Annotation and cluster analysis of spatiotemporal- and sex-related lncRNA expression in Rhesus macaque brain

Siling Liu[1, 2, 9], Zhengbo Wang[1, 9], Dong Chen[3, 9], Bowen Zhang[5, 9], Ren-Rong Tian[1,9], Jing Wu[1], Ying Zhang[1], Kaiyu Xu[1, 2], Liu-Meng Yang[1], Chaocheng[3], Jian Ma[1, 2], Longbao Lv[1,6], Yong-Tang Zheng[1,6], Xintian Hu[1,6,7], Yi Zhang[3,4, *], Xiangting Wang[5, 8, *], and Jiali Li[1, 6, *]

[1] Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences & Yunnan Province

Kunming Institute of Zoology, the Chinese Academy of Sciences

Kunming, Yunnan, 650223, China

[2]Kunming College of Life Science, University of Chinese Academy of Sciences

Kunming 650223, China

[3]Center for Genome Analysis, ABLife Inc., Wuhan 430075, P. R. China

[4]Laboratory for Genome Regulation and Human Health, ABLife Inc., Wuhan 430075, P. R. China

[5]School of Life science, CAS Key Laboratory of Brain Function and Disease, CAS Center for Excellence in Molecular Cell Science, School of Life Science, University

of Science and Technology of China, Hefei , Anhui 230026, China

[6]Kunming Primate Research Center of the Chinese Academy of Sciences

Kunming, Yunnan 650223, China

[7]CAS Center for Excellence in Brain Science and Intelligence Technology, China

[8]CAS Center for Excellence in Molecular Cell Science, China

[9] These authors contributed equally to this work

[*]Co-corresponding author

yizhang@ablife.cc

and

wangxt11@ust.edu.cn

and

lijiali@mail.kiz.ac.cn

Phone: 86-871- 65197043

Facsimile: 86-871- 65197043

Abstract

Long noncoding RNAs (lncRNAs) mediate important epigenetic regulation in a wide range of biological processes and diseases. We applied comprehensive analyses of RNA-seq and CAGE-seq (cap analysis of gene expression and sequencing) to characterize the dynamic changes in lncRNA expression in rhesus macaque (*Macaca mulatta*) brain in four representative age groups. We identified 18 anatomically diverse lncRNA modules and 14 mRNA modules representing spatial, age and sex specificities. Spatiotemporal- and sex-biased changes in lncRNA expression were generally higher than those observed in mRNA expression. A negative correlation between lncRNA and mRNA expression in cerebral cortex was observed and functionally validated. Our findings offer a fresh insight into spatial-, age- and sex-biased changes in lncRNA expression in macaque brain, and suggest that the changes represent a previously unappreciated regulatory system that potentially contributes to brain development and ageing.

Transcriptional dynamics has been suggested to be a major contributor to brain architecture, functional evolution as well as the development process and ageing (Belgard et al. 2011; Aprea et al. 2013; Telley et al. 2016). Long noncoding RNAs (lncRNAs) are a subgroup of RNA longer than 200 nucleotides yet have limited protein coding potential. Many lncRNAs are 5'capped, alternatively spliced and polyadenylated like mRNAs (Rinn and Chang 2012; Sun and Kraus 2013). Despite such similarity, lncRNAs are regulated differently and represent the fastest evolving parts of the primate genome (Pollard et al. 2006; Qureshi and Mehler 2012). LncRNA have a broad range of functions in various physiological and pathological contexts (Huarte and Rinn 2010; Guttman et al. 2011; Gutschner and Diederichs 2012; Rinn and Chang 2012; Batista and Chang 2013; Sauvageau et al. 2013; Sun and Kraus 2013; Necsulea et al. 2014; Sun et al. 2015). LncRNAs are epigenetic and transcriptional regulators that serve as scaffolds for the assembly of chromatin- and gene-regulating complexes, and can take part in directing those complexes to specific loci in the genome (Wang and Chang 2011; Rinn and Chang 2012; Vance and Ponting 2014). Alternatively, lncRNAs can act as molecular sponges that buffer various protein factors and thus regulate the processing and post-transcriptional modifications of mRNAs. Also, relying on base pairing mechanisms, they modulate mRNA stability and affect translational control (Fatica and Bozzoni 2014).

The number of identified lncRNAs is close to the number of the protein encoding mRNAs (GENCODE V25, http://www.gencodegenes.org/). While the majority of the

lncRNAs are poorly conserved and expressed at significantly lower levels than mRNAs (Derrien et al. 2012; Briggs et al. 2015; Ulitsky 2016), their expression patterns are tissue- and stage-specific, suggesting their considerable importance in regulating different biological functions, in particular cellular differentiation and development (Mercer et al. 2009; Ponting et al. 2009; Fatica and Bozzoni 2014; Briggs et al. 2015). The brain is an excellent example of this function. Around 40% of mammalian lncRNAs are expressed in the brain in a precise temporal and spatial pattern. This suggests that lncRNAs are part of the machinery needed to regulate specific neuronal functions (Mercer et al. 2008a; Mercer et al. 2008b; Derrien et al. 2012; He et al. 2014; Necsulea et al. 2014; Briggs et al. 2015). Examples of this function include *Malat1*, *MIAT* and antisense RNAs to *Uchl1* and *Kcna2* (Bernard et al. 2010; Carrieri et al. 2012; Zhao et al. 2013; Barry et al. 2014). In addition, reconstruction of an evolutionarily conserved co-expression network suggested that lncRNAs might be involved in synaptic transmission of neurons and other fundamental biological process like spermatogenesis (Necsulea et al. 2014). Despite these new data, the precise mechanism(s) by which lncRNAs play their roles in defining the complexity of brain functions remains unclear.

A recent microarray analysis of the temporal and anatomical expression of protein coding genes but not of lncRNAs in cortical and subcortical regions associated with human neuropsychiatric diseases has yielded wealthy information of the transcriptional regulation in the primate brain development and function, and the transcriptional link with neurological states (Bakken et al. 2016). However, it remains

uncertain on the mechanism on how lncRNAs play their roles in defining the complexity of brain functions, especially in primate brain during development and ageing.

Results

LncRNA expression in rhesus macaque brain is highly similar to human

We generated cDNA libraries of polyadenylated RNA extracted from eight macro-dissected brain areas, that included the prefrontal cortex (PFC), posterior cingulate cortex (PCC), temporal cortex (TC), parietal cortex (PC) and occipital cortex (OC), hippocampus CA1 and dentate gyrus (DG), and cerebellar cortex (CB) regions from macaques of four different age groups (1-,4-,10- and 20-year-old) (Fig. 1A, and Supplemental Table S1). We generated RNA-seq data sets (1 library per age- & sex-matched pair samples) at a sequencing depth of 148.1 million reads per sample (Supplemental Table S1). After all, we aligned the filtered reads to the reference sequence (Rhesus Macaque Genome et al. 2007) by TopHat2 (Kim et al. 2013) with 2 mismatches, we were able to detect and characterize the expression patterns of ~96.26% of known annotated genes (Fig. 1B, Supplemental Fig. S1B).

In order to identify lncRNAs from different brain regions, we used Cufflinks (Trapnell et al. 2010) to perform *ab initio* transcript assembly and reconstructed a

total of 620,089 transcripts. After a series of filtering as described Supplemental Methods, 19,509 multi-exonic lncRNAs encoded by 9904 genomic loci were identified from the remaining transcripts (Supplemental Table S2), among which 2492 (12.77%) were originated from antisense regions. The lengths of these lncRNAs were generally shorter than mRNAs (Supplemental Fig. S1C-D). In addition, lncRNA genes expressed in macaque brains have much lower GC contents in comparison with mRNA-coding genes (Supplemental Fig. S1E). A larger number of lncRNAs (759) were exclusively expressed in the neocortex compared to those in CB (270) and hippocampus (388) (Fig. 1C). In contrast, there were more mRNA specifically expressed in CB (705, *p*-value = 0.03, Fisher's exact test). Our data also reflected that expression of lncRNA genes was less conserved than mRNA genes among the main brain regions (Fig. 1B-C).

To further explore the conservation of macaque brain lncRNAs, we first downloaded 9,325; 20,785; 141,353; and 117,405 lncRNAs specific for macaque, gorilla, human and mouse respectively from the NONCODE database (Zhao et al. 2016) followed by a comparative analysis. Among them, 19,509 macaque brain lncRNAs were aligned to 4388 of macaque, 4236 of gorilla, 6752 of human and 3036 of mouse lncRNAs, respectively. Note that the homologues identified between macaque brain lncRNAs and the NONCODE primate lncRNAs were not significantly reduced by increasing the BLAST (Altschul et al. 1990) stringency, while those between the macaque brain lncRNA and the mouse lncRNAs were significantly decreased (Fig. 1D). When mammalian brain-related lncRNAs homologous to macaque brain lncRNAs were

aligned with each other, and approximately half of them (2241) were shared by all four species (E-value = 1e-3). The number of lncRNAs shared by macaque, gorilla and human were found to be higher with a stricter threshold (2039, E-value = 1e-10) than with a looser threshold (1093, E-value = 1e-3) (Supplemental Fig.S1F).

The extent of regulation of the expression of brain lncRNAs is higher than that of mRNAs

In order to understand the spatiotemporal expression patterns of all mRNAs and lncRNAs in our data sets, principal component analysis (PCA) was performed. The mRNA expression pattern in CB representing a distinct cluster, those in TC and OC representing another one, and the rest of the five regions represent the third one; whereas for lncRNA expression, only CB cluster was separable from another cluster comprising all other samples (Fig. 2A). Next, Pearson correlation analysis for all pairs of RNA-seq samples was performed, demonstrating similar results (Fig. 2B, and Supplemental Fig. S2A). Expression of mRNAs in each cluster was closer than that of lncRNAs, consistent with a higher expression dynamics of lncRNAs except for the CB cluster (Fig. 2 and Supplemental Fig. S2). The clustering of cerebral lncRNAs showed close similarities in all samples from the 1-year-old age group but a clear divergent expression at later ages (Fig. 2B). To eliminate the influence of expression discrepancy due to any spatiotemporal features of these two classes of RNAs, we performed a similar analysis with filtered lncRNAs and mRNAs having expression

RPKM values ranging from 0.1 to 20. The results showed the same clustering profiles as those of the unfiltered data sets (Supplemental Fig. S2B-C).

LncRNAs are well-known for their tissue-specific expression patterns than protein-coding genes, and Jensen-Shannon (JS) divergence analysis reveals high tissue specificity scores of lncRNAs expressed from different human tissues (Cabili et al. 2011). The same analysis also suggests that considerable cellular specificity of lncRNAs among different neuronal cell types (Molyneaux et al. 2015). We further performed JS divergence analysis for brain mRNAs and lncRNAs from the brain samples and found that majority of lncRNAs and mRNAs scored lower than 0.25, which were lower than the lncRNA scores corresponding to studies mentioned above. Interestingly, no significant difference in specificity scores was observed between lncRNAs and mRNAs (Supplemental Fig. S2D, *p*-value = 0.91, Kolmogorov-Smirnov test, KS test). We also calculated the tissue specificity scores for lncRNAs and mRNAs within similar expression levels (RPKM value ranging from 0.1 to 20). Profiles of cumulative specificity scores between filtered lncRNAs and mRNAs were similar (Supplemental Fig. S2D, *p*-value = 1, KS test).

Next, we determined how lncRNAs were differentially expressed by studying the expression of known lncRNAs (Supplemental Fig. S3). We identified 19 copies of *KCNQ1OT1*, 3 copies of *RMST*, one copy of *XIST* and its antisense non-coding RNA *TSIX*, *SOX21-AS1* and *MIAT*. As being sex-determined, *XIST* was exclusively expressed in high levels among all female macaque brain samples without significant

changes at different ages (Supplemental Fig. S3A). Such female exclusive expression pattern was further confirmed by qPCR (Supplemental Fig. S3A). Interestingly, expression of *TSIX* was highly neocortex-specific, and the expression level was the highest in 1-year-old OC samples (Supplemental Fig. S3B). *RMST* is known to be regulated by the transcription factor REST which then drives the recruitment of the neural transcription factor, SOX2, to turn on key neurogenesis-promoting genes, such as *DLX1* and *ASCL1* (Ng et al. 2013). We observed that among three copies of *RMST*, one was expressed at very low level while the other two expressed in an age- and sex-dependent manner. *RMST-2* (the second copy of *RMST*) was more negatively correlated to age in female samples; while *RMST-3* (the third copy of *RMST*) expression was more temporally regulated in both female and male macaques (Supplemental Fig. S3C-D). The temporal regulation of *MIAT* expression showed more spatial-specific (standard deviation, SD = 12.51) than that of *SOX21-AS1* expression (SD = 0.20) (Supplemental Fig. S3E-F).

Lastly, we determined the differential expression of lncRNAs and mRNAs of the same anatomic structure between any two adjacent age groups (1-, 4-, 10-, and 20-year-old). Stages from 1-year-old to 4-year-old showed that expression of lncRNAs changed the most in all the regions except DG. Such changing pattern was also evident in mRNA expression, with exception that both CB and DG failed to show the most significant changes at 1-year. Substantial changes in expression of both lncRNAs and mRNAs were also observed in the period from 4-year-old to

10-year-old. However, changes observed between 10-year to 20-year-old period were the least (Fig. 2C-D).

Temporal-regulated lncRNAs are populated into spatial-, temporal- and sex-specific classes

To characterize the dynamic changes of lncRNA and mRNA expression, we clustered all their expression patterns (3635 lncRNAs and 7070 mRNAs) by the WGCNA method (Langfelder and Horvath 2008). We identified 18 main lncRNA transcriptional modules, each represented by a characteristic expression pattern (Fig.3A-B). On the other hand, 14 main mRNA transcriptional modules were also identified (Supplemental Fig. S4A-B). We explored each lncRNA and mRNA modules by heatmap graphing and eigengene value graphing (described by "color" corresponding to cluster dendrogram), this allowed us to define the modules into three classes-spatiotemporal, tempo-spatial and sex-temporal. Spatiotemporal modules were characterized by remarkably higher expression in distinct brain structures, while temporal regulation was less remarkable (Figs. 3 and S4). The postnatal dynamic lncRNA modules strongly associated with specific brain architectures includes CB (M1, turquoise, 794 lncRNAs), DG/CA1 (M2, blue, 443 lncRNAs), CA1 (M4, yellow, 369 lncRNAs), neocortex (M7, black, 123 lncRNAs) and OC (M10, purple, 57 lncRNAs) (Fig. 3C).

Tempo-spatial modules demonstrated a more pronounced and patterned regulation by postnatal development and ageing, but were less patterned by structure separation (Fig. 3D). Expression of lncRNA and mRNA genes in sex-temporal modules was specific to both sex and age, but was less associated with specific structure (Fig. 3E, Supplemental Fig. S4E10, and Supplemental Table S3). Surprisingly, reciprocal sex-specific patterns of lncRNAs were observed across the four ages (Fig. 3E1-4). Such a reciprocal sex-specific regulation was also shown by mRNAs (Supplemental Fig. S4E10).

To validate the spatial-specific lncRNAs in macaque brain, we determined the expression levels of three CB-specific lncRNA *RP11-491F9.1*, *Gm37142* and *LINC00670* which were abundant in brain and predicted with potential roles in brain function. Both qPCR and RNA-seq data revealed that *RP11-491F9.1*, *Gm37142* and *LINC00670* exclusively preserved in CB across the four ages (Supplemental Fig. S5A and S6A). ISH data from 10-year old CB slices validated that *RP11-491F9.1* exclusively expressed in CB (Supplemental Fig. S5B and S6B). This phenotype was also confirmed by the CA1- and DG-enriched *NONHSAG047825.1* too (Supplemental Fig. S6C-D).

High dynamics of lncRNA expression in the cerebral cortices

Among the lncRNA co-expression modules, the third largest M3 contains 396 lncRNAs. M3 did not express in CB or DG, but highly expressed in PFC, PCC, TC,

PC and OC in an age-regulated pattern (Fig. 3D). These lncRNAs were mostly expressed in 1-year-old, and their expression was reduced significantly at other ages especially in regions like PCC, PC and TC. We named this class of cerebral lncRNAs as "Early lncRNAs". In CA1 region, these classes of lncRNAs were expressed in a similar pattern. M9 (magenta, 58) represented another class of lncRNAs. Similar to M3, M9 lncRNAs expressed at highest level in 1-year-old macaques but reduced significantly in other ages, especially in regions of PFC, PCC, TC, PC, OC and CA1. In contrast to M3, high M9 lncRNAs expression at 1-year-old was more evident in male than in female except for PFC. M6 represent another class of early lncRNA, featured by being mostly expressed in OC. Interestingly, expression of both M5 (green, 207) and M7 (black, 123) lncRNAs was at minimal at 1-year-old, but their expression was significantly higher in other age groups. We therefore termed M5 and M7 as "Late lncRNAs". Next, we observed that the expression pattern of *AC112693.3*, *HCG11*, *NONMMLT001498.1*; and *AC016757.3*, *lnc-OCM-2*, *NONHSAT163151.1*, resembled early and late lncRNAs with potential brain function, respectively. Data from qPCR and ISH showed that *AC112693.3*, *HCG11* and *NONMMLT001498.1* expression decreased sharply after age 1-year-old, while *AC016757.3*, *lnc-OCM-2*, *NONHSAT163151.1* increased gradually with age (supplemental Fig.S7A-D, and S8A-E).

Sex difference in lncRNA expression

Sex-biased expression of protein coding genes has been reported in the human brain (Kang et al. 2011). Here, we have identified a 4-year-old male mRNA module (81

protein-encoding genes) and 4 sex-temporal lncRNA modules (183 lncRNA genes). The four sex-temporal-specific lncRNA modules were the 1- and 10-year male (M13 and M9), 4- and 20-year female (M8 and M18) (Fig. 3E1-4).

We next used paired *t*-test module (*p*-value<0.01) to identify sex-specific lncRNAs regardless to their temporal regulation. A total of 307 sex-biased lncRNAs were identified, including 148 female-biased and 159 male-biased (Fig. 4A). Among these sex-specific lcnRNAs, 5 and 2 sex-biased lncRNAs were encoded from the X Chromosome of female and male macaques, respectively. The same approach identified 90 female-biased and 129 male-biased mRNAs (Supplemental Fig. S9A-D). Therefore, sex-biased lncRNAs (307/9904, 3.1%) were present at a much higher frequency than mRNAs (219/26654, 0.8%, p-value < 2.2e-16, Fisher's exact test) across all ages and brain structures. The number of overlaps found between sex-biased and sex-temporal mRNA and lncRNAs were very limited, with only 7 mRNAs and 24 lncRNAs that being confirmed as two populations of sex-specific lncRNAs. Finally, we have determined the expression levels of three sex-biased lncRNAs *AC027613.1*, *NONGGOT004660.1* and *AC132825.2* in macaque brain which were abundant in brain. Further analyses of RNA-seq, qPCR and ISH data revealed high correlation of the *AC027613.1*, *NONGGOT004660.1* and *AC132825.2* with sex and age specificities (Fig. 4B-E, Supplemental Fig. S10A-D, and S11).

CAGE-seq analysis of the transcription start sites and 5'-capping dynamics of lncRNAs

Annotation of the transcriptional start sites on lncRNAs is important but this concept is challenged by their diversity in biogenesis and by their low expression and conservation levels. Until recently, serious efforts have been made to gather human lncRNA transcript models with accurate 5' ends by integration of various CAGE-seq data (Hon et al. 2017). In order to identify the more accurate transcription start sites (TSS), we used a modified CAGE-seq technology to analyze lncRNAs identified by RNA sequencing assembly in macaque brains (Fig.1). This technology selected polyadenylated lncRNAs to identify the 5'-cap structures. As a result, full-length lncRNAs with both 5'-cap and 3'-poly (A) tail were enriched. CAGE-seq was used to generate 5'-cap sequencing reads from the very same 64 macaque brain samples used to generate RNA sequencing reads for assembly. Therefore, TSSs were annotated to the same sets of lncRNAs assembled by RNA reads. Detailed sequencing and alignment results were shown in Supplemental Table S4, with an average of 66.83% mapping efficiency. Note that these reads were significantly enriched at the TSSs of known mRNAs (Supplemental Fig. S12A).

With the CAGE transcript start sites (CTSSs) falling within 20 bp being clustered into transcript clusters (TCs) as previously reported (Nepal et al. 2013), each TC may then represent a potential TSS. A total of 103,766 TCs were identified from all brain samples; 52.49% of them were distributed across 15,592 annotated genes. 80% of

TCs had a width of no more than 4-nt (Supplemental Fig. S12B), illustrating the strict usage of TSS in macaque brain. Over 82% of CTSS were grouped into TCs, among which 32.93% TCs were detected from more than one sample. TCs were enriched around the TSSs of both known protein encoding genes as well as lncRNAs identified in this study (Fig. 5A). We found that CAGE-seq reads and TCs were strongly enriched at 5'UTR but not at 3'UTR and intronic regions (Fig. 5B).

A total of 6991 of the intergenic TCs fell into 3084 (31.14%) lncRNAs annotated in this study, among which 2324 lncRNAs were homologous to those of human, showing significant enrichment (p-value < 0.001, Fisher's exact test). A total of 13,269 mRNAs (43.87%) had at least one TC support. Different TC frequencies of lncRNAs and mRNAs were well correlated with their differential expression levels (Supplemental Fig. S12C, p-value=4.369e-08, Mann-Whitney U test). TCs of a gene identified within the gene body or at 2-kb upstream of its previously annotated TSS, or genes containing more than one TCs were assigned alternative promoters. About 66% and 46% of mRNAs and lncRNAs contained more than one TC, respectively, showing alternative promoter evidence (Fig. 5C, and Supplemental Fig. S12D).

We further plotted the ratio of multiple TC-containing genes to all TC-containing genes in all samples. Alternative promoter usage in mRNA genes was dynamically regulated by age in a spatial and sex-dependent manner (Supplemental Fig. S12E). When the same analysis was applied for lncRNA with alternative TCs, similar dynamic patterns were demonstrated for both male and female subjects (Fig. 5D, R =

0.57, Pearson correlation coefficient). The frequency of 5'-capped mRNA and lncRNAs among all lncRNAs and mRNAs demonstrated that the 5' capping efficiency could be regulated spatially in different brain regions, as well as by the degree of brain maturation and age-related degeneration (Fig. 5E, and Supplemental Fig. S12F). We also noticed the presence of sex-dependent regulation of 5' capping efficiency and alternative promoter usage, one example is the *lnc-CHRM3-1* gene, shown in Fig. 5F-G.

LncRNA-mRNA co-expressed network

To explore the functions of brain lncRNAs, a correlation matrix between 9904 lncRNAs and 26654 mRNAs was generated by computing the Pearson correlation coefficient for all pairwise combinations based on their expression in our 64 transcriptomes. At a stringency of p-value ≤ 0.01 and absolute Pearson correlation coefficient abs (PCC) ≥ 0.7, a total of 3341261 co-expression pairs were detected between 5084 lncRNAs and 18418 mRNAs (Supplemental Table S5). For example, 237 mRNAs and 93 lncRNAs were co-expressed with *MIAT*, functional clustering of the interacted mRNAs revealed that this lncRNA is extensively involved in neuroactive ligand-receptor interaction, GABAergic synapse, dopaminergic synapse, glutamatergic synapse and morphine addiction (Fig. 6A).

Enriched Gene Ontology (GO) terms and KEGG pathways were further obtained for all mRNAs interacted with each lncRNA module. Fanconi anemia pathway was the most enriched for M1 lncRNAs (CB module); circadian rhythm pathway was mostly

enriched for M2 lncRNAs (CA1/DG-module); allograft rejection and autoimmune disease were found for M13 lncRNAs (1-year male module). For the neocortex lncRNA modules such as M3, M5, M6, M7 and M9, mRNAs interacted were more enriched in conferring synaptic functions (Fig. 6B).

As an example of illustration, we explored the function of two neocortex lncRNA modules M5 (late lncRNAs) and M6 (early lncRNAs) (Fig. 3 and 6C, Supplemental Fig. S7 and S13A). Both GO and KEGG analyses showed that these two lncRNA modules were enriched in quite divergent functions (Supplemental Fig. S13B). We further generated M5 and M6 co-expression networks with their mRNA partners and mapped their interaction strength. The interaction strength map revealed that mRNA genes strongly correlated with M5 lncRNAs included *ARHGAP9* (rho GTPase activating protein 9), *MAPK13*, *CAMK2N1* (calcium/calmodulin dependent protein Kinase II Inhibitor 1), *HTR2A* (5-hydroxytryptamine receptor 2A) and *NRSN1* (neurensin 1) (Supplemental Fig. S13C). The co-expression map for M6 lncRNAs revealed different classes of strong co-expression genes such as *NGEF* (neuronal guanine nucleotide exchange factor), *KCNH4* (potassium voltage-gated channel subfamily H member 4) and *HCRTR1* (hypocretin receptor 1) (Fig. 6D).

Lastly, we analyzed the number of lncRNA-mRNA pairs between any two module pairs (18 lncRNAs and 14 mRNA modules). We found numbers of strong module-module co-expression pairs, which include exclusive pairs such as L1-M1, L4-M4 as well as multiple pairs such as L2 with M1 and M10, and L6 with M2, 7 and M8 (Fig. 6E, and Supplemental Table S5).

Negatively regulatory networks between mRNA-lncRNA and lncRNA-lncRNA

Co-expression of the gene pairs was then established between all pairs of lncRNA-lncRNA and mRNA-mRNA. We found that positive pairs were the predominant species, as also consistent with the co-expression pattern of most genes involved (D'Haeseleer et al. 2000; Zhang and Horvath 2005). Strikingly, as high as 22.37% lncRNA-mRNA and 25.35% lncRNA-lncRNA were negative pairs, in contrast to the 5.03% observed in mRNA-mRNA pairs. This suggests that the higher population of lncRNA takes part in negative pairs supported the notion that lncRNAs is gene repressing in nature, this includes the repression of both mRNAs and lncRNAs expression (Supplemental Table S5).

After the stringent filtering, 3341261 lncRNA-mRNA pairs remained in our lncRNA-mRNA network containing 5084 lncRNAs and 18418 mRNAs. The network of a million co-expressed pairs contains 92.93% of positive lncRNA-mRNA pairs and 7.07% negative pairs.

We determined to illustrate the resulting negative regulatory network with that of *Ptbp1* gene, a conservative heterogeneous nuclear ribonucleoprotein (hnRNP) that regulates neuronal gene expression. We found that *Ptbp1* level was negatively correlated with 61 mRNA genes. Functions of these genes included neuron differentiation, cell projection organization, neuron and nervous system development (Fig.7A-B). And genes negatively regulated by *Ptbp1* formed extensive co-expression

networks (Fig.7C). Knockdown of *Ptbp1* in mouse cortical neurons significantly increased the levels of several of its targeting genes including *Emx2*, *LhxX2*, *Nr2e1*, *Kif3a* and *Foxg1* expression (Supplemental Fig. S14A-I).

Next, we analyzed the lncRNAs in each distinct module that formed negatively correlates with mRNAs (Fig. 7D). We further identified lncRNAs that target mRNA and lncRNA modules, and then analyzed if there are any overlaps among these lncRNAs. Fig.7E shows the presence of a large number of lncRNAs that controlled both mRNAs and lncRNAs belonged to the M1 module (high expression in CB). Alternatively, we also noticed substantial number of lncRNAs could negatively regulate the expression of both M1 CB-mRNAs and M6/M7 neocortex-lncRNAs. Among such negative regulatory network, lncRNAs of M1 and M5 modules were most extensively regulated by other lncRNAs, the same was also true for mRNAs of M6, M7 and M8 modules.

Discussion

Genomic and transcriptomic profiling of brain tissue data sets of different species reveal alterations in genetic and epigenetic systems underlies the processes brain development, ageing, and even mental disorders (Oldham et al. 2008; Belgard et al. 2011; Qureshi and Mehler 2012; Aprea et al. 2013; Bakken et al. 2016). In this study, using RNA-seq and CAGE-seq we generated complementary data sets that allowed

the identification and confirmation of full-length orthologous lncRNA sequences, novel transcripts from macaque brain across postnatal development and ageing. We expect that our new resource should contribute to the understanding of the importance of lncRNA-mediated regulation, not only in aspects of brain development and ageing, but also to brain-related disorders during at different periods of a life time.

Although the contribution of sex difference in human cognition is well conceived, very limited information were available in the literature which explains their relationship (McCarthy and Arnold 2011). Our analysis of macaque brain lncRNAs is the first to identify hundreds of sex-temporal and sex-biased lncRNAs related to the postnatal development and ageing, indicating that lncRNAs might play significant roles in shaping the cognitive difference observed between male and female subjects.

The alternative promoter usages of both mRNAs and lncRNAs are also expected to play a role during brain development and ageing, which has not been systematically studied so far. Using full-length CAGE-seq approach, we identified not only the potential transcription start sites of a large fraction of macaque brain lncRNAs (31.14%), but also extended this finding to understand how spatial, temporal and sex parameters regulate of brain lncRNAs expression. Alternative promoter usage and capping efficiency associated with the transcription of lncRNA and mRNAs could represent an important mechanism in regulating macaque brain development and ageing, this may also take part in regulating the expression of these two classes of

RNAs.

While mRNA co-expression networks has been described as important code to understand the brain (Cabili et al. 2011; Fertuzinhos et al. 2014; Molyneaux et al. 2015; Zeisel et al. 2015), very few of them appears to reflect the complexity of brain architecture and function. We demonstrated how the dataset can be used to profile trajectories of genes associated with specific neurobiological categories or disorders, many of which are not likely be evident from transcriptomic profiles of commonly studied model systems. Coupled with analysis of co-expressed genes in the dataset, these provide information on specific timing and tissue localization of various genes expressed in the brain, which will also offer insights regarding to their function. Our data enhances genome-wide associations and linkage studies by narrowing the focus to any candidate genes that are specifically expressed during development or restricted to a specific region known to be afflicted in neurological diseases. Additional parameter like how lncRNA represses other lncRNA expression is still currently understudied. Therefore, it is likely that our report on the dynamic changes observed in lncRNA co-expression networks may serve as a regulatory system that truly contribute to the complexity of the brain architectures and function, particularly in primates.

Methods

RNA-seq and CAGE-seq library construction and sequencing

For RNA-seq library, total RNA was extracted from all the brain tissue samples by using TRIzol Reagent (Ambion) following the manufacturer's instructions. After DNA deplection, polyadenylated RNAs were purified and concentrated with oligo (dT)-conjugated magnetic beads (Invitrogen) before being used for directional RNA-seq library preparation. RNA reverse transcription was performed with RT primer harboring 3' adaptor sequence and randomized hexamer. The cDNAs were purified and amplified. And products corresponding to 200-500 bp were purified, quantified and stored at -80℃ before sequencing.

For CAGE-seq, total RNA was treated with RQ1 DNase (Promega) to remove DNA. Polyadenylated RNAs were purified and concentrated with oligo (dT)-conjugated magnetic beads (Invitrogen). The capped mRNA was performed with RT primer , then synthesized DNA with Terminal-Tagging oligo. The cDNAs were purified and amplified with PCR primers (Illumina) and PCR products corresponding to 200-500 bp were purified, quantified and stored at -80℃ until sequencing.

For high-throughput sequencing, the libraries were prepared following the manufacturer's instructions and applied Illumina HiSeq 2000 system for 100 nt paired-end sequencing and NextSeq 500 system for 150 nt paired-end sequencing by ABlife. Inc (Wuhan, China), for RNA-seq and CAGE-seq, respectively.

RNA-seq and CAGE-seq Raw Data filtering and Alignment statistics

Raw reads were first filtered to remove the adaptor and bases of low quality by FASTX-Toolkit (Version 0.0.13). Filtered reads were aligned to the macaque genome by TopHat2 (Kim et al. 2013) with end-to-end method allowing 2 mismatches. Uniquely localized reads were then used to calculate reads number and RPKM value (RPKM represents reads per kilobase and per million) for each gene according to reads and genes genomic location. After getting the Expression level of all genes in all the samples, differentially expressed genes were analyzed by using edgeR (Robinson et al. 2010). See the Supplemental Methods for more details.

CAGE-seq data analysis

After alignment, 5 end of prime of each reads was considered as the CAGE tag-defined transcription start sites (CTSSs). The number of CAGE tags mapping to each CTSS across different samples was normalized to obtain the normalized number of tags per million (TPM). We then combined the TSSs to transcription clusters (TCs) according the known method (Nepal et al. 2013). Only CTSSs supported by a minimum of 0.5 tpm in at least one sample were used for a sample-specific clustering into transcript clusters (TCs). Neighboring CTSSs were clustered if they were <20 bp apart. See the Supplemental Methods for more details.

Data access

RNA-seq and CAGE-seq data have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE87182. The modified WGCNA code and co-expression network and data are available in the Supplemental Material can be downloaded from GitHub (https://github.com/DChenABLife/RhesusLncRNA).

Acknowledgments

We are grateful to Karl Herrup for critical reading the manuscript, and Kim Heiman Chow for editorial assistance. This study was supported by grants from the Chinese Academy of Sciences (Y406541141 and 1100050210 to J.L.), the Ministry of Science and Technology of China (2015CB755605 to J.L.), and the National Science Foundation of China (91649119 to J.L.); and this study is also supported by grants the Chinese Academy of Sciences (KJ2070000031 to X.W.) and the National Science Foundation of China (91440108 to X.W.); this study is also supported by grants from the Strategic Priority Research Program of CAS (XDB02020005 to X.H.) and the National Science Foundation of China (81471312 to X.H.); this study is also supported by grants from the Appreciate the Beauty of Life, Inc. (ABL2015-02007 to Y. Zhang.).

Author Contributions

J.L., X.W., and Yi Zhang initiated and conceived the project, J.L., S.L., Z.W., D.C., B.Z., R.T., designed the experiments. J.L. organized and supervised the whole project.

S.L., Z.W., B.Z., R.T., J.W., Ying Zhang, L.L. and L.Y. collected macaque brain samples and performed RNA extraction; D.C., C.C., Yi Zhang., X.W., and J.L. collected and analyzed the data; S.L., Z.W., B.Z., R.T., K.X. and J. M. performed ISH and qPCR. J.L. drafted the manuscript with input from all authors.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of molecular biology* **215**: 403-410.

Aprea J, Prenninger S, Dori M, Ghosh T, Monasor LS, Wessendorf E, Zocher S, Massalini S, Alexopoulou D, Lesche M et al. 2013. Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *The EMBO journal* **32**: 3145-3160.

Bakken TE, Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, Dalley RA, Royall JJ, Lemon T et al. 2016. A comprehensive transcriptional map of primate brain development. *Nature* **535**: 367-375.

Barry G, Briggs JA, Vanichkina DP, Poth EM, Beveridge NJ, Ratnu VS, Nayler SP, Nones K, Hu J, Bredy TW et al. 2014. The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Molecular psychiatry*

19: 486-494.

Batista PJ, Chang HY. 2013. Long noncoding RNAs: cellular address codes in development and disease. *Cell* **152**: 1298-1307.

Belgard TG, Marques AC, Oliver PL, Abaan HO, Sirey TM, Hoerder-Suabedissen A, Garcia-Moreno F, Molnar Z, Margulies EH, Ponting CP. 2011. A transcriptomic atlas of mouse neocortical layers. *Neuron* **71**: 605-616.

Bernard D, Prasanth KV, Tripathi V, Colasse S, Nakamura T, Xuan Z, Zhang MQ, Sedel F, Jourdren L, Coulpier F et al. 2010. A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J* **29**: 3082-3093.

Briggs JA, Wolvetang EJ, Mattick JS, Rinn JL, Barry G. 2015. Mechanisms of Long Non-coding RNAs in Mammalian Nervous System Development, Plasticity, Disease, and Evolution. *Neuron* **88**: 861-877.

Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development* **25**: 1915-1927.

Carrieri C, Cimatti L, Biagioli M, Beugnet A, Zucchelli S, Fedele S, Pesce E, Ferrer I, Collavin L, Santoro C et al. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* **491**: 454-457.

D'Haeseleer P, Liang S, Somogyi R. 2000. Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* **16**: 707-726.

Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research* **22**: 1775-1789.

Fatica A, Bozzoni I. 2014. Long non-coding RNAs: new players in cell differentiation and development. *Nature reviews Genetics* **15**: 7-21.

Fertuzinhos S, Li M, Kawasawa YI, Ivic V, Franjic D, Singh D, Crair M, Sestan N. 2014. Laminar and temporal expression dynamics of coding and noncoding RNAs in the mouse neocortex. *Cell reports* **6**: 938-950.

Gutschner T, Diederichs S. 2012. The hallmarks of cancer: a long non-coding RNA point of view. *RNA biology* **9**: 703-719.

Guttman M, Donaghey J, Carey BW, Garber M, Grenier JK, Munson G, Young G, Lucas AB, Ach R, Bruhn L et al. 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* **477**: 295-300.

He Z, Bammann H, Han D, Xie G, Khaitovich P. 2014. Conserved expression of lincRNA during human and macaque prefrontal cortex development and maturation. *RNA* **20**: 1103-1111.

Hon CC, Ramilowski JA, Harshbarger J, Bertin N, Rackham OJ, Gough J, Denisenko E, Schmeier S, Poulsen TM, Severin J et al. 2017. An atlas of human long non-coding RNAs with accurate 5' ends. *Nature* **543**: 199-204.

Huarte M, Rinn JL. 2010. Large non-coding RNAs: missing links in cancer? *Hum Mol Genet* **19**: R152-161.

Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, Sousa AM, Pletikos M, Meyer KA, Sedmak G et al. 2011. Spatio-temporal transcriptome of the human brain. *Nature* **478**: 483-489.

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**:

R36.

Langfelder P, Horvath S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics* **9**: 559.

McCarthy MM, Arnold AP. 2011. Reframing sexual differentiation of the brain. *Nat Neurosci* **14**: 677-683.

Mercer TR, Dinger ME, Mariani J, Kosik KS, Mehler MF, Mattick JS. 2008a. Noncoding RNAs in Long-Term Memory Formation. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* **14**: 434-445.

Mercer TR, Dinger ME, Mattick JS. 2009. Long non-coding RNAs: insights into functions. *Nature reviews Genetics* **10**: 155-159.

Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. 2008b. Specific expression of long noncoding RNAs in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 716-721.

Molyneaux BJ, Goff LA, Brettler AC, Chen HH, Brown JR, Hrvatin S, Rinn JL, Arlotta P. 2015. DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85**: 275-288.

Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U, Baker JC, Grutzner F, Kaessmann H. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**: 635-640.

Nepal C, Hadzhiev Y, Previti C, Haberle V, Li N, Takahashi H, Suzuki AM, Sheng Y, Abdelhamid RF, Anand S et al. 2013. Dynamic regulation of the transcription initiation landscape at single nucleotide resolution during vertebrate embryogenesis. *Genome research* **23**: 1938-1950.

Ng SY, Lin L, Soh BS, Stanton LW. 2013. Long noncoding RNAs in development and disease of the central nervous system. *Trends Genet* **29**: 461-468.

Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, Geschwind DH. 2008. Functional organization of the transcriptome in human brain. *Nat Neurosci* **11**: 1271-1282.

Pollard KS, Salama SR, Lambert N, Lambot MA, Coppens S, Pedersen JS, Katzman S, King B, Onodera C, Siepel A et al. 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* **443**: 167-172.

Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell* **136**: 629-641.

Qureshi IA, Mehler MF. 2012. Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nature reviews Neuroscience* **13**: 528-541.

Rhesus Macaque Genome S Analysis C Gibbs RA Rogers J Katze MG Bumgarner R Weinstock GM Mardis ER Remington KA Strausberg RL et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**: 222-234.

Rinn JL, Chang HY. 2012. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* **81**: 145-166.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, Sanchez-Gomez DB, Hacisuleyman E, Li E, Spence M et al. 2013. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* **2**: e01749.

Sun M, Gadad SS, Kim DS, Kraus WL. 2015. Discovery, Annotation, and Functional Analysis of Long

Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells. *Molecular cell* **59**: 698-711.

Sun M, Kraus WL. 2013. Minireview: Long noncoding RNAs: new "links" between gene expression and cellular outcomes in endocrinology. *Molecular endocrinology* **27**: 1390-1402.

Telley L, Govindan S, Prados J, Stevant I, Nef S, Dermitzakis E, Dayer A, Jabaudon D. 2016. Sequential transcriptional waves direct the differentiation of newborn neurons in the mouse neocortex. *Science* **351**: 1443-1446.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **28**: 511-515.

Ulitsky I. 2016. Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. *Nat Rev Genet* **17**: 601-614.

Vance KW, Ponting CP. 2014. Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends Genet* **30**: 348-355.

Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Molecular cell* **43**: 904-914.

Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, Marques S, Munguba H, He L, Betsholtz C et al. 2015. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**: 1138-1142.

Zhang B, Horvath S. 2005. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**: Article17.

Zhao X, Tang Z, Zhang H, Atianjoh FE, Zhao JY, Liang L, Wang W, Guan X, Kao SC, Tiwari V et al. 2013. A long noncoding RNA contributes to neuropathic pain by silencing Kcna2 in primary afferent neurons. *Nature neuroscience* **16**: 1024-1031.

Zhao Y, Li H, Fang S, Kang Y, Wu W, Hao Y, Li Z, Bu D, Sun N, Zhang MQ et al. 2016. NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic acids research* **44**: D203-208.

Figure legends

Figure 1. A comprehensive catalog of lncRNA genes in rhesus monkey Brain

(A) Illustration of the experimental design and bioinformatics analysis pipeline for the identification and functional annotation of lncRNA genes expressed in macaque brain. Macaque brain regions used in this study were colored in red (neocortex), green (hippocampus) and blue (cerebellum). See Methods for more details.

(B-C) Venn diagram of detected mRNA (B) and lncRNA (C) genes in neocortex, hippocampus (CA1 & DG) and cerebellum.

(D) Number of NONCODE lncRNAs in gorilla, human, mouse and rhesus that are homologous to macaque brain lncRNAs identified in this study with loose (e-value < 1e-3) and strict (e-value < 1e-10) threshold by BLASTN. Alignment to mouse showed significant decrease with the strict threshold (***, *p*-value <0.0001 by Fisher's exact test, human as background).

Figure 2.The discrete expression patterns of lncRNAs and mRNAs

(A) Principal component analysis (PCA) of 64-pair distinct samples across the four ages based on normalized mRNAs (top) and lncRNAs (bottom) expression level. The samples were grouped by brain region and the ellipse for each group is the confidence ellipse.

(B) Heatmap of correlation coefficient for 64 samples based on the lncRNA expression level. The samples were grouped by hierarchical clustering and the dendrogram was not shown.

(C-D) Bar plot presentation of differentially expressed mRNAs (C) and lncRNAs (D) based on neighboring age groups.

Figure 3. The discrete expression modules of lncRNA expression by the WGCNA analysis

(A) Hierarchical clustering heatmap of all differentially expressed lncRNAs (Figure 2C, right) by samples. LncRNA modules were arranged from 0 (top) to 18 (bottom);

(B) Hierarchical cluster dendrogram of all differentially expressed lncRNAs modules. Modules corresponding to branches are labeled with colors indicated by the color bands underneath the tree.

(C) Eigengene bar plot of spatiotemporal modules of lncRNAs. Samples were first sorted by brain regions in the order of CB, DG, CA1, OC, TC, PC, PCC and PFC order. In each brain region, samples were then sorted by age (1y to 20y), and by sex (female and male).

(D) Eigengene bar plot of tempo-spatial modules of lncRNAs in the same sample order as in (C).

(E1-4) Eigengene bar plot of sex-temporal modules of lncRNAs in the same sample order as in (C).

Figure 4. Characteristics of sex-biased lncRNA expression

(A) Hierarchical clustering heatmap representation of the sex-biased lncRNA expression level. The sex-biased lncRNAs were obtained by *t*-test analysis (*p*-value < 0.01).

(B-C) Bar plot of a sex-regulated lncRNA *AC027613.1* expression pattern in female (B) and male (C) samples across the four ages with RNA-seq expression level (top) and qRT-PCR level (bottom). (B) represents the female samples, and (C) represents the male samples.

(D-E) Representative ISH images of the *AC027613.1* expression in female D) and male E) PFC across the four ages with 10x amplification (top) and 20x amplification (bottom). The images are representative of replicates of three independent experiments.

Figure 5. CAGE-seq characterization of alternative promoter usage and full-length frequency of lncRNAs

(A) TC number distribution around the annotated TSS of known mRNAs (turquoise) and lncRNAs (red) identified in this study.

(B) Genomic region distribution of RNA-seq reads (control), CAGE-seq reads and TCs to view the CAGE signal enrichment. Enrichment *p*-values were labeled for 5'UTR (Fisher's exact test).

(C) Pie chart of percentage for lncRNA genes with one and more promoters. ">5" indicates lncRNA genes with 6 and more promoters.

(D) Dynamics of alternative promoter profiles during macaque brain development and ageing. Alternative TC reads divided by total TC reads for each of the 64-samples were calculated and plotted. Male and female samples were separately plotted.

(E) Dynamics of full-length frequency profiles of all lncRNAs in all 64 brain samples. Full-length frequency is indicated by the detected fraction of lncRNAs with polyadenylation and 5' capping. The X axis label is the same as in (D).

(F) Distribution of RNA-seq reads density (blue) and CAGE-seq TCs (red) along the *lnc-CHRM3-1* lncRNA. Y-axis indicates the normalized density for RNA-seq and CAGE-seq.

(G) Line plot of *lnc-CHRM3-1* TSS density in PFC samples. The TSS density was represented by TPM.

Figure 6. Co-expression network illustration between lncRNAs and mRNAs

(A) Functional presentation of mRNAs that were co-expressed with lncRNA *MIAT*. Green rhombuses are the functional terms and the size represents the statistical significance. The red circles are the mRNAs. The line thickness represents the correlation coefficient of mRNAs and *MIAT*, and the shade degree of mRNAs represents the total statistical significance of mRNAs and *MIAT*.

(B) Heatmap presentation of the KEGG pathways for mRNAs associated with each lncRNA module. The function of each lncRNA module was annotated by

co-expressed mRNAs. Color degree of each cell represents the statistical significance of pathways [-log10 (Qvalue, see label)].

(C) Hierarchical clustering heatmap presentation for the expression pattern of Module 6 lncRNAs. Higher expression level was observed in 1-year samples for all brain areas except for DG.

(D) The co-expression network of M6 lncRNAs and the co-expressed mRNAs. LncRNAs are in the center and co-expressed mRNAs are on the outside. The numbers on lncRNAs are the last four digits of lncRNA ID. LncRNA shade degree, mRNA circle and word size represent interaction strength (sum of correlation coefficients) between M6 lncRNAs and mRNAs. Genes for different neuronal functions were presented with respective colors.

(E) Circular presentation of module-module interaction between lncRNAs and mRNAs. Scale bars were the same as in B). Shade degree of the cambered bars for each lncRNA module (LM) and mRNA module (MM) represents the log10 value of the co-expressed pair number in each module.

Figure 7. Negative regulatory networks between mRNA-lncRNA and lncRNA-lncRNA

(A) Hierarchical clustering heatmap of the *PTBP1* and the mRNAs negatively regulated by *PTBP1*. Color bar represents the $\log_{10}$ RPKM.

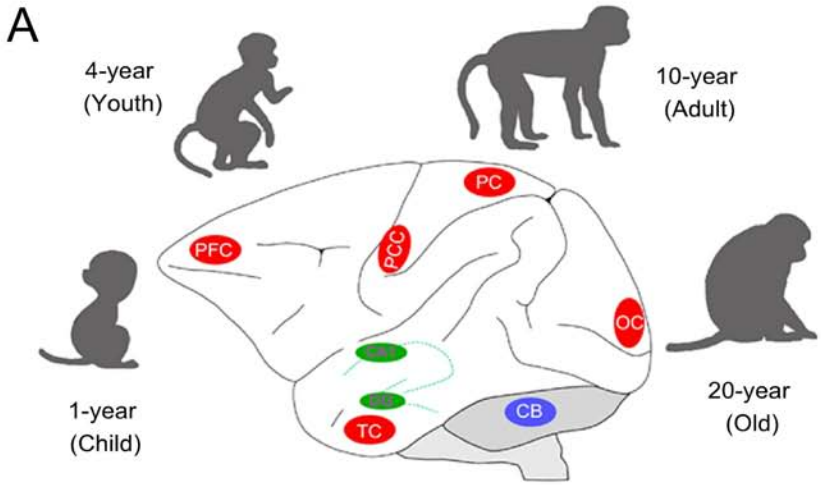
(B) Bar plot presentation of the functional terms of mRNAs negatively regulated by PTBP1. Length of bar represents the statistical significance of pathways [-log10 (Qvalue, see label)].

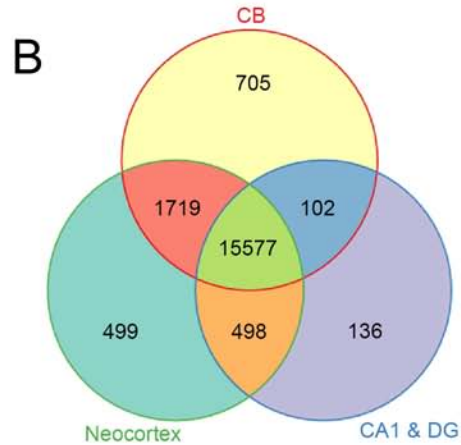
(C) Co-expression network presentation of mRNAs that were negatively regulated by *PTBP1*. Circle and word size of the co-expressed mRNAs represent additive interaction strength (sum of correlation coefficients) among mRNAs.

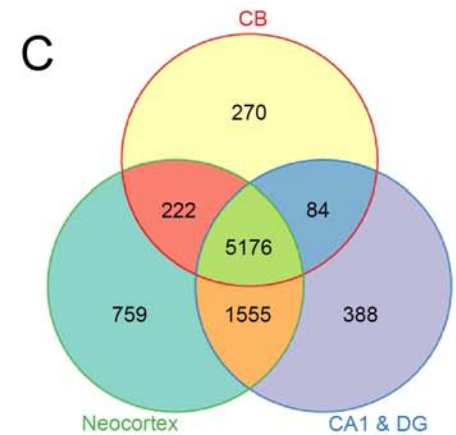
(D) Heatmap presentation of functional clustering by the negatively paired mRNAs of each lncRNA module. Color degree of each cell represents the statistical significance of pathways [-log10 (Qvalue, see label)].

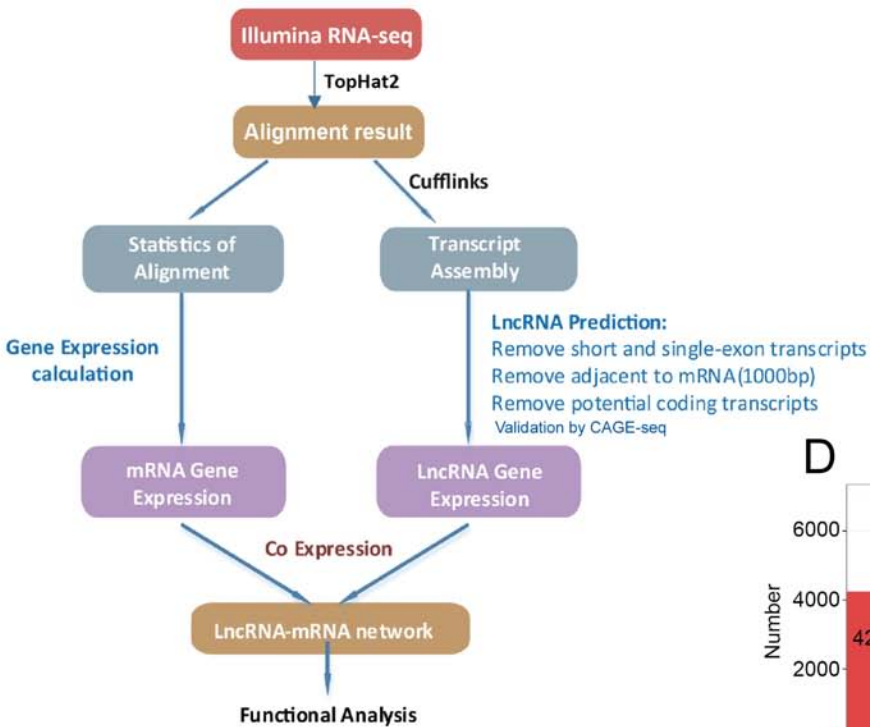
(E) Circular presentation of association between lncRNA and mRNA modules that were negatively co-regulated by lncRNAs. The length of cambered bar represents the regulating lncRNA number between lncRNAs and mRNAs, and the arc color of cambered bar represents the ratio between regulating lncRNAs and the gene number in each module. MM represents mRNA modules, and LM for lncRNA modules.

Figure 1


A
4-year
(Youth)
10-year
(Adult)
1-year
(Child)
20-year
(Old)
PC
PFC
PCC
OC
CB
TC


B
CB
705
1719
102
15577
499
498
136
Neocortex
CA1 & DG


C
CB
270
222
84
5176
759
1555
388
Neocortex
CA1 & DG


Illumina RNA-seq
TopHat2
Alignment result
Cufflinks
Statistics of Alignment
Transcript Assembly
Gene Expression calculation
LncRNA Prediction:
Remove short and single-exon transcripts
Remove adjacent to mRNA(1000bp)
Remove potential coding transcripts
Validation by CAGE-seq
mRNA Gene Expression
LncRNA Gene Expression
Co Expression
LncRNA-mRNA network
Functional Analysis

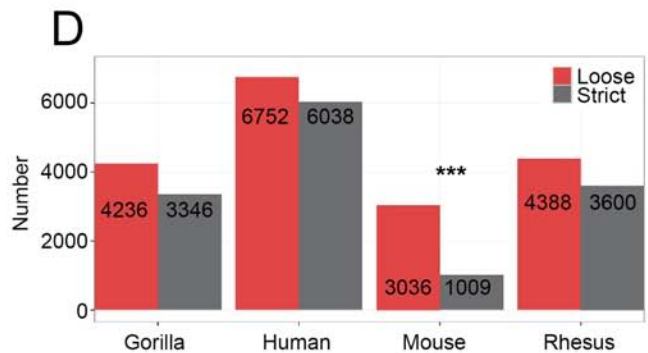

D
Loose
Strict
Number
6000
4000
2000
0
4236
3346
6752
6038
***
3036
1009
4388
3600
Gorilla
Human
Mouse
Rhesus

Figure 2

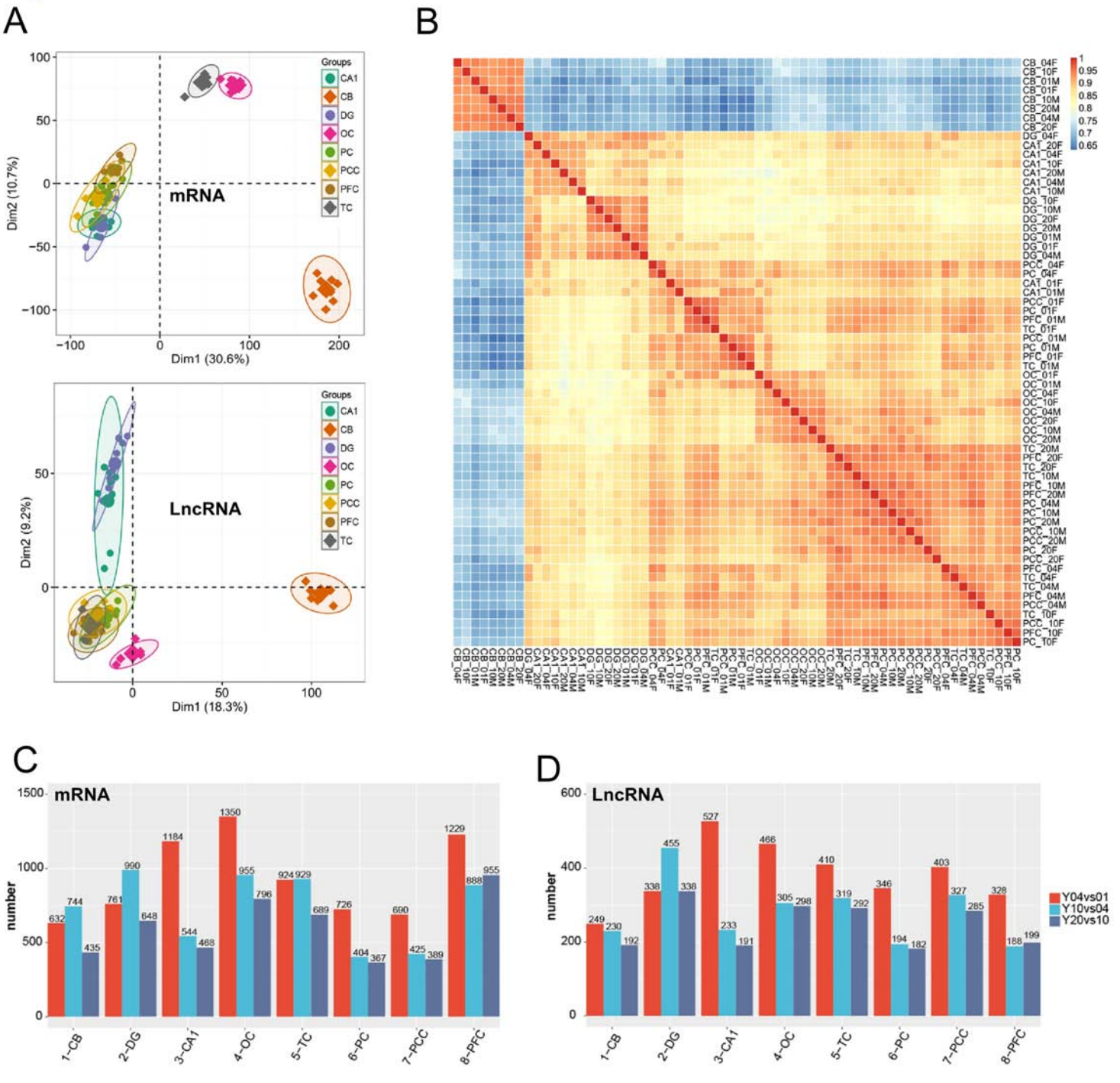

A
mRNA
Dim1 (30.6%)
Dim2 (10.7%)
LncRNA
Dim1 (18.3%)
Dim2 (9.2%)
Groups
CA1
CB
DG
OC
PC
PCC
PFC
TC
B
C
mRNA
number
1-CB
2-DG
3-CA1
4-OC
5-TC
6-PC
7-PCC
8-PFC
D
LncRNA
Y04vs01
Y10vs04
Y20vs10

Figure 3

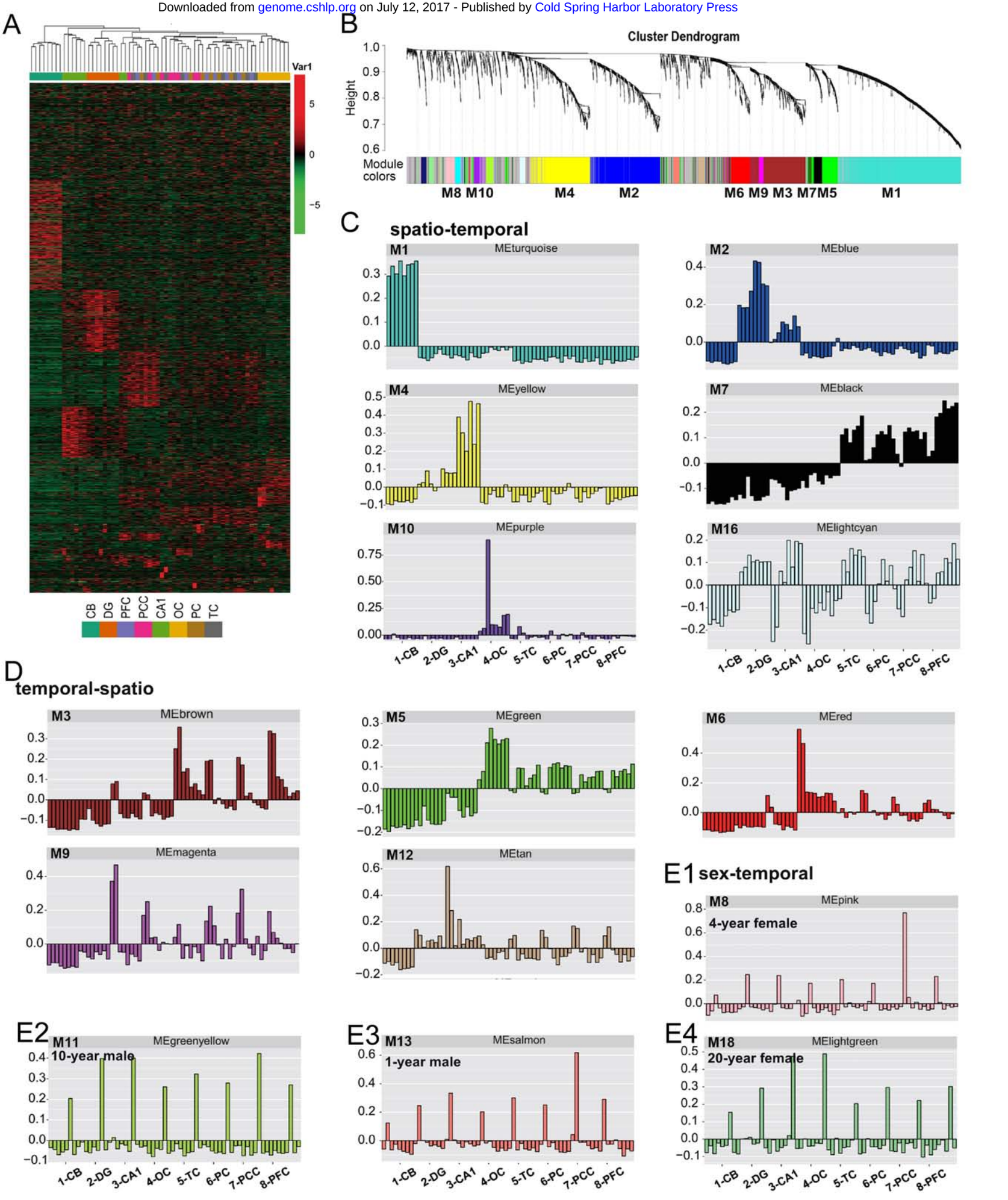


Figure 4

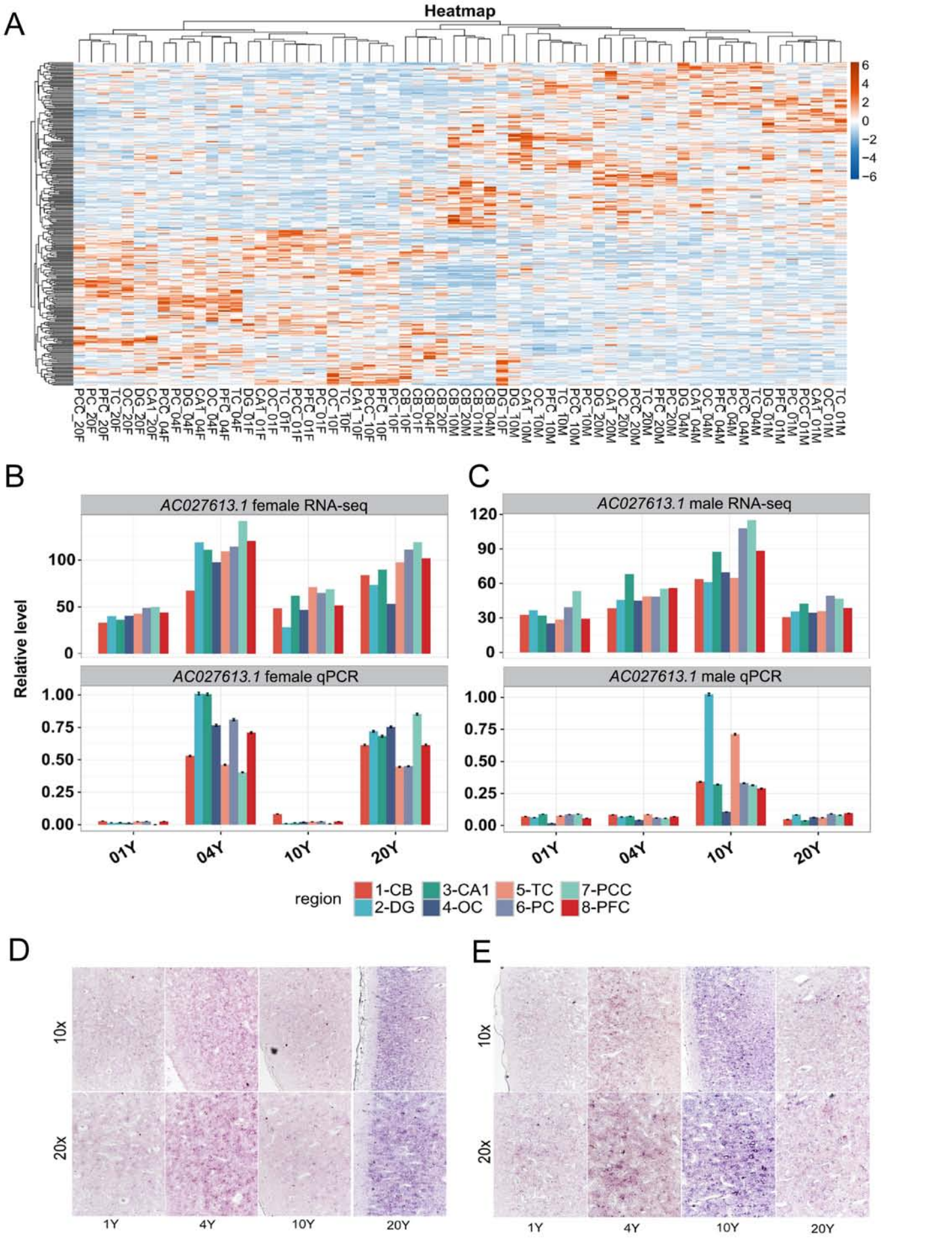

A
Heatmap
6
4
2
0
−2
−4
−6
PCC_20F
PC_20F
PFC_20F
TC_20F
OC_20F
DG_20F
CA1_20F
PCC_04F
PC_04F
DG_04F
CA1_04F
OC_04F
PFC_04F
TC_04F
DG_01F
CA1_01F
OC_01F
TC_01F
PCC_01F
PFC_01F
PC_01F
OC_10F
TC_10F
CA1_10F
PCC_10F
PFC_10F
PC_10F
CB_10F
CB_01F
CB_04F
CB_20F
CB_10M
CB_20M
CB_01M
CB_04M
DG_10F
DG_10M
CA1_10M
OC_10M
PFC_10M
TC_10M
PCC_10M
PC_10M
DG_20M
CA1_20M
OC_20M
PCC_20M
TC_20M
PFC_20M
PC_20M
DG_04M
CA1_04M
OC_04M
PFC_04M
PC_04M
PCC_04M
TC_04M
DG_01M
PFC_01M
PC_01M
PCC_01M
CA1_01M
OC_01M
TC_01M
B
AC027613.1 female RNA-seq
AC027613.1 female qPCR
Relative level
100
50
0
1.00
0.75
0.50
0.25
0.00
01Y
04Y
10Y
20Y
C
AC027613.1 male RNA-seq
AC027613.1 male qPCR
120
90
60
30
0
region
1-CB
2-DG
3-CA1
4-OC
5-TC
6-PC
7-PCC
8-PFC
D
E
10x
20x
1Y
4Y
10Y
20Y

Figure 5

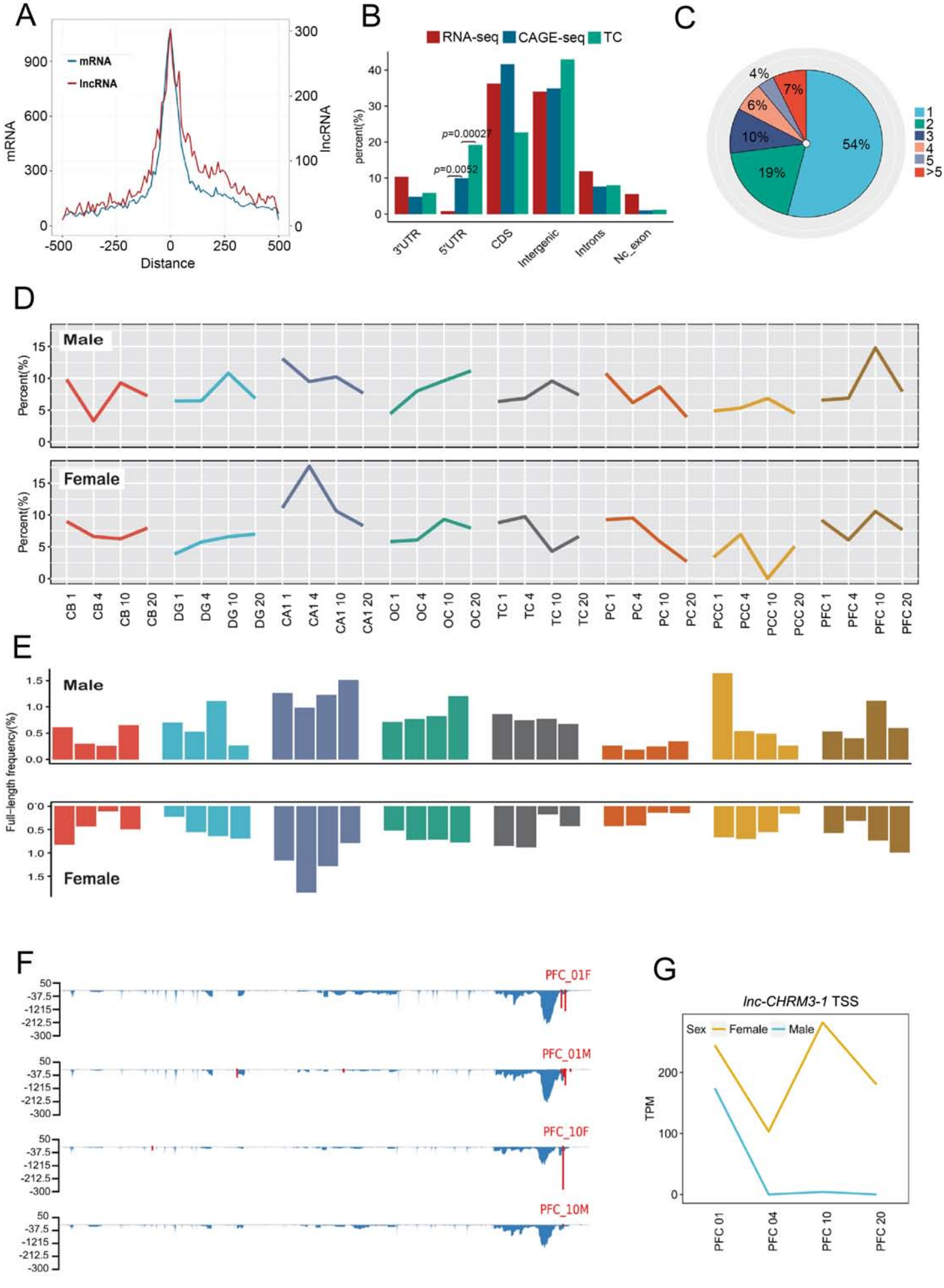


Figure 6

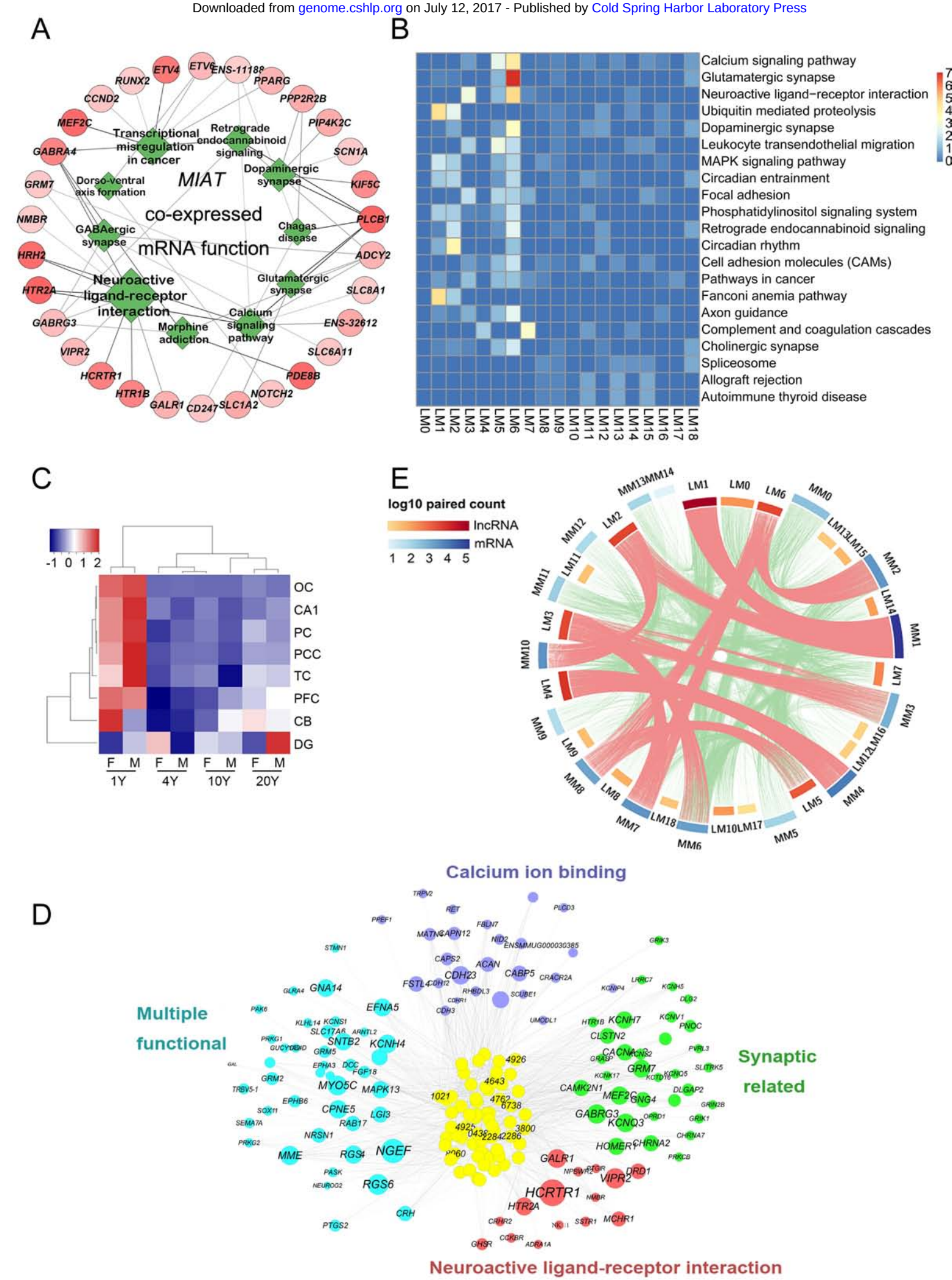
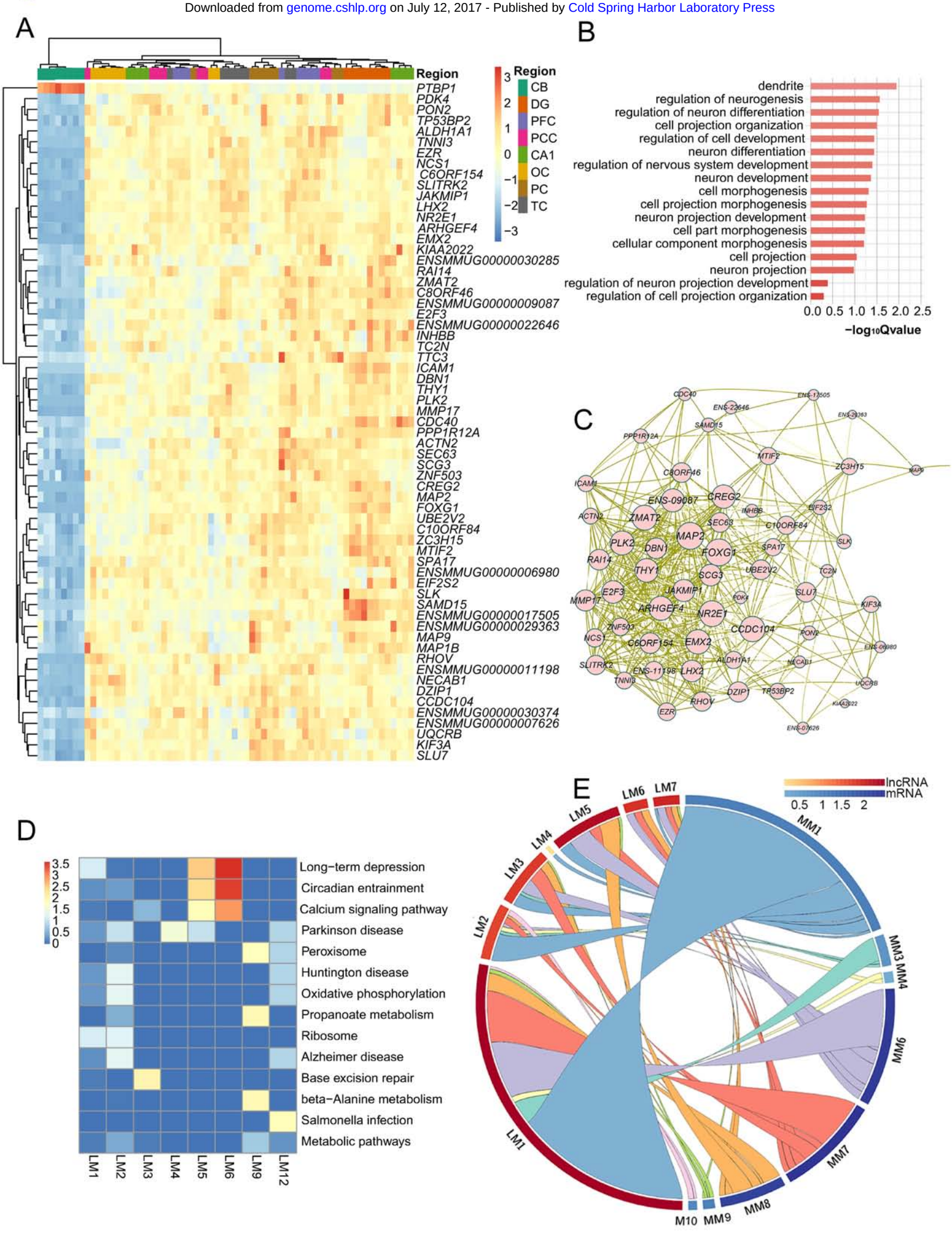


Figure 7



Annotation and cluster analysis of spatiotemporal- and sex-related lncRNA expression in Rhesus macaque brain

Siling Liu, Zhengbo Wang, Dong Chen, et al.

Genome Res. published online July 7, 2017
Access the most recent version at doi:10.1101/gr.217463.116