Genome Analysis

Is there any sense in antisense editing?

Yossef Neeman[1,2], Dvir Dahary[1], Erez Y. Levanon[1,3], Rotem Sorek[4] and Eli Eisenberg[5]

[1]Compugen Ltd, 72 Pinchas Rosen St, Tel-Aviv 69512, Israel
[2]Faculty of life sciences, Bar Ilan University, Ramat Gan 52900, Israel
[3]Department of Pediatric Hematology–Oncology, Chaim Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv 52621, Israel
[4]Department of Human Genetics, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel
[5]School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Several recent studies have hypothesized that sense–antisense RNA-transcript pairs create dsRNA duplexes that undergo extensive A-to-I RNA editing. We studied human and mouse genomic antisense regions and found that the editing level in these areas is negligible. This observation questions the scope of sense–antisense duplexes formation *in-vivo*, which is the basis for several proposed regulatory mechanisms.

RNA editing

Adenosine to inosine (A-to-I) RNA-editing is a post-transcriptional mechanism, resulting in a mature mRNA that is modified relative to its genomic template. A-to-I editing is mediated by members of the double-stranded RNA-specific adenosine deaminases acting on RNA (ADAR) family, and can change codons, create or destroy splice sites, alter RNA structure and affect RNA localization and translation rates (reviewed in Ref. [1]). RNA editing is crucial for normal life and development in both invertebrates and vertebrates [2–4], and altered editing patterns are associated with several pathologies. Until recently only a handful of edited human genes were documented. However, increased levels of inosines are observed in mammalian transcripts, making it clear that the few known editing events were only the 'tip of the iceberg' [5].

Naturally occurring antisense RNA

Sense and antisense transcript pairs are RNAs containing sequences that are complementary to each other. They can be transcribed in *cis*, from opposing DNA strands at the same genomic locus, or in *trans*, from distinct loci. Several independent studies have reported on widespread natural antisense transcripts (NATs). In humans, between 5% and 10% of all genes were found to have a *cis* antisense counterpart [6–8], and similar results were reported for the mouse [9], *Drosophila* [10], *Arabidopsis* [11] and rice genomes [12].

Pioneering studies in several eukaryotic systems have identified several mechanisms by which antisense transcription can regulate gene expression, including transcriptional interference, RNA masking, RNA interference (RNAi) and RNA editing (reviewed in Ref. [13]). Most of these mechanisms assume pairing of antisense transcripts to form double-stranded RNA (dsRNA) structures. Hypothetically, these long (average of >300bp) and perfect inter-molecular duplexes can serve as editing substrates. Indeed, such long and perfectly matching dsRNAs are extensively edited following transfection to mammalian cells *in vitro* [14]. Several studies have, therefore, suggested that naturally occurring antisense dsRNA duplexes are heavily edited *in vivo*, proposing a possible general regulatory role for antisense transcripts [3,15–17]. Nevertheless, only two cases of editing in sense–antisense transcript pairs have been reported so far, none of them in mammals [18,19].

Alu repeats as major RNA-editing substrates in humans

Recently, several independent studies have introduced bioinformatics methods for the detection of A-to-I editing sites. They found that A-to-I editing is extremely abundant in the human genome [20–23], where virtually all of the editing sites reside within *Alu* repeat elements. *Alu* elements are typically 300 nucleotides long, and account for $>10\%$ of the human genome [24]. Being so abundant in the genome, they are likely to have a second nearby *Alu* repeat with reversed orientation. If such an inverted repeat exists, the two repeats can pair together to form the dsRNA hairpin structure that is then targeted by the ADARs [20–23].

Although the findings of abundant *Alu* editing account for the observed increased levels of inosines, the question whether editing of antisense transcripts has a significant role still remains. Because perfectly matching long dsRNAs are extensively edited following transfection to mammalian cells [14], the existence of antisense editing, or lack thereof, might tell us how significant is pairing of antisense transcripts in the nucleus.

Searching for editing sites in antisense loci

The sequencing reaction (and the ribosome) recognizes inosine (I) as a guanosine (G). Therefore, the fingerprints of ADAR editing are genomically encoded adenosines, which are read as guanosines in the RNA sequence. Following the approach used in our earlier work [25], we used alignments of 128 068 mRNA sequences to the genome in UCSC July 2003 assembly and recorded all

Corresponding author: Eisenberg, E. (elieis@post.tau.ac.il).
Available online 15 August 2005

Table 1. Mismatch clusters in all human RNAs and antisense regions[a]

Cluster type	Mismatch type	All RNA sequences (94 858)		Antisense regions only[b]		Antisense regions without repeats	
		Number	Percent (%)	Number	Percent (%)	Number	Percent (%)
Cluster of one mismatch	A to G	102 832	39%	1731	34%	1286	29%
	G to A	52 488	20%	1029	20%	945	21%
	T to C	58 083	22%	1317	25%	1191	27%
	C to T	52 195	19%	1076	21%	1002	23%
Cluster of three mismatches	A to G	11 613	80%	100	78%	18	46%
	G to A	968	6%	9	7%	7	18%
	T to C	1115	8%	10	8%	8	21%
	C to T	853	6%	9	7%	6	15%
Cluster of five mismatches	A to G	4926	96%	37	95%	3	60%
	G to A	48	1%	0	0%	0	0%
	T to C	71	1.5%	2	5%	2	40%
	C to T	74	1.5%	0	0%	0	0%

[a]The number of a single or a cluster of consecutive mismatches, for the four most common mismatches. In the complete sequence of the RNAs, and in antisense regions only, there is a preference for A-to-G mismatches over all other mismatches.
[b]When filtering out *Alu* repeats in antisense regions, the distribution of A-to-G over other mismatches drops sharply, and the number of mismatches becomes negligible.

the mismatches along them. A-to-I-editing sites often occur in clusters, an edited sequence is typically edited in many close-by sites [26]. Therefore, to detect the correct editing sites [as opposed to single nucleotide polymorphisms (SNPs), sequencing and other errors], we retained only those mismatches that are part of a stretch of identical mismatches between the given RNA sequence and the genomic DNA. Applying this to all RNA sequences resulted in a vast over-representation of A-to-G mismatches compared with other common mismatches, suggesting that we detected true editing sites (Table 1). We found 11 613 (~80%) clusters of three consecutive identical A-to-G mismatches, compared with only 968 clusters of G-to-A mismatches (~7%). This means that ~10 600 (>90%) of the detected A-to-G mismatch clusters are probably a result of A-to-I-editing events [25]. The specificity improves as we increase the number of identical mismatches in the cluster: requiring stretches of five consecutive identical mismatches results in 96% A-to-G mismatch clusters (Table 1).

To test whether antisense RNAs are significant substrates for RNA editing, we combined our algorithm for identifying antisense regions [6] with the approach described above for detecting editing sites [25]. Using the Antisensor algorithm [6] (Box 1), we found pairs of overlapping transcriptional units on opposite DNA strands. This approach yielded 9502 genes that are predicted to have a *cis* antisense counterpart. These genomic loci, in which sense and antisense transcripts are predicted to overlap, cover a total of ~4.3 million base-pairs (bp) and are supported by 16 344 RNA sequences (GenBank version GB139; http://www.ncbi.nlm.nih.gov/). The average overlap was 316 bp and the median was 168 bp. We then looked for traces of A-to-I editing in sequences transcribed from antisense loci. This resulted in 4307 RNA sequences that include at least one mismatch within an antisense region.

Antisense regions are not extensively edited

Focusing on antisense regions only, we detected a pattern similar to that detected for the complete RNA sequence (Table 1). For clusters of three consecutive mismatches ~78% are of the A-to-G type (~80% in the complete RNA sequence), and for clusters of five consecutive mismatches

Box 1. The Antisensor algorithm

The antisensor algorithm [6] assigns the correct DNA strand for each expressed sequence (Figure I). The two main sources of information used by the algorithm are the splice-junction consensus sequences (introns begin with GT and end with AG in most known introns), and the polyA tail at the 3′ end of the transcripts. The cluster of sequences at a given genomic locus is then separated according to the strand the sequences are transcribed from. If two distinct reliable sequence clusters are formed, one deduces the existence of antisense transcription.

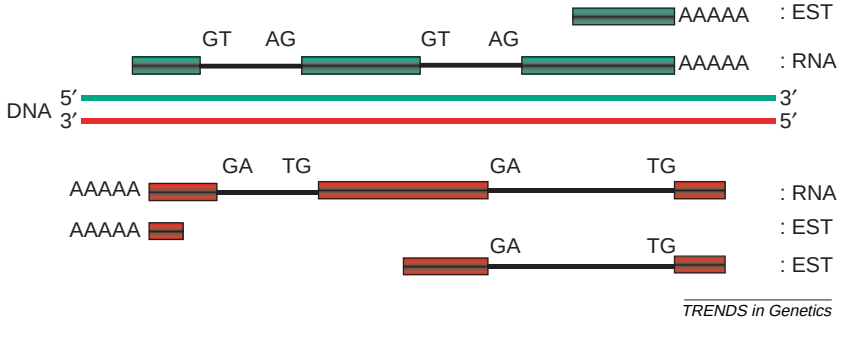

Figure I. Combining different sources of information for antisense prediction. Thin bars represent the two strands of DNA; thicker bars stand for expressed sequences, where lines connect the different exons. The splice junction consensus sequences and polyA tail sequences are shown. Based on these sources of information, the expressed sequences are separated into two clusters, (colored green or red) according to the DNA strand they are transcribed from.

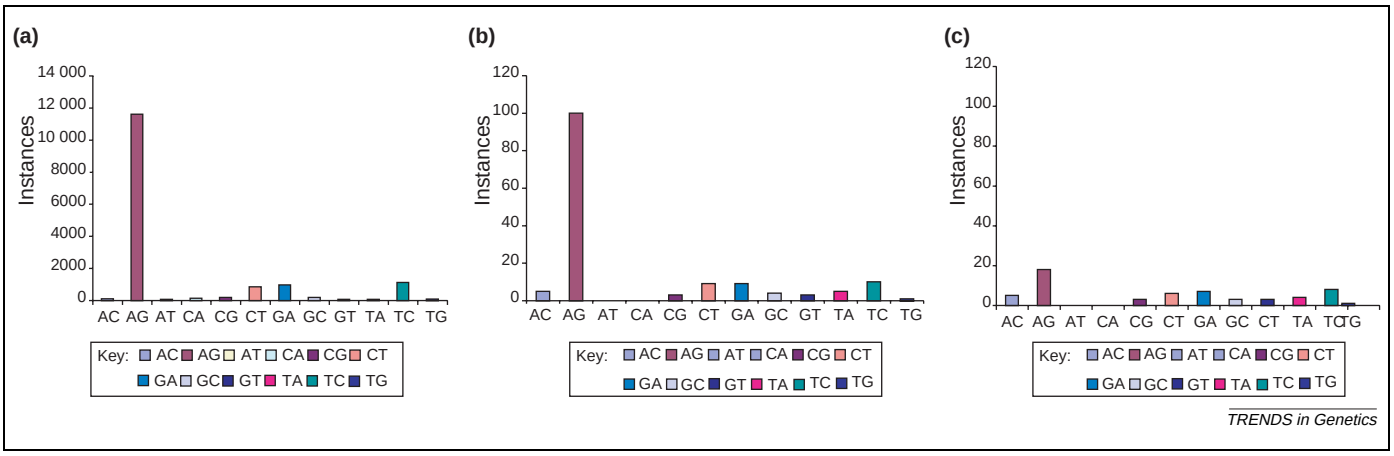

Figure 1. Distributions of clusters of three identical, consecutive mismatches, for the different mismatch types: **(a)** Genomic mismatches in all RNA sequences (UCSC July 2003 assembly); **(b)** genomic mismatches in antisense regions; and **(c)** genomic mismatches in antisense regions without *Alu* repeats. A significant preference of A-to-G mismatches is observed for all RNAs and for antisense regions. However, after filtering out *Alu* repeats regions, there is no significant preference for A-to-G over other mismatches, suggesting that intra-molecular pairing of *Alu* repeats is responsible for the editing signal observed.

~95% are A-to-G mismatches (~96% in the complete RNAs). However, one must take into account the occurrence of editing sites in the antisense loci owing to intra-molecular *Alu–Alu* dsRNAs. Indeed, a recent study by Athanasiadis *et al.* [20] reported that only 1% of editing in *Alu* elements could be attributed to inter-molecular dsRNAs, with the remainder probably guided by intra-molecular *Alu–Alu* dsRNA. Therefore, we filtered out all *Alu* repeats from the data set using Repeatmasker (http://www.repeatmasker.org/). These *Alu* repeats comprise merely 7% of the antisense genomic loci (~320 000 bp). However, excluding this small fraction of the antisense regions, the over-representation of A-to-G mismatch clusters compared with other common mismatches (Figure 1) virtually vanished, suggesting that antisense transcripts, apart from the *Alu* regions within them, are not extensively edited (Figure 2). Our results show a slight preference of A-to-G mismatches over other mismatches (18 examples of A-to-G mismatch clusters compared with eight examples of T-to-C mismatch; Table 1). Some of these mismatches might attest for RNA editing events, either caused by intra-molecular paring other then *Alu–Alu* pairing, or as a result of antisense transcript pairing. Nevertheless, these few examples are negligible compared with the abundance of the global RNA editing phenomenon in the human genome.

To rule out the possibility that RNAs with antisense counterparts have different features than RNAs without such counterparts, we repeated the analysis for the subset of RNAs that are part of antisense pairs, and searched for differences between their overlapping and the non-overlapping regions. The results were essentially identical: after filtering out *Alu* elements, both regions have the same mismatch distribution exhibiting no significant over-representation of A-to-G mismatch clusters. For example, among the clusters that contain three mismatches, 46% and 38% contain A-to-G mismatches in overlapping and non-overlapping regions, respectively, with a low overall mismatches rate (Table 1 supplementary material online).

The widespread natural antisense transcripts (NATs) phenomenon is not human specific: ~2400 sense–antisense gene pairs have been identified in the mouse genome [9] and 1027 in the *Drosophila* genome [10]. Although it was expected that similar levels of editing would be observed for all mammals, two recent studies reported

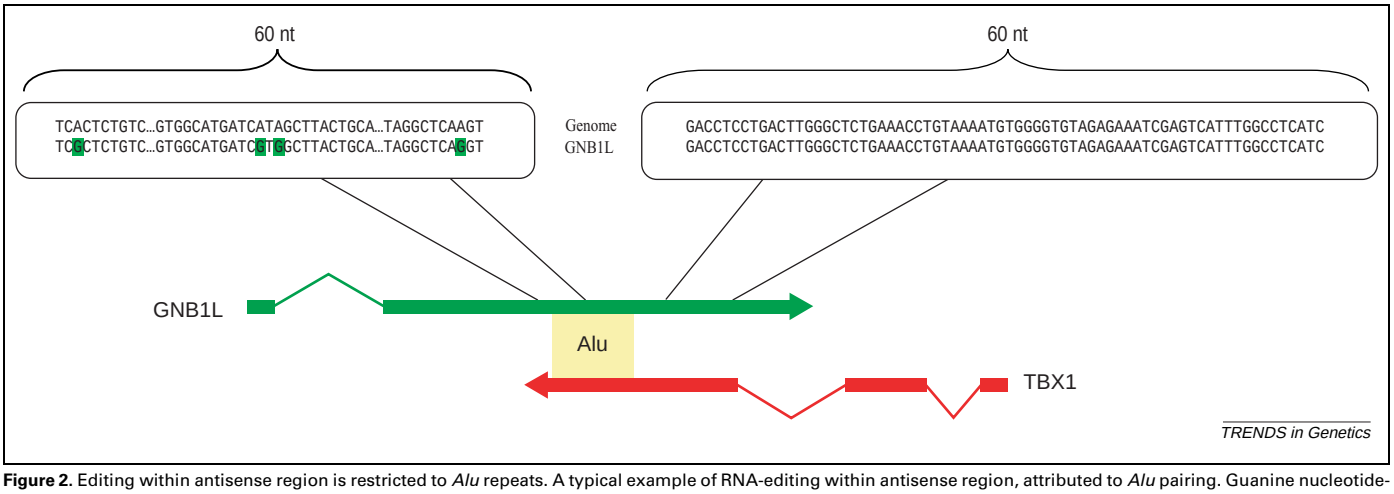

Figure 2. Editing within antisense region is restricted to *Alu* repeats. A typical example of RNA-editing within antisense region, attributed to *Alu* pairing. Guanine nucleotide-binding protein β polypeptide 1-like (GNB1L) has an overlapping antisense region with T-box 1 (TBX1). The genomic overlap region includes an *Alu* repeat. A cluster of A-to-G mismatches (highlighted in green), indicative of A-to-I editing, is found only in the *Alu* region, whereas the remainder of the antisense region is not edited.

that the editing levels in human are at least an order of magnitude greater than those of mouse, with most human editing sites residing within *Alu* elements [21,25]. Therefore, it is possible that antisense editing in human is masked by the extensive amount of *Alu* editing, but will be observable in other organisms. We therefore applied the same search algorithm to the mouse data. Here too, we found no preference for A-to-G mismatches over other common mismatches in antisense regions (Table 2 in supplementary material online).

Concluding remarks

It had been shown that long, perfectly matching dsRNAs are extensively edited following transfection to mammalian cells [14]. However, Athanasiadis *et al.* have recently reported that only 1% of editing in *Alu* sequences could be attributed to inter-molecular dsRNA, suggesting that antisense pairing does not lead to significant editing [20]. We have conducted a systematic search throughout human and mouse antisense loci, looking for traces of A-to-I RNA editing. Although our results cannot exclude the possibility of some antisense genes being modified by editing, we found no evidence for significant RNA editing within antisense regions. These results might lead to the conclusion that inter-molecular sense-antisense RNA pairings do not normally occur after transcription in the nucleus. Alternatively, pairing might actually occur within the cell, but the resulting duplexes, edited or unedited, are either retained in the nucleus [17] or degraded by RNAi or other mechanisms, and are thus not represented in expressed sequence data. The role of the abundant antisense transcripts in gene expression regulation is, therefore, still elusive.

Acknowledgements

R.S. was supported by a fellowship from the Clore Israel Foundation. E.E. was supported by an Alon Fellowship at Tel Aviv University.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2005.08.005

References

1 Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* 71, 817–846
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 Wang, Q. *et al*. (2000) Requirement of the RNA editing deaminase *ADAR1* gene for embryonic erythropoiesis. *Science* 290, 1765–1768
4 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
5 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
6 Yelin, R. *et al*. (2003) Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* 21, 379–386
7 Shendure, J. and Church, G.M. (2002) Computational discovery of sense-antisense transcription in the human and mouse genomes. *Genome Biol.* 3, DOI:10.1186/gb-2002-3-9-research0044 (http://genomebiology.com/2002/3/9/research/0044)
8 Lehner, B. *et al*. (2002) Antisense transcripts in the human genome. *Trends Genet.* 18, 63–65
9 Kiyosawa, H. *et al*. (2003) Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. *Genome Res.* 13(6B), 1324–1334
10 Misra, S. *et al.* (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* 3, DOI:10.1186/gb-2002-3-12-research0083 (http://genomebiology.com/2002/3/12/research/0083)
11 Meyers, B.C. *et al*. (2004) Analysis of the transcriptional complexity of *Arabidopsis thaliana* by massively parallel signature sequencing. *Nat. Biotechnol.* 22, 1006–1011
12 Osato, N. *et al.* (2003) Antisense transcripts with rice full-length cDNAs. *Genome Biol.* 5, R5 (http://genomebiology.com/2003/5/1/R5)
13 Lavorgna, G. *et al*. (2004) In search of antisense. *Trends Biochem. Sci.* 29, 88–94
14 Polson, A.G. and Bass, B.L. (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J.* 13, 5701–5711
15 Tonkin, L.A. and Bass, B.L. (2003) Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science* 302, 1725
16 Kumar, M. and Carmichael, G.G. (1998) Antisense RNA: function and fate of duplex RNA in cells of higher eukaryotes. *Microbiol. Mol. Biol. Rev.* 62, 1415–1434
17 Carmichael, G.G. (2003) Antisense starts making more sense. *Nat. Biotechnol.* 21, 371–372
18 Peters, N.T. *et al*. (2003) RNA editing and regulation of *Drosophila* 4f-rnp expression by sas-10 antisense readthrough mRNA transcripts. *RNA* 9, 698–710
19 Kimelman, D. and Kirschner, M.W. (1989) An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in *Xenopus* oocytes. *Cell* 59, 687–696
20 Athanasiadis, A. *et al*. (2004) Widespread A-to-I RNA editing of *Alu*-containing mRNAs in the human transcriptome. *PLoS Biol.* 2, e391
21 Kim, D.D. *et al*. (2004) Widespread RNA editing of embedded *Alu* elements in the human transcriptome. *Genome Res.* 14, 1719–1725
22 Levanon, E.Y. *et al*. (2004) Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat. Biotechnol.* 22, 1001–1005
23 Blow, M. *et al*. (2004) A survey of RNA editing in human brain. *Genome Res.* 14, 2379–2387
24 Lander, E.S. *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
25 Eisenberg, E. *et al*. (2005) Is abundant A-to-I RNA editing primate-specific? *Trends Genet.* 21, 77–81
26 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

doi:10.1016/j.tig.2005.08.005