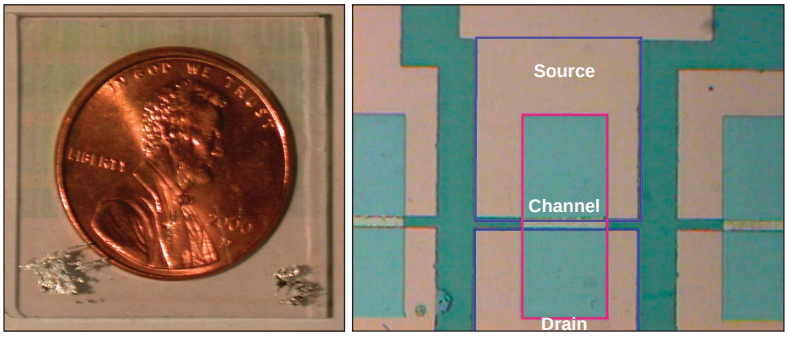
ported by Nomura *et al.* (*1*) meets the on-to-off current ratio constraint, it is unlikely that either the cost or temperature requirement can be met by this approach, given the use of single-crystal substrates, exotic processing, and high-temperature annealing. Perhaps the most technologically significant aspect of the transparent transistor of Nomura *et al.* (*1*) is that it proves the possibility of realizing transparent transistors with very large mobilities. This could be critical for high-end products such as next-generation projector displays.



ZnO-based transparent transistors. (**Left**) Patterned transparent transistor test structures are evident, upon close inspection, in the upper portion of the glass substrate, which sits on a penny. Solder contacts are present near the bottom corners of the glass. (**Right**) Enhanced-contrast, magnified image of a bottom-gate transparent transistor. The glass substrate is coated with a lower layer of indium–tin oxide (ITO), which serves as a gate, and an upper layer of aluminum–titanium oxide (ATO), which functions as the gate dielectric. Dark green regions are coated with ITO and ATO only. ITO source and drain contacts are shown in tan, and a ZnO channel layer in light green.

ZnO-based transparent transistors appear to be better suited to AMLCD select transistor applications. Hoffman *et al.* (*3*) and Carcia *et al.* (*4*) have demonstrated on-to-off current ratios greater than 10^6. Moreover, Carcia *et al.* (*4*) have fabricated ZnO transparent transistors with channel mobilities of more than 2 $cm^2\ V^{-1}\ s^{-1}$ in which the ZnO layer is deposited near room temperature with radio-frequency sputtering. Given that ZnO is inexpensive and radio-frequency sputtering is a proven technology used in commercial manufacturing, no obvious problems seem to preclude the use of ZnO transparent transistors for AMLCD applications.

Future applications of transparent transistors may include a transparent display, through which one could look as through a window. This could be realized by using a transparent electronics backplane of transistor drivers. However, this application requires a transparent optical source. This would necessitate advances in transparent opto-electronics if an inorganic light source is to be employed, although initially, organic or polymer light-emitting devices will certainly be used. In this manner, application needs will drive the development of new materials, devices, and circuits. On the other hand, the availability of new materials or devices will create new application possibilities. In the field of transparent electronics, the limits of opportunity are now difficult to see—after all, the constituent devices are invisible!

References and Notes

1. K. Nomura *et al.*, *Science* **300**, 1269 (2003).
2. S. Masuda *et al.*, *J. Appl. Phys.* **93**, 1624 (2003).
3. R. L. Hoffman, B. J. Norris, J. F. Wager, *Appl. Phys. Lett.* **82**, 733 (2003).
4. P. F. Carcia, R. S. McLean, M. H. Reilly, G. Nunes, *Appl. Phys. Lett.* **82**, 1117 (2003).
5. I thank R. Hoffman, D. Keszler, and D. Hong for helpful discussions. Supported by NSF grant DMR-0071727 and Army Research Office grant MURI E-18-667-G3.

GENOMICS

Not Junk After All

Wojciech Makałowski

From bacteria to mammals, the DNA content of genomes has increased by about three orders of magnitude in just 3 billion years of evolution (*1*). Early DNA association studies showed that the human genome is full of repeated segments, such as Alu elements, that are repeated hundreds of thousands of times (*2*). The vast majority of a mammalian genome does not code for proteins. So, the question is, "Why do we need so much DNA?" Most researchers have assumed that repetitive DNA elements do not have any function: They are simply useless, selfish DNA sequences that proliferate in our genome, making as many copies as possible. The late Sozumu Ohno coined the term "junk DNA" to describe these repetitive elements. On page 1288 of this issue, Lev-Maor and colleagues (*3*) take junk DNA to new heights with their analysis of how Alu elements in the introns of human genes end up in the coding exons, and in so doing influence evolution.

Although catchy, the term "junk DNA" for many years repelled mainstream researchers from studying noncoding DNA. Who, except a small number of genomic clochards, would like to dig through genomic garbage? However, in science as in normal life, there are some clochards who, at the risk of being ridiculed, explore unpopular territories. Because of them, the view of junk DNA, especially repetitive elements, began to change in the early 1990s. Now, more and more biologists regard repetitive elements as a genomic treasure (*4, 5*). Genomes are dynamic entities: New functional elements appear and old ones become extinct. It appears that transposable elements are not useless DNA. They interact with the surrounding genomic environment and increase the ability of the organism to evolve. They do this by serving as recombination hotspots, and providing a mechanism for genomic shuffling and a source of "ready-to-use" motifs for new transcriptional regulatory elements, polyadenylation signals, and protein-coding sequences. The last of these is especially exciting because it has a direct influence on protein evolution.

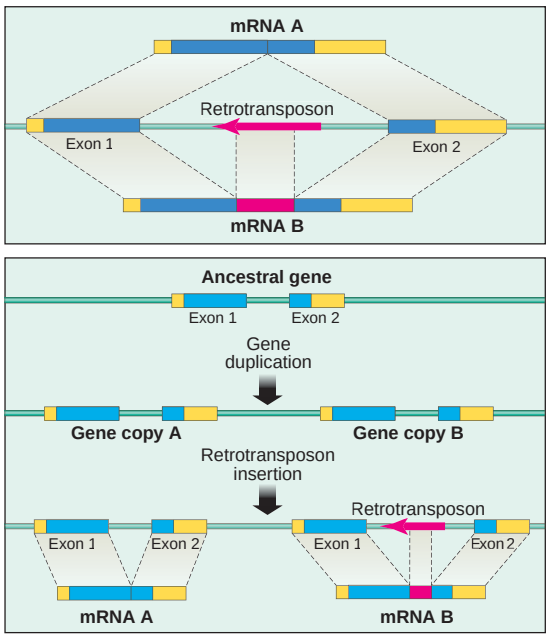
More than a decade ago, Mitchell *et al.* showed that a point mutation in an Alu element residing in the third intron of the ornithine aminotransferase gene activated a cryptic splice site, and consequently led to the introduction of a partial Alu element into an open reading frame (*6*). The in-frame stop codon carried by the Alu element resulted in a truncated protein and ornithine aminotransferase deficiency. This discovery led to the hypothesis that a similar mechanism may result in fast evolutionary changes in protein structure and increased protein variability (*7*). Several genome-wide investigations have shown that all types of mobile elements in all vertebrate genomes can be used in this way. The unsolved mystery is how a genome adapts to the drastic changes conferred on a protein by the insertion of a mobile element into the coding region of its gene. Lev-Maor and co-workers and a second group now demonstrate how this process takes place without disturbing the function of the original protein (see the figure) (*3, 8*).

The author is at the Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, PA 16802, USA. E-mail: wojtek@psu.edu

Last year, Sorek *et al*. (*9*) noticed that about 5% of alternatively spliced internal exons in the human genome originate in an Alu sequence. Interestingly, because Alu elements are primate specific, these exons must be primate or human specific as well as much younger than other exons in a gene. Additionally, they noticed that the vast majority of "Alu exons" are alternatively spliced (that is, there is always another messenger RNA without the Alu element in the coding region). They concluded that "Alu elements have the evolutionary potential to enhance the coding capacity and regulatory versatility of the genome without compromising its integrity" (*9*).

In their new work, this group now shows how alternative splicing of Alu exons is regulated (*3*). It is well established that the precise selection of the 3′ splice site depends on the distance between the branch point site (BPS) and the AG dinucleotide downstream of the BPS. The optimal distance between the BPS and the AG dinucleotide is relatively narrow (19 to 23 nucleotides). Interestingly, if there is another AG dinucleotide closer to the BPS, it will be recognized by a spliceosome even if a second AG located more optimally is used in the transesterification reaction (*10*). A splicing factor, hSlu7, is required to facilitate recognition of the correct AG. Thus, the correct selection of the 3′ splice site is an interplay between AG dinucleotides and certain splicing factors.

It is even more tricky to maintain the delicate balance of signals that cause an exon to be spliced alternatively—you make one mistake (a point mutation) and either a splicing signal becomes too strong and an exon is spliced constitutively, or the signal becomes too weak and an exon is skipped. Lev-Maor and colleagues (*3*) performed a series of experiments to identify an ideal sequence signal surrounding the 3′ splice site within the Alu element that kept the Alu element alternatively spliced. It appears that in addition to the distance between two AG dinucleotides, a nucleotide immediately upstream of proximal AG is also important. Hence, a proximal GAG sequence serves as a signal weak enough to create an alternatively spliced Alu exon. Any mutation of a proximal GAG in the first position results in a constitutive Alu exon. This is an important observation because most of the more than 1 million Alu elements populating the human genome contain such a potential 3′ splice site. Of these, 238,000 are located within introns of protein-coding genes, and each one can become an exon. Unfortunately, most mutations will lead to abnormal proteins and are likely to result in disease. Yet a small number may create an evolutionary novelty, and nature's "alternative splicing approach" guarantees that such a novelty may be tested while the original protein stays intact.



Junk DNA caught in the act. Two ways in which a repetitive DNA element, such as an Alu element, can be incorporated into the coding region of a gene without destroying the gene's function. (**Top**) A TE-cassette is inserted into the mRNA as an alternative exon. (**Bottom**) Insertion of a TE-cassette is preceded by a gene duplication. In both cases, the genome gains two forms of the mRNA transcript—one with and one without the TE-cassette.

Another way to exploit an evolutionary novelty without disturbing the function of the original protein is gene duplication (see the figure). Gene duplication is one of the major ways in which organisms can generate new genes (*11*). After a gene duplication, one copy maintains its original function whereas the other is free to evolve and can be used for "nature's experiments." Usually, this is accomplished through point mutations and the whole process is very slow. However, recycling some modules that already exist in a genome (for example, in transposons) can speed up the natural mutagenesis process tremendously. Several years ago, Iwashita and colleagues discovered a bovine gene containing a piece of a transposable element (called a TE-cassette) in the middle of its open reading frame (*12*). This cassette contributes a whole new domain to the bovine BCNT protein, namely an endonuclease domain native to the ruminant retrotransposable element–1 (RTE-1). Interestingly, the human and mouse homologs of bovine BCNT lack the endonuclease domain but instead contain a different one at their carboxyl terminus. This raised two questions: When did the BCNT protein acquire the endonuclease domain, and how did the bovine genome manage such a drastic rearrangement of BCNT without losing its fitness? Iwashita *et al*. give the answers to both questions in their new study (*8*). They discovered another copy of the bovine *bcnt* gene that resembles mammalian *bcnt* homologs (also called *CFDP1*) just six kilobases downstream of the gene with the TE-cassette. Both copies of the gene are apparently expressed and both proteins are functional. Phylogenetic analysis suggests that shortly after gene duplication in the ruminant lineage, one of the copies acquired an endonuclease domain from an RTE-1 retrotransposon. Not surprisingly, this gene undergoes accelerated evolution.

The reports by Lev-Maor *et al*. and Iwashita and colleagues describe different ways in which genes can be rapidly rearranged and acquire evolutionary novelty through the use of so-called junk DNA. These discoveries wouldn't be so exciting if they didn't show how genomes achieve this without disturbing an original protein. To quote an old Polish proverb: "A wolf is sated and a lamb survived." These two papers demonstrate that repetitive elements are not useless junk DNA but rather are important, integral components of eukaryotic genomes. Risking personification of biological processes, we can say that evolution is too wise to waste this valuable information. Therefore, repetitive DNA should be called not junk DNA but a genomic scrapyard, because it is a reservoir of ready-to-use segments for nature's evolutionary experiments (*13*).

References

1. M. Nei, *Nature* **221**, 40 (1969).
2. R. Britten, D. Kohne, *Science* **161**, 529 (1968).
3. G. Lev-Maor, R. Sorek, N. Shomron, G. Ast, *Science* **300**, 1288 (2003).
4. J. Brosius, *Science* **251**, 753 (1991).
5. R. Nowak, *Science* **263**, 608 (1994).
6. G. A. Mitchell *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **88**, 815 (1991).
7. W. Makalowski, G. A. Mitchell, D. Labuda, *Trends Genet.* **10**, 188 (1994).
8. S. Iwashita *et al*., *Mol. Biol. Evol.*, in press.
9. R. Sorek, G. Ast, D. Graur, *Genome Res.* **12**, 1060 (2002).
10. K. Chua, R. Reed, *Mol. Cell. Biol.* **21**, 1509 (2001).
11. S. Ohno, *Evolution by Gene Duplication* (Springer-Verlag, New York, 1970).
12. T. Nobukuni *et al*., *J. Biol. Chem.* **272**, 2801 (1997).
13. W. Makalowski, *Gene* **259**, 61 (2000).

activation threshold, but also enables the channel to be dynamically modulated by inflammatory products that activate PLC. Finally, it is interesting to note that the C-terminal domain of TRPV3, a warm-sensitive channel with an activation threshold of ~35°C (*21–23*), conspicuously lacks a region corresponding to the 777–792 domain of TRPV1, the minimal essential core of the predicted PIP_2 binding site. Thus, modification of this PIP_2 regulatory domain by genetic, biochemical, or pharmacological mechanisms may have profound effects on sensitivity of primary afferent nerve fibers to chemical and thermal stimuli under normal or pathological conditions.

References and Notes

1. D. W. Hilgemann, S. Feng, C. Nasuhoglu, *Sci. STKE* **2001**, RE19 (2001).
2. R. C. Hardie, *Annu. Rev. Physiol.* **65**, 735 (2003).
3. H. H. Chuang *et al.*, *Nature* **411**, 957 (2001).
4. M. Tominaga, M. Wada, M. Masu, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6951 (2001).
5. L. W. Runnels, L. Yue, D. E. Clapham, *Nature Cell Biol.* **4**, 329 (2002).
6. M. J. Caterina *et al.*, *Nature* **389**, 816 (1997).
7. M. Tominaga *et al.*, *Neuron* **21**, 531 (1998).
8. M. J. Caterina *et al.*, *Science* **288**, 306 (2000).
9. J. B. Davis *et al.*, *Nature* **405**, 183 (2000).
10. E. D. Prescott, D. Julius, unpublished data.
11. H. Zhang, C. He, X. Yan, T. Mirshahi, D. E. Logothetis, *Nature Cell Biol.* **1**, 183 (1999).
12. C. L. Huang, S. Feng, D. W. Hilgemann, *Nature* **391**, 803 (1998).
13. E. Kobrinsky, T. Mirshahi, H. Zhang, T. Jin, D. E. Logothetis, *Nature Cell Biol.* **2**, 507 (2000).
14. C. Wiedemann, T. Schafer, M. M. Burger, *EMBO J.* **15**, 2094 (1996).
15. V. Vellani, S. Mapplebeck, A. Moriondo, J. B. Davis, P. A. McNaughton, *J. Physiol.* (*London*) **534**, 813 (2001).
16. L. S. Premkumar, G. P. Ahern, *Nature* **408**, 985 (2000).
17. M. Numazaki, T. Tominaga, H. Toyooka, M. Tominaga, *J. Biol. Chem.* **277**, 13375 (2002).
18. P. M. Zygmunt *et al.*, *Nature* **400**, 452 (1999).
19. T. Hofmann *et al.*, *Nature* **397**, 259 (1999).
20. S. E. Jordt, D. Julius, *Cell* **108**, 421 (2002).
21. A. M. Peier *et al.*, *Science* **296**, 2046 (2002).
22. G. D. Smith *et al.*, *Nature* **418**, 186 (2002).
23. H. Xu *et al.*, *Nature* **418**, 181 (2002).
24. Materials and methods, fig. S1, and associated references are available as supporting material on *Science* Online.

25. We thank R. Nicoll and D. Minor for many helpful suggestions and encouragement throughout this project, G. Reid for help with the Peltier system, and members of our laboratory for advice. This work was supported by predoctoral fellowships from the NSF and the University of California, San Francisco, Chancellor's Fund (E.D.P.) and by the NIH (D.J.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/300/5623/1284/DC1
Materials and Methods
Fig. S1
References

19 February 2003; accepted 15 April 2003

The Birth of an Alternatively Spliced Exon: 3′ Splice-Site Selection in *Alu* Exons

Galit Lev-Maor,[1]* Rotem Sorek,[1,2]* Noam Shomron,[1] Gil Ast[1]†

Alu **repetitive elements can be inserted into mature messenger RNAs via a splicing-mediated process termed exonization. To understand the molecular basis and the regulation of the process of turning intronic *Alu*s into new exons, we compiled and analyzed a data set of human exonized *Alu*s. We revealed a mechanism that governs 3′ splice-site selection in these exons during alternative splicing. On the basis of these findings, we identified mutations that activated the exonization of a silent intronic *Alu*.**

Alu elements are short (about 300 nucleotides in length), interspersed elements that amplify in primate genomes through a process of retroposition (*1–3*). These elements have reached a copy number of about 1.4 million in the human genome, composing more than 10% of it (*4*). A typical *Alu* is a dimer, built of two similar sequence elements (left and right arms) that are separated by a short A-rich linker. Most *Alu*s have a long poly-A tail of about 20 to 100 bases (*5*).

Parts of *Alu* elements, predominantly on their antisense orientation, can be inserted into mature mRNAs by way of splicing ("exonization"). Presumably, the exonization process is facilitated by sequence motifs that resemble splice sites, which are found within the *Alu* sequence (*6–9*) (see fig. S1 for a model of exonization). Because *Alu*s are found in primate genomes only, *Alu*-derived exons might contribute to some of the characteristically unique features of primates.

We have previously shown that more than 5% of human alternatively spliced exons are *Alu*-derived and that most, if not all, *Alu*-containing exons are alternatively spliced (*9*). We therefore hypothesized that mutations causing a constitutive splicing of intronic *Alu*s would cause genetic diseases, and indeed we found in the literature several instances in which a constitutive *Alu* insertion caused a genetic disorder (*10–12*).

To study the alternative splicing regulation of exonized *Alu*s, we compiled a data set of exonized *Alu*s from the human genome. An analysis of this data set revealed that two positions along the inverted *Alu* sequence are most commonly used as 3′ splice sites (3′SSs) in *Alu* exonizations: position 279 ("proximal AG") and position 275 ("distal AG"). The relationships between two near AGs in a 3′SS were well characterized previously in the context of constitutive splicing (*13*, *14*). To pinpoint the sequence determinants by which the spliceosome selects one of the two possible AGs in the context of alternative splicing, we aligned the exonized *Alu*s that use either of these AGs to their ancestor.

The 3′SS regions of these instances are shown in Fig. 1. Figure 1 also shows that the proximal AG is selected mostly in exonized *Alu*s of S subfamilies (9 times out of 13), whereas the distal AG is mainly selected in exonized *Alu*s belonging to J subfamilies (12 times out of 16). This differential usage of AG selection in *Alu* subfamilies is probably because of the polymorphism between the J and S subfamilies in position 277 (Fig. 1, colored yellow), which eliminates the distal AG in *Alu*s of the S subfamilies. As a result, the proximal AG is selected. Although another polymorphism at position 275 creates a new distal AG in the S subfamilies, this new AG is six nucleotides downstream from the proximal AG, a distance that was shown to be out of the effective range for selecting a distal AG in constitutive splicing (*14*). Indeed, the cases where *Alu*s of the S subfamilies used the distal AG required mutations that shortened the distance between AGs back to four nucleotides (Fig. 1, colored green). This indicates that when the range between the two AGs is four nucleotides or less the distal AG is preferred and when the distance is six nucleotides or more the proximal is preferred.

However, in five cases (Fig. 1, rows 25 to 29), the proximal AG was selected, even though a distal AG existed less than six nucleotides in range; in all these cases, the G in position –7 (colored purple) was mutated to either A (two cases) or T (three cases). Remarkably, a mutation in the same position in intron 5 of the COL4A3 gene leads to exonization of a silent intronic *Alu*. This *Alu* exon is constitutively spliced, resulting in an Alport syndrome phenotype (*10*). This implies that the G in position –7 suppresses the selection of the proximal AG, causing a shift toward selection of the distal AG. When this G is mutated, the proximal AG is preferred.

[1]Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel. [2]Compugen, 72 Pinchas Rosen Street, Tel Aviv 69512, Israel.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: gilast@post.tau.ac.il

This is supported by the finding that GAG triplets at ends of introns are poorly cleaved in vitro and extremely rare in vivo (*15*).

To examine the above hypotheses, we cloned a minigene of the ADAR2 gene (adenosine deaminase, involved in RNA editing) (Fig. 1, row 1). Previously, exon 8 [denoted as exon 5a in (*16*)] of this gene was found to be an alternatively spliced *Alu*-derived exon, adding 40 amino acids in frame to the protein (*17*). In this exon, the distal AG is used as the 3′SS. Trying to characterize the relationship between proximal and distal AGs in the context of alternative splicing, we generated a set of mutations within the 3′SS.

Whereas the *Alu* exon in the wild-type ADAR2 was included in 40% of the transcripts (Fig. 2B, lane 3), replacement of the G in position –7 to A, U, or C (Fig. 2B, lanes 10 to 12) had two effects. First, as predicted from Fig. 1, the replacement shifted the selection from the distal AG to the proximal one. Second, the replacement resulted in a shift from alternative splicing of the *Alu* exon toward a nearly constitutive inclusion of the exon in the mature transcript. Our results point to the important role of the G in position –7 in shifting the selection toward the distal AG, thus maintaining the alternative splicing of the *Alu*-containing exon. Mutation of that G will likely result in a constitutive inclusion of the *Alu* exon and thus might cause a disease, as occurs in the case of Alport syndrome (*10*).

To check whether the proximal AG affects the selection of the distal AG, we mutated the proximal AG to UC or GA (Fig. 2B, lanes 8 and 9, respectively). The GA mutation resulted in a higher ratio of exon inclusion, reaching more than 85% inclusion instead of 40% in the wild type. The UC mutation caused the splicing of the *Alu* exon to become constitutive, possibly because it strengthened the polypyrimidine tract (PPT) that was originally 18 bases long (on average, the PPT length in exonized *Alu*s was 19 bases ±3). These findings indicate that the proximal AG presumably weakens the selection of the distal AG and is therefore required for maintaining alternative rather than constitutive splicing of the *Alu* exon. To summarize, when the distal 3′SS is used, the G at position –7 suppresses the selection of the proximal AG, and the proximal AG maintains the alternative splicing.

We then sought to understand whether the nucleotide composition between the two adjacent AGs affects 3′SS selection and ratio of alternative splicing. The two AGs are separated by an AC dinucleotide (Fig. 2A). A deletion of both these nucleotides (position –3 and –4) or only the C (Fig. 2B, lanes 5 and 7) resulted in an exon skipping, pointing to the importance of the C in position –3. Deletion or mutations of the A in position –4 to G or C changed the ratio between the two isoforms (Fig. 2B, lanes 6, 13, and 14). This indicates that position –4 also affects the inclusion ratio.

To test whether increased distance between the two AGs shifts the selection toward the proximal AG, we introduced additional nucleotides between the two AGs (Fig. 3A). Increasing the distance between the proximal and distal 3′SS to six or eight nucleotides resulted in *Alu* exon skipping (Fig. 3B, lanes 7 and 8). However, when the distance between the two AGs grew to 10 nucleotides, a residual exon inclusion was recovered in a little more than 3% of the spliced transcripts (Fig. 3B,

Gene name	Alu exon number	subfam	position relative to distal 3'ss	290	289	288	287	286	285	284	283	282	-7 281	-6 280	-5 279	-4 278	-3 277	-2 276	-1 275	274	273	272	271
		J		T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T
		S		T	T	T	T	T	T	T	T	G	A	G	A	C	G	G	A	G	T	C	T
1	ADAR2	8	Jb	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T
2	TFB2M	4	Jb	T	C	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T
3	MVK	4	Jb	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T
4	CBFA2T2	3	Jo	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	T
5	NPD002	7	Jb	T	T	T	C	T	T	T	T	G	A	G	A	C	A	G	G	G	T	C	C
6	MOG	3	Jb	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	T	G	T	C	T
7	n/a	5	Jb	T	T	T	A	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T
8	PTGES	2	Jb	T	T	T	T	T	T	T	G	G	A	G	A	C	A	G	A	G	T	C	T
9	DAF*	10	Jb	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	G	T	T	C	T
10	STK2*	n/a	Jb	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	C	T
11	MLANA	4	Jb	T	T	T	T	T	C	T	T	G	A	A	A	C	A	G	G	A	A	A	T
12	n/a	24	Jb	T	T	T	T	T	T	T	T	G	A	A	A	C	A	G	C	G	T	C	T
13	ITGB1*	7	Sx	T	T	T	A	T	T	T	T	G	A	G	A	C	A	G	-	-	T	C	T
14	n/a	2	Sg	T	C	T	T	T	T	T	T	T	T	G	A	C	A	G	A	G	T	C	C
15	MBD3	12	Sx	T	T	T	T	T	T	T	T	G	T	G	A	C	A	G	A	G	T	C	T
16	CNN2	6	Sx	T	T	T	A	T	T	T	T	G	A	G	A	T	A	G	G	A	T	C	T
17	PGT	12	Sp	T	T	T	T	T	T	T	T	G	A	G	A	C	G	G	A	G	T	T	T
18	n/a	2	Sg	C	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T
19	RES4-22	18	Sg	T	T	T	A	T	T	T	C	G	A	G	A	T	G	G	A	G	T	T	T
20	LOC51193	5	Sg	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T
21	n/a	3	Sx	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	C
22	CHRNA3*	6	Sx	T	T	T	T	T	T	T	T	G	A	G	A	T	G	G	A	G	T	C	T
23	PTD011	2	Sx	T	T	T	T	T	T	T	T	G	A	G	A	C	G	C	C	C	A	G	G
24	HCA66	18	Sg	T	T	T	T	T	T	T	T	T	A	G	A	C	G	G	A	G	T	C	T
25	CYP3A43	8	Sg	T	T	T	T	T	T	T	T	T	A	G	A	C	A	G	A	G	T	C	T
26	LCAT*	6	Jo	T	T	T	T	T	T	T	G	T	A	G	A	G	A	C	A	G	G	G	T
27	KIAA1169	24	Jo	T	T	T	T	G	T	T	T	T	A	G	A	G	A	T	G	G	T	A	T
28	SLC3A2	6	Jb	T	T	G	T	T	T	T	T	A	A	G	A	C	A	G	C	A	T	T	T
29	ICAM2	0	Jb	T	T	T	G	T	T	T	T	A	A	G	A	C	A	G	G	G	T	C	T
30	COL4A3	6	Sx	T	T	T	T	T	C	T	T	T	A	G	A	T	G	G	A	G	T	C	T
31	GUSB	9	Sg/x	A	T	T	T	T	T	T	G	G	A	T	A	T	G	C	A	G	T	C	T
32	OAT	4	Jo	T	T	T	T	T	T	T	T	G	A	G	A	C	A	G	A	G	T	T	T
													proximal AG				distal AG						

Fig. 1. The selection of AGs in the 3′SSs of *Alu*-derived exons. Alignment is shown for the region near the two most prevalent 3′SSs in the right arm of exonized *Alu* sequences (in the antisense orientation). Data for 29 exonized *Alu*s, compiled from the results of our previous study (*9*) as well as newly collected data from the literature (*22*–*26*), are shown. The 20 nucleotides presented are positions 290 to 271 in the *Alu* sequence, according to the numbering in (*27*). The two possible AG dinucleotides (distal and proximal to the PPT) are marked in red. The selected AG dinucleotide, defining the end of the intron, is underlined for each exonized *Alu*. Selected AG dinucleotides were inferred with the use of alignments of expressed sequences to the human genome (*9*) (table S1). Those marked by an asterisk next to the gene name are additional *Alu* exons found in the literature scan (*22*–*26*). Consensus sequences of subfamilies S and J appear in the first two rows, with positions differing between subfamilies marked in yellow. Rows 30 to 32 represent the 3′SSs of *Alu* sequences whose constitutive exonization was shown to cause a genetic disease [Alport syndrome (COL4A3), Sly syndrome (GUSB), and OAT deficiency (OAT)]. The mutation causing Alport syndrome is marked light blue (position –7 G to T); exonization in Sly syndrome and OAT deficiency resulted from mutations in the 5′SS. Numbers on top mark the position relative to the distal 3′SS as referred to in this article. Gene names are as given in RefSeq conventions, the *Alu* exon number is the serial number of the *Alu*-containing exon in the gene, and the subfam is the *Alu* subfamily type, inferred with the use of RepeatMasker (*28*).

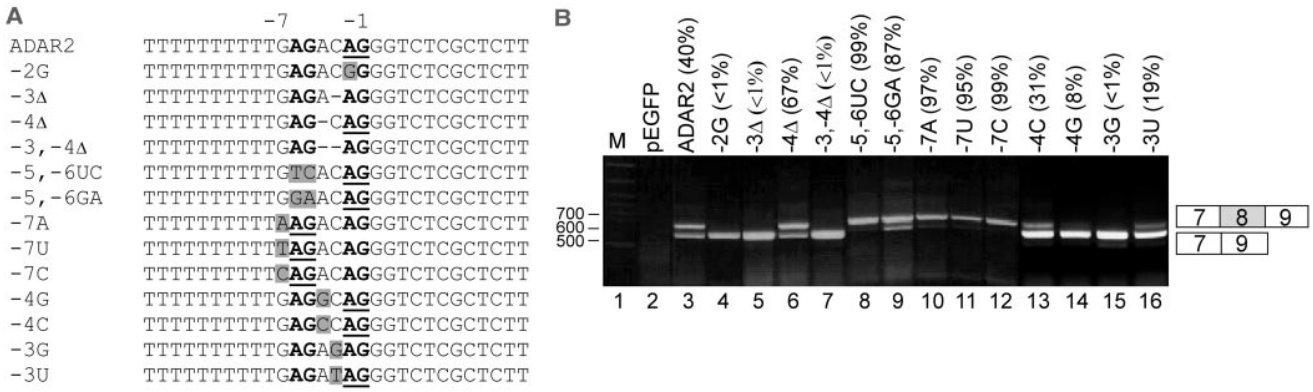


Fig. 2. Splicing assays on ADAR2 minigene mutants. (**A**) The 3′SS sequence of the wild type and 13 mutants of ADAR2. The proximal and distal AGs are in bold, mutations are shaded, and the selected AGs are boldfaced and underlined. (**B**) The indicated plasmid mutants were introduced into 293T cells by transfection, total cytoplasmic RNA was extracted, and splicing products were separated in 2% agarose gel after reverse transcription polymerase chain reaction (RT-PCR) (*18*). Lane 1, DNA size marker; lane 2, vector only (pEGFP); lane 3, splicing products of wild-type (wt) ADAR2; and lanes 4 to 16, splicing products of mutated ADAR2 minigenes, corresponding to the sequences in (A). The two possible minigene mRNA isoforms are shown on the right. Numbers in parentheses indicate percentages of the *Alu*-containing mRNA isoform as determined by quantified RT-PCR (100% corresponds to the total of both mRNA isoforms). Identical results were also obtained with HeLa cells.

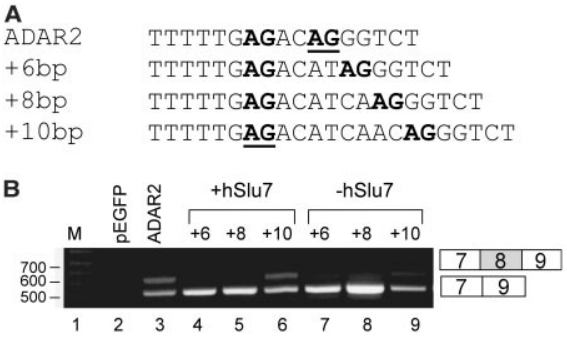


Fig. 3. The effect of hSlu7 on AG selection. (**A**) The sequence of the 3′SS of ADAR2 and the three insertion mutants. Both potential AGs are marked in bold, and the selected AG is boldfaced and underlined. bp, base-pair (**B**) The indicated plasmid mutants were treated as described for Fig. 2B (*18*). Lanes 4 to 6 represent cotransfection of the insertion mutants with a plasmid expressing hSlu7; lanes 7 to 9 represent the insertion mutants without additional hSlu7.

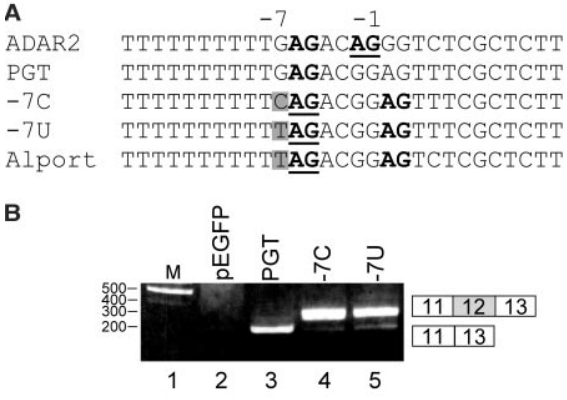


Fig. 4. Splicing assays on wt and mutated PGT. (**A**) Sequences from top to bottom are as follows: wt *Alu* 3′SS of ADAR2; wt *Alu* 3′SS of PGT; mutant *Alu* 3′SS of PGT; and mutant sequence of COL4A3, which causes Alport syndrome. Both potential AGs are marked in bold, and the selected 3′SS is boldfaced and underlined. The mutated position is shaded. (**B**) Transfection was performed in HT1080 cell lines. Total RNA and RT-PCR was performed as mentioned in Fig. 2B (*18*). Lane 1, DNA size marker; lane 2, vector only (pEGFP); lane 3, splicing product of wt PGT; lanes 4 and 5, splicing products of mutated PGT minigenes, corresponding to the sequences in (A). The two possible minigene mRNA isoforms are shown on the right. The results were reproducible in 293T cell lines (*19*) as well.

lane 9). In these transcripts, the proximal AG was selected even though it was preceded by G (Fig. 3A).

We further examined whether hSlu7 (human synergistic lethal with U5 small nuclear RNA), a second-step splicing factor, might be involved in the activation of the proximal AG. This protein is known to be required for correct AG identification when more than one possible AG exists in the 3′SS region (*13*). Cotransfection of 293T cells with plasmids containing the insertion mutants and with hSlu7 (10-fold higher than endogenous hSlu7 concentrations) (*18*) led to an increase in the selection of the proximal AG by 10-fold, reaching 32% inclusion when the distance between the proximal and distal AGs was 10 bases (Fig. 3B, lane 6). Presumably, hSlu7 activation of the weak splice site may depend on the existence of a distal AG, because elimination of the distal AG (mutant –2G, Fig. 2) resulted in an exon skipping that was not reversed by increasing the concentration of hSlu7 (*19*). These results propose that the distal AG can affect the selection of the proximal one negatively when the proximal is preceded by a G nucleotide. The proximal 3′SS can be selected when hSlu7 is present, and the efficiency of this selection is increased when the distal AG is found far enough from the splice site (in our case, 10 nucleotides into the exon). This observation, therefore, indicates that activation of the weak 3′SS (GAG) depends on hSlu7 concentration and suggests a possible role for hSlu7 concentration in alternative-splicing regulation.

Rows 17 to 22 in Fig. 1 show instances in which the proximal AG is selected even though the distal AG is found six nucleotides downstream. However, the +6 base-pair mutant (Fig. 3B, lane 6) resulted in a total exon skipping. The above results suggest that these exonization instances might occur with high hSlu7 concentrations within certain cell types or with high local concentration of hSlu7 within the subregion of the nucleus. From this, we further assumed that in normal conditions these *Alu* exons would be skipped. We therefore chose one of these genes, one encoding a putative glucosyltransferase (PGT) (Fig. 1, row 17), and cloned a minigene of its exons 11 to 13, including the introns in between (the *Alu* exon being exon 12). Indeed, when the PGT minigene was transfected into HT1080 and 293T cell lines, only a single mRNA isoform appeared, corresponding to *Alu*-exon skipping (Fig. 4B, lane 3). Repeating the same experiment with the use of endogenous PGT mRNA also showed *Alu*-exon skipping (*19*).

To test if, as predicted from our results, a mutation in position –7 of a completely silent intronic *Alu* element would result in exoniza-

tion, we mutated this position in the PGT minigene. As seen in Fig. 4B (lanes 4 and 5), this point mutation was enough to activate the nearly constitutive inclusion of the *Alu* exon in the mature transcript. As indicated above, the same mutation in the COL4A3 gene activates a constitutive exonization of a silent intronic *Alu*, resulting in Alport syndrome (*10*). To assess the importance of our findings, we analyzed the entire content of *Alu*s in the human genome and found that there are at least 238,000 antisense *Alu*s located within introns in the human genome (*20*). Of these, 52,935 *Alu*s carry a potential ADAR2-like 3′SS, and 23,012 carry a potential PGT-like 3′SS. Our results suggest that many of these silent intronic *Alu* elements might be susceptible to exonization by the same single point mutation and are thus under strict selective pressure. Such point mutations in human genomic antisense *Alu*s may, therefore, be the molecular basis for predisposition to so-far uncharacterized genetic diseases.

Because all *Alu*-containing exons are alternatively spliced (*9*), they add splice variants to our transcriptome while maintaining the original proteins intact. Exonized *Alu*s can, thus, acquire functionality and become exapted, i.e., adapted to a function different than their original (*21*). When the splicing of an *Alu* exon is constitutive, however, the transcript encoding to the original protein is permanently disrupted, which could provide the basis for a genetic disorder. Identification of genomic *Alu*s that are one point mutation away from exonization might therefore enable the screening for predisposition for genetic diseases that involve *Alu* exonization.

References and Notes

1. A. J. Mighell, A. F. Markham, P. A. Robinson, *FEBS Lett.* **417**, 1 (1997).
2. D. J. Rowold, R. J. Herrera, *Genetica* **108**, 57 (2000).
3. C. W. Schmid, *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 283 (1996).
4. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
5. A. M. Roy-Engel *et al.*, *Genome Res.* **12**, 1333 (2002).
6. W. Makalowski, G. A. Mitchell, D. Labuda, *Trends Genet.* **10**, 188 (1994).
7. W. Makalowski, *Gene* **259**, 61 (2000).
8. A. Nekrutenko, W. H. Li, *Trends Genet.* **17**, 619 (2001).
9. R. Sorek, G. Ast, D. Graur, *Genome Res.* **12**, 1060 (2002).
10. B. Knebelmann *et al.*, *Hum. Mol. Genet.* **4**, 675 (1995).
11. G. A. Mitchell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 815 (1991).
12. R. Vervoort, R. Gitzelmann, W. Lissens, I. Liebaers, *Hum. Genet.* **103**, 686 (1998).
13. K. Chua, R. Reed, *Nature* **402**, 207 (1999).
14. K. Chua, R. Reed, *Mol. Cell. Biol.* **21**, 1509 (2001).
15. R. E. Manrow, S. L. Berger, *J. Mol. Biol.* **234**, 281 (1993).
16. D. Slavov, K. Gardiner, *Gene* **299**, 83 (2002).
17. F. Lai, C. X. Chen, K. C. Carter, K. Nishikura, *Mol. Cell. Biol.* **17**, 2413 (1997).
18. Materials and methods are available as supporting material on *Science* Online.
19. G. Lev-Maor, R. Sorek, N. Shomron, G. Ast, unpublished data.
20. R. Sorek, R. Shalgi, G. Ast, D. Graur, unpublished data.
21. J. Brosius, S. J. Gould, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10706 (1992).
22. M. Mihovilovic *et al.*, *Biochem. Biophys. Res. Commun.* **197**, 137 (1993).
23. M. Miller, K. Zeller, *Gene* **190**, 309 (1997).
24. W. G. Cance, R. J. Craven, T. M. Weiner, E. T. Liu, *Int. J. Cancer* **54**, 571 (1993).
25. I. W. Caras *et al.*, *Nature* **325**, 545 (1987).
26. G. Svineng, R. Fassler, S. Johansson, *Biochem. J.* **330**, 1255 (1998).
27. J. Jurka, A. Milosavljevic, *J. Mol. Evol.* **32**, 105 (1991).
28. RepeatMasker is available online at http://repeatmasker.genome.washington.edu/cgi-bin/RepeatMasker.

29. We thank M. Kupiec for a critical reading; R. Reed for the hSlu7 plasmid; and also F. Belinky, R. Shalgi, T. Dagan, and E. Sharon for assistance in *Alu* data analysis. Supported by a grant from the Israel Science Foundation and, in part, by a grant from the Israel Cancer Association and the Indian-Israeli Scientific Research Corporation to G.A.

Supporting Online Material

www.sciencemag.org/cgi/content/full/300/5623/1288/DC1
Materials and Methods
Fig. S1
Table S1
References and Notes

21 January 2003; accepted 9 April 2003

Essential Role of Fkbp6 in Male Fertility and Homologous Chromosome Pairing in Meiosis

Michael A. Crackower,[1]*† Nadine K. Kolas,[2]* Junko Noguchi,[4] Renu Sarao,[1] Kazuhiro Kikuchi,[4] Hiroyuki Kaneko,[4] Eiji Kobayashi,[5] Yasuhiro Kawai,[6] Ivona Kozieradzki,[1] Rushin Landers,[1] Rong Mo,[7] Chi-Chung Hui,[7] Edward Nieves,[3] Paula E. Cohen,[2] Lucy R. Osborne,[8] Teiji Wada,[1] Tetsuo Kunieda,[6] Peter B. Moens,[9] Josef M. Penninger[1]‡

Meiosis is a critical stage of gametogenesis in which alignment and synapsis of chromosomal pairs occur, allowing for the recombination of maternal and paternal genomes. Here we show that FK506 binding protein (Fkbp6) localizes to meiotic chromosome cores and regions of homologous chromosome synapsis. Targeted inactivation of *Fkbp6* in mice results in aspermic males and the absence of normal pachytene spermatocytes. Moreover, we identified the deletion of *Fkbp6* exon 8 as the causative mutation in spontaneously male sterile *as/as* mutant rats. Loss of Fkbp6 results in abnormal pairing and misalignments between homologous chromosomes, nonhomologous partner switches, and autosynapsis of X chromosome cores in meiotic spermatocytes. Fertility and meiosis are normal in *Fkbp6* mutant females. Thus, Fkbp6 is a component of the synaptonemal complex essential for sex-specific fertility and for the fidelity of homologous chromosome pairing in meiosis.

Meiosis is a fundamental process in sexually reproducing species that allows genetic exchange between maternal and paternal genomes (*1*, *2*). Defects in high-fidelity meiotic chromosome alignment or in genome segregation in germ cells result in aneuploidies such as trisomy 21 in Down syndrome. Aneuploidy is a leading cause of spontaneous miscarriage in humans and a hallmark of many human cancer cells (*2*). Once homologs are paired, the chromosomes are connected by a specific structure: the synaptonemal complex (SC) (*3*). SCs are zipperlike structures assembled along the paired meiotic chromosomes during the prophase of the first meiotic division (*3*). Although SCs were first discovered more than 45 years ago (*4*, *5*), only very few structural meiosis-specific components of the SC have been identified in mammals, such as SC proteins 1, 2, and 3 [Scp1 (also known as Syn1/Sycp1), Scp2, and Scp3 (also known as Cor1)] (*3*). Genetic inactivation of the mouse *Scp3* gene results in male infertility due to a failure to form chromosome synapsis in meiotic prophase (*6*). Female *Scp3*$^{-/-}$ mice have reduced fertility, and embryos from *Scp3*$^{-/-}$ mothers have increased incidents of aneuploidy (*7*). To our

[1]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), c/o Dr. Bohrgasse 7, 1030, Vienna, Austria. [2]Department of Molecular Genetics, and [3]Department of Biochemistry, Laboratory for Macromolecular Analysis and Proteomics, Albert Einstein College of Medicine (AECOM), 1300 Morris Park Avenue, Bronx, NY 10461, USA. [4]Germ Cell Conservation Laboratory, National Institute of Agrobiological Sciences, Kannondai, Tsukuba, Ibaraki 305-8602, Japan. [5]National Livestock Breeding Center, Odakura, Nishigo, Fukushima 961-851, Japan. [6]Graduate School of Natural Science and Technology, Okayama University, Okayama 700-0082 Japan. [7]Program in Developmental Biology, The Hospital for Sick Children, Toronto, ON M5G 1X8, Canada. [8]Departments of Medicine and Molecular and Medical Genetics, University of Toronto, 1 King's College Circle, Toronto, ON M5S 1A8, Canada. [9]Department of Biology, York University, Toronto, ON M3J 1P3, Canada.

*These authors contributed equally to this work.
†Present address: Department of Protein Sciences, Amgen, Thousand Oaks, CA 91320, USA.
‡To whom correspondence should be addressed. E-mail: josef.penninger@oeaw.ac.at

Supporting Online Materials

Materials and Methods

Plasmid Constructs

Oligonucleotide primers were designed to amplify (from human genomic DNA) a mini-gene that contains exons 7, 8, and 9 of the adenosine deaminase (ADAR2) gene and exons 11, 12, and 13 of the putative glucosyltransferase gene (PGT). Each primer contained an additional extension encoding a restriction enzyme sequence. The PCR product of ADAR2 and PGT (2.2kb and 3kb, respectively) was restriction digested and inserted between the KpnI/BglII sites in the pEGFP-C1 vector (Clonthech). The hSlu7 cDNA was a kind gift from Robin Reed and was inserted as described above into pEGFP-C1.

Site-Directed Mutagenesis

Oligonucleotide primers containing the desired mutations were used to amplify a mutation-containing replica of the wild type mini-gene plasmid. The PCR products were treated with 12U DpnI restriction enzyme (New England Biolabs) for 1hr at 37^{o}C. 1-3μl of the DNA was transformed into *E.coli* DH5α strain, followed by colony-picking mini-prep and midi-prep extraction (GIBCO/BRL). All plasmids were confirmed by sequencing.

Transfection, RNA Isolation and RT-PCR Amplification

293T, HeLa and HT1080 cell lines were cultured in Dulbecco's Modification of Eagle Medium, supplemented with 4.5g/ml glucose (Biological Industries) and 10% fetal calf

serum, and cultured in 60mm dish under standard conditions at 37^0C with 5% CO_2. Cells were grown to 50% confluence, and transfection was performed using Metafectene (Biontex) with 10μg of plasmid DNA or using FuGENE6 (Roche) with 6μg of plasmid DNA. Cells were harvested after 48hr. Total cytoplasmic RNA was extracted using Tri Reagent (Sigma), followed by treatment with 1U RNase-free DNase (Promega). Reverse transcription (RT) was preformed on 2μg total cytoplasmic RNA for 1hr at 42°C, using a pEGFP-C1-specific reverse primer and 2U reverse transcriptase of avian myeloblastosis virus (A-AMV, Roche).

The spliced cDNA products derived from the expressed mini-genes were detected by PCR, using the pEGFP-C1-specific reverse primer and an exon 7- or 11-forward primer (ADAR2 and PGT respectively): Amplification was performed for 30 cycles, consisting of 1 min at 94°C, 45 sec at 61°C, and 1 min at 72°C. The products were resolved on 2% agarose gel and confirmed by sequencing. The level of mRNA of the house-keeping gene, Glycerol-3-phosphate dehydrogenase, was used as the internal control for each transfection.

Nonsense-Mediated Decay (NMD)

NMD might affect the concentrations of each of the isoforms after completion of mRNA splicing. We examined this point and found that the levels of the isoforms are unaffected by NMD. This was indicted by incubation of the transfected cells with 300 μg/ml puromycin (Sigma) for 4 hr before RNA collection (as described in ref. *S1*).

Real-Time PCR

The LightCycler PCR and detection system (Roche) was used for quantification of the PCR products. The PCR reaction for each cDNA was performed twice, using specific primers -- one amplified only the upper band (exons 7, 8, and 9) and the other amplified only the lower band (exons 7 and 9). The PCR mixture (Roche) contained Taq DNA polymerase, reaction mix (buffer, SYBR Green I dye, dNTPs with dUTP instead of dTTP, 13mM magnesium chloride) and 12.5 pmol of primers. The samples were run for 45 cycles of repeated 10 sec at 95^{o}C, 10 sec at 66^{o}C, and 10 sec at 72^{o}C. Another reaction was performed, using specific primers for the GAPDH gene, that was used as an endogenous expressed control.

Genomic *Alu*s analysis

*Alu*s in the human genome were scanned using the August 2002 release of RepeatMasker (http://repeatmasker.genome.washington.edu) that was run on the human genome, downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/H_sapiens) on March 2003. GenBank annotation files (gbk files) were used to retrieve the exon-intron borders of genes. *Alu*s found in the reverse orientation inside introns were examined for the existence of the two possible 3' splice sites: (i) ADAR2-like potential 3'SS, defined by p(Y)GAGACAG and (ii) PGT-like 3'ss, defined by p(Y)GAGACGGAG (with the second AG allowed to be in a distance of 6-8 nucleotides from the first AG). Poly pyrimidine tract (p(Y)) was defined by at least 15 nucleotides, 10 of which are T or C.

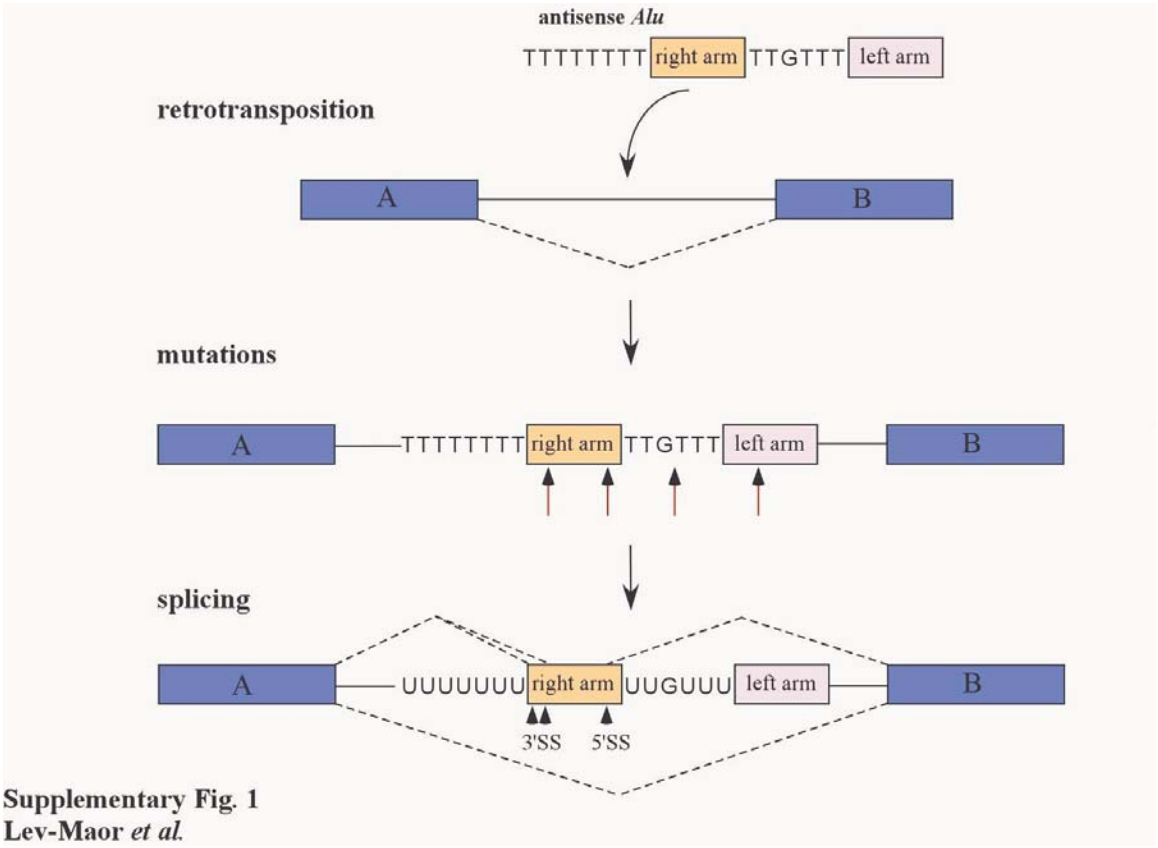


Figure S1:

Schematic model of *Alu* exonization. *Alu* is inserted into introns of primate genes by retrotransposition (upper panel). During evolution, mutations within pseudo splice-sites in the intronic *Alu* activate these sites (middle panel, marked by red arrows), and part of the *Alu* sequence is recognized as a new exon ('exonized'). Most exonizations involve the right arm of *Alu* on its antisense orientation, presumably because of the preceding long polyT that serves as a strong poly-pyrimidine tract necessary for the 3'SS recognition (lower panel). Although many possible pseudo 3'SS are found within *Alu* (*S2*), two of them (positions 279 and 275) are most commonly selected in *Alu* exonizations.

Table S1: Sequences supporting *Alu*-exon insertions*

	EST/RNA confirming exon insertion	Gene name
1	U76421	ADAR2
2	AA460397	TFB2M
3	AF217536	MVK
4	AF069747	CBFA2T2
5	BE898836	NPD002
6	U64570	MOG
7	AB007962	n/a
8	AF217965	PTGES
9	BG542576	DAF
10	n/a	STK2
11	AA071342	MLANA
12	HSM800877	n/a
13	n/a	ITGB1
14	AK021447	n/a
15	AA285195	MBD3
16	BE836938	CNN2
17	AA225691	PGT
18	U92992	n/a
19	AB000460	RES4-22
20	AW954573	LOC51193
21	AK024074	n/a
22	n/a	CHRNA3
23	BE747669	PTD011
24	AI949382	HCA66
25	AF280111	CYP3A43
26	BM558997	LCAT
27	BF087651	KIAA1169
28	AW381165	SLC3A2
29	BE261894	ICAM2

* Confirming accession numbers are presented for each of the genes presented in Figure 1 in the printed article. Row numbers follow the numbers in Figure 1.

References

S1. F. Pagani, *et al.*, *Nat. Genet.* **30**, 426 (2002).

S2. R. Sorek, G. Ast, D. Graur, *Genome Res.* **12**, 1060 (2002).