nature biotechnology

Systematic identification of abundant A-to-I editing sites in the human transcriptome

Erez Y Levanon[1,2,4], Eli Eisenberg[1,4], Rodrigo Yelin[1,4], Sergey Nemzer[1,4], Martina Hallegger[3], Ronen Shemesh[1], Zipora Y Fligelman[1], Avi Shoshan[1], Sarah R Pollock[1], Dan Sztybel[1], Moshe Olshansky[1], Gideon Rechavi[2] & Michael F Jantsch[3]

RNA editing by members of the ADAR (adenosine deaminases acting on RNA) family leads to site-specific conversion of adenosine to inosine (A-to-I) in precursor messenger RNAs. Editing by ADARs is believed to occur in all metazoa, and is essential for mammalian development. Currently, only a limited number of human ADAR substrates are known, whereas indirect evidence suggests a substantial fraction of all pre-mRNAs being affected. Here we describe a computational search for ADAR editing sites in the human transcriptome, using millions of available expressed sequences. We mapped 12,723 A-to-I editing sites in 1,637 different genes, with an estimated accuracy of 95%, raising the number of known editing sites by two orders of magnitude. We experimentally validated our method by verifying the occurrence of editing in 26 novel substrates. A-to-I editing in humans primarily occurs in noncoding regions of the RNA, typically in Alu repeats. Analysis of the large set of editing sites indicates the role of editing in controlling dsRNA stability.

RNA editing by members of the double-stranded RNA (dsRNA)-specific ADAR family leads to site-specific conversion of A-to-I in precursor messenger RNAs[1]. ADAR-mediated RNA editing is essential for normal life and development in both invertebrates and vertebrates[2–5]. ADAR-deficient invertebrates show only behavioral defects[2,3], whereas ADAR1 knockout mice die embryonically and ADAR2 null mice are born at full term but die prematurely[4,5]. High editing levels have been found in inflamed tissues[6], in agreement with a proposed antiviral function for ADARs and their transcriptional regulation by interferon[7]. Altered editing patterns were found in epileptic mice[8], suicide victims suffering chronic depression[9], amyotrophic lateral sclerosis[10] and in malignant gliomas[11]. Until recently only a handful of edited human genes were documented, most of which were discovered serendipitously[12]. A systematic experimental search for inosine-containing RNAs has yielded 19 additional cases[13], and one further example was found using a cross-genome comparison approach[14]. However, quantification of inosine in total RNA suggests that editing affects a much larger fraction of the mammalian transcriptome[15]. In addition, tantalizing hints of abundant editing were observed in high-throughput cDNA sequencing data[16].

Large-scale identification of editing substrates using bioinformatics tools was previously considered practically impossible[17]. In principle, editing may be detected using the large-scale database of expressed sequence tags[18] (ESTs) and RNAs, which currently holds over five million human records. Editing sites show up when a sequence is aligned with the genome: while the DNA reads A, sequencing identifies the inosine in the edited site as guanosine (G). However, the poor sequencing quality of the sequence database (up to 3% sequencing errors[19]) precludes a straightforward application of this approach. Moreover, millions of single nucleotide polymorphisms (SNPs) and mutations are erroneously identified as editing events by this method.

Here we present a computational approach that overcomes these challenges. We mapped 12,723 A-to-I editing sites in 1,637 different genes. Editing was experimentally validated in 26 of these 1,637 genes. The editing sites found are typically located within Alu elements residing in noncoding regions of the RNA. The effect of editing on dsRNA stability is analyzed.

RESULTS

Computational identification of A-to-I editing

ADAR substrates are usually imperfect dsRNA stems formed by base pairing of an exon containing the adenosine to be edited with a complementary portion of the pre-mRNA (up to several thousand nucleotides apart). We therefore restricted the search for mismatches to potential double-stranded regions, in order to remove most of the noise and facilitate the identification of true editing sites. For this purpose, human ESTs and cDNAs were aligned to the genome and assembled into clusters representing genes or partial genes[20]. We then used our algorithm to align the expressed part of the gene with the corresponding genomic region, looking for reverse complement alignments longer than 32 bp with identity levels higher than 85% (**Fig. 1**). About 429,000 candidate dsRNAs were found in 14,512 different genes, mostly resulting from alignment of an exon (including

[1]Compugen Ltd., 72 Pinchas Rosen St., Tel-Aviv 69512, Israel. [2]Department of Pediatric Hematology-Oncology, Chaim Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv 52621, Israel. [3]Max F. Perutz Laboratories, Dept. of Cell Biology and Genetics, University of Vienna, Rennweg 14, A-1030 Vienna, Austria. [4]These authors contributed equally to this work. Correspondence should be addressed to E.Y.L. (erez@compugen.co.il).

Published online 18 July 2004; doi:10.1038/nbt996

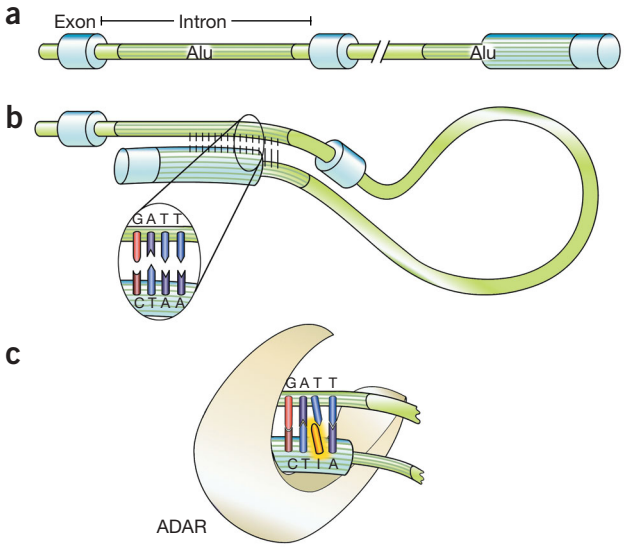


Figure 1 ADAR-mediated editing. (**a**) Pre-mRNA as transcribed from DNA. The gene contains two Alu repeats with opposite orientations, one of which overlaps with an exon. (**b**) The two oppositely oriented Alu sequences form a dsRNA structure. (**c**) An enzyme of the ADAR family edits some of the adenosines in the dsRNA structure into inosines.

the 3′- and 5′-UTRs) to an intron. To further decrease the number of random mismatches, SNPs and mutations, the algorithm then cleaned the sequences supporting the stem region. Because sequencing errors tend to cluster in certain regions, especially in low complexity areas and towards the ends of sequences, we discarded all single-letter repeats longer than 4 bp, as well as 150 bp at both ends of each sequence. In addition, all 50 nucleotide-wide windows in which the total number of mismatches was five or more were considered as having low sequencing quality and were discarded. However, four or more identical sequential mismatches were masked in the count for mismatches in a given window. This exception is intended to retain sequences with many sequential editing sites, which were found to occur in previously documented examples[21]. Mismatches supported by <5% of available sequences were also discarded, and, finally, known SNPs of genomic origin were removed. By using these criteria one finds that the putative editing sites tend to group together[16], a fact that is also supported by the few available known cases[13]. Thus, all mismatches that occur less than three times in an exon were ignored.

This above cleaning procedure resulted almost exclusively in A-to-G mismatches (**Fig. 2a**). We identified 12,723 putative editing sites, belonging to 1,637 different genes by using this procedure. The same approach applied to G-to-A mismatches yielded only 242 sites. Sequencing errors, SNPs and mutations, the three main sources of noise in our analysis, are expected to produce at least as many G-to-A as A-to-G mismatches (**Fig. 2**). This signal-to-noise ratio (12,723:242) suggests that our false positive rate is very low.

Experimental validation

To experimentally validate the predicted editing sites, we chose 30 genes and sequenced matching DNA and RNA samples retrieved from the same specimen, for up to five tissues. We have positively verified editing events in 26 previously unknown editing substrates. PCR products were either cloned followed by sequencing of individual clones, or sequenced as a population, without cloning. When the PCR products were cloned, editing occurrence was detected by comparing the sequences of several clones with the genomic sequence (**Fig. 3** and **Supplementary Fig. 1**). When PCR products were directly sequenced, the occurrence of editing was determined by the presence of an unambiguous trace of guanosine in positions for which the genomic DNA clearly indicated the presence of an adenosine (**Fig. 4**). The full sequencing data are given on our website (see Methods). We show here direct evidence for editing in the liver, lung, kidney, prostate, colon and uterus. For most genes, editing was found in all tissues, with varying relative abundance, but generally the unedited signal dominated the edited signal. Two genes were validated using cell-lines known to have varying levels of ADAR activity (**Fig. 3**). The observed levels of A-to-I conversions correlated well with the reported ADAR activities in these cell lines[22]. Typically, additional editing sites, not present in our list, were found in the same region. The validation set was composed of two subsets: (i) 20 genes for which the EST data suggested many putative editing events, 18 of which were confirmed to be edited; (ii) 17 genes chosen randomly from the list of 1,637 predicted genes. Four

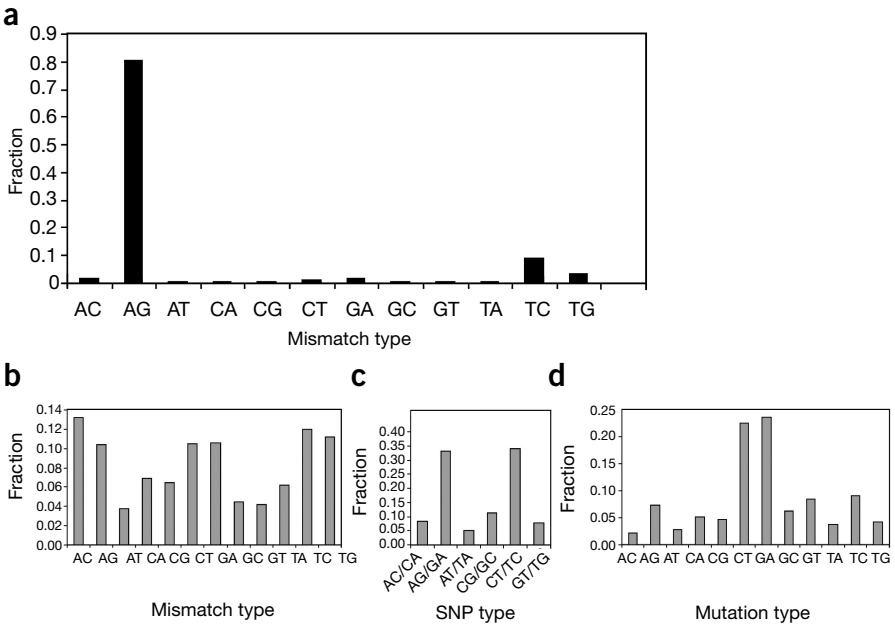


Figure 2 Distribution of mismatches between the DNA and the expressed RNA sequences that pass the cleaning algorithm. (**a**) Results of algorithm application to dsRNAs only. A-to-G mismatches clearly dominate the distribution. Notably, T-to-C mismatches are also overrepresented, likely being A-to-I editing events that were aligned to the opposite strand. (**b**) The distribution of mismatches resulting from applying the algorithm to random expressed sequences covering about 20% of the transcriptome. (**c**,**d**) The distributions for known SNPs[31] (**c**) and mutations[32] (**d**), respectively. A-to-G mismatches do not stand out in the distributions **b**–**d**.

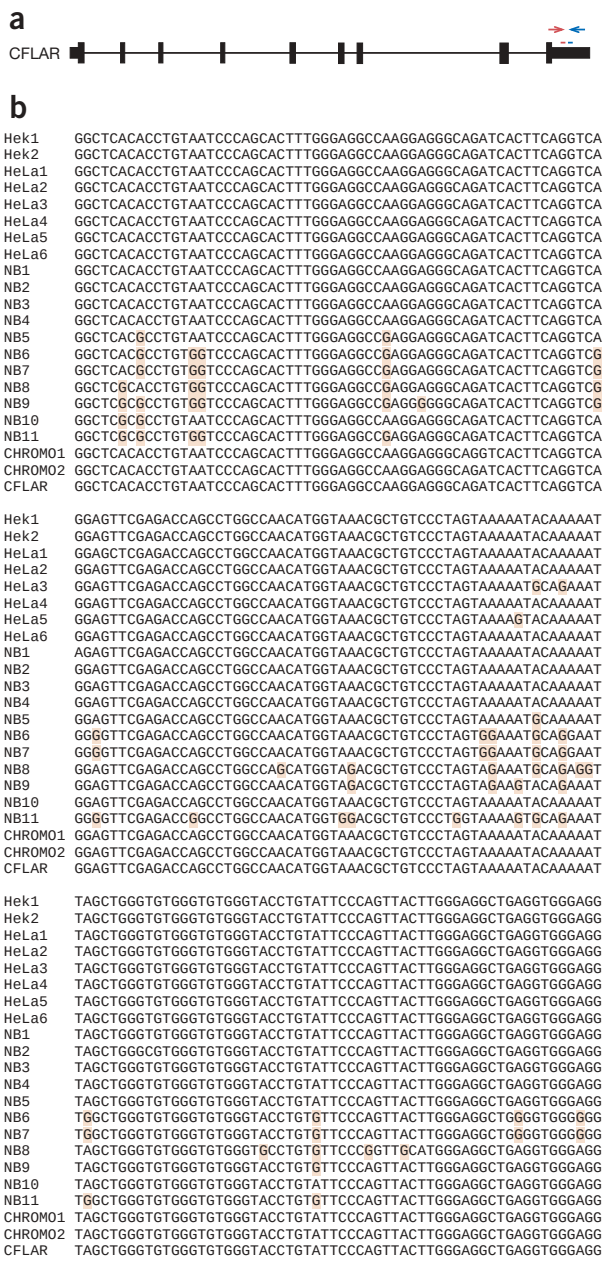


Figure 3 Editing in the CFLAR transcript. A region corresponding to the 3′-UTR of CFLAR was amplified from cDNAs and gDNA of neuroblastoma, HeLa and HEK293 cells. (**a**) Schematic organization of CFLAR with predicted editing in the 3′-UTR (brown shading). There are dozens of Alu elements within the genomic region of the CFLAR gene, and we cannot tell for sure which one pairs with the above 3′-UTR region (marked with a red arrow) to form the dsRNA required for editing. The closest Alu element is located on the 3′-UTR as well, 1,450 bp downstream (marked with a blue arrow). For this dsRNA, virtually all editing events recorded in this figure result in destabilization of the dsRNA. (**b**) Sequences of individually cloned fragments were aligned to the published human genomic sequence. No A-to-I (reads as G in the sequence) conversion is found in HEK293 cells, whereas abundant and moderate editing is seen in neuroblastomas and HeLa cells, respectively. Editing events are highlighted in light brown shading. An additional example is provided in **Supplementary Figure 1**.

of these latter genes were discarded, as they did not allow for designing high-quality amplification primers outside the Alu sequence. We successfully amplified and sequenced 9 of the remaining 13 genes, 8 of which exhibited editing. Note that the success rate in our random subset (89%) is a lower bound to the true accuracy of the list, as either low editing efficiency at a given site or limited variety of tissues in our validation experiment could prevent the detection of editing events in the experimental sample.

Characterization of the editing sites

Interestingly, 92% of sites occur within an Alu repeat, and an additional 1.3% lie within the primate-L1 repeat, in accordance with previous reports[13,16]. This is explicable by the fact that only long, paired RNA molecules were scanned for editing, a structure more likely formed between repetitive elements. The distribution of editing sites within the Alu sequence exhibits a number of preferred, edited adenosines, as well as adenosines unlikely to be edited. In particular, two specific A sites within the Alu repeat, in positions 27 and 28 of Alu, account for ~12% of all editing events (see **Supplementary Notes**).

We have also found that G is underrepresented in the nucleotide upstream to the edited A, and overrepresented in the nucleotide following the editing site (see **Supplementary Notes**), in accordance with previous reports[23,24]. However, the fact that most of the sites occur within Alu repeats strongly biases the identification of additional patterns characterizing the editing site.

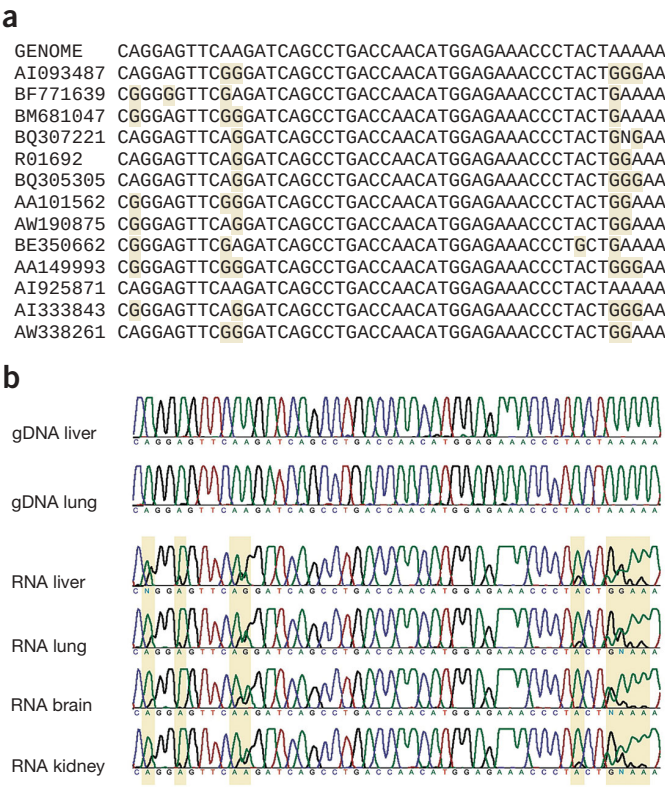
Typically, editing is seen in only a fraction of the supporting expressed sequences (ESTs or cDNAs). In fact, for 83% of the sites only one sequence exhibits editing (after applying our cleaning filters). This suggests that editing does not occur with equal frequency in all tissues and conditions, and is of a probabilistic nature. Our experimental data also support this finding.

No specific expression pattern or Gene Ontology (GO; http://www.geneontology.org/) classification for the edited genes was found. However, we analyzed the EST libraries searching for specific libraries showing an altered editing pattern. The libraries with the most significant (highest *P* value) overediting pattern came from thymus, brain, pancreas, spleen and prostate (see **Supplementary Notes**). Some of these observations support previous reports[15,25].

Editing can extend the proteomic diversity by changing the identity of a particular codon[26], as the ribosome reads inosine as guanosine. Two novel examples of such editing are presented in the **Supplementary Notes**. However, it has been predicted that most pre-mRNA editing in the brain is located in noncoding regions[21]. In agreement with this, virtually all of the editing sites identified by us were located in noncoding regions: of the sites that can be aligned with RefSeq sequences (http://www.ncbi.nlm.nih.gov/RefSeq), 12% were located in the 5′-UTR, 54% in the 3′-UTR and 33% in RefSeq introns. Some of the sites annotated as introns might actually be within an alternative exon not covered by the RefSeq sequence. Note that our strict cleaning procedure definitely misses many true editing sites. In particular, the known examples of A-to-I editing in the glutamate receptor and serotonin receptor were not picked up by our algorithm, as the expressed part of the dsRNA supporting them was not long enough. Thus, it is likely that there are more editing sites within the coding region not detected in this work.

It was suggested[21] that one of the functions of RNA editing is the destabilization of dsRNAs. Our large database of editing sites enabled us to test this prediction. ADAR-mediated editing of an A in an A-U base pair produced the less stable I-U pair, whereas A-C mismatches were edited into the more stable I-C pairs. Looking at the best complementary alignment of the editing regions, we found

Figure 4 Editing in the F11 receptor (JAM1) gene. (**a**) Some of the publicly available expressed sequences covering this gene, together with the corresponding genomic sequence. The evidence for editing is highlighted. (**b**) Results of sequencing experiments. Matching DNA and cDNA RNA sequences for a number of tissues. Editing is characterized by a trace of guanosine in the cDNA RNA sequence, where the DNA sequence exhibits only adenosine signals (highlighted). Twenty-three additional examples are provided on our website (see Methods). Note the variety of tissues showing editing, and the variance in the relative intensity of the edited guanosine signal.



that in 78% of the editing cases an A-U pair was destabilized, whereas in 19% an A-C pair was stabilized. Editing of either A-A or A-G pairs occurred in only 3% of the cases. This suggests that editing is aimed at stabilization and destabilization only, and does not occur in situations where it has no major effect on dsRNA stability. Furthermore, the editing mechanism seems to prefer stabilization: 22% of the editing events target a mismatched base pair, whereas the average frequency of such mismatched base-pairs in the sites adjacent to the editing sites is only 10%, because these sites are all located in double-stranded regions. Thus, although most editing events result in destabilization of the dsRNA, we found many more stabilization events (that is, editing of A-C to I-C) than what would be expected based on a random choice of the editing sites along the dsRNA. The preference towards stabilization editing is in agreement with previous reports[27].

DISCUSSION

This work increases the number of editing substrates by two orders of magnitude, in accordance with prior estimates[15]. This allows a large-scale analysis of the editing phenomenon. The widespread occurrence of editing makes it an important contributor to the diversity of the transcriptome, producing presumably more different transcripts than produced by alternative splicing, while affecting only a small number of nucleotides. Interestingly, the large-scale editing in humans is found to be strongly associated with Alu repeats, which are unique to primates. Thus, one does not expect the corresponding sites to be found in nonprimate mammals. However, other repeats present in these organisms may be associated with the same phenomenon. The pronounced concentration of editing sites in Alu repeats raises the question whether A-to-I editing acts as an antitransposition mechanism by inhibiting the integration of transcribed Alu back into the genome. Such a scenario is in agreement with an antiviral mechanism of editing[28], as retrotransposition of many repetitive elements is very similar to some stages of the retroviral infection. Alternatively, it has been suggested that editing regulates RNAi by protecting the dsRNA from degradation[29]. Our results indicate that these possible mechanisms may be of wide applicability. Finally, we note that there are probably many more sites than those listed in this work. (i) Editing happens in only a fraction of the sequences. Because the expressed sequence coverage of many genes is scarce, many editing sites might be absent from GenBank sequences. (ii) Our filtering parameters were chosen to minimize the noise, but inevitably miss many true sites such as the known sites in the glutamate receptor and serotonin 2c receptor pre-mRNAs[9,26]. (iii) The experimental evidence shows that a typical editing substrate contains more editing sites than the number predicted by us. Thus, the 12,723 sites we listed may still represent only a portion of the actual editing repertoire. The large-scale mapping of editing sites enables the identification of new properties of noncoding regions, and may facilitate the association of mutations in these regions with known pathologies.

METHODS

Alignment of expressed sequences to the genome. Human ESTs and cDNAs were obtained from NCBI GenBank version 136 (June 2003; http://www.ncbi.nlm.nih.gov/dbEST). The genomic sequences were taken from the Human Genome Build 33 (June 2003; http://www.ncbi.nlm.nih.gov/genome/guide/human).

Briefly, sequences were aligned as follows: sequences were cleaned from terminal vector sequences, and low-complexity stretches and repeats (including Alu repeats) in the expressed sequences were masked. Then, expressed sequences were compared with the genome to find likely high-quality hits. They were then aligned to the genome by use of a spliced alignment model that allows long gaps. Only sequences having >94% identity to a stretch in the genome were used in further stages. Further details can be found in ref. 20.

Experimental methods. Total RNA and genomic DNA (gDNA) isolated simultaneously from the same tissue sample were purchased from Biochain Institute. In this work we used samples of liver, prostate, uterus, kidney, colon, normal and tumorous lung, brain tumor (glioma), cerebellum and frontal lobe.

The total RNA underwent oligo-dT primed reverse transcription using Superscript II (Invitrogen) according to manufacturer's instructions. The cDNA and gDNA (at 0.1 μg/μl) were used as templates for PCR reactions. We aimed at high sequencing quality and thus amplified rather short genomic sequences (roughly 200 bp). The amplified regions chosen for validation were selected only if the fragment to be amplified maps to the genome at a single site. PCR reactions were done with TaKaRa Ex Taq Hot Start (Takara Bio) using the primers and annealing conditions as detailed in the **Supplementary Methods**. The PCR products were run on 2% agarose gels and if a single clear band of the correct approximate size was obtained, it was excised and sent to Hy-labs laboratories for purification and direct sequencing without cloning.

Poly A RNA from tissue culture cells was isolated using Trifast (PeqLab) and poly-A was selected using magnetic oligo dT beads (Dynal). We reverse transcribed 1 μg of poly A RNA using random hexamers as primers and RNAseH-deficient M-MLV reverse transcriptase (Promega). Genomic DNA from tissue culture cells was isolated as described[30].

First strand cDNAs or corresponding genomic regions were amplified with suitable primers using *Pfu* polymerase, to minimize mutation rates during

amplification. Amplified fragments were A-tailed using *Taq* polymerase, purified by agarose gel electrophoresis and cloned into pGem-T easy (Promega). After transformation in *Escherichia coli*, individual plasmids were sequenced and aligned using ClustalW.

We used Contig Express software from Vector NTI 6.0 Suite (Informax) for multiple-alignment of the electropherograms. Typically, the extent of A-I editing is variable; for example, the levels of the guanosine trace sometimes is only a fraction of the adenine trace, whereas in some instances the conversion from A to I is almost complete. For each gene tested, we sequenced the three tissues in which the expression was the highest. The RT-PCR and gDNA-PCR of one of these tissues were sequenced from both ends to ensure the consistency of the resulting electropherograms.

Information concerning the editing sites is available online at: http://www.cgen.com/research/Publications/AtoIEditing. This site includes the chromatograms of 23 additional editing substrates, a FASTA file containing the flanking sequences of 12,723 identified sites, and a database with additional annotation for the editing sites.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

We thank A. Diber, E. Shuster and S. Zevin for technical help and P. Akiva, A. Amit and R. Sorek for critical reading of the manuscript. The work of E.Y.L. was done in partial fulfillment of the requirements for a PhD degree from the Sackler Faculty of Medicine, Tel Aviv University, Israel. Part of this work was supported by the Austrian Science Foundation grant SFB1706 to M.J.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 5 April; accepted 24 May 2004
Published online at http://www.nature.com/naturebiotechnology/

1. Polson, A.G., Crain, P.F., Pomerantz, S.C., McCloskey, J.A. & Bass, B.L. 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 (1991).
2. Tonkin, L.A. *et al.* RNA editing by ADARs is important for normal behavior in *Caenorhabditis elegans*. *EMBO J.* **21**, 6025–6035 (2002).
3. Palladino, M.J., Keegan, L.P., O'Connell, M.A. & Reenan, R.A. A-to-I pre-mRNA editing in *Drosophila* is primarily involved in adult nervous system function and integrity. *Cell* **102**, 437–449 (2000).
4. Wang, Q., Khillan, J., Gadue, P. & Nishikura, K. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* **290**, 1765–1768 (2000).
5. Higuchi, M. *et al.* Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* **406**, 78–81 (2000).
6. Yang, J.H. *et al.* Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. *Immunology* **109**, 15–23 (2003).
7. Patterson, J.B. & Samuel, C.E. 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 (1995).
8. Brusa, R. *et al.* Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice. *Science* **270**, 1677–1680 (1995).
9. Gurevich, I. *et al.* Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims. *Neuron* **34**, 349–356 (2002).
10. Kawahara, Y. *et al.* Glutamate receptors: RNA editing and death of motor neurons. *Nature* **427**, 801 (2004).
11. Maas, S., Patt, S., Schrey, M. & Rich, A. Underediting of glutamate receptor GluR-B mRNA in malignant gliomas. *Proc. Natl. Acad. Sci. USA* **98**, 14687–14692 (2001).
12. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu. Rev. Biochem.* **71**, 817–846 (2002).
13. Morse, D.P. & Bass, B.L. Long RNA hairpins that contain inosine are present in *Caenorhabditis elegans* poly(A)+ RNA. *Proc. Natl. Acad. Sci. USA* **96**, 6048–6053 (1999).
14. Hoopengardner, B., Bhalla, T., Staber, C. & Reenan, R. Nervous system targets of RNA editing identified by comparative genomics. *Science* **301**, 832–836 (2003).
15. Paul, M.S. & Bass, B.L. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J.* **17**, 1120–1127 (1998).
16. Kikuno, R., Nagase, T., Waki, M. & Ohara, O. HUGE: a database for human large proteins identified in the Kazusa cDNA sequencing project. *Nucleic Acids Res.* **30**, 166–168 (2002).
17. Seeburg, P.H. A-to-I editing: new and old sites, functions and speculations. *Neuron* **35**, 17–20 (2002).
18. Boguski, M.S., Lowe, T.M. & Tolstoshev, C.M. dbEST–database for "expressed sequence tags." *Nat. Genet.* **4**, 332–333 (1993).
19. Hillier, L.D. *et al.* Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6**, 807–828 (1996).
20. Sorek, R., Ast, G. & Graur, D. Alu-containing exons are alternatively spliced. *Genome Res.* **12**, 1060–1067 (2002).
21. Morse, D.P., Aruscavage, P.J. & Bass, B.L. 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. USA* **99**, 7906–7911 (2002).
22. Maas, S. *et al.* Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J. Biol. Chem.* **271**, 12221–12226 (1996).
23. Polson, A.G. & Bass, B.L. Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J.* **13**, 5701–5711 (1994).
24. Lehmann, K.A. & Bass, B.L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* **39**, 12875–12884 (2000).
25. Kim, U., Wang, Y., Sanford, T., Zeng, Y. & Nishikura, K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc. Natl. Acad. Sci. USA* **91**, 11457–11461 (1994).
26. Higuchi, M. *et al.* RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* **75**, 1361–1370 (1993).
27. Wong, S.K., Sato, S. & Lazinski, D.W. Substrate recognition by ADAR1 and ADAR2. *RNA* **7**, 846–858 (2001).
28. Lei, M., Liu, Y. & Samuel, C.E. Adenovirus VAI RNA antagonizes the RNA-editing activity of the ADAR adenosine deaminase. *Virology* **245**, 188–196 (1998).
29. Tonkin, L.A. & Bass, B.L. Mutations in RNAi rescue aberrant chemotaxis of ADAR mutants. *Science* **302**, 1725 (2003).
30. Ausubel, F.M. *et al. Current Protocols in Molecular Biology* (John Wiley & Sons, Inc., New York, 1987).
31. Jiang, R. *et al.* Genome-wide evaluation of the public SNP databases. *Pharmacogenomics* **4**, 779–789 (2003).
32. Antonarakis, S.E., Krawczak, M. & Cooper, D.C. in *The Genetic Basis of Human Cancer*, edn. 2 (eds. Vogelstein, B. & Kinzler, K.W.) 7–41 (McGraw-Hill, New-York, 2002).