References

1 Concha, M.L. and Adams, R.J. (1998) Oriented cell divisions and cellular morphogenesis in the zebrafish gastrula and neurula: a time-lapse analysis. *Development* 125, 983–994
2 Gong, Y. *et al*. (2004) Planar cell polarity signalling controls cell division orientation during zebrafish gastrulation. *Nature* 430, 689–693
3 Keller, R. *et al*. (2000) Mechanisms of convergence and extension by cell intercalation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 355, 897–922
4 Irvine, K.D. and Wieschaus, E. (1994) Cell intercalation during *Drosophila* germband extension and its regulation by pair-rule segmentation genes. *Development* 120, 827–841
5 Wallingford, J.B. *et al*. (2002) Convergent extension: the molecular control of polarized cell movement during embryonic development. *Dev. Cell* 2, 695–706
6 Keller, R. (2002) Shaping the vertebrate body plan by polarized embryonic cell movements. *Science* 298, 1950–1954
7 Tada, M. *et al*. (2002) Non-canonical Wnt signalling and regulation of gastrulation movements. *Semin. Cell Dev. Biol.* 13, 251–260
8 Strutt, D.I. (2002) The asymmetric subcellular localisation of components of the planar polarity pathway. *Semin. Cell Dev. Biol.* 13, 225–231
9 Zallen, J.A. and Wieschaus, E. (2004) Patterned gene expression directs bipolar planar polarity in *Drosophila*. *Dev. Cell* 6, 343–355
10 Ninomiya, H. *et al*. (2004) Antero-posterior tissue polarity links mesoderm convergent extension to axial patterning. *Nature* 430, 364–367
11 Bertet, C. *et al*. (2004) Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* 429, 667–671
12 Fristrom, D. (1988) The cellular basis of epithelial morphogenesis. A review. *Tissue Cell* 20, 645–690
13 Weliky, M. and Oster, G. (1990) The mechanical basis of cell rearrangement I. Epithelial morphogenesis during Fundulus epiboly. *Development* 109, 373–386
14 Knust, E. and Bossinger, O. (2002) Composition and formation of intercellular junctions in epithelial cells. *Science* 298, 1955–1959
15 Sahai, E. and Marshall, C.J. (2002) ROCK and Dia have opposing effects on adherens junctions downstream of Rho. *Nat. Cell Biol.* 4, 408–415
16 Munro, E.M. and Odell, G.M. (2002) Polarized basolateral cell motility underlies invagination and convergent extension of the ascidian notochord. *Development* 129, 13–24

0168-9525/$ - see front matter
doi:10.1016/j.tig.2004.12.002

Genome Analysis

Alternative splicing of conserved exons is frequently species-specific in human and mouse

Qun Pan[1], Malina A. Bakowski[1], Quaid Morris[1,2], Wen Zhang[1], Brendan J. Frey[2], Timothy R. Hughes[1] and Benjamin J. Blencowe[1]

[1]CH Best Institute, University of Toronto, 112 College Street, Toronto, Ontario, Canada M5G 1L6
[2]Department of Electrical and Computer Engineering, University of Toronto, King's College Road, Toronto, Ontario, Canada M5S 34G

In this article, we provide evidence that a frequent source of diversity between mammalian transcripts occurs as a consequence of species-specific alternative splicing (AS) of conserved exons. Using a highly predictive computational method, we estimate that >11% of human and mouse cassette alternative exons undergo skipping in one species but constitutively splicing in the other. These species-specific AS events are predicted to modify conserved domains in proteins more frequently than other classes of AS events. The results thus provide evidence that species-specific AS of conserved exons constitutes an additional potential source of complexity and species-specific differences between mammals.

Alternative splicing (AS), the process by which exons in primary transcripts are spliced in different combinations, is a crucial step in the regulation of many vertebrate genes and is thought to have an important role in the generation of proteomic diversity [1–3]. Recent computational analyses of AS have estimated that one-to-two-thirds of mammalian genes are alternatively spliced [4,5]. These and other studies have also provided evidence that mammalian AS events frequently arise from the evolutionarily rapid loss or gain of exons from genomes (referred to as 'genome-specific AS events') [5–8]. However, in these previous studies, it was not determined to what extent conserved human and mouse exons might also undergo species-specific AS (i.e. skipping in one species and constitutive splicing in another species). Using a highly predictive computational approach, we provide evidence that conserved human and mouse exons frequently undergo species-specific AS in normal cells and tissues. These species-specific AS events target conserved protein domains more often than conserved AS events or AS events involving genome-specific exons.

Computational identification of AS events in human and mouse

We used computational methods initially to identify and analyze different types of AS events, including cassette, alternative 5′–3′ and mutually exclusive exons in 2603 unique human and mouse ortholog pairs (see the supplementary material online for the methods used). These ortholog pairs were selected from an initial starting set of 11 738 on the basis of several criteria including their representation in sequence databases by comparable

Corresponding author: Blencowe, B.J. (b.blencowe@utoronto.ca).
Available online 19 December 2004

numbers of corresponding transcript sequences (supplementary information online). The 2603 ortholog pairs represent a wide range of mammalian gene functions because they contain annotations for 1112 of 1873 Gene Ontology Biological Process (GO-BP) categories (http://www.geneontology.org/), obtained from combined human and mouse annotations (supplementary information online). The AS events were identified by aligning similar numbers of human and mouse expressed sequence tag (EST) and mRNA sequences to contigs of exons that were extracted from the ortholog pairs. By comparing mRNA and EST sequences in our total data sets, we detected a greater percentage of AS genes in human compared with mouse (65.3% versus 55.3%), a result that is consistent with previous analyses [4,5]. We also detected an increased frequency of AS events per gene in human compared with mouse (1.92 versus 1.22).

Frequencies of conserved and species-specific alternative exons in human and mouse

Next, we analyzed the features of human and mouse cassette alternative exons, the most common type of AS. By comparing the genomic exon structures and mRNA and EST sequences of orthologs in our data sets, we estimated the relative proportions of AS events that are conserved between human and mouse versus those that are species-specific. Species-specific AS events were differentiated into two types: (i) those that are represented by conserved exons ('species-specific AS of conserved exons'); and (ii) those that are represented by genome-specific alternative exons ('genome-specific AS'; Figure 1a). We found that 46.4% of human and 46.5% of mouse exons that undergo cassette AS are represented by exons that are conserved in the other species (Figure 1b). Of these conserved exons, 7.4% in human and 12.9% in mouse are detected as alternative exons in the other species ('conserved AS'), whereas AS is detected only in human and only in mouse for the remaining 38.9% and 33.6% conserved exons, respectively.

Because computational detection of AS depends on the level of mRNA and EST coverage, we applied a stringent mRNA and EST-scoring scheme to identify 'high-confidence' cases of species-specific AS of conserved exons (supplementary information online). This scoring scheme takes into account the number of transcripts displaying inclusion and exclusion of a conserved exon in one species, and the relative number of transcripts displaying inclusion only in the other species. For example, the presence of numerous transcripts displaying inclusion and exclusion in one species, and the presence of a comparable total number of transcripts displaying inclusion only in the other species, leads to high confidence score. Using this scoring scheme on the mRNAs and ESTs in our entire data sets, which include sequences from tumor and cell line sources, we estimate that $>11.1\%$ of human and mouse cassette exons represent conserved exons that undergo species-specific AS (Figure 1b). A similar frequency is measured when sampling mRNAs and ESTs sorted from normal cell and tissue sources (Figure 1b), indicating that species-specific AS of conserved exons occurs frequently and under physiologically relevant conditions in mammals.

Validation of computationally detected species-specific AS of conserved exons

To test the accuracy of our computational detection method and estimated frequencies of 'high confidence' species-specific AS of conserved exons, we performed RT–PCR assays on ten examples each of the human and mouse-specific AS events identified by our scoring scheme (Figure 1c). To facilitate unbiased and direct comparisons, the 20 examples were randomly selected from the high confidence set, with the additional criterion that they are represented by human and mouse mRNAs and ESTs from the same adult tissue source (i.e. brain in this case). Accordingly, we performed RT–PCR assays using polyA+ mRNA isolated from human and mouse adult whole brain. These polyA+ mRNA samples were obtained from pooled sources, thereby reducing the possibility of measuring individual-specific variation in AS. Remarkably, species-specific exon skipping was detected in all 20 cases, thereby confirming that our computational method for detection of species-specific AS of conserved exons is accurate (compare adjacent lanes indicated by the horizontal lines below the accession numbers in Figure 2c). Thus, 11.1% appears to represent a reliable lower-bound estimate for the percent of cassette AS events representing conserved exons that undergo species-specific AS in human and mouse.

Functional significance of species-specific AS of conserved exons

To what extent is species-specific AS of conserved exons likely to represent a process that is under evolutionary selection pressure and, therefore, of functional importance, as opposed to representing possible splicing errors? In the vast majority of the examples identified using our scoring scheme, exon skipping was detected in 5–95% of the mRNA and EST sequences sampled, and, in most cases, these examples were represented by sequences from more than one cell or tissue type. Thus, it is unlikely that many of these events arise as a consequence of splicing errors, which are estimated to occur in less than a few percent of total transcripts [9]. This conclusion is also supported by the experimental verification of computationally detected species-specific AS of conserved exons using pooled mRNA samples, which are from sources that are independent of those used for cDNA and EST library construction. Consistent with the RT–PCR results, in which the skipped isoform is generally the less abundant isoform, a survey of the mRNA and EST clusters corresponding to the set of the high confidence species-specific AS events indicates that these AS exons are frequently included (Figure 2a). By contrast, consistent with previous estimates [5], genome-specific alternative exons are more often skipped and conserved alternative exons are included and skipped at comparable frequencies (Figure 2a).

We next asked to what extent species-specific AS of conserved exons targets structural and functional domains in proteins. Notably, in both human and mouse there is a two-to-threefold increase in the frequency at which 'high confidence' species-specific AS of conserved exons is expected to result in alteration (i.e. modification, disruption or creation) of conserved protein domains, compared with the frequency of domain alteration that occurs

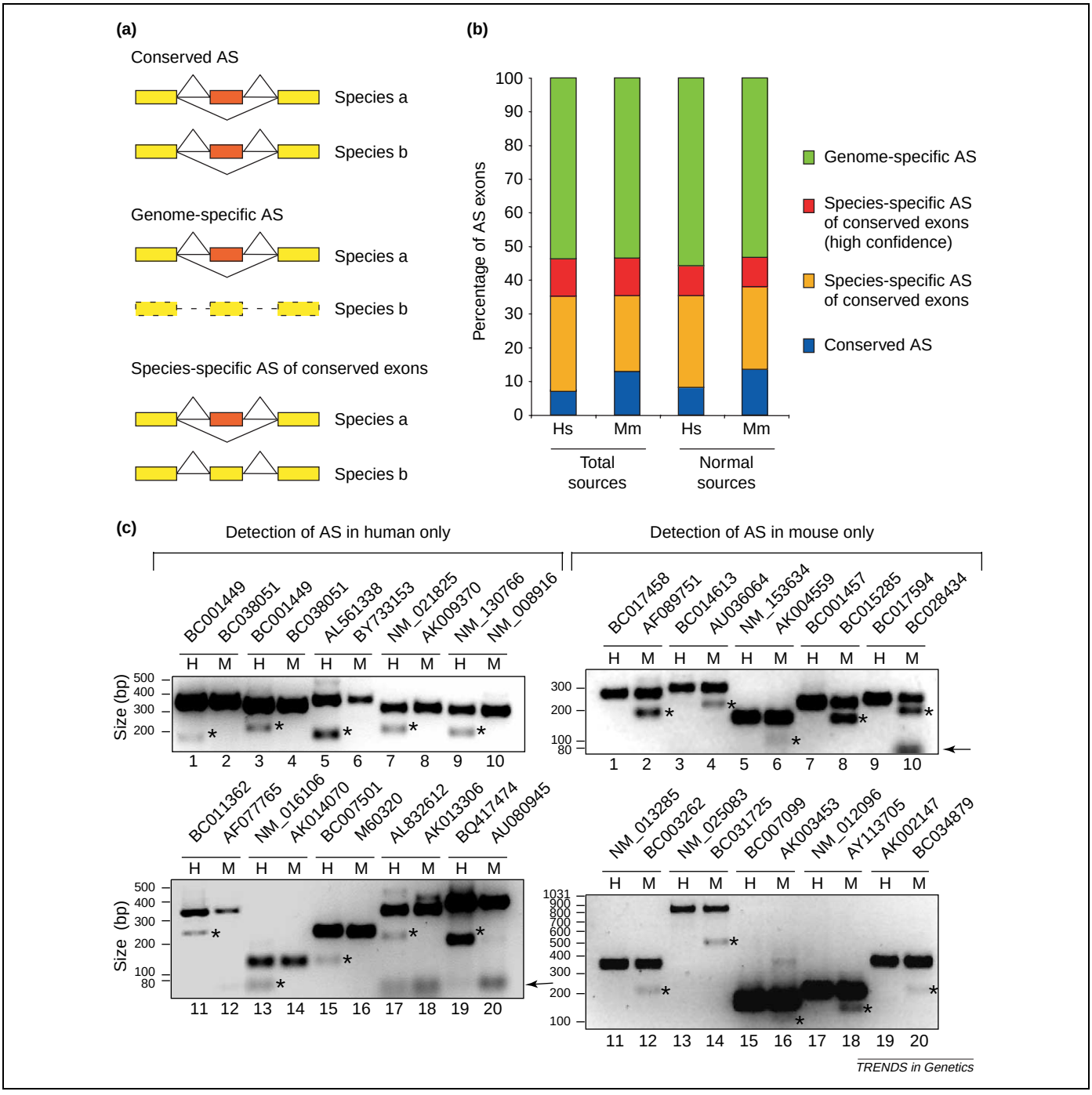

Figure 1. Analysis of alternative splicing (AS) in human and mouse. **(a)** Types of cassette AS identified and analyzed by comparisons of human and mouse mRNA and EST sequences corresponding to 2603 unique orthologous gene pairs. An exon can be detected as alternative in both species ('Conserved AS' top panel), detected only in one of the two species ('Genome-specific AS' middle panel), or an exon can be conserved but detected as alternative only in one species ('Species-specific AS of conserved exons' lower panel) (Box 1). **(b)** The percentages of the total cassette AS exons shown in (a) that were detected using either total mRNA and EST sequences (from all sources including tumor and cell lines), or mRNA and EST sequences from 'normal' cell and tissue sources only. Further information on the procedures used to sort mRNAs and ESTs according to cell and tissue library sources is given in the supplementary material online. **(c)** RT–PCR analysis of 20 examples of computationally detected, high confidence, species-specific AS of conserved exons. RT–PCR assays were performed using polyA+ mRNA isolated from pooled whole adult human (H) and mouse (M) brain tissue, using species-specific primer pairs targeted to the immediate flanking constitutive exon sequences (for more information, see the supplementary material online). Bands indicated by asterisks are consistent with the sizes expected for skipped exon products; approximately half of these were sequenced and in every case their identity was confirmed. The arrows on the bottom-right of two of the gel panels indicate non-specific primer amplification products that were also detected in the absence of mRNA (not shown). The RefSeq ID or accession number representing the 'included' isoform for each alternatively spliced gene is shown. Horizontal lines below the accession numbers indicate pairs of human and mouse orthologs.

following AS of conserved or genome-specific alternative exons (Figure 2b). Consistently, a smaller (twofold) increase in the frequency of domain disruption is observed for species-specific AS of conserved exons that do not fall within the 'high confidence' set (Figure 2b). This is expected given the probable higher false-positive detection rate for species-specific AS among this group (i.e. it is more likely that a larger number of these are conserved AS exons, which alter protein domains at a reduced rate). Domain alteration by species-specific AS of conserved

Box 1. The types of cassette alternative splicing (AS) events and their distinguishing features

Conserved AS

The same AS event is detected in transcripts in more than one species (in the present study human and mouse were compared). These AS events are detected at a significantly lower frequency than species-specific AS events and do not appear to show a preferential trend towards either high or low inclusion levels across mammalian tissues.

Genome-specific AS

This type of AS event is represented by exons that are detected in only one of two or more species being compared. Genomic-specific AS exons appear to represent the most common type of cassette AS events in mammalian species, and these exons are on average frequently skipped in mammalian tissues (for more details, see Ref. [5]).

Species-specific AS of conserved exons

This class of AS event is represented by exons that undergo AS in one species but that are constitutively included in another species. We provide the first estimate of the frequency of this type of AS in human and mouse, using an experimentally validated scoring scheme (see Figure 1c and supplementary material online). Our conservative estimate is that at least 11% of cassette alternative exons belong to this class of AS event. Conserved exons that undergo species-specific AS are on average frequently included across human and mouse tissues, however, they appear to target conserved protein sequences more often than the other types of AS events listed above (see Figure 2b).

exons was also found to be significantly more frequent than the hypothetical rate (17.8% in human and 16.3% in mouse) of domain alteration measured by simulated skipping of constant exons (data not shown). There is a close correlation between the frequencies at which the human and mouse AS events alter protein domains, indicating that these measured frequencies are reliable and significant. Thus, although species-specific AS of conserved exons does not appear to often result in pronounced changes in exon inclusion levels, it is expected to often result in the regulation of activities that are attributed to conserved protein domains. These AS events could therefore provide an important contribution to the evolution of novel selected characteristics, which ultimately define fundamental differences between species. In this context, it is noteworthy that the domains that are affected by species-specific AS of conserved exons are associated with genes in diverse GO-BP categories (Table 4 in the supplementary material online).

Concluding remarks

We have provided evidence that at least 11% of human and mouse cassette AS events represent conserved exons that undergo species-specific AS. These species-specific AS events have the potential to modulate frequently the structural and functional properties of proteins that are attributed to conserved domains. Therefore, they could have an important role in the evolutionary differences between mammalian species. Future experimental studies will be required to determine the precise contribution of these AS events to proteomic diversity and to the fundamental differences that exist between mammalian species.

Update

While this article was in preparation, the authors confirmed experimentally the differences in global inclusion-level properties of conserved and species-specific alternative exons shown in Figure 2a, using a new quantitative AS microarray profiling system [10].

Acknowledgements

We thank Jim Friesen, Alan Cochrane and members of the Blencowe Laboratory for their helpful comments on this article. This work was supported by an operating grant from the CIHR to B.J.B.

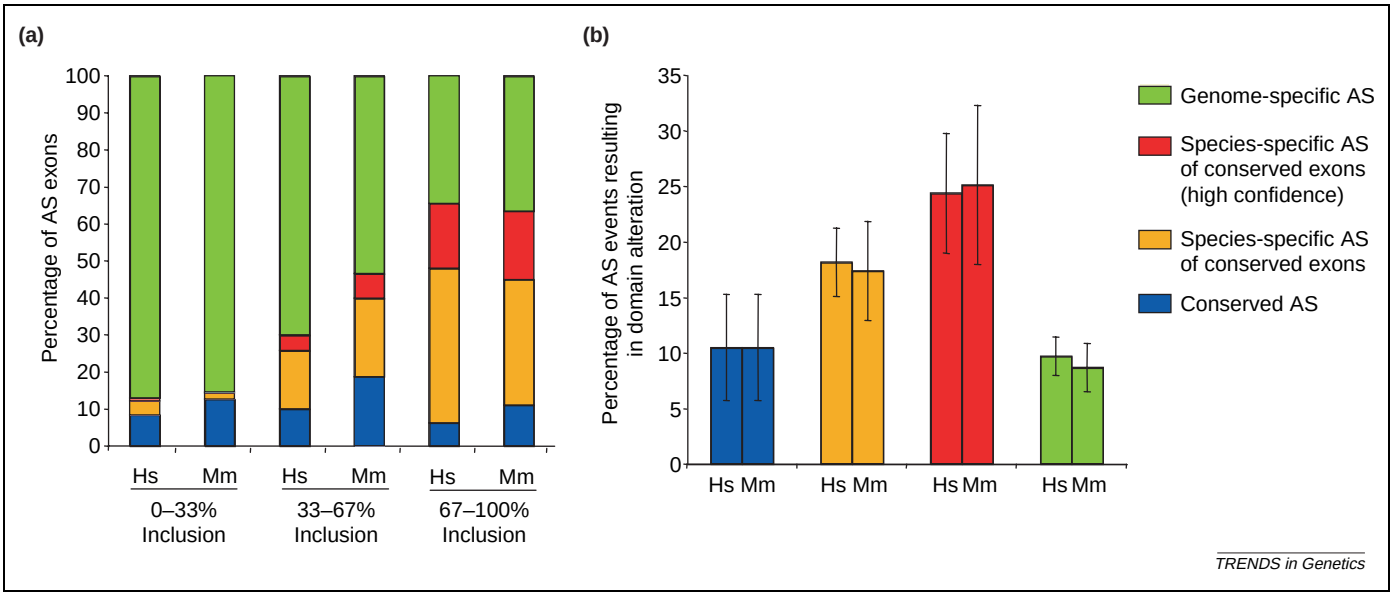

Figure 2. The proportion of different cassette AS exons in human and mouse transcripts and predicted effect on conserved protein domains. **(a)** Relative levels of inclusion estimated for cassette AS exons belonging to each of the classes shown in Figure 1. Inclusion levels were estimated by the percentage of supporting mRNAs and ESTs that show inclusion versus skipping of the alternative exon. Similar results were obtained from comparing mRNA and EST sequences from total (shown) and normal cell and tissue sources (not shown). **(b)** The percentage of AS events predicted to result in alteration (i.e. modification, disruption or creation) of conserved protein domains. Measurements were determined for the AS events identified using mRNA and EST sequences from total cell and tissue sources, and similar results were obtained using mRNA and EST sequences that were sorted from normal cell and tissue sources only (data not shown). Error bars represent two standard errors of the measured proportion. Abbreviations: Hs, human; Mm, mouse.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2004.12.004

References

1 Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107
2 Caceres, J.F. and Kornblihtt, A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* 18, 186–193
3 Maniatis, T. and Tasic, B. (2002) Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* 418, 236–243
4 Brett, D. *et al*. (2002) Alternative splicing and genome complexity. *Nat. Genet.* 30, 29–30
5 Modrek, B. and Lee, C.J. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet.* 34, 177–180
6 Thanaraj, T.A. *et al*. (2003) Conservation of human alternative splice events in mouse. *Nucleic Acids Res.* 31, 2544–2552
7 Nurtdinov, R.N. *et al*. (2003) Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum. Mol. Genet.* 12, 1313–1320
8 Sorek, R. *et al*. (2004) How prevalent is functional alternative splicing in the human genome? *Trends Genet.* 20, 68–71
9 Kan, Z. *et al*. (2002) Selecting for functional alternative splices in ESTs. *Genome Res.* 12, 1837–1845
10 Pan, Q. *et al*. Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* (in press)

doi:10.1016/j.tig.2004.12.004

Is abundant A-to-I RNA editing primate-specific?

Eli Eisenberg[1,2], Sergey Nemzer[1], Yaron Kinar[1], Rotem Sorek[1], Gideon Rechavi[3] and Erez Y. Levanon[1,3]

[1]Compugen Ltd, 72 Pinchas Rosen St, Tel-Aviv 69512, Israel
[2]School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel
[3]Department of Pediatric Hematology-Oncology, Chaim Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv 52621, Israel

A-to-I RNA editing is common in all eukaryotes, and is associated with various neurological functions. Recently, A-to-I editing was found to occur frequently in the human transcriptome. In this article, we show that the frequency of A-to-I editing in humans is at least an order of magnitude higher than in the mouse, rat, chicken or fly genomes. The extraordinary frequency of RNA editing in human is explained by the dominance of the primate-specific *Alu* element in the human transcriptome, which increases the number of double-stranded RNA substrates.

Introduction

A-to-I RNA editing is the site-specific modification of adenosine to inosine in stem-loop structures within precursor mRNAs, catalyzed by members of the double-stranded-RNA (dsRNA) specific ADAR (adenosine deaminase acting on RNA) family [1]. ADAR-mediated RNA editing is essential for the normal development of both invertebrates and vertebrates [2–5]. Altered editing patterns are associated with inflammation [6], epilepsy [7], depression [8], amyotrophic lateral sclerosis (ALS) [9] and malignant gliomas [10]. In a few known examples, editing changes an amino acid in the translated protein, resulting in a change in its function. However, it was suggested that this might not be the primary role of editing by ADARs [4] because most documented editing events occur within intronic and untranslated regions (UTRs) [11]. These editing events can affect splicing, RNA localization, RNA stability and translation [12], however, a full understanding of the purpose of editing in these regions is still elusive.

Using a bioinformatics approach to search for potential stem-loop structures in transcripts combined with differences between EST and genomic sequences, we have recently reported the identification of abundant A-to-I editing in human, affecting >1600 different genes [13]. Most of these editing sites reside in *Alu* elements within UTR regions [13,14]. *Alu* elements are short interspersed elements (SINEs), typically 300 nucleotides long, which comprise >10% of the human genome. Despite being considered genetically functionless, *Alu* elements were suggested to have broad evolutionary impacts [15,16]. They are found in all primates but in no other organism [17,18]. Therefore, they were suggested to have a role in primate evolution [19,20]. However, the nature of this role is still under debate. The question thus arises whether the abundance of A-to-I editing sites in humans is related to some special characteristics of the *Alu* repeat, and therefore unique to primates, or whether similar editing

Corresponding author: Eisenberg, E. (elie@compugen.co.il).
Available online 16 December 2004