Cytogenet Genome Res 110:108–116 (2005)
DOI: 10.1159/000084942

Cytogenetic and Genome Research

How repeated retroelements format genome function

R. von Sternberg,[a] J.A. Shapiro[b]

[a] National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD and Department of Systematic Biology, National Museum of Natural History, Smithsonian Institution, Washington, DC; [b] Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL (USA)

Manuscript received 1 August 2004; accepted in revised form for publication by J.-N. Volff 3 August 2004.

Abstract. Genomes operate as sophisticated information storage systems. Generic repeated signals in the DNA format expression of coding sequence files and organize additional functions essential for genome replication and accurate transmission to progeny cells. Retroelements comprise a major fraction of many genomes and contain a surprising diversity of functional signals. In this article, we summarize some features of the taxonomic distribution of retroelements, especially mammalian SINEs, tabulate functional roles documented for different classes of retroelements, and discuss their potential roles as genome organizers. In particular, the fact that certain retroelements serve as boundaries for heterochromatin domains and provide a significant fraction of scaffolding/matrix attachment regions (S/MARs) suggests that the reversed transcribed component of the genome plays a major architectonic role in higher order physical structuring. Employing an information science model, the "functionalist" perspective on repetitive DNA leads to new ways of thinking about the systemic organization of cellular genomes and provides several novel possibilities involving retroelements in evolutionarily significant genome reorganization.

In the 21st Century, it is appropriate to think about DNA as a data storage medium and about genomes as sophisticated computational information storage systems (Sternberg, 1996, 2000, 2002; Shapiro, 1999, 2002; Shapiro and Sternberg, 2004). Like electronic computational storage systems, DNA molecules contain not only data files (coding sequences) but also generic repeated signals. These repetitive signals format the genome for expression, replication, transmission, repair and restructuring. They serve as the physical basis for integrating different segments of genomic DNA into computationally accessible systems and subsystems for the execution of complex cellular routines, such as cell division and differentiation. Applying the informatics metaphor allows us to understand the functional significance of the surprisingly abundant fraction of repetitive DNA sequences found in virtually all genomes, including prokaryotes (data summarized in Shapiro and Sternberg, 2004).

Request reprints from James A. Shapiro
Department of Biochemistry and Molecular Biology
University of Chicago, 920 E. 58th Street
Chicago, IL 60637 (USA); telephone: 773-702-1625
fax: 773-947-9345; e-mail: jsha@uchicago.edu

Since retroelements constitute a large part (often the majority) of genomic repetitive DNA, this review summarizes the documented functional properties of reverse transcribed DNA sequences. We omit the roles of retroelements in genome restructuring because that has been well covered in recent reviews (Kazazian, 2000; Deininger et al., 2003).

Diverse genomic functions associated with retroelements

Table 1 presents over 30 examples where functional activity has been assigned to a particular retroelement. The genome functions range from providing promoter and enhancer activity to modulating transcript elongation, targeting mRNA to specific tissues, stimulating mRNA translation, providing replication origin recognition sequences, contributing to pericentromeric heterochromatin, serving as telomere caps, nucleating heterochromatin in chromosome arms, supplying chromatin boundary signals, and providing S/MAR attachment sites.

The list is far from exhaustive. Many well-documented examples of repeat element function are probably excluded because the origin of the cognate repeats by reverse transcription may not be known. From the cases cited in Table 1, it is

1424–8581/05/1104–0108$22.00/0

Table 1. Selected examples of retroelement functions

Function	Structural class	Example	Comment	Reference
Transcription				
Promoters	Transposable elements		TE sequences in almost a quarter of human promoter sequences	Jordan et al., 2003
	LINE	Human LINE-1	1.6% of 2004 examined human promoters include LINEs; the 5'-untranslated region of L1 has both an internal (sense) promoter and an antisense promoter (ASP); L1 ASP chimeric transcripts are highly represented in expressed-sequence tag (EST) databases.	Speek, 2001 Zaiss and Kloetzel, 1999 Nigumann et al., 2002 Jordan et al., 2003
	SINE	Human Alus; mouse B2 elements	Genomic synonyms; RNA polymerase II promoter elements-5.3% of 2004 examined human promoters have SINEs as components	Ferrigno et al., 2001 Jordan et al., 2003
Enhancers and silencers	LINE	Human LINE-1	Positive transcriptional regulatory element; binding sites for SRY, YY1 factors	Yang et al., 1998 Becker et al., 1993 Tchénio et al., 2000
	SINE	Subgroup II-III of human AluSx subfamily	Nuclear hormone receptor binding sites for thyroid hormone receptor, retinoic acid receptor and estrogen receptor	Norris et al., 1995 Vansant and Reynolds, 1995 Babich et al., 1999
		Jo, Jb, Sq, Sp, Sx, and Sg subfamilies of human Alus; subset of rodent B1 elements;	Genomic synonyms; Pax6 binding sites	Zhou et al., 2000, 2002
Transcription attentuation	LINE	LINE-1	Retards transcript elongation in a strand asymmetric manner	Han et al., 2004
Regulatory RNAs	LTR retrotransposon	Mouse VL30 elements	Non-protein coding transcripts of VL30 elements selectively bind to PSF repressor, allowing transcription of genes controlled by insulin-like growth factor response elements (IGFRE); the VL30 transcripts are causally involved in steroidogenesis and oncogenesis	Song et al., 2004
Post-transcriptional RNA processing				
mRNA targeting	SINE	Rodent ID elements	Target mRNAs to neuronal dendrites; genomic synonym	Chen et al., 2003
		Rodent BC200 and primate homologue	Neuronal targeting; genomic synonym	Skyrabin et al., 1998
		Primate Alu	Neuronal targeting; genomic synonym	Watson and Sutcliffe, 1987
RNA editing	SINE	Human Alu	92% of adenosine to inosine editing of pre-mRNAs in the human transcriptome occurs in Alu elements	Levanon et al., 2004
Translation				
Selective enhancement of mRNA translation	SINE	Human Alus; mouse B1, B2 elements.	Genomic synonyms	Rubin et al., 2002
DNA replication, localization and movement				
Origins	LTR and unclassified	*S. cerevisiae* LTR, subtelomeric X and Y' repeats	Contain 20% of *S. cerevisiae* sequences that immunoprecipitate with origin recognition proteins	Wyrick et al., 2001
Centromeres	LTR	Cereal centromere repeats		Aragon-Alcaide et al., 1996
		CRR in rice		Cheng et al., 2002
		maize CRM; CentA, Huck and Prem2	CENH3 interacts specifically with CRM	Ananiev et al., 1998b Zhong et al., 2002 Nagaki et al., 2003
		Arabidopsis Athila element		Pelissier et al., 1996
		250, 301 bp repeats in wheat and rye	Ty3/gypsy-related	Cheng and Murata, 2003
		Ty3/gypsy family Sorghum elements;	Ty3/gypsy-related sequences present exclusively in the centromeres of all sorghum chromosomes; Ty1/copia-related DNA sequences are not specific to the centromeric regions	Miller et al., 1998
Telomeres	non-LTR retrotransposons	*Drosophila* HeT-A, TART elements	HeT-A units work in pairs: the 5' element has a promoter in the 3' UTR that allows transcription of the adjacent template-unit.	Pardue and DeBaryshe, 2003
		Plasmodium telomere associated repetitive elements (TAREs)	*Plasmodium falciparum* telomere-associated sequences of the 14 linear chromosomes display a similar higher order organization and form clusters of four to seven telomeres localized at the nuclear periphery.	Figueiredo et al., 2000, 2002
		Giardia telomere retroposons		Arkhipova and Morrison, 2001
	LTR retrotransposon	Yeast Ty1	Ty1 activated when normal telomere function impaired	Scholes et al., 2003

Table 1 (continued)

Function	Structural class	Example	Comment	Reference
S/MARs (scaffold/matrix associated regions)	LTR	Human LTR retrotransposons	7.0% of examined human S/MARs derived from LTR retrotransposons	Jordan et al., 2003
		Drosophila gypsy	Elements determining intranuclear gene localization/nuclear pore association	Nabirochkin et al., 1998 Gerasimova et al., 2000 Labrador and Corces, 2002
	LINEs	Human LINE-1	39.4% of human S/MARs are LINE sequences; 98 LINE1 consensus sequences were found to contain 14 distinct S/MAR recognition signatures; the distribution of Alu and LINE repetitive DNA are biased to positions at or adjacent to apoptotic cleavage sites.	Chimera and Musich, 1985 Rollini et al., 1999 Khodarev et al., 2000 Jordan et al., 2003
Chromatin organization and epigenetic modification				
Heterochromatin	LTR and non-LTR retroposons	*Drosophila* transposable elements	Nine transposable elements (copia, gypsy, mdg-1, blood, Doc, I, F, G, and Bari-1) are preferentially clustered into one or more discrete heterochromatic regions in chromosomes of the Oregon-R laboratory stock; P and hobo elements, recent invaders of the *D. melanogaster* genome exhibit heterochromatic clusters in certain natural populations.	Cryderman et al., 1998 Pimpinelli et al., 1995
	LTR retroposon	Hamster IAP elements	In Syrian hamster, over half of the genomic IAP elements are accumulated in heterochromatin, including the entire Y chromosome.	Dimitri and Junakovic, 1999
		Maize Grande, Prem2, RE-10, RE-15, and Zeon	Abundant in heterochromatic knob regions; blocks of tandem 180-bp repeats interrupted by insertions of full size copies of retrotransposable elements; about 30% of cloned knob DNA fragments.	Ananiev et al., 1998a
		Arabidopsis Athila	Athila elements in the *Arabidopsis* genome are concentrated in or near heterochromatic regions. Most of the heterochromatic elements retrotransposed directly into 180 bp satellite clusters.	Pelissier et al.,1996
		Several *Drosophila* LTR families	LTR elements represent 61% of euchromatic transposable elements and approximately 78% of heterochromatic elements. LINE elements represent 24% of the euchromatic and 17% of the heterochromatic transposable element sequence. TIR elements represent 15% in euchromatin and 5% in heterochromatin.	Hoskins et al., 2002
	LINE	Human LINE-1	X inactivation, monoallelic expression, imprinting	Bailey et al., 2000 Lyon, 2000 Parish et al., 2002 Allen et al., 2003
	Numerous LTR retrotransposon elements and SINEs	*Arabidopsis* retroelements	Interstitial (knob) heterochromatin is formed by the interaction of clusters of retroelements and related tandem repeats (with DNA transposons), the chromatin remodeling ATPase DDM1, and small interfering RNAs that are similar to the retroelements.	Lippman et al., 2004
Epigenetic memory elements	LTR retrotransposon	Mouse IAP; several maize and *Arabidopsis* LTR families	LTR elements near or in genes modify gene expression in a heritably metastable manner.	Chong and Whitelaw, 2004 Lippman et al., 2004 Lane et al., 2003
Methylation	SINE	Mouse B1	B1 elements methylated de novo to a high level after transfection into embryonal carcinoma cells; B1 elements acted synergistically.	Yates et al., 1999
Insulator/boundary elements	LTR retrotransposon	*Drosophila* gypsy element	The gypsy insulator blocks propagation of silencing and alters the nuclear localization of adjacent DNA.	Gerasimova et al., 2000 Chen and Corces, 2001 Labrador and Corces, 2002 Byrd and Corces, 2003

clear that the reverse-transcribed repetitive component of the genome carries a wide variety of generic signals that help organize the genome functionally and architecturally within the nucleus.

We are becoming increasingly aware of how the genome is organized at higher levels into multi-locus chromatin domains (van Driel et al., 2003). An architectural role for dispersed retroelements agrees with the conservation detected by comparative genomics in the positions and orientations of shared elements (Zhu et al., 2003; Silva et al., 2003). The observations on conserved repeats suggest that high numbers of "framework elements" may be retained in disparate mammalian genomes, with more derived subfamilies of LINEs, SINEs, and LTR elements being restricted to particular families and genera.

Taxonomically-specific genome system architecture

Our view of the genome as a hierarchically organized data storage system formatted by repetitive DNA sequence elements implies that each organism has a genome system architecture, in the same way that each computer has a characteristic architecture. In the computer example, architecture depends upon the operating system and hardware that are used, not upon the

content of each data file. Macintosh, Windows and Unix machines can all display the same images and text files, even though the data retrieval paths are operationally quite distinct. Similarly, many protein and RNA sequences (data files) are conserved through evolution, but different taxa organize and format their genomes in quite different ways for replication, transmission and expression. An overall system architecture is required since these processes must be coordinated to operate without mutual interference. DNA segments must be in the right place at the right time for function. In other words, the genome must be organized in space and time for operation.

A basic aspect of genome system architecture is the nature of signals that regulate transcription of different genetic loci. These signals include promoters and enhancers as well as determinants of epigenetic states that are either permissive or restrictive for transcription. Genome analysis is beginning to provide evidence of functional roles related to imprinting for evolutionarily "recent" LINE insertions. Nonorthologous L1 elements are similarly positioned asymmetrically in the X inactivation centers of human, mouse, and cow (Chureau et al., 2002), and L1 elements are significantly associated with monoallelically expressed loci in both human and mouse genomes (Allen et al., 2003).

From a perspective postulating that changes in repetitive elements may be important events in establishing specific new genome architectures, it is significant to note that each order of mammals has its own characteristic set of SINE elements (Table 2). Since these highly iterated SINEs are independently derived from cellular sequences, such as different tRNA or 7S RNA sequences, it is clear that taxonomic diversification among mammals involved many thousands of independent SINE element amplification and insertion events. Similarly, plant species can be discriminated by their pericentromeric repeats, a number of which are LTR retrotransposons (Table 1).

A frequently ignored feature of genome system architecture associated with repeat elements is overall genome size (Cavalier-Smith, 1985). In plants, genome size correlates with an increase in repetitive DNA abundance, particularly LTR retrotransposons (Meyers et al., 2001; Zhang and Wessler, 2004). Plant molecular geneticists have suggested that the total length of each genome is an important functional characteristic, which influences replication time, a characteristic that correlates with the length of the life cycle (Bennett, 1998; Bennetzen, 2000; Petrov, 2001). It makes sense that amplification of retroelements is an efficient method of altering total DNA content in the genome. Similarly, distance between regulatory and coding sequences may be an important control parameter (Zuckerkandl, 2002).

Evolutionary implications of retroelement formatting

The proposal that genomes have taxon-specific system architectures formatted by retroelements and other repeats mandates a serious examination of the morphological, physiological and reproductive effects of changes in the widely neglected repetitive component of the genome. Phenotypic variation that gives rise to adaptations is usually conceived in terms of altered gene products produced by mutations in protein-coding sequences. This view is too limited for two reasons. First, the organization of proteins can change without coding sequence modifications through the alteration of splicing patterns or the rearrangement of exons. Segmental duplications that generate new exon combinations are one group of such alterations (Eichler, 2001), and changes in RNA splicing patterns via the integration of retroelements into introns is another (Nekrutenko and Li, 2001). Retrotransduction and ectopic recombination between dispersed retroelements can lead to permutations of protein and RNA domains by duplications, deletions, and shuffling (Moran et al., 1999; Kazazian, 2000; Bailey et al., 2003; Deininger et al., 2003). Novel RNA splicing and modification codes are in turn provided by the integration of these information-rich sequences into coding regions. In the human transcriptome, for example, 92% of adenosine to inosine RNA editing sites occurs within Alu sequences (Levanon et al., 2004). The integration of an Alu element into a genetic locus can thus expand the array of transcripts produced by at least two distinct mechanisms (splicing, editing). In addition, of course, retroelements can be exapted into "neogenes," as seen in some olfactory receptor loci that lack introns, telomerase and syncytin (Brosius, 1999; Blackburn, 2000; Mallet et al., 2004).

A second way the classical evolutionary genetic view is unnecessarily limited is by restricting adaptive variation to mutations in an organism's repertoire of protein and RNA sequences. It has long been apparent that changes in the regulatory formatting of conserved coding sequences can result in novel developmental patterns, leading to new traits using the same assemblage of proteins and RNAs (Britten and Davidson, 1971). Ample evidence of this can be found in the *Drosophila* literature where major morphological changes have been tied to retroelement alterations. Developmental genetic studies of model organisms such as the mouse and *Arabidopsis* have made it clear that retroelements play a role in the epigenetic settings of the genome, both globally (in the form of heterochromatin arrays) and locally as gene control elements. When these settings are modified phenotypic changes result (Chong and Whitelaw, 2004; Lippman et al., 2004).

Comparative analysis of genomic sequences indicates that rearrangement of retroelements has played a significant role in reorganizing multicellular development circuits. The human genome provides many instances of regulatory regions embedded in the remnants of retroelements (Britten, 1996; Brosius, 1999; Jordan et al., 2003), and detailed studies have documented the participation of retroelements in regulation of coding sequence expression (e.g. Mozer and Benzer, 1994; Song et al., 2004). As can be seen in Table 1, retroelements play a role in every aspect of chromatin formatting and nuclear organization, which regulate the developmental expression of large chromosomal regions (van Driel et al., 2003). Thus, it is evident that the turnover of retroelement sequences can have a profound impact on genome structure and expression. Since retroelements have the potential *globally* to modulate heritable epigenetic states, is likely to prove rewarding to investigate how turnover of repeats involved in imprinting has led to taxonomic differences in developmental patterns involving the same protein and RNA-coding cassettes.

Table 2. Presence of SINE element families in mammalian taxa

SINE family	Element organization	*Mammalia*																														References
				Placentalia																												
				Boreoeutheria																												
				Afrotheria										*Laurasiatheria*												*Euarchontoglires*						
							Insectivora						*Cetartiodactyla*																			
												Cetacea				*Artiodactyla*				*Perissodactyla*	*Carnivora*	*Chiroptera*		*Primates*	*Rodentia*					*Lagomorpha*	*Dermoptera*	
		Monotremata	*Marsupialia*	*Hyracoidea*	*Proboscoidea*	*Sirenia*	*Erinaceidae*	*Soricidae*	*Talpidae*	*Macroscelidea*	*Tubulidentata*	*Mysticetes*	*Odontocetes*	*Ruminantia*	*Tylopoda*	*Suina*	*Hippopotamidae*			*Microchiroptera*		*Glitridae*	*Sciuridae*	*Castoridae*	*Hystricidae*	*Caviidae*						
															Camelidae	*Suidae*																
MIR/CORE	260 nt monomer consisting of 5’ tRNA-like region, non-tRNA segment, and A-rich 3’ end	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	Gilbert and Labuda, 1999 Jurka et al., 1995
AfroSINE-Anc	180 nt monomer consisting of 5’ tRNA-like region, non-tRNA segment, and A-rich 3’ end	–	–	+	+	+	+	+	+	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Nikaido et al., 2003
AfroSINE-Ad	155 nt monomer resulting from 3’-terminal deletion of AfroSINE-Anc	–	–	+	+	+	+	+	+	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Nikaido et al., 2003
ERI-1	~170-200 nt monomer consisting of 5’ tRNALys-like region, non-tRNA segment, polypyrimidine stretch, and A-rich 3’ end	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Borudulina and Kramerov, 2001
ERI-2	~200-250 nt monomer consisting of 5’ tRNA-like region, non-tRNA segment, polypyrimidine stretch, and A-rich 3’ end	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Borudulina and Kramerov, 2001
SOR	~190 nt monomer consisting of 5’ tRNA-like region, non-tRNA segment, and A-rich 3’ end	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Borudulina and Kramerov, 2001
TAL	~280 nt monomer consisting of 5’ tRNA-like region, non-tRNA segment, and A-rich 3’ end	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Borudulina and Kramerov, 2001
CHR-1	109 nt monomer consisting of 5’ tRNA-like region, non-tRNA segment, and A-rich 3’ end, derived from CHRS	–	–	–	–	–	–	–	–	–	–	+	+	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999
CHR-2-FL	321 nt monomer consisting of 5’ tRNA-like region, non-tRNA segment, and A-rich 3’ end, derived from CHR-1	–	–	–	–	–	–	–	–	–	–	+	+	+	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999
CHR-2-CD	CHR-2 with central deletion	–	–	–	–	–	–	–	–	–	–	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Nikaido et al., 2001
CHR-2-CDO	CHR-2-CD with 3’ deletion	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Nikaido et al., 2001
CHRS	~110-123 nt monomer consisting of 5’ tRNAGlu-like region, non-tRNA segment, and A-rich 3’ end	–	–	–	–	–	–	–	–	–	–	+	+	+	–	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999
CHR-S	Sequence variant of CHRS	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999
Bov-A	~115 nt monomer	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999
Bov-A2	~290 nt dimer: 2 Bov-A elements conncted by a middle segment	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999

Table 2 (continued)

SINE family	Element organization	*Mammalia*																														References
		Monotremata	*Marsupialia*	*Placentalia*																												
				Boreoeutheria																												
				Afrotheria										*Laurasiatheria*												*Euarchontoglires*						
				Hyracoidea	*Proboscoidea*	*Sirenia*	*Insectivora*			*Macroscelidea*	*Tubulidentata*	*Cetartiodactyla*				*Perissodactyla*	*Carnivora*	*Chiroptera*		*Primates*	*Rodentia*						*Lagomorpha*	*Dermoptera*		
							Erinaceidae	*Soricidae*	*Talpidae*			*Cetacea*		*Artiodactyla*				*Microchiroptera*										
												Mysticetes	*Odontocetes*	*Ruminantia*	*Suina*	*Hippopotamidae*					*Gliridae*	*Sciuridae*	*Castoridae*	*Hystricidae*	*Caviidae*				
														Tylopoda	*Suidae*															
														Camelidae																
Bov-tA	~217 nt 'dimer': fusion of two elements: CHR-2 (5'-half) and Bov-A (3'-half)	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999
vic-1	150 nt monomer consisting of 5' tRNAAla-like region, non-tRNA segment, and A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Lin et al., 2001
PRE-1	223 nt composite region and 3' A-rich end; CHR-S with tRNAARG insertion, plus an insertion and duplication of smaller sequences	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	Shimamura et al., 1999
ERE-1	~152 nt monomer consisting of 5' tRNAGlu-like region, non-tRNA segment, and A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	Sakagami et al., 1999
Can	265 nt monomer consisting of 5' tRNALys-like region, non-tRNA segment, and A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	Zehr et al., 2001
VES	~215 nt monomer consisting of 5' tRNA-like region, non-tRNA segment, and A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	Kawai et al., 2002 (Absent from the *Rhinolophidae*)
Alu	280 nt 'dimer' comprised of two nonidentical monomers derived from 7SL RNA, a middle A-rich region, and an A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	Schmid, 1996
B1	~140 nt monomer derived from 7SL RNA	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	–	–	–	–	Vassetzky et al., 2003
B1-dID	Fusion of B1-like unit with an ID-like monomer, connected by a short middle spacer	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	–	–	–	–	Kramerov et al., 1999
g/sB1-dID	*Glirid/sciurid*-specific variant of B1-dID	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	–	–	–	–	–	–	–	Kramerov et al., 1999
B2	~200 nt monomer consisting of 5' tRNALys-like region, non-tRNA segment, and A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	+	–	–	–	–	Mayorov et al., 2000
ID	~75 nt monomer derived from BC1 RNA, plus A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	+	+	+	+	–	–	–	–	Kass et al., 1996
C	~316 nt monomer consisting of 5' tRNAGly-like region, non-tRNA segment, and A-rich 3' end	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	Krane et al., 1991
t-SINE	Dimers and trimers of tRNAIle-like regions of variable length divisible into 3 subfamilies (a, b, and g)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	Piskurek et al., 2003

The genome system architecture concept further indicates that changes in retroelement profiles can alter genome transmission without affecting the somatic phenotype. Members of cryptic and sibling species complexes often have no detectable morphological, physiological, or adaptive differences and yet have distinct distributions of heterochromatin or chromosome structures that cause mating incompatibilities (see Shapiro and Sternberg, 2004). The data summarized in Table 1 show that retroelements are important components of the cellular apparatus for chromosome replication and distribution (e.g. origins, centromeres and telomeres). Accordingly, we predict that significant changes in the retroelements that format genome maintenance and transmission can lead to reproductive isolation, thereby setting the stage for subsequent clade-restricted changes in phenotype. In other words, we suggest that key adaptive events can *initiate* within the retroelement portion of the genome. There need be no correlation with mutations in the coding sector. Thus, another potentially fruitful area of investigation concerns differences in retroelement distributions between sibling species, particularly in centromeric and telomeric locations. We know that at least some centromeric repeats originate from retrotransposons (Cheng and Murata, 2003).

A more integrative view of the genome

In the era of "systems biology," it helps to recall that a system is more than a collection of components. Those components need to integrate functionally so they can accomplish systemic tasks requiring cooperative action. Retroelements and other DNA repeats provide the physical basis within the genome for functional integration. Dispersed regulatory sites of the kind provided by retrotransposons connect unlinked coding sequences into coordinately controlled subsystems. Similarly, replication and genome transmission processes are organized by elements carrying generic signals for origins, telomeres, centromeres and other structures essential to genome maintenance (Table 1). Signals delineating chromatin domains provide a higher level of organization for both transcription and replication, and distributed sites for attachment to cellular or nuclear structures provide a dynamic overall physical organization of the genome whose operation we are just beginning to comprehend.

As we increasingly apply computational metaphors to cellular function, we expect that a deeper understanding of retroelements and other repeats, the integrative fraction of cellular DNA, will lead to increased understanding of the logical architecture inherent to genome organization. In the era of biocomputing and systems biology, the study of cellular information processing promises to revolutionize not only the life sciences but also the information sciences. We anticipate learning powerful new computational paradigms as we come to understand how cells use myriad molecular components to regulate millions of biochemical events that occur every minute of every cell cycle. Our expectation is that, one day, we will think of what used to be called "junk DNA" as a critical component of truly "expert" cellular control regimes.

References

Allen E, Horvath S, Tong F, Spiteri E, Riggs AD, Marahrens Y: High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. Proc Natl Acad Sci USA 100:9940–9945 (2003).

Ananiev EV, Phillips RL, Rines HW: Complex structure of knob DNA on maize chromosome 9: retrotransposon invasion into heterochromatin. Genetics 149:2025–2037 (1998a).

Ananiev EV, Phillips RL, Rines HW: Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA 95:13073–13078 (1998b).

Aragon-Alcaide L, Miller T, Schwarzacher T, Reader S, Moore G: A cereal centromeric sequence. Chromosoma 105:261–268 (1996).

Arkhipova IR, Morrison HG: Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead. Proc Natl Acad Sci USA 98:14497–14502 (2001).

Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, Tomilin N: Association of some potential hormone response elements in human genes with Alu family repeats. Gene 239:341–349 (1999).

Bailey JA, Carrel L, Chakravarti A, Eichler EE: Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. Proc Natl Acad Sci USA 97:6634–6639 (2000).

Bailey JA, Liu G, Eichler EE: An Alu transposition model for the origin and expansion of human segmental duplications. Am J Hum Genet 73:823–834 (2003).

Becker KG, Swergold GD, Ozata K, Thayer RE: Binding of the ubiquitous nuclear transcription factor YY1 to a *cis* regulatory sequence in the human LINE-1 transposable element. Hum Mol Genet 2:1697–1702 (1993).

Bennett MD: Plant genome values: how much do we know? Proc Natl Acad Sci USA 95:2011–2016 (1998).

Bennetzen JL: Transposable elements contributions to plant gene and genome evolution. Plant Mol Biol 42:251–269 (2000).

Blackburn EH: The end of the (DNA) line. Nat Struct Biol 7:847–850 (2000).

Borodulina OR, Kramerov DA: Short interspersed elements (SINEs) from insectivores. Two classes of mammalian SINEs distinguished by A-rich tail structure. Mammal Genome 12:779–786 (2001).

Britten RJ: DNA sequence insertion and evolutionary variation in gene regulation. Proc Natl Acad Sci USA 93:9374–9377 (1996).

Britten RJ, Davidson EH: Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart Rev Biol 46:111–138 (1971).

Brosius J: RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 238:115–134 (1999).

Byrd K, Corces VG: Visualization of chromatin domains created by the gypsy insulator of *Drosophila*. J Cell Biol 162:565–574 (2003).

Cavalier-Smith T: The evolution of genome size (John Wiley & Sons Ltd., Chichester 1985).

Chen D, Kunlin J, Kawaguchi K, Nakayama M, Zhou X, Xiong Z, Zhou A, Mao XO, Greenberg DA, Graham SH, Simon RP: *Ero1-L*, an ischemia-inducible gene from rat brain with homology to global ischemia-induced gene 11 *(Giig11)*, is localized to neuronal dendrites by a dispersed identifier (ID) element-dependent mechanism. J Neurochem 85:670–679 (2003).

Chen S, Corces VG: The gypsy insulator of *Drosophila* affects chromatin structure in a directional manner. Genetics 159:1649–1658 (2001).

Cheng ZF, Dong T, Langdon S, Ouyang CR, Buell FR, Blattner F, Jiang J: Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691–1704 (2002).

Cheng Z-J, Murata M: A centromeric tandem repeat family originating from a part of Ty3/gypsy-retroelement in wheat and its relatives. Genetics 164:665–672 (2003).

Chimera JA, Musich PR: The association of the interspersed repetitive KpnI sequences with the nuclear matrix. J Biol Chem 260:9373–9379 (1985).

Chong S, Whitelaw E: Murine metastable alleles and transgenerational epigenetic inheritance. Cytogenet Genome Res 105:311–315 (2004).

Chureau C, Prissette M, Bourdet A, Barbe V, Cattolico L, Jones L, Eggen A, Avner P, Duret L: Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. Genome Res 12:894–908 (2002).

Cryderman DE, Cuaycong MH, Elgin SCR, Wallrath LL: Characterization of sequences associated with position-effect variegation at pericentric sites in *Drosophila* heterochromatin. Chromosoma 107: 277–285(1998).

Deininger PL, Moran JV, Batzer MA, Kazazian HH: Mobile elements and mammalian genome evolution. Curr Opin Genet Devel 13:651–658 (2003).

Dimitri P, Junakovic N: Revising the selfish DNA hypothesis: new evidence on accumulation of transposable elements in heterochromatin. Trends Genet 15:123–124 (1999).

Eichler EE: Recent duplication, domain accretion and the dynamic mutation of the human genome. Trends Genet 17:661–669 (2001).

Ferrigno O, Virolle T, Djabari Z, Ortonne JP, White RJ, Aberdam D: Transposable B2 SINE elements can provide mobile RNA polymerase II promoters. Nat Genet 28:77–81 (2001).

Figueiredo LM, Pirrit LA, Scherf A: Genomic organisation and chromatin structure of *Plasmodium falciparum* chromosome ends. Mol Biochem Parasitol 106:169–174 (2000).

Figeuiredo LM, Freitas-Junior LH, Bottius E, Olivo-Martin J-C, Schert A: A central role for *Plasmodium falciparum* subtelomeric regions in spatial positioning and telomere length regulation. EMBO J 21:815–824 (2002).

Gerasimova TI, Byrd K, Corces VG: A chromatin insulator determines the nuclear localizations of DNA. Mol Cell 6:1025–1035 (2000).

Gilbert N, Labuda D: CORE-SINEs: Eukaryotic short interspersed retroposing elements with common sequence motifs. Proc Natl Acad Sci USA 96: 2869–2874 (1999).

Han JS, Szak ST, Boeke JD: Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. Nature 429:268–274 (2004).

Hoskins RA, Smith CD, Carlson JW, Carvalho AB, Halpern A, Kaminker JS, Kennedy C, Mungall CJ, Sullivan BA, Sutton GG, Yasuhara JC, Wakimoto BT, Myers EW, Celniker SE, Rubin GM, Karpen GH: Heterochromatic sequences in a *Drosophila* whole-genome shotgun assembly. Genome Biol 3:0085.1–0085.16 (2002).

Jordan IK, Rogozin IB, Glazko GV, Koonin EV: Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet 19:68–72 (2003).

Jurka J, Zietkiewicz E, Labuda D: Ubiquitous mammalian-wide interspersed repeats (MIR) are molecular fossils from the Mesozoic era. Nucleic Acids Res 23:170–175 (1995).

Kass DH, Kim J, Deininger PL: Sporadic amplification of ID elements in rodents. J Mol Evol 42:7–14 (1996).

Kawai K, Nikaido M, Harada M, Matsumura S, Lim L-K, Wu Y, Hasegawa M, Okada N: Intra- and interfamily relationships of Vespertilionidae inferred by various molecular markers including SINE insertion data. J Mol Evol 55:284–301 (2002).

Kazazian HH: L1 retrotransposons shape the mammalian genome. Science 289:1152–1153 (2000).

Khodarev NN, Bennett T, Shearing N, Sokolova I, Koudelik J, Walter S, Villalobos M, Vaughn ATM: LINE L1 retrotransposable element is targeted during the initial stages of apoptotic DNA fragmentation. J Cell Biochem 79:486–495 (2000).

Kramerov D, Vassetzky N, Serdobova I: The evolutionary position of dormice (Gliridae) in Rodentia determined by a novel short retroposon. Mol Biol Evol 16:715–717 (1999).

Krane DE, Clark AG, Cheng J-F, Hardison R: Subfamily relationships and clustering of rabbit C repeats. Mol Biol Evol 8:1–30 (1991).

Labrador M, Corces VG: Setting the boundaries of chromatin domains and nuclear organization. Cell 111:151–154 (2002).

Lane N, Dean W, Erhardt S, Hajkova P, Surani A, Walter J, Reik W: Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis 35:88–93 (2003).

Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D, Olshansky M, Rechavi G, Jantsch MF: Systematic identification of abundant A-to-I editing sites in the human transcriptome. Nat Biotechnol 22:1001–1005 (2004).

Lin Z, Nomura O, Hayashi T, Wada Y, Yasue H: Characterization of a SINE species from vicuna and its distribution in animal species including the family Camelidae. Mamm Genome 12:305–308 (2001).

Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R: Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471–476 (2004).

Lyon MF: LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? Proc Natl Acad Sci USA 97:6248–6249 (2000).

Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, Bonnaud B, Lucotte G, Duret L, Mandrand B: The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. Proc Natl Acad Sci USA 101:1731–1736 (2004).

Mayorov VI, Rogozin IB, Elisaphenko EA, Adkison LR: B2 elements are present in the human genome. Mamm Genome 11:177–179 (2000).

Meyers BC, Tingey SV, Morgante M: Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res 11: 1660–1676 (2001).

Miller JT, Dong F, Jackson SA, Song J, Jiang J: Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics 150: 1615–1623 (1998).

Moran JV, DeBerardinis RJ, Kazazian HH: Exon shuffling by L1 retrotransposition. Science 283:1530–1534 (1999).

Mozer BA, Benzer S: Ingrowth by photoreceptor axons induces transcription of a retrotransposon in the developing *Drosophila* brain. Development 120: 1049–1058 (1994).

Nabirochkin S, Ossokina M, Heidmann T: A nuclear matrix/scaffold attachment region co-localizes with the gypsy retrotransposon insulator sequence. J Biol Chem 273:2473–2479 (1998).

Nagaki K, Song J, Stupar RM, Parokonny AS, Yuan Q, Ouyang S, Liu J, Hsiao J, Jones KM, Dawe RK, Buell CR, Jiang J: Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics 163:759–770 (2003).

Nekrutenko A, Li W-H: Transposable elements are found in a large number of human protein coding regions. Trends Genet 17:619–625 (2001).

Nigumann P, Redik K, Matlik K, Speek M: Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics 79: 628–634 (2002).

Nikaido M, Matsuno F, Abe H, Shimamura M, Hamilton H, Matsubayashi H, Okada N: Evolution of CHR-2 SINEs in cetartiodactyl genomes: possible evidence for the monophyletic origin of toothed whales. Mamm Genome 12:909–915 (2001).

Nikaido M, Nishihara H, Hukumoto Y, Okada N: Ancient SINEs from African endemic mammals. Mol Biol Evol 20:522–527 (2003).

Norris J, Fan D, Aleman C, Marks JR, Futreal PA, Wiseman RW, Inglehart JD, Deininger PL, McDonnell DP: Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. J Biol Chem 270:22777–22782 (1995).

Pardue ML, DeBaryshe PG: Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. Ann Rev Genet 37: 485–511 (2003).

Parish DA, Vise P, Wichman HA, Bull JJ, Baker RJ: Distribution of LINEs and other repetitive elements in the karyotype of the bat *Carollia:* implications for X-chromosome inactivation. Cytogenet Genome Res 96:191–197 (2002).

Pelissier T, Tutois S, Tourmente S, Deragon JM, Picard G: DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in Athila retroelement sequences. Genetica 97:141–151 (1996).

Petrov DA: Evolution of genome size: new approaches to an old problem. Trends Genet 17:23–28 (2001).

Pimpinelli S, Berloco M, Fanti L, Dimitri P, Bonaccorsi S, Marchetti E, Caizzi R, Caggese C, Gatti M: Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. Proc Natl Acad Sci USA 92:3804–3808 (1995).

Piskurek O, Nikaido M, Boeadi, Baba M, Okada N: Unique mammalian tRNA-derived repetitive elements in dermopterans: the t-SINE family and its retrotransposition through multiple sources. Mol Biol Evol 20:1659–1668 (2003).

Rollini P, Namciu SJ, Marsden MD, Fournier REK: Identification and characterization of nuclear matrix-attachment regions in the human serpin gene cluster at 14q32.1. Nucleic Acids Res 27:3779–3791 (1999).

Rubin CM, Kimura RH, Schmid CW: Selective stimulation of translational expression by Alu RNA. Nucleic Acids Res 30:3253–3261 (2002).

Sakagami M, Hiromura K, Chemnick LG, Ryder OA: Distribution of the ERE-1 family in Perissodactyla. Mamm Genome 10:930–933 (1999).

Schmid C: Alu: structure, origin, evolution, significance, and function of one-tenth of human DNA. Prog Nucl Acids Res Mol Biol 53:283–319 (1996).

Scholes DT, Kenny AE, Gamache ER, Mou Z, Curcio MJ: Activation of an LTR-retrotransposon by telomere erosion. Proc Natl Acad Sci USA 100:15736–15741 (2003).

Shapiro JA: Genome system architecture and natural genetic engineering in evolution. Ann NY Acad Sci 870:23–35 (1999).

Shapiro JA: Genome organization and reorganization in evolution: formatting for computation and function. Ann NY Acad Sci 981:111–134 (2002).

Shapiro JA, Sternberg RV: Why repetitive DNA is essential for genome function. Biol Rev (in press 2004).

Shimamura M, Abe H, Nikaido M, Ohshima K, Okada N: Genealogy of families of SINEs in cetaceans and artiodactyls: the presence of a huge superfamily of tRNA(Glu)-derived families of SINEs. Mol Biol Evol 16:1046–1060 (1999).

Silva JC, Shabalina SA, Harris DG, Spouge JL, Kondrashovi AS: Conserved fragments of transposable elements in intergenic regions: evidence for widespread recruitment of MIR- and L2-derived sequences within the mouse and human genomes. Genet Res 82:1–18 (2003).

Skyrabin BV, Kremerskothen J, Vassilacopoulou D, Disotell TR, Kapatinov VV, Jorka J, Brosius J: The BC200 RNA gene and its neural expression are conserved in Anthropoidea (primates). J Mol Evol 47:677–685 (1998).

Song X, Sui A, Garen A: Binding of mouse VL30 retrotransposon RNA to PSF protein induces genes repressed by PSF: Effects on steroidogenesis and oncogenesis. Proc Natl Acad Sci USA 101:621–626 (2004).

Speek M: Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. Mol Cell Biol 21:1973–1985 (2001).

Sternberg RV: The role of constrained self-organization in genome structural evolution. Acta Biotheor 44:95–118 (1996).

Sternberg RV: Genomes and form. The case for teleomorphic recursivity. Ann NY Acad Sci 901:224–236 (2000).

Sternberg RV: On the roles of repetitive DNA elements in the context of a unified genomic-epigenetic system. Ann NY Acad Sci 981:154–188 (2002).

Tchénio T, Casella J-F, Heidmann T: Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res 28:411–415 (2000).

van Driel R, Fransz PF, Verschure PJ: The eukaryotic genome: a system regulated at different hierarchical levels. J Cell Sci 116:4067–4075 (2003).

Vansant G, Reynolds WF: The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element. Proc Natl Acad Sci USA 92:8229–8233 (1995).

Vassetzky NS, Kramerov DA: CAN – a pan-carnivore SINE family. Mamm Genome 13:50–57 (2002).

Vassetzky NS, Ten OA, Kramerov DA: B1 and related SINEs in mammalian genomes. Gene 319:149–160 (2003).

Watson JB, Sutcliffe JG: Primate brain-specific cytoplasmic transcript of the Alu repeat family. Mol Cell Biol 7:3324–3327 (1987).

Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM: Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. Science 14:2357–2360 (2001).

Yang Z, Bofelli D, Boonmark N, Schwartz K, Lawn R: Apolipoprotein(a) gene enhancer resides within a LINE element. J Biol Chem 273:891–897 (1998).

Yates PA, Burman RW, Mummaneni P, Krussel S, Turker MS: Tandem B1 elements located in a mouse methylation center provide a target for de novo DNA methylation. J Biol Chem 274:36357–36361 (1999).

Zaiss DMW, Kloetzel P-M: A second gene encoding the mouse proteosome activator b subunit is part of a LINE1 element and is driven by a LINE1 promoter. J Mol Biol 287:829–835 (1999).

Zehr SM, Nedbal MA, Flynn JJ: Tempo and mode of evolution in an orthologous Can SINE. Mamm Genome 12:38–44 (2001).

Zhang X, Wessler SR: Genome-wide comparative analysis of the transposable elements in the related species *Arabidopsis thaliana* and *Brassica oleracea*. Proc Natl Acad Sci USA 101:5589–5594 (2004).

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK: Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836 (2002).

Zhou Y-H, Zheng JB, Gu X, Li W-H, Saunders GF: A novel Pax-6 binding site in rodent B1 repetitive elements: coevolution between developmental regulation and repeated elements? Gene 245:319–328 (2000).

Zhou Y-H, Zheng JB, Gu X, Saunders GF, Yung W-KA: Novel Pax6 binding sites in the human genome and the role of repetitive elements in the evolution of gene regulation. Genome Res 12:1716–1722 (2002).

Zhu L, Swergold GD, Seldin MF: Examination of sequence homology between human chromosome 20 and the mouse genome: intense conservation of many genomic elements. Hum Genet 113:60–70 (2003).

Zuckerkandl E: Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. Genetica 115:105–129 (2002).