Journal of Molecular Cell Biology Advance Access published July 17, 2015

Integrative analyses of RNA-editing, alternative splicing, and expression of young genes in human brain transcriptome by deep RNA-sequencing

Dong-Dong Wu[1,2,†], Ling-Qun Ye[1,2,5,†], Yan Li[1,2,†], Yan-Bo Sun[1,2], Yi Shao[4], Chunyan Chen[4], Zhu Zhu[3], Li Zhong[3], Lu Wang[3], David M. Irwin[1,6,7], Yong E. Zhang[4], and Ya-Ping Zhang[1,2,3,*]

[1]State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[2]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, China
[3]Laboratory for Conservation and Utilization of Bio-resource, Yunnan University, Kunming 650091, China
[4]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, China
[5]University of Chinese Academy of Sciences, Beijing, China
[6]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada
[7]Banting and Best Diabetes Centre, University of Toronto, Toronto, Canada

[*]Correspondence to: Ya-Ping Zhang, E-mail: zhangyp@mail.kiz.ac.cn

[†]These authors contributed equally to this work.

Key words: RNA-editing, human brain transcriptome, young gene, human brain evolution

Abstract

Next-generation RNA-sequencing has been successfully used for identification of transcript assembly, evaluation of gene expression levels, and detection of post-transcriptional modifications. Despite these large-scale studies, additional comprehensive RNA-seq data from different subregions of the human brain are required to fully evaluate the evolutionary patterns experienced by the human brain transcriptome. Here, we provide a total of 6.5 billion RNA-seq reads from different subregions of the human brain. A significant correlation was observed between the levels of alternative splicing and RNA-editing, which might be explained by a competition between the molecular machineries responsible for the splicing and editing of RNA. Young human protein-coding genes demonstrate biased expression to the neocortical and non-neocortical regions during evolution on the lineage leading to humans. We also found that a significantly greater number of young human protein-coding genes are expressed in the putamen, a tissue that was also observed to have the highest level of RNA-editing activity. The putamen, which previously received little attention, plays an important role in cognitive ability, and our data suggest a potential contribution of the putamen to human evolution.

Introduction

During evolution in the lineage leading to humans, the brain has expanded greatly in size and complexity, particularly the cerebral cortex (Carroll, 2003; Hill and Walsh, 2005; Rakic, 2009; Geschwind and Rakic, 2013). Owing to the brain, our species posses high cognitive features that make us unique with special characters such as the ability to speak languages, strong prosocial and cooperative behavior, and abstract thinking (Somel et al., 2013). Large scale studies have suggested many potential genetic mechanisms for these changes in brain function including changes in gene expression levels (Cáceres et al., 2003; Preuss et al., 2004; Khaitovich et al., 2006a, b), recruitment of new genes (Li et al., 2010; Wu et al., 2011; Zhang et al., 2011), and changes in alternative splicing (Calarco et al., 2007; Johnson et al., 2009; Lin et al., 2010). However, the functional and phenotypic consequences of these changes largely remain unclear. Changes in gene expression, which has been considered to be the major driving force of phenotypic evolution, has been systematically studied for the evolution of different regions of the brain (Oldham et al., 2008; Johnson et al., 2009; Miller et al., 2010; Kang et al., 2011; Hawrylycz et al., 2012; Konopka et al., 2012). The conclusions of many of these studies, however, are affected by some technological limitations. For example, many young new genes lack functional annotation and are not represented in gene expression microarrays (Zhang et al., 2012). It is also difficult to detect post-transcriptional modifications, such as alternative splicing and RNA-editing, by many of the technologies used in these previous studies. Next generation RNA-sequencing provides sequence information for every transcript in a sample, including RNA molecules revised by post-transcriptional modification, which can include alternative splicing, alternative polyadenylation, and RNA-editing. While this technology has been used to capture transcriptomic

information from the human brain (http://www.brainspan.org), complementary RNA-seq data from different subregions of the human brain is needed to better explore the role of post-transcriptional modifications in the human brain transcriptome in the evolution of brain function.

Here, we sequenced poly $(A)^+$ RNA transcriptomes for 4 subregions of the neocortex and 10 subregions of non-neocortical areas of the brain to study the evolutionary diversity and complexity of the human brain transcriptome, which allowed us to evaluate the roles of processes such as alternative splicing, changes in gene expression, and RNA-editing. The functions of these regions are described in Supplementary material.

Results

Identification of A-to-I RNA-editing sites

A-to-I RNA editing is a widespread post- or co-transcriptional modification of nucleotides in RNA that is catalyzed by the adenosine deaminase acting on RNA (ADAR) family of enzymes (Bass, 2001; Nishikura, 2010; Li and Church, 2013). Many previous studies have found RNA-edited sites in mRNAs that result in changed amino acids in genes involved in the nervous system, and have proposed potential roles for RNA-editing in the nervous system (Li and Church, 2013). However, most observed A-to-I editing sites occur in non-coding regions of genes, and are pervasive in Alu repeats, where double stranded RNAs is commonly formed and targeted by the ADARs. With the advent of second-generation genome sequencing, RNA editing sites can be called with high confidence using RNA sequencing data (Bahn et al., 2012; Peng et al., 2012; Ramaswami et al., 2012, 2013). Here, we developed a pipeline, based on previous approaches (Ramaswami et al., 2012, 2013), that included many

steps to filter potential false positive sites, attributable to sequencing and mapping errors, to detect A-to-I RNA editing sites in human transcriptomes (Supplementary Figure S1A). RNA-editing sites were grouped into Alu sites, repetitive non-Alu sites, and non-repetitive sites according to their location, as previously described (Ramaswami et al., 2012, 2013). A comparison of our results to the datasets generated by Ramaswami et al. (2013), Pinto et al. (2014), and the DARNED database (Kiran et al., 2012) was also performed (Supplementary Figure S1B and C). We validated our approach by Sanger sequencing 12 exonic RNA-editing sites (Supplementary Figure S2). Our data shows that the numbers of RNA-editing sites is positively correlated with the expression levels of the *ADAR2* and *ADAR3* genes (Supplementary Figures S3 and S4), supporting the roles of these two genes in regulating RNA-editing. However, the correlation between the expression level of *ADAR1* and the number of RNA-editing sites was not significant. Gene annotation indicated that most of the RNA-editing occurred in the introns of genes (Supplementary Figure S5), where the level of editing was also higher than that seen in other regions (e.g. UTR or coding exon) (Supplementary Figure S6). However, the editing levels calculated here may be higher than those of previous studies, e.g. Bazak et al. (2014), because we required that the frequency of the read containing the variants (i.e. RNA-editing level) be ≥ 0.1, as described in the method section, which is important for the detection of RNA-editing sites with high confidence.

We found that the number of RNA-editing sites was significantly higher in RNAs sequenced from nervous system tissues compared to RNAs obtained from other non-nervous system tissues, based on random samples of 10, 20, 30, 40, and 50 M reads (Figure 1A-C). The high number of RNA-editing sites in the nervous system may be attributable to the higher expression level of genes in the nervous system, since there

is a positive correlation between the number of RNA-editing sites and the expression level of a gene (Supplementary Figure S7). To clarify this, we grouped genes into groups based on their expression levels (FPKM). For each group, the number of RNA-editing sites found in RNA from nervous system tissues was higher than the number of RNA-editing sites in RNA from other tissues (Supplementary Figure S7), indicating that the higher numbers of RNA-editing sites in the nervous system is not simply attributable to increased gene expression levels. This finding is consistent with previous observations on the pivotal role of RNA editing in the nervous system (Nishikura, 2010; Li and Church, 2013). Gene ontology analysis of the genes possessing significantly greater numbers of RNA-editing sites (top 5%) in the nervous system also suggested roles in neuronal processes, such as neurogenesis and the synapse related processes. In contrast, gene enrichment analysis of edited genes from non-nervous system tissues did not find an association between the presence of RNA-editing and neuronal processes.

Highly edited genes show significant interaction based on the BioGRID data, including both physical and genetic interactions, in different subregions of the brain (Figure 2A, P= 4.24E-06). The mechanism underlying the interactions of the highly edited genes is unclear. Concurrence of these genes might be consistent with the finding that *ADAR* genes regulate gene expression (Wang et al., 2013), and suggest a possible regulatory role of *ADAR* genes in gene-gene interaction. However, more experimental evidence is needed to address this question.

The term “alternative splicing” is significantly enriched in the SP_PIR_KEYWORDS for highly edited genes. Previous case studies found a potential association among the different kinds of post-transcriptional modifications, such as alternative splicing and RNA-editing (Rueter et al., 1999; Penn et al., 2013; Solomon et al., 2013). Here, we

found that the number of mutually exclusive exons is significantly negatively correlated with the number of RNA-editing sites, including non-repetitive (Figure 1D), Alu (Figure 1E), and repetitive non-Alu (Figure 1F) sites when using a random sample of 20 million reads. Similar correlations were found between the numbers of skipped exons and RNA-editing sites (Supplementary Figure S8). In contrast, the number of retained intron events was positively correlated with the number of RNA-editing sites (Figure 1D-F). These correlations support a hypothesis of cross-talk existing between A-to-I editing and alternative splicing. When RNA-editing sites were annotated to genomic location, the number of intronic RNA-editing sites was found to be positively correlated with the number of retained intron, and negatively correlated with the number of mutually exclusive exons (Supplementary Figure S9). In contrast, the number of RNA-editing at other genomic locations, such as the UTR or exons, did not show a correlation with alternative splicing.

To better study the function of A-to-I RNA-editing, we retrieved sites that yield changed amino acids in genes expressed in the nervous system (Table 1, lists all genes having at least 2 sites, Supplementary Table S2 shows all genes having amino acid-changing sites). As previously reported (Burns et al., 1997; Seeburg and Hartner, 2003; Gardiner and Du, 2006; Streit et al., 2011; Gonzalez et al., 2013), five amino acid-changing sites were detected in *5-HTR2C* and one site in *KCNA1*. In addition to these genes, we also identified six glutamate receptors, i.e. *GRIK2*, *GRIA2*, *GRIA3*, *GRIA4*, *GRIK1*, and *GRM4*, which contain amino acid-changing A-to-I RNA-editing sites (Supplementary Table S2). These examples illustrate the strength of our pipeline for detecting RNA-editing sites. We also identified amino acid changing sites in many other genes, although the functional consequences of these RNA-edits are unclear. We performed a gene enrichment analysis of these genes, and found that these genes are

enriched in the categories “glutamate receptor signaling pathway”, and also in two immunity related pathways, “acute inflammatory response” (with 9 genes, GO: 0002526, *P*=0.031), and “Complement and coagulation cascades” (9 genes, KEGG: 04610, *P*=0.00015) (Supplementary Table S3). These results suggest that RNA-editing might play a role in the immune system. The function of most identified RNA-editing sites remained unknown, but we believe that these sites may be promising targets for further experimental studies.

Comparison of transcriptome complexity among different tissues

Expression levels of genes/transcripts were calculated by implementing the Tophat-Cufflinks pipeline (Trapnell et al., 2009, 2010) (see Materials and methods). We first evaluated transcriptomic complexity by determining the number of genes that were expressed and the number of alternative splicing events, for each sample. Consistent with previous reports (Ramsköld et al., 2009; Soumillon et al., 2013), the testis presented the highest complexity, with the highest number of expressed genes (*Figure 3A*), including lincRNA (*Figure 3B*) and protein-coding genes (*Figure 3C*) revealed using different cutoff values (FPKM>1, 2, 4, 8, and 16), and the highest levels of alternative splicing, revealed by the number of alternative splicing events, including mutually exclusive exons, skipped exons, retained introns, alternative donor sites, and alternative acceptor sites (*Figure 3D*). Using additional data from the different subregions of the brain, we did not observe any significant difference in the mean transcription level (*Figure 3A-C*), or a significant difference in the level of alternative splicing between the nervous system and other tissues (excluding nervous system and testis) (*Figure 3D*). In contrast, we observed that many modules (gene co-expression network) showed tissue specific patterns, representing core gene networks operating

in each tissue (presented below). Genes do not work independently, but instead interact with each other to form complicated networks, to determine a phenotype. These data suggest a more remarkable contribution of gene networks to evolution, than of single genes, and perhaps in phenotypic evolution, gene networks might play a more significant role.

Gene co-expression network analysis of the human brain transcriptome

Weighted gene co-expression network analysis (WGCNA) was performed to identify modules of highly co-regulated genes that correlate with specific tissues. This type of analysis has successfully been employed to deduce a neurogenic disease causing network (Parikshak et al., 2013; Willsey et al., 2013), and regulatory networks in the human brain that are distinct from those of other primates (Oldham et al., 2008; Johnson et al., 2009; Kang et al., 2011; Hawrylycz et al., 2012; Konopka et al., 2012). The unsupervised and unbiased analysis identified a total of 40 distinct co-expression modules corresponding to clusters of correlated genes (Figure 4A), which can indicate regulating relationships or functional relatedness. Most of the modules showed tissue specific patterns, representing core gene networks operating in each tissue (Supplementary Figure S10).

We found that the proportion of expressed lincRNA genes is higher in the nervous system and the testis than in other tissues (Figure 4B) (P=7.41E-80 and 6.62E-74 by χ^2 test), consistent with previous observations of pervasive transcription of lincRNAs in these tissues (Mercer et al., 2008; Cabili et al., 2011; Derrien et al., 2012; Necsulea et al., 2014). Using ChIP-seq transcription-factor-binding data, we found that the promoters of the lincRNA genes were more frequently associated with transcription factors than regions 100 kb, 1 Mbp, and 2 Mbp away from the genes (Figure 4C,

P=0.006 by Wilcoxon Signed Ranks Test, Supplementary Figure S11), supporting the active regulation and function of the lincRNA genes. Conservation of exon sequences in protein-coding and lincRNA genes was stronger in the nervous system than in other non-nervous system tissues (Figure 4D, *P*=0.0005 and 0.027 by Mann-Whitney U test), supporting the existence of stronger selective constraints operating on these genes due to their important functions in the nervous system. Sequence conservation was lowest in the testis, consistent with the rapid evolution of genes specifically expressed in the testis (Figure 4D). Introns and gene sequences also demonstrate similar patterns of conservation (Supplementary Figure S12).

Genes with higher degrees of connectivity within a module are expected to play more important functional roles in biological networks, and those with the highest degree of within-module connectivity are termed hub genes. LincRNA genes showed significantly higher levels of connectivity than protein-coding genes in modules correlated with the nervous system (*P*=7.38E-06 by Mann-Whitney U test) and the testis (*P*=6.33E-48 by Mann-Whitney U test) (Figure 4E). In contrast, lincRNAs had lower connectivity than protein-coding genes in other tissues, although the difference was not statistically significant (*P*=0.192 by Mann-Whitney U test) (Figure 4E). However, since lincRNAs show strong tissue-specific expression with high tissue specificity in the brain and testis (Figure 4F), this might consequently lead to higher levels of connectivity with other genes. To exclude the confounding issue of tissue specificity, we divided the lincRNAs and protein-coding genes into different bins according to their tissue specificity (0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, and 0.9-1). At lower values of tissue specificity fewer lincRNAs were found, and no lincRNA was found with tissue specificity lower than 0.5 (Figure 4G). Within each of the five bins examined, protein-coding genes demonstrated higher value of connectivity than

lincRNA genes (Figure 4G), suggesting a more important function of protein-coding genes in the regulatory network.

Hub genes can be grouped into date hubs that display low co-expression with their partners, and party hubs having high co-expression (Han et al., 2004). It was proposed that party hubs are local coordinators whereas date hubs are global connectors. Party hubs interact with most of their partners simultaneously, while date hubs bind different partners at different locations and times (Han et al., 2004). We classified our hub genes according to (Han et al., 2004), and then calculated the mean Pearson correlation Coefficient (PCC) of expression values of these hub genes (connectivity >0.95 and 0.99) with other genes. We found that lincRNA hub genes harbor significantly higher PCC values than protein-coding hub genes (P= 6.87E-41 and 0.029, respectively, Figure 4H), which means that the lincRNA hubs tend to be party hubs, while the protein-coding hubs tend to be date hubs. This result is consistent with the tissue and developmental stage-specific function of lincRNA acting as party hubs and local coordinators.

Contribution of young protein-coding genes to the evolution of the human brain

New genes provide crucial material for the evolutionary innovation of phenotypes such as development, behavior, and reproduction (Chen et al., 2013; Long et al., 2013). Our previous meta analyses showed that primate-specific new genes tend to be upregulated in the fetal neocortex relative to non-neocortical regions and tend to be transcribed more in the fetal brain compared to the adult brain (Wu et al., 2011; Zhang et al., 2011). Although previous work provided the important insight that new genes are implicated in human brain evolution, questions remained as to which subregion(s), if any, had the most important role(s). Here, by taking advantage of our relatively

greater comprehensive sampling of brain subregions and the tissue profiling data generated by the Illumina RNA-seq platform, we asked again whether primate-specific genes showed enriched transcription in the brain. By inferring the presence and absence of orthologs on the vertebrate phylogenetic tree based on syntenic genomic alignment (Zhang et al., 2010, 2011), we identified 929 primate-specific genes in the Ensembl database released on September 2013, which were referred as the young gene dataset (Figure 5A). Compared to an older dataset of 2000 primate-specific genes, based on Ensembl released in 2008, this dataset is more conservative but possibly more reliable (Zhang et al., 2012).

Young genes harbor lower levels of alternative splicing, and lower levels of connectivity than old genes (Supplementary Figure S13). Interestingly, young genes harbor more RNA-editing sites in the 3'UTR than old genes (Supplementary Figure S13), which is not attributable to the length of 3'UTR, as young protein-coding genes harbor shorter 3'UTR than old genes.

As expected, young genes showed a higher level of expression in the testis (Figure 5B) (P=0.0002). However, we did not find a higher level of expression for the neocortex compared to other parts of the nervous system, demonstrating that the previous observation of biased expression towards the fetal neocortex (Zhang et al., 2011, 2012) could not be generalized to adult tissue (Figure 5B). We also used another strategy to evaluate expression levels by using the percentile rank of FPKM value of each gene within each tissue, which yielded a consistent result (Supplementary Figure S14). Young genes show higher mean FPKM in the fetal brain compared to adult brain (Figure 5C), which is likely partially due to the upregulation of young genes in the fetal neocortex. To trace when this bias in expression first appeared, we calculated the relative expression levels of genes that originated at different stages of primate

evolution. The bias was defined as the proportion of expression in the fetal brain relative to total expression across all tissues for one gene of interest. We found that the excess in fetal brain expression relative to adult brain expression began in the common ancestor of eutherian mammals, and continued on almost all stages of evolution towards modern humans (Supplementary Figure S15). Upregulated relative gene expression was observed on the primate lineages leading to humans for the neocortex, and in the catarrhine lineage leading to humans for the non-neocortex (Figure 5D). In contrast, in the testis, relative expression levels decreased with the age of the genes, and in other tissues, the relative expression levels increased with age (Figure 5D).

To better explore the expression patterns of young genes, we analyzed the enrichment levels of genes within gene co-expression modules (Figure 5E, F). Young genes showed relatively higher levels of enrichment in modules correlated with nervous system tissues compared to other tissues (Figure 5E, P=0.033 by Mann-Whitney Test). We also calculated the enrichment levels of genes with different ages in the modules correlated with nervous system, testis, and other tissues (Supplementary Figure S16). Consistent with the above, enrichment levels of genes in modules correlated with the nervous system increased in the primate lineage, enrichment levels of genes in modules correlated with the testis decreased with the age of genes, and for other tissues, enrichment levels increased with age.

The turquoise module, which correlated significantly with the testis, was enriched with the highest number of young genes, with three other modules, labeled in greenyellow, magenta, and orange that correlated with the insula, putamen, and nucleus accumbens, respectively, also containing significantly increased numbers of young genes (36, 29, and 9 genes, respectively) (Figure 5F). The relative expression

level of the primate-specific genes in the insula and putamen increased in the primate lineages leading to humans (P= 0.03, and P=0.006, respectively by Pearson Correlation, Supplementary Figure S17). Gene ontology analysis of genes in the module correlated with the putamen found a significant enrichment of 21 genes involved in "G-protein coupled receptor signaling pathway", 20 genes in "behavior", 6 genes in "dopamine receptor signaling pathway", 7 genes in "learning", 6 genes in "memory", and 27 genes in "synaptic transmission" (Supplementary Table S4). Similar results were obtained using the GOseq tool to correct for gene length in an enrichment analysis (http://bioconductor.org/packages/release/bioc/html/goseq.html) (Supplementary Table S5). In addition, the regulatory network found in the putamen module is also supported by the significantly frequent interactions among these genes based on the BioGRID interaction data (Figure 2B). Using more fine scale transcriptomic data reported previously (Hawrylycz et al., 2012), we also found that young protein-coding genes harbored their highest level of expression in the putamen (Supplementary Figure S18). Young primate-specific genes in the putamen module might have contributed to the evolution of cognitive ability during the evolution of humans. In contrast, gene ontology analysis of the modules correlated with the insula did not find an enrichment of categories associated with the nervous system, but instead with categories associated with olfactory, consistent with the role of insula in sensory systems (e.g. olfactory, gustatory). Similarly, 16 and 29 primate specific genes are found to be significantly enriched in modules correlated with the adrenal gland and the colon, respectively. The functions of these young genes are unclear, but might also contribute to phenotypic evolution in primates.

In addition, we also studied the expression patterns of 176 human-specific protein-coding genes. These new genes, which originated on the human lineage after

divergence from the chimpanzee (branch 13 in Figure 5A), showed significantly higher relative expression levels in the nervous system relative to other tissues, than genes that originated on the lineage leading to humans and chimpanzee (branch 12 in Figure 5A) (Supplementary Figure S19, P= 0.022 by Mann-Whitney U test). Significant enrichment was observed in the testis-correlated module (71 genes), with another 16 new genes significantly enriched in the insula-correlated module. However, all of these new genes showed very low levels of expression in all tissues, and the function of these genes remains largely unexplored.

Discussion

In this study, we provide deep RNA-sequencing of poly $(A)^+$ RNAs from different subregions of the human brain, which we used to study the expression pattern of long non-coding RNA, newly originated genes, and the post-transcriptional modification, including alternative splicing and RNA-editing, of RNA transcripts.

Genetic changes that contributed to the rapid evolution of the human brain and the evolution of human cognitive abilities remain largely unclear, although research has identified many genetic processes, including changes in gene expression, rapid evolution of regulatory elements, protein-coding substitutions, and the emergence of young new genes, which may be involved. The increased size of the human brain particularly affects the cerebral cortex, the largest brain structure and the seat of higher cognitive functions, therefore, most studies have focused on genetic novelties that appear in this region of the brain (Hill and Walsh, 2005; Rakic, 2009).

In contrast to rodent-specific genes, a significantly larger proportion of primate-specific genes are expressed in the brain, with an excess of the new genes recruited to the fetal neocortex (Zhang et al., 2011). Wu et al. (2011) also reported a relatively

high level of expression for human-specific de novo genes in the adult cerebral cortex. Both of these studies lacked a comprehensive sampling that covered the different subregions of the brain as their data was generated from a single transcriptome platform. Here, we confirm that young genes are biased in their transcription to the fetal brain compared to the adult brain (Figure 5C), which can partially be explained by a bias in the expression of young genes to the fetal neocortex. In addition, we discovered a novel pattern with the expression of young genes being upregulated in expression in the adult neocortex on the primate lineage leading to humans and to non-neocortical areas on the catarrhine lineage.

We found 29 young human protein-coding genes that are significantly enriched in the module, labeled by the greenyellow color, which is correlated with the putamen. The relative expression levels of the primate-specific genes in the putamen module increased along the lineages leading to humans (Supplementary Figure S17). This conclusion is also supported by our analysis of previously reported fine scale transcriptomic data (Hawrylycz et al., 2012) (Supplementary Figure S18). Among the 29 young genes in the putamen module, 6 encode zinc finger proteins and 5 encode proteins containing ankyrin repeat domains, which likely allowed these genes to join gene-gene interaction networks due to their ability to interact with other proteins, DNA, or RNA. Previous studies have also noted that some of the young genes that originated during human evolution are zinc finger protein genes (Zhang et al., 2011). The putamen is located at the base of the forebrain and forms the dorsal striatum together with the caudate nucleus (Kreitzer and Malenka, 2008). The dorsal striatum is known for its role in many kinds of cognitive functions, including learning, decision-making (e.g. action selection and initiation), and motor behaviors including regulation of motor activity, motor skill learning and motor response (Balleine et al.,

2007; Surmeier et al., 2009; Lovinger, 2010). The dorsal striatum is also known for its contribution to neurological diseases, such as Parkinson's disease, Alzheimer's disease, and Huntington's disease (De Jong et al., 2008; Kreitzer and Malenka, 2008). For example, a strongly reduced volume of the putamen is found in Alzheimer's disease (De Jong et al., 2008). Consistent with this, we found that *HTT* (huntingtin), a disease gene linked to the neurodegenerative disease Huntington's disease characterized by loss of striatal neurons, is expressed in the putamen module.

The dorsal striatum serves as the primary gateway to the basal ganglia, and receives excitatory afferents from the cortex and the thalamus (Kreitzer and Malenka, 2008). The vast majority of the neurons in the striatum are medium spiny neurons (MSNs), which are categorized into two different types, i.e. striatonigral MSNs that highly express the dopamine receptor D1, and striatopallidal MSNs that exhibit a high expression level of dopamine receptor D2 (Kreitzer and Malenka, 2008). Dopamine is a neurotransmitter, and the dopamine signal plays an essential role in the function of the dorsal striatum. Consistent with the above, in the module correlated with the putamen, we also found specific expression of the dopamine receptors D1, D2, and D3. In addition, the gene gephyrin, which encodes a neuronal assembly protein that anchors inhibitory neurotransmitter receptors to the postsynaptic cytoskeleton, is also involved in the putamen module.

Evidence for RNA-editing in some of these genes had been found in previous studies. For example, RNA-editing in the serotonin 2C receptor, *5-HT2RC,* alters the amino acid sequence of the encoded protein, with a total of 5 potential editing sites yielding 32 mRNA sequence variants, which encode 24 protein isoforms, that vary in biochemical and pharmacological properties (Gardiner and Du, 2006). Changes in the level of *5-HT2RC* RNA editing have been reported in individuals suffering from

schizophrenia, major depression, or committed suicide (Gardiner and Du, 2006). *KCNA1* encodes a voltage-gated delayed potassium channel, and an RNA editing event (I400V) alters the inner permeation pathway of human KV1.1, modifying the kinetics of fast inactivation, which should greatly influence the action potential shape, signal propagation and the firing pattern of a neuron (Gonzalez et al., 2013). Among the different subregions of the brain, the putamen has the largest number of detected RNA-editing sites at non-repetitive (Figure 1A) and repetitive non-Alu (Figure 1C) sites, and the second largest number at Alu sites (Figure 1B). The cerebellum has the largest number at Alu sites. Gene ontology analysis of the top 5% genes containing the highest numbers of RNA-editing sites in the putamen, found that the highly edited genes were enriched at many neuronal categories, such as “transmission of nerve impulse”, “synaptic transmission”, “neurogenesis”, and “glutamatergic synapse”. However, the functional consequences of the A-to-I RNA-editing sites, particularly the many sites that change amino acids, are unclear.

We observed a significant correlation between the numbers of alternative splicing events and the numbers of RNA-editing sites. Competition between the splicing machinery and the molecular mechanisms involved in RNA-editing may occur on pre-RNA molecules. Here we propose a competition model that explains the correlation between alternative splicing events and RNA-editing (Figure 6). RNA-editing sites in genic regions are located mostly in introns (Supplementary Figure S5), suggesting that the RNA-editing machinery is targeted most frequently to introns. When the RNA-editing machinery competitively targets an intron, then that intron should be less likely to be spliced from the pre-mRNA by the splicing machinery, which should result in an increase in the number of retained introns. This model is supported by the positive correlation between the numbers of retained introns and RNA-editing sites.

However, skipped exons would be less likely to be spliced from the pre-mRNA, which is supported by the negative correlation between the numbers of skipped or mutually exclusive exons and RNA-editing sites. In addition, the positive correlation between retained introns and intronic RNA-editing sites, and negative correlation between mutually exclusive exons and intronic RNA-editing sites also supported the model of targeting competitively on intron.

These same processes might regulate the competition between the machineries for editing and splicing. The activity of RNA molecules or proteins involved in alternative splicing may be reduced when the corresponding RNAs were subjected to RNA-editing, and vice versa. For example, for the gene *RBFOX1*, the largest numbers of RNA-editing sites were detected in the nervous system. RBFOX1 is a member of the Fox-1 family of RNA-binding proteins and regulates tissue-specific alternative splicing. The correlation between RNA-editing and alternative splicing was further validated by the finding of the increased number of alternative splicing events when *ADAR2* gene was knock down that caused decreased RNA-editing (Supplementary Figure S20).

Materials and methods

Full materials and methods are described in Supplementary material.

RNA from the brain used in this study was obtained from Clontech (Supplementary Table S1). A single sequencing library was constructed for the poly $(A)^+$ RNA as described in the Illumina manual, and sequenced on the Hiseq 2000 sequencing platform. Sequence reads were mapped by Tophat (v 2.0.4) (Trapnell et al., 2009) to the human reference genome, and Cufflinks (v 2.1.1) was used to assemble transcripts, and cuffcompare was used to merge the newly assembled transcripts among the

different tissues (Trapnell et al., 2010).

Weighted gene co-expression network analysis (WGCNA) was applied to construct gene co-expression networks across tissues (Langfelder and Horvath, 2008)http:///. ChIP-seq peaks for transcription-factor-binding sites were downloaded from the ENCODE project (Consortium, 2012). Phastcons conservation score data for the primates was downloaded from UCSC. The tissue specificity of each gene was calculated as described in (Yanai et al., 2005).

Protein-coding genes from 23 vertebrate species were downloaded from Ensembl (www.ensembl.org, version 73) and were used for this analysis. We dated the origin of the human protein-coding genes based on the presence and absence of orthologs across the vertebrate phylogenetic tree (Figure 5A) based on the UCSC syntenic genomic alignments, as described previously (Zhang et al., 2010, 2011).

Acknowledgements

We thank Dr Rui Zhang from Stanford University for his suggestions and comments on the manuscript.

Funding

This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13000000) and the National Natural Science Foundation of China (31271339, 31301042).

Conflict of interest: none declared.

References

Bahn, J. H., Lee, J. H., Li, G., et al., (2012). Accurate identification of A-to-I RNA editing in human by transcriptome sequencing. Genome Res. *22*, 142-50.

Balleine, B. W., Delgado, M. R., and Hikosaka, O., (2007). The role of the dorsal striatum in reward and decision-making. J. Neurosci. *27*, 8161-8165.

Bass, B. L., (2001). RNA editing by adenosine deaminases that act on RNA. Annu. Rev. Biochem. *71*, 817-846.

Bazak, L., Haviv, A., Barak, M., et al., (2014). A-to-I RNA editing occurs at over a hundred million genomic sites, located in a majority of human genes. Genome Res. *24*, 365-376.

Burns, C. M., Chu, H., Rueter, S. M., et al., (1997). Regulation of serotonin-2C receptor G-protein coupling by RNA editing. Nature *387*, 303-308.

Cáceres, M., Lachuer, J., Zapala, M. A., et al., (2003). Elevated gene expression levels distinguish human from non-human primate brains. Proc. Natl Acad. Sci. USA *100*, 13030-13035.

Cabili, M. N., Trapnell, C., Goff, L., et al., (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915-1927.

Calarco, J. A., Xing, Y., Cáceres, M., et al., (2007). Global analysis of alternative splicing differences between humans and chimpanzees. Genes Dev. *21*, 2963-2975.

Carroll, S. B., (2003). Genetics and the making of Homo sapiens. Nature *422*, 849-857.

Chen, S., Krinsky, B. H., and Long, M., (2013). New genes as drivers of phenotypic evolution. Nat. Rev. Genet. *14*, 645-660.

Consortium, T. E. P., (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

De Jong, L. W., Van der Hiele, K., Veer, I. M., et al., (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. Brain *131*, 3277-3285.

Derrien, T., Johnson, R., Bussotti, G., et al., (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775-1789.

Gardiner, K., and Du, Y., (2006). A-to-I editing of the 5HT2C receptor and behaviour. Brief. Funct. Genomic Proteomic *5*, 37-42.

Geschwind, D. H., and Rakic, P., (2013). Cortical evolution: judge the brain by its cover. Cell *80*, 633-647.

Gonzalez, C., Lopez-Rodriguez, A., Srikumar, D., et al., (2013). Editing of human KV1.1 channel mRNAs disrupts binding of the N-terminus tip at the intracellular cavity. Nat. Commun. *2*, 436.

Han, J.-D. J., Bertin, N., Hao, T., et al., (2004). Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature *430*, 88-93.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., et al., (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391-399.

Hill, R. S., and Walsh, C. A., (2005). Molecular insights into human brain evolution. Nature *437*, 64-67.

Johnson, M. B., Kawasawa, Y. I., Mason, C. E., et al., (2009). Functional and evolutionary insights into human brain development through global transcriptome analysis. Neuron *62*, 494-509.

Kang, H. J., Kawasawa, Y. I., Cheng, F., et al., (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483-489.

Khaitovich, P., Enard, W., Lachmann, M., et al., (2006a). Evolution of primate gene expression. Nat. Rev. Genet. *7*, 693-702.

Khaitovich, P., Tang, K., Franz, H., et al., (2006b). Positive selection on gene expression in the human brain. Curr. Biol. *16*, R356-R358.

Kiran, A. M., O'Mahony, J. J., Sanjeev, K., et al., (2012). Darned in 2013: inclusion of model organisms and linking with Wikipedia. Nucleic Acids Res. *41*, D258-D261.

Konopka, G., Friedrich, T., Davis-Turak, J., et al., (2012). Human-specific transcriptional networks in the brain. Neuron *75*, 601-617.

Kreitzer, A. C., Malenka, R. C., (2008). Striatal plasticity and basal ganglia circuit function. Cell *60*, 543-554.

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

Li, C.-Y., Zhang, Y., Wang, Z., et al., (2010). A human-specific de novo protein-coding gene associated with human brain functions. PLoS Comput. Biol. *6*, e1000734.

Li, J. B., and Church, G. M., (2013). Deciphering the functions and regulation of brain-enriched A-to-I

RNA editing. Nat. Neurosci. *16*, 1518-1522.
Lin, L., Shen, S., Jiang, P., et al., (2010). Evolution of alternative splicing in primate brain transcriptomes. Hum. Mol. Genet. *19*, 2958-2973.
Long, M., VanKuren, N. W., Chen, S., et al., (2013). New gene evolution: little did we know. Ann. Rev. Genet. *47*, 307-333.
Lovinger, D. M., (2010). Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. Neuropharmacology *58*, 951-961.
Mercer, T. R., Dinger, M. E., Sunkin, S. M., et al., (2008). Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl Acad. Sci. USA *105*, 716-721.
Miller, J. A., Horvath, S., and Geschwind, D. H. (2010). Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc. Natl Acad. Sci. USA *107*, 12698-12703.
Necsulea, A., Soumillon, M., Warnefors, M., et al., (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. Nature *505*, 635-640.
Nishikura, K., (2010). Functions and regulation of RNA editing by ADAR deaminases. Annu. Rev. Biochem. *79*, 321-349.
Oldham, M. C., Konopka, G., Iwamoto, K., et al., (2008). Functional organization of the transcriptome in human brain. Nat. Neurosci. *11*, 1271-1282.
Parikshak, N. N., Luo, R., Zhang, A., et al., (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell *155*, 1008-1021.
Peng, Z., Cheng, Y., Tan, B. C., et al., (2012). Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. Nat. Biotechnol. *30*, 253-60.
Penn, A. C., Balik, A., and Greger, I. H. (2013). Reciprocal regulation of A-to-I RNA editing and the vertebrate nervous system. Front. neurosci. *7*, 61.
Preuss, T. M., Cáceres, M., Oldham, M. C., et al., (2004). Human brain evolution: insights from microarrays. Nat. Rev. Genet. *5*, 850-860.
Rakic, P., (2009). Evolution of the neocortex: a perspective from developmental biology. Nat. Rev. Neurosci. *10*, 724-735.
Ramaswami, G., Lin, W., Piskol, R., et al., (2012). Accurate identification of human Alu and non-Alu RNA editing sites. Nat. Methods *9*, 579-81.
Ramaswami, G., Zhang, R., Piskol, R., et al., (2013). Identifying RNA editing sites using RNA sequencing data alone. Nat. Methods *10*, 128-32.
Ramsköld, D., Wang, E. T., Burge, C. B., et al., (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol. *5*, e1000598.
Rueter, S. M., Dawson, T. R., and Emeson, R. B. (1999). Regulation of alternative splicing by RNA editing. Nature *399*, 75-80.
Seeburg, P. H., and Hartner, J., (2003). Regulation of ion channel/neurotransmitter receptor function by RNA editing. Curr. Opin. Neurobiol. *13*, 279-283.
Solomon, O., Oren, S., Safran, M., et al., (2013). Global regulation of alternative splicing by adenosine deaminase acting on RNA (ADAR). RNA *19*, 591-604.
Somel, M., Liu, X., and Khaitovich, P. (2013). Human brain evolution: transcripts, metabolites and their regulators. Nat. Rev. Neurosci. *14*, 112-127.
Soumillon, M., Necsulea, A., Weier, M., et al., (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. *3*, 2179-2190.
Streit, A. K., Derst, C., Wegner, S., et al., (2011). RNA editing of Kv1. 1 channels may account for reduced ictogenic potential of 4-aminopyridine in chronic epileptic rats. Epilepsia *52*, 645-648.
Surmeier, D. J., Plotkin, J., and Shen, W. (2009). Dopamine and synaptic plasticity in dorsal striatal circuits controlling action selection. Curr. Opin. Neurobiol. *19*, 621-628.
Trapnell, C., Pachter, L., and Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.
Trapnell, C., Williams, B. A., Pertea, G., et al., (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotech. *28*, 511-515.
Wang, I. X., So, E., Devlin, J. L., et al., (2013). ADAR regulates rna editing, transcript stability, and gene expression. Cell Rep. *5*, 849–860.
Willsey, A. J., Sanders, S. J., Li, M., et al., (2013). Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell *155*, 997-1007.
Wu, D.-D., Irwin, D. M., and Zhang, Y.-P., (2011). De novo origin of human protein-coding genes. PLoS. Genet. *7*, e1002379.
Yanai, I., Benjamin, H., Shmoish, M., et al., (2005). Genome-wide midrange transcription profiles

reveal expression level relationships in human tissue specification. Bioinformatics *21*, 650-659.

Zhang, Y. E., Landback, P., Vibranovski, M., et al., (2012). New genes expressed in human brains: Implications for annotating evolving genomes. Bioessays *34*, 982-991.

Zhang, Y. E., Landback, P., Vibranovski, M. D., et al., (2011). Accelerated recruitment of new brain development genes into the human genome. PLoS Biol. *9*, e1001179.

Zhang, Y. E., Vibranovski, M. D., Landback, P., et al., (2010). Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. PLoS Biol. *8*, e1000494.

Figure legends

Figure 1 A-to-I RNA-editing in different tissues. Numbers of RNA-edited sites at non-repetitive (**A**, $\times 10^3$), Alu (**B**, $\times 10^4$), and repetitive non-Alu (**C**, $\times 10^3$) sites based on sets of randomly chosen reads (10, 20, 30, 40, and 50 M). Correlation of the number of RNA-editing sites at the non-repetitive (**D**, $\times 10^3$), Alu (**E**, $\times 10^4$), and repetitive non-Alu (**F**, $\times 10^3$) sites with the number of mutually exclusive exons (left) and retained introns (right). Numbers of alternative splicing events and RNA-editing sites were detected using sets of 20 M randomly selected reads.

Figure 2 Gene−gene interactions in the putamen. (**A**) Interaction map of the top 5% genes having the highest A-to-I RNA-editing sites in the putamen. (**B**) Interaction map of genes in the module correlated with the putamen.

Figure 3 Transcriptome complexity analysis. (**A**) Numbers of genes, including lincRNA and protein-coding genes, with expression values (FPKM) higher than 1, 2, 4, 8, and 16 in the testis, nervous system, and other tissues. (**B**) Numbers of lincRNA genes with expression values (FPKM) higher than 1, 2, 4, 8, and 16 in the testis, nervous system, and other tissues. (**C**) Numbers of protein-coding genes with expression values (FPKM) higher than 1, 2, 4, 8, and 16 in the testis, nervous system, and other tissues. (**D**) Numbers of mutually exclusive exons, skipped exons, retained exons, and alternative splice acceptor and donor sites in the testis, nervous system, and other tissues.

Figure 4 Comparison of expression patterns in the nervous system and other tissues. (**A**) Weighted gene co-expression network analysis (WGCNA) identified 40 modules labeled by different colors. (**B**) Proportion of different types of genes in the modules correlated with different tissues. The dashed horizontal line was used to present the value of the proportion of lincRNAs in other tissues. (**C**) Mean numbers of transcription factors (± standard errors) (i.e. mean numbers of ChIP-seq peaks)

targeting the promoters (0-2 kb upstream) and the regions 100-102 kb upstream (treated as random regions) of lincRNA genes in the modules correlated with the nervous system, testes, and other tissues. (**D**) PhastCons scores of exons of the lincRNA genes (left Y-axis), and protein-coding genes (right Y-axis), in the modules correlated with the nervous system, testes, and other tissues. (**E**) Connectivity (± standard errors) of the lincRNA and protein-coding genes in the modules correlated with the nervous system, testes, and other tissues. (**F**) Tissues specificity (± standard errors) of the lincRNA and protein-coding genes in the modules correlated with the nervous system, testes, and other tissues. Colors of the lincRNA and protein-coding genes are the same as in **B**, **D**, **E**, **F**. (**G**) Connectivity and number of lincRNA and protein coding genes within different bins of tissue specificity. (**H**) PCC (Pearson correlation coefficient) of lincRNA and protein coding hub genes (with connectivity >0.95 and 0.99).

Figure 5 Expression analysis of young protein-coding genes. (**A**) Phylogenetic tree used for dating the ages of genes in the human genome. Genes originating on branch 13 are human-specific genes, and genes on the branch 8-13 are primate-specific genes. (**B**) Normalized expression level of young protein-coding genes calculated from the mean expression value of young genes divided by the mean expression value of whole genome wide genes in each tissue. (**C**) Mean expression level of primate specific genes in the whole fetal brain and whole adult brain. (**D**) Mean relative expression level of genes originating on different phylogenetic branches described in **A**, in the neocortex (left Y-axis), non-neocortex (left Y-axis), testes (right Y-axis), and other tissues (left Y-axis). Relative expression level of each gene in each tissue is calculated by the expression level in the tissue divided by the total expression level of the gene in the sampled tissues. (**E**) Mean enrichment level (+ the largest enrichment value) of

primate-specific genes in modules correlated with the nervous system, testes, and other tissues. Enrichment level of young protein-coding genes in each module is calculated by the proportion of young protein-coding genes in the module divided by the proportion of other genes in the module. (**F**) Enrichment level of primate-specific genes in each module. * represents modules containing significant enrichment in young genes and correlated tissues ($P<0.01$ by χ^2 test corrected by bonferroni correction). Numbers in parentheses are the numbers of primate-specific genes.

Figure 6 A competition model to explain the correlation between alternative splicing and RNA-editing. When the RNA-editing machinery targets an intron, the intron should then be less likely to be spliced from the pre-mRNA by the splicing machinery, thus inducing a retained intron. This is supported by the positive correlation between the numbers of retained introns and RNA-editing sites. However, skipped exons are less likely to be spliced from the pre-mRNA, which is supported by the negative correlation between the numbers of skipped or mutually exclusive exons and RNA-editing sites.

Table 1 Numbers of A-to-I RNA-editing sites that change amino acid in edited genes.

SORBS1	6	3	8	0	3	1	0	4	3	8	4	2	0	7	8	0
CDK13	4	3	0	1	0	0	0	0	0	0	1	3	4	3	0	1
C11orf80	3	2	0	0	0	0	0	0	1	0	1	0	1	0	2	0
CCNJL	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GRIK2	3	3	0	1	1	3	2	3	1	1	3	3	1	3	3	0
SLC38A6	3	1	0	0	0	0	0	0	0	0	1	2	1	0	3	1
ZNF587B	3	3	0	1	2	0	0	1	2	1	1	2	2	1	2	1
C16orf89	2	1	0	0	0	0	0	1	0	0	0	3	1	1	0	1
CACNA1D	2	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0
RHBG	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HIF3A	1	1	0	0	1	0	2	0	0	0	2	1	2	0	0	0
HSD11B1L	1	1	0	0	2	1	1	0	1	2	1	0	2	2	0	1
MICAL3	1	0	0	0	0	0	0	1	0	0	1	0	0	1	2	0
TAF1L	1	3	0	0	0	2	1	2	1	0	0	1	0	2	1	0
TUBGCP2	1	2	3	0	0	0	0	1	0	0	1	1	2	0	2	1
ZNF891	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
HTR2C	0	1	3	0	0	0	0	0	2	1	0	0	0	2	5	2
MAGEL2	0	1	0	0	0	0	0	0	0	0	1	2	1	0	0	0
PDCD7	0	1	0	0	0	0	0	0	0	0	1	1	1	2	0	0
TRIM54	0	4	0	0	0	0	0	0	0	0	0	0	1	0	0	0
NOP2	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
SDHD	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
CCNE2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
PDE5A	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
FGB	0	0	0	0	0	1	0	1	0	1	0	1	1	1	2	0
FGG	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0
ADAM20	0	0	0	0	0	0	0	1	0	0	0	4	0	0	1	0
FAM193A	0	0	0	0	0	0	0	2	0	0	1	0	1	1	2	0
MAT1A	0	0	0	0	0	0	0	0	1	0	2	0	1	1	0	0
ZNF586	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
SLC4A8	0	0	0	0	0	0	0	0	0	0	1	0	2	1	0	0

Note: Tissues from left to right are cerebellum, cerebral cortex, corpus callosum, dura mater, fetal brain, frontal lobe, hippocampus, insula, nucleus accumbens, oblongata, parietal lobe, pons, posterior central gyrus, precentral gyrus, putamen, and substantia nigra, respectively. Only genes having more than one site in at least one tissue are shown. Genes having amino-acid changing RNA-editing site are presented in Supplementary Table S2.