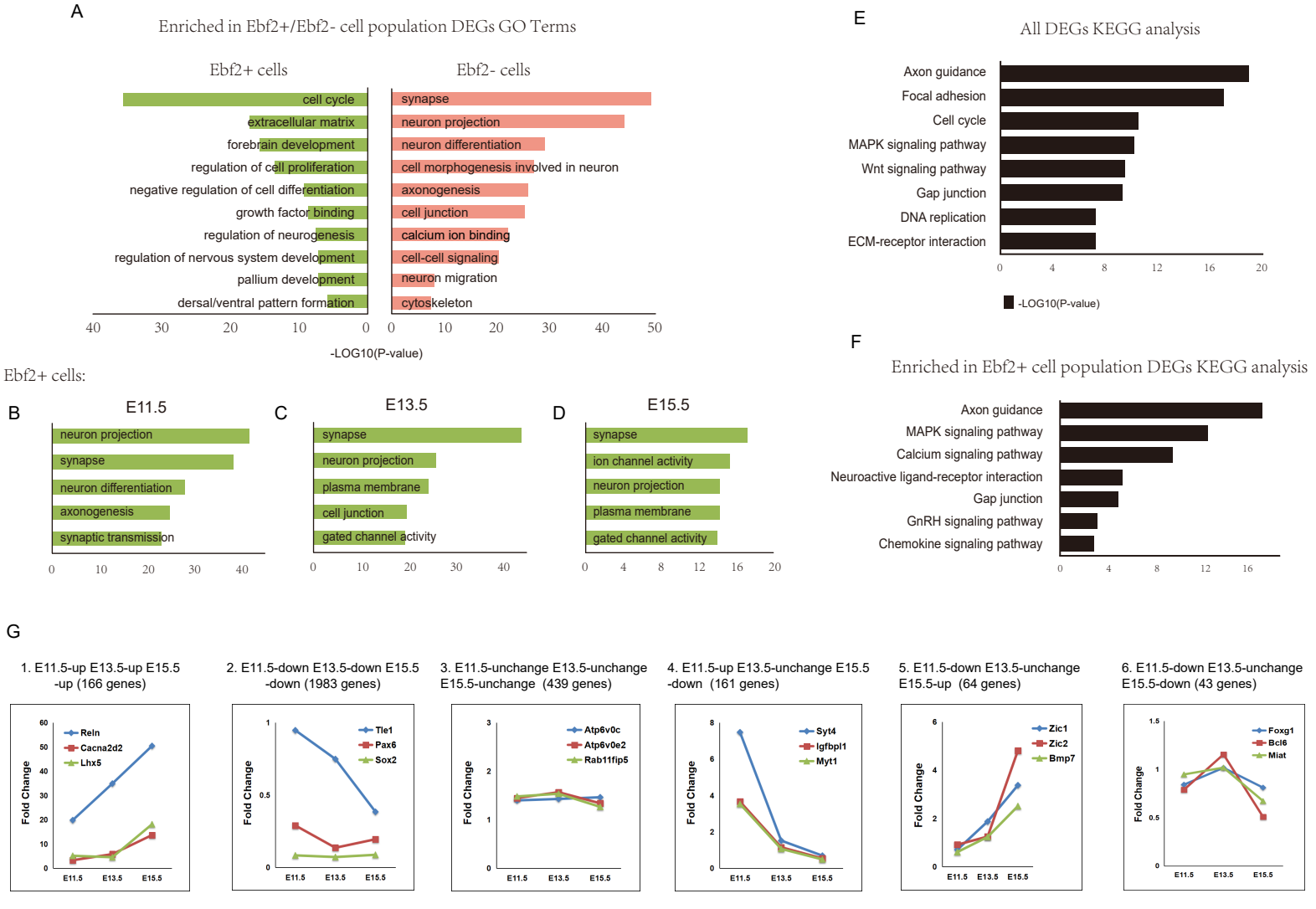

A
Enriched in Ebf2+/Ebf2- cell population DEGs GO Terms
Ebf2+ cells
Ebf2- cells
cell cycle
extracellular matrix
forebrain development
regulation of cell proliferation
negative regulation of cell differentiation
growth factor binding
regulation of neurogenesis
regulation of nervous system development
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)
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
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
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
Miat
Fold Change
E11.5
E13.5
E15.5

Figure 3

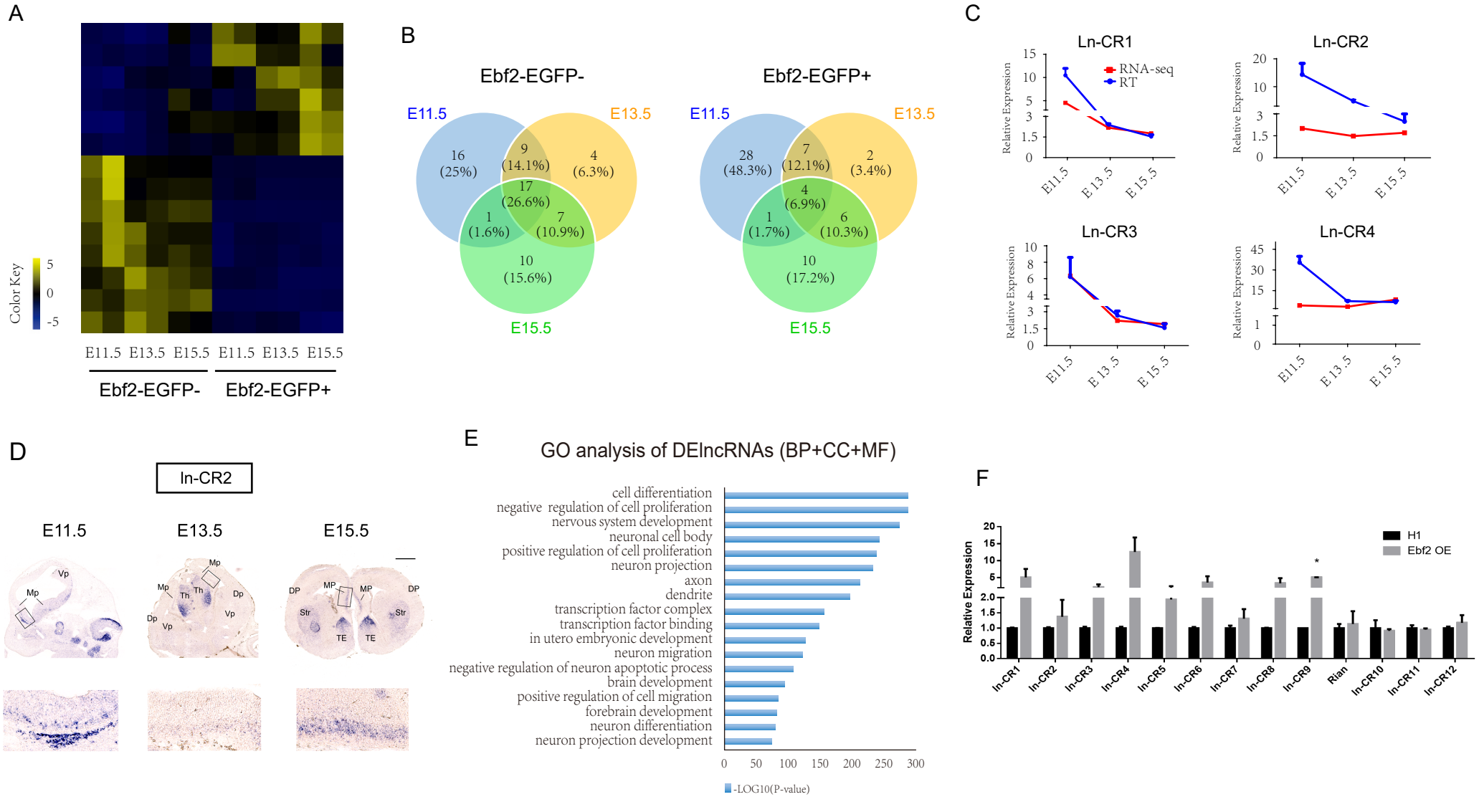

A
Color Key
5
0
-5
E11.5 E13.5 E15.5 E11.5 E13.5 E15.5
Ebf2-EGFP- Ebf2-EGFP+
B
Ebf2-EGFP-
E11.5
E13.5
E15.5
16 (25%)
9 (14.1%)
4 (6.3%)
17 (26.6%)
1 (1.6%)
7 (10.9%)
10 (15.6%)
Ebf2-EGFP+
E11.5
E13.5
E15.5
28 (48.3%)
7 (12.1%)
2 (3.4%)
4 (6.9%)
1 (1.7%)
6 (10.3%)
10 (17.2%)
C
Ln-CR1
Ln-CR2
Ln-CR3
Ln-CR4
RNA-seq
RT
Relative Expression
E11.5
E13.5
E15.5
D
lncR2
E11.5
E13.5
E15.5
Vp
Mp
Th
Dp
DP
MP
Str
TE
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
lncR1
lncR2
lncR3
lncR4
lncR5
lncR6
lncR7
lncR8
lncR9
Rian
lncR10
lncR11
lncR12

Figure 4

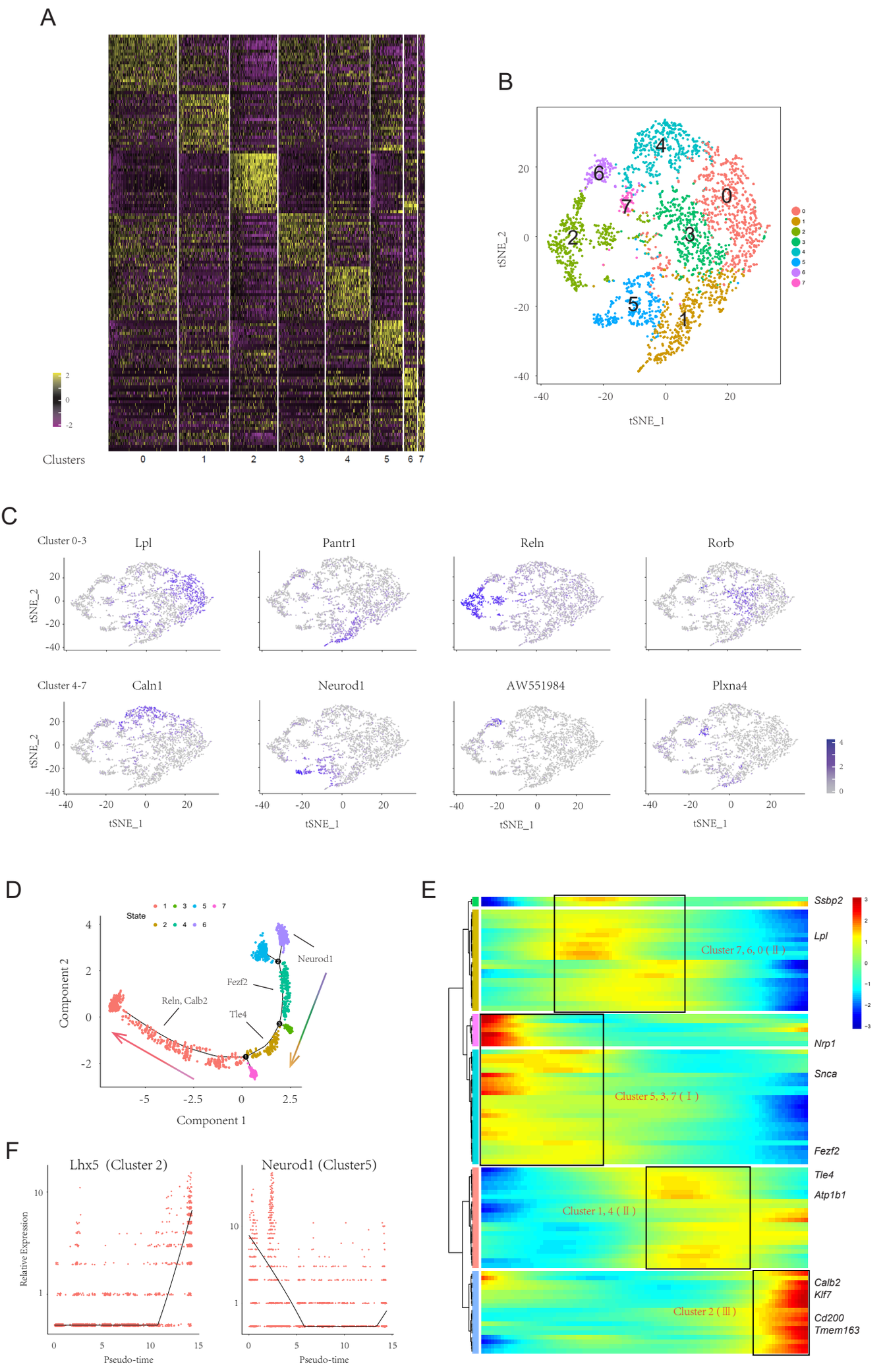
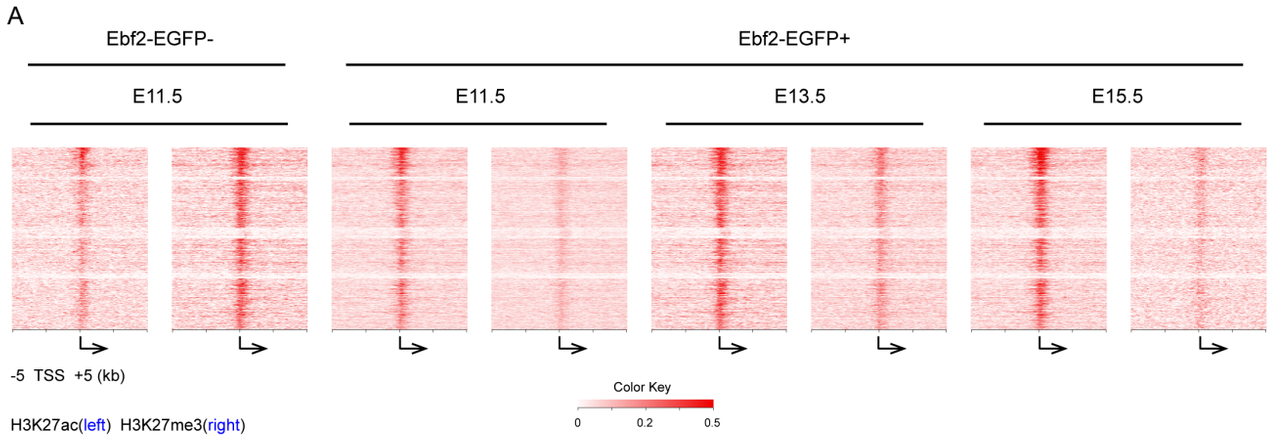
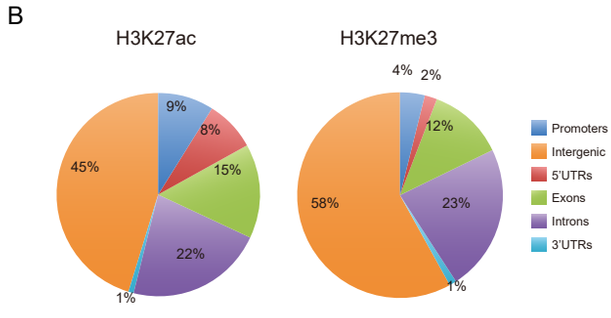
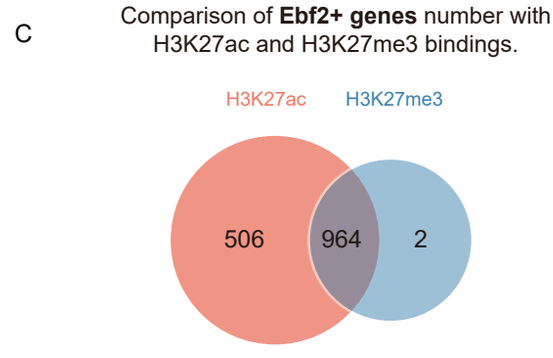
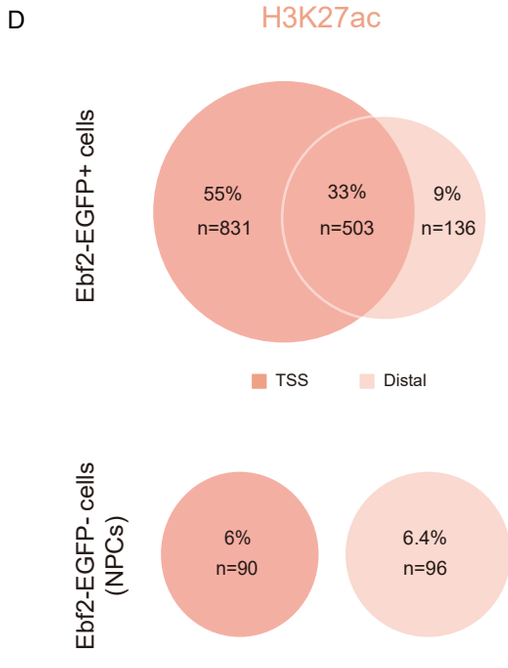

A
Clusters
0 1 2 3 4 5 6 7
B
tSNE_1
tSNE_2
C
Cluster 0-3
Lpl
Pantr1
Reln
Rorb
Cluster 4-7
Caln1
Neurod1
AW551984
Plxna4
D
State
Component 1
Component 2
Neurod1
Fezf2
Tle4
Reln, Calb2
E
Ssbp2
Lpl
Cluster 7, 6, 0 (Ⅱ)
Nrp1
Snca
Cluster 5, 3, 7 (Ⅰ)
Fezf2
Tle4
Atp1b1
Cluster 1, 4 (Ⅱ)
Calb2
Klf7
Cd200
Tmem163
Cluster 2 (Ⅲ)
F
Lhx5 (Cluster 2)
Neurod1 (Cluster5)
Relative Expression
Pseudo-time

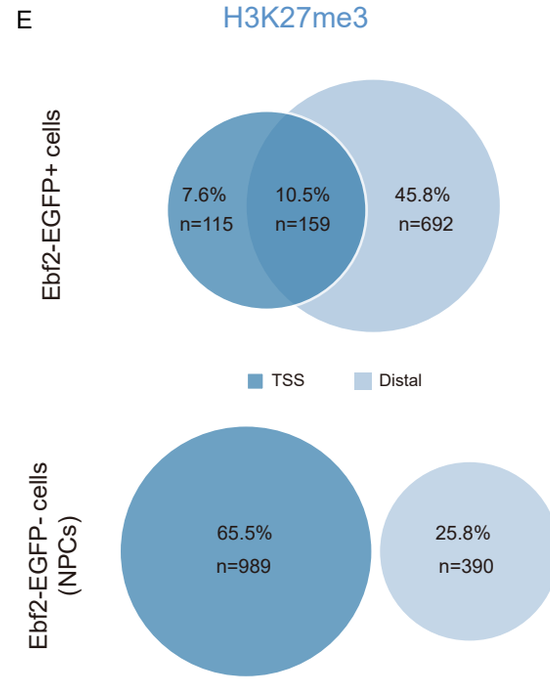
Figure 5


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

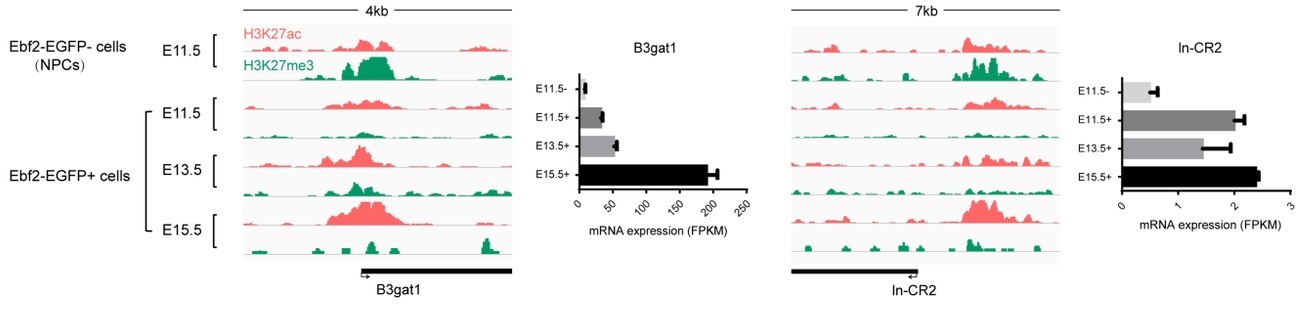
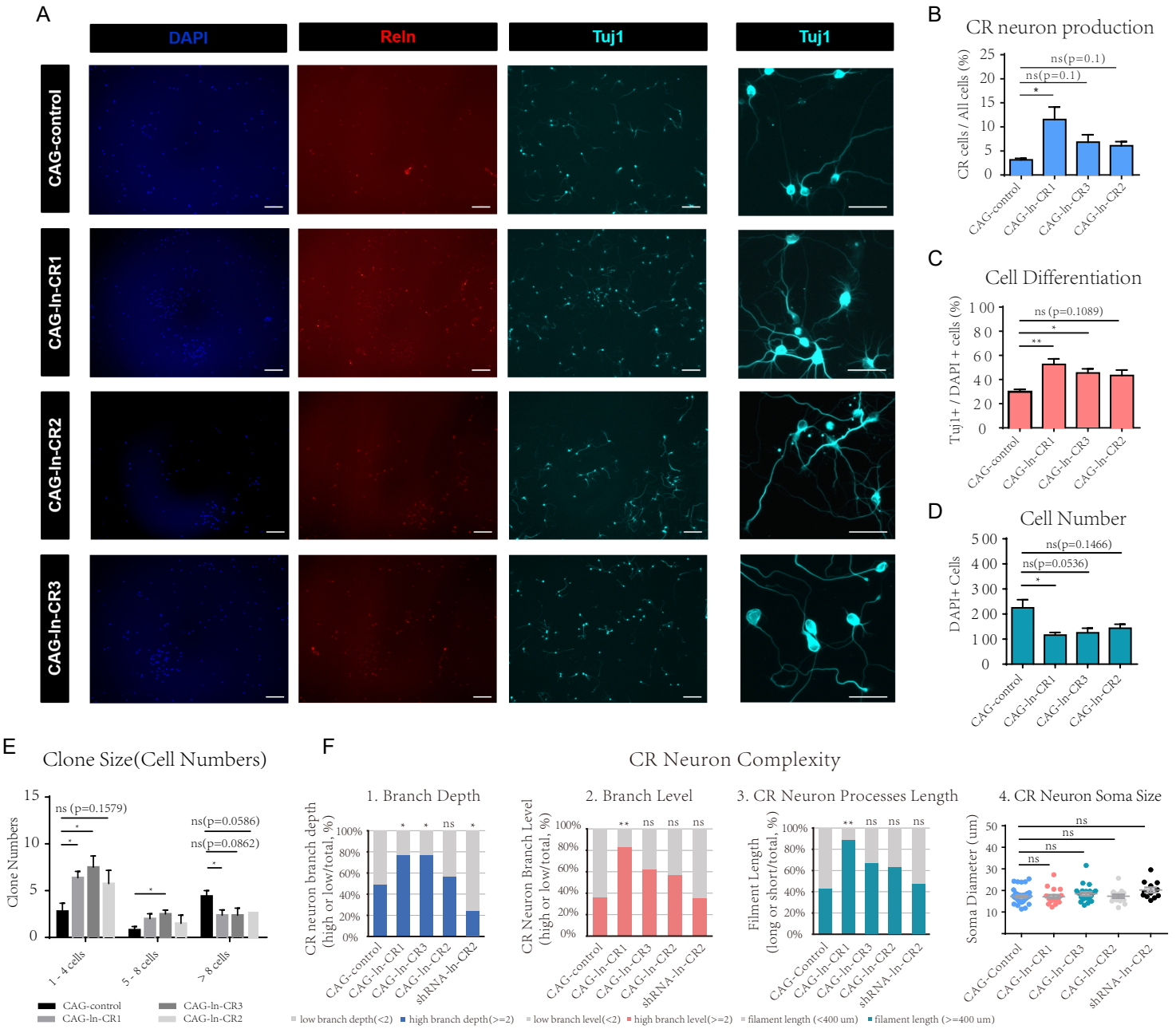

F
4kb
Ebf2-EGFP- cells (NPCs)
E11.5
H3K27ac
H3K27me3
Ebf2-EGFP+ cells
E11.5
E13.5
E15.5
B3gat1
E11.5-
E11.5+
E13.5+
E15.5+
0
50
100
150
200
250
mRNA expression (FPKM)
7kb
In-CR2
0
1
2
3
mRNA expression (FPKM)

Figure 6


A
DAPI
Reln
Tuj1
Tuj1
CAG-control
CAG-In-CR1
CAG-In-CR2
CAG-In-CR3
B
CR neuron production
CR cells / All cells (%)
ns(p=0.1)
ns(p=0.1)
*
CAG-control
CAG-In-CR1
CAG-In-CR3
CAG-In-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-In-CR1
CAG-In-CR3
CAG-In-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-In-CR1
CAG-In-CR3
CAG-In-CR2
shRNA-In-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-In-CR1
Merge
Ebf2-EGFP
CAG-control-RFP
CAG-In-CR1-RFP