Minireview

# SINEs point to abundant editing in the human genome

Joshua DeCerbo and Gordon G Carmichael

Address: Department of Genetics and Developmental Biology, University of Connecticut Health Center, Farmington, CT 06030-3301, USA.

Correspondence: Gordon G Carmichael. E-mail: carmichael@nso2.uchc.edu

Published: 31 March 2005

*Genome* ***Biology*** 2005, **6:**216 (doi:10.1186/gb-2005-6-4-216)

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2005/6/4/216



## Abstract

Recent bioinformatic analyses suggest that almost all human transcripts are edited by adenosine deaminases (ADARs), converting adenosines to inosines. Most of this editing is in *Alu* element transcripts, which are unique to primates. This editing might have no function or might be involved in functions such as the regulation of splicing, chromatin or nuclear localization of transcripts.

## Editing of double-stranded RNAs

Many double-stranded RNAs (dsRNAs) in cells, especially those in the nucleus, are susceptible to base editing in which adenosines are deaminated to inosines by enzymes known as dsRNA-specific adenosine deaminases (ADARs) [1]. This editing leads to a recoding of the genetic information, because inosines are translated as if they were guanosines. Thus, RNA editing can have dramatic consequences for the expression of genetic information, and in a number of cases it has been shown to lead to the expression of proteins not only with altered amino-acid sequences from those predicted from the DNA sequence, but also with altered biological functions [1,2].

It seems there are two types of RNA editing, selective and promiscuous. Selective editing (Figure 1a) results in the conversion of one or a few adenosines in a transcript to inosines; it is generally associated with the expression of proteins with altered functions. These editing events usually occur within relatively short and incompletely base-paired sequences that form between the edited exon and a nearby intron, and they are directed to specific adenosine residues (for example, see [2]). Promiscuous editing, on the other hand, involves the deamination of numerous adenosines in RNA duplexes that are generally longer than 30 base-pairs (bp; Figure 1b) [1]. This type of editing is thought to be the result of aberrant production of dsRNA and has been suggested to lead to RNA degradation [3], nuclear retention [4] or even gene silencing [5].

In the past several years, interest in the prevalence of editing in the human genome and in the identity of endogenous editing substrates has grown. Recent work using computational approaches has provided intriguing and unexpected results. Independently and almost simultaneously, four groups have made remarkably similar and provocative observations [6-9]. Many thousands of sites of mRNA editing have now been revealed in more than 1,600 human genes. But a remarkable additional finding has emerged: in each of these studies, a very high proportion of the editing sites discovered (90% or more) are found in a single class of repetitive sequences called *Alu* elements, which generally lie within noncoding segments of transcripts, such as introns and 5′ and 3′ untranslated regions.

## What are *Alu* elements?

Of the 3 billion bp of the haploid human genome, only 3-5% encode proteins, but as much as 45% of the genome is composed of repetitive and transposable elements [10]. One of the most abundant and important of these classes is the short interspersed nuclear repetitive DNA elements, SINEs. Almost all of the human SINEs belong to a single family and are known as *Alu* elements. There are up to 1.4 x $10^6$ copies of these 300 bp elements in the genome, corresponding to more than one *Alu* element for about every 3,000 bp of genomic DNA. As these elements are not randomly distributed throughout the genome but rather are biased toward gene-rich regions [11], the conclusion can be drawn that the

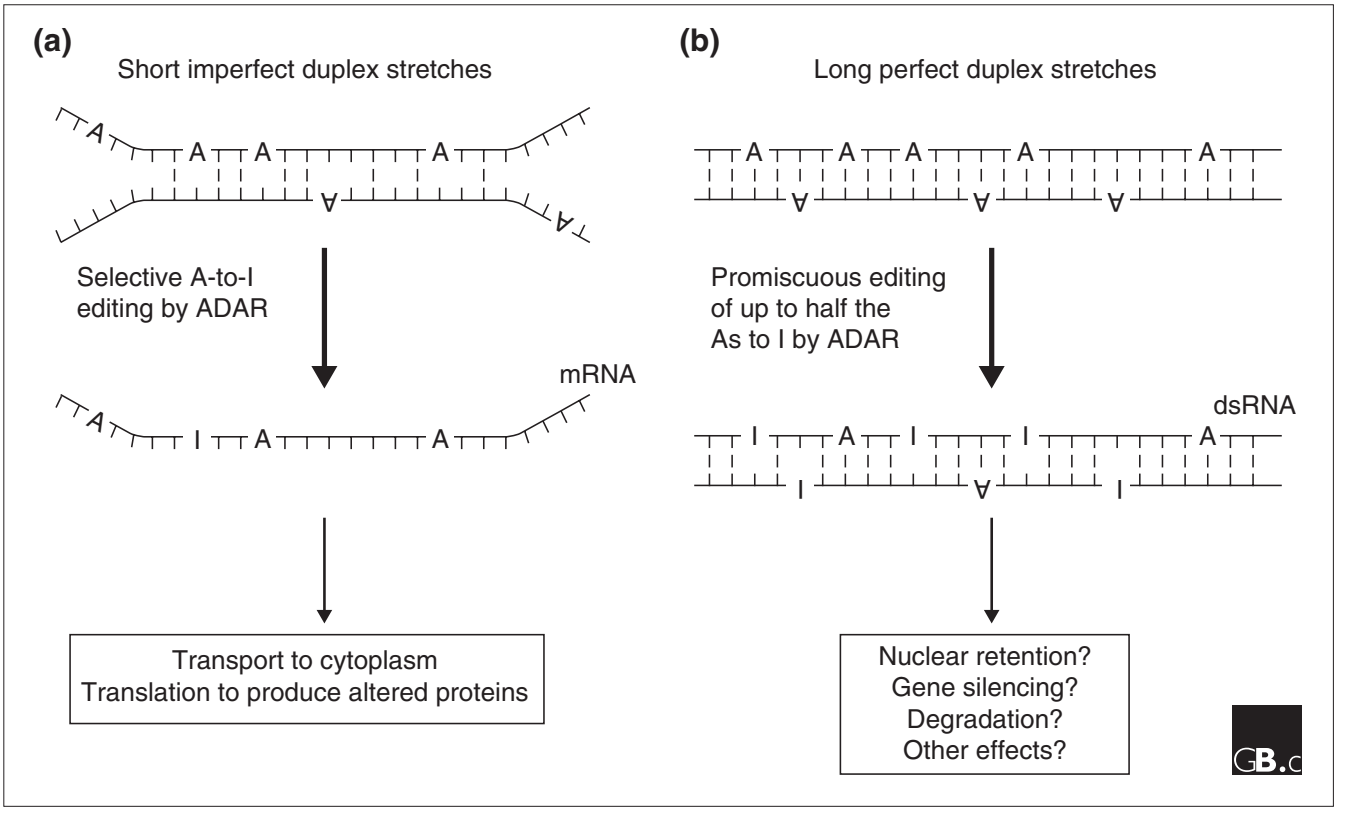


**Figure 1**
Double-stranded RNAs can be edited by ADARs by **(a)** selective or **(b)** promiscuous editing. **(a)** Short, imperfect dsRNA duplexes can be edited selectively at precise locations, which are determined by both the sequence and structure of the RNAs. When this occurs in mRNAs, the inosines (I) are translated as guanosines, thus generating proteins with altered amino-acid sequences. **(b)** Long perfect duplexes (over 30 bp) can be promiscuously edited, with up to half of the adenosines (A) on each strand being deaminated to I in an almost random fashion. These edited RNAs are not destined for translation in the cytoplasm; editing may lead to a number of distinct consequences.

average human pre-mRNA molecule might contain the surprisingly high number of more than 16 *Alu* elements (see Figure 2a). *Alu* elements are conserved along their sequence and do not encode any protein. They can act as insertional mutagens, but the vast majority appear to be genetically inert. Although many *Alu*s are almost identical to one another, others have diverged somewhat over time into distinct evolutionary lineages [12].

The data reported by Athanasiadis *et al.* [9] serve to illustrate many of the key recent findings on *Alu* elements. By comparing cDNA sequences with genomic sequences and searching for clusters of A-to-G changes as indicators of editing, 1,445 human mRNAs were identified that might be edited, and for several of them this was confirmed experimentally. The vast majority of the editing is located within *Alu* elements. Importantly, however, each edited *Alu* has an oppositely oriented partner nearby, which also appears to be edited (Figure 2b). The authors [9] went one step further in this analysis - instead of examining existing cDNA sequences for evidence of editing, they asked whether the existence of oppositely oriented *Alu* elements in a gene actually predicts that editing will be observed. Strikingly, this appears to be the case. Thus, there may remain many editing events that are not yet represented in existing cDNA datasets.

*Alu* elements can insert into the genome in either orientation relative to gene transcription. Given the abundance and uniformity in sequence of *Alu* elements, Athanasiadis *et al.* [9] argue that about 90% of human genes in fact contain *Alu* sequences that can form intramolecular dsRNA structures that are subject to ADAR editing. Thus, in the past year we have progressed from thinking of ADAR editing as affecting only a small subset of human genes to now having to accept that it may affect almost all of them! This situation stands in stark contrast to that found for non-primate mammals. A similar computational analysis of mouse mRNAs revealed no widespread editing [13]. Rodent genomes have a density of SINEs similar to that of primates, but because in these mammals there are numerous distinct families of SINEs [14], the potential for significant intramolecular base pairing in pre-mRNA molecules is far lower than it is in humans, and so also is the potential for editing.

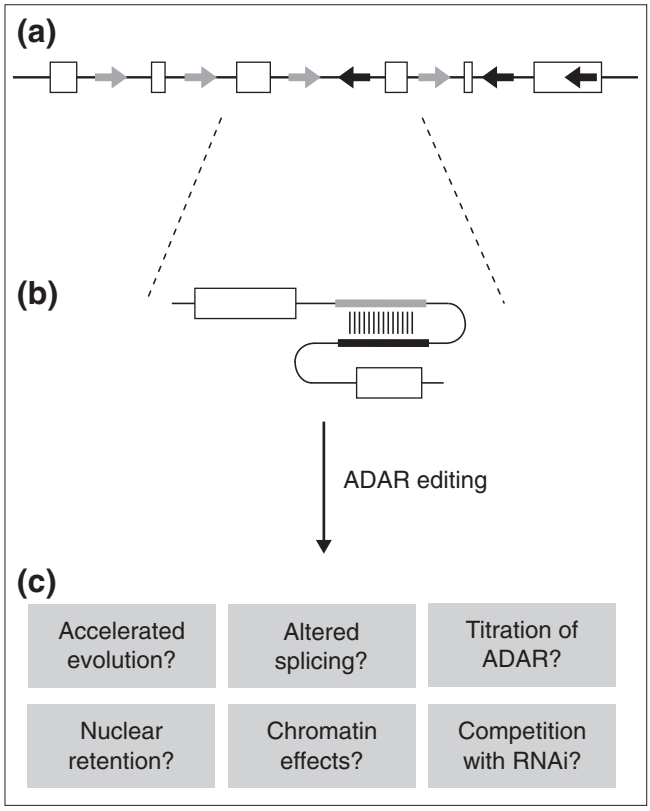


**Figure 2**
*Alu* elements in human genes. **(a)** A typical gene, with exons as boxes and introns as lines. *Alu* elements (arrows) are found at multiple locations, primarily in introns, and in either orientation (black or gray shading). **(b)** Part of the gene from (a) is shown after transcription; inverted *Alu* elements can base-pair to give dsRNA structures that serve as substrates for ADAR editing. **(c)** Editing may lead to one or more consequences (see text for details).

## Is *Alu* editing important?

So what is the significance of the high level of *Alu* editing that appears to be restricted to primates? *Alu* editing may serve no particular purpose but may simply result from the abundance of *Alu* elements in the human genome. In this scenario, dsRNA duplexes in pre-mRNAs would be edited by the ADAR in the nucleus, but other kinds of mRNA processing and function would be largely unaffected. The large difference in editing between primates and rodents may simply reflect the fact that in humans the SINEs are all almost identical, whereas in rodents there are multiple classes that are less likely to base-pair with one another.

Alternatively, *Alu* editing may be functional, and we can suggest six different but not mutually exclusive interpretations for the significance of the high level of *Alu* editing observed (Figure 2c). Firstly, it may provide a rich additional source of genetic recoding that can influence protein function and evolution. Although exonic *Alu* elements are generally in noncoding regions, some lie within coding regions, and editing of these can lead to amino-acid changes. Athanasiadis *et al*. [9] illustrated this principle for the gene encoding the G-protein-coupled receptor LUSTR1, which contains an *Alu*-related element within an alternatively spliced exon. Editing was observed at several sites in this exonic element, and the editing varied significantly in different tissues. Thus, *Alu* editing might serve as a novel source of functional diversity for proteins. If transcripts containing edited exonic *Alu* elements were mobilized for transposition in germ cells (probably a very rare event), genetic variation could be enhanced by a route other than random mutagenesis, thus serving as a mechanism to speed evolution.

Secondly, *Alu* editing might help to regulate splicing. In the human genome there is an enormous amount of alternative splicing of pre-mRNAs. Furthermore, at least 5% of all known human alternative exons are derived from *Alu* elements, and even single-base mutations in these elements can lead to splicing effects [15]. Thus, editing of *Alu* elements could possibly influence RNA splicing, for example by creating new splicing signals; this has in fact been observed [16]. As most of the observed *Alu* editing is of the promiscuous type, however, such regulation is likely to be relatively rare in human populations.

*Alu* editing could alternatively lead to titration of ADAR activity: inverted *Alu* elements would attract ADAR to harmless intronic sites and thereby titrate the activity of the enzyme away from important targets of selective editing, perhaps thereby modulating the levels of selective editing. Consistent with this model, some recent work has shown that the subnuclear localization of ADAR2 can be influenced by the concentration of its substrates [17]. Also, numerous researchers have observed that all forms of ADAR editing vary significantly from tissue to tissue, as does *Alu* editing.

In another model, editing could perform a quality-control function, to prevent promiscuously edited mRNAs from reaching the cytoplasm. Interestingly, most of the edited RNAs reported in the recent studies [6-9] contain edited introns that have not been removed. These incompletely processed mRNAs may represent non-functional transcripts that were detected only because they have inosines in them. It has been reported that promiscuously edited RNAs can be retained in the nucleus through a strong and specific interaction with a protein complex associated with the nuclear matrix [4]. Therefore, the bulk of mRNAs containing edited *Alu* sequences, and certainly those with edited intronic *Alu*s, might remain in the nucleus and thus not interfere with normal gene expression.

An intriguing possibility concerns the effects of *Alu* editing on chromatin. Even though *Alu* elements are found primarily within transcribed genes, they appear to be associated with aspects of more condensed chromatin, such as CpG methylation [18] and histone H3 lysine 9 methylation [19].

Could they therefore contribute to chromatin domains that might influence transcriptional activity? If so, could this be related to editing? This possibility is supported by the recent observation [5] that edited RNAs bind tightly to a protein, vigilin, which is closely associated with and important for the formation of heterochromatin.

Finally, we must consider the possibility that *Alu* editing reflects a competition between distinct cellular dsRNA response pathways that may be active in the nucleus. In the past few years, increasing evidence has suggested that dsRNA can, in some cases, lead to heterochromatic gene silencing through a pathway related to RNA interference (RNAi) [20], but editing of dsRNAs inhibits the RNAi response [21]. Given that most human genes contain *Alu* elements with the potential to form dsRNA structures, and that such duplexes could potentially lead to gene silencing by the RNAi machinery, editing might serve to save the cell from silencing most of its own genes by modulating an RNAi-mediated gene-silencing response. There has been a report of splicing regulation that is dependent on the dsRNA-activated kinase, PKR [22]. As some PKR is nuclear [23], it is possible that *Alu* hybrids can influence the local or even global activity of this important enzyme, and that editing can modulate this influence.

There is currently insufficient evidence for us to decide which of these models reflects the real function(s) of *Alu* editing; some or all of them may be true. It is clear, however, that RNA editing is far more widespread in the human genome than previously imagined, and it now appears to have the potential to impact the expression of almost every single gene. Future work may help to determine whether this in fact happens and whether *Alu* elements confer on primates a novel genetic advantage not available to other mammals.

## Acknowledgements

The authors thank A. Athanasiadis, M. Blow, A. Gabriel and E. Levanon for helpful discussions and acknowledge NIH grant GM066816 for support.

## References

1. Bass BL: **RNA editing by adenosine deaminases that act on RNA.** *Annu Rev Biochem* 2002, **71:**817-846.
2. Hoopengardner B, Bhalla T, Staber C, Reenan R: **Nervous system targets of RNA editing identified by comparative genomics.** *Science* 2003, **301:**832-836.
3. Scadden AD, Smith CW: **Specific cleavage of hyper-edited dsRNAs.** *EMBO J* 2001, **20:**4243-4252.
4. Zhang Z, Carmichael GG: **The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs.** *Cell* 2001, **106:**465-475.
5. Wang Q, Zhang Z, Blackwell K, Carmichael GG: **Vigilins bind to promiscuously A-to-I edited RNAs and are involved in the formation of heterochromatin.** *Curr Biol* 2005, **15:**384-391.
6. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, *et al.*: **Systematic identification of abundant A-to-I editing sites in the human transcriptome.** *Nat Biotechnol* 2004, **22:**1001-1005.
7. Blow M, Futreal PA, Wooster R, Stratton MR: **A survey of RNA editing in human brain.** *Genome Res* 2004, **14:**2379-2387.
8. Kim DD, Kim TT, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A: **Widespread RNA editing of embedded alu elements in the human transcriptome.** *Genome Res* 2004, **14:**1719-1725.
9. Athanasiadis A, Rich A, Maas S: **Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome.** *PLoS Biol* 2004, **2:**e391.
10. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409:**860-921.
11. Versteeg R, van Schaik BD, van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AH: **The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes.** *Genome Res* 2003, **13:**1998-2004.
12. Deininger PL, Batzer MA: **Mammalian retroelements.** *Genome Res* 2002, **12:**1455-1465.
13. Eisenberg E, Nemzer S, Kinar Y, Sorek R, Rechavi G, Levanon EY: **Is abundant A-to-I RNA editing primate-specific?** *Trends Genet* 2005, **21:**77-81.
14. Gibbs RA, Weinstock GM, Metzker ML, Muzny DM, Sodergren EJ, Scherer S, Scott G, Steffen D, Worley KC, Burch PE, *et al.*: **Genome sequence of the Brown Norway rat yields insights into mammalian evolution.** *Nature* 2004, **428:**493-521.
15. Kreahling J, Graveley BR: **The origins and implications of Aluternative splicing.** *Trends Genet* 2004, **20:**1-4.
16. Lev-Maor G, Sorek R, Shomron N, Ast G: **The birth of an alternatively spliced exon: 3′-splice site selection in Alu exons.** *Science* 2003, **300:**1288-1291.
17. Sansam CL, Wells KS, Emeson RB: **Modulation of RNA editing by functional nucleolar sequestration of ADAR2.** *Proc Natl Acad Sci USA* 2003, **100:**14018-14023.
18. Yoder JA, Walsh CP, Bestor TH: **Cytosine methylation and the ecology of intragenomic parasites.** *Trends Genet* 1997, **13:**335-340.
19. Kondo Y, Issa JP: **Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells.** *J Biol Chem* 2003, **278:**27658-27662.
20. Ekwall K: **The RITS complex-A direct link between small RNA and heterochromatin.** *Mol Cell* 2004, **13:**304-305.
21. Scadden AD, Smith CW: **RNAi is antagonized by A-to-I hyper-editing.** *EMBO Rep* 2001, **2:**1107-1111.
22. Osman F, Jarrous N, Ben-Asouli Y, Kaempfer R: **A cis-acting element in the 3′-untranslated region of human TNF-alpha mRNA renders splicing dependent on the activation of protein kinase PKR.** *Genes Dev* 1999, **13:**3280-3293.
23. Jeffrey IW, Kadereit S, Meurs EF, Metzger T, Bachmann M, Schwemmle M, Hovanessian AG, Clemens MJ: **Nuclear localization of the interferon-inducible protein kinase PKR in human cells and transfected mouse cells.** *Exp Cell Res* 1995, **218:**17-27.