Symbiosis

www.symbiosisonline.org
www.symbiosisonlinepublishing.com

Research Article | International Journal of Structural and Computational Biology | Open Access

Transposable elements: a comparative study in the introns and UTRs of the homologous mitochondrial solute carrier genes of Human, Mouse and Zebrafish

Antonia Cianciulli, Rosa Calvello and Maria Antonietta Panaro*

Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, via Orabona, 4, I-70126 Bari, Italy

Received: March 20, 2017; Accepted: April 7, 2017; Published: April 25, 2017

***Corresponding author:** Maria Antonietta Panaro, Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, via Orabona, 4, I-70126 Bari, Italy; E-mail: mariaantonietta.panaro@uniba.it*

Abstract

We studied the localization of transposable elements (TEs) in the 5' and 3' flanking regions and in the introns of Mouse, Human and Zebrafish mitochondrial solute carrier genes. The canonical transcripts of these highly homologous genes exhibit superimposable patterns of exon-intron alternation in the three species. In introns, two sections approximately corresponding to the first and last 20 nucleotides are relatively, but not completely, depleted of TEs. The distributions of the distances of the TEs which are the closer to the start and stop codons are right-skewed, with lower values for short distances, a peak at 750 nt in Mouse and Human and 450 nt in Zebrafish, followed by an exponential decay. Taken together these results suggest that the exon/intron structuring of the mitochondrial solute carrier genes has been under strict evolutionary control from fish to mammals and variants with potentially dangerous inserts (such as LTR elements) nearing the regulatory sequences have been selected against, whilst the extant TEs do not exert any significant action not even at relatively short distances.

Keywords: Mitochondrial solute carrier genes; Transposable elements; UTRs; Introns; Human; Mouse; Zebrafish.

Introduction

Transposable elements (Transposons, TEs) are mobile segments of genetic material which may be found in DNA non-coding sections, either intergenic or intragenic (introns). TEs are widespread in the genomes of all Vertebrates representing a very variable, yet always significant, percentage of the genome [1]. TEs account for about 32% of the genome in Mouse, 45% in Human and 55 % in Zebrafish [1,2,3]. Their biological success is owed to their ability to reproduce several copies of themselves, which may settle in the same and other genes. When TEs settle in germline cells they are thenceforth transmitted to the offspring in a Mendelian manner. TEs are heterogeneous as regards their origin and mode of propagation. The DNA transposons are derived from segments of endogenous retroviruses (ERVs) or pseudogenes or other DNAs of different origin and may move directly to new genomic loci without being reverse-transcribed. On the contrary, the retrotransposons, which are DNAs derived from ancient RNA precursors, need to be transcribed into RNA before being retrotranscribed into DNA and inserted into another site of the host genome. However, the retrotransposons of the SINE (Short Interspersed Elements) class are unable to retrotranspose autonomously and are thought to borrow the enzymatic machinery required for their amplification (reverse transcriptase and endonuclease) from their "partner" autonomous retrotransposon LINE (Long Interspersed Element) L1 [4,5].

TEs at many locations are possibly biologically inactive, but in specific instances they have been shown to play important roles in gene regulation or have been implicated in various diseases.

When settling at specific sites, TEs could interfere with the gene regulatory sequences residing in the noncoding DNA of the 5' and 3' flanking regions and the introns of individual genes [6,7,8,9]. Amongst the different players controlling gene expression, TEs are thought to interfere especially with some microRNAs (miRNAs) and some transcription factors. miRNAs are short (about 22 nt) endogenous RNAs [10,11]. Which regulate gene expression mainly by binding specific sequences at both UTRs or AUGs upstream of the start codon (uAUGs in the 5'UTRs)[12,13]. Furthermore, the Primate SINE Alu elements are possible microRNA targets [14,15,16]. In addition, thanks to their numerous binding sites both Alus and the Rodent SINE B1 elements could regulate the gene expression by directly binding

some transcription factors [1,3,17,18].

In other instances the gene protein product may be modified by TE insertions through the activation of alternative promoters and non-canonical start codons and/or the activation of alternative splice sites [18,20,21,22,23]. It has been demonstrated that some of the protein products generated by these alternative transcripts have been adapted to serve some essential physiological function, but in most cases the splice variants generated by TE inserts result in short-lived protein products [20,24,25].

TEs may also have had a significant impact on genome evolution, including the formation of new genes by some form of "reshuffling" of genetic material or the exonization of some TE sections [3,19,26].

Finally, TE insertions in somatic cells have been implicated in various diseases, including cancer, [3,19].when they interfere with critical sequences in the 5'- and 3'-untranslated regions of mRNAs [28,29,30].

The TE distribution in the genome of extant organisms is determined by both the original insertion site preferences and the succeeding natural selection. Actually TEs are not evenly distributed in the non-coding DNA of Mouse and Human; for instance, Alu and B1 are more represented in the upstream and intronic regions of genes of specific functional classes, but highly expressed genes or loci that might require subtle regulation of transcript levels or precise activation timing tend to be TE-depleted, possibly as a result of a purifying selection [30,31,32].

A few wide scale investigations have been dedicated to the study of the general TE distribution in Mouse and Human genes. A preferential location/persistence of some TE types has been demonstrated, e.g., of Alus in GC-rich isochores [30,33,34]. Zhang et al. have addressed the issue of the likely harmful effects of intronic TE insertions in Mouse and Human [35].

In the present investigation we address the specific issue of TE distribution in the introns and the UTRs of the mitochondrial solute carrier (SLC25) genes of Zebrafish, Mouse and Human. In this very homogeneous family of genes the exon conservation is approximately 70% throughout vertebrates, while the intronic Mouse/Human conservation averages 15% and the Zebrafish/Mouse or Zebrafish/Human conservation is extremely low (at most 0.2%)[36,37]. Despite such divergence, the relative positions of exons and introns have often been kept unaltered during vertebrate evolution,so that the introns of the three species may be considered as strictly homologous.

Methods

Sequences of all available homologous Mouse, Human and Zebrafish SLC25 genes (A1 to A54; 52 Mouse genes, 53 Human genes and 44 Zebrafish genes) were retrieved from NCBI GenBank (http://www.ncbi.nlm.nih.gov/homologene/; http://www.ncbi.nlm.nih.gov/); the transcript variants with superimposable exon-intron arrangements in the three species were preferentially selected for the present analysis. All TEs of the different classes in the 5'- and 3'-flanking regions (each 10,000 nt long) and in the introns of Mouse, Human and Zebrafish genes were identified using the CENSOR tool (http://www.girinst.org/censor , accepting TEs with a score 400 or higher. The significance of the pairwise alignments between the relevant gene sequences and the corresponding TE sequences was further checked with the NCBI BLAST tool [http://blast.ncbi.nlm.nih.gov/Blast.cgi] with settings: "Nucleotide Blast", "Align Two or More Sequences" and "Somewhat similar sequences (blastn)"; only alignments with Expect equal to or lower than 9e-6 were considered [38,39].

In particular, at the flanking regions we recorded at 5' the distance (LAG, measured in nucleotides) between the last upstream significant TE insert and the Start codon [5'UTR-LAG] and at 3' the distance between the Stop codon and the first downstream TE insert [3'UTR-LAG]. In each intron we recorded the number of nucleotides intervening between the last nucleotide of the preceding exon and the first nucleotide of the TE (if any) [INTRON-5'-LAG] and the number of nucleotides intervening between the last nucleotide of the TE (if any) and the first nucleotide of the following exon [INTRON-3'-LAG]. In the case of introns, only LAGs 0 to 100 were considered.

Whenever possible we studied the positions of the proximal promoters or silencers relative to the last TE in the 5' flanking region and of the polyA signal relative to the first TE in the 3' flanking region. Only a small number of experimentally validated promoters could be retrieved from published reports or the EPD Eukaryotic Promoter Database [http://epd.vital-it.ch/]. The positions of the polyA signal were derived from PubMed Nucleotide (http://www.ncbi.nlm.nih.gov/pubmed)[40].

Statistical analyses were made according to the standard methods for comparisons of percentages [41].

Results

General

All introns began with the canonical dinucleotide GT and ended with the canonical dinucleotide AG, except the first intron of SLC25A42 which began with GC in Mouse and Human. However, the initial and terminal hexanucleotides of the introns were highly polymorphic, exhibiting 75 and 102 different configurations, respectively, in our material.

Although about 65 % of the Mouse and Human introns carried at least one TE insert, only 12% circa of Mouse and Human introns exhibited an insert which was 100 nt or less distant from the end of the preceding exon or the beginning of the following exon. In Zebrafish about 80% of the introns carried TE inserts and 29% of the introns exhibited inserts which were less than 100 nt distant from the end of the preceding exon or the beginning of the following exon (including 7% of introns which carried nearing TEs at both ends).

The TEs were usually shorter than 400 nt, but could vary in length from 70 nt up to 600 nt in some ERV-derived TEs.

In the Supplementary Material section are listed, for the different species, the TEs found in the 100 upstream and downstream nucleotides of the introns and the UTR TEs which are the more closer to the Start and Stop codons (Suppl-Tables 1-12).

Suppl-Table 1: 5' end of Mouse Introns. Distances of the first TE from the intron 5' end (INTRON-5'-LAGs)

Carrier	Intron	Distance (nt)	TE name	TE class	Direction	Position
slc25a3	5	34	B1_Mur2	NonLTR/SINE/SINE1	c	TM5
slc25a12	5	48	HAL1	NonLTR/L1	c	upstream of TM1
	9	73	B3	NonLTR/SINE/SINE2	d	upstream of TM1
slc25a16	2	16	MTC	ERV/ERV3	c	between TM1 and TM2
	8	56	B2_Mm1a	NonLTR/SINE/SINE2	c	between TM5 and TM6
slc25a20	7	64	ID_B1	NonLTR/SINE/SINE2	c	between TM5 and TM6
slc25a24	4	61	RMER5	ERV/ERV1	c	upstream of TM1
	7	55	MLT1F1	ERV/ERV3	c	between TM3 and TM4
slc25a26	5	92	B2_Rat3	NonLTR/SINE/SINE2	d	TM4
slc25a27	6	61	B3A	NonLTR/SINE/SINE2	c	nd
slc25a31	3	64	B1_Rn	NonLTR/SINE/SINE1	c	between TM3 and TM4
slc25a36	5	20	B1F	NonLTR/SINE/SINE1	c	between TM3 and TM4
slc25a38	2	91	PB1D10	NonLTR/SINE/SINE1	c	between TM1 and TM2
slc25a39	3	99	B1_Mus1	NonLTR/SINE/SINE1	d	between TM1 and TM2
slc25a40	3	83	Lx_3end	NonLTR/L1	c	between TM1 and TM2
slc25a41	1	55	B1_Mus2	NonLTR/SINE/SINE1	d	upstream of TM1
	5	97	B1_Mus1	NonLTR/SINE/SINE1	c	between TM4 and TM5
slc25a42	1	24	URR1	DNA/hAT	c	upstream of TM1
	2	50	B1F	NonLTR/SINE/SINE1	c	between TM1 and TM2
slc25a46	6	16	LX5	NonLTR/L1	c	TM3
Slc25a50	7	58	B3	NonLTR/SINE/SINE2	c	nd
	11	39	B4A	NonLTR/SINE/SINE2	c	nd

Suppl-Table 2: 5' end of Human Introns. Distances of the first TE from the intron 5' end (INTRON-5'-LAGs)

Carrier	Intron	Distance (nt)	TE name	TE class	Direction	Position
SLC25A3	4	89	AluYd2	NonLTR/SINE/SINE1	c	between TM3 and TM4
SLC25A6	2	9	MLT1B	ERV/ERV3	c	TM4
SLC25A12	3	68	AluJ_Mim	NonLTR/SINE/SINE1	d	upstream of TM1
	12	69	Alu2_TS	NonLTR/SINE/SINE1	d	TM2
SLC25A13	5	33	AluYb3a2	NonLTR/SINE/SINE1	d	upstream of TM1
SLC25A14	5	5	MIRb	NonLTR/SINE/SINE2	d	between TM3 and TM4
SLC25A15	1	3	GOLEM_A	DNA/Mariner	d	TM1
	2	53	Alu2_TS	NonLTR/SINE/SINE1	d	between TM2 and TM3
	5	55	AluYb3a1	NonLTR/SINE/SINE1	d	downstream of TM6
SLC25A16	2	80	LTR10C	ERV/ERV1	d	between TM1 and TM2
	4	84	MER4CL34	ERV/ERV1	d	between TM2 and TM3
SLC25A17	2	78	AluSq2	NonLTR/SINE/SINE1	c	between TM1 and TM2
	5	84	L1-2_Cja	NonLTR/L1	c	between TM3 and TM4
SLC25A20	3	73	AluSx1	NonLTR/SINE/SINE1	d	between TM2 and TM3
	4	94	AluJo	NonLTR/SINE/SINE1	d	between TM3 and TM4
SLC25A21	6	78	AluSp	NonLTR/SINE/SINE1	c	between TM3 and TM4
	7	52	MIRb	NonLTR/SINE/SINE2	c	between TM4 and TM5
SLC25A24	7	39	MLT1F1	ERV/ERV3	c	between TM3 and TM4
	9	100	MER106B	DNA/hAT	c	between TM5 and TM6
SLC25A38	6	56	MER5B	DNA/hAT	d	between TM5 and TM6
SLC25A39	2	61	AluSx	NonLTR/SINE/SINE1	d	between TM1 and TM2
SLC25A40	1	70	MER11A	ERV/ERV2	c	TM1
SLC25A46	5	48	L1PA16	NonLTR/L1	c	TM2
SLC25A50	9	75	AluSx	NonLTR/SINE/SINE1	d	between TM2 and TM3

Suppl-Table 3: 5' end of Zebrafish Introns. Distances of the first TE from the intron 5' end (INTRON-5'-LAGs)

Carrier	Intron	Distance (nt)	TE name	TE class	Direction
slc25a1	3	72	I-3_DR	NonLTR/Nimb	d
slc25a11	6	54	HE1_DR1	NonLTR/SINE/SINE2	d
	7	35	TC1DR3	DNA	c
slc25a12	8	54	DNA-1-5_DR	DNA	d
	9	38	IS3EU-4_DR	DNA/IS3EU	d
	11	36	hAT-N88_DR	DNA/hAT	c
	14	85	hAT-N66B_DR	DNA/hAT	c

	15	13	ANGEL	DNA/Kolobok	c
	17	60	EnSpm-14_HM	DNA/EnSpm/CACTA	d
slc25a16	6	42	DNA11TA1_DR	DNA	c
slc25a18	4	45	I-3_DR	NonLTR/Nimb	d
slc25a19	3	46	TDR2	DNA/Mariner	c
	5	39	Rex1-40_DRe	NonLTR/Rex1	d
slc25a21	2	30	ANGEL	DNA/Kolobok	c
	3	41	Mariner-N7_DR	DNA/Mariner	d
slc25a23	6	90	Kolobok-N10B_DR	DNA/Kolobok	c
slc25a24	7	62	TDR7	DNA	d
	8	67	HATN10_DR	DNA/hAT	c
slc25a25	3	67	Transib-6_HM	DNA/Transib	d
slc25a26	2	16	Tc1-6_AFC	DNA/Mariner	c
slc25a27	5	33	DNAX-16_DR	DNA	d
slc25a29	3	21	Helitron-N2_DR	DNA/Helitron	d
slc25a32	1	61	L2-1B_DR	NonLTR/L2	c
	2	41	Helitron-N3_DR	DNA/Helitron	d
	3	32	DNA-6-N2_DR	DNA	c
	6	70	hAT-N129_DR	DNA/hAT	c
slc25a33	2	82	Kolobok-N3_DR	DNA/Kolobok	d
slc25a38	3	18	DNAX-16_DR	DNA	d
	4	41	Gypsy-223_DR-LTR	LTR/Gypsy	d
	6	91	TDR2	DNA/Mariner	d
slc25a40	2	57	ANGEL	DNA/Kolobok	c
	6	17	DNAX-1B_DR	DNA	d
	7	40	ANGEL	DNA/Kolobok	c
	9	11	TDR12	DNA/Mariner	d
slc25a42	1	23	HE1_DR1	NonLTR/SINE/SINE2	c
	4	65	Kolobok-1N2_DR	DNA/Kolobok	c
	5	17	CR1-24_DR	NonLTR/CR1	d
	6	10	Harbinger-2_DR	DNA/Harbinger	d
slc25a43	3	93	TDR3	DNA/hAT	c
slc25a44	1	69	SINE2-1B_DR	NonLTR/SINE/SINE2	d
slc25a45	1	42	DNA-N15_DR	DNA	d
	2	27	TC1DR3	DNA	c
	5	76	Kolobok-N1_DR	DNA/Kolobok	c
slc25a46	4	49	L2-2_DRe	NonLTR/L2	d
slc25a47	4	17	Zator-N1_DR	DNA/Zator	d

Suppl-Table 4: 3' end of Mouse Introns. Distances of the last TE end from the intron 3' end (INTRON-3'-LAGs)

Carrier	Intron	Distance (nt)	TE name	TE class	Direction	Position
slc25a12	2	25	B3A	NonLTR/SINE/SINE2	c	upstream of TM1
	6	41	B1_Mus1	NonLTR/SINE/SINE1	d	upstream of TM1
slc25a14	4	39	B2_Rat4	NonLTR/SINE/SINE2	c	between TM2 and TM3
slc25a16	3	29	B3	NonLTR/SINE/SINE2	c	between TM2 and TM3
	7	90	ID_B1	NonLTR/SINE/SINE2	c	between TM4 and TM5
slc25a17	1	93	B3	NonLTR/SINE/SINE2	d	TM1
slc25a19	4	68	B3A	NonLTR/SINE/SINE2	c	between TM4 and TM5
slc25a20	3	83	B1_Mus1	NonLTR/SINE/SINE1	c	between TM2 and TM3
	4	64	B4	NonLTR/SINE/SINE2	c	between TM3 and TM4
slc25a23	2	68	RSINE2A	NonLTR/SINE/SINE2	d	upstream of TM1
slc25a27	4	95	RMER4B_LTR	ERV/ERV2	d	nd
slc25a30	1	48	B4	NonLTR/SINE/SINE2	c	TM1
	7	33	MTB	ERV/ERV3	c	between TM5 and TM6
slc25a32	5	49	B1_Mus1	NonLTR/SINE/SINE1	d	between TM4 and TM5
slc25a36	6	49	B1_Rn	NonLTR/SINE/SINE1	d	between TM5 and TM6
slc25a40	8	66	ORR1C2_LTR	ERV/ERV3	d	between TM5 and TM6
slc25a41	2	54	ZP3AR	Simple/Sat	d	between TM1 and TM2
slc25a43	4	44	B3A	NonLTR/SINE/SINE2	d	TM6
slc25a46	7	34	B3	NonLTR/SINE/SINE2	d	between TM3 and TM4
Carrier	**Intron**	**Distance (nt)**	**TE name**	**TE class**	**Direction**	**Position**
slc25a12	2	25	B3A	NonLTR/SINE/SINE2	c	upstream of TM1
	6	41	B1_Mus1	NonLTR/SINE/SINE1	d	upstream of TM1
slc25a14	4	39	B2_Rat4	NonLTR/SINE/SINE2	c	between TM2 and TM3
slc25a16	3	29	B3	NonLTR/SINE/SINE2	c	between TM2 and TM3
	7	90	ID_B1	NonLTR/SINE/SINE2	c	between TM4 and TM5
slc25a17	1	93	B3	NonLTR/SINE/SINE2	d	TM1
slc25a19	4	68	B3A	NonLTR/SINE/SINE2	c	between TM4 and TM5
slc25a20	3	83	B1_Mus1	NonLTR/SINE/SINE1	c	between TM2 and TM3
	4	64	B4	NonLTR/SINE/SINE2	c	between TM3 and TM4
slc25a23	2	68	RSINE2A	NonLTR/SINE/SINE2	d	upstream of TM1
slc25a27	4	95	RMER4B_LTR	ERV/ERV2	d	nd
slc25a30	1	48	B4	NonLTR/SINE/SINE2	c	TM1
	7	33	MTB	ERV/ERV3	c	between TM5 and TM6
slc25a32	5	49	B1_Mus1	NonLTR/SINE/SINE1	d	between TM4 and TM5
slc25a36	6	49	B1_Rn	NonLTR/SINE/SINE1	d	between TM5 and TM6
slc25a40	8	66	ORR1C2_LTR	ERV/ERV3	d	between TM5 and TM6
slc25a41	2	54	ZP3AR	Simple/Sat	d	between TM1 and TM2
slc25a43	4	44	B3A	NonLTR/SINE/SINE2	d	TM6
slc25a46	7	34	B3	NonLTR/SINE/SINE2	d	between TM3 and TM4

Suppl-Table 5: 3' end of Human Introns. Distances of the last TE end from the intron 3' end (INTRON-3'-LAGs)

Carrier	Intron	Distance (nt)	TE name	TE class	Direction	Position
SLC25A7	2	93	AluSx	NonLTR/SINE/SINE1	d	between TM2 and TM3
	5	42	AluJr	NonLTR/SINE/SINE1	d	TM6
SLC25A12	3	27	L1MC5	NonLTR/L1	c	upstream of TM1
SLC25A14	3	93	L1MC1_EC	NonLTR/L1	c	between TM1 and TM2
SLC25A15	2	94	AluJr	NonLTR/SINE/SINE1	c	between TM2 and TM3
SLC25A16	8	73	AluSx	NonLTR/SINE/SINE1	d	between TM5 and TM6
SLC25A17	1	65	AluJr	NonLTR/SINE/SINE1	d	TM1
	2	56	AluS	Interspersed_Repeat	d	between TM1 and TM2
SLC25A20	5	27	AluJ	Interspersed_Repeat	c	TM4
SLC25A24	3	80	Charlie25	DNA/hAT	d	upstream of TM1
	6	35	MLT1A0	ERV/ERV3	c	between TM2 and TM3
SLC25A27	7	96	AluY	NonLTR/SINE/SINE1	c	between TM5 and TM6
SLC25A30	7	72	AluY	NonLTR/SINE/SINE1	c	between TM5 and TM6
SLC25A33	1	94	AluYb3a1	NonLTR/SINE/SINE1	c	TM1
	6	85	AluSz	NonLTR/SINE/SINE1	c	between TM5 and TM6
SLC25A38	3	53	LTR16C	ERV/ERV3	d	TM2
SLC25A39	2	71	AluSx	NonLTR/SINE/SINE1	d	between TM1 and TM2
SLC25A50	4	40	AluS	Interspersed_Repeat	d	nd
	6	39	AluSq2	NonLTR/SINE/SINE1	c	nd

Suppl-Table 6: 3' end of Zebrafish Introns. Distances of the last TE end from the intron 3' end (INTRON-3'-LAGs).

Carrier	Intron	Distance (nt)	TE name	TE class	Direction
slc25a3	1	94	EnSpm-N21_DR	DNA/EnSpm/CACTA	c
slc25a8	5	39	hAT-27N1_DR	DNA/hAT	c
slc25a9	1	89	SINE2-3_DR	NonLTR/SINE/SINE2	d
slc25a11	1	75	hAT-N86_DR	DNA/hAT	d
	2	48	Gypsy41-LTR_DR	LTR/Gypsy	c
	5	37	LOOPERN4B_DR	DNA/Kolobok	c
	7	75	TDR24	DNA	d
slc25a12	5	19	Mariner-N8_EL	DNA/Mariner	d
	11	93	hAT-N88_DR	DNA/hAT	c
	14	53	hAT-N66B_DR	DNA/hAT	c
	15	68	Kolobok-N3_DR	DNA/Kolobok	c

	17	85	Harbinger-N11_DR	DNA/Harbinger	d
slc25a16	4	48	Rex1-38_DRe	NonLTR/Rex1	c
	6	32	DNA11TA1_DR	DNA	c
slc25a18	2	92	TDR7	DNA	d
	7	42	hAT-N48B_DR	DNA/hAT	c
slc25a19	4	59	hAT-N157_DR	DNA/hAT	d
slc25a21	8	72	ANGEL	DNA/Kolobok	c
	9	26	Kolobok-N10B_DR	DNA/Kolobok	d
slc25a22	2	37	Kolobok-N5_DR	DNA/Kolobok	d
slc25a23	2	82	Mariner-7_DR	DNA/Mariner	c
	8	59	SINE2-4_DR	NonLTR/SINE/SINE2	d
slc25a25	6	65	SINE3-1	NonLTR/SINE/SINE3	c
slc25a27	5	46	EnSpm-N4_DR	DNA/EnSpm/CACTA	d
slc25a29	3	72	DNAX-1_DR	DNA	c
slc25a32	5	53	Kolobok-N5_DR	DNA/Kolobok	c
slc25a36	3	42	HATN12B_DR	DNA/hAT	d
slc25a37	1	84	Kolobok-N1_DR	DNA/Kolobok	c
	3	60	HATN6_DR	DNA/hAT	c
slc25a38	1	85	TDR25	DNA	c
	3	72	HATN4_DR	DNA/hAT	c
	4	27	L2-2_DRe	NonLTR/L2	d
	6	45	CryptonV-N5_DR	DNA/Crypton/CryptonV	d
slc25a39	10	90	EnSpm1_SB	DNA/EnSpm/CACTA	c
slc25a42	4	15	Kolobok-1N2_DR	DNA/Kolobok	c
	6	22	Helitron-N3_DR	DNA/Helitron	c
slc25a44	2	70	hAT-N48_DR	DNA/hAT	c
slc25a45	1	3	Polinton-1_DR	DNA/Polinton	c
	2	52	TC1DR3	DNA	c
	3	9	hAT5-N2_DR	DNA/hAT	d
	5	65	DNA-8-3_DR	DNA	d
slc25a46	3	90	DNA13TA1_DR	DNA	d
	4	88	TDR17B	DNA	d
slc25a47	1	71	HE2_DR	NonLTR/SINE/SINE2	d
slc25a50	12	51	hAT-27N1_DR	DNA/hAT	d

Suppl-Table 7: Mouse 5'UTR. Distances from the start codon to the end of the immediately upstream TE (5'UTR-LAGs).

Carrier	Distance (nt)	TE name	TE class	Direction
slc25a1	3125	B1_Mus1	NonLTR/SINE/SINE1	d
slc25a2	596	PB1D9	NonLTR/SINE/SINE1	c
slc25a3	1405	B3A	NonLTR/SINE/SINE2	c
slc25a4	2029	Lx7_3end	NonLTR/L1	d

slc25a5	1415	MTD	ERV/ERV3	d
slc25a7	1558	B1_Mur4	NonLTR/SINE/SINE1	c
slc25a8	4462	RLTR14	ERV/ERV1	c
slc25a9	617	B1	NonLTR/SINE/SINE1	d
slc25a10	2966	B1	NonLTR/SINE/SINE1	c
slc25a11	1549	B3	NonLTR/SINE/SINE2	d
slc25a12	754	ID_B1	NonLTR/SINE/SINE2	c
slc25a13	1355	MLT1A1	ERV/ERV3	c
slc25a14	784	LX5c	NonLTR/L1	c
slc25a15	1512	LX9	NonLTR/L1	c
slc25a16	351	B1F1	NonLTR/SINE/SINE1	c
slc25a17	506	B3A	NonLTR/SINE/SINE2	c
slc25a18	400	B2_Mm1a	NonLTR/SINE/SINE2	c
slc25a19	185	B1_Rn	NonLTR/SINE/SINE1	c
slc25a20	1076	B3	NonLTR/SINE/SINE2	c
slc25a21	158	B1F	NonLTR/SINE/SINE1	c
slc25a22	2242	B1_Rn	NonLTR/SINE/SINE1	d
slc25a23	4119	B1_Mus1	NonLTR/SINE/SINE1	d
slc25a24	1422	ORR1F_Str	ERV/ERV2	c
slc25a25	6901	B1_Mus2	NonLTR/SINE/SINE1	c
slc25a26	714	RSINE1	NonLTR/SINE/SINE2	c
slc25a27	1516	RSINE1	NonLTR/SINE/SINE2	c
slc25a28	2642	B1_Rn	NonLTR/SINE/SINE1	c
slc25a29	3625	B2	NonLTR/SINE/SINE2	c
slc25a30	103	ORR1C2_LTR	ERV/ERV3	d
slc25a31	1176	LX8	NonLTR/L1	d
slc25a32	755	B1_Mus1	NonLTR/SINE/SINE1	c
slc25a33	0	BEL-636_AA-I	LTR/BEL	d
slc25a34	497	B1_Mur1	NonLTR/SINE/SINE1	d
slc25a35	2205	B1_Mur4	NonLTR/SINE/SINE1	d
slc25a36	1676	B1_Mus1	NonLTR/SINE/SINE1	c
slc25a37	1417	B1F	NonLTR/SINE/SINE1	c
slc25a38	335	B1_Mus1	NonLTR/SINE/SINE1	c
slc25a39	2210	B1_Mm	NonLTR/SINE/SINE1	c
slc25a40	937	B1_Mur2	NonLTR/SINE/SINE1	c
slc25a41	267	ID_B1	NonLTR/SINE/SINE2	d
slc25a42	1290	B1_Mur2	NonLTR/SINE/SINE1	c
slc25a43	1251	RSINE2A	NonLTR/SINE/SINE2	c
slc25a44	1137	B1_Mus2	NonLTR/SINE/SINE1	c
slc25a45	1552	L1MA1	NonLTR/L1	c
slc25a46	1977	B1_Mm	NonLTR/SINE/SINE1	c
slc25a47	1646	B1_Mus1	NonLTR/SINE/SINE1	d

slc25a48	504	BGLII_LTR	ERV/ERV2	d
slc25a49	1641	ORR1A0	ERV/ERV3	c
slc25a50	606	B1_Mur1	NonLTR/SINE/SINE1	c
slc25a51	706	B2_Rat1	NonLTR/SINE/SINE2	d
slc25a53	3300	B1_Rn	NonLTR/SINE/SINE1	c
slc25a54	786	B1_Rn	NonLTR/SINE/SINE1	c

Suppl-Table 8: Human 5'UTR. Distances from the start codon to the end of the immediately upstream TE (5'UTR-LAGs)

Carrier	Distance (nt)	TE name	TE class	Direction
SLC25A1	6650	AluY	NonLTR/SINE/SINE1	c
SLC25A2	364	AluJ	Interspersed_Repeat	c
SLC25A3	1242	AluSx	NonLTR/SINE/SINE1	d
SLC25A4	1832	AluSx	NonLTR/SINE/SINE1	c
SLC25A5	620	MER96B	DNA/hAT	c
SLC25A6	1435	AluSp	NonLTR/SINE/SINE1	d
SLC25A7	2689	MLT1H1	ERV/ERV3	c
SLC25A8	886	AluSx	NonLTR/SINE/SINE1	c
SLC25A9	395	AluJr	NonLTR/SINE/SINE1	d
SLC25A10	599	AluS	Interspersed_Repeat	c
SLC25A11	1541	MIRb	NonLTR/SINE/SINE2	c
SLC25A12	1014	AluSx1	NonLTR/SINE/SINE1	d
SLC25A13	1192	THE1C	ERV/ERV3	d
SLC25A14	894	AluSz	NonLTR/SINE/SINE1	c
SLC25A15	769	MER5A1	DNA/hAT	d
SLC25A16	357	AluSz	NonLTR/SINE/SINE1	c
SLC25A17	1120	AluSz	NonLTR/SINE/SINE1	c
SLC25A18	558	AluSx1	NonLTR/SINE/SINE1	c
SLC25A19	82	L2	NonLTR/CR1	c
SLC25A20	310	L2	NonLTR/CR1	d
SLC25A21	705	MER5A	DNA/hAT	c
SLC25A22	3681	AluSz	NonLTR/SINE/SINE1	c
SLC25A23	1299	AluSq	NonLTR/SINE/SINE1	d
SLC25A24	1553	AluS	Interspersed_Repeat	d
SLC25A25	2908	AluSq	NonLTR/SINE/SINE1	c
SLC25A26	1196	MIR	NonLTR/SINE/SINE2	c
SLC25A27	1600	L1ME4A	NonLTR/L1	d
SLC25A28	1702	AluJb	NonLTR/SINE/SINE1	c
SLC25A29	222	EnSpm-2_PGr	DNA/EnSpm/CACTA	c
SLC25A30	414	AluJo	NonLTR/SINE/SINE1	c
SLC25A31	769	MER5B	DNA/hAT	c
SLC25A32	1932	Charlie7_Aves	DNA/hAT	c
SLC25A33	111	MuDR-9_SBi	DNA/MuDR	d

SLC25A34	652	MIRb	NonLTR/SINE/SINE2	c
SLC25A35	1126	AluY	NonLTR/SINE/SINE1	d
SLC25A36	586	L1ME3F_3end	NonLTR/L1	c
SLC25A37	1444	L1ME4A	NonLTR/L1	c
SLC25A38	702	CHARLIE5	DNA/hAT	c
SLC25A39	93	AluSc	NonLTR/SINE/SINE1	d
SLC25A40	445	AluY	NonLTR/SINE/SINE1	c
SLC25A41	260	AluY	NonLTR/SINE/SINE1	d
SLC25A42	505	AluSx1	NonLTR/SINE/SINE1	d
SLC25A43	648	MIRb	NonLTR/SINE/SINE2	c
SLC25A44	1097	Alu2_TS	NonLTR/SINE/SINE1	c
SLC25A45	669	L2	NonLTR/CR1	c
SLC25A46	1104	L1MA3	NonLTR/L1	d
SLC25A47	812	MIRc	NonLTR/SINE/SINE2	d
SLC25A48	8043	MIRb	NonLTR/SINE/SINE2	c
SLC25A49	2322	AluSq	NonLTR/SINE/SINE1	c
SLC25A50	640	AluSg	NonLTR/SINE/SINE1	d
SLC25A51	228	CHARLIE1	DNA/hAT	c
SLC25A52	641	AluJr	NonLTR/SINE/SINE1	c
SLC25A53	696	MER20	DNA/hAT	d

Suppl-Table 9: Zebrafish *5'UTR.* Distances from the start codon to the end of the immediately upstream TE (5'UTR-LAGs)

Carrier	Distance (nt)	TE name	TE class	Direction
slc25a1	575	Ginger1-10_HM	DNA/Ginger1	d
slc25a3	455	GYCUME1_LTR	LTR/Gypsy	d
slc25a4	2793	HATN4_DR	DNA/hAT	d
slc25a5	1868	L2-5_DRe	NonLTR/L2	d
slc25a6	382	ANGEL	DNA/Kolobok	c
slc25a8	175	ANGEL	DNA/Kolobok	d
slc25a9	413	DNA-X-8_DR	DNA	d
slc25a11	215	Tx1-36_DR	NonLTR/Tx1	c
slc25a12	139	Tx1-17_DR	NonLTR/Tx1	c
slc25a13	307	DNA-N15_DR	DNA	c
slc25a14	1037	DNA-5-8B_DR	DNA	c
slc25a15	373	HE1_DR1	NonLTR/SINE/SINE2	d
slc25a16	936	Kolobok-N7_DR	DNA/Kolobok	d
slc25a17	2089	EnSpm-N4_DR	DNA/EnSpm/CACTA	d
slc25a18	727	TDR7	DNA	c
slc25a19	443	Ginger1-N1_DR	DNA/Ginger1	c
slc25a20	268	MuDR-N1_DR	DNA/MuDR	c
slc25a21	1070	HATN10_DR	DNA/hAT	c

slc25a22	981	DNA-CCGG-1_DR	DNA	c
slc25a23	1469	EnSpm-N6_DR	DNA/EnSpm/CACTA	d
slc25a24	2165	TDR14	DNA	d
slc25a25	879	ANGEL	DNA/Kolobok	c
slc25a26	185	Tx1-23_DR	NonLTR/Tx1	c
slc25a27	575	Kolobok-2_DR	DNA/Kolobok	d
slc25a28	575	DNA-8-13_DR	DNA	d
slc25a29	442	EnSpm-N13_DR	DNA/EnSpm/CACTA	c
slc25a32	700	DNA-8-28_DR	DNA	c
slc25a33	500	Tx1-5_DR	NonLTR/Tx1	c
slc25a35	410	L2-5_DRe	NonLTR/L2	d
slc25a36	703	DNA-6-N4_DR	DNA	d
slc25a37	1211	L1-72_DR	NonLTR/L1	d
slc25a38	617	EnSpm-N6_DR	DNA/EnSpm/CACTA	c
slc25a39	584	DNA8-7_DR	DNA/hAT	d
slc25a40	1405	SINE2-4_DR	NonLTR/SINE/SINE2	c
slc25a42	112	HATN10_DR	DNA/hAT	c
slc25a43	417	DNAX-16_DR	DNA	d
slc25a44	315	EnSpm-N13_DR	DNA/EnSpm/CACTA	d
slc25a45	542	hAT-N137_DR	DNA/hAT	c
slc25a46	1944	hAT-N25_DR	DNA/hAT	d
slc25a47	774	ANGEL	DNA/Kolobok	d
slc25a48	2284	DNA2-10_DR	DNA	d
slc25a50	335	SINE2-5_DR	NonLTR/SINE/SINE2	d
slc25a51	341	ANGEL	DNA/Kolobok	c
slc25a52	354	ANGEL	DNA/Kolobok	d

Suppl-Table 10: Mouse 3'UTR. Distances from the stop codon to the first downstream TE (3'UTR-LAGs)

Carrier	Distance (nt)	TE name	TE class	Direction
slc25a1	5566	ID_B1	NonLTR/SINE/SINE2	d
slc25a2	619	L1MdTf_II	NonLTR/L1	c
slc25a3	1325	B1_Mus1	NonLTR/SINE/SINE1	d
slc25a4	1381	B3	NonLTR/SINE/SINE2	d
slc25a5	1394	B2_Mm2	NonLTR/SINE/SINE2	c
slc25a7	1332	RMER17A	ERV/ERV2	c
slc25a8	855	B1F	NonLTR/SINE/SINE1	c
slc25a9	1324	ORR1D2_LTR	ERV/ERV3	d
slc25a10	2821	RSINE2	NonLTR/SINE/SINE2	d
slc25a11	624	B1_Mus1	NonLTR/SINE/SINE1	d
slc25a12	2152	B1F	NonLTR/SINE/SINE1	c

slc25a13	1165	B2_Rat4	NonLTR/SINE/SINE2	c
slc25a14	457	B2_Rat2	NonLTR/SINE/SINE2	c
slc25a15	2059	ORR1C2_LTR	ERV/ERV3	d
slc25a16	229	B1F2	NonLTR/SINE/SINE1	c
slc25a17	757	ORR1D1_LTR	ERV/ERV	d
slc25a18	717	B3A	NonLTR/SINE/SINE2	c
slc25a19	1277	MTC	ERV/ERV3	c
slc25a20	1508	RSINE1	NonLTR/SINE/SINE2	d
slc25a21	2103	B1_Mus1	NonLTR/SINE/SINE1	d
slc25a22	5628	B1_Mm	NonLTR/SINE/SINE1	d
slc25a23	2160	Gypsy-14B_ATr-LTR	LTR/Gypsy	d
slc25a24	1197	B1_Rn	NonLTR/SINE/SINE1	d
slc25a25	2395	B2_Mm2	NonLTR/SINE/SINE2	d
slc25a26	3011	RSINE2A	NonLTR/SINE/SINE2	d
slc25a27	2494	B3	NonLTR/SINE/SINE2	c
slc25a28	2711	PB1D7	NonLTR/SINE/SINE1	c
slc25a29	1971	RSINE1	NonLTR/SINE/SINE2	c
slc25a30	1726	B1_Rn	NonLTR/SINE/SINE1	c
slc25a31	1762	LX8	NonLTR/L1	c
slc25a32	1297	MER99	DNA	d
slc25a33	559	PB1D10	NonLTR/SINE/SINE1	c
slc25a34	859	ID_B1	NonLTR/SINE/SINE2	d
slc25a35	2922	B1_Mus2	NonLTR/SINE/SINE1	c
slc25a36	4294	RLTR11A2	ERV/ERV2	d
slc25a37	3883	B3	NonLTR/SINE/SINE2	d
slc25a38	790	B2_Rat4	NonLTR/SINE/SINE2	d
slc25a39	326	PB1D7	NonLTR/SINE/SINE1	d
slc25a40	982	RSINE1	NonLTR/SINE/SINE2	c
slc25a41	4240	B1_Mus1	NonLTR/SINE/SINE1	d
slc25a42	1253	B1_Mus1	NonLTR/SINE/SINE1	c
slc25a43	225	B2_Mm1a	NonLTR/SINE/SINE2	d
slc25a44	1418	ID_B1	NonLTR/SINE/SINE2	c
slc25a45	950	B3	NonLTR/SINE/SINE2	d
slc25a46	2374	IAPEY3_LTR	ERV/ERV2	d
slc25a47	2037	B1_Mur1	NonLTR/SINE/SINE1	c
slc25a48	2092	B1_Mus1	NonLTR/SINE/SINE1	c
slc25a49	899	MLTR25A	ERV/ERV2	d
slc25a50	1212	Charlie16a	DNA/hAT	c
slc25a51	1342	PB1D10	NonLTR/SINE/SINE1	c
slc25a53	1743	LTRIS_Mm	ERV/ERV1	d
slc25a54	1090	L1MB3	NonLTR/L1	d

Suppl-Table 11: Human 3'UTR. Distances from the stop codon to the first downstream TE (3'UTR-LAGs)

Carrier	Distance (nt)	TE name	TE class	Direction
SLC25A1	8119	MER81	DNA/hAT	c
SLC25A2	462	AluSx1	NonLTR/SINE/SINE1	c
SLC25A3	642	AluSp	NonLTR/SINE/SINE1	c
SLC25A4	725	AluSz6	NonLTR/SINE/SINE1	d
SLC25A5	730	AluSc	NonLTR/SINE/SINE1	c
SLC25A6	747	MLT1A0	ERV/ERV3	c
SLC25A7	1147	LTR16C	ERV/ERV3	d
SLC25A8	963	AluSz6	NonLTR/SINE/SINE1	c
SLC25A9	1610	AluJr	NonLTR/SINE/SINE1	d
SLC25A10	1151	ALU	NonLTR/SINE/SINE1	c
SLC25A11	2125	AluJb	NonLTR/SINE/SINE1	c
SLC25A12	1670	L1ME4A	NonLTR/L1	c
SLC25A13	3498	AluSp	NonLTR/SINE/SINE1	c
SLC25A14	1111	L1MC1	NonLTR/L1	c
SLC25A15	508	AluSz	NonLTR/SINE/SINE1	c
SLC25A16	579	PB1	NonLTR/SINE/SINE1	d
SLC25A17	2496	MIRc	NonLTR/SINE/SINE2	c
SLC25A18	1312	AluY	NonLTR/SINE/SINE1	d
SLC25A19	3572	AluJb	NonLTR/SINE/SINE1	d
SLC25A20	1526	Alu2_TS	NonLTR/SINE/SINE1	d
SLC25A21	2227	MER126_Crp	DNA	d
SLC25A22	6060	Alu2_TS	NonLTR/SINE/SINE1	d
SLC25A23	2129	AluSz	NonLTR/SINE/SINE1	c
SLC25A24	3075	THE1B	ERV/ERV3	d
SLC25A25	2203	L2	NonLTR/CR1	d
SLC25A26	7663	AluSz6	NonLTR/SINE/SINE1	c
SLC25A27	1861	MIR3	NonLTR/SINE/SINE2	c
SLC25A28	731	MIRc	NonLTR/SINE/SINE2	c
SLC25A29	2427	AluSz	NonLTR/SINE/SINE1	c
SLC25A30	3188	AluSq	NonLTR/SINE/SINE1	d
SLC25A31	900	CR1_Mam	NonLTR/CR1	d
SLC25A32	505	AluY	NonLTR/SINE/SINE1	c
SLC25A33	441	AluS	Interspersed_Repeat	d
SLC25A34	316	MIRb	NonLTR/SINE/SINE2	c
SLC25A35	1178	AluJb	NonLTR/SINE/SINE1	c
SLC25A36	4268	Alu2_TS	NonLTR/SINE/SINE1	c
SLC25A37	2332	AluSz	NonLTR/SINE/SINE1	d
SLC25A38	863	L1ME1	NonLTR/L1	c
SLC25A39	601	AluSc	NonLTR/SINE/SINE1	d
SLC25A40	2386	L1PA13	NonLTR/L1	d

SLC25A41	1763	BEL-636_AA-I	LTR/BEL	c
SLC25A42	2802	AluJo	NonLTR/SINE/SINE1	d
SLC25A43	471	AluJo	NonLTR/SINE/SINE1	c
SLC25A44	3035	AluJ	Interspersed_Repeat	c
SLC25A45	314	AluY	NonLTR/SINE/SINE1	c
SLC25A46	3422	L1MB4_5	NonLTR/L1	d
SLC25A47	2166	AluY	NonLTR/SINE/SINE1	c
SLC25A48	498	MamGypLTR2c	LTR/Gyps	c
SLC25A49	2362	L1MB8	NonLTR/L1	d
SLC25A50	1361	Charlie16a	DNA/hAT	c
SLC25A51	324	AluSg	NonLTR/SINE/SINE1	c
SLC25A52	2447	AluSz	NonLTR/SINE/SINE1	d
SLC25A53	1181	AluSq	NonLTR/SINE/SINE1	c

Suppl-Table 12: Zebrafish 3'UTR. Distances from the stop codon to the first downstream TE (3'UTR-LAGs)

Carrier	**Distance (nt)**	**TE name**	**TE class**	**Direction**
slc25a1	854	hAT-N22_DR	DNA/hAT	d
slc25a3	395	DNA-5-8_DR	DNA	d
slc25a4	982	Helitron-N2_DR	DNA/Helitron	c
slc25a5	409	CryptonV-N5_DR	DNA/Crypton/CryptonV	c
slc25a6	386	EnSpm-N4_DR	DNA/EnSpm/CACTA	c
slc25a8	616	TC1DR3	DNA	d
slc25a9	243	hAT-N145_DR	DNA/hAT	d
slc25a11	30	TDR16	DNA	c
slc25a12	885	DNAX-1_DR	DNA	d
slc25a13	1614	DNA-X-5_DR	DNA	d
slc25a14	725	Kolobok-N10_DR	DNA/Kolobok	c
slc25a15	821	TDR2	DNA/Mariner	d
slc25a16	949	hAT-N141_DR	DNA/hAT	c
slc25a17	332	DNA15TA1_DR	DNA	c
slc25a18	1395	SINE3-1a	NonLTR/SINE/SINE3	c
slc25a19	957	TC1DR3B	DNA/Mariner	c
slc25a20	403	TDR24	DNA	d
slc25a21	3805	TDR7	DNA	c
slc25a22	844	TDR23	DNA	d
slc25a23	417	hAT-N137_DR	DNA/hAT	c
slc25a24	824	ANGEL	DNA/Kolobok	c
slc25a25	194	Mariner-N5_HSal	DNA/Mariner	d
slc25a26	460	TDR13	DNA	c
slc25a27	657	Mariner-N38_DR	DNA/Mariner	d
slc25a28	289	hAT-N134_DR	DNA/hAT	d

slc25a29	103	DNAX-1C_DR	DNA	c
slc25a32	502	Mariner-N38_DR	DNA/Mariner	d
slc25a33	979	EXPANDER1_DR	NonLTR/RTE	d
slc25a35	220	P-30_HM	DNA/P	d
slc25a36	1741	Kolobok-N10B_DR	DNA/Kolobok	c
slc25a37	692	TDR3C	DNA	c
slc25a38	955	hAT-N161_DR	DNA/hAT	c
slc25a39	1018	ASAT_CY	Simple/Sat/SAT	c
slc25a40	762	DNA11TA1_DR	DNA	d
slc25a42	221	hAT-N22_DR	DNA/hAT	c
slc25a43	3940	CryptonV-N5_DR	DNA/Crypton/CryptonV	c
slc25a44	720	SINE3-1a	NonLTR/SINE/SINE3	c
slc25a45	89	DNAX-1_DR	DNA	d
slc25a46	834	CryptonV-N5_DR	DNA/Crypton/CryptonV	c
slc25a47	34	TDR17B	DNA	d
slc25a48	536	Daphne-27_DRe	NonLTR/Daphne	c
slc25a50	1106	SINE3-1	NonLTR/SINE/SINE3	d
slc25a51	121	EnSpm-N21_DR	DNA/EnSpm/CACTA	c
slc25a52	230	Mariner-N7_DR	DNA/Mariner	d

For each TE the type, class and direction are reported. For intronic TEs, the number of the hosting intron and the distance from its upstream or downstream end are indicated and in Mouse and Human is also reported the position of the intron concerned with reference to the six transmembrane (TM) coding segments. For UTR TEs, the distance in nucleotides from the Start or Stop codon is reported.

The nature of TEs

In the text Table 1 are reported the per cent incidences of the different classes (and subclasses) of TEs at the different locations (intronic 5' and 3' ends, UTR 5' and 3') in the different species.

LTR Retrotransposons, Non-LTR Retrotransposons and DNA transposons were found in all three species, but Endogenous Retrovirus-derived TEs were not found in Zebrafish, although represented in Mouse and Human. In our material, the other main difference between Mammals and Zebrafish is that in Mammals Non-LTR Retrotransposons are highly prevalent on DNA transposons (78-81% vs. 2-11%), while in Zebrafish the majority of TEs are DNA transposons (81%) and the Non-LTR Retrotransposon are 17% only.

The SINE1 subclass of Non-LTR Retrotransposons, typically represented in Mammals by the widespread primate Alu and the rodent B1 [both derived from the 7SL RNA], is not represented in Zebrafish. The SINE3 subclass of Non-LTR Retrotransposons [derived from the 5S rRNA] is represented in Zebrafish but not in Mammals. On the contrary, the SINE2 subclass (derived from tRNA) is represented in all species [5,42,43].

DNA transposons are present in all three species, but TEs of the subclass Kolobok were not found in Mammals and those of the subclasses EnSpm/CACTA and Mariner were absent in the Human, while expressed in Mouse.

In Mammals the overall frequency of SINE1 is about twice of that of SINE2, but, at all locations, the incidence of SINE1 is higher in Human than in Mouse (on average, 60% vs. 36%; $p<0.01$), while the incidence of SINE2 is significantly higher in Mouse as compared to Human (on average, 37% vs. 7%; $p<0.01$).

TE type composition may vary significantly between the two UTRs in the same species. In Mouse the SINE1 subclass is significantly more represented at 5' than at 3' (51.9% vs 34.6%; $p<0.01$), whereas the SINE2 subclass is significantly more represented at 3' than at 5' (36.5% vs. 23.1%; $p<0.01$). In Human, no significant 5'/3' differences were detected. In Zebrafish the DNA transposon Mariner is represented at the 3' UTR, but not at UTR 5' ($p<0.05$).

The incidence differences in TE direction (whether direct or complement) are not statistically significant, except for the SINE1 subclass in Mouse 5' flanking region, in which the antisense direction prevails (74%).

We also compared the frequency of the different TE classes in the proximity of the coding regions with the estimated general frequency in the corresponding genomes, as reported by Chalopin et al. (2015; for Mouse, Human and Zebrafish) and Howe et al. (2013; for Zebrafish) (Table 2). In all the three species at intronic and UTR sites nearing coding sequences the overall percentage of Non-LTR Retrotransposons is higher and that of LTR Retrotransposons + Endogenous Retroviruses is lower, as compared to the average global TE distribution in the corresponding genome.

Exon-next TE distances at the 5' end of the introns [INTRON-5'-LAGs]

In this section the results will be treated cumulatively for Mouse, Human and Zebrafish since, despite the differences in TE types, no remarkable difference was observed between these species as regards the statistical distribution of the TE distances from the intron 5' ends.

The individual INTRON-5'-LAGs are listed in the Suppl-Tables 1-3. The cumulative frequency distribution of TE LAGs 0-100 at the 5' end of introns is shown in Figure: 1, both as a column bar graph and as a line graph of the central moving average of five consecutive data.

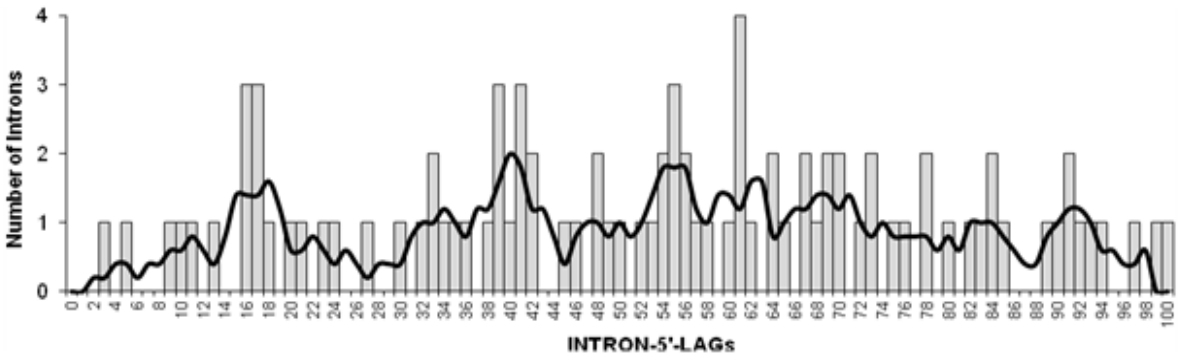

Figure 1: Distribution of INTRON-5'-LAGs 0 to 100. Number of introns (Mouse + Human + Zebrafish) = 91. INTRON-5'-LAG is the distance (in nucleotides) between the TE first nucleotide and the last nucleotide of the preceding exon. Column bar graph: actual data. Line graph: central moving average of five consecutive data (values for Lags 0 and 1 and 99 and 100 are not-valid).

The average TE frequency per Lag in Lags 0 to 20 is 0.67 ± 0.13 and in Lags 21 to 100 is 0.96 ± 0.03. Thus, the TE frequency is significantly ($p<0.01$) lower in a segment approximately 20-nt long at the 5' end of the intron than at longer distances from the preceding exon.

TE-next Exon distances at the 3' end of the introns [INTRON-3'-LAGs]

As for the INTRON-5'-LAGs, the INTRON-3'-LAGs are treated cumulatively for Mouse, Human and Zebrafish.

The individual INTRON-3'-LAGs are listed in the Suppl-Tables 4-6. The cumulative frequency distribution of TE LAGs 0-100 is shown in Figure: 2, both as a column bar graph and as a line graph of the central moving average of five consecutive data.

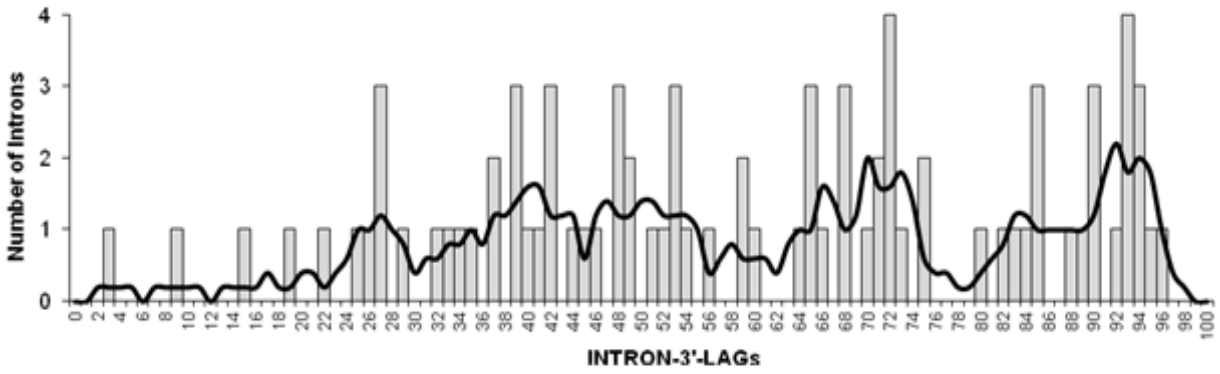

Figure 2: Distribution of INTRON-3'-LAGs 0 to 100. Number of introns (Mouse + Human + Zebrafish) = 83. INTRON-3'-LAG is the distance (in nucleotides) between the TE last nucleotide and the first nucleotide of the succeeding exon. Column bar graph: actual data. Line graph: central moving average of five consecutive data (values for Lags 0 and 1 and 99 and 100 are not-valid).

The average TE frequency per Lag is 0.19 ± 0.11 in Lags 0 to 20 and 0.99 ± 0.02 in Lags 21 to 100. Thus, the TE frequency is significantly ($p<0.01$) lower in a segment approximately 20-nt long at the 3' end of the intron than at longer distances from the next exon.

As shown previously, the TE frequency is significantly lower at both the 20-nucleotide ends of the introns, but a comparison between the relative frequencies at the upstream end (0.67 ± 0.13) and the downstream end (0.19 ± 0.11) demonstrates that the TE frequency is significantly ($p<0.01$) more reduced at the 3' than at the 5'.

On the whole, in Mouse and Human introns with TE inserts in the proximity of their ends were apparently randomly distributed in the different sections of the genes, i.e., upstream of TM 1, within all the six TM-coding segments and the intervening segments, and downstream of TM 6 (Suppl-Tables 1 and 2 and 4 and 5).

TEs at the 5' flanking region [5'UTR-LAGs] in Mouse and Human (the relative position of regulatory 5' sequences)

The individual 5'UTR-LAGs are listed in the Suppl-Tables 7 and 8. The distances recorded were widely variable in the different genes, but the overall cumulative (Mouse + Human) distribution was unimodal and distinctly right-skewed with Excel asymmetry index = 2.66 (Figure: 3). The distribution peaked at 750 nt and the average was 1350 nt. However, 10% of distances were less than 300 nt and in one exceptional case (Mouse slc25a33) a BEL-636_AA-I (LTR/BEL class) TE immediately preceded the start codon.

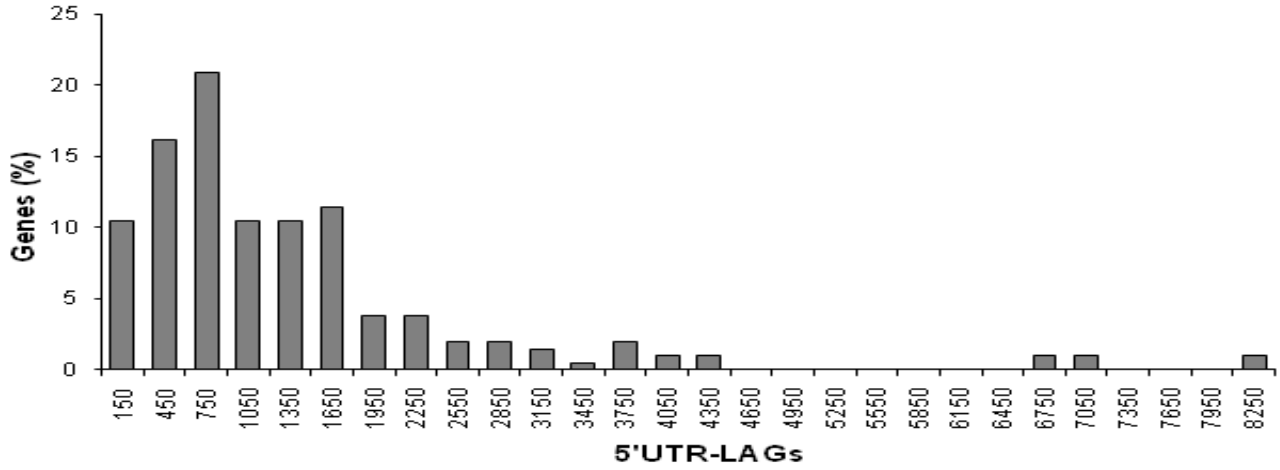

Figure 3: Distribution of 5'UTR-LAGs. Cumulative distribution of all SLC25 Mouse and Human genes. 5'UTR-LAGs are the distances (in nucleotides) between the start codon and the end of the immediately upstream TE.

There is little currently known about experimentally validated proximal promoter sites. In human SLC25As members 1 4 (EPD database), 5 (EPD database)in which the 5'UTR-LAG interval is 620 nt or more long, the proximal promoter site was found to be located within this region. In addition, the 5'UTR-LAG regions of SLC25A1 and SLC25A3 also comprised a silencer sequence [27,32,44,45,46,47]. In SLC25A20, in which the 5'UTR-LAG interval is 310 nt long, the promoter is likely to reside at the very upstream limit of the LAG region [48,49]. In SLC25A19, in which the 5'UTR-LAG region is only 82 nt long, the basal promoter is reported to reside about 2,500 nt upstream the LAG region [50].

TEs at the 3' flanking region [3'UTR-LAGs] in Mouse and Human (the relative position of polyA signals)

The individual 3'UTR-LAGs are listed in the Suppl-Tables 10 and 11. Like at the 5', the distances recorded were widely variable. The overall cumulative (Mouse + Human) distribution was unimodal and markedly right-skewed with Excel asymmetry index = 1.97 (Figure: 4). The distribution peak was at 750 nt and the average was 1850 nt, but 2% of distances were less than 300 nt.

The Mouse/Human UTR-LAGs (5' + 3') in corresponding genes are positively correlated ($r = 0.45$; $p<0.01$).

Data on the polyA signal position were available for some genes only. In 10 Mouse genes (slc25a members 3, 5, 8, 10, 21, 23, 29, 31, 32, and 37) and 22 Human genes (SLC25A members 1, 3, 4, 5, 6, 8, 9, 10, 12, 13, 14, 18, 19, 22, 24, 25, 29, 36, 38, 42, 47, and 49) the polyA signal was located in the TE-free area following the stop codon. In one Mouse gene (slc25a16) and four Human genes (SLC25A members 15, 16, 32, and 43) the polyA signal was located downstream of the first significant TE insert.

The position of the polyA signal is very variable, but, in Human, it regularly fell in the TE-free region in all genes in which this region was at least 640 nt long.

In the special case of the Mouse slc25a15 gene the polyA signal (ATTAAA) was entirely provided by the first TE, an ORR1C2_LTR insert derived from the ERV3 murine endogenous retrovirus.

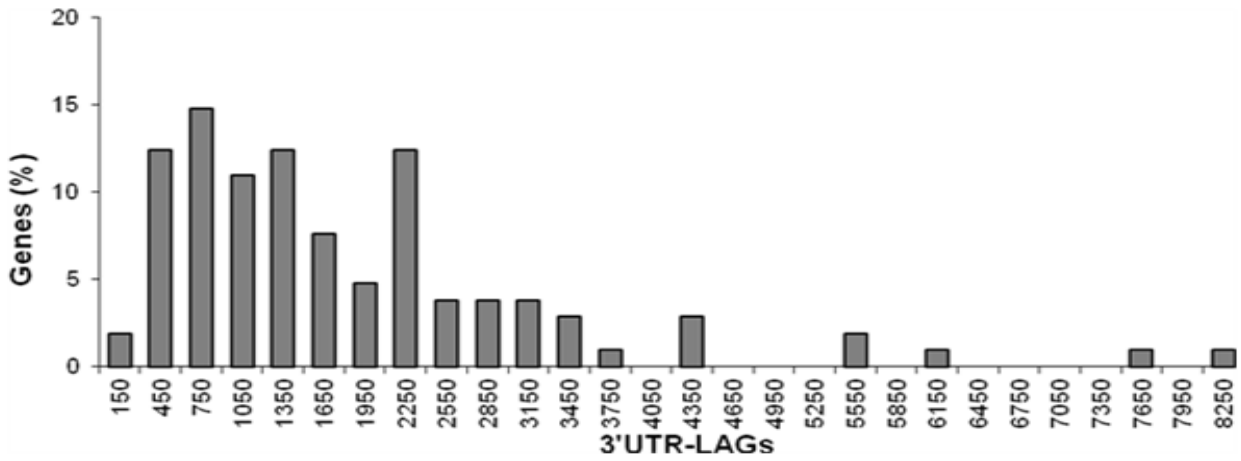

Figure 4: Distribution of 3'UTR-LAGs. Cumulative distribution of all SLC25 Mouse and Human genes. 3'UTR-LAGs are the distances (in nucleotides) from the stop codon to the beginning of the first downstream TE.

Modeling the TE distributions at the 5' and 3' flanking regions in Mouse and Human

We attempted a general modeling of the spatial distribution of the most proximal TEs at the 5' and 3' flanking regions. If the TEs did insert in the flanking regions at purely random sites, the frequency distribution of the minimum TE distances would be expected to decay exponentially from shorter distances to longer distances, as shown in Figure: 5 by the dot-and-line graph of the function y = 30.5 e-0.233 (LAG/300) (where 30.5 is a scaling factor). In Figure:5 the gray columns represent the actual combined distribution of the 5' and 3'UTR-LAGs. In this simulation, the theoretical frequencies match the actual values for the UTR-LAGs 750 and higher, while at lower LAGs the actual frequency is reduced by comparison to the theoretical frequency by 77 % at LAG 150 and 32 % at LAG 450. Thus, the spatial distribution of TEs at UTRs seem to be essentially random at a certain distance from the start and stop codons, whereas the TE settlement seems to be the more "inhibited" the more a specific UTR section is close to the coding regions.

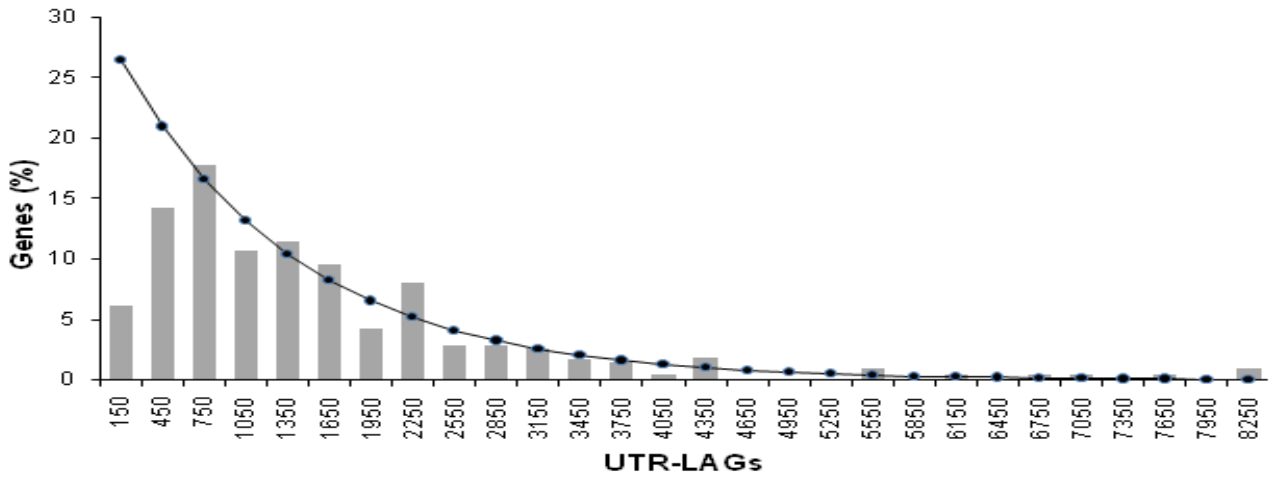

Figure 5: Dots and exponential interpolating line: the theoretical frequency distribution of genes according to the 5'UTR-LAGs or 3'UTR-LAGs in the hypothesis of TEs inserting at purely random sites. Column graph: the combined actual distributions of the Mouse and Human 5' and 3'UTR-LAGs.

TEs at 5' and 3' flanking regions [5'UTR-LAGs and 3'UTR-LAGs] in Zebrafish and modelin

The individual 5'UTR-LAGs and 3'UTR-LAGs are listed in the Suppl-Tables 9 and 12, respectively. The frequency distributions are very similar at 5' and 3' and they will be treated cumulatively. The resulting distribution was unimodal and right-skewed as in Mouse and Human, but peaked at 450 nt and the average was about 800 nt; the Excel asymmetry index (1.66) was also lower than in the corresponding Mouse and Human distributions (Figure: 6).

The actual distribution may be modeled as previously explained (Figure:6; dots and exponential interpolating line, plotting the function y = 70.0 e-0.493 (LAG/300), where 70.0 is a scaling factor). The theoretical frequencies match the actual values for the UTR-LAGs 450 and higher, while at LAG 150 the actual frequency is reduced in comparison to the theoretical frequency by about 65 %. While the Mouse/Human UTR-LAGs in corresponding genes are positively correlated, the Mouse/Zebrafish and Human/Zebrafish UTR-LAGs are not significantly correlated.

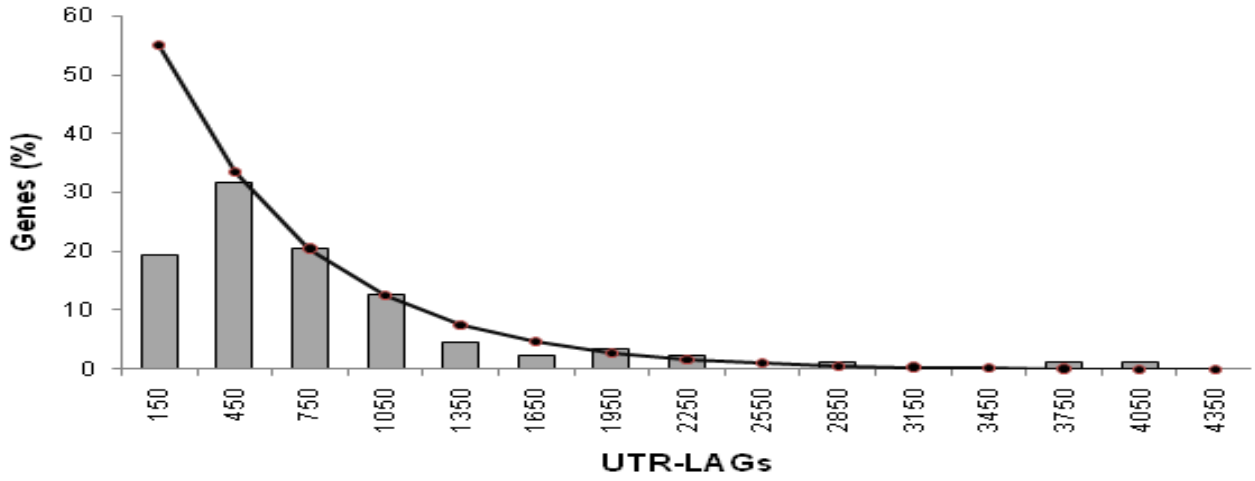

Figure 6: Column graph: cumulative distribution of the 5'UTR-LAGs and 3'UTR-LAGs in Zebrafish genes. Dots and exponential interpolating line: the theoretical frequency distribution of genes according to the UTR-LAGs in the hypothesis of TEs inserting at purely random sites.

Discussion

TE type and direction

In Human and Mouse, LTR elements have been reported to be underrepresented in the proximity to genes and in the introns, as compared to the intergenic regions [35,51]. This may be due to a negative selection at these sites, since the LTRs can impact the genome in many ways, being extremely recombinogenic and carrying transcriptional regulatory signals very similar to those in genes [52,53]. Indeed, there is evidence of transcriptional disruption of some Mouse genes by LTR element insertions [35]. We have found that, like in mammals, also in Zebrafish LTR elements are underrepresented in the proximity of coding sections (Results: The nature of TEs and Table 2).

At some sites the direction (sense/antisense) of TEs may exhibit significant biases. At the intron 3' end, SINEs have been reported to be strongly antisense-biased in Mouse, but not in Human [35]. We were unable to confirm this finding, but we found that the SINE1 TEs at the 5' UTR of Mouse genes are oriented with a significant antisense bias (Results: The nature of TEs). Selection against sense-oriented elements is likely due to the greater chance that such elements will disrupt gene transcript processing [51]. In Zebrafish, according to a previous report and our own data, no direction bias in the different TE classes could be detected. Obviously, this issue deserves to be further investigated [2].

In the first intron of Mouse and Human SLC25A42 the initial GT is changed into GC (Results: General). This is a relatively rare event occurring in less than 1% of cases [54]. In Zebrafish the corresponding intron starts by GT, indicating that the mutation occurred relatively late in the Mouse and Human common ancestor.

Intronic TEs and the control of splicing

TEs appear to be randomly distributed in the different intronic sections of the genes. Lack of preference for specific introns is also indicated by the observation that the number of TE-bearing introns shared by the different species is lower than the expectance based on random coincidences (calculated from the percentages of introns bearing significant TEs, Results: General): Mouse and Human (Suppl-Tables 1 and 4 and 2 and 5) 4 vs. expected 4.8; Mouse and Zebrafish (Suppl-Tables 1 and 4 and 3 and 6) 6 vs. 8.8; Human and Zebrafish (Suppl-Tables 2 and 5 and 3 and 6) 3 vs. 8.5.

It had been already reported that in Mouse and Human the frequency of TEs is in general lower at both intron ends, especially so at the 3' end [35]. In these species most of TEs are non-LTR Retrotransposons and DNA transposons are poorly represented (Table 2). Our present results confirm this finding for the family of mitochondrial carrier genes of Mouse and Human and add novel data on the homologous introns of Zebrafish, a spe-

cies distantly related to mammals and in which the vast majority of TEs belongs to the class of DNA transposons (Table 2). Despite such differences, the Zebrafish pattern of distribution of the TE inserts paralleled the mammalian distribution. In all, in our material, the TE frequency was 0.67 per unit distance (nt) in the initial 0-20 nucleotides of the intron 5' end and 0.96 in the more downstream 21-100 nucleotides (Results: Exon-next TE distances at the 5' end of the introns [INTRON-5'-LAGs]). At the intronic 3' end the corresponding figures were 0.19 and 0.99 (Results: TE/next-exon distances at the 3' end of the introns [INTRON-3'-LAGs]). Statistical analysis demonstrates that at both 20-nt intron ends the TE density is reduced as compared to the adjoining sections and, in addition, the density is significantly lower at the 3' 20-nt end than at the corresponding 5' end.

The pre-mRNA splicing of introns is a very complex and not fully understood process in which there is thought to be an interplay among several components. The effector machine is the spliceosome, a massive ribonucleoprotein complex which responds to an ensemble of signals which likely originate from both the intron concerned and the flanking exons harboring specific splicing enhancer and silencer motifs. The seemingly more robust signals originate from the nucleotide sequences at the 5' and 3' ends of each intron, the so called 5′ and 3′ splice site (5′ss and 3'ss, or splice donor and acceptor site, respectively)[55].

Based on sequence conservation in homologous genes it has been estimated that the effective 5'ss signal stretches for a maximum of 5 or 6 nt beyond the initial GT, i.e., from +1 to +7 or +8 [56,58]. However, the nucleotide composition of this segment is very variable: in a study on the first six nucleotides in 216 couples of orthologous Mouse/Human cytokine receptor gene introns we found 51 different configurations and 75 different configurations in the present study [59].

The arrangement at the 3' end of introns is more complicated. A "branch site" (a very short sequence including an adenine nucleotide) is localized at a variable distance (averaging 33–34 nt in Mammals); from the terminal AG. Further downstream there is a short polypyrimidine tract (rich in Ts and Cs) followed by a few variable nucleotides up to the AG end. The more evident T positive bias seems to extend in Vertebrates approximately from -5 to -20 with a peak at -10 [58,60]. On the whole, the typical sequence span of the 3'ss signal has been estimated to stretch approximately from -16 down to the AG end [56]. The nucleotide composition at 3' end is even more variable than at the 5' end: in our study on the last six nucleotides in 216 couples of orthologous Mouse/Human cytokine receptor gene introns we found as many as 94 different configurations and 102 different configurations in the present study [61].

TE elements and host genes co-evolved and selection was responsible for balancing the tension between retrotransposon proliferation and host survival so that the actual TE distribution may depend largely upon a secondary purging selection [35,62,63]. The relative TE depletion at the intron ends observed in Mouse, Human and Zebrafish likely results from a negative selection of genes harboring TEs which could overlap the 5' or 3'ss. The evolutionary control would be more stringent at the 3' where the splicing signal involves a longer series of nucleotides.

However, in some instances the TEs were likely to fell partly into the 5' splicing signal regions. In Human SLC25A14 (Suppl-Table 2) the INTRON-5'-LAG following the fifth exon is 5 and the initial hexanucleotide is GTAAGA. In Human SLC25A15 (Suppl-Table 2) the INTRON-5'-LAG following the first exon is 3 and the initial hexanucleotides is GTACAG. Both hexanucleotides are often found at the 5' end of other introns. Thus, due to the high degeneracy of the short 5'ss, the leftmost part of a TE could happen to integrate a genuine splicing signal.

At 3', in the Zebrafish slc25a42 the INTRON-3'-LAG upstream of the fifth exon is 15 and in slc25a45 the INTRON-3'-LAG upstream of the second exon is only 3 (Suppl-Table 6). The terminal segments of the corresponding TEs are not sufficiently rich in Ts and Cs to integrate a polypyrimidine tract (details not shown). On the contrary, in Zebrafish slc25a45, where the INTRON-3'-LAG upstream of the forth exon is 9 (Suppl-Table 6), the terminal segments of the corresponding TE is TCTTCTCT.

The comparative analysis demonstrates that the genetically fixed Zebrafish/Mouse/Human scheme of exon/intron alternation keeps unaltered even when a TE settles near (< 100 nt) the splicing sites. It is thus concluded that the extant TEs do not exert any short-range disturbing activity on the splicing process. The extant TEs which are at a very short distance from the intron ends in certain cases may be even integrated in the splicing signal, while in the other cases there seem to be a relatively wide tolerance.

TEs in 5' and 3' flanking UTRs

The issues of TE differences between the 5' and 3' UTRs and of TE distances from the start and stop of the translation have not been specifically addressed so far.

Minding that our results make reference only to the TEs more nearing the coding region, it may be of interest that some TEs (SINE1 and SINE2 in Mouse and Mariner in Zebrafish) are asymmetrically represented in the two UTRs (Results: The nature of TEs).

Should TEs in the proximity of start or stop codons randomly located, the distribution of the minimum TE distances would have a maximum next to these codons, then declining exponentially. However, the actual distribution of the minimum distances exhibits relatively low frequencies near start and stop codons, then rises to a maximum (at about 750 nt in Mouse and Human and 450 nt in Zebrafish) to decline progressively thereafter. Superimposition of a suitable exponentially-decaying distribution shows that at the shortest distances the actual frequency is much lower than the theoretical frequency, but for higher distances the gap decreases and eventually the two distributions become similar. Therefore, it appears that the theoretical stochastic distribution is modified by a deterministic component which acts abating the short distances, viz., preventing TEs to settle too close to starts and stops of the translation region, this inhibition being the stronger the closer to the translation region, then vanishing at longer distances.

We were unable to correlate the extent of these upstream and downstream "relative inhibition regions" with the position of the regulatory sequences which may be present up-

stream and downstream of the translation region. Information on the proximal promoter sites is too scanty to allow general conclusions. In Mouse and Human genes for which the polyA signal position was available this sequence fell in the TE-free region in all genes in which this region was at least 640 nt long, but in other instances the polyA signal was located downstream of the first significant TE insert.

In summary, at both UTRs the net balance between insertion site preferences and selection tends to determine in mammals and Zebrafish a TE-free area of very varying extent, from a few tens up to a few thousands of nucleotides. On average, the TE-free area is narrower in Zebrafish than in mammals (Suppl-Tables 7-12). In addition, it is remarkable that the TE-free area at the UTRs is much wider than the TE-underrepresentation area at the intron ends. The present results shed no clear light on the relationships between these TE-free areas and the position of the regulating sequences hosted in the UTRs. Possibly the recorded more proximal TEs do not interfere with the regulatory sequences and the positioning of TEs and these sequences are independently regulated, as suggested by the TE/polyA signal relative positioning. But the hypothesis that TEs may in some way modulate the main regulatory signals cannot be ruled out. Obviously, all these poorly understood issues need further investigation.

As already noticed, a strong deterministic component regulates the exon/intron alternation in the mitochondrial solute carrier genes of vertebrates; furthermore, in these genes the lengths of the individual homologous Mouse/Human introns are strongly positively correlated [61]. The present results show that the Mouse/Human UTR-LAGs in corresponding genes are positively correlated, while the Mouse/Zebrafish and Human/Zebrafish UTR-LAGs are not correlated. These results illustrate the existence of a residual common deterministic component regulating the structure of Mouse and Human introns; such component, likely inherited from the common ancestor, had persisted despite the extensive intron re-editing and the incorporation of different TE species after the rodent/primate divergence, about 100 million years ago.

We conclude that, unlike some large intronic retroviral insertions which have been found to affect the expression of coding sequences even at a certain distance the extant TE inserts present in mitochondrial solute carrier genes, despite the differences in location and type in different species, do not alter the structural exon/intron plane of the individual genes, which has been kept conserved from fish to mammals along about 400-450 million years [64,65,66]. Such strict evolutionary control is possibly to be put in relation to the essential role of these genes in cell activity and survival.

References

1. Chalopin D, Naville M, Plard F, Galiana D, Volff JN. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. Genome Biol. Evol. 2015;7(2):567-580. doi:10.1093/gbe/evv005
2. Howe K, Clark MD, Torroja CF, Carlos F. Torroja, James Torrance, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature 2013;496(7446):498-503. doi: 10.1038/nature12111
3. Ayarpadikannan S, Kim HS. The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. Genomics Inform. 2014;12(3):98-104. doi: 10.5808/GI.2014.12.3.98
4. Kramerov DA, Vassetzky NS. Short retroposons in eukaryotic genomes. Int. Rev. Cytol. 2005;247:165-221. doi: 10.1016/S0074-7696(05)47004-7
5. Ichiyanagi K. Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. Genes Genet. Syst. 2013;88(1):19-29. doi.org/10.1266/ggs.88.19
6. Dye MJ, Gromak N, Haussecker D, West S, Proudfoot NJ. Turnover and function of noncoding RNA polymerase II transcripts. Cold Spring Harb. Symp. Quant. Biol. 2006;71:275-284. doi: 10.1101/sqb.2006.71.040
7. Barrett LW, Fletcher S, Wilton SD. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. Cell. Mol. Life Sci. 2012;69(21):3613-3634. doi: 10.1007/s00018-012-0990-9
8. Jia J, Yao P, Arif A, Fox PL. Regulation and dysregulation of 3'UTR-mediated translational control. Curr. Opin. Genet. Dev. 2013;23(1):29-34. doi: 10.1016/j.gde.2012.12.004
9. Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science 2016; 351(6274): aac7247. doi: 10.1126/science.aac7247
10. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116(2):281-297. doi.org/10.1016/S0092-8674(04)00045-5
11. Ameres SL, Zamore PD. Diversifying microRNA sequence and function. Nat. Rev. Mol. Cell. Biol. 2013;14(8):475-488. doi: 10.1038/nrm3611
12. Lee I, Ajay SS, Yook JI, Kim HS, Hong SH, Kim NH, Dhanasekaran SM, Chinnaiyan AM, Athey BD. New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. Genome Res. 2009;19(7):1175-1183. doi: 10.1101/gr.089367.108
13. Ajay SS, Athey BD, Lee I. Unified translation repression mechanism for microRNAs and upstream AUGs. BMC Genomics. 2010;11:155. doi: 10.1186/1471-2164-11-155
14. Smalheiser NR, Torvik VI. Alu elements within human mRNAs are probable microRNA targets. Trends Genet. 2006;22(10):532-536. doi:10.1016/j.tig.2006.08.007
15. Hoffman Y, Dahary D, Bublik DR, Oren M, Pilpel Y. The majority of endogenous microRNA targets within Alu elements avoid the microRNA machinery. Bioinformatics 2013;29(7):894-902. doi: 10.1093/bioinformatics/btt044
16. Chen YA, Aravin AA. Non-Coding RNAs in Transcriptional Regulation: The review for Current Molecular Biology Reports. Curr. Mol. Biol. Rep. 2015;1(1):10-18. doi: 10.1007/s40610-015-0002-6
17. Polak P, Domany E. Alu elements contain many binding sites for transcription factors and may play a role in regulation of developmental processes. BMC Genomics. 2006;7:133. doi: 10.1186/1471-2164-7-133
18. Häsler J, Strub K. Alu elements as regulators of gene expression. Nucleic Acids Res.2006;34(19):5491-5497. doi: 10.1093/nar/gkl706
19. Deininger P. Alu elements: know the SINEs. Genome Biol. 2011;12(12):236. doi: 10.1186/gb-2011-12-12-236
20. Peaston AE, Evsikov AV, Graber JH, de Vries WN, Holbrook AE, Solter D, Knowles BB. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev. Cell. 2004;7(4):597-606. doi: 10.1016/j.devcel.2004.09.004
21. Sorek R, Lev-Maor G, Reznik M, Dagan T, Belinky F, Graur D, et al. Minimal conditions for exonization of intronic sequences: 5' splice site formation in alu exons. Mol. Cell. 2004;14(2):221-231. doi:10.1016/S1097-2765(04)00181-9
22. Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene. 2009;448(2):105-114. doi: 10.1016/j.gene.2009.06.020
23. Hughes DC. Alternative splicing of the human VEGFGR-3/FLT4 gene as a consequence of an integrated human endogenous retrovirus. J. Mol. Evol. 2001;53(2):77-79. doi: 10.1007/s002390010195
24. Sorek R, Ast G, Graur D. Alu-containing exons are alternatively spliced. Genome Res. 2002;12(7):1060–1067. doi: 10.1101/gr.229302

25. Del Arco A. Novel variants of human SCaMC-3, an isoform of the ATP-Mg/Pi mitochondrial carrier, generated by alternative splicing from 3'-flanking transposable elements. Biochem. J. 2005;389(Pt3):647-655. doi: 10.1042/BJ20050283
26. Buzdin AA. Retroelements and formation of chimeric retrogenes. Cell. Mol. Life Sci. 2004;61(16):2046-2059. doi: 10.1007/s00018-004-4041-z
27. Schmitz J, Brosius J. Exonization of transposed elements: A challenge and opportunity for evolution. Biochimie 2011; 93(11):1928-1934. doi: 10.1016/j.biochi.2011.07.014
28. Chatterjee S, Pal JK. Role of 5'- and 3'-untranslated regions of mRNAs in human diseases. Biol. Cell. 2009;101(5): 251-262. doi: 10.1042/BC20080104
29. Vislovukh A, Vargas TR, Polesskaya A, Groisman I. Role of 3'-untranslated region translational control in cancer development, diagnostics and treatment. World J. Biol. Chem. 2014;5(1):40-57. doi: 10.4331/wjbc.v5.i1.40
30. Sironi M, Menozzi G, Comi GP, Cereda M, Cagliani R, Bresolin N, et al. Gene function and expression level influence the insertion/fixation dynamics of distinct transposon families in mammalian introns. Genome Biol. 2006;7:R120. doi: 10.1186/gb-2006-7-12-r120
31. Tsirigos A, Rigoutsos I. Alu and B1 repeats have been selectively retained in the upstream and intronic regions of genes of specific functional classes. PLoS Comput. Biol. 2009;5(12), e1000610
32. Ho JW, Ho PW, Zhang WY, Liu HF, Kwok KH, Yiu DC, Chan KH, Kung MH, Ramsden DB, Ho SL. Transcriptional regulation of UCP4 by NF-kappaB and its role in mediating protection against MPP+ toxicity. Free Radic. Biol. Med. 2010;49(2):192-204. doi: 10.1016/j.freeradbiomed.2010.04.002
33. Medstrand P, van de Lagemaat LN, Mager DL. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res. 2002;12(10):1483- 1495. doi: 10.1101/gr.388902
34. Costantini M, Auletta F, Bernardi G. The distributions of "new" and "old" Alu sequences in the human genome: the solution of a "mystery". Mol. Biol. Evol. 2012;29(1):421-427. doi: 10.1093/molbev/msr242
35. Zhang Y, Romanish MT, Mager DL. Distributions of transposable elements reveal hazardous zones in mammalian introns. PLoS Comput. Biol. 2011;7(5),e1002046. doi: 10.1371/journal.pcbi.1002046
36. Palmieri F. The mitochondrial transporter family SLC25: identification, properties and physiopathology. Mol. Aspects Med. 2013;34(2-3):465-484. doi: 10.1016/j.mam.2012.05.005
37. Palmieri F. Mitochondrial transporters of the SLC25 family and associated diseases: a review. J. Inherit. Metab. Dis. 2014;37(4):565-575. doi: 10.1007/s10545-014-9708-5
38. Kohany O, Gentles AJ, Hankus L, Jurka J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. BMC Bioinformatics. 2006;7:474. doi: 10.1186/1471-2105-7-474
39. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990;215(3): 403-410.doi:10.1016/S0022-2836(05)80360-2
40. Dreos R, Ambrosini G, Périer R, Bucher P. EPD and EPDnew, high-quality promoter resources in the next-generation sequencing era. Nucleic Acids Res. 2013;41(Database issue):D157-D164. doi: 10.1093/nar/gks1233
41. Samuels ML, Witmer JA. Statistics for the Life Sciences, 3rd edition. Prentice Hall, Upper Saddle River, NJ, USA 2002.
42. Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J. Evolutionary history of 7SL RNA-derived SINEs in Supraprimates. Trends Genet. 2007;23(4):158-161. doi: 10.1016/j.tig.2007.02.002
43. Kapitonov VV, Jurka J. A novel class of SINE elements derived from 5S rRNA. Mol. Biol. Evol. 2003;20(5):694-702. doi: 10.1093/molbev/msg075
44. Iacobazzi V, Infantino V, Palmieri F. Epigenetic mechanisms and Sp1 regulate mitochondrial citrate carrier gene expression. Biochem. Biophys. Res. Commun. 2008;376(1):15-20. doi: 10.1016/j.bbrc.2008.08.015
45. Iacobazzi V, Convertini P, Infantino V, Scarcia P, Todisco S, Palmieri F. Statins, fibrates and retinoic acid upregulate mitochondrial acylcarnitine carrier gene expression. Biochem. Biophys. Res. Commun. 2009a; 388(4):643-647. doi: 10.1016/j.bbrc.2009.08.008
46. Iacobazzi V, Infantino V, Costanzo P, Izzo P, Palmieri F. Functional analysis of the promoter of the mitochondrial phosphate carrier human gene: identification of activator and repressor elements and their transcription factors. Biochem. J. 2005;391(Pt3):613-621. doi: 10.1042/BJ20050776
47. Biswas A, Senthilkumar SR, Said HM. Effect of chronic alcohol exposure on folate uptake by liver mitochondria. Am. J. Physiol. Cell Physiol. 2012;302(1):C203-C209. doi: 10.1152/ajpcell.00283.2011
48. Iacobazzi V, Infantino V, Convertini P, Vozza A, Agrimi G, Palmieri F. Transcription of the mitochondrial citrate carrier gene: identification of a silencer and its binding protein ZNF224. Biochem. Biophys. Res. Commun. 2009b;386(1):186-191. doi: 10.1016/j.bbrc.2009.06.003
49. Convertini P, Infantino V, Bisaccia F, Palmieri F, Iacobazzi V. Role of FOXA and Sp1 in mitochondrial acylcarnitine carrier gene expression in different cell lines. Biochem. Biophys. Res. Commun. 2011;404(1):376-381. doi: 10.1016/j.bbrc.2010.11.126
50. Nabokina SM, Valle JE, Said HM. Characterization of the human mitochondrial thiamine pyrophosphate transporter SLC25A19 minimal promoter: a role for NF-Y in regulating basal transcription. Gene. 2013;528(2): 248-255. doi: 10.1016/j.gene.2013.06.073
51. van de Lagemaat LN, Medstrand P, Mager DL. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. Genome Biol. 2006;7(9):R86. doi: 10.1186/gb-2006-7-9-r86
52. Kumar CS, Qureshi SF, Ali A, Satyanarayana ML, Rangaraju A, Venkateshwari A, Nallari P. Hidden magicians of genome evolution. Indian J. Med. Res. 2013;137(6):1052– 1060
53. Zaratiegui M. Influence of long terminal repeat retrotransposons in the genomes of fission yeasts. Biochem. Soc. Trans. 2013;41(6),1629-1633. doi: 10.1042/BST20130207
54. Kralovicova J, Hwang G, Asplund AC, Churbanov A, Smith CI, Vorechovsky I. Compensatory signals associated with the activation of human GC 5' splice sites. Nucleic Acids Res. 2011;39(16):7077-7091. doi: 10.1093/nar/gkr306
55. Nilsen TW. The spliceosome: the most complex macromolecular machine in the cell? BioEssays. 2003;25(12): 1147-1149. doi: 10.1002/bies.10394
56. Abril JF, Castelo R, Guigó R. Comparison of splice sites in mammals and chicken. Genome Res. 2005;15(1):111–119. doi: 10.1101/gr.3108805
57. Perincheri S, Dingle RW, Peterson ML, Spear B.T. Hereditary persistence of alpha-fetoprotein and H19 expression in liver of BALB/cJ mice is due to a retrovirus insertion in the Zhx2 gene. Proc. Natl. Acad. Sci. USA. 2005;102(2):396-401. doi: 10.1073/pnas.0408555102
58. Schwartz SH, Silva J, Burstein D, Pupko T, Eyras E, Ast G. Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. Genome Research. 2008;18(1):88–103. doi: 10.1101/gr.6818908.
59. Calvello R, Cianciulli A, Panaro MA. Conservation/Mutation in the Splice Sites of Cytokine Receptor Genes of Mouse and Human. Int. J. Evol. Biol. 2013;2013:818954. doi: 10.1155/2013/818954
60. Kol G, Lev-Maor G, Ast G. Human–mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. Hum. Mol. Genet. 2005;14(11):1559- 1568. doi: 10.1093/hmg/ddi164
61. Cianciulli A, Calvello R, Panaro MA. Determinism and randomness in the evolution of introns and SINE inserts in mouse and human mitochondrial solute carrier and cytokine receptor genes. Comput. Biol. Chem. 2015;55:49-59. doi: 10.1016/j.compbiolchem. 2015.02.012
62. Beauregard A, Curcio MJ, Belfort M. The take and give between retrotransposable elements and their hosts. Annu. Rev. Genet. 2008;42:587-617. doi: 10.1146/annurev.genet.42.110807.091549
63. Brady T, Lee YN, Ronen K, Malani N, Berry CC, Bieniasz PD, Bushman FD. Integration target site selection by a resurrected human endogenous retrovirus. Genes Dev. 2009;23(5):633–642. doi: 10.1101/gad.1762309
64. Chang CM, Coville JL, Coquerelle G, Gourichon D, Oulmouden A, Tixier-Boichard M. Complete association between a retroviral insertion in the tyrosinase gene and the recessive white mutation in chickens. BMC Ge-

nomics. 2006;7:19. doi: 10.1186/1471-2164-7-19

65. Nobrega MA, Pennacchio LA. Comparative genomic analysis as a tool for biological discovery. J. Physiol. 2004; 554(Pt1):31-39. doi: 10.1113/jphysiol.2003.050948

66. Broughton RE, Betancur-R R, Li C, Arratia G, Ortí G. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLoS Curr. 2013;5. pii:ecurrents.tol.2ca8041495ffafd0c92756e75247483e. doi: 10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e