

# Supporting Information

Ayoub et al. 10.1073/pnas.1112213108

## SI Materials and Methods

**Animals.** C57BL/6J mice were housed and bred in the animal care facility at Yale University School of Medicine. Mouse embryos were collected from timed pregnancies, in which the vaginal plug date was considered E0.5. Brains were dissected and quickly frozen over isopentane/dry ice slurry for LMD experiments. For histological procedures, brains were immersed in 4% paraformaldehyde for 4 h, cryoprotected in 30% sucrose, and then frozen as before.

**LMD.** LMD was performed on E14.5 coronal cryosections (18  $\mu$ m thick) mounted on PEN-slides (Leica Microsystems) and chosen by systematic random sampling. The VZ was dissected only from the dorsolateral and the medial pallium. To collect the SVZ–IZ region, we consistently placed the laser cut line on the border with the VZ and excluded subplate neurons near the border of the CP. All differentiated neuronal subtypes that reside in thin subjacent layers, such as Cajal–Retzius cells and layer VB neurons, were grouped with the CP region. LMD was performed using a fully motorized LMD6000 system (Leica) under a 20 $\times$  objective. Slides were processed for 20 to 30 s in each of the following steps: twice in 70% ethanol, 0.2% toluidine blue in 70% ethanol, then dehydrated in 70%, 90%, and 100% ethanol. Dissected cells were collected from the neocortical zones in 30-min intervals, then stored at  $-80^{\circ}\text{C}$  until further processing. RNA was pooled from three brains for each replicate and then extracted by using the RNeasy Microkit (Qiagen) according to the protocol provided for LMD tissue.

**ISH.** Antisense RNA probes were designed against the *Mus musculus* RefSeq by using Primer Blast with amplicon size set to 600 to 1,000 bp. Probes were synthesized from PCR products and then labeled by using the DIG labeling kit (Roche). Hybridization on 30- $\mu$ m cryosections was performed following a standard protocol applied in our laboratory as previously described (1). Imaging was performed on an Axioplan2e with an MRm camera (Carl Zeiss Microimaging). All primers are included in Table S2.

**Quantitative Real-Time PCR.** Primers for mouse genes were designed by using Primer Blast against the *M. musculus* mRNA RefSeq. Template cDNA was generated from total RNA by using the SuperScript III First Strand Synthesis System (Invitrogen). qRT-PCR reactions were performed in quadruplets on 6 ng of template cDNA in a total of 20  $\mu$ L reaction volume per well by using the Power SYBR Green PCR Master Mix (Applied Biosystems) on an ABI PRISM 7900 (Applied Biosystems). The mouse GADPH gene was used as a reference for all samples. Fold change was determined by using the DDct method. All primers are included in Table S2.

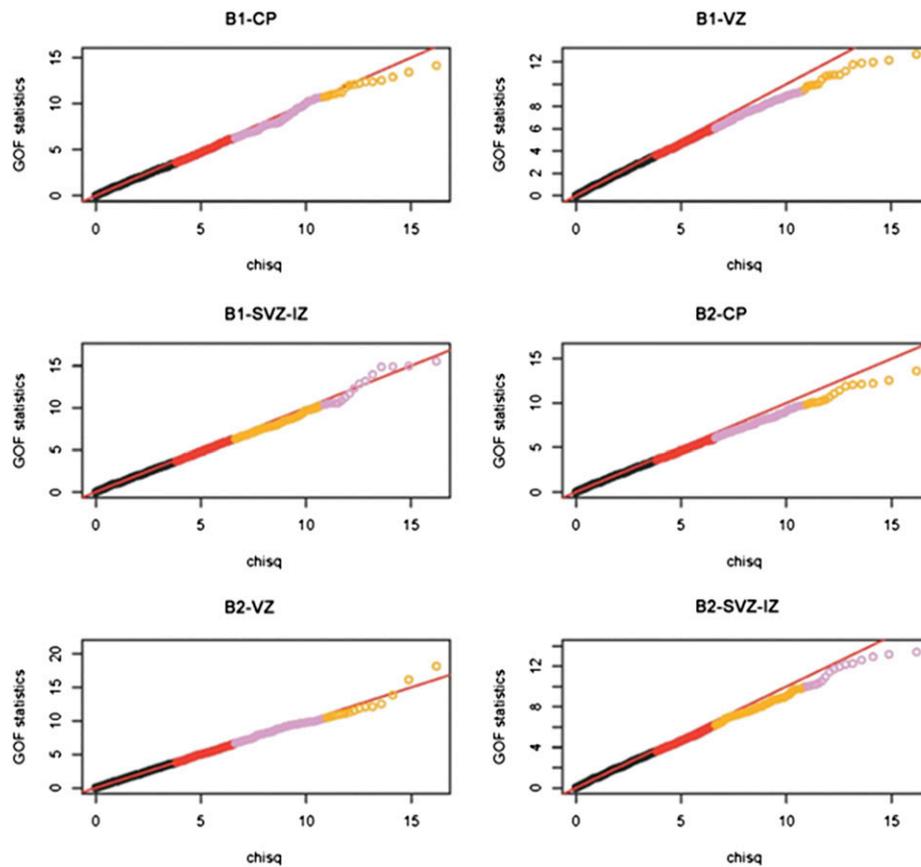
**High-Throughput Sequencing and Quantification of Gene Expression.** The quality of total RNA was checked on a BioAnalyzer (Agilent). mRNA-seq libraries were constructed according to the standard Illumina protocol. Two lanes of single-end 75-bp sequencing were performed for each library by using the Illumina Genome Analyzer IIX at the Yale Center for Genome Analysis. Sequencing reads were trimmed and uniquely mapped to the UCSC mouse reference genome (mm9) and a custom splice junction library in parallel using Bowtie (bowtie-trim5 10-trim3 4 -l 35 -m 1-solexa1.3-quals-sam-sam-nohead) (2). The splice junction library based on the UCSC Known Genes annotation,

consisting of 110-bp splice junction sequences incorporating 55 bp of transcript sequence flanking either side of each splice site, was built by using RSEQTools (3). mRNA-seq generated more than 50 million sequencing reads, 60% mapped uniquely to the mouse genome (Fig. 1B). The number of genes with mapped reads in all cellular zones was 14,582 for B1 and 14,570 for B2, respectively. Genes were considered not expressed if they had zero reads in all technical and biological replicates for each cellular zone. Mapped reads in Bowtie SAM format were converted to SAM format required by Cufflinks (4), which reports RPKM values for individual annotated transcripts. Gene-level RPKM was calculated as the sum of RPKMs of all annotated isoforms for that gene. nTARs were called based on minimum 0.5 $\times$  coverage at each base per million reads, with gaps less than 20 bp and minimum 100 bp length. We merged nTARs from both replicates and then calculated RPKM values. All computations were completed on the Yale Biomedical HPC clusters. We used Galaxy (5) for genomic analyses of DEGs, DEIs, p300-enriched forebrain peaks, and nTARs, as well as computing overlap with various genomic intervals, such as promoters and intergenic regions including gene deserts. Raw and processed data can be obtained from GSE30765 or the author's Web site (<http://rakiclab.med.yale.edu/transcriptome.php>).

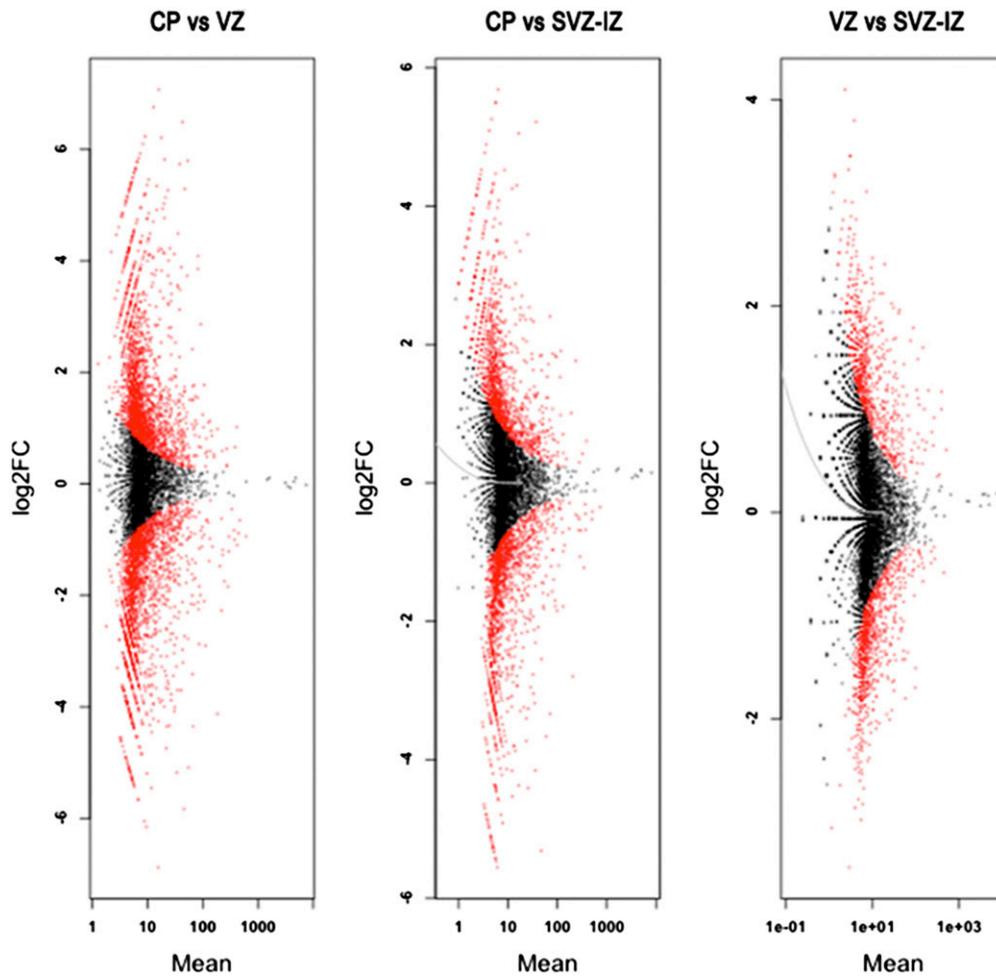
**Statistical Methods.** First, we performed exploratory analyses to investigate differences specific to biological tissue, biological replicates, and sequencing lanes (Fig. 1C and D). We prepared mean-difference (M-D) plots (Figs. S1–S3) to investigate any bias in the distribution of data and quantile-quantile goodness-of-fit plots to determine any sequencing bias (Fig. S2). Subsequently, we applied a Fisher exact test for individual DEG data and the likelihood ratio test from the log-linear Poisson model to identify DEGs in data pooled from two biological replicates (6, 7). We used a two-by-two contingency table Fisher exact test with rows corresponding to each gene and columns to tissue type. The *P* value was computed by summation of probabilities less than or equal to the probability of the observed table based on the hypergeometric distribution. We tested three pairwise tissue-specific differences—(i) CP vs. VZ, (ii) CP vs. SVZ–IZ, and (iii) VZ vs. SVZ–IZ—in two biological replicates. Additionally, we compared DEG lists obtained by using individual Fisher exact tests to the edgeR package (Bioconductor), which returned no difference in detecting statistical significance at  $P < 0.001$  (Table S1). We applied the likelihood ratio test based on the log-linear Poisson model for data pooled from two biological replicates (B1 and B2). We applied a conservative Benjamini–Hochberg *P* value ( $\text{bhp} < 10^{-3}$ ) and a  $\log_2$  fold change to the list of 27,189 expressed genes, which resulted in 875 DEGs in the CP, 622 DEGs in the SVZ–IZ, and 1,180 DEGs in the VZ. Among 49,152 isoforms, 5,135 passed a  $\text{bhp} < 10^{-5}$  and only 3,498 exhibited more than twofold change between zones. We detected 219,528 exons expressed in the neocortex, of which 11,584 are differentially expressed. We used the same criteria as described earlier to select differentially expressed exons (DEEs). We found 5,098 DEEs enriched in the CP, 2,431 DEEs in the SVZ–IZ, and 4,055 DEEs in the VZ. Because some genes showed large variability between biological replicates, we performed log-linear analysis with the negative binomial model instead of the Poisson model. However, we report only the results based on the log-linear Poisson model because both tests were consistent in detecting statistically significant differences at a  $P \leq 0.001$  significance level (Table S1). Tables containing gene and transcript







**Fig. S2.** Quantile-quantile goodness-of-fit plot presents two probability distributions of technical replicates for each zone. By comparing a pair of sequencing lanes within tissue, for each gene, a  $P$  value is computed to test a deviation between two lanes with the null hypothesis that the read counts in the gene and one lane resembled a random sample from the reads in both lanes. The distribution of these  $P$  values across genes is a uniform distribution in  $\{0,1\}$ , whereas deviations from linearity showing any skewed pattern may represent a lane effect. Labels starting with "B1" and "B2" indicate biological replicas 1 and 2, respectively; C, CP; S, SVZ-IZ; V, VZ.



**Fig. S3.** M-D plot for nTARs. The y axis shows  $\log_2$  transformation for fold change and overall expression on the x axis. The general assumption is that most of the nTARs would not show any expression differences between zones. The pattern of M-D plot would be that most of points are located on the horizontal line, at  $M = 0$ , unless there is bias in the distribution. To display a number for nTARs with zero read counts in the M-D plot, we added one to all data points.



**Table S1. Principal components analysis data details and comparison of different statistical models**

[Table S1 \(XLS\)](#)

**Table S2. List of qRT-PCR and ISH primers and summary of nTARs and DEIs**

[Table S2 \(XLS\)](#)

**Table S3. Summary of GO functional analysis by biological and disease processes (complete DAVID analysis and summary of GeneGO pathway analysis)**

[Table S3 \(XLS\)](#)

**Table S4. SOM and network data**

[Table S4 \(XLS\)](#)

**Table S5. Candidate enhancers near DEGs**

[Table S5 \(XLS\)](#)