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.

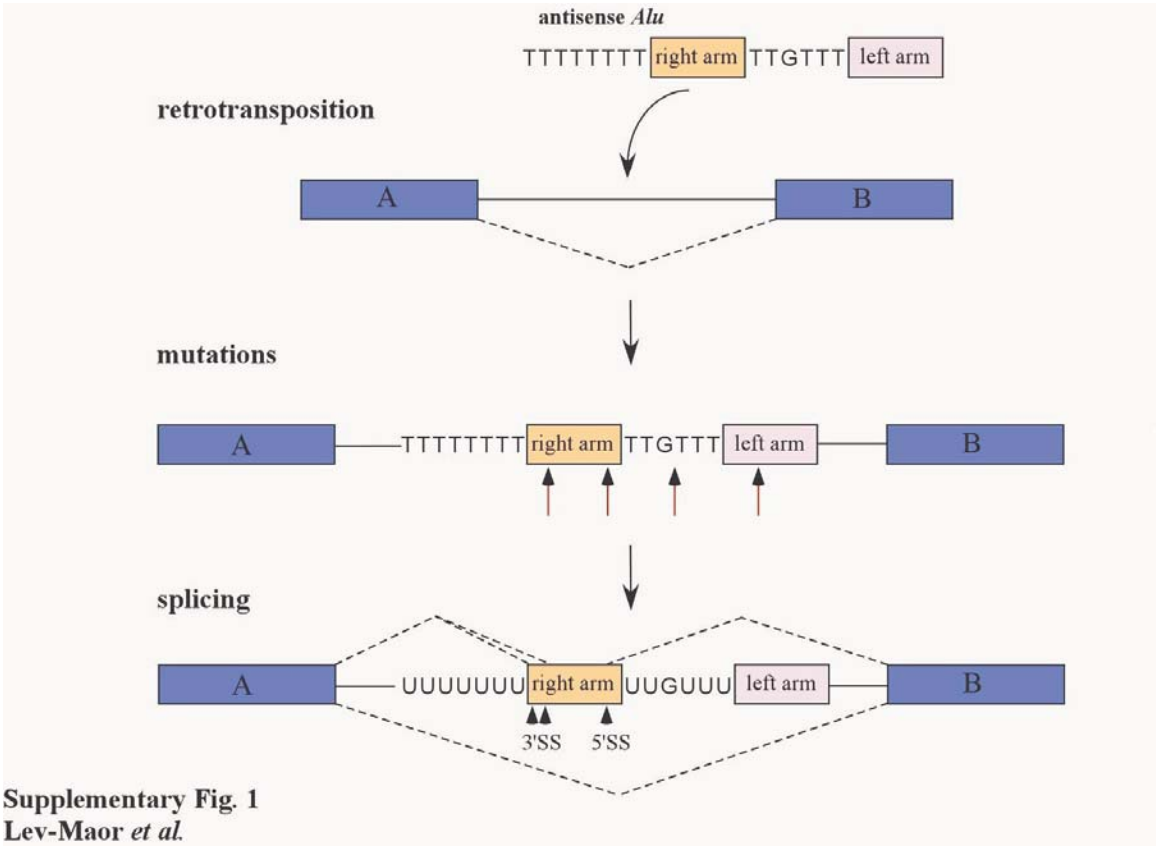

Figure S1:

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).