

Changes in snoRNA and snRNA Abundance in the Human, Chimpanzee, Macaque, and Mouse Brain

Bin Zhang^{1,2,†}, Dingding Han^{1,†}, Yuriy Korostelev³, Zheng Yan^{1,4}, Ningyi Shao¹, Ekaterina Khrameeva⁵, Boris M. Velichkovsky⁶, Yi-Ping Phoebe Chen⁴, Mikhail S. Gelfand^{3,7,*}, and Philipp Khaitovich^{1,5,8,9,*}

¹CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai, China

²Graduate School of Chinese Academy of Sciences, Beijing, China

³Department of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia

⁴Department of Computer Science and Computer Engineering, La Trobe University, Melbourne, VIC, Australia

⁵Skolkovo Institute for Science and Technology, Skolkovo, Russia

⁶Department of Neuroscience, National Research Center "Kurchatov Institute", Moscow, Russia

⁷Institute for Information Transmission Problems, RAS, Moscow, Russia

⁸Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

⁹School of Life Science and Technology, ShanghaiTech University, Shanghai, China

*Corresponding author: E-mail: mikhail.gelfand@gmail.com; khaitovich@eva.mpg.de.

[†]These authors contributed equally to this work.

Accepted: February 18, 2016

Data deposition: This project has been deposited at NCBI Sequence Read Archive (SRA) under the accession SRP059657.

Abstract

Small nuclear and nucleolar RNAs (snRNAs and snoRNAs) are known to be functionally and evolutionarily conserved elements of transcript processing machinery. Here, we investigated the expression evolution of snRNAs and snoRNAs by measuring their abundance in the frontal cortex of humans, chimpanzees, rhesus monkeys, and mice. Although snRNA expression is largely conserved, 44% of the 185 measured snoRNA and 40% of the 134 snRNA families showed significant expression divergence among species. The snRNA and snoRNA expression divergence included drastic changes unique to humans: A 10-fold elevated expression of *U1* snRNA and a 1,000-fold drop in expression of *SNORA29*. The decreased expression of *SNORA29* might be due to two mutations that affect secondary structure stability. Using in situ hybridization, we further localized *SNORA29* expression to nucleolar regions of neuronal cells. Our study presents the first observation of snoRNA abundance changes specific to the human lineage and suggests a possible mechanism underlying these changes.

Key words: snRNA, snoRNA, evolution, brain, human.

Introduction

Gene expression differences among species, including differences between humans and closely related primates, constitute one of the main sources of phenotypic and functional divergence (King and Wilson 1975)

Recent advances in sequencing technology have allowed the investigation of expression-level evolution for different RNA types, including messenger RNA (mRNA), long noncoding RNA (lncRNA), and microRNA, which all show substantial evolutionary divergence among species, including humans (Brawand et al. 2011; Meunier et al. 2013; Necsulea et al. 2014).

A group of transcripts, mid-size RNA with length between 100 and 200 nt, have so far remained outside the scope of evolutionary transcriptomic studies. This length range mainly contains two noncoding RNA types: Small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA). Both snRNAs and snoRNAs constitute conserved elements of basic posttranscriptional regulation machinery. In the human genome, there are ten snRNA gene families (*U1*, *U2*, *U4*, *U5*, *U6*, *U7*, *U4atac*, *U6atac*, *U11*, and *U12*), each represented by multiple copies. Functionally, all snRNAs, except *U7*, serve as essential elements of the spliceosome complexes (Matera et al. 2007).

© The Author 2016. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

SnoRNA are a more diverse type of transcripts, represented in the human genome by 1,740 genes, grouped into 226 families based on their Rfam classification (Burge, et al. 2013). SnoRNAs function as part of ribo-protein complexes, guiding them to perform methylation or pseudouridylation at specific sites of ribosomal RNA (rRNA), transfer RNA, and snRNA (Kiss-Laszlo et al. 1996; Ganot et al. 1997; Jady and Kiss 2001; Kiss 2002). The target modification sites for approximately 80% of human snoRNAs can be predicted based on sequence complementarity (Lestrade and Weber 2006); however, the remaining snoRNAs have no known targets. This suggests that the full spectrum of functional roles played by snoRNA remains to be determined. In support of this notion, some snoRNAs have recently been identified to play roles in the regulation of transcript processing and transcript stability (Kishore and Stamm 2006; Ender et al. 2008; Kishore et al. 2010; Brameier et al. 2011; Yin et al. 2012).

Mechanisms of transcriptional regulation differ substantially for snRNAs and snoRNAs. Although snRNAs are transcribed from their own promoters, snoRNA, especially conserved snoRNA, do not commonly have independent promoter elements and are instead transcribed as a part of a larger transcript, commonly residing within an intron of mRNA or lncRNA (Hoepfner et al. 2009).

Taken together, functional conservation, peculiarities of transcriptional regulation, as well as technical questions related to unusual RNA size range make snRNAs and snoRNAs an unconventional target for evolutionary studies. Here, by employing a size-fractionation procedure coupled with high-throughput sequencing, we nonetheless explored snRNAs and snoRNAs expression in the frontal cortex of four mammalian species: Humans, chimpanzees, rhesus monkeys, and mice. To our surprise, we show that expression of these transcripts has undergone a number of drastic changes on the primate and mouse lineages, including changes during the past 6–8 Myr of human evolutionary history.

Materials and Methods

Ethics Statement

This study was reviewed and approved by the Institutional Animal Care and Use ethics committee at the Shanghai Institutes for Biological Sciences.

Tissues and RNA Library Preparation

Human samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland (USA). All subjects were defined as healthy controls by forensic pathologists at the corresponding tissue bank. Chimpanzee samples were obtained from the Anthropological Institute & Museum of the University of Zürich-Irchel (Switzerland), and the Biomedical Primate Research Centre (the Netherlands). Rhesus macaque samples

were obtained from the Suzhou Experimental Animal Center (China). All nonhuman primates used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the tissue used. Mouse samples were obtained from the mouse facility at Shanghai Institutes for Biological Sciences (China). All mouse individuals were from the C57BL/6 strain with no genetic modifications. The PFC samples were dissected from the anterior part of the superior frontal gyrus, a cortical region approximately corresponding to Brodmann area 10. All tissues were snap-frozen after dissection and stored at -80°C without thawing. Total RNA was extracted by TRIzol (Invitrogen, USA) according to the manufacturer's instructions. All RNA quality was assessed using an Agilent 2100 Bioanalyzer. Only samples with RNA Integrity Number (RIN) values greater than or equal to 8 were used in this study. For the pooled sample, 4 μg total RNA from each of five individuals from the same species was merged into 20 μg pool sample.

Sequencing libraries were prepared as previously described (Hu et al. 2009). Briefly, 20 μg total RNA was purified by RNeasy MinElute Cleanup Kit (Qiagen, Germany). Ribosome RNA was then removed using the RiboMinus Eukaryote Kit (Ambion, USA). Purified RNA was loaded to a denaturing 15% TBE-PAGE (Tris-Borate-EDTA Polyacrylamide Gel Electrophoresis) gel (Invitrogen), and the range of the gel containing 100- to 200-nt-long RNA was extracted, ligated to the adapters, amplified, and sequenced following the Small RNA Sample Preparation Protocol (Illumina, USA).

Transcriptome Data Analysis

The pooled samples were sequenced with a read length of 36 bp, while the individual samples were sequenced on Illumina HiSeq 2000 with a read length of 50 bp. The reads were then mapped to a cumulative rRNA sequence constructed based on the rRNA gene sequences from SILVA and UCSC (Quast et al. 2013; Karolchik et al. 2014) databases. The mapping was performed by Bowtie2 with default parameters (Langmead and Salzberg 2012). Reads that mapped to rRNA were removed from further analyses. The remaining reads were mapped to the respective genome: Human, hg19; chimpanzee, panTro4; rhesus monkey, rheMac3; and mouse, mm10. Gene annotation was constructed using the human and mouse GENCODE annotation as a base (hg19: v19; mm10: m2) (Harrow et al. 2012). We combined the two annotations using LiftOver, and searched for homologs of each exon in the chimpanzee and rhesus genomes (Karolchik et al. 2014).

The unique and multiple mapped read counts for each transcript were quantified with BEDtools (Quinlan and Hall 2010). For multiple mapped reads, one genomic locus with the best alignment score was randomly selected. We further examined the reads proportion from different categories of transcripts, including snoRNAs, snRNAs, protein-coding transcripts, retained introns, other transcript types, and

unannotated genomic regions. We used uniquely mapped reads for snoRNA expression quantification, and both uniquely and multiply mapped reads for snRNA expression quantification. Specifically, for each snRNA, including *U1*, *U2*, *U4*, *U5*, *U6*, *U7*, *U11*, *U12*, *U4atac*, and *U6atac*, we summed the read counts of all copies.

The snoRNA expression heatmap was constructed based on the log₂-transformed RPKM (Reads Per Kilo bases per Million reads) values of snoRNA genes using R package “pheatmap.” For each snoRNA gene, we conducted lineage analysis as follows: We used one-way Analysis of Variance (ANOVA) followed by Tukey’s Honestly Significant Difference (HSD) test to calculate the significance of gene expression difference between each pair of species, and estimated the significance and the False Discovery Rate (FDR) by permutation of sample labels 100 times (permutation $P < 0.05$, FDR < 0.05). Based on their expression patterns, changes in snoRNA expression were assigned to ten groups: Human lineage (HL), chimpanzee lineage (CL), rhesus lineage (RL), mouse lineage (ML), human–chimpanzee shared lineage, human–rhesus shared lineage, human–mouse shared lineage, chimpanzee–rhesus shared lineage, chimpanzee–mouse shared lineage, and rhesus–mouse shared lineage. The expression changes assigned to a specific lineage should meet the following criteria: The expression in the corresponding species is significantly different from that in the other three species, and the expression distance to the average of all four species is the largest. The expression changes assigned to a shared lineage should show no expression difference between the two species sharing the lineage, but be significantly different from the expression level in the other two species. We further performed analysis using DESeq2 (Love et al. 2014) to identify snoRNA differentially expressed among species. We then used the abovementioned procedure to assign observed changes to the lineages using the adjusted *P* value ($P < 0.05$) and the normalized expression distance based on the size factor estimated by DESeq2. The expression analysis of snoRNA families was conducted using either uniquely mapped reads only or combing the uniquely and multiply mapped reads.

The assignment of the snoRNA to the corresponding family was carried out based on its Rfam ID, which was acquired from the Ensembl (Ensembl 78) annotation. The cumulative read counts for all snoRNA genes within a same family were used as the measure of snoRNA family expression. The heatmap construction, expression difference analysis, and the lineage assignment procedure were carried out using the same procedure as for individual snoRNAs.

To compare the transcriptional conservation of individual snoRNA genes with its family-level expression, we defined the dissimilarity value ($1 - \text{Rho}$, spearman’s correlation coefficient) across species. The significance of this value’s difference between the gene and family was estimated using a pairwise Wilcoxon test.

SnoRNA Sequence and Secondary Structure Analysis

The alignment of snoRNA sequences was generated using the MUSCLE tool (Edgar 2004), with default parameters, from which we constructed a consensus sequence. The secondary structure of snoRNA was predicted and its minimum free energy (MFE) was calculated using the RNAfold tool from the Vienna RNA package2 (Lorenz et al. 2011). To evaluate the effect of human-specific mutations in the *SNORA29* sequence, we artificially mutated the *SNORA29* human–chimpanzee–rhesus consensus sequence using an identical number and type of nucleotide substitutions.

To clarify the role of mutation in snoRNA essential motif in regulating its expression, we investigated the occurrence of motifs in all the snoRNAs we analyzed. We searched the C box ([AG]JUGAUGA), D box (CUGA), H box (ANANNA), and ACA box (ACA) for each snoRNA in each species. Meanwhile, the position and the order of two motifs had to be appropriate for eligible motif, such as the C box should be upstream of D box while the H box should be upstream of ACA box, as well as these motifs were not located within the stem of hairpin structure of snoRNA.

To determine whether secondary structure could influence the snoRNA abundance, we divided snoRNAs into four groups based on their expression in our data or in published data: Nonabundant or defined as not expressed ($\text{RPKM} \leq 2$); low abundant ($2 < \text{RPKM} \leq 50$); medium abundant ($50 < \text{RPKM} \leq 500$), and highly abundant ($\text{RPKM} > 500$). Subsequently, we compared the distribution of the minimal free energy normalized by length across these four groups, as well as compared these distributions to their corresponding genomic background using the Kolmogorov–Smirnov test, and visualized the results using the R package “beanplot.” The genomic background was derived from randomly selecting the genomic fragment with the same length distribution of corresponding snoRNAs and subsequently calculating the MFE for each fragment. The effect of MFE at the ribonuclear–protein complex level was investigated using public snoRNP abundance data from the PAR-CLIP experiments conducted for FBL, NOP56, and NOP58 components of C/D box snoRNPs, as well as DKC1 (Kishore, et al. 2013), the core protein of H/ACA snoRNAs.

The human and mouse intronic snoRNAs were determined by the intersection of genomic coordinates between introns and snoRNA genes using BEDtools. The Pearson correlation was calculated based on zero-mean-normalized, log-transformed snoRNA expression, host gene expression, and MFE values.

In Situ Hybridization

The in situ hybridization was performed as described previously (Somel et al. 2011). In brief, the full length of *SNORA29* was cloned into the pGEM-T easy vector (Promega, USA). Digoxigenin (DIG)-labeled sense and antisense riboprobes were synthesized using SP6/T7 Transcription Kit (Roche,

Switzerland) following manufacturer' instructions. The 14- μ m thick brain tissue cryosections were mounted on positively charged slides, fixed with formaldehyde and acetylated. After prehybridization for 4 h, sections were hybridized overnight and washed in Saline Sodium Citrate (SSC) buffer six times. Alkaline phosphatase-conjugated anti-DIG antibody (Roche) was used to detect the hybridization signals, followed by development in buffer with NBT/BCIP for 16 h. Hybridization with the sense probe was performed in parallel and no signal was detected in all cases (data not shown). Immunohistochemical staining was performed after in situ hybridization. Developed sections were blocked with 10% normal goat serum and subsequently incubated with monoclonal anti-NeuN antibody (1:100, MAB377, EMD Millipore, USA).

Results

We characterized the expression of transcripts with length between 100 and 200 nt in the frontal cortex of humans, chimpanzees, rhesus monkeys, and mice by sequencing the fraction of total RNA of the corresponding length. For each species RNA sequencing (RNA-seq) was conducted in three individual samples, as well as one pooled sample of five individuals (supplementary table S1, Supplementary Material online). This resulted in a total of 45,234,672 available sequence reads, 69.5% of which could be uniquely mapped to the corresponding genome (supplementary table S2, Supplementary Material online). Half of these reads mapped to annotated snRNA (1.3%) and snoRNA (52.9%). The remaining reads mapped to protein-coding transcripts (10.9%), retained introns (2.5%), other annotated transcript types (13.5%), and unannotated genomic regions (18.9%) (fig. 1A). Allowing multiple mapping further increased the proportion of reads that mapped to snRNA genes, which are represented by multiple gene copies, to 11.4% (fig. 1A).

Among the ten annotated snRNA families conserved across the four species, nine were expressed in our data (average RPKM ≥ 1) when allowing for multiple mapping (Hart et al. 2013). Expression of most snRNA was conserved among species (supplementary fig. S1, Supplementary Material online). The only exception was the expression of U1 snRNA: U1 expression was approximately 10-fold higher in the human brain compared with the brains of chimpanzees, macaques, and mice (Tukey-HSD test, $P < 0.05$, FDR < 0.05 after multiple testing correction) (fig. 1B).

Of the 208 annotated snoRNA genes conserved across the 4 species, 183 were found to be expressed in our data (average RPKM ≥ 1) using uniquely mapped reads. In contrast to snRNAs, 83 (44%) were expressed differently among species (Tukey-HSD test, $P < 0.05$, FDR < 0.05 after multiple test correction). Assignment of these expression differences to evolutionary lineages resulted in 5 human-specific expression changes, 4 chimpanzee-specific ones, 12 changes assigned to

the ape lineage, 13 to the macaque lineage, and 46 to the lineage connecting primates and mice (fig. 1C and supplementary fig. S2, Supplementary Material online). Similar results were obtained using DESeq2 (Love et al. 2014) (supplementary fig. S3, Supplementary Material online), as well as considering male samples only (supplementary fig. S4, Supplementary Material online).

Within a family, snoRNAs often share the same functionality (Hoepfner and Poole 2012). Thus, it has been suggested that changes in the expression of individual snoRNAs could be compensated for by expression changes of other snoRNAs from the same family (Kehr et al. 2014). We indeed found substantially higher conservation of snoRNA expression at the family level compared with individual snoRNA (fig. 1D).

SNORD116 (*HBI-85*) snoRNA cluster can be used as an example of such conservation. The cluster is formed by tandem repeats, resulting in a snoRNA family cluster containing 29 members. Of these, 25 have detectable expression in our data. At the individual gene level these snoRNAs show substantial expression variation among species: Nine show significant lineage-specific expression changes among primates, including two on the human evolutionary lineage (fig. 1E). In contrast, at the family level, *SNORD116* expression is completely conserved (fig. 1F).

Species-specific changes in expression of snoRNA were nonetheless still observed after grouping snoRNA genes into families. Specifically, of the 134 snoRNA families expressed in brain, we identified 2 that showed expression change on the HL, 1 on the CL, 6 on the ape lineage, 6 on the macaque lineage, and 39 on the lineage connecting mouse and primates (fig. 1G and supplementary fig. S5, Supplementary Material online). Consistent results were obtained based on using both uniquely and multiply mapped reads (supplementary fig. S6, Supplementary Material online).

Each of the two snoRNA families shows human-specific change containing only one member: *SNORA29* and *SNORA51*, respectively (fig. 2A). Of the two snoRNAs, *SNORA29* expression changes much more drastically: While this snoRNA is highly expressed in mouse, macaque, and chimpanzee brains, its expression drops as much as 1,000-fold in the human brain compared with the average of the other three species (fig. 2B and supplementary fig. S7, Supplementary Material online).

Similar to many other snoRNAs, *SNORA29* has intronic localization: Its sequence resides within the fourth intron of the *TCP1* gene. We assessed the expression of the *TCP1* gene in the frontal cortex of adult humans, chimpanzees, macaques, and mice using a publicly available polyA-plus RNA-seq data set (Bozek et al. 2014). There are no substantial differences in *TCP1* expression among the four species (supplementary fig. S7, Supplementary Material online). Furthermore, expression of another snoRNA located within an intron of the *TCP1* gene, *SNORA20*, is also highly conserved (supplementary fig. S7, Supplementary Material online). Thus, the drastic drop in

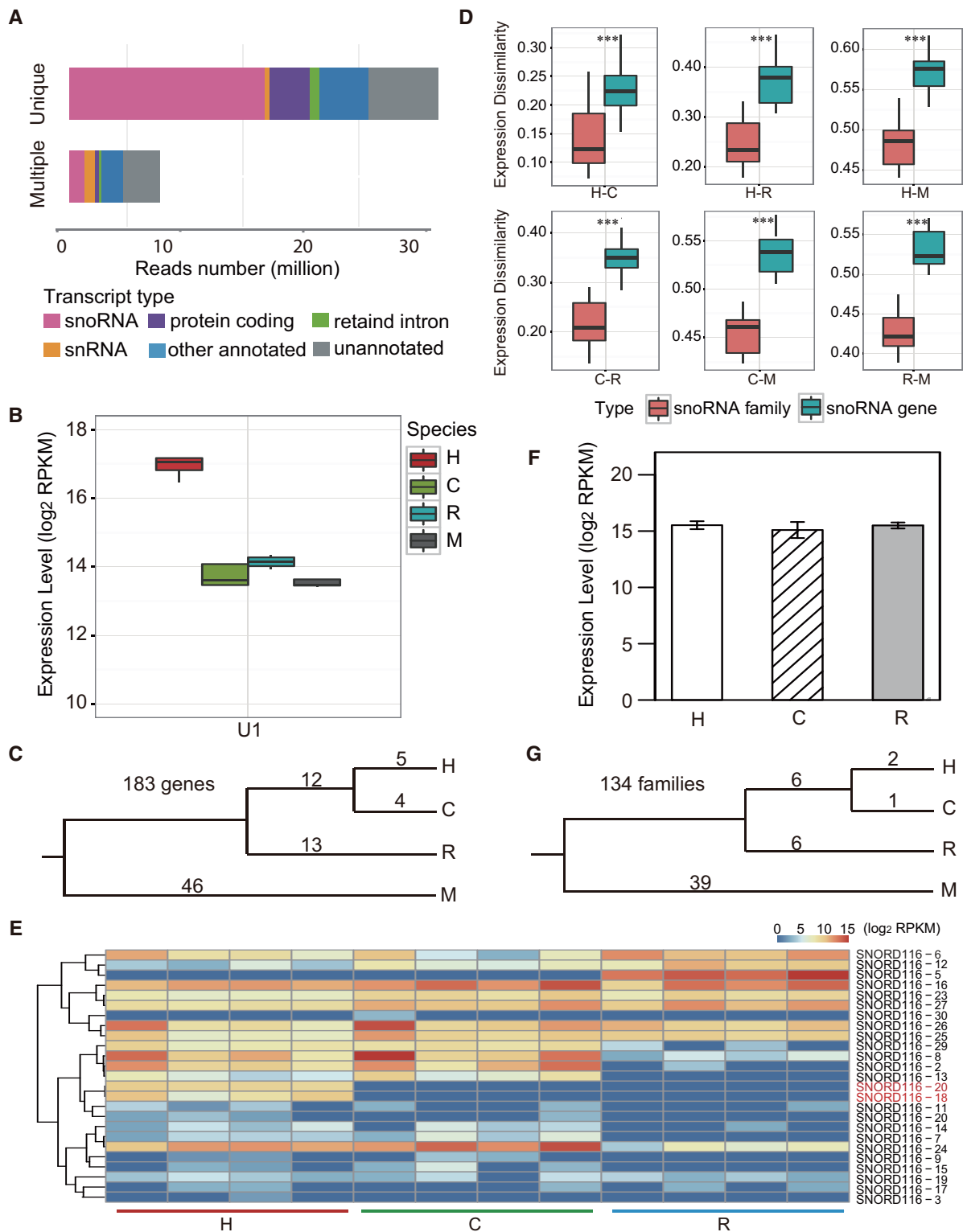


Fig. 1.—Expression of snRNA and snoRNA genes and families across four species. (A) Proportions of sequence reads in different transcript categories. (B) Expression of *U1* snRNA in four species. The species labels on this and the following panels: H, human; C, chimpanzee; R, rhesus monkey; M, mice. (C) Distribution of snoRNA gene expression changes on the evolutionary lineages (187 genes—total number of expressed snoRNA genes). (D) The expression dissimilarity value distribution of snoRNA genes and families for each pairwise species comparison (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$). (E) Expression of individual *SNORD116* genes. The colors indicate log₂-transformed RPKM values. The labels of *SNORD116* genes showing human-specific expression are shown in red. (F) The cumulative expression level of *SNORD116* cluster. (G) Distribution of snoRNA family expression changes on the evolutionary lineages (134 families—total number of expressed snoRNA families).

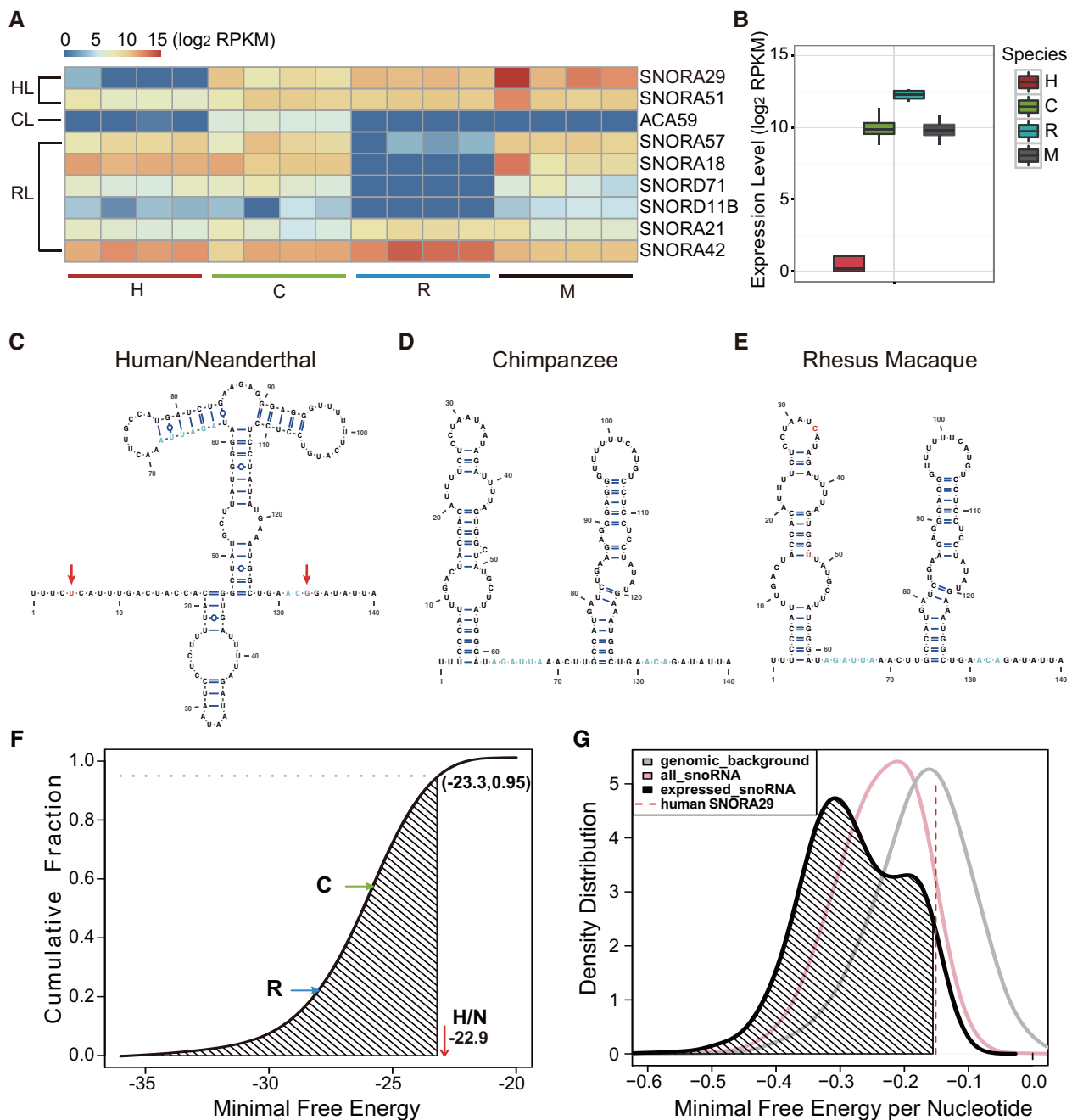


FIG. 2.—The human-specific expression pattern of *SNORA29* and the underlying genetic mechanism. (A) Expression of snoRNA families showing expression changes specific to human, chimpanzee, and rhesus evolutionary lineages. (B) Expression of *SNORA29* in four species. (C–E) Predicted *SNORA29* secondary structure in humans/Neanderthals, chimpanzees, and rhesus monkeys. Red arrows indicate human-specific mutations. (F) MFE of *SNORA29* in humans/Neanderthals (H/N), chimpanzees (C), and rhesus monkeys (R). The cumulative curve is based on MFE values of *SNORA29* mutants artificially constructed based on the *SNORA29* human–chimpanzee–rhesus consensus sequence. (G) The MFE distribution for all snoRNA annotated in the four species; all snoRNA expressed in at least one of the four species (RPKM > 2), and the human genomic background. The vertical dashed red line shows MFE of human *SNORA29*.

the abundance level of *SNORA29* on the human lineage cannot be explained by a change in expression of its host gene.

To assess *SNORA29* expression in human tissues other than brain, we analyzed *SNORA29* expression based on available

public data: Direct RNA sequencing and PAR-Clip of snoRNP core proteins and AGO proteins in human HEK293 cell line. We only detected low expression of *SNORA29* (RPKM = 1.21) in one of the three data sets (AGO PAR-Clip), while expression

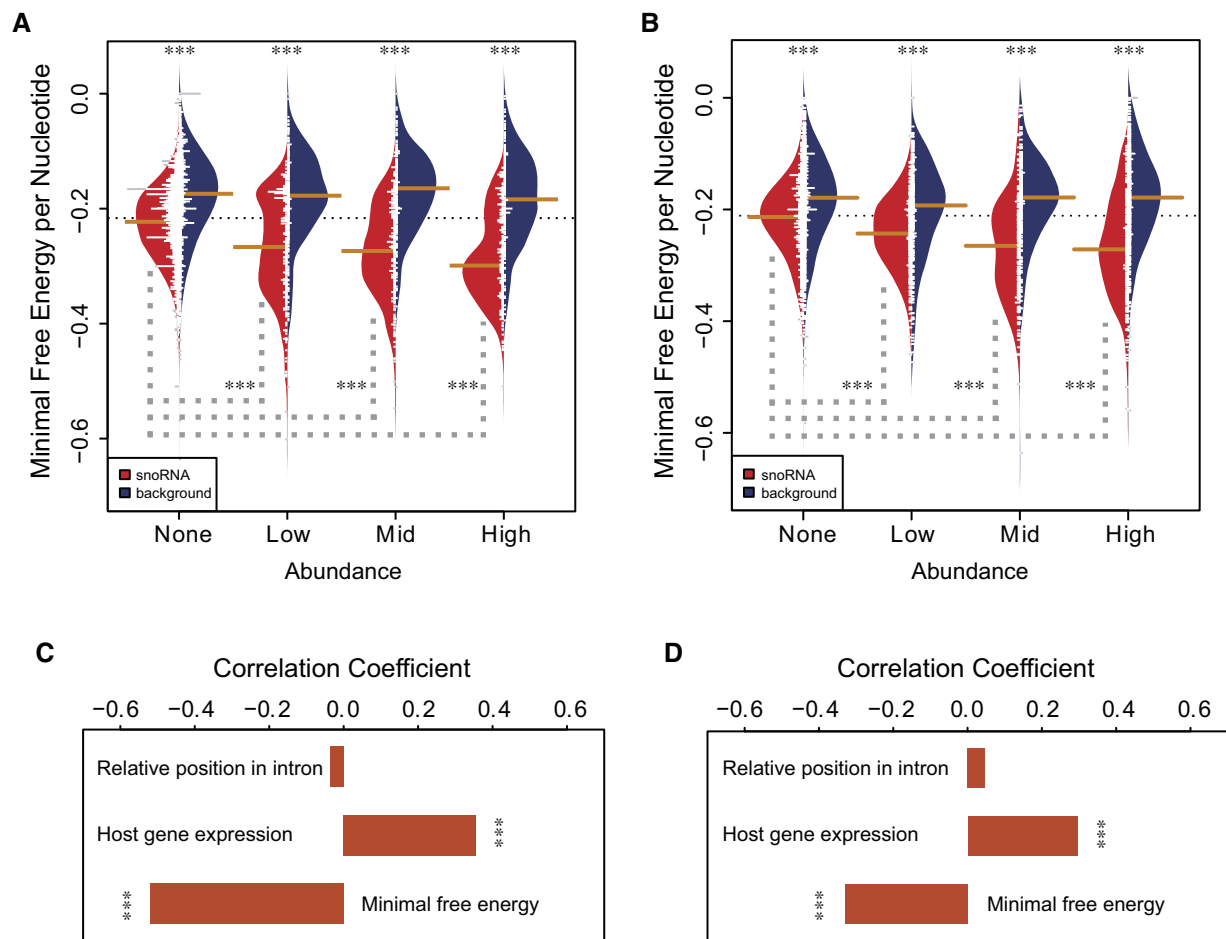


FIG. 3.—Relationship between snoRNA abundance and its structure stability. (A and B) Distributions of MFE of snoRNAs and their genetic background in four snoRNA expression level groups. The expression of snoRNAs was quantified based on our data (A) and published PAR-CLIP data (B). (C and D) The Pearson correlation between intronic snoRNA expression and its relative position within intron, host gene expression, and MFE in human (C) and mouse (D). * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

of other snoRNAs was substantially higher. At the same time, *TCF1* expression in HEK293 cell line is relatively high (RPKM = 100). This further confirms that decreased expression of *SNORA29* on the HL cannot be explained by a change in expression of its host gene.

By comparing the human *SNORA29* sequence with the consensus sequence of its primate homologs, we identified two nucleotide changes: An A to G substitution at position 133, and a C to T substitution at position 5 (fig. 2C–E and supplementary fig. S8, Supplementary Material online). These changes affect a conserved snoRNA element, the ACA box, and the first stem, respectively, and substantially affect *SNORA29* secondary structure (fig. 2C). Notably, the two human-specific mutations result in an unusually large increase in free energy of the *SNORA29* secondary structure, indicating a sharp reduction in its stability (fig. 2F and G).

We hypothesized that the reduction in *SNORA29* secondary structure stability, rather than human-specific regulation of

SNORA29 expression, may result in its decreased abundance in the human brain. To test this, we investigated the general relationship between snoRNA structural stability and expression abundance. Indeed, the abundance of all 4,765 snoRNAs detected in our RNA-seq data in at least one species is negatively correlated with the minimal free energy of their secondary structure ($r = -0.29$, $P < 0.0001$). This relationship between the minimal free energy and snoRNA expression abundance was also clearly observed in a comparison of snoRNA groups separated based on their expression level in brain tissue (fig. 3A). Furthermore, the same relationship can be observed using another snoRNA expression data set: A combined PAR-CLIP data set containing abundance levels of 1532 C/D box and H/ACA box snoRNPs (Kishore et al. 2013) (fig. 3B). Notably, for intronic snoRNAs, the amplitude of the correlation between the expression abundance and the minimal free energy was comparable or greater than the correlation between the expression abundance and the host gene

expression (fig. 3C and D). Taken together, these observations suggested that the decreased abundance of *SNORA29* in the human brain might be caused by the decreased stability of its secondary structure.

To get a more precise estimate of the evolutionary time during which the two mutations affecting *SNORA29* secondary structure stability occurred, we investigated the genomes of Neanderthals and Denisovans (Lazaridis et al. 2014). In both genomes, the nucleotides in these two loci were identical to that of modern humans (supplementary fig. S8, Supplementary Material online). Thus, the mutations must have occurred after the human/chimpanzee split approximately 8 Ma, but before the human/Neanderthal and Denisovan separation a bit more than 500,000 years ago (Prufer et al. 2014).

SNORA29 is not unique in carrying lineage-specific mutations. Based on snoRNA sequence alignment among 4 species, 3 out of 5 (60%) snoRNA showing expression change on the HL, 2 out of 4 (50%) on the CL, 8 out of 13 (61.5%) on the rhesus monkey lineage, and 49 out of 49 (100%) on the ML have species-specific base substitutions. Taken together, 35 of these 58 substitutions (60%) change snoRNA's MFE in direction corresponding with the direction of the expression change. Among all snoRNAs showing species-specific expression change, mutations occurring within the essential snoRNA motifs are only present in *SNORA29*.

To gain insights into *SNORA29* functionality in the brain, we assessed its localization in prefrontal cortex tissue of humans and rhesus monkeys using in situ hybridization. As shown in figure 4A, we observed strong hybridization signals concentrated at a small region within nuclei in macaque, but not in human brain tissue. Such localization is a common feature of snoRNAs and marks their location in the nucleolar region. Notably, the *SNORA29* signal was found in only a fraction of all nuclei. Using NeuN antibodies, which mark neuronal nuclei, we showed that *SNORA29* localization is restricted to neurons (fig. 4B and C).

Discussion

In this study, we estimated the evolutionary conservation of snRNA and snoRNA expression in a specific tissue, frontal neocortex, of four mammalian species including humans. Our customized sequencing strategy succeeded in capturing the expression of these two mid-size RNA classes, allowing us to determine expression level divergence of individual transcript and transcript families among the four species.

Surprisingly, despite the presumed conservation of snRNA and snoRNA functionality, we observe a substantial divergence of their expression levels among the four mammalian species used in our study. Furthermore, we find striking examples of snRNA and snoRNA expression level changes that have occurred within the past 6–8 Myr of human evolutionary history: *U1* snRNA showing approximately 10-fold higher

expression, and *SNORA29* showing approximately 1,000-fold lower expression in human brain cortex compared with other species.

The case of elevated expression of *U1* in human brain stands out, as this is the only significant snRNA expression change detected in our study on any of the four evolutionary lineages. It is appealing to speculate that elevated human-specific expression of *U1* might be linked with additional roles played by this RNA in transcriptional regulation and RNA processing (Kaida et al. 2010; Berg et al. 2012).

For snoRNAs expression, only moderate skew toward evolutionary acceleration on the human evolutionary lineage was observed: Five versus four snoRNAs show changes on the HL and CL at the gene level, two versus one at the family level. Still, the example of *SNORA29* snoRNA expression divergence on the HL is particularly striking. This snoRNA shows the same stable and high expression on the mouse, macaque, and chimpanzee evolutionary lineages that together covers approximately 110 Myr of conserved expression (Nei et al. 2001). Yet, this expression drops approximately 1,000-fold on the human evolutionary lineage, sometime between 8 and 0.5 Ma. Intriguingly, in contrast to other H/ACA snoRNAs that are known to guide pseudouridylation of rRNA and snRNA (Jady and Kiss 2001), *SNORA29* is an orphan snoRNA with no known targets (Lestrade and Weber 2006). Our in situ experiments show no detectable *SNORA29* hybridization signal in human brain cortex, confirming the RNA-seq analysis results. The same experiments demonstrate that *SNORA29* expression is restricted to the nucleolus of neural cells in the macaque cortex, suggesting potential importance of this snoRNA to modulation of neuronal functionality. Intriguingly, in line with this notion, the *SNORA29* gene is localized within a chromosome region linked to schizophrenia (Edgar et al. 2000; Pulver 2000). It has to be noted that these findings, although suggestive, certainly do not provide information about *SNORA29* role in evolution of human-specific cognitive features. Furthermore, *SNORA29* expression change is likely not to be restricted to brain. Still, a drastic decrease in *SNORA29* expression abundance on the HL after more than 110 Myr of conservatively high expression suggests contribution to the emergence of functional features unique to humans.

The *SNORA29* gene, similar to many other snoRNA genes, is located in the intron of a long mRNA transcript and has no independent promoter. Yet, we show that change in expression of *SNORA29* in human brain is not caused by host gene inactivation. Instead, our result suggests that *SNORA29* expression change is caused by two mutations, shared by humans and Neanderthals, which drastically destabilize the *SNORA29* secondary structure. Our further analysis of intronic snoRNA expression, conducted using our data set, as well as published data, shows that the minimal free energy of snoRNA's secondary structure is often a better predictor of snoRNA expression than the expression level of its host

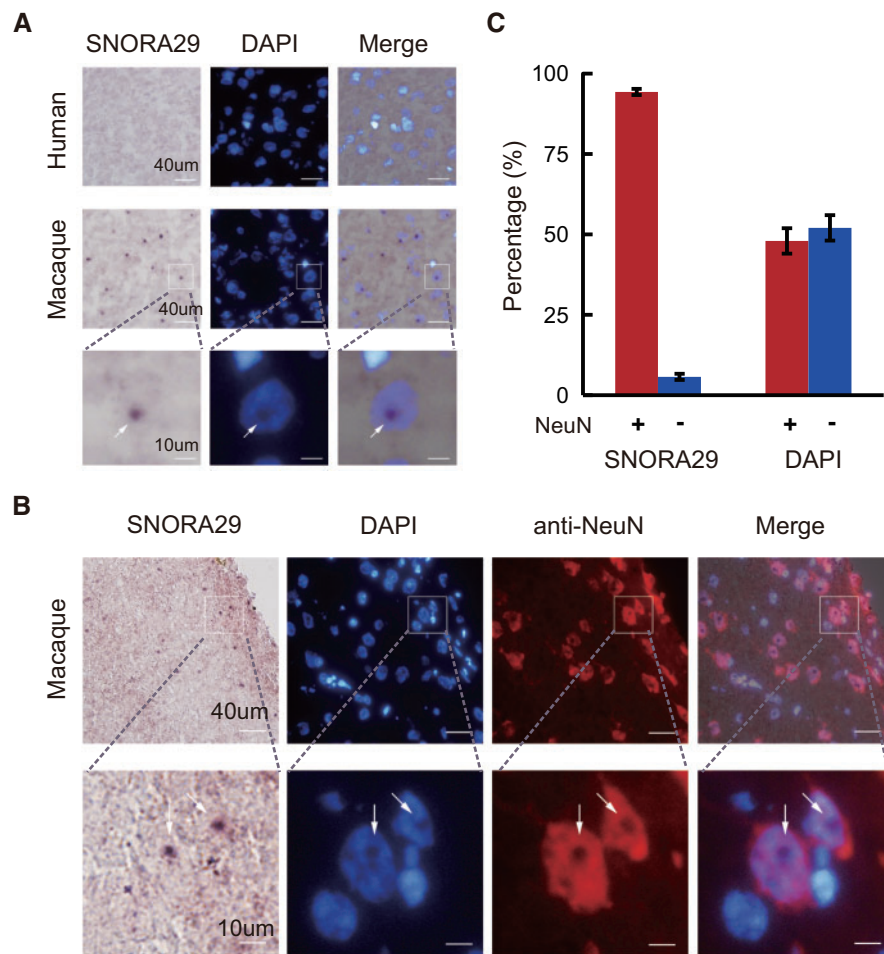


Fig. 4.—In situ hybridization of *SNORA29*. (A) In situ hybridization with *SNORA29* antisense probe in human (upper panel) and macaque (middle and bottom panels) prefrontal cortex. Nuclei were stained with DAPI (4',6-diamidino-2-phenylindole). A merged image showed that *SNORA29* localizes in nucleolus and is not detectable in human tissue. (B) In situ hybridization with *SNORA29* antisense probe in macaque prefrontal cortex. Nuclei were stained with DAPI. Neuronal nuclei were stained with anti-NeuN antibody. The merged image shows that *SNORA29* is expressed in neurons. (C) Counts of neural (NeuN+) and nonneural (NeuN-) cells with *SNORA29* probe and DAPI staining.

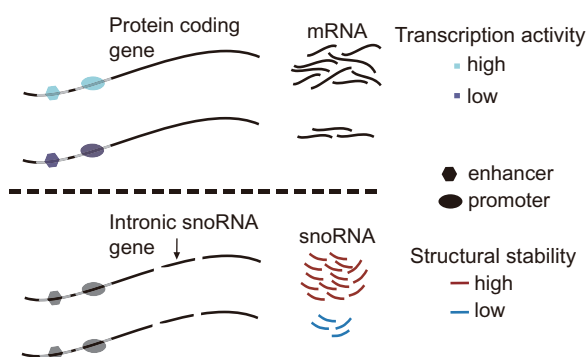


Fig. 5.—Schematic representation of suggested abundance regulation mechanisms for mRNA and snoRNA.

gene. This suggests that in the absence of their own promoters, changes in snoRNA secondary structure stability represent a major mechanism of independent snoRNA expression regulation (fig. 5).

Comparison of snoRNA expression among species shows substantially higher conservation at the family level than at the individual snoRNA level. This observation is compatible with the notion that changes in the expression of individual snoRNA are compensated by expression changes of other snoRNA from the same family (Kehr et al. 2014). Still, we observe substantial snoRNA expression divergence among examined species both for snoRNA families and individual snoRNAs. Of the 187 snoRNA genes and 134 snoRNA families present in the four species and expressed in brain, 44% and 40%, respectively, show divergent expression among species.

Such a high extent of expression-level divergence contradicts sharply the presumed conservation of snoRNA functionality.

As for other RNA types, such as mRNA, microRNA, or lncRNA (Brawand et al. 2011; Meunier et al. 2013; Necsolea et al. 2014), the extent of snoRNA expression divergence among species is largely proportional to the phylogenetic divergence. This might indicate that snoRNA expression divergence merely reflects the accumulation of genetic divergence among species and has no functional effect. This notion, however, is hardly compatible with the essential role played by snoRNA in the modification of highly conserved functional RNA molecules. Thus, the unexpectedly rapid pace of snoRNA expression evolution suggests that snoRNAs play a more dynamic role in the RNA-based regulatory and functional networks than is currently appreciated.

Supplementary Material

Supplementary tables S1–S3 and figures S1–S8 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We thank Mingshuang Li for help with RNA library preparation and Gangcai Xie for transferring the RNA-seq data; Xintian You for the construction of rRNA gene sequence collection; Jerome Boyd-Kirkup for help with manuscript preparation; Haiyang Hu and Zhisong He for helpful suggestions and discussions. This work was supported by CAS Strategic Priority Research Program (grant no. XDB13010200); National Natural Science Foundation of China (grant nos. 91331203, 31420103920, 31501047); Bureau of International Cooperation, Chinese Academy of Sciences (grant no. GJHZ201313); Foreign Expert 1000 Talents Plan Program (grant no. WQ20123100078 to P.K.); Russian Science Foundation (grant 14-28-00234 to N.R.C. “Kurchatov Institute”); and China Postdoctoral Science Foundation (grant no. 2015M571610 to H.D.D.).

Literature Cited

- Berg MG, et al. 2012. *U1* snRNP determines mRNA length and regulates isoform expression. *Cell* 150:53–64.
- Bozek K, et al. 2014. Exceptional evolutionary divergence of human muscle and brain metabolomes parallels human cognitive and physical uniqueness. *PLoS Biol.* 12:e1001871.
- Brameier M, Herwig A, Reinhardt R, Walter L, Gruber J. 2011. Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. *Nucleic Acids Res.* 39:675–686.
- Brawand D, et al. 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478:343–348.
- Burge SW, et al. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res.* 41:D226–D232.
- Edgar PF, et al. 2000. Comparative proteome analysis of the hippocampus implicates chromosome 6q in schizophrenia. *Mol Psychiatry.* 5:85–90.
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32:1792–1797.
- Ender C, et al. 2008. A human snoRNA with microRNA-like functions. *Mol Cell.* 32:519–528.
- Ganot P, Bortolin ML, Kiss T. 1997. Site-specific pseudouridine formation in preribosomal RNA is guided by small nucleolar RNAs. *Cell* 89:799–809.
- Harrow J, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22:1760–1774.
- Hart T, Komori HK, LaMere S, Podshivalova K, Salomon DR. 2013. Finding the active genes in deep RNA-seq gene expression studies. *BMC Genomics* 14:778.
- Hoepfner MP, Poole AM. 2012. Comparative genomics of eukaryotic small nucleolar RNAs reveals deep evolutionary ancestry amidst ongoing intragenomic mobility. *BMC Evol Biol.* 12:183.
- Hoepfner MP, White S, Jeffares DC, Poole AM. 2009. Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes. *Genome Biol Evol.* 1:420–428.
- Hu HY, et al. 2009. Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* 10:413.
- Jady BE, Kiss T. 2001. A small nucleolar guide RNA functions both in 2'-O-ribose methylation and pseudouridylation of the *U5* spliceosomal RNA. *EMBO J.* 20:541–551.
- Kaida D, et al. 2010. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* 468:664–668.
- Karolchik D, et al. 2014. The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* 42:D764–D770.
- Kehr S, Bartschat S, Tafer H, Stadler PF, Hertel J. 2014. Matching of Soulmates: coevolution of snoRNAs and their targets. *Mol Biol Evol.* 31:455–467.
- King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* 188:107–116.
- Kishore S, Stamm S. 2006. The snoRNA *HBII-52* regulates alternative splicing of the serotonin receptor 2C. *Science* 311:230–232.
- Kishore S, et al. 2010. The snoRNA *MBII-52* (*SNORD 115*) is processed into smaller RNAs and regulates alternative splicing. *Hum Mol Genet.* 19:1153–1164.
- Kishore S, et al. 2013. Insights into snoRNA biogenesis and processing from PAR-CLIP of snoRNA core proteins and small RNA sequencing. *Genome Biol.* 14:R45.
- Kiss T. 2002. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. *Cell* 109:145–148.
- Kiss-Laszlo Z, Henry Y, Bachellerie JP, Caizergues-Ferrer M, Kiss T. 1996. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* 85:1077–1088.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359.
- Lazaridis I, et al. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 513:409–413.
- Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* 34:D158–D162.
- Lorenz R, et al. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol.* 6:26.
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:550.
- Matera AG, Terns RM, Terns MP. 2007. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol.* 8:209–220.
- Meunier J, et al. 2013. Birth and expression evolution of mammalian microRNA genes. *Genome Res.* 23:34–45.
- Necsolea A, et al. 2014. The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505:635–640.
- Nei M, Xu P, Glazko G. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and

- several distantly related organisms. *Proc Natl Acad Sci U S A*. 98:2497–2502.
- Prufer K, et al. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505:43–49.
- Pulver AE. 2000. Search for schizophrenia susceptibility genes. *Biol Psychiatry*. 47:221–230.
- Quast C, et al. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res*. 41:D590–D596.
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- Somel M, et al. 2011. MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biol*. 9:e1001214.
- Yin QF, et al. 2012. Long noncoding RNAs with snoRNA ends. *Mol Cell*. 48:219–230.

Associate editor: Mary O'Connell