Phylogenetic and Molecular Analyses Identify SNORD116 Targets Involved in the Prader–Willi Syndrome

Laeya Baldini, Anne Robert, Bruno Charpentier, and Stéphane Labialle [ID]*

Université de Lorraine, CNRS, Nancy, France

*Corresponding author: E-mail: stephane.labialle@univ-lorraine.fr.
Associate editor: Ilya Ruvinsky

Abstract

The eutherian-specific SNORD116 family of repeated box C/D snoRNA genes is suspected to play a major role in the Prader–Willi syndrome (PWS), yet its molecular function remains poorly understood. Here, we combined phylogenetic and molecular analyses to identify candidate RNA targets. Based on the analysis of several eutherian orthologs, we found evidence of extensive birth-and-death and conversion events during SNORD116 gene history. However, the consequences for phylogenetic conservation were heterogeneous along the gene sequence. The standard snoRNA elements necessary for RNA stability and association with dedicated core proteins were the most conserved, in agreement with the hypothesis that SNORD116 generate genuine snoRNAs. In addition, one of the two antisense elements typically involved in RNA target recognition was largely dominated by a unique sequence present in at least one subset of gene paralogs in most species, likely the result of a selective effect. In agreement with a functional role, this ASE exhibited a hybridization capacity with putative mRNA targets that was strongly conserved in eutherians. Moreover, transient downregulation experiments in human cells showed that Snord116 controls the expression and splicing levels of these mRNAs. The functions of two of them, diacylglycerol kinase kappa and Neuroligin 3, extend the description of the molecular bases of PWS and reveal unexpected molecular links with the Fragile X syndrome and autism spectrum disorders.

Key words: box C/D snoRNA, snord116, Prader–Willi syndrome.

Introduction

Prader–Willi syndrome (PWS) is a neurobehavioral disorder characterized by hypotonia, suck and feeding difficulties, and failure to thrive in infancy followed by developmental delay, short stature, hyperphagia that may cause morbid obesity and behavioral and cognitive troubles (Bennett et al. 2015; Butler et al. 2019; Muscogiuri et al. 2019). PWS is caused by the absence of paternally expressed genes in the 15q11-13q region. The ∼ 2.5 Mb-long PWS locus is controlled by parental genomic imprinting, an epigenetic phenomenon where genes are mostly or exclusively expressed from one parental allele. The majority of PWS patients harbor large genetic deletions on the paternal chromosome, whereas 20–30% have maternal uniparental disomy and 1–2% have imprinting disorders leading to the absence of expression of the paternal genes (Ohta et al. 1999). Apart from several protein-coding genes, the PWS locus exhibits two large tandem repeats of C/D box snoRNA genes called SNORD116 and SNORD115, each copy being hosted in one intron of the long noncoding SNHG14 gene (fig. 1A). Although the genetic organization and epigenetic control of these genes was identified early (Cavaillé et al. 2000; de los Santos et al. 2000; Meguro et al. 2001), their repetitive nature has made it difficult to identify precisely their number, but accuracy increased with improvements in genome sequencing. Interestingly, it is now hypothesized that the number of SNORD115 and SNORD116 gene copies varies among individuals, as recently reported in mouse (Keshavarz et al. 2021). C/D box snoRNAs represent an ancient family of small noncoding RNAs that typically function as guides for the 2′-O-methylation of ribosomal RNAs and small nuclear RNAs in Archaea and Eukaryotes (fig. 1B). However, the eutherian-specific Snord116 and Snord115 belong to the class of orphan snoRNAs that lack apparent base complementarity with usual RNA targets. Yet SNORD115 genes possess conserved sequence complementarity with 5htr2c mRNA (Cavaillé et al. 2000) that codes for a seven-transmembrane G-protein-coupled receptor involved in serotonin signaling. Interestingly, molecular studies evidenced that Snord115 snoRNAs promote alternative splicing and editing of this mRNA (Vitali et al. 2005; Kishore and Stamm 2006), which could contribute to the PWS phenotype (Doe et al. 2009; Morabito et al. 2010). Conversely, no clear molecular function has emerged for SNORD116 genes, whereas their function might be of relevance for PWS: they are located in the minimal region that is absent in PWS patients harboring microdeletions (Sahoo et al. 2008; de Smith et al. 2009; Duker et al. 2010; Tan et al. 2020) and the knockout of the Snord116 cluster in mouse models largely recapitulates the PWS phenotype (Skryabin et al. 2007; Ding et al. 2008; Polex-Wolf et al. 2018; Adhikari et al. 2019). Accordingly, efforts have been made to elucidate their

Open Access

molecular function. Transcriptome analyses of *Snord116* knockout mice have revealed hundreds of differentially expressed genes (Bochukova et al. 2018; Coulson et al. 2018; Pace et al. 2020). In addition, transient overexpression of Snord115 and/or Snord116 via artificial constructs influenced the expression level of numerous genes in cell lines (Falaleeva et al. 2015). Although these data are consistent with a complex pathological condition, to date they provided little information about the molecular targets of Snord116 snoRNAs. On the other hand, computational predictions of snoRNA targets such as realized by SNOTARGET (Bazeley et al. 2008) or PLEXY (Kehr et al. 2011) have proposed that Snord116 can hybridize with multiple cellular RNAs. Of note, SNOTARGET predictions included the Ankrd11 mRNA that encodes a chromatin regulator essential for neural development (Gallagher et al. 2015) and a recent study suggested that its expression correlates with the number of *Snord116* gene copies in mouse (Keshavarz et al. 2021). Another study computing snoRNA targets has predicted that Snord116 methylate human 18S at position U1162 (Kehr et al. 2014), but an experimental validation is pending. In the last years, several molecular studies have interrogated RNA–RNA interactions using high-throughput methods. A comprehensive collection of these data has been recently released, called the RISE database (Gong et al. 2018), which proposes candidate interactions of human Snord116 snoRNAs with several C/D and H/ACA snoRNAs, mRNAs, and lncRNAs. However, again, no experimental validation has been provided to date. In addition, defects in prohormone processing were suspected in PWS after observation of decreased expression of the prohormone convertase gene PCSK1 and its associated regulator NHLH2 in patient-derived induced pluripotent stem cells (iPSCs) and in *Snord116* knockout mice (Burnett et al. 2017). Yet the existence of altered Pcsk1 expression in the hypothalamus of *Snord116* knockout mice was not confirmed in another work (Polex-Wolf et al. 2018). The possibility of a direct interaction with the Nhlh2 mRNA was nevertheless proposed recently; if the theoretical interaction energy is modest in Human and questionable in mouse, the hypothesis was partially supported by overexpression experiments in mouse cells (Kocher et al. 2021). Alternatively, it can be hypothesized that a function of the *SNORD116* cluster other than snoRNA production is involved in the pathology. First, it is suspected that C/D snoRNA gene clusters help elicit parental genomic imprinting at the local level in association with their repetitive structure (Labialle and Cavaillé 2011). The *SNORD116* cluster hosts several binding sites for the ZNF274 protein that are important for local epigenetic regulation during development (Cruvinel et al. 2014). Very interestingly, a knockout of the *ZNF274* gene or of the ZNF274 protein binding sites at the SNORD116 locus partially rescued expression of the silent maternal *SNORD116* alleles in neurons derived from PWS iPSCs (Langouët et al. 2018, 2020). Furthermore, the *SNORD116* cluster generates a set of long noncoding RNAs (lncRNAs) including sno-lncRNAs that are thought to sequester nuclear proteins in human pluripotent cells (Yin et al. 2012; Wu et al. 2016), but whether these RNA species are involved in PWS has not yet been tested.

Several rare genetic conditions share features with PWS including maternal uniparental disomy of chromosome 14, Xq27-qter disomy, 1p36 monosomy, deletion of 6q, of 2pter, of 3p26.3, of 10q26, duplication of Xq21, of Xq23-q25, and fragile X syndrome (FXS), among others (Cheon 2016). The main clinical manifestations include hypotonia, obesity, autism spectrum disorders (ASD), and intellectual disability. As the genetic basis of these disorders differs, one or several dysregulated genes are expected to be involved in pathways that control the development of the PWS phenotype. FXS is caused by the expansion of a trinucleotide repeat in the *FMR1* gene that codes for the FMRP protein, is characterized by intellectual disability, ASD, and has distinctive physical features (Hagerman et al. 2017). FMRP potentially regulates the translation of hundreds of mRNAs, many of which are involved in neuronal synaptic connections. The diacylglycerol kinase kappa (Dgkk) mRNA was relatively recently identified as a major effector of FMRP function (Tabet et al. 2016). The *DGKK* gene controls the balance between diacylglycerol and phosphatidic acid signaling pathways and its deficit leads to synaptic and dendritic alterations reminiscent of FXS symptoms in mouse (Tabet et al. 2016). Until now, the proximity between FXS and PWS conditions has been linked to the genetic location of the *CYFIP1* gene at the proximal border of the PWS locus. This biallelically expressed gene codes for a cytoplasmic protein that interacts with FMRP and mediates its translational effects (Napoli et al. 2008; De Rubeis et al. 2013) and *Cyfip1* haploinsufficiency has been reported to provoke abnormal neurogenesis and Fragile X-like phenotypes in mouse models (Bozdagi et al. 2012; Haan et al. 2021). In agreement, the paternal copy of *CYFIP1* could be lost, or not, in patients harboring 15q11-q13 deletions (Chai et al. 2003). Conversely, around 10% of FXS patients harbor a Prader–Willi phenotype (PWP-FXS) including obesity and hyperphagia, delayed puberty, infant hypotonia, and ASD with no evidence of a 15q11-q13 defect (Nowicki et al. 2007; Juriaans et al. 2021). Interestingly, one study reported a decrease in *CYFIP1* gene expression in some PWP-FXS patients (Nowicki et al. 2007). However, the mechanisms underlying PWP-FXS and Prader–Willi-like disorders remain unclear.

The molecular functions of the SNORD116 genes and their involvement in PWS are still enigmatic. Whether all or only a certain number of SNORD116 gene copies are functional is also unclear. The aim of the present study was thus to address these questions using a combination of phylogenetic and functional approaches.

Results

Phylogenetics of *SNORD116* Genes

To better understand the evolutionary constraints that shaped *SNORD116* history, we conducted phylogenetic analysis of the 394 gene sequences found at PWS loci in 16 species. The sequences are listed in supplementary table S1, Supplementary Material online. The species were chosen for their distribution over the eutherian tree as well as for the reliability of the nucleotide sequences obtained from genomic and transcriptomic data. Remarkably, the number of
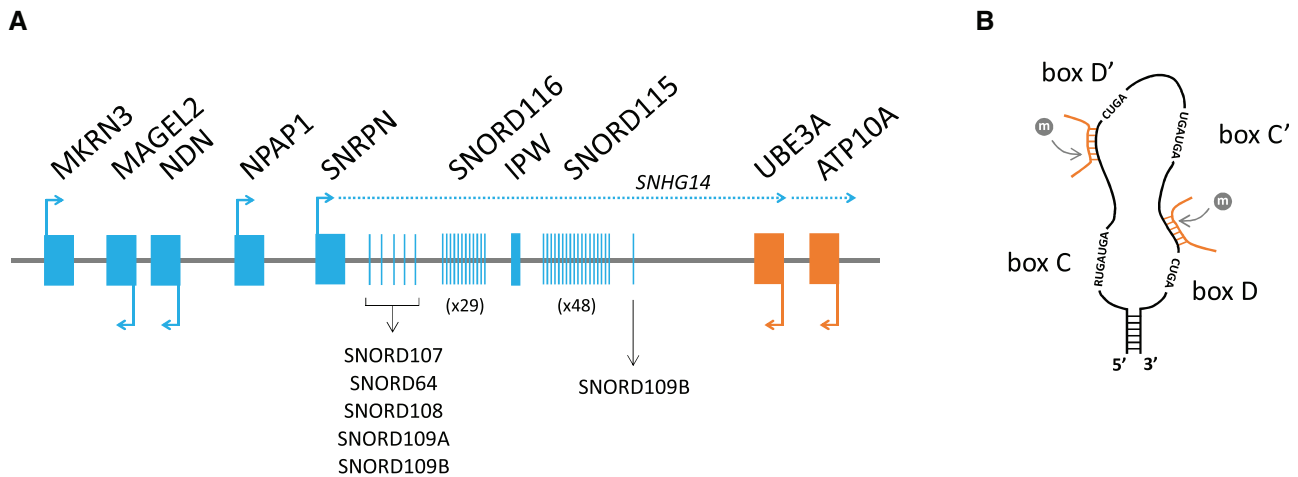
Fig. 1. (A) The human PWS locus contains maternally expressed (orange) and paternally expressed (blue) genes. Protein-coding genes are represented as boxes and arrows indicate the sense of transcription. The C/D snoRNA genes are represented as thin lines. The drawing is not to scale. (B) Representation of a box C/D snoRNA in standard interaction with RNA targets (in orange), that is, involving hybridization with the ASEs positioned on the flanking 5′ side of boxes D and D′. The interaction usually allows modification of the target RNA(s) by the methyltransferase fibrillarin (FBL), one of the core proteins associated with the snoRNA.

paralogs varied between six in pig and 79 in mouse (fig. 2A). The variation concerns closely related species such as Human and chimpanzee (29 vs. 22 genes; last common ancestor ∼6.65 Ma) or mouse and rat (79 vs. 18 genes; last common ancestor ∼20.9 Ma), which supports the hypothesis that a gene birth-and-death process has been extremely active, as previously proposed (Zhang et al. 2014). The p-distance between paralog sequences ranges from very low in hedgehog to the highest score in pig, reinforcing the hypothesis of a complex evolutionary history that included species-specific events. Still, paralog diversity tends to be higher in Primates than in non-Primates ($d = 0.186 \pm 0.034$ vs. 0.099 ± 0.066, unpaired t-test $P = 0.0072$). Despite this, the mean p-distance between human genes and genes from other species is globally independent of the species analyzed ($d = 0.238 \pm 0.015$ and 0.224 ± 0.016 for Primates and non-Primates, respectively; unpaired t-test $P = 0.1541$) and is therefore poorly related to evolutionary distance. To better describe this feature, we generated an unrooted tree showing the relatedness of the 394 homolog sequences and revealed an interlaced pattern of orthologs that is prominent in Primates and Glires (fig. 2B and supplementary fig. S1, Supplementary Material online). In contrast, paralogs from mouse, rat, hedgehog, and bat remain largely monophyletic, suggesting a surge of specific gene copies in these species. To analyze the distribution of nucleotide variation across gene sequences, we first generated a consensus of the 394 homologs and reported the nucleotide variability per position, that is, the percentage of occurrence of nucleotides that differ from the main one at each position (fig. 2C). The regions that classically contribute to snoRNA biogenesis and stability, that is, the basal stem, the C/D boxes, and the C′/D′ boxes are the most conserved ones, suggesting that the SNORD116 genes form bona fide snoRNPs as already proposed (Bortolin-Cavaillé and Cavaillé 2012). The variability of sequences flanked by box C and box D′ and of those flanked by box C′ and box D is similar (∼0.18 variation per

nucleotide) but its distribution differs. Regular C/D snoRNAs use stretches of nucleotides called antisense elements (ASEs) located upstream of box D and box D′ to hybridize with their target RNAs (fig. 1B). Generally, ASEs form 7- to 24-bp long hybrids with their RNA target (Chen et al. 2007; Yang et al. 2016), but most of the interactions range between 10 and 17 bp. Thus, we defined the Snord116 ASEs ASE1 and ASE2 as the 17-mer directly upstream of box D′ and box D, respectively. According to this definition, ASE1 sequences are less variable than ASE2 sequences (0.077 vs. 0.197 variations per nucleotide, unpaired t-test $P = 0.002$), which opens the possibility that the two elements are not subject to the same evolutionary constraints.

SNORD116 Homologs Organize in Subfamilies

The diversification of gene paralogs in subfamilies is an interesting feature that could be linked to a process of pseudogenization or neofunctionalization. In these cases, only a subset of gene copies may still perform the ancestral function. Human SNORD116 genes have been previously grouped in subfamilies based on sequence similarity (Runte et al. 2001). In order to test the conservation of these subfamilies, we used pairwise sequence alignment and identity calculation of all gene homologs. The p-distances are listed in supplementary table S2, Supplementary Material online. First, we grouped human genes using an inclusion threshold of $d < 0.1$, which generated group I (SNORD116-1 to SNORD116-9), group II (SNORD116-12 and SNORD116-14 to SNORD116-24), and group III (SNORD116-25 and SNORD116-26). The remaining gene copies constituted the outgroup. The prioritization of grouping human genes was suggested by the fact that, as described above, Primates exhibit a greater paralog diversity than other clades. We then attributed each ortholog sequence to the closest human group using the same threshold. Using this approach, group I harbored the highest number of genes from the highest number of species, followed by group
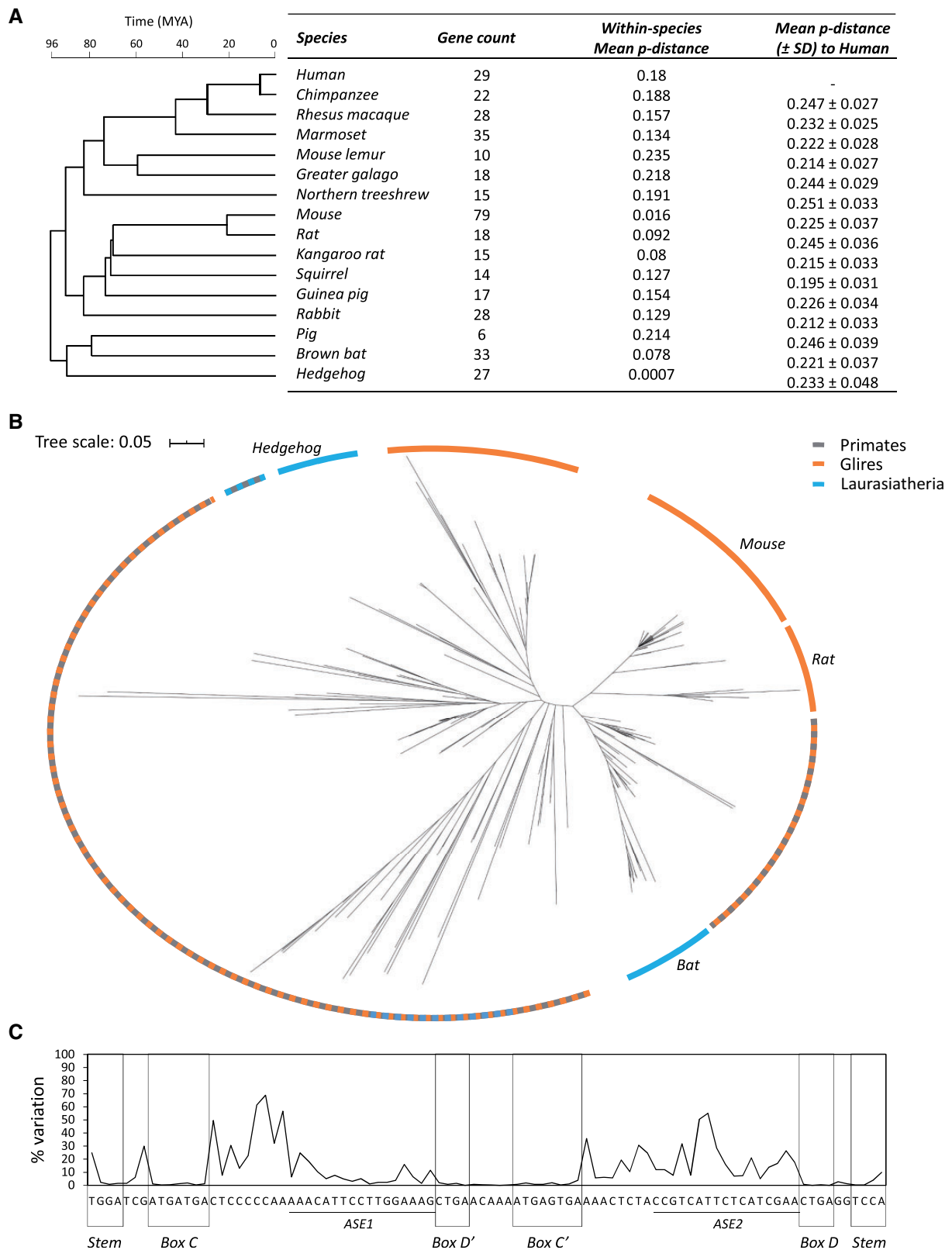
FIG. 2. Inter- and intraspecies phylogenetic comparison of the *SNORD116* genes in 16 eutherian species. (*A*) The number of paralogs, within p-distances and mean p-distance to Human are given for each species (SD, standard deviation). The dendrogram was generated by TimeTree (Kumar et al. 2017). (*B*) Phylogenetic tree of the 394 gene homologs. Dotted lines indicate the presence of gene copies from different clades. (*C*) Consensus sequence of the 394 gene homologs. The most frequent nucleotide is given at each position, and the graph shows the percentage occurrence of the other nucleotides. The basal stem, the C, C′, D, and D′ boxes are boxed and the ASE1 and ASE2 sequences are underlined.
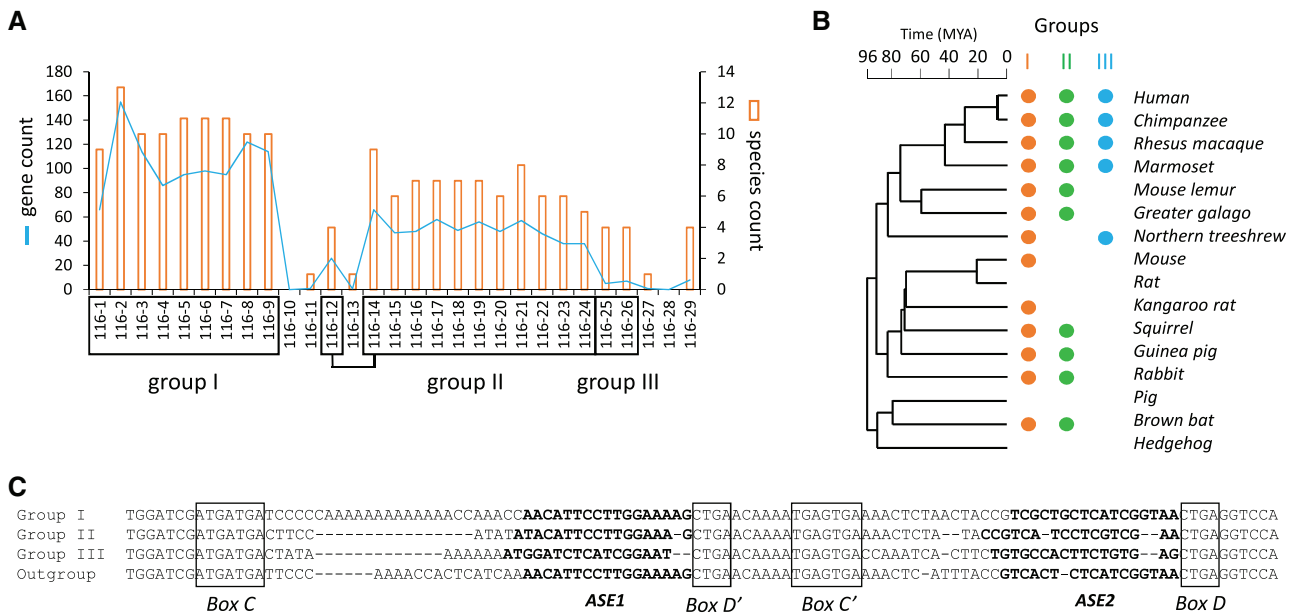
FIG. 3. Identification of SNORD116 gene subfamilies. (A) Number of homolog genes for each human gene copy and the number of species to which they belong. (B) Distribution of gene groups in the species tree. (C) Alignment of the consensus sequences of group I (181 genes), group II (51 genes), group III (8 genes), and the 154 outgroup genes. ASE1 and ASE2 sequences are in bold.

II (fig. 3A). Conversely, group III was found only in Primates and likely appeared more recently (fig. 3B). Therefore, if it is still not totally clear due to their complex evolutionary history, it is plausible that the SNORD116 genes originated from an ancestor related to current group I. Interestingly, most sequences in the outgroup—including those coming from Laurasiatherias—were closer to group I than to the groups II and III (supplementary table S2, Supplementary Material online), this was confirmed by aligning the consensus sequences of the four groups (fig. 3C). Alignment also showed that the main differences between groups come from the region between box C and ASE1, dominated by a variation in the size of an A-rich stretch, and from the region between box C' and ASE2. To be noted, consensus ASE1 sequences are identical in group I and in the outgroup, whereas the consensus ASE2 sequences share the ten last nucleotides, which mainly explains why these two groups are located close to one another.

Genes from each group are distributed from proximal to distal position on human chromosome 15 in a rather orderly manner (Runte et al. 2001). A similarity-based comparison with other primate genomes revealed that the consecutive distribution of the gene groups is largely conserved (supplementary fig. S2, Supplementary Material online), suggesting its presence in a primate ancestor. It also suggests the constitution and maintenance of the subfamilies by nearby duplication and/or another mode of local exchange of gene copies. In this sense, the evolutionary history of SNORD116 genes in Primates appears to be more conservative than previously thought.

To investigate whether the existence of SNORD116 subfamilies could be linked to functional diversification, we analyzed the level of expression of SNORD116 copies in human tissues (supplementary fig. S3, Supplementary Material

online). High-throughput quantification of small- and medium-size noncoding RNAs is often complicated by the fact that sequence reads correspond to incomplete gene annotations. This could be the result of endogenous processing of full-length snoRNAs that generated stable fragments, but also of technical biases including poor reverse transcription efficiency due to the presence of secondary structures or nucleotide modifications on RNA templates. Consequently, we used recent data generated by the thermostable group II intron reverse transcriptase (TGIRT)-seq method where only reads that cover full-length gene annotations were considered (Fafard-Couture et al. 2021). TGIRT reverse transcriptases exhibit higher fidelity and processivity than conventional enzymes (Nottingham et al. 2016). Moreover, the data pipeline included a read assignment correction (Deschamps-Francoeur et al. 2019) that addresses the challenge of multi-mapping issues concerning repeated sequences. The cumulative expression level from seven adult tissues revealed marked variation among SNORD116 copies (supplementary fig. S3A, Supplementary Material online). This pattern of expression cannot be explained by a difference in RNA stability coming from the basal stem or the C/D and C'/D' boxes, as these elements are strictly conserved in all human copies except for two nucleotides in the SNORD116-12 copy. Therefore, with the exception of the latter, whose poor expression could be due to the presence of a C instead of an A that destabilizes the basal stem, it seems that the regulation of expression of each gene copy may be partially independent. Globally, the expression of genes in groups I and II dominated in each tissue analyzed, whereas expression of the other genes was weak or absent (supplementary fig. S3B, Supplementary Material online). Whether the latter are robustly expressed in a subset of tissues that has not yet been tested or are poorly expressed in a constitutive manner remains an open

question. Interestingly, group II genes were significantly more expressed than group I genes in prostate and liver while similarly expressed in the other tissues. Although the data should be interpreted with caution due to the limited number of samples, this variable pattern of expression supports the hypothesis of a process of neofunctionalization of the human *SNORD116* genes.

Microevolution of the *SNORD116* Genes in Human

To gain more insights into the history of the *SNORD116* genes, we analyzed the occurrence of single-nucleotide polymorphisms (SNPs) in human populations. We collected data from the high-quality 1000 Genome data set that contains 121 single polymorphic sites concerning *SNORD116* copies. Of these, only four SNPs are shared by African, American, East Asian, European, and South Asian populations, whereas there was a high prevalence of rare variants, as 93 SNPs (77%) have a minor allele frequency <0.001 (for more details, see supplementary table S3, Supplementary Material online). Accordingly, numerous SNPs were found to be singletons or specific to a population (fig. 4A). If not due to sequencing errors and according to evolutionary considerations, the presence of numerous SNPs that are infrequent could have several causes, for example, purifying selection, selective sweep, population expansion or a combination of these events. As selection could reduce the level of polymorphism in functionally important regions, we then compared the SNP density of different *SNORD116* regions (fig. 4B). The basal stem and the C/D and C′/D′ boxes harbored a low level of polymorphism as could be expected for important structural elements. Strikingly, SNP density was lower at ASE1 than at ASE2 (21 vs. 42, Khi test $P = 0.008$). To better understand this difference, we investigated the origin of the polymorphisms by analyzing their position on an alignment of the human gene copies. Interestingly, of the 99 events that occurred at conserved positions, 66 corresponded to a paralogous sequence variant (PSV) in another copy (supplementary fig. S4, Supplementary Material online). Considering the 46 diallelic positions of the alignment, that is, positions with only two different nucleotides on all paralogs, we found significant overrepresentation of 31 PSVs (Fisher's exact test $P = 0.0016$). Likewise, the difference in the SNP load on the ASE1 and ASE2 sequences was mainly due to overaccumulation of PSVs (11 vs. 32, Khi test $P = 0.001$) but not of point mutations (8 vs. 5). These data suggest that SNPs do not arise only by point mutation but also via gene conversion events that transfer them from donor to acceptor copies. Indeed, one feature of gene conversion is the prevalence of shared nucleotides at paralogous positions. Strikingly, the frequency of SNPs along the consensus gene sequence correlated positively with the level of nucleotide variation between paralogs (fig. 4C; Pearson correlation $r^2 = 0.482$, $P = 4E-7$). Again, this observation fulfills the criteria of gene conversion events whose frequency is likely homogeneous along the gene sequence but whose detection depends on the level of nucleotide variation between donor and acceptor copies.

Strong Conservation of a Subset of ASE1 Sequences

The difference in the level of sequence variation at ASE1 and ASE2 is not only found in Human but in most species analyzed (supplementary fig. S5A, Supplementary Material online, paired *t*-test $P = 0.0005$). To go further, we evaluated if these sequences are conserved between species. We first performed a pairwise comparison between species for each ASE. Overall, we found greater p-distances between ASE2 than between ASE1 elements (fig. 5A and supplementary table S4, Supplementary Material online). Unrooted trees presenting the relatedness of the ASE homologs confirmed the more disperse pattern of ASE2 sequences compared with ASE1 sequences, including the presence of several species-specific leaves (supplementary fig. S5B, Supplementary Material online). In theory, a duplicated pair of paralog genes could tolerate sequence variation, leading to pseudogenization or neofunctionalization of one copy whereas the other maintains the ancestral function. This prompted us to investigate whether interspecies conservation of ASE sequences exists. On the 394 homolog genes, we found 62 and 107 unique ASE1 and ASE2 sequences, respectively, that are highly heterogeneous in their distribution pattern: although 35 ASE1 and 55 ASE2 sequences occur only once, a limited number of sequences are found in many genes and species, and only six sequences are shared by at least four species (fig. 5B–D and supplementary table S5, Supplementary Material online). By far the most frequent sequence corresponds to an ASE1 element shared by 184 genes belonging to group I or to the outgroup, whereas the other sequences shared by at least four species are found in Primates (plus one ASE1 sequence in guinea pig and bat). As an example of the extreme conservation of this sequence, it was found to be present in all 79 paralogs in mouse, whereas they harbored 14 different ASE2 sequences. Overall, the sequence was present in 13 out of the 16 species studied and, in the three other species, a single-nucleotide substitution was found in all *SNORD116* copies in hedgehog and in *SNORD116-1* copy in mouse lemur, whereas two substitutions and one deletion are present in the *SNORD116-1*, *SNORD116-2*, *SNORD116-5*, and *SNORD116-6* copies in pig (fig. 5D). As it is widely found in eutherian species, we named this sequence ASE1-Euth. In conclusion, it is likely that the ASE2 elements undergo relaxed selection compared with ASE1 thereby enabling the formation of a larger repertoire of sequences that are mainly monophyletic. Conversely, one ASE1 sequence dominated, likely the result of a selective effect. Taken together, the data suggest a scenario where the selective maintenance of an ASE1 sequence combined with the horizontal transfer by gene birth-and-death and conversion events explain the fact that the repertoire of ASE1 sequences is drastically reduced compared with the ASE2 repertoire, with the ASE1-Euth sequence behaving as a stable attractor.

Conservation of ASE1–RNA Hybridization Potentials

The conservation of primary structures in noncoding RNAs often underlies their ability to hybridize with complementary RNAs, and not only regular C/D snoRNAs but also orphan ones may use this strategy to affect RNA targets in various
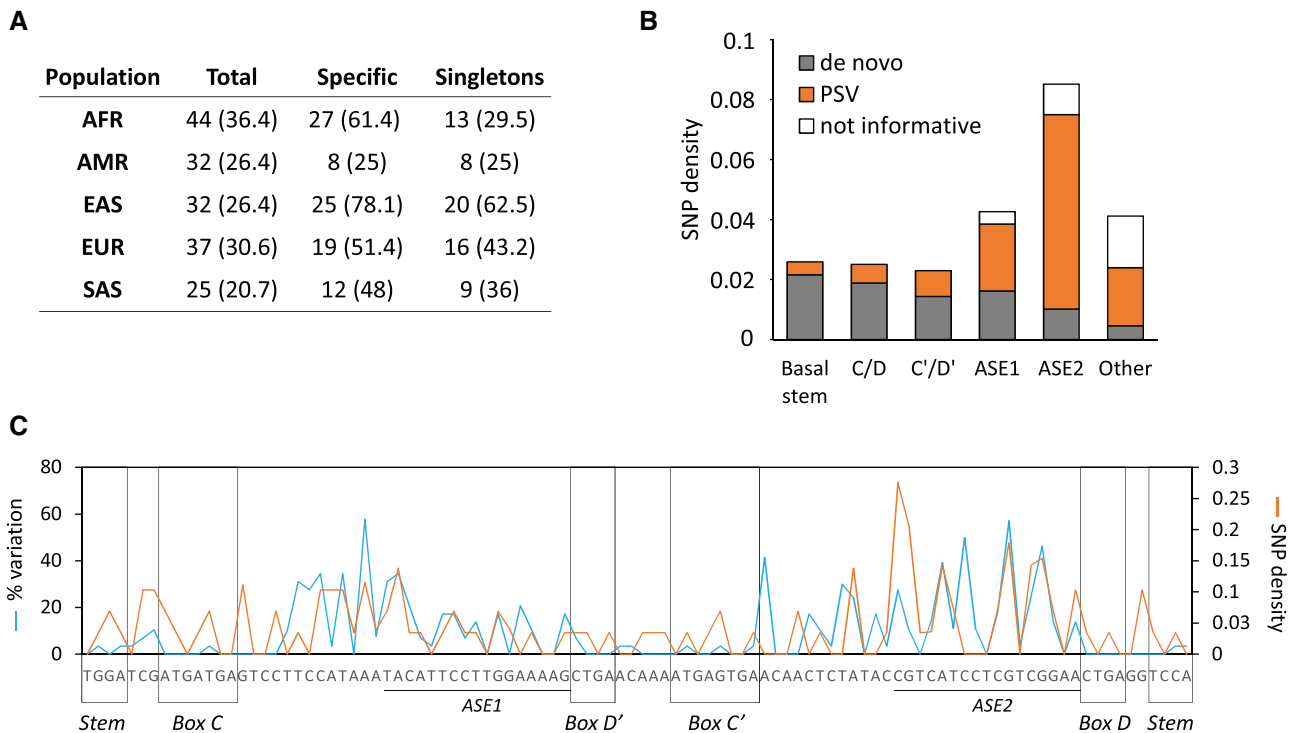
A

Population	Total	Specific	Singletons
AFR	44 (36.4)	27 (61.4)	13 (29.5)
AMR	32 (26.4)	8 (25)	8 (25)
EAS	32 (26.4)	25 (78.1)	20 (62.5)
EUR	37 (30.6)	19 (51.4)	16 (43.2)
SAS	25 (20.7)	12 (48)	9 (36)

B

C

Fig. 4. Characterization of the 1000 Genome data set SNPs present on the *SNORD116* genes. (*A*) SNP occurrence in five human populations. SNP count per population (and the percentage relative to the total SNP count), SNP count specific to each population or occurring as singletons (and the percentage relative to the specific SNP count) are given. (*B*) SNP density at the different snoRNA regions. The density of paralogous variants (PSV) is shown in orange, the density of non-PSVs (de novo) is shown in gray. (*C*) The proportion of sequence variation between human paralogs and SNP density is reported on the consensus sequence of human genes (with the exception of the *SNORD116-10* sequence).
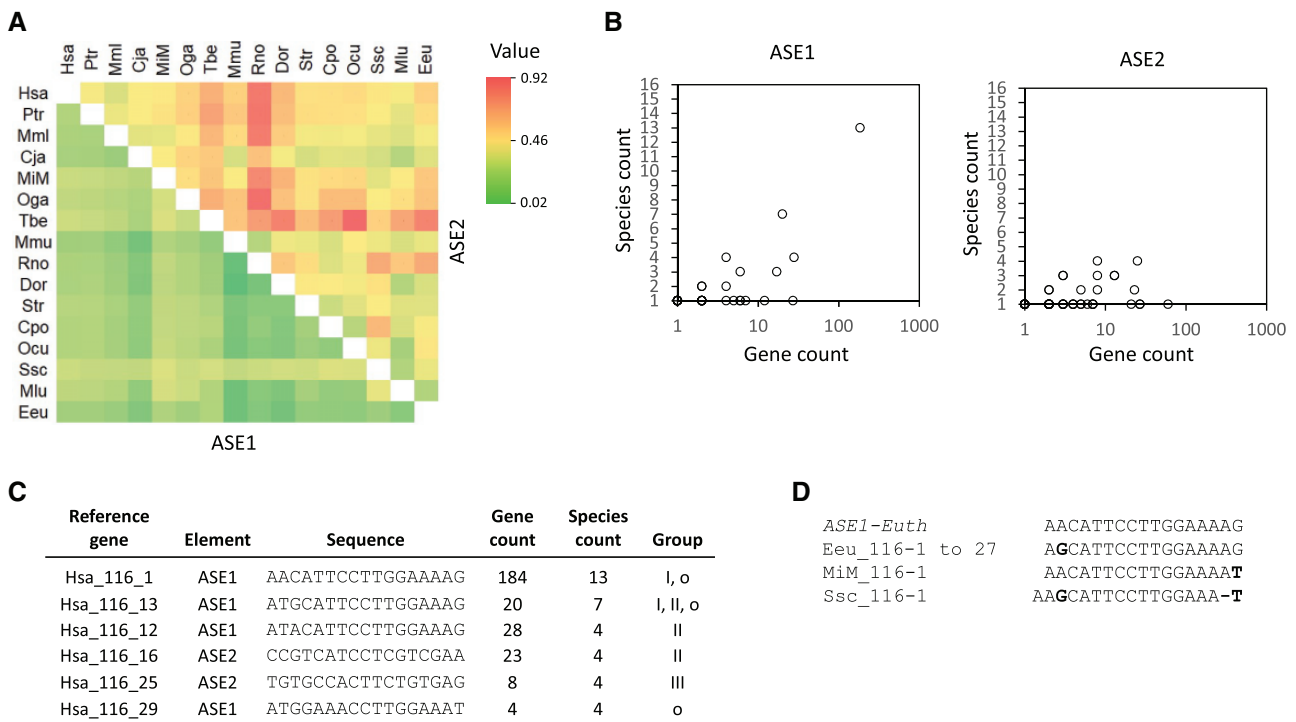
A

B

C

Reference gene	Element	Sequence	Gene count	Species count	Group
Hsa_116_1	ASE1	AACATTCCTTGGAAAAG	184	13	I, o
Hsa_116_13	ASE1	ATGCATTCCTTGGAAAG	20	7	I, II, o
Hsa_116_12	ASE1	ATACATTCCTTGGAAAG	28	4	II
Hsa_116_16	ASE2	CCGTCATCCTCGTCGAA	23	4	II
Hsa_116_25	ASE2	TGTGCCACTTCTGTGAG	8	4	III
Hsa_116_29	ASE1	ATGGAAACCTTGGAAAT	4	4	o

D

```
ASE1-Euth        AACATTCCTTGGAAAAG
Eeu_116-1 to 27  AGCATTCCTTGGAAAAG
MiM_116-1        AACATTCCTTGGAAAAT
Ssc_116-1        AAGCATTCCTTGGAAA-T
```

Fig. 5. Conservation of ASE1 and ASE2 sequences of the *SNORD116* genes. (*A*) P-distances between ASE1 or ASE2 sequences grouped per species. (*B*) Occurrence of ASE1 or ASE2 sequences per genes and per species. The gene count includes paralog and ortholog genes. (*C*) List of the ASE1 and ASE2 sequences found in at least four out of the 16 species analyzed. The reference gene is defined as the proximal one in the human *SNORD116* cluster or in the species closest to Human in the gene tree. The homolog count includes paralog and ortholog genes. O, outgroup. (*D*) Alignment of the ASE1-Euth sequence found in 13 species and its closest variant in hedgehog (Eeu), mouse lemur (MiM), and pig (Ssc).

ways. As C/D snoRNA–rRNA interactions largely involve hybridization between perfect or close complementary sequences, we used a simple BLAST approach to test the potential of Snord116 ASE sequences to hybridize with cellular RNAs. We used two high-quality RNA data sets, the Ensembl human and mouse transcript collections. Indeed, we hypothesized that some *SNORD116* copies have similar molecular functions in the two species, as the phenotypic consequences of *SNORD116/Snord116* deficiencies in patients and mouse models largely overlap. As a control experiment, we performed the same analysis with Snord115 ASE elements and found only one conserved complementarity between most ASE2 sequences and the 5htr2c mRNA, as documented previously (Cavaillé et al. 2000). Concerning Snord116, we found 80 and 111 RNAs that could hybridize with the ASE1 sequences and 25 and 214 RNAs that could hybridize with the ASE2 sequences in Human and mouse, respectively. We did not consider the ASE1 sequence of the human *SNORD116-10* gene (5′-TTTTTTTTTTTTGGAAAG-3′) that exhibited low complexity and, in consequence, association with 453 RNAs. More information concerning these RNAs is available in supplementary table S6, Supplementary Material online. We found three RNAs shared by the two species: Dgkk, Neuroligin 3 (Nlgn3) and the Round spermatid basic protein 1 like (Rsbn1l) mRNAs. The Dgkk interaction sites are located in the middle of exon 8 at a position that could also be used to generate a circular RNA by back splicing (circBase, hsa_circ_0140367), close to the intron2–exon3 junction for Nlgn3 at a position that could also be used for the production of a circular RNA (circBase, hsa_circ_0090986), and on the last exon for Rsbn1l (fig. 6A). Interestingly, the theoretical stability of these interactions is in the same range as that observed in regular C/D snoRNA–rRNA hybrids (fig. 6B). In addition, the interactions are largely conserved: in most species, the best interaction in terms of energy of hybridization and conservation occurs between the ASE1-Euth sequence and the Dgkk mRNA. In hedgehog and pig where ASE1-Euth is absent, a silent U to C substitution in the Dgkk mRNA sequence is offset by A to G substitution in the ASE1 sequence (for all Snord116 copies in hedgehog and for the Snord116-1, -2, -5, and -6 in pig; fig. 6C). Concerning Nlgn3 mRNA, the best interaction involves the ASE1 from Group II genes, except in Glires and bat where it involves the ASE1-Euth sequence. It should be noted that guinea pig is the only species in which hybridization with Rsbn1l was not found. Finally, to test whether other Snord116 regions could have a hybridization potential, we repeated the analysis by scanning entire snoRNA sequences. We performed Blast analyses of 17-mers using a sliding window of 1 nt covering gene sequences close to the consensus of paralog copies: the human *SNORD116-3* gene and the mouse *Snord116-2* gene. However, no region other than ASE1s displayed a hybridization potential with ortholog RNAs from the two species (data not shown).

Snord116 snoRNAs Control the Expression Level of the mRNA Targets

To test the existence of a functional effect, we used a human HeLa S3 cell line to evaluate the capacity of Snord116 snoRNAs to affect the expression of the candidate mRNA targets. We transiently transfected chimeric RNA–DNA antisense oligonucleotides (ASO) called Gapmers to knockdown *SNORD116* expression (fig. 6D, top left). Upon interaction by base-pair complementarity, Gapmers elicit potent RNase H-dependent cleavage of the RNA target. We observed a ∼50% decrease in Snord116 expression 24 h after transfection expression (fig. 6D, bottom left). This effect was similar for snoRNAs produced by group I genes (*SNORD116-1* to *SNORD116-9*) and produced by group II genes (genes *SNORD116-14* to *SNORD116-22*). However, we observed no significant alteration of the level of Snhg14 RNAs produced by the host *SNHG14* gene or of the Snord115 snoRNAs whose gene cluster sited in the close vicinity of the *SNORD116* cluster, suggesting that the destabilization of Snord116 did not affect the expression of the surrounding genes. In contrast, we observed a significant increase in the level of Dgkk, Nlgn3, and Rsbn1l mRNAs. We also observed a significant increase in exon3 inclusion concerning Nlgn3 mRNA (fig. 6D, right), whereas the inclusion of Dgkk exon 8 was not affected (data not shown). The Snord116 interaction site at the 5′ side of Nlgn3 exon 3 overlaps a predicted exon splicing enhancer (ESE). To confirm this status, we constructed an Nlgn3 mini-gene vector and by mutating this sequence, we observed that it indeed promotes exon 3 inclusion (supplementary fig. S6, Supplementary Material online). As we failed to find a cell line that expresses the circular forms of the Dgkk and Nlgn3 RNAs, we were unable to test the effect of Snord116 on these isoforms. In conclusion, these experiments confirmed that the mRNAs identified in the interaction screen can be considered as robust candidate effectors of Snord116 function.

Discussion

Today, knowledge of the molecular functions of the Snord116 snoRNAs remains poor despite the considerable attention they have received since they were shown to belong to the minimal region deleted in PWS patients (Gallagher et al. 2002). Our phylogenetic analysis confirmed the complex history of the *SNORD116* genes dominated by birth-and-death processes as already identified (Zhang et al. 2014), but also provides evidence of pervasive gene conversion events. Our analysis confirms the existence of three subfamilies in Primates and, based on their relative conservation, it could be hypothesized that the ancestral *SNORD116* gene relates to group I. The syntenic position of the subfamilies is largely maintained, suggesting that events of gene conversion between highly similar copies have dominated events between more divergent copies. Such an inverse relationship between the rate of gene conversion and the distance between duplicates has been already documented in the human genome (Harpak et al. 2017).

Despite these events, the core snoRNA elements (i.e., basal stem and C/D boxes) as well as a subset of ASE1 sequences exhibit strong conservation, suggesting functionality. These observations prompted us to test three hypotheses to identify *SNORD116* functions: 1) Snord116 use their ASEs (at least ASE1) to hybridize with one or several RNA targets, as do regular C/D snoRNAs or orphan snoRNAs such as Snord115,
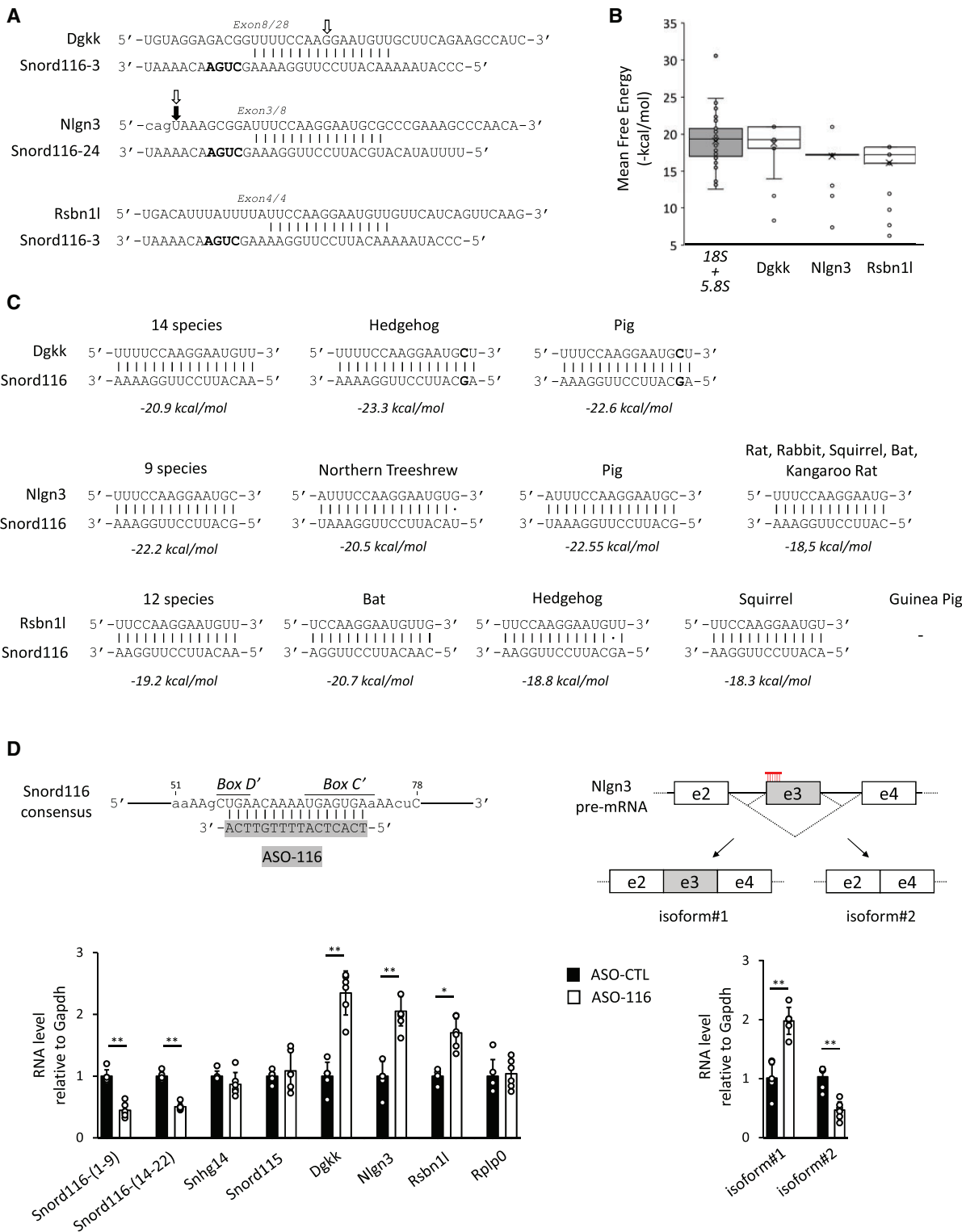
Fig. 6. Expression of *SNORD116* affects the expression and splicing of the predicted mRNA targets. (*A*) Predicted Snord116-target RNA interactions. The D′ box is shown in bold. On the mRNA side, the nucleotide involved in regular splicing and circularization are indicated by closed and open arrows, respectively. The sequence corresponding to Nlgn3 intron 2 is written in lower case. (*B*) Theoretical hybridization energies for regular snoRNA–rRNA interactions and for the interaction between snord116 snoRNAs and the indicated mRNAs in Human. (*C*) Predicted interactions and their energy of hybridization per species. The sequences harboring nucleotide substitutions in hedgehog and pig are shown in bold. (*D*) Upper left panel, position of complementarity between the central part of human *SNORD116* consensus sequence and the ASO-116 Gapmer sequence (highlighted in grey); for a complete view of the alignment and consensus, see supplementary figure S4, Supplementary Material online. Lower left panel, RNA levels following 24 h of treatment with control ASO or ASO-116 (mean±SD of six biological replicates; *$P < 0.05$, **$P < 0.01$, two-tailed Wilcoxon–Mann–Whitney test). Upper right panel, representation of the two Nlgn3 isoforms generated by inclusion or exclusion of exon 3. The snord116 (in red) could hybridize with the 5′ extremity of exon 3. Lower right panel, RT–qPCR assays of the Nlgn3 isoforms.

2) variations in ASE sequences allow different Snord116 copies to target different RNAs; some interactions are hypothesized to be conserved, whereas other could be species-specific, and 3) some RNA targets are found in both Human and mouse, as the absence of *SNORD116/snord116* gene expression leads to largely overlapping phenotypes in these species. Following these hypotheses, we indeed identified a conserved hybridization potential with the Dgkk, Nlgn3, and Rsbn1l mRNAs. The best interaction in terms of conservation and hybridization energy involves the Dgkk mRNA and the highly conserved ASE1-Euth sequence. In pig and hedgehog, this interaction is maintained via reciprocal substitutions, suggesting that a subset of ASE1 sequences has undergone adaptation to the modification of the mRNA sequence. Unlike the Dgkk and Rsbn1l mRNAs, in Primates the best interaction potential with the Nlgn3 mRNA involves an ASE1 variant mostly found in group II genes and only to a lesser extent the ASE1-Euth found in group I genes. Moreover, these two gene groups may have variable expression levels in human tissues as suggested by the analysis of TGIRT-seq data. Therefore, these two hallmarks of neofunctionalization open the possibility that the effects on mRNA targets vary depending on the *SNORD116* copies that are expressed, enabling fine tuning. To gain in consistency, this hypothesis merits studies to scrutinize *SNORD116* functions at a subcluster scale.

We transiently repressed the expression of *SNORD116* genes in human cells by using chimeric RNA–DNA ASO called Gapmers to confirm the target status of the candidate mRNAs. Gapmers allow rapid depletion thereby favoring the detection of direct effects and have already demonstrated their efficiency and selectivity in several categories of small noncoding RNAs (Liang et al. 2011) including C/D snoRNAs such as Snord83b in a human cell line (Sharma et al. 2016) and Snord116 in mouse (Meng et al. 2015). We selected the cervix carcinoma HeLa S3 cell line to perform these experiments because it allows high transfection efficiency and expresses, even modestly, all the RNAs of interest. It should be stressed that the Snord116 snoRNAs are highly and similarly expressed in cerebral and uterine tissues in Human (Cavaillé et al. 2002). The alteration of mRNA levels caused by *SNORD116* downregulation needs to be confirmed using complementary approaches that goes beyond the scope of the present study. It will be important to identify direct RNA–RNA interactions in tissues that are relevant for *SNORD116* physiopathology, as well as to confirm a functional effect at the organismal level, for example, in mouse models. To go further, we invite readers to consult recent reviews that provide an extended discussion on the identification of snoRNA functions (Bergeron et al. 2020; Bratkovič et al. 2020; Baldini et al. 2021). These approaches could represent a long but necessary effort. Indeed, it is not the first time that a candidate interaction has been identified between a C/D snoRNA and an mRNA: the Snord115 snoRNAs have been proposed to regulate splicing and/or editing of 5htr2c mRNAs based on studies largely dependent on artificial overexpression approaches (Vitali et al. 2005; Kishore and Stamm 2006;

Raabe et al. 2019). However, clear in vivo evidence using functional invalidation approaches is pending (Hebras et al. 2020).

We also observed that Snord116 levels affect the pattern of expression of Nlgn3 isoforms. This effect could be the result of two processes: interaction with Nlgn3 exon 3 in the close vicinity of the splicing site could decrease its usage if it occurred on the pre-mRNA, for example, by masking an ESE element, and/or the RNA isoforms that possess a Snord116 hybridization site (i.e., that include exon 3) may be destabilized by the interaction, as is likely the case for the Dgkk and Rsbn1l mRNAs. Both scenarios deserve further study including dissection of the molecular mechanisms used by a snoRNA to alter mRNA stability and splicing. It would also be interesting to test the role played by Snord116 in the production of circular forms of Dgkk and Nlgn3 RNAs in appropriate biological models. Indeed, circular RNA is an emerging class of RNA with a large set of functions (Chen 2020) whose expression is particularly enriched in the brain (Gokool et al. 2020) and may therefore be implicated in PWS etiology.

DGKK and *NLGN3* are important genes for cerebral functions whereas the *RSBN1L* gene has no identified function. *DGKK* has been recently proposed to play a major role in FXS as the main target of the FMRP protein (Tabet et al. 2016). *NLGN3*, a member of the neuroligin family involved in the formation of functional synapses, is a candidate gene for autism (Jamain et al. 2003; Ellegood et al. 2015; Quartier et al. 2019). A *Nlgn3* knockout model in mouse exhibits phenotypic hallmarks of FXS (Baudouin et al. 2012) and it was recently suggested that an *NLGN3/CYFIP1/FMR1* pathway contributes to ASD (Sledziowska et al. 2020). In addition, it has been shown that loss of *Nlgn3* impacts oxytocin signaling in dopaminergic neurons leading to altered behavioral responses to social novelty tests in mouse (Hörnberg et al. 2020). Interestingly, dysfunction of the oxytocin system has been also reported in PWS patients (Kabasakalian et al. 2018) and has been the target of recent clinical trials (Rice et al. 2018). In addition, the function of Nlgn3 splice isoforms is starting to be evaluated and exclusion of exon 3 was recently proposed to increase inhibitory synaptic transmission (Uchigashima et al. 2020) that is reminiscent of the imbalance between inhibitory and excitatory neural circuits that underlies some of the clinical manifestations of PWS (Ates et al. 2019) and FXS (Hagerman et al. 2017). Therefore, the data provided here suggest that both syndromes involve deregulation of a common set of mRNAs, which warrants further analyses.

Finally, we believe that the strategy applied here could help identify the molecular targets of other tandem repeat orphan C/D snoRNA genes, such as the eutherian-conserved *SNORD113* and *SNORD114* clusters located at the imprinted *DLK1-DIO3* locus (human chromosome 14q32). This locus is associated with Temple syndrome (TS) and Kagami–Ogata syndrome (Prasasya et al. 2020) and the identification of *SNORD113* and *SNORD114* targets is considered a priority for understanding these syndromes (Prasasya et al. 2020). Moreover, another tandem repeat family found only in rat (Cavaillé et al. 2001) opens the possibility that other

unsuspected clusters remain to be discovered in little-studied genomes. Identifying their molecular functions will also be important to understand the puzzling presence of this unusual feature of eutherian genomes.

Materials and Methods

Identification of SNORD116 Gene Sequences

We selected 16 eutherians species: Human (*Homo sapiens*, Hsa), chimpanzee (*Pan troglodytes*, Ptr), rhesus macaque (*Macaca mulatta*, Mml), marmoset (*Callithrix jacchus*, Cja), mouse lemur (*Microcebus murinus*, MiM), greater galago (*Otolemur garnettii*, Oga), northern treeshrew (*Tupaia belangeri*, Tbe), mouse (*Mus musculus*, Mmu), rat (*Rattus norvegicus*, Rno), kangaroo rat (*Dipodomys ordii*, Dor), squirrel (*Spermophilus tridecemlineatus*, Str), guinea pig (*Cavia porcellus*, Cpo), Rabbit (*Oryctolagus cuniculus*, Ocu), pig (*Sus scrofa*, Ssc), brown bat (*Myotis lucifugus*, Mlu), and hedgehog (*Erinaceus europeus*, Eeu). To obtain *SNORD116* gene sequences, we combined data from whole-genome annotations using the UCSC (genome.ucsc.edu) and Ensembl (www.ensembl.org/index.html) genome browsers and from the snoRNA databases snoRNA-LBME-db (Lestrade and Weber 2006) and snOPY (Yoshihama et al. 2013). We only collected sequences from *SNRPN-UBE3A* loci. The genome assemblies used were GRCh38.p13 for Human, Pan_tro_3.0 for chimpanzee, Mmul_10 for rhesus macaque, ASM275486v1 for marmoset, Mmur_2.0 for mouse lemur, OtoGar3 for greater galago, TupBel1 for northern treeshrew, GRCm38 for mouse, Rnor_6.0 for rat, DipOrd1 for kangaroo rat, SpeTri2.0 for squirrel, Cavpor3 for guinea pig, OryCun2.0 for Rabbit, Sscrofa11.1 for pig, Myoluc2.0 for brown bat, and EriEur2 for hedgehog. For each species, we checked the accuracy of the annotations and identified gene copies that could have been omitted in data sets using the BLAT/BLAST option of the UCSC and Ensembl browsers. We also performed manual curation to remove obvious pseudogenes. When the accuracy of the genomic data allowed it, *SNORD116* gene copies were numbered from the proximal to the distal position on the tandem repeat. On rare occasions, we removed short 5′ and/or 3′ extensions from transcripts or gene annotations to avoid distortion in gene sequence comparisons. The *SNORD116* gene sequences are listed in supplementary table S1, Supplementary Material online.

Phylogenetic Analyses

The sequences were aligned using the MUSCLE application in EMBL-EBI (Madeira et al. 2019) with default parameters. The number of base substitutions per site was estimated using the MEGA X software by averaging between sequence groups or overall sequence pairs in each group (Kumar et al. 2018) using the Kimura 2-parameter model. The p-distance corresponds to the proportion (p) of nucleotide sites at which two sequences being compared differ and is obtained by dividing the number of nucleotide differences by the total number of nucleotides being compared. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Variance was estimated using the

bootstrap method and 1,000 replicates. The phylogenetic network was constructed with the SIMPLE application in EMBL-EBI using a neighbor-joining clustering method and the phylogenetic trees were generated using the iTOL tool (Letunic and Bork 2019). Consensus sequences were generated from sequence alignments and the percentage of nucleotide variation was calculated as $100-Nt/(Nt-Nm)$, where Nt is the total nucleotide count and Nm is the major nucleotide count at a given position in the alignment. For the sake of clarity, nucleotide positions where gaps were equal to or exceeded 75% of NT were removed. The proportion of each nucleotide in the alignment is shown in supplementary table S7, Supplementary Material online.

Prediction of SNORD116–RNA Interactions

We used the Ensembl interface to perform a BLASTN (ensembl.org/Multi/Tools/Blast) search with distant homologies (maximum hits: 5,000; maximal E-value: 10,000; word size for seeding alignment: 5; match/mismatch: 1, −1; gap penalties: opening: 0, extension: 2) testing human and mouse ASE sequences against human and mouse cDNA (transcripts/splice variants) collections, respectively.

Theoretical Energy of Hybridization

The energy of hybridization between RNA sequences was calculated using the IntaRNA application (Mann et al. 2017). Bona fide interactions between human C/D snoRNAs and 18S or 5.8S rRNAs were obtained at the snoRNA-LBME-db.

Identification of Exon Splicing Enhancers/Silencers and Circular RNA Splicing Sites

The presence of ESE and exon splicing silencers was estimated using the ACESCAN2 web server (Yeo et al. 2005) and human Splicing Finder software (Desmet et al. 2009). The sequence of human circular RNAs was found using circBase (Glažar et al. 2014).

Cell Culture

Human cervix carcinoma HeLa S3 and embryonic kidney Hek293 cells were grown in Dulbecco's-modified Eagle's medium-high glucose (Sigma–Aldrich) supplemented with 10% fetal bovine serum (Dutscher), 1% penicillin/streptomycin (Sigma–Aldrich) and 1% L-glutamine (Sigma–Aldrich).

Gapmer Transfection

A total of 200, 000 HeLa S3 cells were seeded per well in six-well plates 24 h before transfection by Gapmer oligonucleotides targeting human Snord116 RNAs (sequence 5′-TCACTCATTTTGTTCA-3′) or control Gapmers (QIAGEN). Transfection was performed using Lipofectamine 2000 (Invitrogen) at a final Gapmer concentration of 16 nM. Samples were collected 24 h posttransfection for total RNA extraction.

Total RNA Extraction and RT–qPCR

Total RNAs were collected using the miRNeasy mini kit (QIAGEN) and extracted following the manufacturer's

recommendations. The extracted RNAs were quantified using nanodrop 2000. After a DNase step, RNAs were reverse transcribed using the Superscript III kit (Thermo Fisher Scientific) following the manufacturer's recommendations, using a mix of oligo d(T), random hexanucleotides, and oligonucleotides specific for Dgkk and Nlgn3 as primers. RNA expression level was quantified by real-time quantitative PCR (RT–qPCR) using the iTaq Universal SYBR Green Supermix (Bio-Rad) and StepOne Real-Time PCR system (Applied Biosystems). The RT–qPCR primers used in this study are listed in supplementary table S8, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Data Availability

The data underlying this article will be shared on reasonable request to the corresponding author.

References

Adhikari A, Copping NA, Onaga B, Pride MC, Coulson RL, Yang M, Yasui DH, LaSalle JM, Silverman JL. 2019. Cognitive deficits in the Snord116 deletion mouse model for Prader–Willi syndrome. *Neurobiol Learn Mem*. 165:106874.

Ates T, Oncul M, Dilsiz P, Topcu IC, Civas CC, Alp MI, Aklan I, Ates Oz E, Yavuz Y, Yilmaz B, et al. 2019. Inactivation of Magel2 suppresses oxytocin neurons through synaptic excitation-inhibition imbalance. *Neurobiol Dis*. 121:58–64.

Baldini L, Charpentier B, Labialle S. 2021. Emerging data on the diversity of molecular mechanisms involving C/D snoRNAs. *Noncoding RNA*. 7(2):30.

Baudouin SJ, Gaudias J, Gerharz S, Hatstatt L, Zhou K, Punnakkal P, Tanaka KF, Spooren W, Hen R, De Zeeuw CI, et al. 2012. Shared synaptic pathophysiology in syndromic and nonsyndromic rodent models of autism. *Science* 338(6103):128–132.

Bazeley PS, Shepelev V, Talebizadeh Z, Butler MG, Fedorova L, Filatov V, Fedorov A. 2008. snoTARGET shows that human orphan snoRNA targets locate close to alternative splice junctions. *Gene* 408(1–2):172–179.

Bennett JA, Germani T, Haqq AM, Zwaigenbaum L. 2015. Autism spectrum disorder in Prader–Willi syndrome: a systematic review. *Am J Med Genet A*. 167A(12):2936–2944.

Bergeron D, Fafard-Couture É, Scott MS. 2020. Small nucleolar RNAs: continuing identification of novel members and increasing diversity of their molecular mechanisms of action. *Biochem Soc Trans*. 48(2):645–656.

Bochukova EG, Lawler K, Croizier S, Keogh JM, Patel N, Strohbehn G, Lo KK, Humphrey J, Hokken-Koelega A, Damen L, et al. 2018. A transcriptomic signature of the hypothalamic response to fasting and BDNF deficiency in Prader–Willi syndrome. *Cell Rep*. 22(13):3401–3408.

Bortolin-Cavaillé ML, Cavaillé J. 2012. The SNORD115 (H/MBII-52) and SNORD116 (H/MBII-85) gene clusters at the imprinted Prader–Willi locus generate canonical box C/D snoRNAs. *Nucleic Acids Res*. 40(14):6800–6807.

Bozdagi O, Sakurai T, Dorr N, Pilorge M, Takahashi N, Buxbaum JD. 2012. Haploinsufficiency of Cyfip1 produces fragile X-like phenotypes in mice. *PLoS One* 7(8):e42422.

Bratkovič T, Božič J, Rogelj B. 2020. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res*. 48(4):1627–1651.

Burnett LC, LeDuc CA, Sulsona CR, Paull D, Rausch R, Eddiry S, Carli JF, Morabito MV, Skowronski AA, Hubner G, et al. 2017. Deficiency in prohormone convertase PC1 impairs prohormone processing in Prader–Willi syndrome. *J Clin Invest*. 127(1):293–305.

Butler MG, Miller JL, Forster JL. 2019. Prader–Willi syndrome—clinical genetics, diagnosis and treatment approaches: an update. *Curr Pediatr Rev*. 15(4):207–244.

Cavaillé J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Hüttenhofer A. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A*. 97(26):14311–14316.

Cavaillé J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP. 2002. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader–Willi/Angelman syndrome region. *Hum Mol Genet*. 11(13):1527–1538.

Cavaillé J, Vitali P, Basyuk E, Hüttenhofer A, Bachellerie JP. 2001. A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J Biol Chem*. 276(28):26374–26383.

Chai JH, Locke DP, Greally JM, Knoll JH, Ohta T, Dunai J, Yavor A, Eichler EE, Nicholls RD. 2003. Identification of four highly conserved genes between breakpoint hotspots BP1 and BP2 of the Prader–Willi/Angelman syndromes deletion region that have undergone evolutionary transposition mediated by flanking duplicons. *Am J Hum Genet*. 73(4):898–925.

Chen CL, Perasso R, Qu LH, Amar L. 2007. Exploration of pairing constraints identifies a 9 base-pair core within box C/D snoRNA-rRNA duplexes. *J Mol Biol*. 369(3):771–783.

Chen LL. 2020. The expanding regulatory mechanisms and cellular functions of circular RNAs. *Nat Rev Mol Cell Biol*. 21(8):475–490.

Cheon CK. 2016. Genetics of Prader–Willi syndrome and Prader–Will-like syndrome. *Ann Pediatr Endocrinol Metab*. 21(3):126–135.

Coulson RL, Powell WT, Yasui DH, Dileep G, Resnick J, LaSalle JM. 2018. Prader–Willi locus Snord116 RNA processing requires an active endogenous allele and neuron-specific splicing by Rbfox3/NeuN. *Hum Mol Genet*. 27(23):4051–4060.

Cruvinel E, Budinetz T, Germain N, Chamberlain S, Lalande M, Martins-Taylor K. 2014. Reactivation of maternal SNORD116 cluster via SETDB1 knockdown in Prader–Willi syndrome iPSCs. *Hum Mol Genet*. 23(17):4674–4685.

de los Santos T, Schweizer J, Rees CA, Francke U. 2000. Small evolutionarily conserved RNA, resembling C/D box small nucleolar RNA, is transcribed from PWCR1, a novel imprinted gene in the Prader–Willi deletion region, which is highly expressed in brain. *Am J Hum Genet*. 67(5):1067–1082.

De Rubeis S, Pasciuto E, Li KW, Fernández E, Di Marino D, Buzzi A, Ostroff LE, Klann E, Zwartkruis FJ, Komiyama NH, et al. 2013. CYFIP1 coordinates mRNA translation and cytoskeleton remodeling to ensure proper dendritic spine formation. *Neuron* 79(6):1169–1182.

Deschamps-Francoeur G, Boivin V, Abou Elela S, Scott MS. 2019. CoCo: RNA-seq read assignment correction for nested genes and multi-mapped reads. *Bioinformatics* 35(23):5039–5047.

Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res*. 37(9):e67.

de Smith AJ, Purmann C, Walters RG, Ellis RJ, Holder SE, Van Haelst MM, Brady AF, Fairbrother UL, Dattani M, Keogh JM, et al. 2009. A deletion of the HBII-85 class of small nucleolar RNAs (snoRNAs) is associated with hyperphagia, obesity and hypogonadism. *Hum Mol Genet*. 18(17):3257–3265.

Ding F, Li HH, Zhang S, Solomon NM, Camper SA, Cohen P, Francke U. 2008. SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* 3(3):e1709.

Doe CM, Relkovic D, Garfield AS, Dalley JW, Theobald DE, Humby T, Wilkinson LS, Isles AR. 2009. Loss of the imprinted snoRNA mbii-52 leads to increased 5htr2c pre-RNA editing and altered 5HT2CR-mediated behaviour. *Hum Mol Genet*. 18(12):2140–2148.

Duker AL, Ballif BC, Bawle EV, Person RE, Mahadevan S, Alliman S, Thompson R, Traylor R, Bejjani BA, Shaffer LG, et al. 2010. Paternally inherited microdeletion at 15q11.2 confirms a significant role for the SNORD116 C/D box snoRNA cluster in Prader–Willi syndrome. *Eur J Hum Genet*. 18(11):1196–1201.

Ellegood J, Anagnostou E, Babineau BA, Crawley JN, Lin L, Genestine M, DiCicco-Bloom E, Lai JK, Foster JA, Peñagarikano O, Geschwind DH, et al. 2015. Clustering autism: using neuroanatomical differences in 26 mouse models to gain insight into the heterogeneity. *Mol Psychiatry*. 20(1):118–125.

Fafard-Couture É, Bergeron D, Couture S, Abou-Elela S, Scott MS. 2021. Annotation of snoRNA abundance across human tissues reveals complex snoRNA-host gene relationships. *Genome Biol*. 22(1):172.

Falaleeva M, Surface J, Shen M, de la Grange P, Stamm S. 2015. SNORD116 and SNORD115 change expression of multiple genes and modify each other's activity. *Gene* 572(2):266–273.

Gallagher D, Voronova A, Zander MA, Cancino GI, Bramall A, Krause MP, Abad C, Tekin M, Neilsen PM, Callen DF, et al. 2015. Ankrd11 is a chromatin regulator involved in autism that is essential for neural development. *Dev Cell*. 32(1):31–42.

Gallagher RC, Pils B, Albalwi M, Francke U. 2002. Evidence for the role of PWCR1/HBII-85 C/D box small nucleolar RNAs in Prader–Willi syndrome. *Am J Hum Genet*. 71(3):669–678.

Glažar P, Papavasileiou P, Rajewsky N. 2014. circBase: a database for circular RNAs. *RNA* 20(11):1666–1670.

Gokool A, Loy CT, Halliday GM, Voineagu I. 2020. Circular RNAs: the brain transcriptome comes full circle. *Trends Neurosci*. 43(10):752–766.

Gong J, Shao D, Xu K, Lu Z, Lu ZJ, Yang YT, Zhang QC. 2018. RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res*. 46(D1):D194–D201.

Haan N, Westacott LJ, Carter J, Owen MJ, Gray WP, Hall J, Wilkinson LS. 2021. Haploinsufficiency of the schizophrenia and autism risk gene Cyfip1 causes abnormal postnatal hippocampal neurogenesis through microglial and Arp2/3 mediated actin dependent mechanisms. *Transl Psychiatry*. 11(1):313.

Hagerman RJ, Berry-Kravis E, Hazlett HC, Bailey DB Jr, Moine H, Kooy RF, Tassone F, Gantois I, Sonenberg N, Mandel JL, et al. 2017. Fragile X syndrome. *Nat Rev Dis Primers*. 3:17065.

Harpak A, Lan X, Gao Z, Pritchard JK. 2017. Frequent nonallelic gene conversion on the human lineage and its effect on the divergence of gene duplicates. *Proc Natl Acad Sci U S A*. 114(48):12779–12784.

Hebras J, Marty V, Personnaz J, Mercier P, Krogh N, Nielsen H, Aguirrebengoa M, Seitz H, Pradere JP, Guiard BP, et al. 2020. Reassessment of the involvement of Snord115 in the serotonin 2c receptor pathway in a genetically relevant mouse model. *Elife* 9:e60862.

Hörnberg H, Pérez-Garci E, Schreiner D, Hatstatt-Burklé L, Magara F, Baudouin S, Matter A, Nacro K, Pecho-Vrieseling E, Scheiffele P. 2020. Rescue of oxytocin response and social behaviour in a mouse model of autism. *Nature* 584(7820):252–256.

Jamain S, Quach H, Betancur C, Råstam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, et al.; Paris Autism Research International Sibpair Study. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet*. 34(1):27–29.

Juriaans AF, Kerkhof GF, kkken-Koelega ACS. 2021. The spectrum of the Prader–Willi-like pheno- and genotype: a review of the literature. *Endocr Rev*. bnab026. doi: 10.1210/endrev/bnab026.

Kabasakalian A, Ferretti CJ, Hollander E. 2018. Oxytocin and Prader–Willi syndrome. *Curr Top Behav Neurosci*. 35:529–557.

Kehr S, Bartschat S, Stadler PF, Tafer H. 2011. PLEXY: efficient target prediction for box C/D snoRNAs. *Bioinformatics* 27(2):279–280.

Kehr S, Bartschat S, Tafer H, Stadler PF, Hertel J. 2014. Matching of soulmates: coevolution of snoRNAs and their targets. *Mol Biol Evol*. 31(2):455–467.

Keshavarz M, Savriama Y, Refki P, Reeves RG, Tautz D. 2021. Natural copy number variation of tandemly repeated regulatory SNORD RNAs leads to individual phenotypic differences in mice. *Mol Ecol*. 30(19):4708–4722.

Kishore S, Stamm S. 2006. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* 311(5758):230–232.

Kocher MA, Huang FW, Le E, Good DJ. 2021. Snord116 post-transcriptionally increases Nhlh2 mRNA stability: implications for human Prader–Willi syndrome. *Hum Mol Genet*. 30(12):1101–1110.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35(6):1547–1549.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 34(7):1812–1819.

Labialle S, Cavaillé J. 2011. Do repeated arrays of regulatory small-RNA genes elicit genomic imprinting? Concurrent emergence of large clusters of small non-coding RNAs and genomic imprinting at four evolutionarily distinct eutherian chromosomal loci. *Bioessays* 33(8):565–573.

Langouët M, Glatt-Deeley HR, Chung MS, Dupont-Thibert CM, Mathieux E, Banda EC, Stoddard CE, Crandall L, Lalande M. 2018. Zinc finger protein 274 regulates imprinted expression of transcripts in Prader–Willi syndrome neurons. *Hum Mol Genet*. 27(3):505–515.

Langouët M, Gorka D, Orniacki C, Dupont-Thibert CM, Chung MS, Glatt-Deeley HR, Germain N, Crandall LJ, Cotney JL, Stoddard CE, et al. 2020. Specific ZNF274 binding interference at SNORD116 activates the maternal transcripts in Prader–Willi syndrome neurons. *Hum Mol Genet*. 29(19):3285–3295.

Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res*. 34(90001):D158–D162.

Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 47(W1):W256–W259.

Liang XH, Vickers TA, Guo S, Crooke ST. 2011. Efficient and specific knockdown of small non-coding RNAs in mammalian cells and in mice. *Nucleic Acids Res*. 39(3):e13.

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, et al. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res*. 47(W1):W636–W641.

Mann M, Wright PR, Backofen R. 2017. IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic Acids Res*. 45(W1):W435–W439.

Meguro M, Mitsuya K, Nomura N, Kohda M, Kashiwagi A, Nishigaki R, Yoshioka H, Nakao M, Oishi M, Oshimura M. 2001. Large-scale evaluation of imprinting status in the Prader–Willi syndrome region: an imprinted direct repeat cluster resembling small nucleolar RNA genes. *Hum Mol Genet*. 10(4):383–394.

Meng L, Ward AJ, Chun S, Bennett CF, Beaudet AL, Rigo F. 2015. Towards a therapy for Angelman syndrome by targeting a long non-coding RNA. *Nature* 518(7539):409–412.

Morabito MV, Abbas AI, Hood JL, Kesterson RA, Jacobs MM, Kump DS, Hachey DL, Roth BL, Emeson RB. 2010. Mice with altered serotonin 2C receptor RNA editing display characteristics of Prader–Willi syndrome. *Neurobiol Dis*. 39(2):169–180.

Muscogiuri G, Formoso G, Pugliese G, Ruggeri RM, Scarano E, Colao A. 2019. Prader–Willi syndrome: an uptodate on endocrine and metabolic complications. *Rev Endocr Metab Disord*. 20(2):239–250.

Napoli I, Mercaldo V, Boyl PP, Eleuteri B, Zalfa F, De Rubeis S, Di Marino D, Mohr E, Massimi M, Falconi M, et al. 2008. The fragile X syndrome protein represses activity-dependent translation through CYFIP1, a new 4E-BP. *Cell* 134(6):1042–1054.

Nottingham RM, Wu DC, Qin Y, Yao J, Hunicke-Smith S, Lambowitz AM. 2016. RNA-seq of human reference RNA samples using a thermostable group II intron reverse transcriptase. *RNA* 22(4):597–613.

Nowicki ST, Tassone F, Ono MY, Ferranti J, Croquette MF, Goodlin-Jones B, Hagerman RJ. 2007. The Prader–Willi phenotype of fragile X syndrome. *J Dev Behav Pediatr*. 28(2):133–138.

Ohta T, Gray TA, Rogan PK, Buiting K, Gabriel JM, Saitoh S, Muralidhar B, Bilienska B, Krajewska-Walasek M, Driscoll DJ, et al. 1999. Imprinting-mutation mechanisms in Prader–Willi syndrome. *Am J Hum Genet*. 64(2):397–413.

Pace M, Falappa M, Freschi A, Balzani E, Berteotti C, Lo Martire V, Kaveh F, Hovig E, Zoccoli G, Amici R, et al. 2020. Loss of Snord116 impacts lateral hypothalamus, sleep, and food-related behaviors. *JCI Insight*. 5(12):e137495.

Polex-Wolf J, Lam BY, Larder R, Tadross J, Rimmington D, Bosch F, Cenzano VJ, Ayuso E, Ma MK, Rainbow K, et al. 2018. Hypothalamic loss of Snord116 recapitulates the hyperphagia of Prader–Willi syndrome. *J Clin Invest*. 128(3):960–969.

Prasasya R, Grotheer KV, Siracusa LD, Bartolomei MS. 2020. Temple syndrome and Kagami-Ogata syndrome: clinical presentations, genotypes, models and mechanisms. *Hum Mol Genet*. 29(R1):R107–R116.

Quartier A, Courraud J, Thi Ha T, McGillivray G, Isidor B, Rose K, Drouot N, Savidan MA, Feger C, Jagline H, et al. 2019. Novel mutations in NLGN3 causing autism spectrum disorder and cognitive impairment. *Hum Mutat*. 40(11):2021–2032.

Raabe CA, Voss R, Kummerfeld DM, Brosius J, Galiveti CR, Wolters A, Seggewiss J, Huge A, Skryabin BV, Rozhdestvensky TS. 2019. Ectopic expression of Snord115 in choroid plexus interferes with editing but not splicing of 5-Ht2c receptor pre-mRNA in mice. *Sci Rep*. 9(1):4300.

Rice LJ, Einfeld SL, Hu N, Carter CS. 2018. A review of clinical trials of oxytocin in Prader–Willi syndrome. *Curr Opin Psychiatry*. 31(2):123–127.

Runte M, Hüttenhofer A, Gross S, Kiefmann M, Horsthemke B, Buiting K. 2001. The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for UBE3A. *Hum Mol Genet*. 10(23):2687–2700.

Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, Garnica A, Cheung SW, Beaudet AL. 2008. Prader–Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet*. 40(6):719–721.

Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. 2016. Global mapping of human RNA-RNA interactions. *Mol Cell*. 62(4):618–626.

Skryabin BV, Gubar LV, Seeger B, Pfeiffer J, Handel S, Robeck T, Karpova E, Rozhdestvensky TS, Brosius J. 2007. Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. *PLoS Genet*. 3(12):e235.

Sledziowska M, Galloway J, Baudouin SJ. 2020. Evidence for a contribution of the Nlgn3/Cyfip1/Fmr1 pathway in the pathophysiology of autism spectrum disorders. *Neuroscience* 445:31–41.

Tabet R, Moutin E, Becker JA, Heintz D, Fouillen L, Flatter E, Krężel W, Alunni V, Koebel P, Dembélé D, et al. 2016. Fragile X Mental Retardation Protein (FMRP) controls diacylglycerol kinase activity in neurons. *Proc Natl Acad Sci U S A*. 113(26):E3619–E3628.

Tan Q, Potter KJ, Burnett LC, Orsso CE, Inman M, Ryman DC, Haqq AM. 2020. Prader–Willi-Like Phenotype Caused by an Atypical 15q11.2 Microdeletion. *Genes (Basel)* 11(2):128.

Uchigashima M, Leung M, Watanabe T, Cheung A, Le T, Pallat S, Dinis ALM, Watanabe M, Kawasawa YI, Futai K. 2020. Neuroligin3 splice isoforms shape inhibitory synaptic function in the mouse hippocampus. *J Biol Chem*. 295(25):8589–8595.

Vitali P, Basyuk E, Le Meur E, Bertrand E, Muscatelli F, Cavaillé J, Huttenhofer A. 2005. ADAR2-mediated editing of RNA substrates in the nucleolus is inhibited by C/D small nucleolar RNAs. *J Cell Biol*. 169(5):745–753.

Wu H, Yin QF, Luo Z, Yao RW, Zheng CC, Zhang J, Xiang JF, Yang L, Chen LL. 2016. Unusual Processing Generates SPA LncRNAs that Sequester Multiple RNA Binding Proteins. *Mol Cell*. 64(3):534–548.

Yang Z, Lin J, Ye K. 2016. Box C/D guide RNAs recognize a maximum of 10 nt of substrates. *Proc Natl Acad Sci U S A*. 113(39):10878–10883.

Yeo GW, Van Nostrand E, Holste D, Poggio T, Burge CB. 2005. Identification and analysis of alternative splicing events conserved in human and mouse. *Proc Natl Acad Sci U S A*. 102(8):2850–2855.

Yin QF, Yang L, Zhang Y, Xiang JF, Wu YW, Carmichael GG, Chen LL. 2012. Long noncoding RNAs with snoRNA ends. *Mol Cell*. 48(2):219–230.

Yoshihama M, Nakao A, Kenmochi N. 2013. snOPY: a small nucleolar RNA orthological gene database. *BMC Res Notes*. 6:426.

Zhang YJ, Yang JH, Shi QS, Zheng LL, Liu J, Zhou H, Zhang H, Qu LH. 2014. Rapid birth-and-death evolution of imprinted snoRNAs in the Prader–Willi syndrome locus: implications for neural development in Euarchontoglires. *PLoS One* 9(6):e100329.