Supplemental Information

Laminar and Temporal Dynamics of Coding and Noncoding RNAs in the Mouse Neocortex

Sofia Fertuzinhos, Mingfeng Li, Yuka Imamura Kawasawa, Vedrana Perkov, Daniel Franjić, Darshani Singh, Michael Crair and Nenad Šestan

SUPPLEMENTAL FIGURES

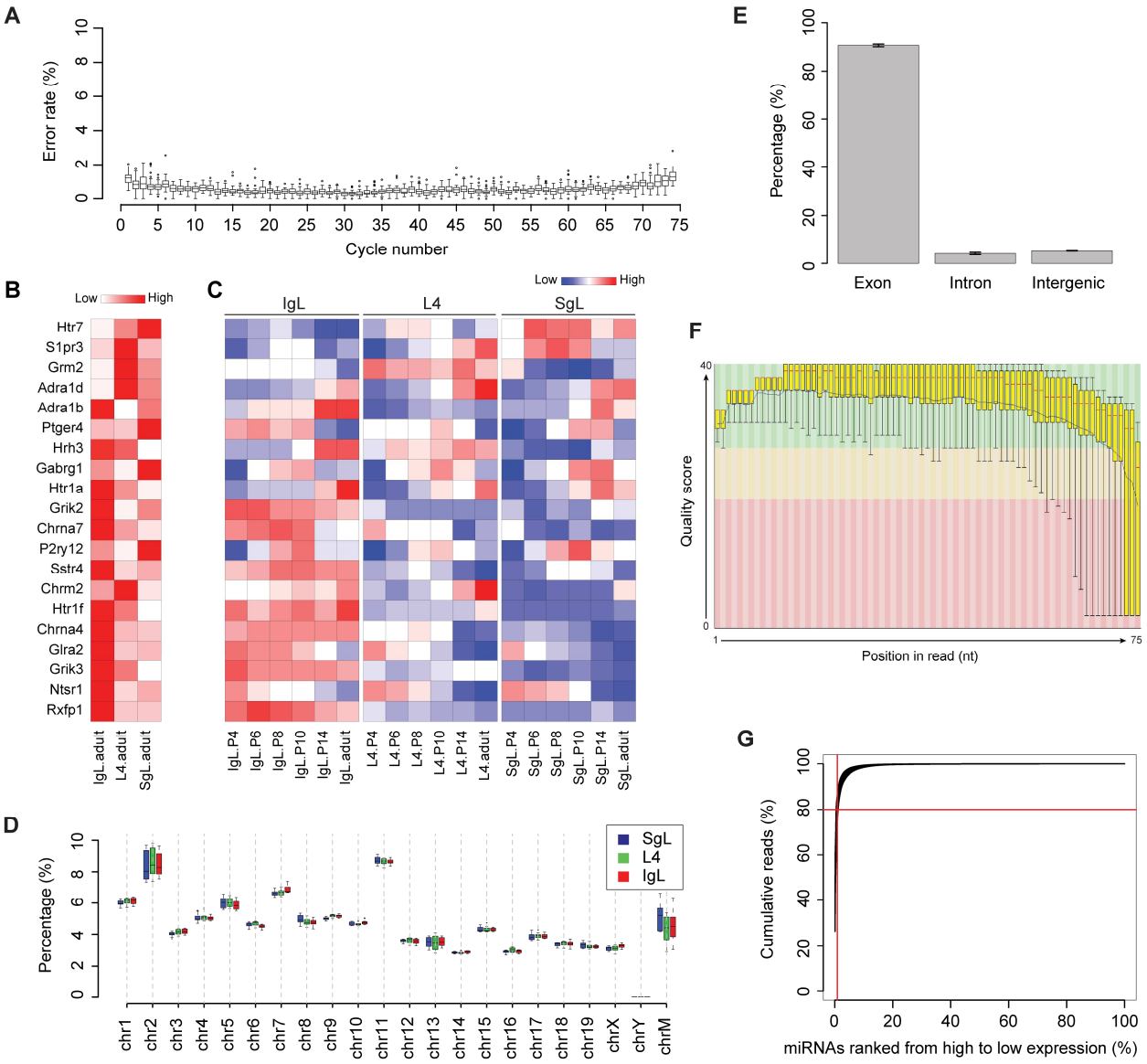

Figure S1. Quality Assessment of mRNA-seq and smRNA-seq Sequencing Data, Related to Figure 1

(A) Sequencing error rate as a function of cycle number for mRNA-seq data. The sequencing error rate was estimated by the ratio of the mismatch of the aligned spike-in reads from the reference spike-in RNA sequences. The distribution of the error rates in each sequencing cycle was shown by box plots across each sequencing cycle.

(B) The 20 novel significant layer-pattern genes reported in Belgard et al. 2011 were plotted by using 6 adult samples, showing consistent patterns.

(C) Heat map of novel layer-patterned genes reported in Belgard et al. 2011. These genes were re-plotted using the 36 samples from all ages. Most genes showed temporal dynamics of

enrichment. To plot the heat map, the log2 transformed gene RPKM+1 was normalized by *normalize* function in R software.
(D) Reads distribution across chromosomes. The box plots represent the proportions of reads uniquely mapped to each chromosome in SgL (blue), L4 (green) and IgL (red) samples.
(E) Reads classification to exon, intron and intergenic regions. Reads were assigned to exon, intron and intergenic regions according to their mapping coordinates. Bar plots represent the proportions of reads residing in each annotation entry. Error bars represent s.e.m. (n=36).
(F) Example of smRNA-seq base-quality distribution. The SgL of MMB1 smRNA-seq data was chosen as an example. The quality scores of the smRNA-seq reads were summarized by software FastQC. The yellow box plots show an overview of the range of quality scores across all bases at each position. The background of the graph divides the y axis into very good quality (green), reasonable quality (orange), and poor quality (red). Notably, all 36 smRNA-seq samples show a pattern similar to this example. nt, nucleotides.
(G) Cumulative dependence between smRNA-seq reads and miRNAs. The accumulated percentage of uniquely mapped reads was plotted against the accumulated percentage of miRNAs ranked from highest to lowest expression for 36 samples. The vertical red line indicates the top 10 highly expressed miRNAs and the horizontal red line indicates the 80% uniquely mapped reads.

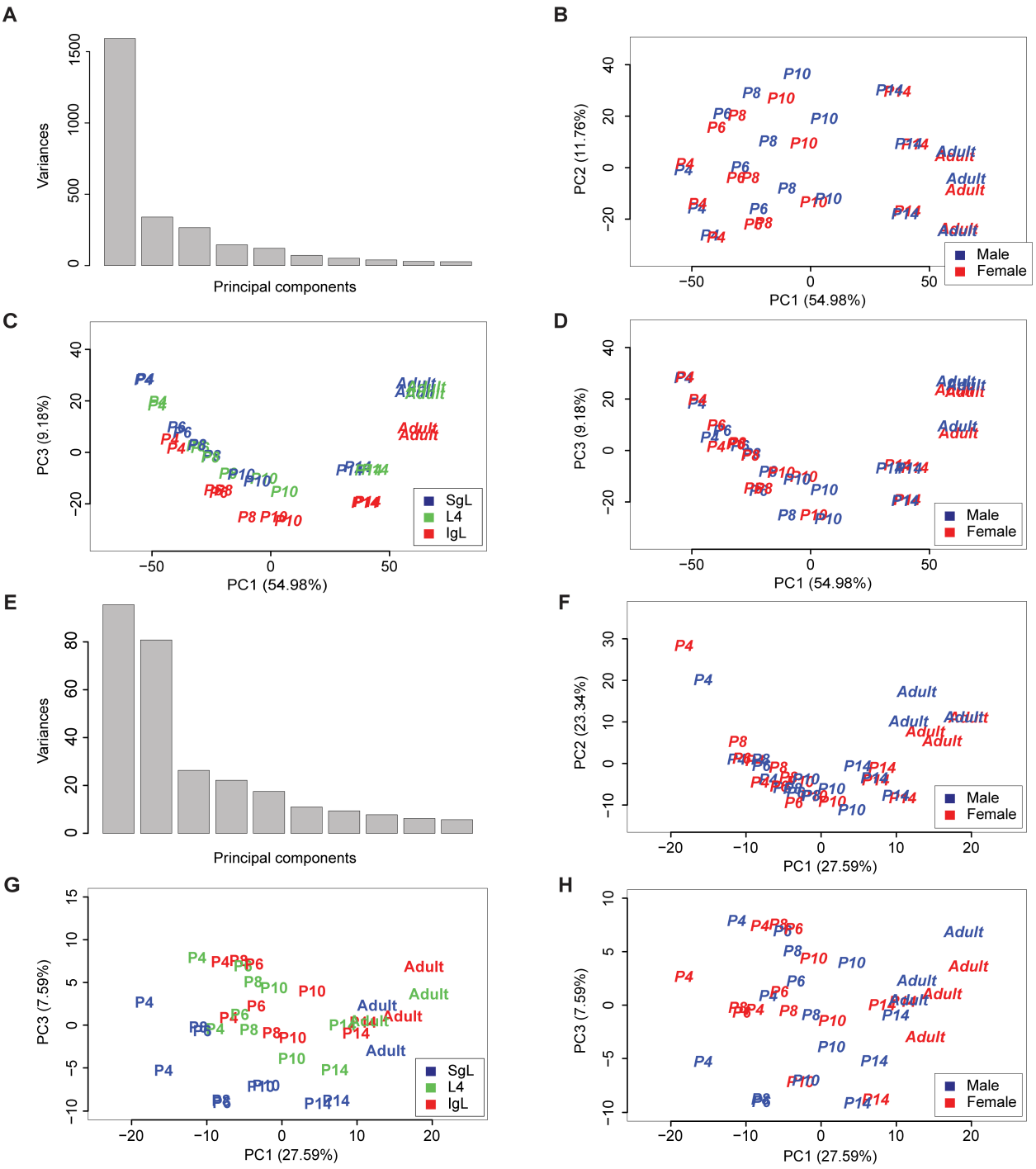

Figure S2. Principal Component Analyses of mRNA-seq and smRNA-seq Samples, Related to Figure 2.

(A and E) Distribution of variances as a function of principal components.

(B and F) Two-dimensional plot of PC1 and PC2 representing the PCA analysis of 36 mRNA-seq (B) and smRNA-seq (F) samples. Each character string represents one sample and indicates its age. The male (blue) and female (red) samples were differentially colored.

(C and G) Two-dimensional plot of PC1 and PC3 representing the PCA analysis of 36 mRNA-seq (C) and smRNA-seq (G) samples. Each character string represents one sample and indicates its age. The SgL (blue), L4 (green) and IgL (red) samples were differentially colored.
(D and H) Two-dimensional plot of PC1 and PC3 representing the PCA analysis of 36 mRNA-seq (D) and smRNA-seq (H) samples. Each character string represents one sample and indicates its age. The male samples are blue, and female samples are red.

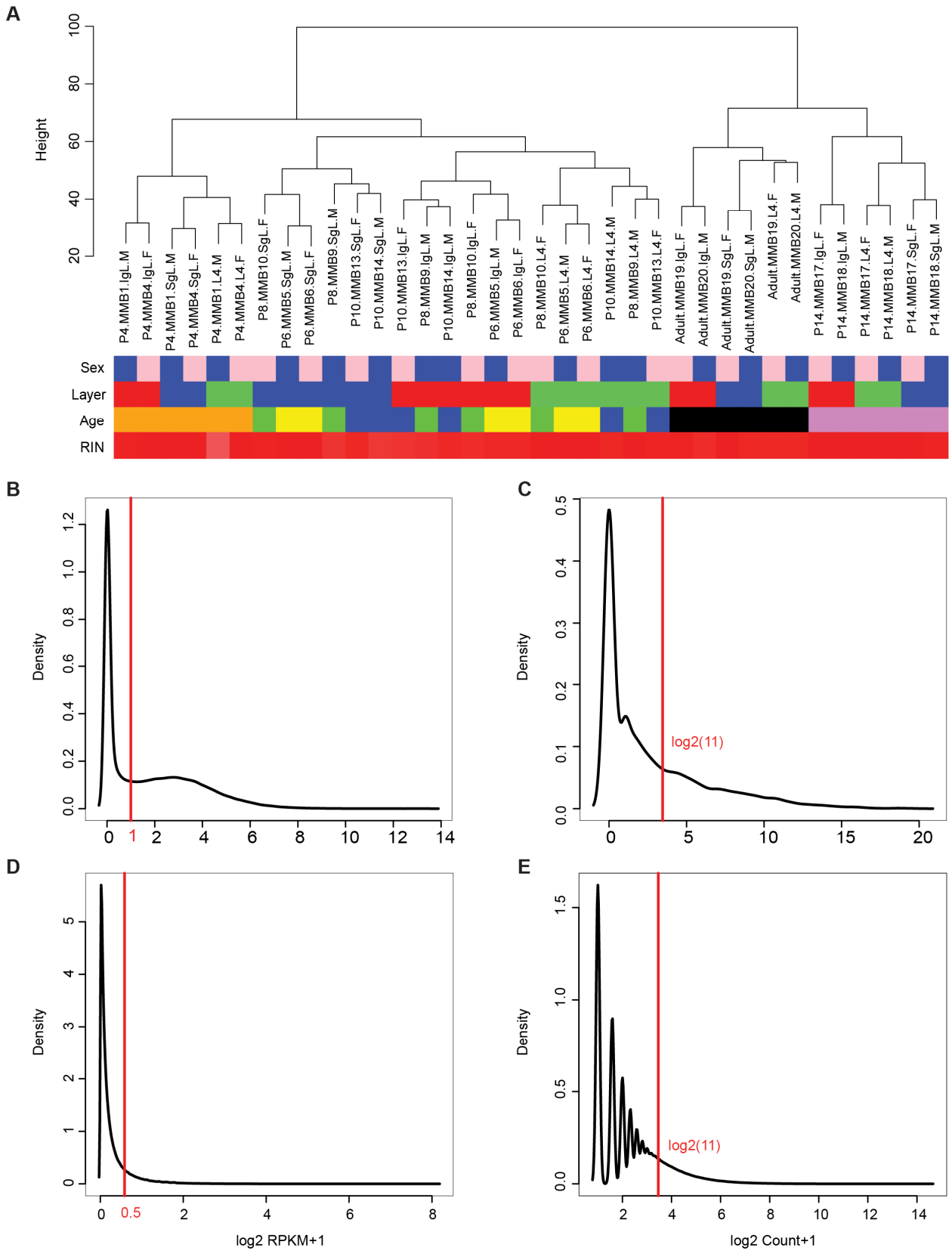
A
Height
100
80
60
40
20
P4.MMB1.lgL.M
P4.MMB4.lgL.F
P4.MMB1.SgL.M
P4.MMB4.SgL.F
P4.MMB1.L4.M
P4.MMB4.L4.F
P8.MMB10.SgL.F
P6.MMB5.SgL.M
P6.MMB6.SgL.F
P8.MMB9.SgL.M
P10.MMB13.SgL.F
P10.MMB14.SgL.M
P10.MMB13.lgL.F
P8.MMB9.lgL.M
P10.MMB14.lgL.M
P8.MMB10.lgL.F
P6.MMB5.lgL.M
P6.MMB6.lgL.F
P8.MMB10.L4.F
P6.MMB5.L4.M
P6.MMB6.L4.F
P10.MMB14.L4.M
P8.MMB9.L4.M
P10.MMB13.L4.F
Adult.MMB19.lgL.F
Adult.MMB20.lgL.M
Adult.MMB19.SgL.F
Adult.MMB20.SgL.M
Adult.MMB19.L4.F
Adult.MMB20.L4.M
P14.MMB17.lgL.F
P14.MMB18.lgL.M
P14.MMB17.L4.F
P14.MMB18.L4.M
P14.MMB17.SgL.F
P14.MMB18.SgL.M
Sex
Layer
Age
RIN
B
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
1
2
4
6
8
10
12
14
C
Density
0.0
0.1
0.2
0.3
0.4
0.5
log2(11)
0
5
10
15
20
D
Density
0
1
2
3
4
5
0
0.5
2
4
6
8
log2 RPKM+1
E
Density
0.0
0.5
1.0
1.5
log2(11)
2
4
6
8
10
12
14
log2 Count+1

Figure S3. Dendrogram of mRNA-seq Samples and Expression Distribution of Genes, Introns and miRNAs, Related to Figure 2 and Figure 3

(A) Dendrogram of mRNA-seq samples in terms of sex, layer, age and RIN. Dendrogram branches show hierarchical clustering of 36 mRNA-seq samples. The colors underneath label the potential confounders: sex (male - blue, female - pink), layer (SgL - blue, L4 - green, IgL - red), age (P4 - orange, P6 - yellow, P8 - green, P10 - blue, P14 - violet, adult - black), RIN (low to high representing pink to red, ranging from 8 to 10). The samples were clustered predominately by layer and age, but not by sex or RIN.

(B) The density plot of log2 transformed RPKM+1 for all annotated genes in mm9 Ensembl v63. The red vertical line indicates the threshold used for choosing reliably expressed protein-coding mRNA genes, RPKM ≥ 1 in at least 2 samples.

(C) The density plot of log2 transformed reads count+1 for all annotated miRNAs in mm9 miRBase v18. The red vertical line indicates the threshold used for choosing reliably expressed miRNAs, reads count ≥ 10 in at least 2 samples.

(D) The density plot of log2 transformed RPKM+1 for all annotated introns in mm9 Ensembl v63.

(E) The density plot of log2 transformed read count+1 for all annotated introns in mm9 Ensembl v63. The definition of reliably expressed intron was based on the combination of RPKM ≥ 0.5 and reads count ≥ 10 in at least 2 samples, being indicated by the red vertical lines in D and E.

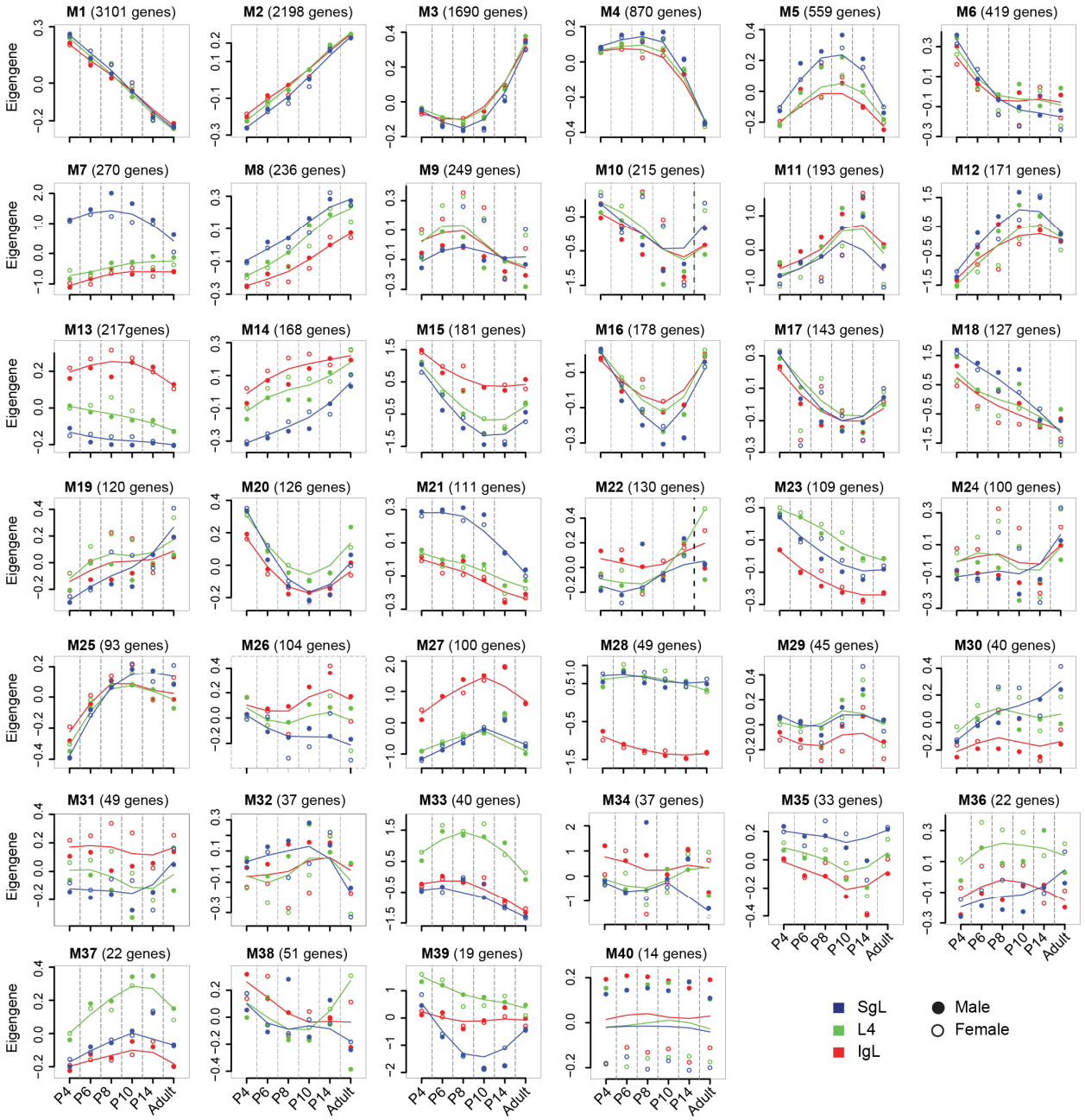

Figure S4. Trajectories of Modular Eigengenes, Related to Figure 4

The spatiotemporal pattern for each module was summarized by the trajectory of the modular eigengene values, in which the eigengene was plotted against age. The trajectories for SgL (blue), L4 (green) and lgL (red) samples were separated and smoothed by loess function in R software. The full and empty circles represent the male and female samples respectively, and are differentially colored according to layers. The module index and the number of modular genes were indicated above the plot.

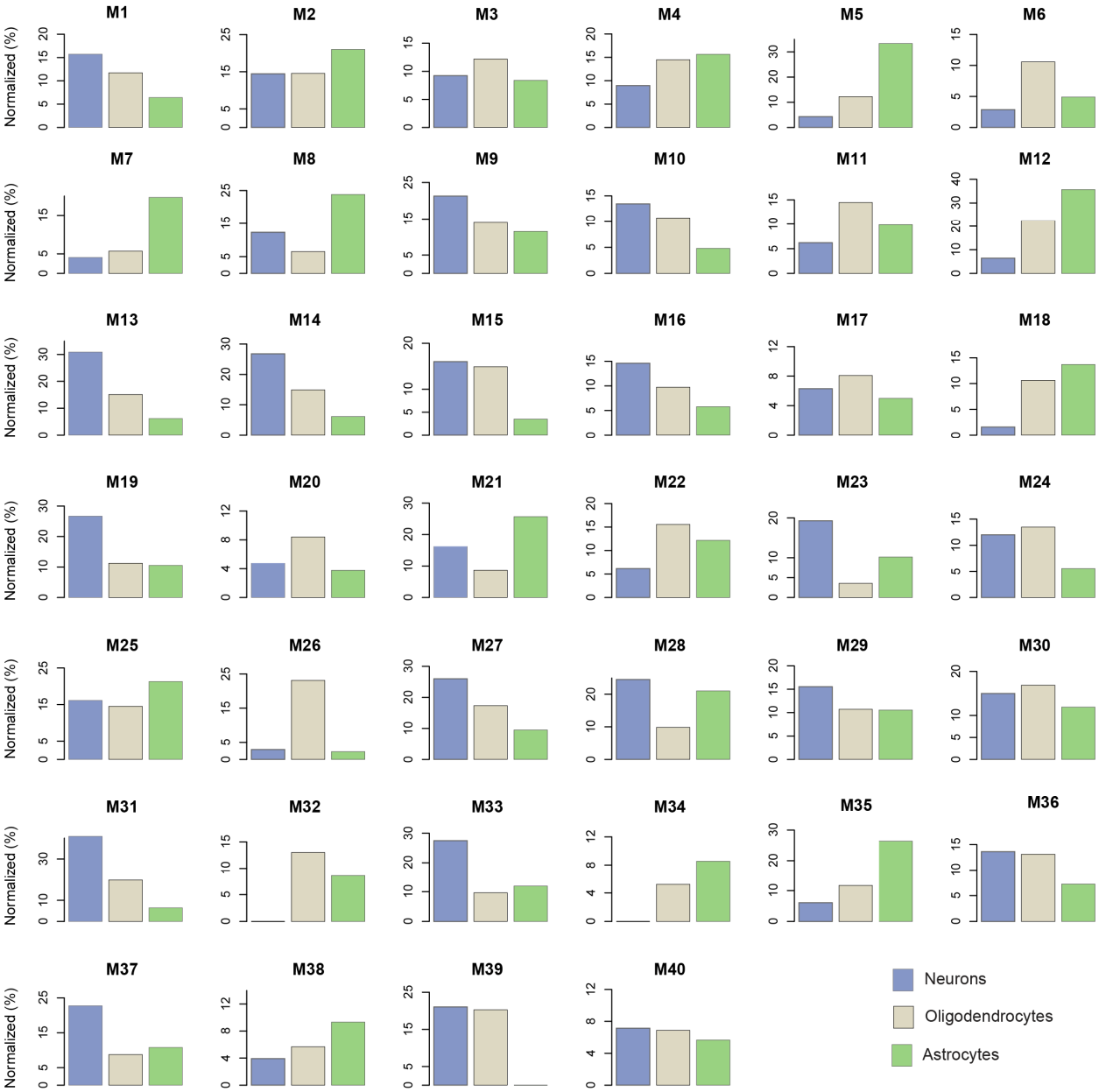

Figure S5. Neural Cell Type Enrichment per Module, Related to Figure 4

In Cahoy et al. 2008, around six thousand genes were reported to be enriched in neurons, oligodendrocytes and astrocytes. These genes were intersected with our gene set of each module. The proportions of intersected genes were normalized to the same scale across cell types to facilitate the visualization of cell type specific enrichment per module. Additionally, the enrichment for each and between cell types were estimated by binomial and chi-square statistic test.

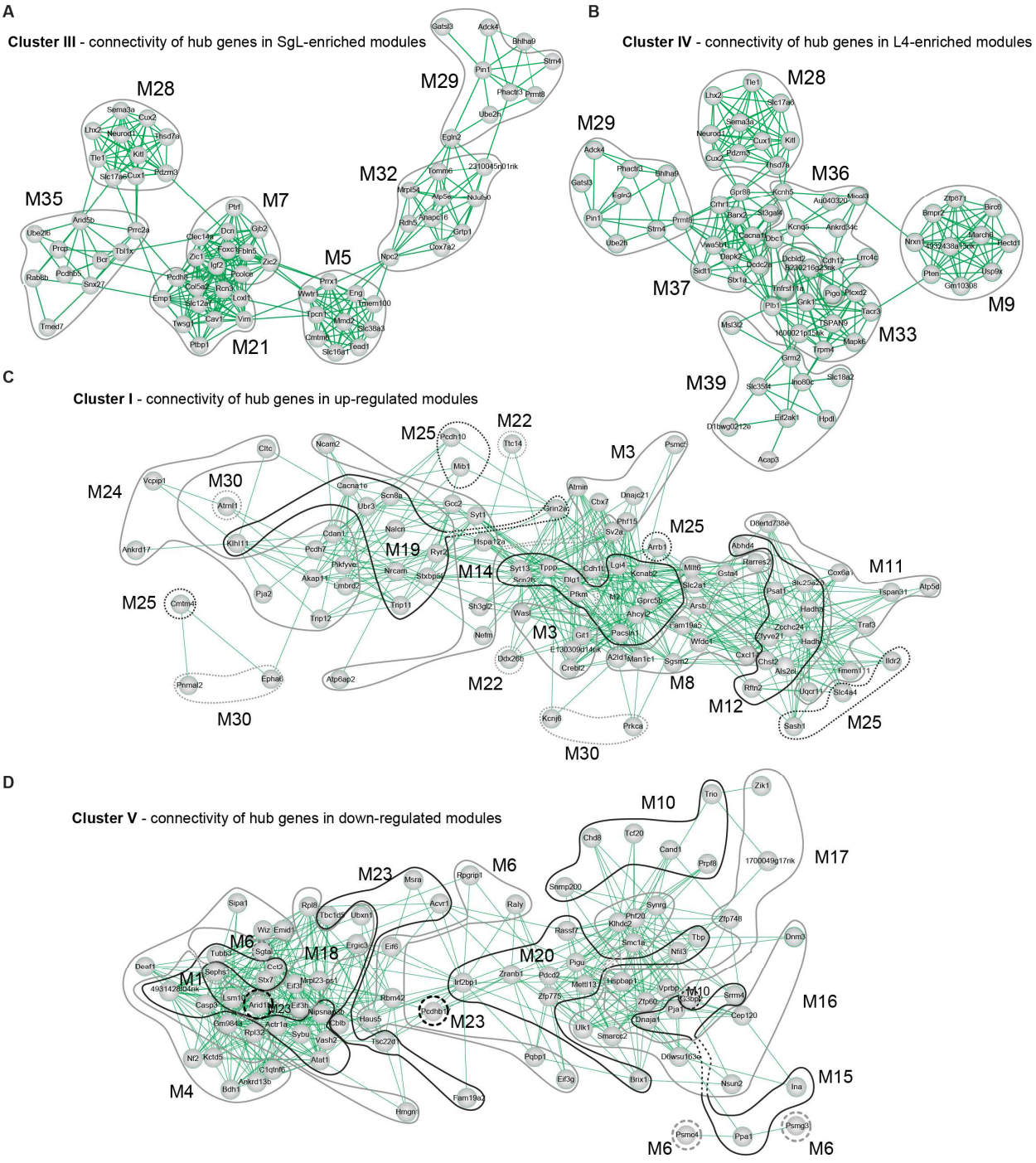

Figure S6. Inter- and Intra-Modular Hub Genes Connectivity per Cluster, Related to Figure 5

(A-D) The inter- and intra-modular hub genes connectivity for four additional clusters as labeled. In each cluster, we calculated the Pearson correlation between hub genes. Gene pairs with correlation coefficients larger than 0.7 were chosen for network visualization. Cytoscape (Smoot et al., 2011) was used to visualize this network, in which genes were depicted as circles and the correlated genes were connected by lines. The "un-weighted force-directed layout" parameter was used to optimize the network visualization.

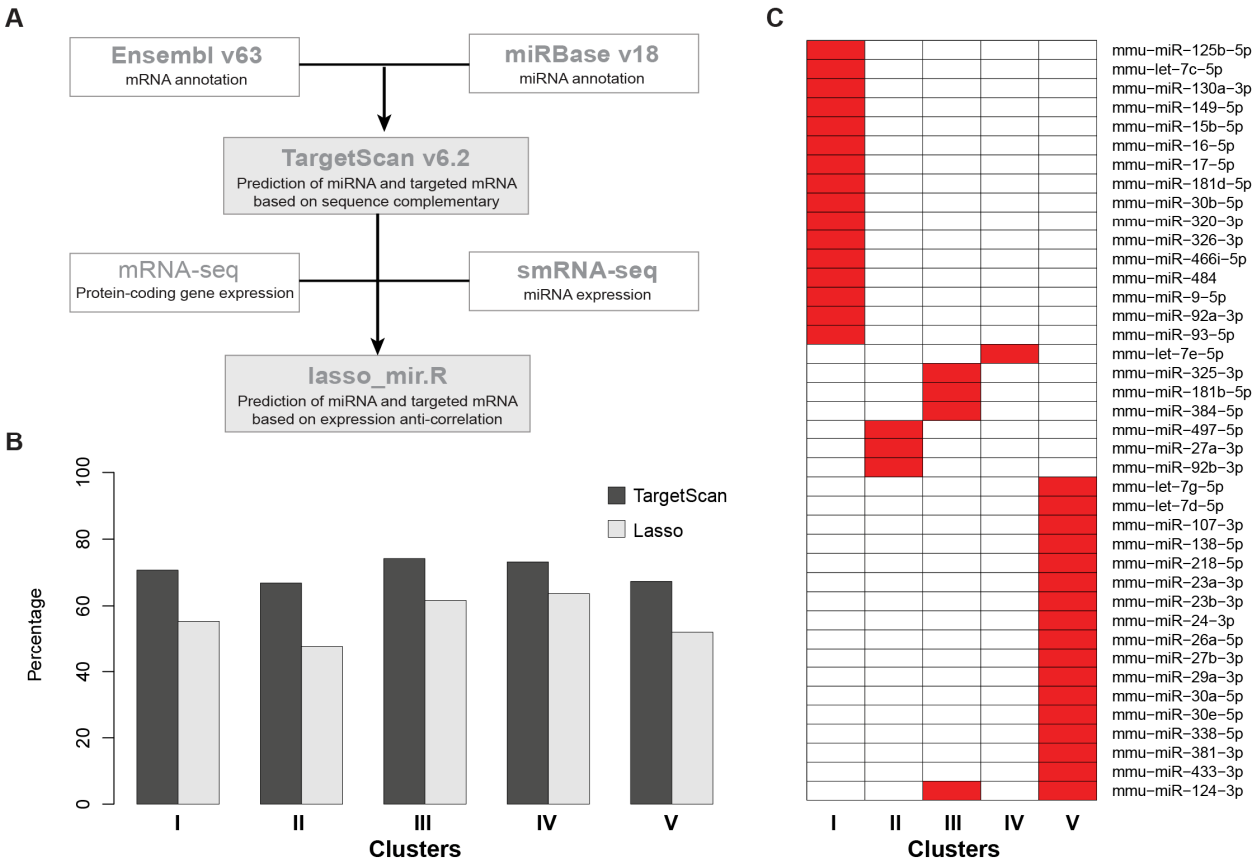

Figure S7. miRNA-mRNA Regulation Prediction, Related to Figure 6

(A) Pipeline for miRNA-mRNA regulation prediction. The white squares indicate the required data for running pipeline. The gray squares represent the integrated software packages.

(B) Percentage of mRNAs targeted by miRNA per cluster. The black-colored bars show the proportions of mRNAs targeted by miRNA in each cluster using TargetScan database. The grey-colored bars show the proportions of mRNAs targeted by miRNA in each cluster using software *lasso-mir.R*, which predicted the miRNA-mRNA regulation in the expression level with the combination of sequence-based prediction from TargetScan database.

(C) miRNA cluster enrichment. The red-colored squares correspond to the enriched miRNA in vertical axis and to the associated cluster in horizontal axis.

SUPPLEMENTAL TABLES

Supplemental Tables are provided in a single Microsoft Excel file

Table S1. Sample Metadata, Spike-In RNAs and Human Donor Metadata, Related to Figure 1

(A)This table summarizes the information for 36 samples and high throughput sequencing, providing age (range is P4 to adult), layer (SgL, L4 and IgL), sex (6 males and 6 females), and RNA integrity number (RIN, range is 8.5 to 10) for each individual sample. For the sequencing information, it provides the spike-in master (tagging mRNA-seq samples), TruSeq index (barcode), HiSeq lane and uniquely mapping reads for mRNA-seq (median > 10 million) and smRNA-seq (median > 1 million) data.

(B) List of spike-in RNAs used for mRNA-seq sample tagging and sequencing quality control. In total, 10 spike-in RNAs with different concentrations were used. The length and GC content were provided for each spike-in RNA. Additionally, the 18 pairs of spike-in RNAs used for sample tagging were provided.

(C) Metadata from human tissue is presented, including specimen ID, age (range is 12PCW to 30 year), sex (4 males and 6 females), postmortem interval (PMI, which ranged from 2 to 20 hours), pH of cerebellum (range is 6 to 6.92), ethnicity, cause of death, and medical history.

Table S2. Differentially Expressed Genes and miRNAs, Related to Figure 2

(A) Differentially expressed protein-coding genes between layers at any age and between ages at any layer. These DEX genes were identified from reliably expressed protein-coding genes (RPKM ≥ 1 in at least 2 samples) and the difference threshold is adjusted p-value (FDR < 0.01).

(B) Differentially expressed miRNAs between layers at any age and between ages at any layer. These DEX miRNAs were identified from reliably expressed miRNAs (reads count ≥ 10 in at least 2 samples) and the difference threshold is adjusted p-value (FDR < 0.01).

Table S3. Gene Function Enrichment Analyses for sDEX, tDEX, stDEX and DEX for Three Stages, Related to Results and Discussion

(A-C) The tables summarize the gene function enrichment analyses for 3034 spatiotemporal differentially expressed genes. It was generated using DAVID Bioinformatics Resources (http://david.abcc.ncifcrf.gov/). The associated gene list and statistical enrichment significance assessment were provided for each function term. (D)The hierarchical sample clustering analysis showed three isolated stages: P4, P6 to P10, and P14 and adult (Figure S3A). The gene function enrichment analyses, generated using DAVID Bioinformatics Resources, were performed for spatial DEX genes in each layer and each stage. The associated gene list and statistic enrichment significance assessment were provided for each function term.

Table S4. Alternative Spliced Genes, Related to Figure 3

Investigation of five basic splicing modalities: cassette exon(s), mutually exclusive exon(s), alternative donors or acceptors, alternative first or last exon, and intron retention. The first four modalities were identified by software JuncBASE with the requirement of > 25 supported reads and FDR < 0.01 as the difference threshold. The intron retention events were identified by software RSEQtools with the requirement of intronic RPKM ≥ 0.5, intronic reads count ≥ 10, and RPKM fold change > 2 as the difference threshold. The differentially alternative spliced (DAS) genes were surveyed between layers at any age and between ages at any layer.

Table S5. WGCNA Modules, Related to Figure 4

List of 40 modules identified by weighted gene co-expression network analysis (WGCNA). The reliability of detected modules was checked/re-organized by custom R script to ensure that a gene had its largest correlation with its modular eigengene. The top 10 genes highly correlated with modular eigengenes were chosen as the modular hub genes. The gene function enrichment analyses for 40 modules generated using DAVID Bioinformatics Resources. The associated gene list and statistic enrichment significance assessment were provided for each function term.

Table S6. miRNA-mRNA Interaction Pairs, Related to Figure 6

List of predicted interaction pairs of miRNA-mRNA by software *lasso_mir.R* with using the sequenced-based prediction from TargetScan database and the expression values for mRNA and miRNA. The pairs with rank score ≥ 80 were chosen as reliable miRNA-mRNA interactions.

Table S7. Signature Trajectory of Neocortical Developmental Events, Related to Figure 7

List of genes used to study the signature trajectory of neocortex developmental events concerning neuronal morphology, synaptogenesis and activity. Genes found to be enriched in certain layer(s) are indicated accordingly.

EXTENDED EXPERIMENTAL PROCEDURES

1. Introduction

In this Supplemental Information we provided technical descriptions of data generation and analyses. We also made available additional data that were discussed in the main manuscript. Finally, we presented supplemental figures and tables generated from sample metadata and specific gene lists.

2. Laminar Microdissection

We took advantage of the *Dcdc2a-Gfp* reporter mouse generated by the GENSAT project (Heintz, 2004) in which GFP was preferentially expressed in the postnatal L4 starting around P2. Mice were decapitated and the brain was removed immediately from the skull and submerged in ice-cold oxygenated artificial cerebrospinal fluid (ACSF) for 5 min. Using a vibratome (Leica VT 1200 S) we prepared live sagittal brain slices (250 μm thick), keeping the cutting brain and cut slices submerged in oxygenated ACSF at ice-cold temperature at all times to reduce enzymatic activity and RNA degradation. The dissection of SgL (L2/3 including marginal zone or L1, and pia), L4 and IgL (L5/6 including subplate or white matter) was performed on each 250 μm slice, while keeping the section submerged in ice-cold ACSF inside a petri dish, using a scalpel and a dissecting needle under a fluorescence stereoscope (Discovery V8, Zeiss). Each sample was collected into a safe-lock microcentrifuge tube (Eppendorf) with 30 μL of RNAlater®, for subsequent total RNA extraction.

3. RNA Isolation, Library Preparation and Sequencing

3.1 RNA Extraction

A bead mill homogenizer (Bullet Blender, Next Advance) was used to homogenize the tissue. Each tissue sample was transferred to a safe-lock microcentrifuge tube (Eppendorf). A mass of stainless steel beads (Next Advance, cat# SSB14B) equal to the mass of the tissue was added to the tube. Two volumes of lysis buffer were added to the tube. Samples were mixed in the Bullet Blender for 1 min at a speed of six. Samples were visually inspected to confirm desired homogenization and then incubated at 37 ℃ for 5 min. The lysis buffer was added up to 0.6 ml, and samples were mixed in the Bullet Blender for 1 min. Total RNA was extracted using RNeasy Plus Mini Kit (Qiagen). Optical density values of extracted RNA were measured using NanoDrop (Thermo Scientific) to confirm an A_{260}:A_{280} ratio above 1.9. RIN was determined for each sample using Bioanalyzer RNA 6000 Nano Kit or Bioanalyzer RNA 6000 Pico Kit (Agilent) (Table S1A), depending upon the total amount of RNA.

3.2 Library Preparation and Sequencing for mRNA

The cDNA libraries were prepared using the TruSeq mRNA Sample Prep Kit (Illumina) as per the manufacturer's instructions with some modifications. Briefly, polyA RNA was purified from 1 μg of total RNA using oligo (dT) beads. Quaint-IT RiboGreen RNA Assay Kit (Invitrogen) was used to quantitate purified mRNA with the NanoDrop 3300 (Thermo Scientific). Following mRNA quantitation, 2.5 μl spike-in master mixes, containing five different types of RNA molecules at varying amount (2.5×10^{-7} to 2.5×10^{-14} mol), were added per 100 ng of mRNA (Jiang et al., 2011), (Tables S1A and S1B). The spike-in RNAs were synthesized by External RNA Control Consortium (ERCC) by *in vitro* transcription of *de novo* DNA sequences or of DNA derived from the *B. subtilis* or the deep-sea vent microbe *M. jannaschii.* Both genomes were generous gifts of Dr. Mark Salit at The National Institute of Standards and Technology (NIST). Each sample was tagged by adding a pair of spike-in RNAs unique to the region from which the sample was taken. Also, an additional three common spike-ins were added for controlling sequencing error rates. Spike-in sequences are available at http://www.hbatlas.org/files/spike_in.fa. The mixture

of mRNA and spike-in RNAs was subjected to fragmentation, reverse transcription, end repair, 3'– end adenylation, and adaptor ligation, followed by PCR amplification and SPRI bead purification (Beckman Coulter). The unique barcode sequences were incorporated in the adaptors for multiplexed high-throughput sequencing. The final product was assessed for its size distribution and concentration using Bioanalyzer DNA 1000 Kit (Agilent). For mRNA-sequencing, 6 libraries were pooled per HiSeq lane sequencing (Table S1A) and diluted to 10 nM in EB buffer (Qiagen) and then denatured using the Illumina protocol. The denatured libraries were diluted to 15 pM, followed by cluster generation on a single-end HiSeq flow cell (v1.5) using an Illumina cBOT, according to the manufacturer's instructions. The HiSeq flow cell was run for 75 cycles using a single-read recipe (v2 sequencing kit) according to the manufacturer's instructions.

3.3 Library Preparation and Sequencing for Small RNA

The TruSeq Small RNA Sample Kit (Illumina) was used to prepare cDNA libraries as per the manufacturer's instructions. Briefly, 1 µg of total RNA was ligated with 3'– and then 5'– adapters, followed by reverse transcription and PCR amplification. The PCR utilizes 48 different types of primer that will add 48 different index sequences to the adapters. Each library was assessed for the presence of desired miRNA population by Bioanalyzer High Sensitivity DNA Kit. We pooled 20 and 19 samples (Table S1A) with distinct indexes, which allow subsequent retrieval of each sample from multiplexed sequencing runs. Each pooled library was size-selected by gel excision. The final product was assessed for its size distribution and concentration using Bioanalyzer DNA 1000 Kit. The library was diluted to 10 nM in EB buffer and then denatured using the Illumina protocol. The denatured libraries were diluted to 15 pM, followed by cluster generation on a single-end HiSeq flow cell (v1.5) using an IlluminacBOT, according to the manufacturer's instructions. The HiSeq flow cell was run for 75 cycles using a single-read recipe (v2 sequencing kit) according to the manufacturer's instructions.

4. mRNA Sequencing Data Analyses

4.1 Reads Filtering, Processing and Alignment

Reads passed the default purify filtering of Illumina CASAVA pipeline (released version 1.7) and were aligned to mouse reference genome (NCBI37/mm9, July 2007), which was downloaded from the UCSC genome browser. We trimmed one nucleotide in 5'-end, leaving 74 nucleotide long reads for sequence alignment. We used Tophat (v1.0.13) to align the reads onto mouse reference genome (Trapnell et al., 2009). We set multiple hit parameter (-g) equal to 1. Reads having multiple genomic hits were excluded, using solely uniquely mapped reads for all downstream analyses. We additionally changed the default 25 segment length to a larger 35 segment length to decrease the possibility of reads having multiple mapping. The segments are mapped independently, allowing up to 2 mismatches in each segment alignment.

4.2 Quality Assessment

As described below, several quality control measures were implemented throughout sample preparation and transcriptome data generation steps.

4.2.1 Spike-In Control RNAs

As mentioned in section 3.2, we used different combinations of two spike-in RNAs to tag different samples (Table S1B). This tagging avoids mixing samples during library preparation or sample loading into the sequencer. In addition, three more common spike-in RNAs were added, combining with the two tagging spike-in RNAs to estimate the sequencing error rate of the sequencer. Since we knew the exact sequences of spike-in RNAs, the mismatches within sequenced reads from the reference spike-in RNAs were considered sequencing errors by the sequencer. Thus, the percentage of mismatches was used to quantify the sequencing error rate.

In principal, the sequencing error rate depends on the number of sequencing cycles. It is expected that the longer the read length is, the higher the error rate will be. The results showed that the median values of the percentage of mismatches in each sequencing cycle were less than 2%, which indicated the high quality of the sequenced reads. Also, we found slightly higher error rates at both ends of the reads, which meant that both PCR and other machine related issues were negligible (Figure S1A).

4.2.2 Transcribed Fraction of Genome/Exon across Each Chromosome

The mouse reference genome is composed of nineteen autosomal chromosomes, one mitochondrial chromosome and two sex chromosomes. The sequenced reads were aligned onto the mouse reference genome to calculate the percentage of reads mapping to each chromosome. In principal, the percentage should be proportional to chromosome length because of the expected uniformity of reads mapping to the genome (Figure S1D). Furthermore, we considered a genomic nucleotide to be transcribed when it was aligned by at least one read. The genome transcription ratio per chromosome was quantified by the proportion of transcribed nucleotides to the chromosome length (Figure 1F). Similarly, the ratio of transcribed exons per chromosome was quantified by the proportion of transcribed nucleotides to the sum of exon length within the chromosome (Figure 1F). The non-exon mapping reads were assigned to introns and intergenic regions (Figure S1E). Together, these data enabled the investigation of mRNA-seq genome uniformity and transcriptome coverage.

4.2.3 Layer Specific Marker Genes

A list of known and characterized layer specific marker genes were previously published (Chen et al., 2005; Kwan et al., 2008; Nakagawa and O'Leary, 2003; Nieto et al., 2004). Additionally, we used a recent mRNA-seq data from Belgard et al, 2011 for further characterization of layer specific markers (Figures S1B and S1C) (Belgard et al., 2011)

4.2.4 Hierarchical Clustering and Principal Component Analysis

We used two approaches to do clustering analysis in order to ensure high quality of mRNA-seq data. Firstly, the "reads per kilobase of exon model per million mapped reads" (RPKM) values of reliably expressed protein-coding genes (see section 4.3) were log2 transformed and quantile normalized across all samples by R bioconductor *limma* (Smyth, 2004). Subsequently, we used R package WGCNA to perform average-linkage hierarchical clustering (Hclust) of all samples based on the processed gene RPKM data by using 1-*r* as distance (or dissimilarity), where *r* was Spearman correlation. Hierarchical clustering revealed notable clustering of the samples firstly by ages and then by layers, which were represented by the different color bars under the cluster (age: orange (P4) yellow (P6) green (P8) blue (P10) violet (P14) black (adult); layer: blue (SgL), green (L4), red (IgL) (Figure S3A). There was no obvious sex clustering. The RIN values of the samples were all high and do not show any clustering (Figure S3A). On the other hand, we used R package *prcomp* to perform principal components analysis (PCA) on the processed gene RPKM data. Taking account of the comparable amount of attribution between PC2 and PC3, the plots between PC1 and PC3 were provided (Figure S2) in addition to the plots between PC1 and PC2 (Figures 2A and S2). Furthermore, each sample was differentially colored according to its sex, i.e. female (red) and male (blue). There was no significant clustering by sex (Figure S2).

4.3 Gene and Exon Expression

The *gtf* formatted gene/exon/transcript annotations of mm9 reference genome were retrieved from Ensembl (NCBI37/mm9, released version 63). In total, 36,814 genes were annotated, including 22,667 protein-coding genes and 14,147 non-protein-coding genes. In this work, the mRNA-seq data analyses only focused on protein-coding genes because of the greater

complexity and lower global expression values of non-protein-coding genes (Cabili et al., 2011).. Hereafter, "gene" refers to a protein-coding gene. We used RSEQTools to calculate the gene expression (Habegger et al., 2011). Briefly, we first used the *mergeTranscripts* program to generate the composite gene model in which the exons from different transcript isoforms were merged into the longest gene. Next, we used the program *mrfQuantifier* to calculate the reads count per gene, as well as the RPKM for each gene (Mortazavi et al., 2008). To reduce false positive results, we focused on reliably expressed genes, which were defined by the requirement of RPKM ≥1 in at least 2 samples (Figure S2B). After these filters we obtained 12,729 reliably expressed protein-coding genes for the downstream analyses.

4.4 Spatiotemporal DEX Genes

We used R package DESeq to identify differentially expressed (DEX) genes (Anders and Huber, 2010). The DEX genes were detected from the reliably expressed protein-coding genes. We did not investigate genes with low expression because they were prone to be affected by background noise. As mentioned in section 4.3, we generated two types of gene expression value. The reads count per gene served as the input for DESeq. Also, we used one male sample and one female sample for each layer at age. They were treated as biological replicates to improve the reliability of DEX genes because DESeq was more reliable at comparing groups with replicates. When performing the comparison, DESeq first gets the mean expression level as a joint estimate for both groups, and then calculates the difference as well as the p-value for the statistical significance of this change. The adjusted p-value was also calculated based on multiple testing with the Benjamini-Hochberg procedure, estimating the false discovery rate (FDR). In this work we set one stringent criterion, FDR < 0.01, so as to detect statistically significant DEX genes.

The DEX genes can be split into two types, spatial DEX (sDEX) genes and temporal DEX (tDEX) genes. Genes that were differentially expressed in at least one layer at any given age were defined as spatial DEX genes. Similarly, genes that were differentially expressed in at least one age in any given layer were defined as temporal DEX genes. To identify spatial DEX genes, we performed pairwise comparisons across layers in P4, P6, P8, P10, P14 and adult. The total spatial DEX genes were the summation of all pairwise-layer DEX genes. To identify temporal DEX genes, we did pairwise comparisons across ages in SgL, L4 and IgL. The total temporal DEX genes were the summation of all pairwise-age DEX genes (Table S2A). We found that most of spatial DEX genes were temporally regulated, and half of temporal DEX genes were spatially regulated. Only a small number of spatial DEX genes were evenly expressed throughout all ages. These genes were reported as layer specific markers (see section 4.2.3). So, the sDEX genes were purified to genes that were differentially expressed between layers but not between ages, the tDEX genes were purified to genes differentially expressed between ages but not between layers. We introduced the third type of DEX genes, spatiotemporal DEX (stDEX) genes, to represent genes that were differentially expressed among layers and ages.

4.5 Alternative Splicing and Intron Retention

Reads mapping to exon-exon junctions allowed us to study gene splicing events. Although some papers propose the use of statistical tools to estimate the expression of the alternate gene transcripts, this cannot compare in accuracy to the use of detected exon-exon junction reads providing evidence of splicing events that actually occurred. In this work, we used JuncBASE (Junction Based Analysis of Splicing Events) to identify the six different types of alternative splicing (AS) events: cassette exon, alternative donor, alternative acceptor, mutually exclusive exons, alternative first exons, and alternative last exons (Brooks et al., 2011). After performing the JuncBASE pipeline, we set two criteria to select statistically significant alternative splicing events. For the AS events expressed in both conditions, we used a threshold of FDR < 0.01. For the AS events expressed only in one condition, we used the supported reads count > 25 as the

threshold. Similar to the analysis of DEX, we performed alternative splicing analysis (AS) across layers and ages. We combined the biological replicates to one sample because JuncBASE currently cannot handle replicates as separate inputs. We performed pairwise comparisons between different layers and between different ages. The differentially alterative splicing (DAS) found between layers but not between ages were defined as spatial DAS (sDAS), the DAS found between ages but not between layers were defined as temporal DAS (tDAS), and the DAS found between ages as well as between layers were defined as spatiotemporal DAS (stDAS) (Table S4).

We used RSEQtools to build intron annotations and calculated the RPKM and reads count. The expression values were used to estimate the retention of introns. To set the proper threshold, we simulated the distribution of intronic RPKM and reads count for the all potentially expressed introns, of which at least one read hits in the introns. The combination of RPKM ≥ 0.5 and raw reads count ≥ 10 were used as the threshold (Figures S2D and S2E). In additional, we did the comparison of intron retention events between layers and between ages. We used RPKM fold change > 2 to detect spatial and temporal different intron retention events (Table S4).

4.6 Weighted Gene Co-Expression Network Analysis

We used R package WGCNA to perform weighted gene co-expression network analyses for the reliably expressed protein-coding genes (Langfelder and Horvath, 2008). In terms of the calculation details, the gene RPKM values were firstly processed by log2 transformed quantile normalization across all samples by R bioconductor *limma (*Smyth, 2004). Next, we used *pickSoftThreshold* function to analyze the network topology and chose 4 as soft-threshold power. In order to do automatic network construction and module detection, we used *blockwiseModules* function to generate signed network. Modules with fewer than 10 genes were merged to their closest larger neighbor module. In total, this analysis produced 40 modules (Table S5). For each module, WGCNA generated the eigengene to characterize the modular feature. To check the reliability of detected modules, the custom R script was used to calculate the correlation between gene and modular eigengene. Genes would be re-assigned to another module to ensure each gene having the largest correlation coefficient with its own modular eigengene. Overall, only a small number of genes were re-assigned to other modules. We used the *moduleEigengenes* function to re-calculate the eigengene for the changed modules. To better understand the modular feature, we plotted the trajectories of the modular eigengenes (Figure S4), which were shown the same spatiotemporal patterns as heat map characterization of gene expression in each module. The module M40 with 14 genes clearly showed a male bias (i.e. Y chromosome or male enriched genes). For the other 39 modules, we found some modules showed similar patterns. So, we performed hierarchical clustering analysis and found the 39 modules can be classified into 5 clusters (Figures 4A, 4B, S4 and the following table). In summary of the 5 clusters, the first and fifth clusters were temporal related clusters, i.e. up and down regulated along development, and the other three clusters were spatial related clusters, i.e. separately enriched in certain layers.

Cluster	Feature	Module
I	Up regulated	M3, M22, M2, M19, M14, M24, M8, M30, M25, M12, M11
II	IgL enriched	M26, M34, M31, M13, M27
III	SgL enriched	M7, M21, M35, M28, M29, M5, M32
IV	L4 enriched	M36, M37, M33, M39, M9
V	Down regulated	M38, M15, M16, M1, M18, M4, M23, M17, M20, M6, M10

Moreover, we used network methods to visualize the relationships between modules within a cluster. In each module, we chose the top 10 genes highly correlated with the eigengene as the module hub genes (Table S5). In each cluster, we calculated pairwise Pearson correlations among hub genes. Gene pairs with correlation coefficients larger than 0.7 were chosen for network visualization. The free software Cytoscape (Smoot et al., 2011) was used to present networks in which genes were depicted as circles and, correlated genes were connected by lines. The "un-weighted force-directed layout" parameter was used to optimize network visualization (Figures 5C and S6). When looking closer at the connectivity network, we noticed that in each cluster, one or two modules, typically associated with a specific biological process, received more inter-modular connections than others, as if becoming the focus of the molecular interactions happening in that given spatial or temporal cluster. We define them as "core module" of cluster. The core modules are: cluster I (M2), cluster II (M13) cluster III (M7 and M21), cluster IV (M37), and cluster V (M1 and M16) (Figures 5B and S6).

5. Small RNA Sequencing Data Analysis

5.1 Sequencing Quality Assessment

Reads passed the default purify filtering of Illumina CASAVA pipeline (released version 1.7), and were input to software FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to perform the sequencing quality assessment of small RNA-seq (smRNA-seq) samples. FastQC generated a comprehensive report on the composition and quality of the sequenced reads. The per-base quality plot shows an overview of the range of quality scores across all bases at each position and then provides an easy way to justify the sequencing quality. In our samples the majority Phred scores were above 30 in each sequencing cycle, suggesting high sequencing quality of the smRNA-seq samples (Figure S1F).

5.2 Adapter Trimming and Sequence Alignment

The FASTX-Toolkit provides a series of commands to preprocess the sequencing reads (http://hannonlab.cshl.edu/fastx_toolkit/download.html). We used *fastx_clipper* command to clip the Illumina small RNA 3' adapter (TGGAATTCTCGGGTGCCAAGG) in 3'– end of reads. We discarded these clipped reads with lengths shorter than 15 nucleotides and larger than 49 nucleotides. The length distribution of the clipped reads was definitely enriched for miRNA length, i.e. 22 nucleotides (Figure 1G). Subsequently, we used software *miRanalyzer* (released version 0.2) and some PERL scripts together to build the miRNA analysis pipeline (Hackenberg et al., 2011; Hackenberg et al., 2009). First, we used PERL script *groupGAreads.pl* to collapse the clipped reads to one multi-fasta format file. Second, we retrieved the annotation of miRNA and pre-miRNA from miRBase (released version 18) (Kozomara and Griffiths-Jones, 2011) and used *bowtie-build* function in software Bowtie to build annotation libraries. Finally, *miRanalyzer* drove Bowtie to orderly align the collapsed reads to miRNA and pre-miRNA. The reads uniquely aligned to either library or unique annotation entries were taken for downstream analyses.

5.3 Expression of miRNA

The mouse miRNA annotations were retrieved from miRBase (released version 18), in which 741 pre-miRNA and 1,157 miRNA were deposited. We found 29 miRNA with identical sequences. They were merged to 13 unique mature miRNA, and then 1,141 miRNA were used as the annotation database. The sequence of miRNA was the consensus sequence in many different experiments (http://www.mirbase.org/). Reads aligned to pre-miRNA but not miRNA were attributed to divergence from the consensus sequence. To better estimate the expression levels of miRNA, reads aligned to pre-miRNA were assigned to miRNA based on its aligned locus, i.e. reads closer to the 3'– end were assigned to the annotated 3'– miRNA, and so on. In this work we focused only on the known miRNA, and reads aligned to other genomic regions were ignored. Moreover, we mainly focused on reliably expressed miRNA, which were defined

as having reads count ≥ 10 in at least 2 samples (Figure S2C). The analysis of miRNA with fewer aligned reads would increase the possibility of false positive results. After these stringent filters, we obtained 436 reliably expressed miRNA.

5.4 Principal Components Analysis

We performed principal component analysis to ensure the high quality of the smRNA-seq data. To remove any possible external effects, the reads per million mapped reads (RPM) values of all reliably expressed miRNA were processed by log2 transformed quantile normalization across all samples by R bioconductor *limma* (Smyth, 2004). We used R package *prcomp* to perform principal components analysis (PCA) on the processed miRNA RPM data. The plots between PC1 and PC3 were provided in addition to the plots between PC1 and PC2 (Figures 2B and S2). Furthermore, each sample was differentially colored according to sex (female red, and male blue). The mixing of female and male samples was the evidence of litter sex effect (Figure S2).

5.5 Spatiotemporal DEX miRNA

We used R package DESeq to identify differentially expressed (DEX) miRNA (Anders and Huber, 2010). The DEX miRNA was detected from the reliably expressed miRNA. We did not investigate miRNA with low expression because they were prone to be affected by background noise. The reads count per miRNA served as the input for DESeq. Also, we used one male sample and one female sample for each layer in any age. They were treated as biological replicates to improve the reliability of DEX miRNA identification because DESeq was more reliable to compare groups with replicates. When performing the comparison, DESeq firstly gets the mean expression level as a joint estimate for both groups, and then calculates the difference as well as the p-value for the statistical significance of this change. The adjusted p-value was also calculated based on multiple testing with the Benjamini-Hochberg procedure, estimating the false discovery rate (FDR). In this paper, we set one stringent criteria, FDR < 0.01, so as to detect reliable DEX miRNA.

The DEX miRNA can be split into two types, spatial DEX miRNA and temporal DEX miRNA. miRNA that were differentially expressed in one layer at any given age were defined as spatial DEX miRNA. Similarly, miRNA that was differentially expressed in one age in any given layer were defined as temporal DEX miRNA. To identify spatial DEX miRNA, we did pairwise comparison across layers in P4, P6, P8, P10, P14 and adult. The total spatial DEX miRNA were the summation of all pairwise-layer DEX miRNA. To identify temporal DEX miRNA, we did pairwise comparison across ages in SgL, L4 and IgL. The total temporal DEX miRNA were the summation of all pairwise-age DEX miRNA (Table S2B). We found that most of spatial DEX miRNAs were temporally regulated, and some temporal DEX miRNAs were spatially regulated. Similar to the classification of DEX genes, the sDEX miRNAs were purified to miRNAs that were differentially expressed between layers but not between ages, the tDEX miRNAs were purified to miRNAs differentially expressed between ages but not between layers. We introduced the third type DEX miRNAs, spatiotemporal DEX (stDEX) miRNAs, to represent miRNAs that differentially expressed between layers as well as between ages.

5.6 miRNA-mRNA Regulation Prediction

In general the targets of miRNA in mammalians are predicted by seed sequence of miRNA complementary to the 3'- UTR regions of mRNA (Lewis et al., 2005). It is also recommended that Pearson correlation analysis be used to dig the anti-correlation relationship between miRNA and its targets in the expression level (Hsu et al., 2011; Xiao et al., 2009). In this paper, we combined the sequence and expression information together to investigate the regulation relationship between miRNA and its mRNA protein-coding gene targets (mRNA hereafter) (Figure S7A). In terms of the analysis details, we first retrieved miRNA-mRNA regulation pairs

from TargetScan database (released version 6.2) (Grimson et al., 2007), in which all regulated pairs were predicted based on the complementary sequence. To reduce false positive results, we restricted the analyses to miRNA and its conserved targets, which were deposited in the file of "Conserved_Site_Context_Scores.txt ". In total, 12,558 mRNAs and 771 miRNAs were included. Second, we used software *lasso_mir.R* to predict the regulation between mRNA and miRNA expression levels (Lu et al., 2011). This approach allowed us to use the Lasso regression model for the identification of miRNA-mRNA targeting relationships that combines sequence based prediction information, miRNA co-regulation, RISC availability, and mRNA/miRNA abundance data. We used a stringent criteria with rank score ≥ 80 to choose reliable miRNA-mRNA regulation pairs (Table S6). To visualize the regulation relationship between the miRNA and all of its targeted mRNAs, the RPM values of miRNAs and the RPKM values of mRNAs were normalized to the same scale by using *normalize* function in R package *som*. The R package *lowess* were used to build smoothed curves for the normalized values of miRNA and targeted mRNAs, profiling the anti-correlation relationship.

To further understand miRNA-mRNA regulation relationship, we analyzed the co-expressed mRNAs. We had already identified 5 clusters, in which mRNAs showed the similar spatiotemporal pattern (see section 4.6). We then tried to determine how these patterns were contributed to by miRNA regulation. We counted how many mRNAs were potentially regulated by at least one miRNA in each cluster. The proportion of targeted mRNAs was calculated separately using TargetScan database and Lasso prediction (Figure S7B). Moreover, we tried to find the cluster-enriched miRNAs in the hope of identifying the main regulators/contributors to the characteristic spatiotemporal pattern of the cluster (Figure S7C). Briefly, we first collected the miRNAs per cluster in which miRNA has at least one targeted mRNA included in the cluster. Next, we performed an enrichment test by comparing the targeted mRNAs in the cluster with its expected targeted mRNAs. The p-values were calculated for each miRNA to quantify statistical significance of enrichment. The adjusted p-value was calculated based on multiple testing with the Benjamini-Hochberg procedure, estimating the false discovery rate (FDR). We used one stringent criterion, FDR < 0.01, to detect reliable cluster-enriched miRNAs (Figure S7C).

6. Validation of Gene/Exon Expression by PCR

6.1. Quantitative Real Time RT-PCR

An aliquot of the total RNA that was previously extracted from each brain region was used for secondary validation through real-time PCR analysis. One µg of total RNA was used for cDNA synthesis using SuperScript III First-strand synthesis Supermix (Invitrogen) and subsequently diluted with nuclease-free water. TaqMan Gene Expression Assay was used for each gene of interest along with TaqMan Universal Master Mix (Applied Biosystems). PCR reactions were conducted on an ABI 7900 Sequence Detection System (Applied Biosystems) and the resulting Ct value (cycle number at threshold) was used to calculate the relative amount of mRNA molecules. The Ct value of each target gene was normalized by subtraction of the Ct value from Gapdh to obtain the ΔCt value. The relative gene expression level was shown as $2^{-\Delta Ct}$.

6.2. Exon Specific PCR

The same cDNA solution was used for the mouse *Dlg2* PCR. For human *DLG2*, cDNA was synthesized in the same manner for human total RNA of the somatosensory neocortical area at comparative developmental periods to the current study (period3: Early fetal: 10-13 postconceptional weeks (PCW), period6: Late mid-fetal: 19-24 PCW, period7: Late fetal: 24-38 PCW, period 10: Early childhood: 1-6 years old, period13: young adulthood: 20-40 years old, sample details are described elsewhere (Kang et al., 2011). Exon-specific high-melting temperature primers were designed using NCBI/Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) and nucleotide sequences for each primer set are shown in the table below.

PCR was performed using Phusion High-Fidelity DNA Polymerase (NEB) under the following conditions: activation at 98ºC for 30 sec, followed by 30 cycles at 98℃ for 15 seconds, 68ºC for 30 sec, and 72ºC for 15 sec. PCR products were applied to an Agilent Bioanalyzer or Tapestation for quantification of each band that is specific to either inclusion or exclusion of an alternative exon.

Gene	Forward primer (5' -> 3')	Reverse primer (5' -> 3')
mouse Dlg2	CCCCGGATTAGGTGACGACGGT	TCCTGCCTCGTGACAGGTTCA
human DLG2	ATCCCCGGATTAGGTGACGA	CCTGCCTTGTAACAGGCTCA

SUPPLEMENTAL REFERENCES

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 25, 1915-1927.

Hackenberg, M., Sturm, M., Langenberger, D., Falcon-Perez, J.M., and Aransay, A.M. (2009). miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. Nucleic Acids Res. 37, W68-76.

Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M., et al. (2011). miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. 39, D163-169.

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. Genome Res 21, 1543-1551.
Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15-20.
Lu, Y., Zhou, Y., Qu, W., Deng, M., and Zhang, C. (2011). A Lasso regression model for the construction of microRNA-target regulatory networks. Bioinformatics 27, 2406-2413.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621-628.

Nakagawa, Y., and O'Leary, D.D. (2003). Dynamic patterned expression of orphan nuclear receptor genes RORalpha and RORbeta in developing mouse forebrain. Dev. Neurosci. 25, 234-244.

Nieto, M., Monuki, E.S., Tang, H., Imitola, J., Haubst, N., Khoury, S.J., Cunningham, J., Gotz, M., and Walsh, C.A. (2004). Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the cerebral cortex. J. Comp. Neurol. 479, 168-180.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3, Article3.

Xiao, F., Zuo, Z., Cai, G., Kang, S., Gao, X., and Li, T. (2009). miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. 37, D105-110.