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)

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

patterns could also be observed in other organisms with a different, yet similar, composition of SINEs.

Comparative search for RNA-editing sites

In this study, we searched for A-to-I editing sites in human, mouse (*Mus musculus*), rat (*Rattus norvegicus*), chicken (*Gallus gallus)* and fly (*Drosophila melanogaster*). We have found that the frequency of predicted A-to-I RNA editing in human is at least an order of magnitude greater than in other organisms. For this purpose, we used the algorithm described by Levanon *et al.* [13]. Briefly, because ADARs bind to double-stranded RNAs (dsRNAs), the algorithm searches for potential dsRNA structures within the genomic sequence of each gene, and enumerates all deviations of the expressed sequence from the genomic sequence in these dsRNA regions. The sequencing reaction (and the ribosome) recognizes inosine as guanosine (G). Therefore, the fingerprints of ADAR editing are genomically encoded adenosines that are read as guanosines in the RNA sequence. A strict cleaning procedure is used to remove sequencing artifacts (noisy expressed sequence) and known single nucleotide polymorphisms (SNPs), resulting in a clean set of A-to-G mismatches, which, with high confidence, confirms A-to-I editing (for more details, see Ref. [13]).

Application of this algorithm to the human transcriptome (comprising more than five million ESTs and RNAs) has yielded 12 723 editing sites (with an estimated accuracy of >95%) in 1637 different genes [13]. A subset of the results has been experimentally validated, resulting in the observation of editing in 26 novel substrates and confirming the computationally estimated accuracy. We applied this algorithm to the mouse transcriptome (comprising more than four million ESTs and mRNAs). The mouse is similar to human in terms of the quality of the genomic sequence, genome size, number of genes and the amounts of expressed sequences, making a comparison applicable. Using the same algorithm that we used for the human genome, we found only 302 A-to-I editing sites (estimated accuracy 90%) in 87 different mouse genes. The detected level of editing in mouse is thus 40-times lower than that in human.

In addition, we have conducted another independent search for RNA editing that does not use the above-mentioned algorithm, and is based on a different sequence-alignment method. We used the UCSC alignments of human and mouse RNA sequences to their genomes (http://genome.ucsc.edu) [21] and recorded all of the mismatches along them. We scanned 128 068 human RNA sequences (259Mb in length) and 102 895 mouse RNA sequences (198Mb in length). The distribution of mismatches in these sequences is presented in Figure 1a and Table 1. Even a simple count of all mismatches exhibits a vast overrepresentation of A-to-G mismatches in human sequences, suggesting that there are ∼50 000 inosines in these sequences (∼one per 5200bp). However, the number of A-to-G mismatches in excess of the noise background level in mouse RNAs is only ∼3000 (∼one per 66 000bp), reflecting a 17-fold increase of inosines in human compared with mouse. These results point to a striking difference in editing patterns between human and mouse.

A-to-I editing sites often occur in clusters, an edited sequence typically being edited in many close-by sites [11]. Therefore, we searched for sequences that exhibit three or more consecutive identical mismatches (Table 2). A-to-G consecutive mismatches are vastly overrepresented compared with other types of consecutive mismatches in human but not in mouse (Figure 1b,c). We found a set of 4864 human RNA sequences with three or more consecutive A-to-G mismatches, with an estimated accuracy of 80%, suggesting ∼4000 RNA sequences are actually edited at multiple sites (3% of the total number of RNA sequences). By contrast, the same analysis applied to mouse RNA sequences yielded an estimate of only ∼220 (0.2%) multi-edited mouse RNA sequences (Table 1). Here too, the number of edited sequences found in human is 20-fold higher than in the mouse. We thus conclude that A-to-I editing levels in human are at least an order of magnitude higher than in mouse. (Note that after the submission of this article, similar results were published by Kim *et al.* [14].)

To check whether the detected differences are primate- or rodent-specific, we repeated the two analyses (single and multi mismatch counts in RNAs relative to genome) on the genomes of rat (10 999 RNA sequences), chicken (19 218 RNA sequences) and fly (14 632 RNA sequences). These genomes showed editing patterns similar to the mouse genome (Figure 1), suggesting that the differences seen between human and mouse stem from unique primate- (or human-) specific factors.

Editing levels vary between different tissues [13,22]. Thus, differences between the tissue distributions of available human and mouse RNA sequences could lead to a bias in our results. To rule out this possibility, we repeated the human–mouse comparison for RNA sequences of the same homogeneous tissue origin. We used three different tissues that have a significant and similar number of sequences for both organisms: brain, thymus and testis. We found that the level of editing in human is significant in all three tissues (at least 3% of sequences are edited), whereas in mouse the editing is undetectable for such small data sets (Tables 1–3 in the online supplementary material). Notably, the editing level in RNAs that originate from the human thymus is exceptionally high: ∼17% of sequences (1bp per 1000bp) are apparently edited. Taken together, our results strongly suggest that A-to-I editing patterns differ significantly between human and other organisms.

RNA editing and the primate-specific *Alu* element

The vast majority of A-to-I editing detected by our algorithm in human occurs within *Alu* elements, which are the most abundant SINEs in primates and are frequently contained within part of transcripts. The total number of SINEs in the human and rodent genome is similar [23]. Nevertheless, forming dsRNA from two consecutive and oppositely orientated SINEs in human is more probable than in mouse because only one SINE is dominant in human. Furthermore, the dsRNAs formed in human are longer (thus contain more adenosines to be
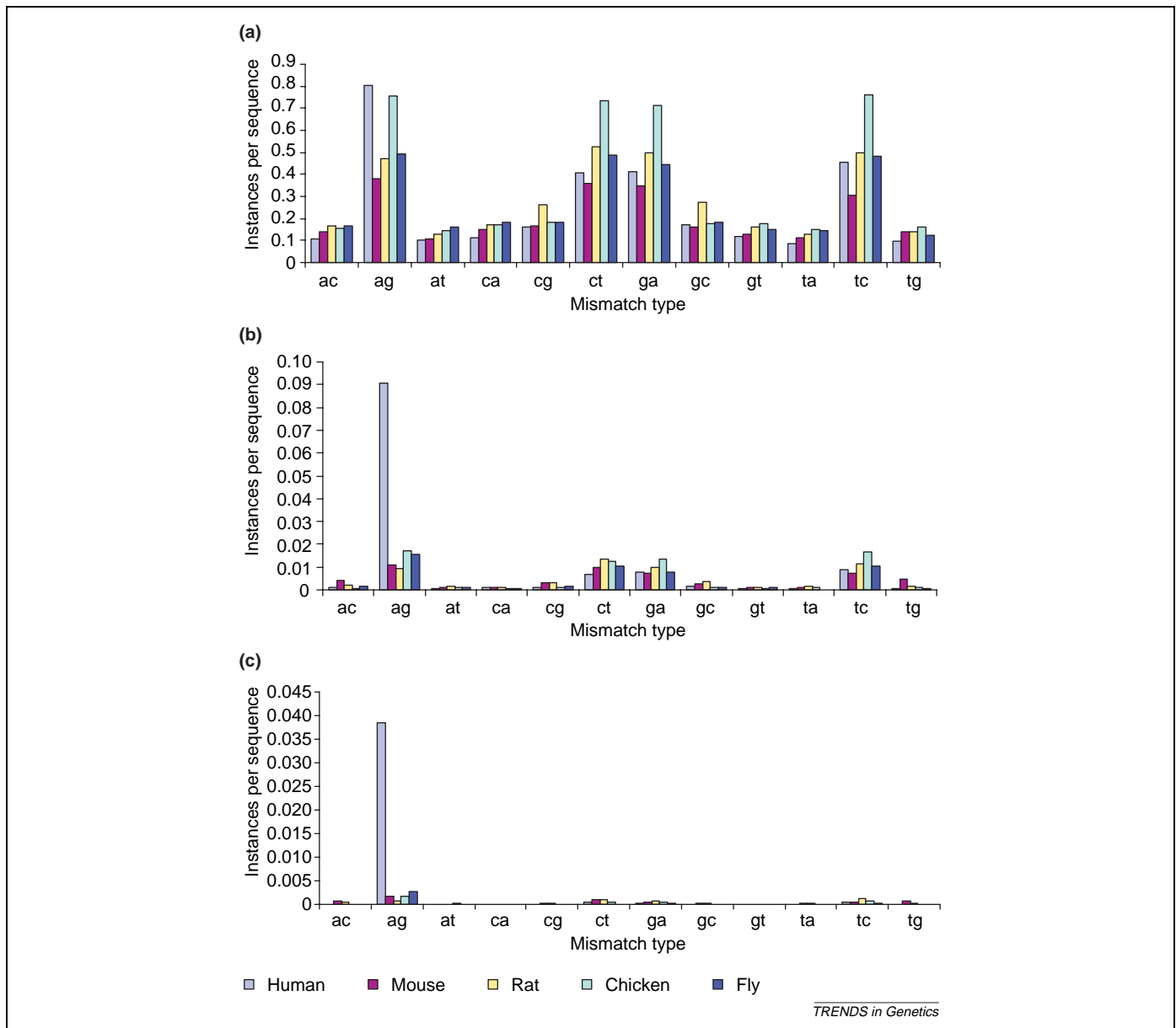
Figure 1. Mismatches distributions: a multi-species comparison. **(a)** Single mismatches; **(b)** instances of consecutive three identical mismatchs per sequence; and **(c)** instances of consecutive five identical mismatchs per sequence. A comparison of the average number of mismatches per sequence in human and mouse RNA sequences shows a significant overabundance of A-to-G mismatches, most probably representing A-to-I editing, in human but not in the other four organisms that were examined. The distribution of other types of mismatches is similar in all of the organisms examined. Note that the background mismatch rate in chicken RNAs is significantly higher, presumably as a result of the lower quality of its current genome draft. *X*-axis: types of mismatch. The label xy refers to x in the DNA sequence and y in the expressed sequence [e.g. ac refers to genomic A(s) that read as C(s) in the expressed data]. *Y*-axis: the average number of instances of consecutive one, three or five mismatches per sequence.

Table 1. Mismatches in genome-aligned RNAs[a]

Mismatch	Human			Mouse		
	Number of consecutive mismatchs			**Number of consecutive mismatchs**		
	1	3	5	1	3	5
A-to-G	102 832 (43 965)	11 613 (4864)	4926 (2513)	38 910 (22 010)	1112 (901)	181 (127)
G-to-A	52 488 (34 180)	968 (914)	48 (44)	35 876 (20 226)	748 (681)	64 (57)
C-to-T	52 195 (34 449)	853 (774)	74 (70)	36 750 (20 116)	1033 (915)	98 (92)
T-to-C	58 083 (37 854)	1115 (1045)	71 (57)	31 630 (17 207)	742 (681)	49 (46)
Percentage of A-to-G mismatches	26.5%	75.4%	95.0%	15.1%	20.2%	28.0%

[a]The number of single (or stretches of consecutive) mismatches, for the most common mismatches. The numbers in parentheses are the number of distinct RNA sequences in which the corresponding single (or stretches or consecutive) mismatches occur. The last row shows the percentage of A-to-G mismatches among the total number of all 12 possible mismatches.

Table 2. Five consecutive editing sites in the 3′ UTR of the human *ASAM* gene[a,b]

Sequence	Genome
GAGGTTGCAGTGAGCCAAGATTATGCCATTGCACTCCAGCCTGGGTGACAAAGCAAGACTCCATCTC	Genome
GAGGTTGCAGTGAGCCGAGATTATGCCATTGCACTCCAGCCTGGGTGACAGAGCGGGACTCCGTCTC	AY358340
GAGGTTGCAGTGAGCCGGGATTATGCCATTGCACTCCGGCCTGGGTGACAGAGCGAGACTCCATCTC	AI001922
GAGGTTGCAGTGAGCCGAGATTATGCCATTGCACTCCAGCCTGGGTGACAAAGCAAGACTCCATCTC	BU740253
GAGGTTGCAGTGAGCCAGGATTGTGCCATTGCACTCCAGCCTGGGTGACAGAGCAGGACTCCATCTC	AW614786
GAGGTTGCAGTGAGCCAAGATTATGCCATTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCGTCTC	BM680006

[a]Abbreviations: *ASAM*, adipocyte-specific adhesion molecule; UTR, untranslated region.
[b]The RNA sequence AY358340 was compared with the genomic sequence (shown in red). ESTs supporting this editing site are presented. The A-to-G mismatches are highlighted. There are 2513 human mRNAs with five consecutive A-to-G mismatches (and only 44 with five or more consecutive G-to-A mismatches), compared with only 127 such mouse mRNAs (57 with G-to-A mismatches).

edited) because *Alu* is longer than the equivalent rodent B1 (online supplementary material). An additional possibility is that the *Alu* repeat could be preferentially targeted by ADARs [13]. We thus suggest that the introduction of *Alu* elements into the ancestral primate genome is responsible for the large differences in editing patterns between human and other genomes. In addition, the observed difference in editing activity in humans could be related to a human-specific splice variant of ADAR2 (also known as ADARB1, which encodes an adenosine deaminase) [24,25]. Intriguingly, this splice variant, which accounts for 40% of all ADAR2 transcripts in humans, was created as a result of an *Alu*-derived exon in intron seven of ADAR2 [16,24].

Possible implications of abundant RNA editing in human

The vast majority of RNA editing events occurs within non-coding regions of the mRNA. The role of inosines within these regions is still a mystery. It is possible that the editing of these non-coding regions is meaningless. However, it was already suggested that RNA editing could regulate the triggering of RNA interference (RNAi) and cause or prevent the degradation of the RNA by stabilizing or destabilizing dsRNA stems [26]. Thus, editing in *Alu* sequences within UTRs might add a powerful mechanism to regulate RNA turnover in primates. In addition, RNA editing was suggested to affect RNA stability [12], localization [27] and translation rate [12]. Moreover, editing at the vicinity of a splice site can affect the splicing pattern. In particular, editing can create a new splice site, thus enabling the introduction of new exons [28]. Although the primary role of editing in the UTRs has yet to be revealed, it is already clear that its widespread occurrence in the human transciptome provides evolution with an additional, post-transcriptional means that enable the fine-tuning of gene expression at the cellular and organism level.

Concluding remarks

In this article, we have shown that abundant editing is unique to primates and results from the properties of the *Alu* element repeat. This finding, accompanied by the observations that A-to-I editing is abundant in brain tissues [13,22] and aberrant in a number of neurological disorders [2,7–10], makes it tempting to speculate that widespread editing, as a result of the introduction of *Alu* elements, might have had a role in the evolution of primates.

Update

After this article was accepted, two additional independent studies reporting abundant A-to-I editing within *Alu* elements in human were published [29,30].

References

1　Polson, A.G. *et al*. (1991) The mechanism of adenosine to inosine conversion by the double-stranded RNA unwinding/modifying activity: a high-performance liquid chromatography-mass spectrometry analysis. *Biochemistry* 30, 11507–11514
2　Palladino, M.J. *et al*. (2000) A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity. *Cell* 102, 437–449
3　Higuchi, M. *et al*. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* 406, 78–81
4　Wang, Q. *et al*. (2004) Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *J. Biol. Chem.* 279, 4952–4961
5　Hartner, J.C. *et al*. (2004) Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *J. Biol. Chem.* 279, 4894–4902
6　Patterson, J.B. and Samuel, C.E. (1995) Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol. Cell. Biol.* 15, 5376–5388
7　Brusa, R. *et al*. (1995) Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science* 270, 1677–1680
8　Gurevich, I. *et al*. (2002) Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. *Neuron* 34, 349–356
9　Kawahara, Y. *et al*. (2004) Glutamate receptors: RNA editing and death of motor neurons. *Nature* 427, 801
10　Maas, S. *et al*. (2001) Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. *Proc. Natl. Acad. Sci. U. S. A.* 98, 14687–14692
11　Morse, D.P. *et al*. (2002) RNA hairpins in noncoding regions of human brain and *Caenorhabditis elegans* mRNA are edited by adenosine deaminases that act on RNA. *Proc Natl Acad Sci U S A* 99, 7906–7911
12　Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71, 817–846
13　Levanon, E.Y. *et al*. (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* 22, 1001–1005

14 Kim, D.D. *et al*. (2004) Widespread RNA editing of embedded *Alu* elements in the human transcriptome. *Genome Res.* 14, 1719–1725

15 Szmulewicz, M.N. *et al*. (1998) Effects of *Alu* insertions on gene function. *Electrophoresis* 19, 1260–1264

16 Lev-Maor, G. *et al*. (2003) The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. *Science* 300, 1288–1291

17 Schmid, C.W. (1996) *Alu*: structure, origin, evolution, significance and function of one-tenth of human DNA. *Prog. Nucleic Acid Res. Mol. Biol.* 53, 283–319

18 Kapitonov, V. and Jurka, J. (1996) The age of *Alu* subfamilies. *J. Mol. Evol.* 42, 59–65

19 Deininger, P.L. *et al*. (2003) Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* 13, 651–658

20 Sorek, R. *et al*. (2002) Alu-containing exons are alternatively spliced. *Genome Res.* 12, 1060–1067

21 Karolchik, D. *et al*. (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.* 31, 51–54

22 Paul, M.S. and Bass, B.L. (1998) Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* 17, 1120–1127

23 Gibbs, R.A. *et al*. (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428, 493–521

24 Gerber, A. *et al*. (1997) Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an *Alu* cassette. *RNA* 3, 453–463

25 Lai, F. *et al*. (1997) Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Mol. Cell. Biol.* 17, 2413–2424

26 Tonkin, L.A. and Bass, B.L. (2003) Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science* 302, 1725

27 Zhang, Z. and Carmichael, G.G. (2001) The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* 106, 465–475

28 Rueter, S.M. *et al*. (1999) Regulation of alternative splicing by RNA editing. *Nature* 399, 75–80

29 Athanasiadis, A. *et al*. (2004) Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome. *PLoS Biol* 2, e391. DOI: 10.1371/journal.pbio.0020391

30 Blow, M. *et al*. (2004) A survey of RNA editing in human brain. *Genome Res.* 14, 2379–2387