U1-like snRNAs lacking complementarity to canonical 5' splice sites

CHRISTINA KYRIAKOPOULOU,^{1,3,4} PONTUS LARSSON,^{1,3} LEI LIU,^{1,3} JENS SCHUSTER,^{1,3}
FREDRIK SÖDERBOM,² LEIF A. KIRSEBOM,¹ and ANDERS VIRTANEN¹

¹Department of Cell and Molecular Biology, Uppsala University, SE-75124 Uppsala, Sweden

²Department of Molecular Biology, Swedish University of Agricultural Sciences, SE-75124 Uppsala, Sweden

ABSTRACT

We have detected a surprising heterogeneity among human spliceosomal U1 small nuclear RNA (snRNA). Most interestingly, we have identified three U1 snRNA variants that lack complementarity to the canonical 5' splice site (5'SS) GU dinucleotide. Furthermore, we have observed heterogeneity among the identified variant U1 snRNA genes caused by single nucleotide polymorphism (SNP). The identified snRNAs were ubiquitously expressed in a variety of human tissues representing different stages of development and displayed features of functional spliceosomal snRNAs, i.e., trimethylated cap structures, association with Sm proteins and presence in nuclear RNA–protein complexes. The unanticipated heterogeneity among spliceosomal snRNAs could contribute to the complexity of vertebrates by expanding the coding capacity of their genomes.

Keywords: U1 snRNP; mRNA splicing; noncanonical 5' splice site; SNP; genome complexity

INTRODUCTION

The small nuclear ribonucleoprotein particles (snRNPs) U1, U2, U4/U6, and U5 are essential for pre-messenger RNA (pre-mRNA) splicing, and form the major spliceosome in combination with a multitude of protein factors (for review, see Jurica and Moore 2003; Patel and Steitz 2003; Matlin et al. 2005). Several sequence elements within the intron are required for the spliceosome catalyzed splicing reaction; among these, the 5' splice site (5'SS), the 3' splice site (3'SS), and the branch point (BP) sequences are essential. In order to assure correct splicing, the spliceosome needs to be highly flexible, recognizing a broad variety of 5'SS, 3'SS, and BP sequences (Burge et al. 1999). Spliceosomal snRNAs are encoded by genes present in multiple copies within the human genome (see Lander et al. 2001; Waterston et al. 2002; Gibbs et al. 2004; Hillier et al. 2004), and early studies revealed that, besides the bona fide snRNAs, lower abundant variants of some snRNAs are expressed (Manser and Gesteland 1982;

Monstein et al. 1983; Westin et al. 1984; Lund 1988). In the majority of analyzed introns the canonical 5'SS, which is recognized by U1 snRNA (Mount et al. 1983; Zhuang and Weiner 1986, 1990; Zhuang et al. 1989; Rosbash and Seraphin 1991; Patel and Steitz 2003), is conserved and contains the dinucleotide GU at the splice junction. However, several cases of noncanonical 5'SS sequences have been observed where the dinucleotide is replaced by, e.g., GC, GG, CU, or GA (Burset et al. 2000, 2001). In this respect, lower abundant U1 snRNA variants could play a role in recognizing noncanonical 5'SSs. We were therefore encouraged to investigate the multitude of expressed spliceosomal U1 snRNA variants and identified three unique U1-like snRNP variants that lack complementarity to the canonical 5'SS GU dinucleotide. Interestingly, the newly identified snRNP variants could potentially interact with noncanonical 5'SSs.

RESULTS AND DISCUSSION

To identify U1-like snRNA genes, we searched the human genome for sequences similar to U1A snRNA (encoded by the *HUMURIA* gene) using the BLAST-like Alignment Tool (BLAT) in combination with the RepeatMasker tool. One hundred eighty-eight putative genes were identified and aligned to the U1A snRNA sequence (Supplemental Fig. S1, Table S1, available at http://www.icm.uu.se/molcell/virtanen/kyriakopoulou_2005/supplementary.php).

³These authors contributed equally to this work.

⁴**Present address:** European Commission, Office-CDMA 02/161, 1049 Brussels, Belgium.

Reprint requests to: Anders Virtanen, Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, SE-75124 Uppsala, Sweden; e-mail: anders.virtanen@icm.uu.se; fax: +46 18 530396.

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.26506>.

One hundred sixty-one of the putative 188 genes were disregarded for further analysis because they showed a high sequence similarity to U1A snRNA or lacked characteristics of expressed snRNA genes, e.g., promoter/enhancer motifs, 3' processing signals or Sm-binding sites (Ciliberto et al. 1986; Mattaj et al. 1988). Subsequent Northern blot analyses indicated that eight of the remaining genes were expressed as RNA in HeLa cells (Supplemental Table S1). Finally, rapid amplification of cDNA ends (5'- and 3'-RACE) and molecular cloning confirmed that at least three of the eight candidate genes were expressed (Supplemental Fig. S1; Table S1). The identified snRNA variants showed 75%, 72%, and 80% sequence identity to the U1A snRNA sequence and were named U1A5, U1A6, and U1A7 snRNA, respectively (Fig. 1). In addition to these three snRNAs, detected by Northern blot analysis, one snRNA variant was found in an EST database and named U1A4 (Supplemental Table S1). The sequence of U1A4 snRNA showed over 90% sequence similarity to U1A snRNA, and was therefore disregarded for further analysis.

We performed several sets of analyses to characterize U1A5, U1A6, and U1A7 snRNA in more detail. A quantitative Northern blot analysis of HeLa cell RNA indicated that they were present in ~2% (U1A6 and U1A7 snRNA) or 7% (U1A5 snRNA) relative to the level of U1A snRNA (Fig. 2A; Supplemental Fig. S2), suggesting that the identified snRNA molecules are present in ~20–70 thousand copies per cell, in amount comparable to the minor spliceosomal U11 and U12 snRNAs (Yu et al. 1999). The expression pattern was investigated by Northern blot analysis and revealed that they all were ubiquitously expressed in a wide range of human tissues representing distinct developmental stages (Fig. 2B). We determined the expression levels of the snRNA variants relative to U1A snRNA in each tissue and observed that the snRNA levels are comparable between the different tissues, although U1A5 and U1A6 snRNA seem to be slightly higher expressed in embryonic tissue (see Materials and Methods for details; Supplemental Fig. S2). Immunoprecipitation experiments showed that the snRNA variants displayed features of functional spliceosomal snRNAs (Will and Luhrmann 2001) as they contained trimethylated cap-structures, were associated with Sm-proteins and were present in nuclear RNA–protein complexes (Fig. 2C,D). We conclude that (1) significant amounts of stable variants of at least three U1-like snRNA variants are present in human cells, and (2) these three variants have been transported to the cytoplasm, hypermethylated, assembled into RNA–protein complexes, and subsequently reimported into the nucleus.

The structural and functional features of U1A snRNA have been extensively characterized, and it adopts a typical cloverleaf-like structure (Fig. 1B), which contains a four-way junction, four stem-loop structures (numbered I to IV) and a central stem (stem A) (Krol et al. 1990). The

binding sites for two U1A snRNP specific proteins are located in the loop regions of stem-loops I and II (Fig. 1B) (Stark et al. 2001). To investigate whether these features were present in U1A5, U1A6, and U1A7 snRNA, we performed a multiple alignment together with the U1A snRNA sequence (Fig. 1A). Differences between U1A snRNA and the U1A5, U1A6, and U1A7 snRNA variants could be observed (Fig. 1), most notably in the 5'SS recognition motif that lacked complementarity to the canonical GU dinucleotide of the 5'SS and in the region surrounding the U1–70K binding site. Based on the multiple sequence alignment we performed secondary structure modeling, assuming that the putative 5'SS recognition motif and the Sm-binding site reside in single-stranded regions of the molecules. The predicted structures of U1A6 and U1A7 snRNA resembled the U1A snRNA structure, whereas U1A5 snRNA displayed a distinct structure (Fig. 1B). Thus, it seems reasonable that at least U1A6 and U1A7 could form snRNP particles similar to U1A snRNP in both structure and protein composition, whereas U1A5 may adopt a unique structure that might even represent a distinct class of spliceosomes. A detailed analysis of the composition and structure of U1A5, U1A6, and U1A7 snRNPs is required to unambiguously resolve this issue.

During cloning and sequencing of the U1A6 snRNA, we observed a discrepancy between the obtained cDNA sequence and the reference sequence of the human genome. Therefore, we resequenced the human U1A6 snRNA gene and searched the NCBI Single Nucleotide Polymorphism (SNP) database for SNPs. Eight variable positions in the gene for U1A6 snRNA were identified (Fig. 1B; Supplemental Fig. S3). Out of these eight variable positions, the nucleotide at position 8 (Y₈) is of particular interest, since it could influence the recognition of the 5'SS by U1A6 snRNA (Fig. 1B–D). In addition, we identified a stretch of 11 nucleotides within stem-loop II of U1A6 snRNA that was either present or absent (Fig. 1B). The high frequency of variable positions within the U1A6 snRNA gene prompted us to search the SNP database for variable positions within the U1A4, U1A5, and U1A7 snRNA genes, and one SNP each was found in the genes for U1A4 and U1A7 snRNA (Fig. 1B; Supplemental Fig. S3). Taken together, the observed repertoire of snRNA variants seems to be further expanded by sequence variation.

The observed differences in the putative 5'SS recognition motifs of the U1A5, U1A6, and U1A7 snRNA variants show that these variants lack complementarity to the canonical GU dinucleotide at the 5'SS. This could imply that these variants will not efficiently recognize a canonical 5'SS. To investigate this, we first analyzed the U1A snRNA:5'SS interaction following the approach of (Carmel et al. 2004) using a data set of human 5'SS sequences extracted from 252,159 canonical GU-AG introns. A Gaussian-like distribution centered around seven base pairs was observed, and

FIGURE 1. U1A snRNA variants. (A) Alignment of the RNA sequences of the U1A, U1A4, U1A5, U1A6, and U1A7 snRNA variants. Nucleotides in U1A that are conserved in a variant snRNA are highlighted in green. Regions corresponding to important features within U1A snRNA are indicated by colored boxes beneath the alignment: 5' SS recognition motif (5' SS, blue); stem A (A, white); U1-70K protein binding site (U1-70K, yellow); U1-A protein binding site (U1-A, orange); Sm-binding site (Sm, gray). The numbers to the right refer to the number of nucleotides counted from the 5' end of the individual snRNA. (B) Predicted secondary structures for the U1A, U1A5, U1A6, and U1A7 snRNA variants. Regions corresponding to important features within U1A snRNA are color coded as specified above. Identified variable positions are highlighted by black boxes. The regions of the snRNAs targeted by the sequence specific probes used for Northern blot analysis are indicated by black lines. (C) Interaction of U1A, U1A5, U1A6, and U1A7 snRNA variants with canonical GU dinucleotide 5' SS sequences. The 5' SS is shown as a sequence logo constructed from 252,159 canonical GU dinucleotide 5' SS sequences. The putative 5' SS recognition motifs of the U1A5, U1A6, and U1A7 snRNA variants are aligned to the 5' SS recognition motif of U1A snRNA. Nucleotides differing from the corresponding nucleotides in U1A snRNA are indicated in bold. The variable nucleotide position Y₈ in U1A6 snRNA is highlighted by a black box. (D) Distributions illustrating the interaction between the U1A, U1A5, U1A6, and U1A7 snRNA variants, respectively, and the above-specified set of canonical GU dinucleotide 5' SS sequences. 5' SSs have been grouped according to the number of possible base pairs.

FIGURE 2. The U1-like snRNA variants are ubiquitously expressed, and show features characteristic of spliceosomal snRNAs. (A) Northern blot analysis of snRNA variants in HeLa cells. The following probes were used: u1a (lane 1), u1a5 (lane 2), u1a6 (lane 3), u1a7 (lane 4), and 5s (lane 5). Positions of RNA size markers are indicated on the left. (B) RNA samples prepared from HeLa cells and a range of human tissues obtained from distinct developmental stages were analyzed by Northern blot analysis. The U1 snRNA variants detected by the specific oligonucleotide probes are indicated on the left and tissue types are indicated above the panels. RNA isolated from HeLa cells (lanes 1,10,11,19,20,28), adult tissues (lanes 2–9), 6-week embryonic tissue (lanes 12–18), and 12-week embryonic tissue (lanes 21–27). (C) Northern blot analysis of immunoprecipitation (IP) experiments. HeLa nuclear extract was subjected to IP using an anti-m₃G/m₃G-cap monoclonal antibody (mAbH20; lane 1), an anti-snRNA-m₃G-cap polyclonal antiserum (r-R1131; lane 3), or an anti-Sm-protein monoclonal antibody (mAbY12; lane 5), respectively. Controls were included utilizing mouse immunoglobulin (mIgG, lanes 2,6) and normal rabbit serum (NRS, lane 4), respectively. Immunoprecipitated RNA was purified and analyzed by Northern blot analysis. U1 snRNA variants or 5S RNA detected by the specific oligonucleotide probes are indicated. (D) Northern blot analysis of electrophoretic mobility shift assays (EMSA) using nuclear extract preparations; nuclear extract (NE; lanes 1,3,5,7); nuclear RNA (RNA; lanes 2,4,6,8). The U1 snRNA variants detected by specific oligonucleotide probes are indicated above the panels. The position of RNA–protein complexes (snRNP) and free snRNA molecules (snRNA) are indicated.

pairs between U1A snRNA and the 5'SS (Zhuang and Weiner 1986; Ketterling et al. 1999; Carmel et al. 2004; Freund et al. 2005). Next, we used the same data set of human canonical introns to investigate possible base pairing interactions between the U1A5, U1A6, and U1A7 snRNA variants and 5'SS sequences (Fig. 1D). The number of base pair interactions between the 5'SS and U1A7, U1A6, and U1A5 snRNA were centered around six, five, and four base pairs, and only in 71%, 26%, and 14% of the cases could six or more base pairs be formed, respectively. In all three cases the decreased number of possible base pairs formed is caused by at least one mismatch to the conserved GU dinucleotide in the 5'SS. These data indicate that the interactions between the U1A snRNA variants and canonical 5'SS sequences are not optimal. To investigate if the variant U1A snRNAs could act on noncanonical 5'SSs, we searched the human AceView database for noncanonical 5'SS sequences that could be recognized by U1A5, U1A6, or U1A7 snRNA through formation of six or more base pairs, including base-pair interactions with the noncanonical dinucleotide at the 5'SS. We found 301 (CU dinucleotide; U1A5 snRNA), 344 (249 AA dinucleotide and 95 GA dinucleotide; U1A6 snRNA), and 16 (UU dinucleotide; U1A7 snRNA) noncanonical 5'SS sequences that fulfilled this criteria (Supplemental Table S2). Thus, it appears likely that the three identified snRNA variants could play a role in recognizing noncanonical 5'SS sequences. The identification of functional substrates for U1A5, U1A6, and U1A7 snRNP will be a primary objective for future studies.

To investigate the evolutionary conservation of the human loci *RNU1A5*, *RNU1A6*, and *RNU1A7* in other organisms we first investigated genome sequences of different species focusing for each locus on an ~500 nucleotides-long DNA fragment that contained the snRNA coding region and flanking sequences (see Materials and Methods for details and a complete list of species). We could by this strategy identify the loci corresponding to *RNU1A6* and *RNU1A7*

six or more base pairs could be formed in 93% of the cases (Fig. 1D). This is in agreement with previous studies, which show that splicing requires the minimal number of six base

sequences (see Materials and Methods for details and a complete list of species). We could by this strategy identify the loci corresponding to *RNU1A6* and *RNU1A7*

in the genomes of the cow, dog, and several primates, and the locus corresponding to *RNU1A5* in primates (Fig. 3A). Notably, all three loci were located in the sense orientation within the first intron of genes classified as testis-expressed genes (*TEX*); *RNU1A5* was within the *TEX27* gene (Lopez-Fernandez and del Mazo 1996), and *RNU1A6* and *RNU1A7* were both within the *TEX14* gene (Wu et al. 2003). We could not convincingly identify any of the loci in any of the other genomes that we investigated, which included vertebrates (rodents, birds, amphibians, and fishes), invertebrates (insects and worms), or unicellular eukaryotes (yeasts), even if we searched the corresponding region of the *TEX* gene when it was present. We next aligned the identified loci with each other (Fig. 3B; Table 1). In the case of *RNU1A5*, it is clear that this locus, including both flanking and coding sequences, is highly conserved between

Homo sapiens (*H.s.*), *Pan troglodytes* (*P.t.*), and *Macaca mulatta* (*M.m.*). However, in *Callithrix jacchus* (*C.j.*), the coding sequence has diverged significantly, most likely due to the appearance of a large deletion within the coding region that should inactivate a putative snRNA. In the case of *RNU1A6* and *RNU1A7*, a slightly different picture emerged, the *RNU1A6* locus being highly conserved between *H.s.* and *P.t.* and the *RNU1A7* locus being highly conserved between *H.s.*, *P.t.*, and *Pongo pygmaeus* (*P.p.*). Stunningly, both these loci in *Canis familiaris* (*C.f.*) and *Bos taurus* (*B.t.*) contain perfect copies of coding regions representing bona fide U1A snRNA, with the exception of one nucleotide difference each in the *RNU1A6* locus of *C.f.* and the *RNU1A7* locus of *B.t.* We also note that the loci *RNU1A6* of *P.p.* and *RNU1A7* of *M.m.*, just as the *RNU1A5* locus of *C.j.*, contain coding regions that have diverged significantly most likely due to the appearances of large deletions that inactivated these loci.

Taken together, our evolutionary analyses strongly suggest that all these loci have recently appeared during evolution, and that they represent rapidly evolving sequences. The analyses also suggest that the variant snRNAs have evolved from bona fide U1 snRNA encoding genes, at least in the case of the *RNU1A6* and *RNU1A7* loci. Furthermore, the fast divergence of the snRNA coding sequences that have acquired deletions supports the conclusion that the expressed human U1A5, U1A6, and U1A7 snRNA variants are functional, since their genes have evolved without becoming transcriptionally inactive and losing properties of functional snRNA. Finally, the evolutionary analyses imply that fast evolution of U1 snRNA genes could be linked to speciation.

FIGURE 3. Evolutionary analysis of the *RNU1A5*, *RNU1A6*, and *RNU1A7* loci. (A) Loci corresponding to the human *RNU1A5*, *RNU1A6*, and *RNU1A7* locus as identified in the chimpanzee (*Pan troglodytes*, *P.t.*), orangutan (*Pongo pygmaeus*, *P.p.*), rhesus macaque (*Macaca mulatta*, *M.m.*), common marmoset (*Callithrix jacchus*, *C.j.*), dog (*Canis familiaris*, *C.f.*), and cow (*Bos Taurus*, *B.t.*). The presence and similarity of an snRNA sequence at a specific locus is indicated by colored squares: similar to U1A5 snRNA (green), U1A6 snRNA (blue), U1A7 snRNA (yellow), or bona fide U1A snRNA (orange); presumably nonfunctional snRNA (dark gray); not detected (ND). Phylogenetic relationships are depicted according to the NCBI taxonomy database (not to scale). An asterisk next to a species name indicates that the loci were found in the NCBI trace archives. (B) Schematic representation of multiple alignments of the identified *RNU1A5*, *RNU1A6*, and *RNU1A7* loci. The locations of the putative coding sequences are indicated by arrows above the alignment and are color coded as above. Flanking regions consist of 150 nucleotides upstream and downstream. The similarities of the sequences relative to the human sequence are indicated: identical nucleotides (boxes), mismatch or deletion (thin line), and insertion (thick line).

CONCLUDING REMARKS

The comparably low number of protein-coding genes in vertebrates relative to lower eukaryotes and invertebrates has been one of the major surprises during recent years (Lander et al. 2001; Waterston et al. 2002; Gibbs et al. 2004). Alternative splicing is thought to be, particularly in multicellular organisms, one of the key mechanisms that contributes to the structural and functional complexity of proteins

TABLE 1. Sequence conservation

Locus	Species	Percent identity			U1A ^b
		Upstream ^a	Coding ^a	Downstream ^a	
<i>RNU1A5</i>	<i>H. sapiens</i>	100%	100%	100%	75%
	<i>P. troglodytes</i>	97%	99%	97%	75%
	<i>M. mulatta</i>	93%	91%	91%	71%
	<i>C. jacchus</i>	90%	63%	75%	58%
<i>RNU1A6</i>	<i>H. sapiens</i>	100%	100%	100%	72%
	<i>P. troglodytes</i>	97%	91%	94%	74%
	<i>P. pygmaeus</i>	89%	58%	89%	51%
	<i>C. familiaris</i>	61%	72%	53%	99%
	<i>B. taurus</i>	61%	73%	51%	100%
<i>RNU1A7</i>	<i>H. sapiens</i>	100%	100%	100%	80%
	<i>P. troglodytes</i>	99%	94%	97%	80%
	<i>P. pygmaeus</i>	95%	85%	90%	82%
	<i>M. mulatta</i>	88%	58%	57%	58%
	<i>C. familiaris</i>	67%	80%	52%	100%
	<i>B. taurus</i>	63%	80%	57%	99%

^aPercent identity relative to the sequence of the corresponding locus in *H. sapiens*. Upstream and downstream correspond to sequence fragments containing 150 nt that are located upstream or downstream of the coding sequence, respectively. Coding refers to the snRNA coding, predicted coding, or corresponding sequence at each locus.

^bPercent identity refers to the coding sequence of the indicated variant U1 snRNA relative to the coding sequence of the bona fide U1 snRNA gene, i.e., *HUMUR1A* in *H. sapiens*.

(Graveley 2001; Black 2003; Sharp 2005). However, even if the importance of alternative splicing is widely recognized, very little is known about the molecular mechanisms controlling the splicing reaction, including both constitutive and alternative splicing events. Compelling evidence suggests that RNA binding proteins, which associate with the pre-mRNA, influence splice site recognition either by enhancing or repressing the ability of the spliceosome to interact and subsequently catalyze the splicing reaction (Jurica and Moore 2003; Park et al. 2004; Hertel and Graveley 2005). RNA-binding proteins are therefore believed to be primary regulators of the splicing reaction. In light of this, our study opens up the possibility for at least two previously unanticipated strategies by which splicing could be regulated by U1 snRNA variants. First of all, we envisage that the heterogeneity among the U1 snRNAs will cause a similar heterogeneity of the major spliceosome. This heterogeneity could play a role in recognizing 5'SSs that are not optimally suited to interact with the predominant U1A snRNA. Sontheimer and Steitz (1992) have previously detected heterogeneity among human U5 snRNAs that is reflected in affinity purified spliceosomes. Heterogeneity among expressed U5 snRNAs has recently been observed in *Drosophila* (Chen et al. 2005). Thus, the observed heterogeneity among U1-like and U5 snRNAs might reflect a strategy to accommodate a broad variety of 5'SS sequences and still assure correct splicing. Second, it has been observed that noncanonical 5'SSs are overrepresented among alternatively spliced introns (Johnson et al. 2003). Thus, it seems plausible that the variant U1 snRNAs

that we have identified could play a role in alternative splicing by facilitating the use of noncanonical 5'SS sequences. Finally, we note that the repertoire of variant spliceosomal snRNAs could be exceptionally large in higher eukaryotes due to the presence of multiple copies of snRNA genes in combination with single nucleotide polymorphism. The identification and characterization of this putatively large repertoire of snRNP variants could provide deeper insight into the complex picture of alternative splicing and explain the complexity of vertebrates by expanding the coding capacity of the genome.

MATERIALS AND METHODS

Bioinformatical analysis

The NCBI Build 35 reference sequence of the *H.s.* draft genome (Lander et al. 2001) was searched using the BLAT Web server (<http://genome.ucsc.edu>) (Kent 2002) and the RepeatMasker software *RepeatMasker Open-3.0* (<http://www.repeatmasker.org>). The human *HUMUR1A* gene (DDBJ/EMBL/GenBank accession number K00788) was used as the query. Multiple alignments manually curated using the BioEdit Sequence Alignment Editor v7.0.4.1 (Thompson et al. 1994; Hall 1999). Secondary structure predictions were performed using the mfold v3.1 Web server, <http://www.bioinf.rpi.edu/applications/mfold> (May 2005) (Mathews et al. 1999; Zuker 2003). Splice site coordinates were obtained from the *H.s.* Aug05 release of the AceView database, <http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly> (D. Thierry-Mieg, J. Thierry-Mieg, M. Potdevin, and M. Sienkiewicz, unpubl.). The 5' splice site sequence logo was constructed from 252,159 human introns carrying the canonical GT donor and AG acceptor dinucleotides using WebLogo v2.8 (Crooks et al. 2004). The nucleotide sequences of the human *RNU1A5*, *RNU1A6*, and *RNU1A7* loci, including the coding region and 150 nt upstream and downstream of the coding region, were used to search for evolutionary conservation. Sequence similarity searches were performed using the BLAT Web server against the following genome assemblies available as of March 2006; UCSC release numbers are indicated in parentheses: *Pan troglodytes* (panTro1; The Chimpanzee Sequencing and Analysis Consortium 2005), *Macaca mulatta* (rheMac2), *Canis familiaris* (canFam2) (Lindblad-Toh et al. 2005), *Bos taurus* (bosTau2), *Mus musculus* (mm7), *Rattus norvegicus* (rn3), *Monodelphis domestica* (monDom1), *Gallus gallus* (galGal2), *Xenopus tropicalis* (xenTro1), *Tetraodon nigroviridis* (tetNig1), *Takifugu rubripes* (fr1), *Danio rerio* (danRer3), *Ciona intestinalis* (ci2), *Strongylocentrotus purpuratus* (strPur1), *Anopheles gambiae* (anoGam1), *Apis mellifera* (apiMel2), *Drosophila ananassae* (droAna2), *Drosophila erecta* (droEre1), *Drosophila grimshawi* (droGri1), *Drosophila melanogaster* (dm2), *Drosophila mojavensis* (droMoj2), *Drosophila persimilis* (droPer1), *Drosophila pseudoobscura* (dp3), *Drosophila sechellia* (droSec1), *Drosophila simulans* (droSim1),

Drosophila yakuba (droYak2), *Drosophila virilis* (droVir2), *Caenorhabditis elegans* (ce2), *Caenorhabditis briggsae* (cb1), and *Saccharomyces cerevisiae* (sacCer1). Furthermore, we explored the chained alignment tracks of *C. familiaris*, *B. taurus*, *M. musculus*, *R. norvegicus*, *G. gallus*, *X. tropicalis*, *T. rubripes*, *T. nigroviridis*, and *D. rerio* in the UCSC genome browser, <http://genome.ucsc.edu> (March 2006) (Kent 2002; Karolchik et al. 2003). In addition, using discontinuous Mega BLAST (Zhang et al. 2000), we searched the March 2006 NCBI trace archives of the *Gorilla gorilla*, *Pongo pygmaeus*, *Nomascus leucogenys*, *Papio anubis*, *Callithrix jacchus*, *Lemur catta*, and *Otolemur garnettii* genome projects. Sequence data for *P. pygmaeus* were produced by the Genome Sequencing Center at Washington University School of Medicine in St. Louis (WUGSC), <http://genome.wustl.edu> and the Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC), <http://www.hgsc.bcm.tmc.edu>. Sequence data for *M. mulatta* were produced by BCM-HGSC and the Rhesus Macaque Genome Sequencing Consortium. Sequence data for *C. jacchus* were produced by WUGSC and the NIH Intramural Sequencing Center, <http://www.nisc.nih.gov>. Sequence data for *B. taurus* were produced by BCM-HGSC.

Analysis of snRNA 5'- and 3'-termini

Total RNA from HeLa cells was used to perform 5'-RACE, using the GeneRacer kit (Invitrogen). To determine 3'-ends of snRNAs, total RNA from HeLa cells was ligated to a DNA oligonucleotide carrying one RNA nucleotide at the 5'-end (5'-pU-ATACTCATGGTCATAGCTGTT-3'). cDNA synthesis and PCR amplification (Schuster et al. 2005) were carried out using primer 5'-AACAGCTATGACCATG-3' and a snRNA specific primer. See Supplemental Material (Materials and Methods) for sequences of snRNA specific primers.

Sequence variation analysis

Sequence variation within the snRNA genes was investigated by searching the NCBI SNP database, <http://www.ncbi.nlm.nih.gov/SNP/> (February 2005). PCR products encompassing the U1A6 RNA gene, amplified from human genomic DNA using oligonucleotides 5'-ATGTAGATAGGGGCGCAGTG-3' and 5'-AAAACAGACCGTAACCTAAGAAGAC-3', were cloned into a TA-cloning vector (Invitrogen). Sequences of single clones were determined and analyzed.

Northern blot analysis

Northern blot analysis was performed according to standard procedures (Sambrook et al. 1989). Sequences of oligonucleotides used as probes to detect the candidate snRNA variants were: u1a (U1A snRNA), 5'-CGCGAACGCGAGTCCCCACTACCAAAA-3'; u1a5 (U1A5 snRNA), 5'-GGATAAGCCCAAGGTAGCAAAC-3'; u1a6 (U1A6 snRNA), 5'-CAAACACATAGTAAAAACCTC-3'; u1a7 (U1A7 snRNA), 5'-TCCACAATGCAAGAGACAAACCT-3'; 5s (5S RNA), 5'-TCCAAGTACTAACCAGGCCCGACC-3'; additional probe sequences are listed in the Supplemental Material (Table S1). Oligonucleotide probes were ³²P-5' end-labeled to equal specific activities and the decay rate for each individual probe was determined. Hybridization temperatures were: u1a (54°C), u1a5 (42°C), u1a6 (40°C), u1a7 (43°C), and 5s (50°C).

To determine the relative expression levels of U1A5, U1A6, and U1A7 snRNA in HeLa cells, 20 µg of total RNA were separated on a 10% polyacrylamide gel (triplicates) and subsequently transferred to a Zeta-Probe GT Blotting membrane. The membrane was probed with labeled oligonucleotide probes, and the resulting signals were quantified by autoradiography (ImageQuant, GE Healthcare). 5S RNA was used as a loading control and the expression levels for the snRNA variants relative to the expression of U1A snRNA were determined taking the decay rate of the specific probe into account. To determine the relative expression levels of the snRNA variants in distinct human tissues, total RNA prepared from HeLa cells and a range of human tissues representing distinct developmental stages were separated and transferred as described above. The resulting membranes were subsequently probed with the different oligonucleotide probes and quantification was carried out as described above. The expression level of the snRNA variants relative to U1A snRNA in HeLa cells was used as a reference to compare results from different experiments.

To test the sequence specificity of the U1A snRNA variant probes, increasing amounts (0.005 µg to 4 µg) of in vitro transcribed and gel purified U1A snRNA were fractionated by gel electrophoresis and transferred to a Zeta-Probe GT Blotting membrane. The degree of cross-hybridization of each U1A snRNA variant probe to U1A snRNA relative the U1A snRNA specific probe was analyzed and showed that the U1A snRNA variant probes were specific, i.e., the probes did not detect U1A snRNA unless U1A snRNA was present in >8000-fold excess relative the amount required to be detected by the U1A specific probe.

Immunoprecipitation assay

Antibodies or serum (mAbH20, mAbY12, mIgG, r-R1131, normal rabbit serum [NRS]) were incubated overnight at 4°C with protein G-Sepharose or A-Sepharose beads, as required (Bochnig et al. 1987; Luhrmann et al. 1982). HeLa cell nuclear extract (15 mg/mL), cleared by centrifugation, was added to the preincubated beads/antibody suspension and incubated overnight at 4°C. Beads were washed, treated with proteinase K, and extracted with phenol/chloroform. RNA was recovered by precipitation and subsequently analyzed by Northern blot analysis.

Electrophoretic mobility shift assay (EMSA)

EMSA were performed according to (Gunzl et al. 2002). Nuclear RNA was extracted from HeLa cell nuclear extract using TRIZOL Reagent (Invitrogen).

DATA DEPOSITION

The sequences of the cloned snRNA variants have been deposited in the GenBank database: U1A4 snRNA (contained in AI972570.1), U1A5 snRNA (DQ058355), U1A6 snRNA (DQ058356), and U1A7 snRNA (DQ058357).

SUPPLEMENTAL MATERIAL

Supplemental material consists of Figure S1, Figure S2, Figure S3, Table S1, Table S2, and Supporting Materials and Methods, and is

available at http://www.icm.uu.se/molcell/virtanen/kyriakopoulou_2005/supplementary.php.

ACKNOWLEDGMENTS

We thank G. Akusjärvi, N. Balatsos, N. Henriksson, P. Nilsson, H. Nordvarg, S. Stier, L. Thuresson, and J. Vogel for helpful discussions throughout completing this work, and Dr. R. Lührmann and Dr. G. Akusjärvi for providing antibodies and plasmid pUC-U1. We are grateful to Dr. M. García-Blanco, Dr. H. Schaal, and Dr. D. Schindler for comments on the manuscript. This work was financially supported by the Wallenberg Consortium North, the Swedish Strategic Research Found, and the Swedish Research Council.

Received January 18, 2006; accepted May 22, 2006.

REFERENCES

- Black, D.L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**: 291–336.
- Bochnig, P., Reuter, R., Bringmann, P., and Lührmann, R. 1987. A monoclonal antibody against 2,2,7-trimethylguanosine that reacts with intact, class U, small nuclear ribonucleoproteins as well as with 7-methylguanosine-capped RNAs. *Eur. J. Biochem.* **168**: 461–467.
- Burge, C.B., Tuschl, T., and Sharp, P.A. 1999. Splicing of precursors to mRNAs by the spliceosome. In: *The RNA world* (eds. R.F. Gesteland et al.), pp. 525–560. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Burset, M., Seledtsov, I.A., and Solovyev, V.V. 2000. Analysis of canonical and non-canonical splice sites in mammalian genomes. *Nucleic Acids Res.* **28**: 4364–4375.
- . 2001. SpliceDB: Database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res.* **29**: 255–259.
- Carmel, I., Tal, S., Vig, I., and Ast, G. 2004. Comparative analysis detects dependencies among the 5' splice-site positions. *RNA* **10**: 828–840.
- Chen, L., Lullo, D.J., Ma, E., Celniker, S.E., Rio, D.C., and Doudna, J.A. 2005. Identification and analysis of U5 snRNA variants in *Drosophila*. *RNA* **11**: 1473–1477.
- The Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**: 69–87.
- Ciliberto, G., Dathan, N., Frank, R., Philipson, L., and Mattaj, I.W. 1986. Formation of the 3' end on U snRNAs requires at least three sequence elements. *EMBO J.* **5**: 2931–2937.
- Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. 2004. WebLogo: A sequence logo generator. *Genome Res.* **14**: 1188–1190.
- Freund, M., Hicks, M.J., Konermann, C., Otte, M., Hertel, K.J., and Schaal, H. 2005. Extended base pair complementarity between U1 snRNA and the 5' splice site does not inhibit splicing in higher eukaryotes, but rather increases 5' splice site recognition. *Nucleic Acids Res.* **33**: 5112–5119.
- Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493–521.
- Graveley, B.R. 2001. Alternative splicing: Increasing diversity in the proteomic world. *Trends Genet.* **17**: 100–107.
- Gunzl, A., Palfi, Z., and Bindereif, A. 2002. Analysis of RNA–protein complexes by oligonucleotide-targeted RNase H digestion. *Methods* **26**: 162–169.
- Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**: 95–98.
- Hertel, K.J. and Graveley, B.R. 2005. RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem. Sci.* **30**: 115–118.
- Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A., Delany, M.E., et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.
- Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**: 2141–2144.
- Jurica, M.S. and Moore, M.J. 2003. Pre-mRNA splicing: Awash in a sea of proteins. *Mol. Cell* **12**: 5–14.
- Karolchik, D., Baertsch, R., Diekhans, M., Furey, T.S., Hinrichs, A., Lu, Y.T., Roskin, K.M., Schwartz, M., Sugnet, C.W., Thomas, D.J., et al. 2003. The UCSC genome browser database. *Nucleic Acids Res.* **31**: 51–54.
- Kent, W.J. 2002. BLAT—The BLAST-like alignment tool. *Genome Res.* **12**: 656–664.
- Ketterling, R.P., Drost, J.B., Scaringe, W.A., Liao, D.Z., Liu, J.Z., Kasper, C.K., and Sommer, S.S. 1999. Reported in vivo splice-site mutations in the factor IX gene: Severity of splicing defects and a hypothesis for predicting deleterious splice donor mutations. *Hum. Mutat.* **13**: 221–231.
- Krol, A., Westhof, E., Bach, M., Lührmann, R., Ebel, J.P., and Carbon, P. 1990. Solution structure of human U1 snRNA. Derivation of a possible three-dimensional model. *Nucleic Acids Res.* **18**: 3803–3811.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas 3rd, E.J., Zody, M.C., et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**: 803–819.
- Lopez-Fernandez, L.A. and del Mazo, J. 1996. Characterization of genes expressed early in mouse spermatogenesis, isolated from a subtractive cDNA library. *Mamm. Genome* **7**: 698–700.
- Lührmann, R., Appel, B., Bringmann, P., Rinke, J., Reuter, R., Rothe, S., and Bald, R. 1982. Isolation and characterization of rabbit anti-m3 2,2,7G antibodies. *Nucleic Acids Res.* **10**: 7103–7113.
- Lund, E. 1988. Heterogeneity of human U1 snRNAs. *Nucleic Acids Res.* **16**: 5813–5826.
- Manser, T. and Gesteland, R.F. 1982. Human U1 loci: Genes for human U1 RNA have dramatically similar genomic environments. *Cell* **29**: 257–264.
- Mathews, D.H., Sabina, J., Zuker, M., and Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**: 911–940.
- Matlin, A.J., Clark, F., and Smith, C.W. 2005. Understanding alternative splicing: Towards a cellular code. *Nat. Rev. Mol. Cell Biol.* **6**: 386–398.
- Mattaj, I.W., Dathan, N.A., Parry, H.D., Carbon, P., and Krol, A. 1988. Changing the RNA polymerase specificity of U snRNA gene promoters. *Cell* **55**: 435–442.
- Monstein, H.J., Hammarstrom, K., Westin, G., Zabielski, J., Philipson, L., and Pettersson, U. 1983. Loci for human U1 RNA: Structural and evolutionary implications. *J. Mol. Biol.* **167**: 245–257.
- Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A., and Steitz, J.A. 1983. The U1 small nuclear RNA–protein complex selectively binds a 5' splice site in vitro. *Cell* **33**: 509–518.
- Park, J.W., Parisky, K., Celotto, A.M., Reenan, R.A., and Graveley, B.R. 2004. Identification of alternative splicing regulators

- by RNA interference in *Drosophila*. *Proc. Natl. Acad. Sci.* **101**: 15974–15979.
- Patel, A.A. and Steitz, J.A. 2003. Splicing double: Insights from the second spliceosome. *Nat. Rev. Mol. Cell Biol.* **4**: 960–970.
- Rosbash, M. and Seraphin, B. 1991. Who's on first? The U1 snRNP-5' splice site interaction and splicing. *Trends Biochem. Sci.* **16**: 187–190.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press. Cold Spring Harbor, NY.
- Schuster, J., Betat, H., and Morl, M. 2005. Is yeast on its way to evolving tRNA editing? *EMBO Rep.* **6**: 367–372.
- Sharp, P.A. 2005. The discovery of split genes and RNA splicing. *Trends Biochem. Sci.* **30**: 279–281.
- Sontheimer, E.J. and Steitz, J.A. 1992. Three novel functional variants of human U5 small nuclear RNA. *Mol. Cell. Biol.* **12**: 734–746.
- Stark, H., Dube, P., Luhrmann, R., and Kastner, B. 2001. Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* **409**: 539–542.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Westin, G., Zabielski, J., Hammarstrom, K., Monstein, H.J., Bark, C., and Pettersson, U. 1984. Clustered genes for human U2 RNA. *Proc. Natl. Acad. Sci.* **81**: 3811–3815.
- Will, C.L. and Luhrmann, R. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* **13**: 290–301.
- Wu, M.H., Rajkovic, A., Burns, K.H., Yan, W., Lin, Y.N., and Matzuk, M.M. 2003. Sequence and expression of testis-expressed gene 14 (Tex14): A gene encoding a protein kinase preferentially expressed during spermatogenesis. *Gene Expr. Patterns* **3**: 231–236.
- Yu, Y.-T., Scharl, E., Smith, C.M., and Steitz, J.A. 1999. The growing world of small nuclear ribonucleoproteins. In: *The RNA world* (eds. R.F. Gesteland et al.), pp. 487–524. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. 2000. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**: 203–214.
- Zhuang, Y. and Weiner, A.M. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell* **46**: 827–835.
- . 1990. The conserved dinucleotide AG of the 3' splice site may be recognized twice during in vitro splicing of mammalian mRNA precursors. *Gene* **90**: 263–269.
- Zhuang, Y.A., Goldstein, A.M., and Weiner, A.M. 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc. Natl. Acad. Sci.* **86**: 2752–2756.
- Zuker, M. 2003. Mfold Web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406–3415.

U1-like snRNAs lacking complementarity to canonical 5' splice sites

Christina Kyriakopoulou, Pontus Larsson, Lei Liu, et al.

RNA 2006 12: 1603-1611

References

This article cites 50 articles, 9 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/12/9/1603.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>