Repeat performance: how do genome packaging and regulation depend on simple sequence repeats?

Ram Parikshan Kumar,[1] *Ramamoorthy Senthilkumar,*[1,2] *Vipin Singh,*[1] *and Rakesh K Mishra*[1]*

[1]Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India
[2]Centre for Computational Natural Sciences and Bioinformatics, International Institute of Information Technology, Gachibowli, Hyderabad, 500 032, India

Non-coding DNA has consistently increased during evolution of higher eukaryotes. Since the number of genes has remained relatively static during the evolution of complex organisms, it is believed that increased degree of sophisticated regulation of genes has contributed to the increased complexity. A higher proportion of non-coding DNA, including repeats, is likely to provide more complex regulatory potential. Here, we propose that repeats play a regulatory role by contributing to the packaging of the genome during cellular differentiation. Repeats, and in particular the simple sequence repeats, are proposed to serve as landmarks that can target regulatory mechanisms to a large number of genomic sites with the help of very few factors and regulate the linked loci in a coordinated manner. Repeats may, therefore, function as common target sites for regulatory mechanisms involved in the packaging and dynamic compartmentalization of the chromatin into active and inactive regions during cellular differentiation.

Keywords: coordinated expression of genes; genome packaging; long-range interactions; repeats; simple sequence repeats; transposon

Introduction

Comparison of sequences of eukaryotic genomes has revealed that the number of protein coding genes does not increase appreciably with the increase in the complexity of organisms. A simple organism such as *Caenorhabditis elegans* with only about 1,000 somatic cells contains more than 19,000 protein coding genes that account for ~25% of its genome (100 Mb), while a far more complex organism like the human has a similar number of genes, ~20,000, but they account for only ~2% of the genome (3,000 Mb).[1–3] This indicates that biological complexity has evolved not by addition of more genes to the genome, but by more sophisticated regulation of preexisting genes. Alternate splicing and post-translational modifications are additional mechanisms that amplify the proteome complexity of higher eukaryotes. However, transcriptional regulation remains the major mechanism to execute developmental programs. Much of the transcriptional regulation is achieved by a variety of *cis*-regulatory elements present in the non-coding part of the genome. It is now generally accepted that the selective advantage conferred by the accumulation of non-coding DNA is a factor in the increase in genome size in complex organisms[3] (Fig. 1).

Eukaryotes, compared to simpler organisms, have to contend with two more levels of regulation: (i) they lack the simple operon system for coordinated expression of metabolically related proteins; genes comprising a metabolic or developmental pathway are typically dispersed on different chromosomes, and (ii) the nucleus is organized into distinct ''active'' and ''inactive'' compartments. Thus eukaryotes need

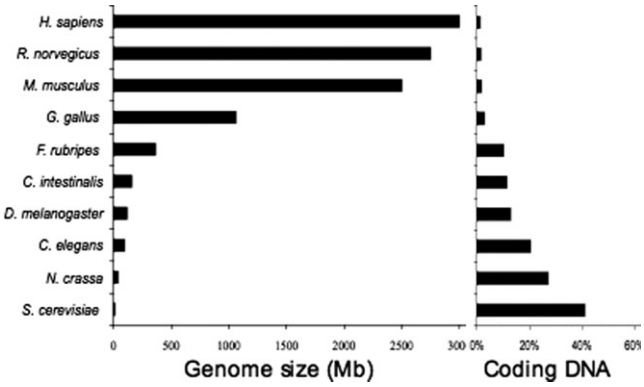

Figure 1. Eukaryotic genome size and proportion of non-coding DNA from lower to higher eukaryotes. Minimum genome size consistently increases from simple eukaryotes like yeast to complex vertebrates like human. The proportion of coding DNA is very small in the bigger genomes, indicating that much of the genome acquired by these organisms is non-coding.

***Correspondence to:** R. K Mishra, Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad 500 007, India.
E-mail: mishra@ccmb.res.in

mechanisms to identify coordinately regulated genes and move them to appropriate regions to regulate gene expression in a cell type-specific manner. It is likely that these functions are also carried out by the non-coding regions in the genome. Repetitive elements represent a major component of the non-coding sequences in higher eukaryotes; *e.g.*, repeat elements comprise ~60% of the non-coding fraction of the human genome.[4] Since repeats constitute the major component of complex genomes, their functional relevance has been a matter of speculation for several decades.[5,6] Several observations indicate that repeat elements are under selection pressure and, therefore, have functional significance, *e.g.*, nature of accumulated repetitive elements, their non-random distribution pattern and association with different nuclear compartments.[7–12] The biological utility of repetitive elements is further indicated by the presence of several repeat-specific DNA-binding proteins[13–15] and the observation that many repeats are transcribed.[9,16–20]

Recent investigations have provided evidence for a role of repeat elements in a variety of functions, such as higher-order organization of chromatin structure,[8,19–24] regulation of gene expression at transcriptional[7,10,12,25–35] as well as post-transcriptional levels,[17,36–40] cell division,[41,42] and human aging and disease.[43–46] While these investigations implicate repeat elements in a variety of nuclear processes, they do not provide evidence for a direct role for repeats in such processes. Here we hypothesize that certain repeat elements, particularly the simple sequence repeats (SSRs), participate in the packaging of the genome and in the mediation of long-range interactions that can lead to clustering of coordinately regulated loci. Such repeats may, therefore, be regarded as the "punctuation marks" to facilitate expression of the language of the genome sequence consisting of four letters.

Repetitive elements

A typical eukaryotic genome harbors a rich variety of repeat elements that differ in their length, sequence, distribution in the genome, and their implicated function (summarized in Table 1). Despite these differences, repeat elements can be categorized into two broad groups: (i) interspersed repeats that include the DNA/RNA transposons, and (ii) tandem repeats that include satellite DNA and SSR. In addition to the differences in the repeat pattern and sequence, these elements also vary in their origins, mechanism of expansion/maintenance, abundance and length of the repeat unit in a given genome.[47,48] As discussed below, several recent studies suggest that these elements may have important functions in genome organization and expression of associated genes (Table 1).

Transposon-derived sequences

The most abundant repeat sequences in eukaryotes are the transposon-derived sequences (henceforth referred to as transposons), comprising ~5% of the *Drosophila* genome and ~50% of the human genome.[48] Transposons can be divided into two broad classes, *i.e.*, the DNA-transposon and RNA- or retro-transposons. Retro-transposons are of several kinds, *i.e.*, long interspersed elements (LINEs), short interspersed elements (SINEs), and long terminal repeat (LTR) retro-transposons (Table 1). SINEs are abundant in animal genomes and the number of SINE families differs significantly among species. The human genome has Alu as the single predominant SINE, while the mouse genome has two predominant SINEs known as B1 and B2. It has been shown that SINE elements are transcribed both in human (Alu) and mouse (B2) and that the ensuing non-coding RNAs (ncRNAs) interact with RNA pol-II at the promoter of heat-shock non-responsive genes and regulate mRNA transcription in *trans*.[17,18,36,37] Even though Alu and B2 SINEs share similar functions, they are not related in sequence. Several regulatory functions have been proposed for these transposons, ranging from mammalian-specific brain patterning[27] and siRNA,[49] to coordinated and chromatin-mediated transcriptional regulation.[12] Intriguingly, Alu sequences in humans are a major site for RNA editing and their abundance correlates with a massive increase in the extent of RNA editing relative to other mammals.[50,51] Another transposon, LINE, has been shown to be involved in chromosome organization,[52,53] imprinting,[54] and regulation of gene expression.[55–58] Like SINEs and LINEs, LTRs have also been implicated in nuclear processes, *e.g.*, heterochromatin formation at centromere and telomere, maintenance of nucleolus organizing region (NOR),[59,60] and chromatin domain boundary functions that also mediate long-range interaction.[10,33]

DNA-transposons are widely distributed but less abundant compared to the retro-transposons (Table 1). These elements have often been found to be associated with new alleles primarily due to disruption of regulatory regions.[61–63] Different interconnected mechanisms have been implicated in the silencing mediated by these elements,[64–66] and they are thought to act as moving targets for local heterochromatin formation due to their repetitive nature.[20]

Satellite DNA

Satellite sequences are frequent and commonly organized as large clusters on chromosomes in animals and plants. Several studies indicate that satellite repeats are important and can directly influence the chromosome organization, transcriptional as well as post-transcriptional aspects of gene regulation (Table 1). For example, satellite repeats have been shown

Table 1. Repeat elements of eukaryotic genome

Class[73]	Order	Repeat length	[a]Abundance (%) copy number[4]	Occurrence[73]	Structural features,[73] special features, and functions
Retro-transposons	LTR	10–25 kb	Moderate (8) 450 000	P, M, F, O	Encodes LTRs, *gag*, *pol*, and *env* genes, TSDs Present throughout genome Common in plants, only transposon found in *S. cerevisiae* Regulation of gene expression[29,74–79]
	LINE	L1 6–7 kb (mammals)	High (21) 850 000	P, M, F, O	Contain TSDs, reverse transcriptase/endonuclease domains, pol-II promoter Present in heterochromatin. R2, RTE and Jockey superfamilies found only in metazoans X inactivation[80] Telomere maintenance[53,75,81] Alteration of cell fate[82]
	SINE	80–500 bp	High (13) 1 500 000	P, M, F	Contain TSDs, derived from 7SLRNA, tRNA/5SRNA, pol III promoter
		Alu-300 bp			Present in euchromatin, 5S RNA derived SINES not reported in plants Nucleosome positioning[83] Regulation of gene expression[7,12,17,18,27,36–38,49,84,85]
	SVA	Variable 1.5–2.8 kb	Very low 3000	M	Hominid specific, composite, non-autonomous retrotransposons[86] Present in euchromatin Implicated in gene duplication[87]
	DIR	DIRS 1–4.7 kb	–	P, M, F, O	Unusual termini-either split direct repeats or inverted repeats
	PLE	Variable	–	P, M, F, O	Contain TSDs, LTRs, and encodes reverse transcriptase/endonuclease domains Present in euchromatin
DNA-transposons	TIR	1–10 kb	Low (3) 300 000	P, M, F, O	Contain TSDs, TIR, encodes DDE transposase Present throughout genome P elements in *Drosophila* Mutator in maize and Tc3 in nematodes prefer insertion in proximity of genes Evolution of new genes[88]
	CRY	4 kb[88]	–	F	Contain TSDs, no TIR, encode tyrosine recombinase
	HEL	5.5–17 kb[89]	–	P, M, F	Contain no TSDs, ends-TC/CTRR (R = purine), encode Y2 tyrosine recombinase, 0–3% in mammals, >2% in *C. elegans* and *Arabidopsis thaliana* Actively transposing in *Myotis lucifugus*[90]
	MAV	15–25 kb[88]	–	M, F, O	Contain long TIRs, TSDs, encodes 11 proteins (DNA pol B and integrase)
Satellite		α-170 bp β-68 bp γ-220 bp	–	P, M, O	Present in heterochromatin, Transcribed from both the strands[16,67–69] Heterochromatin formation and chromosome segregation[19,20,67,71,72]
		Sat1-25–48 bp Sat2 and 3–5 bp			Nucleolar integrity[21] Regulation of gene expression[40,91,92]
SSR		1–7 bp	– (3)	P, M, F, O	Present non-randomly Regulation of gene expression[11,93–99]

[a]For human genome.

CRY, crypton; DIRS, *Dictyostellium* interspersed repeats; HEL, helitron; MAV, maverick; LINEs, long interspersed nuclear elements; SINEs, small interspersed nuclear elements; LTRs, long terminal repeat; P, plants; M, metazoans; F, fungi; O, others; PLE, penelope like elements; SVA, SINE VNTR and Alu (VNTR, variable number tandem repeat); TIR, terminal inverted repeats; TSDs, target site duplications.

to be critically important for centromere heterochromatin formation and chromosome segregation.[20,22] Satellite repeats are transcribed from both the strands in plants and animals,[16,67–69] and are important for the establishment and maintenance of heterochromatin.[21–24,70] Additionally, long ncRNA encompassing a few satellite monomers are components of the kinetochore.[67] Some satellites, like satellite III, are transcribed upon stress and can influence splicing.[40] Similar to the centromere satellite repeats, the telomere satellite repeats are also required for maintaining proper chromosome length and constitutive heterochromatin.[19,71,72]

SSRs

SSRs or microsatellites, very common in the complex genomes, are short sequences of nucleotides (1–6 bp in length) that are repeated in tandem. These are among the most variable types of DNA sequence (Table 1). The genomic distribution of SSRs is non-random, and these repeats have accumulated in the genome in parallel with the evolution of complexity in plants and animals, whereas in prokaryotes their abundance is relatively low.[47] In the case of the human genome, iteration of SSR motifs are found at hundreds of thousands of places along the entire length of chromosomes.[4] SSRs have been shown to be clustered around pericentromere, sub-telomeric, and some euchromatic regions of chromosomes.[100–103] Recent studies indicate a role for SSRs in a variety of nuclear functions.[45,46,94,95,104,105]

Functional relevance of repetitive DNA – a journey from junk to jewel

The function of the ubiquitous repeat sequences has been a matter of much speculation. The unusual abundance of repetitive elements has attracted the view that this part of genome may have no function and was termed selfish or junk DNA.[106–108] Views have varied from one extreme of ''junk DNA'' to other extreme of ''every nucleotide is at least infinitesimally functional.''[109] The idea that repeats may have function has been proposed many times, starting from the time of McClintock.[6] For several decades it has been speculated that repeats may be implicated in protection from mutations, fuel for evolutionary mechanisms and regulation of genes.[5,109–114] Most of these ideas have, however, been restricted primarily, although not exclusively, to transposable elements and, although they provided an exciting prospect for interesting repeat DNA biology, very little experimental evidence existed until recently in their support. Much less has been said about the functional significance of SSRs, which contributing ~3% of the human genome far exceed the protein-coding component.

Here we discuss the potential role of SSRs in packaging the genome and mediating long-range chromosomal interactions to cluster together a number of loci for coordinated expression. We revisit recent literature to support this view with a variety of experimental evidence.

SSR-based genome packaging code

Abundance and distribution

The first clue about functional significance of SSRs comes from their distribution pattern in different genomes. While in some organisms SSRs are clustered in specific regions of the chromosomes, *e.g.*, pericentromeric or sub-telomeric regions,[100–103] in complex genomes these repeats are generally dispersed throughout the genome in a non-random manner, highly populating the intergenic region.[47] In addition to the non-random distribution pattern, there are several other observations that point to a selection pressure favoring acquisition and maintenance of these repeats. For example, only few of all the possible SSRs have accumulated in complex genomes,[115] in some cases longer stretches of repeat are more abundant than shorter stretches,[100] several proteins that bind specifically to repeats are known and, finally, many repeats are transcribed (see below). These observations strongly suggest that the SSR component is unlikely to be junk DNA or a neutral or random evolutionary relic.

Repeats as punctuation marks of genome packaging code

We propose that SSRs (and perhaps other types of repeat elements) are components of a ''genome packing code'' to condense the genome in a cell type-specific manner. Being repetitive in nature, such elements can function as landmarks for similar/concerted chromatinization process to operate on select regions of genome. For example, a large number of loci linked to a repeat can be targeted by a repeat-binding protein or transcript that can function as a guide for targeted landing of specific functions. This immediately offers the possibility of coordinated expression of a large number of genes associated with repeats in different regions of the chromosome, and even on different chromosomes, with the help of a very small number of DNA-binding proteins. Such proteins can bind and recruit other partners to gather genome regions linked to these repeats to a few nuclear compartments, and lead to coordinated expression of the genes marked by such repeats (Fig. 2). Differential regulation of such factors or post-translational modifications can facilitate a cell type-specific switch that can operate at a large number of loci. Regulated expression of such targeting factors acting on specific repeat elements can greatly simplify the complex regulation of a

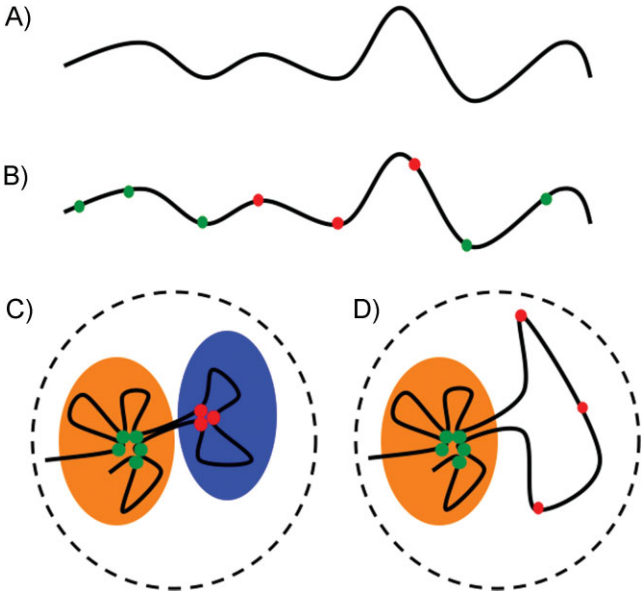

Figure 2. Repeat-dependent cell type-specific genomic packaging code. **A**: Genomic DNA. **B**: Genome marked with different repeat elements shown as different colored spheres. Repeat-specific DNA-binding proteins can allow chromatin loop formation by bringing the repeats together. **C**: Different kinds of repeat can interact with the help of different set of proteins and cluster the linked loci to different nuclear compartments. **D**: In a different cell type where a particular repeat-specific DNA-binding proteins is absent, the clustering can be altered, leading to a different conformation. Such altered packaging can have direct and coordinated regulatory consequences.

large number of genes in a cell type- and stage-specific manner. Repeat elements can function like punctuation marks that are used repeatedly in different genomic contexts to facilitate expression of genetic information.

Long-range interactions mediated by SSRs

One of the key features of this model is that repeat elements mediate long-range interactions for coordinated regulation of gene expression (Fig. 2). Several independent observations point to such mechanisms. One such example is that of *brown Dominant* (bw^D) mutation of *Drosophila*. This mutation is caused by the insertion of an ~2 Mb block of GAGA satellite DNA at the *brown* locus of chromosome 2. In heterozygous context, the satellite DNA drags not only the affected *brown* locus, but also its normal homolog into centromeric heterochromatin. This heterochomatinization of the wild-type *brown* locus manifests as variegated expression of the *brown* gene in which the "tagged" locus effectively causes *in trans* silencing of its normal counterpart on the homologous chromosome.[92,116] Another example of widespread long-range interactions is that of the odorant receptor (*OR*) genes. The *OR* gene family in mouse comprises ~1,300 members scattered on 12 different chromosomes. The expression of this gene family is controlled by a single odorant H-enhancer.[11] Thus, the expression of an *OR* gene is possible only when its promoter is brought to the vicinity of the enhancer as a consequence of inter- or intra-chromosomal interactions. While the molecular basis for these interactions is not known, it is interesting to note that several *OR* loci are associated with retrotransposons and SSRs, and that the H-enhancer is also embedded in SSRs.[117] However, a direct role for the repeat elements in the expression of *OR* gene remains to be established.

One possible candidate for an SSR element acting as a marker for genome compartmentalization is the GATA repeat element. We have earlier shown that the GATA tetranucleotide SSR has several unique features.[100] For example, longer stretches of the repeat, GATA10–12, are more abundant that shorter stretches, and the longer stretches are mostly intergenic and associated with sequences with a very high matrix association potential. We also observed that longer GATA repeats on human Y chromosome are associated with genes that are expressed early during development, indicating that these repeats might be marking loci for coordinated expression by long-range chromosomal interactions that clusters the GATA repeat-linked loci.[100] These studies support the idea that SSRs may function as genome packaging tools and coordinate expression of genes by facilitating long-range interactions.

To assess the patterns of GATA repeats in other genomes, we compared the length and abundance of this repeat in 15 eukaryotic genomes. Longer repeats, GATA10–12, are more favored in most of the vertebrates, but not the invertebrates; an exception is the Zebrafish in which even longer repeats are more abundant (Fig. 3). It will be interesting to map the context of these longer repeats and investigate their function. Extensive SSR analysis of genomes will reveal if elements other than GATA have a similar unusual abundance of longer repeats. Different kinds of repeats in different genomic context can have a differential functional consequence for coordinated regulation of gene expression (Fig. 4).

One of the best characterized examples of spatial organization of chromatin comes from repeat elements *Saccharomyces cerevisiae*. Telomeric repeats that mark the end of the chromosomes have been implicated not only in telomere length maintenance, but also in nuclear organization and regulation of genes located in the proximity of telomeres and long range functional interactions.[118,119] Telomeres in *S. cerevisiae* are known to be clustered in several distinct foci that lie near the nuclear membrane.[120] The observations with yeast telomeric repeats suggest that repeat element-mediated higher order organization of chromatin emerged early in the evolution of eukaryotic genomes, at a stage where the repeat component of the genome was minimal or moderate, and the putative

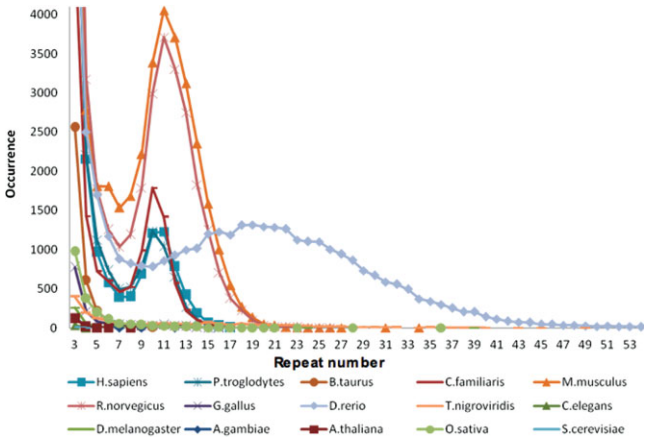

Figure 3. Abundance of different size GATA repeats in the genome of different species. Longer GATA repeats (10–12 times in tandem) are enriched in several vertebrates including human. Zebrafish has the distinction of possessing a higher abundance of longer GATA repeats (peaking in around 20 repeats in tandem). This trend of higher abundance of longer repeats is not seen in several other eukaryotes including, yeast, insects, worms, plants, and some other higher eukaryotes.

evolutionary advantage offered by such systems was the driving force for the accumulation of repeat elements in the genomes of more complex eukaryotes.

Apart from the SSR, there is increasing evidence pointing to the functionality associated with the evolution and enrichment of other class of repetitive elements. Transposable elements like *Idefix* and *Gypsy* retro-transposons in *Drosophila* have been shown to be involved in higher order organization of genomic loci in nuclear compartments and coordinate gene regulation.[77,78] Several studies indicate that SINE elements may promote transcription by looping to specific promoters to activate those genes.[7,27,121,122] In humans, tandem repeat of the DXZ4 (3 kb) array is organized as heterochromatin on active X and euchromatin on inactive X-chromosome.[9] These repeats are critical for developmental control of X-inactivation, and the importance of this stretch of repeats is indicated by the finding that contraction in number of the array is associated with facioscapulohumeral muscular dystrophy.[123] These studies thus provide a functional link between repeats and the regulation of associated loci.

Repeats as chromatin domain boundary elements

Eukaryotic chromosomes harbor both active and inactive genes along their length. The active regions must be insulated from neighboring repressed or inactive regions by domain boundaries. Such chromatin elements that divide genomes into functional domains are expected to occur several thousand times in a complex genome.[124] Boundary elements are, therefore, likely to be associated with, or even overlap with, at least some repeats. As indicated earlier,

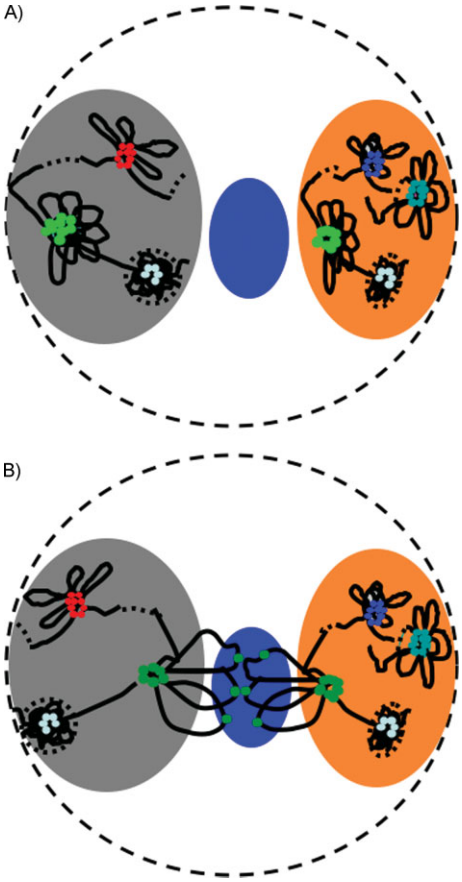

Figure 4. Long-range interactions mediated by repeats. Repeats can organize chromatin into specific compartments. **A**: Similar kinds of repeat-associated loci are shown in same color. **B**: Associated loci, due to differential expression of corresponding repeat-binding protein (*e.g.*, repeat-interacting protein, green) can allow movement of the associated loci to a common compartment. Such chromatin movement can have regulatory consequences for the associated genes.

elements like SINE B2 in mouse, *Gypsy* and *Idefix* in *Drosophila* have been linked to boundary activities.[7,10,25] Our earlier studies indicated that SSRs like GATA repeats can have the potential to function as chromatin domain boundaries.[100] Marking of domains by such boundaries may be relevant to genome packaging with multiple functional consequence, like prevention of inappropriate spread of heterochromatin into euchromatin, restricting enhancers and silencers to their legitimate target promoters and defining functionally independent regions within genomes (Fig. 5).

Mechanisms of SSR-mediated global regulatory events

Repeat-binding protein-mediated events

The clustering-based networking requires protein factors that specifically bind to repeats. Repeat-binding proteins asso-

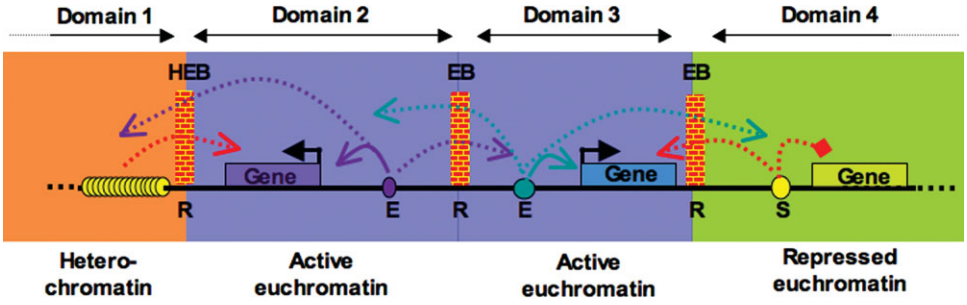

Figure 5. Repeats mediate chromatin domain formation. Different repeat elements are shown to punctuate the genome into different domains. The hetero-euchromatic boundary (HEB) can demarcate transition between heterochromatin (domain 1) and euchromatin (domains 2–4). Within euchromatic regions, euchromatic boundaries (EB) can sub-divide chromatin into functional distinct domains (active domains 2 and 3; inactive domain 4). These repeats can prevent the spread of repressive effect into the euchromatin from heterochromatin. Boundaries can also allow legitimate interaction between the regulatory elements, activation in domain 2–3 (solid arrow) and repression in domain 4 (dotted arrow), and disallow inappropriate interactions (dotted arrow). R, repeat; S, silencer; E, enhancer.

ciated with their distant target sites in the genome can bring such target sites into physical proximity by protein-protein interactions. Interestingly, several repeat-binding proteins have been isolated.[125,126] For example, GAGA repeats, which are abundantly present in the eukaryotic genome, have been shown to be bound by GAGA-associated factor (GAF) in *Drosophila melanogaster*, BBR in barley, GBP in soybean and rice, and angiogenin in mammal.[125,127–129] Such repeat-binding proteins are likely to control large numbers of loci associated with these repeats, although much remains to be done to fully understand their function. Any sequence element that marks a large number of loci is automatically identified as repetitive in nature. As described above, one way to use them as switches of regulatory mechanisms is through the sequence-specific DNA-binding proteins that recognize them. Differential regulation of a few such proteins can influence expression of a large number of loci linked to their target repeats. Clustering of multiple loci involving the cohesin complex that cooperates with the boundary function of CTCF has recently been shown, and involves bundling of distant chromatin fibers in a ring like structure.[130] Any "repeat connection" to such a mechanism, however, has yet to be demonstrated.

RNA-mediated events

An alternative to protein-mediated clustering of coordinately expressed loci and/or regulation of gene expression is the use of RNA transcribed from repeat elements. Even if only a few repeat loci produce transcripts, these ncRNA molecules can diffuse throughout the nucleus and potentially target all the loci linked to the repeat and act as "guide molecules" for their clustering or compartmentalization.[131,132] This suggests that regulation of one or few repeat sites can serve as source loci by which the diffusible repeat transcript can target the other loci. Such repeat RNAs can act as scaffold for recruitment and maintenance of stable protein complexes at target loci. This mechanism will be similar to the one in which roX1 and roX2 ncRNA molecules in *D. melanogaster* serve as "nucleation sites" at the transcription initiation site as well as ~150 other high-affinity "entry sites" on the X chromosome.[133,134] These ncRNA molecules interact with the dosage compensation complex and target them to nucleation/entry sites on the X chromosome, leading to sustained hyper activation of the X chromosome in male *Drosophila*. In this context, it is interesting to note that a large number of repeats, including satellite repeats, are transcribed. It remains to be seen if only a subset of these loci is actually transcribed and what the function is of these RNA molecules.

Concluding remarks

A lot remains to be understood about transcriptional regulation in eukaryotes that involves intra- and inter-chromosomal long-range interactions. New techniques that address the 3D analysis of the nuclear architecture and chromatin interactions have infused fresh excitement into the field of the spatial organization of chromatin networks, long-range interactions and functional consequences of such interactions. Repetitive elements are strong contenders for a key role in these mechanisms by marking sites of interaction of the linked loci, leading to their clustering/compartmentalization and coordinated regulation of gene expression as the functional consequence of this organization. Association or overlap of repeats with a variety of other regulatory elements can significantly expand the repertoire and efficiency of such a mechanism. One indication of a direct role of repeat elements in genomic packaging and coordinated expression is emerging from the non-random distribution of such elements and their association with sets of genes that are functionally related.[135–140] Identification and analysis of repeats in context of the regulatory feature of the genes linked to such elements and characterization of proteins that can

interact with repeats and the transcripts emerging from repetitive DNA will allow us to gain a better understanding of genome organization and regulation. The potential of repeat elements to function as switches operated by a very small number of interacting factors (RNA/proteins) as a part of genomic packaging code and gene regulation processes provides a new perspective on evolutionary mechanisms that have shaped complex eukaryotic genomes.

Acknowledgments: We thank R. K. M. laboratory members for advice and assistance. Research in the laboratory of R. K. M. is supported by the CSIR, India. R. P. K., R. S., and V. S. acknowledge research fellowship from CSIR, India. R. K. M. acknowledges research grants from CEFIPRA, Department of Science and Technology, India and Department of Biotechnology, India. We would like to thank D. P. Kasbekar and Vidya Jonnalagadda, Krishnaveni Mishra, and K. Satish for critical reading of the manuscript and suggestions.

References

1. **Clamp M, Fry B, Kamal M,** *et al.* 2007. Distinguishing protein-coding and noncoding genes in the human genome. *Proc Natl Acad Sci USA* **104**: 19428–33.
2. **Goodstadt L, Ponting CP.** 2006. Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *PLoS Comput Biol* **2**: e133.
3. **Taft RJ, Pheasant M, Mattick JS.** 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**: 288–99.
4. **Lander ES, Linton LM, Birren B,** *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
5. **Britten RJ, Davidson EH.** 1969. Gene regulation for higher cells: a theory. *Science* **165**: 349–57.
6. **McClintock B.** 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA* **36**: 344–55.
7. **Lunyak VV, Prefontaine GG, Nunez E,** *et al.* 2007. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317**: 248–51.
8. **Thakar R, Gordon G, Csink AK.** 2006. Dynamics and anchoring of heterochromatic loci during development. *J Cell Sci* **119**: 4165–75.
9. **Chadwick BP.** 2008. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res* **18**: 1259–69.
10. **Brasset E, Bantignies F, Court F,** *et al.* 2007. Idefix insulator activity can be modulated by nearby regulatory elements. *Nucleic Acids Res* **35**: 2661–70.
11. **Lomvardas S, Barnea G, Pisapia DJ,** *et al.* 2006. Interchromosomal interactions and olfactory receptor choice. *Cell* **126**: 403–13.
12. **Apostolou E, Thanos D.** 2008. Virus infection induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell* **134**: 85–96.
13. **Liu X, Wu B, Szary J,** *et al.* 2007. Functional sequestration of transcription factor activity by repetitive DNA. *J Biol Chem* **282**: 20868–76.
14. **Hu HG, Scholten I, Gruss C,** *et al.* 2007. The distribution of the DEK protein in mammalian chromatin. *Biochem Biophys Res Commun* **358**: 1008–14.
15. **Vashakidze RP, Chelidze MG, Mamulashvili NA,** *et al.* 1988. Nuclear proteins from *Drosophila melanogaster* embryos which specifically bind to simple homopolymeric sequences poly [(dT-dG) (dC-dA)]. *Nucleic Acids Res* **16**: 4989–94.
16. **Lee HR, Neumann P, Macas J,** *et al.* 2006. Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. *Mol Biol Evol* **23**: 2505–20.
17. **Mariner PD, Walters RD, Espinoza CA,** *et al.* 2008. Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock. *Mol Cell* **29**: 499–509.
18. **Espinoza CA, Goodrich JA, Kugel JF.** 2007. Characterization of the structure, function, and mechanism of B2 RNA, an ncRNA repressor of RNA polymerase II transcription. *RNA* **13**: 583–96.
19. **Schoeftner S, Blasco MA.** 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* **10**: 228–36.
20. **Grewal SI, Jia S.** 2007. Heterochromatin revisited. *Nat Rev Genet* **8**: 35–46.
21. **Peng JC, Karpen GH.** 2007. H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability. *Nat Cell Biol* **9**: 25–35.
22. **Morris CA, Moazed D.** 2007. Centromere assembly and propagation. *Cell* **128**: 647–50.
23. **White SA, Allshire RC.** 2008. RNAi-mediated chromatin silencing in fission yeast. *Curr Top Microbiol Immunol* **320**: 157–83.
24. **Riddle NC, Elgin SC.** 2008. A role for RNAi in heterochromatin formation in *Drosophila*. *Curr Top Microbiol Immunol* **320**: 185–209.
25. **Kim JH, Ebersole T, Kouprina N,** *et al.* 2009. Human gamma-satellite DNA maintains open chromatin structure and protects a transgene from epigenetic silencing. *Genome Res* **19**: 533–44.
26. **Gdula DA, Gerasimova TI, Corces VG.** 1996. Genetic and molecular analysis of the gypsy chromatin insulator of Drosophila. *Proc Natl Acad Sci USA* **93**: 9378–83.
27. **Sasaki T, Nishihara H, Hirakawa M,** *et al.* 2008. Possible involvement of SINEs in mammalian-specific brain formation. *Proc Natl Acad Sci USA* **105**: 4220–5.
28. **Cai HN, Levine M.** 1997. The gypsy insulator can function as a promoter-specific silencer in the *Drosophila* embryo. *EMBO J* **16**: 1732–41.
29. **Huh JW, Ha HS, Kim DS,** *et al.* 2008. Placenta-restricted expression of LTR-derived NOS3. *Placenta* **29**: 602–8.
30. **Bacher CP, Guggiari M, Brors B,** *et al.* 2006. Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nat Cell Biol* **8**: 293–9.
31. **Savitskaya E, Melnikova L, Kostuchenko M,** *et al.* 2006. Study of long-distance functional interactions between Su(Hw) insulators that can regulate enhancer-promoter communication in *Drosophila melanogaster*. *Mol Cell Biol* **26**: 754–61.
32. **Gerasimova TI, Byrd K, Corces VG.** 2000. A chromatin insulator determines the nuclear localization of DNA. *Mol Cell* **6**: 1025–35.
33. **Sigrist CJ, Pirrotta V.** 1997. Chromatin insulator elements block the silencing of a target gene by the *Drosophila* polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes. *Genetics* **147**: 209–21.
34. **Shirohzu H, Yokomine T, Sato C,** *et al.* 2004. A 210-kb segment of tandem repeats and retroelements located between imprinted subdomains of mouse distal chromosome 7. *DNA Res* **11**: 325–34.
35. **Gao Q, Reynolds GE, Innes L,** *et al.* 2007. Telomeric transgenes are silenced in adult mouse tissues and embryo fibroblasts but are expressed in embryonic stem cells. *Stem Cells* **25**: 3085–92.
36. **Espinoza CA, Allen TA, Hieb AR,** *et al.* 2004. B2 RNA binds directly to RNA polymerase II to repress transcript synthesis. *Nat Struct Mol Biol* **11**: 822–9.
37. **Allen TA, Von Kaenel S, Goodrich JA,** *et al.* 2004. The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat Struct Mol Biol* **11**: 816–21.
38. **Chen LL, DeCerbo JN, Carmichael GG.** 2008. Alu element-mediated gene silencing. *EMBO J* **27**: 1694–705.
39. **Berretta J, Pinskaya M, Morillon A.** 2008. A cryptic unstable transcript mediates transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev* **22**: 615–26.
40. **Metz A, Soret J, Vourc'h C,** *et al.* 2004. A key role for stress-induced satellite III transcripts in the relocalization of splicing factors into nuclear stress granules. *J Cell Sci* **117**: 4551–8.

41. **Raff JW, Kellum R, Alberts B.** 1994. The *Drosophila* GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. *EMBO J* **13**: 5977–83.
42. **Platero JS, Csink AK, Quintanilla A,** *et al.* 1998. Changes in chromosomal localization of heterochromatin-binding proteins during the cell cycle in *Drosophila*. *J Cell Biol* **140**: 1297–306.
43. **Ju Z, Rudolph KL.** 2006. Telomeres and telomerase in stem cells during aging and disease. *Genome Dyn* **1**: 84–103.
44. **Siwach P, Ganesh S.** 2008. Tandem repeats in human disorders: mechanisms and evolution. *Front Biosci* **13**: 4467–84.
45. **Kersting C, Agelopoulos K, Schmidt H,** *et al.* 2008. Biological importance of a polymorphic CA sequence within intron 1 of the epidermal growth factor receptor gene (EGFR) in high grade central osteosarcomas. *Genes, Chromosomes Cancer* **47**: 657–64.
46. **Kumari D, Usdin K.** 2009. Chromatin remodeling in the non-coding repeat expansion diseases. *J Biol Chem* **284**: 7413–7.
47. **Subramanian S, Madgula VM, George R,** *et al.* 2002. MRD: a microsatellite repeats database for prokaryotic and eukaryotic genomes. *Genome Biol* **3**: Preprint0011.
48. **Hua-Van A, Le Rouzic A, Maisonhaute C,** *et al.* 2005. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. *Cytogenet Genome Res* **110**: 426–40.
49. **Watanabe T, Takeda A, Tsukiyama T,** *et al.* 2006. Identification and characterization of two novel classes of small RNAs in the mouse germline: retrotransposon-derived siRNAs in oocytes and germline small RNAs in testes. *Genes Dev* **20**: 1732–43.
50. **Eisenberg E, Nemzer S, Kinar Y,** *et al.* 2005. Is abundant A-to-I RNA editing primate-specific? *Trends Genet* **21**: 77–81.
51. **Mattick JS, Mehler MF.** 2008. RNA editing, DNA recoding and the evolution of human cognition. *Trends Neurosci* **31**: 227–33.
52. **Mazzuchelli J, Martins C.** 2009. Genomic organization of repetitive DNAs in the cichlid fish *Astronotus ocellatus*. *Genetica* **136**: 461–9.
53. **Maxwell PH, Belote JM, Levis RW.** 2006. Identification of multiple transcription initiation, polyadenylation, and splice sites in the *Drosophila melanogaster* TART family of telomeric retrotransposons. *Nucleic Acids Res* **34**: 5498–507.
54. **Allen E, Horvath S, Tong F,** *et al.* 2003. High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc Natl Acad Sci USA* **100**: 9940–5.
55. **Hansen RS.** 2003. X inactivation-specific methylation of LINE-1 elements by DNMT3B: Implications for the Lyon repeat hypothesis. *Hum Mol Genet* **12**: 2559–67.
56. **Chen J, Rattner A, Nathans J.** 2006. Effects of L1 retrotransposon insertion on transcript processing, localization and accumulation: lessons from the retinal degeneration 7 mouse and implications for the genomic ecology of L1 elements. *Hum Mol Genet* **15**: 2146–56.
57. **Han JS, Boeke JD.** 2005. LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? *Bioessays* **27**: 775–84.
58. **Tomilin NV.** 2008. Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *Bioessays* **30**: 338–48.
59. **Heslop-Harrison JS, Brandes A, Taketa S,** *et al.* 1997. The chromosomal distributions of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* **100**: 197–204.
60. **Brandes A, Heslop-Harrison JS, Kamm A,** *et al.* 1997. Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol Biol* **33**: 11–21.
61. **Miura A, Yonebayashi S, Watanabe K,** *et al.* 2001. Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* **411**: 212–4.
62. **Singer T, Yordan C, Martienssen RA.** 2001. Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). *Genes Dev* **15**: 591–602.
63. **Bourque G, Leong B, Vega VB,** *et al.* 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**: 1752–62.
64. **Lippman Z, Gendrel AV, Black M,** *et al.* 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**: 471–6.
65. **Slotkin RK, Martienssen R.** 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272–85.
66. **Sijen T, Plasterk RH.** 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426**: 310–4.
67. **Wong LH, Brettingham-Moore KH, Chan L,** *et al.* 2007. Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res* **17**: 1146–60.
68. **Deryusheva S, Krasikova A, Kulikova T,** *et al.* 2007. Tandem 41-bp repeats in chicken and Japanese quail genomes: FISH mapping and transcription analysis on lampbrush chromosomes. *Chromosoma* **116**: 519–30.
69. **Yamagishi Y, Sakuno T, Shimura M,** *et al.* 2008. Heterochromatin links to centromeric protection by recruiting shugoshin. *Nature* **455**: 251–5.
70. **May BP, Lippman ZB, Fang Y,** *et al.* 2005. Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. *PLoS Genet* **1**: e79
71. **Yin H, Lin H.** 2007. An epigenetic activation role of Piwi and a Piwi-associated piRNA in *Drosophila melanogaster*. *Nature* **450**: 304–8.
72. **Azzalin CM, Reichenbach P, Khoriauli L,** *et al.* 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* **318**: 798–801.
73. **Wicker T, Sabot F, Hua-Van A,** *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8**: 973–82.
74. **Piriyapongsa J, Polavarapu N, Borodovsky M,** *et al.* 2007. Exonization of the LTR transposable elements in human genome. *BMC Genomics* **8**: 291.
75. **Johnson R, Gamblin RJ, Ooi L,** *et al.* 2006. Identification of the REST regulon reveals extensive transposable element-mediated binding site duplication. *Nucleic Acids Res* **34**: 3862–77.
76. **Lunyak VV.** 2008. Boundaries. Boundaries. Boundaries? *Curr Opin Cell Biol* **20**: 281–7.
77. **Conte C, Dastugue B, Vaury C.** 2002. Promoter competition as a mechanism of transcriptional interference mediated by retrotransposons. *EMBO J* **21**: 3908–16.
78. **Bushey AM, Dorman ER, Corces VG.** 2008. Chromatin insulators: Regulatory mechanisms and epigenetic inheritance. *Mol Cell* **32**: 1–9.
79. **Dorman ER, Bushey AM, Corces VG.** 2007. The role of insulator elements in large-scale chromatin structure in interphase. *Semin Cell Dev Biol* **18**: 682–90.
80. **Lyon MF.** 2000. LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? *Proc Natl Acad Sci USA* **97**: 6248–9.
81. **Casacuberta E, Pardue ML.** 2005. HeT-A and TART, two *Drosophila* retrotransposons with a bona fide role in chromosome structure for more than 60 million years. *Cytogenet Genome Res* **110**: 152–9.
82. **Muotri AR, Chu VT, Marchetto MC,** *et al.* 2005. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435**: 903–10.
83. **Englander EW, Howard BH.** 1995. Nucleosome positioning by human Alu elements in chromatin. *J Biol Chem* **270**: 10091–6.
84. **Kim DD, Kim TT, Walsh T,** *et al.* 2004. Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res* **14**: 1719–25.
85. **Lev-Maor G, Ram O, Kim E,** *et al.* 2008. Intronic Alus influence alternative splicing. *PLoS Genet* **4**: e1000204.
86. **Wang H, Xing J, Grover D,** *et al.* 2005. SVA elements: a hominid-specific retroposon family. *J Mol Biol* **354**: 994–1007.
87. **Xing J, Wang H, Belancio VP,** *et al.* 2006. Emergence of primate genes by retrotransposon-mediated sequence transduction. *Proc Natl Acad Sci USA* **103**: 17608–13.
88. **Feschotte C, Pritham EJ.** 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* **41**: 331–68.
89. **Goodwin TJ, Butler MI, Poulter RT.** 2003. Cryptons: a group of tyrosine-recombinase-encoding DNA transposons from pathogenic fungi. *Microbiology* **149**: 3099–109.
90. **Pritham EJ, Feschotte C.** 2007. Massive amplification of rolling-circle transposons in the lineage of the bat Myotis lucifugus. *Proc Natl Acad Sci USA* **104**: 1895–900.
91. **Valgardsdottir R, Chiodi I, Giordano M,** *et al.* 2008. Transcription of Satellite III non-coding RNAs is a general stress response in human cells. *Nucleic Acids Res* **36**: 423–34.

92. **Dernburg AF, Broman KW, Fung JC,** *et al.* 1996. Perturbation of nuclear architecture by long-distance chromosome interactions. *Cell* **85**: 745–59.
93. **Al-Mahdawi S, Pinto RM, Ismail O,** *et al.* 2008. The Friedreich ataxia GAA repeat expansion mutation induces comparable epigenetic changes in human and transgenic mouse brain and heart tissues. *Hum Mol Genet* **17**: 735–46.
94. **Baralle M, Pastor T, Bussani E,** *et al.* 2008. Influence of Friedreich ataxia GAA noncoding repeat expansions on pre-mRNA processing. *Am J Hum Genet* **83**: 77–88.
95. **Soragni E, Herman D, Dent SY,** *et al.* 2008. Long intronic GAA*TTC repeats induce epigenetic changes and reporter gene silencing in a molecular model of Friedreich ataxia. *Nucleic Acids Res* **36**: 6056–65.
96. **Kharrat N, Al'Fadhli S, Rebai A.** 2008. Screening and validation of dinucleotide repeats in intron 1 of the human EGFR gene and its paralog in the HER2 gene. *J Recept Signal Transduct Res* **28**: 475–83.
97. **Hodgson JW, Argiropoulos B, Brock HW.** 2001. Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid polycomb group response element-dependent silencing. *Mol Cell Biol* **21**: 4528–43.
98. **Arnold R, Maueler W, Bassili G,** *et al.* 2000. The insulator protein CTCF represses transcription on binding to the (gt)(22)(ga)(15) microsatellite in intron 2 of the HLA-DRB1(*)0401 gene. *Gene* **253**: 209–14.
99. **Alvarez M, Ballantyne J.** 2009. Identification of four novel developmentally regulated gamma hemoglobin mRNA isoforms. *Exp Hematol* **37**: 285–93.
100. **Subramanian S, Mishra RK, Singh L.** 2003. Genome-wide analysis of Bkm sequences (GATA repeats): predominant association with sex chromosomes and potential role in higher order chromatin organization and function. *Bioinformatics* **19**: 681–5.
101. **Arens P, Odinot P, van Heusden AW,** *et al.* 1995. GATA- and GACA-repeats are not evenly distributed throughout the tomato genome. *Genome* **38**: 84–90.
102. **Cuadrado A, Cardoso M, Jouve N.** 2008. Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. *Cytogenet Genome Res* **120**: 210–9.
103. **McNeil JA, Smith KP, Hall LL,** *et al.* 2006. Word frequency analysis reveals enrichment of dinucleotide repeats on the human X chromosome and [GATA]n in the X escape region. *Genome Res* **16**: 477–84.
104. **Zhang Z, Xue Q.** 2005. Tri-nucleotide repeats and their association with genes in rice genome. *Biosystems* **82**: 248–56.
105. **Botta A, Rinaldi F, Catalli C,** *et al.* 2008. The CTG repeat expansion size correlates with the splicing defects observed in muscles from myotonic dystrophy type 1 patients. *J Med Genet* **45**: 639–46.
106. **Orgel LE, Crick FH.** 1980. Selfish DNA: the ultimate parasite. *Nature* **284**: 604–7.
107. **Doolittle WF, Sapienza C.** 1980. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**: 601–3.
108. **Ohno S.** 1972. So much ''junk'' DNA in our genome. *Brookhaven Symp Biol* **23**: 366–70.
109. **Zuckerkandl E.** 2002. Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine. *Genetica* **115**: 105–29.
110. **Zuckerkandl E.** 1992. Revisiting junk DNA. *J Mol Evol* **34**: 259–71.
111. **Zuckerkandl E, Cavalli G.** 2007. Combinatorial epigenetics, ''junk DNA'', and the evolution of complex organisms. *Gene* **390**: 232–42.
112. **Li YC, Korol AB, Fahima T,** *et al.* 2004. Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* **21**: 991–1007.
113. **Li YC, Korol AB, Fahima T,** *et al.* 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* **11**: 2453–65.
114. **Kashi Y, King DG.** 2006. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet* **22**: 253–9.
115. **Katti MV, Ranjekar PK, Gupta VS.** 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol Biol Evol* **18**: 1161–7.
116. **Csink AK, Henikoff S.** 1996. Genetic modification of heterochromatic association and nuclear organization in *Drosophila*. *Nature* **381**: 529–31.
117. **Fuss SH, Omura M, Mombaerts P.** 2007. Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell* **130**: 373–84.
118. **O'Sullivan JM, Sontam DM, Grierson R,** *et al.* 2009. Repeated elements coordinate the spatial organization of the yeast genome. *Yeast* **26**: 125–38.
119. **Josse T, Maurel-Zaffran C, de Vanssay A,** *et al.* 2008. Telomeric trans-silencing in Drosophila melanogaster: tissue specificity, development and functional interactions between non-homologous telomeres. *PLoS One* **3**: e3249.
120. **Taddei A, Hediger F, Neumann FR,** *et al.* 2004. The function of nuclear architecture: a genetic approach. *Annu Rev Genet* **38**: 305–45.
121. **Apostolou E, Thanos D.** 2008. Linking differential chromatin loops to transcriptional decisions. *Mol Cell* **29**: 154–6.
122. **Schoenfelder S, Fraser P.** 2008. Interchromosomal huddle kickstarts antiviral defense. *Cell* **134**: 14–6.
123. **Dmitriev P, Lipinski M, Vassetzky YS.** 2009. Pearls in the junk: dissecting the molecular pathogenesis of facioscapulohumeral muscular dystrophy. *Neuromuscul Disord* **19**: 17–20.
124. **Iqbal H, Mishra RK.** 2007. Chromatin domain boundaries: defining the functional domains in genome. *Proc Indian Nat Sci Acad* **73**: 239–253.
125. **Adkins NL, Hagerman TA, Georgel P.** 2006. GAGA protein: a multi-faceted transcription factor. *Biochem Cell Biol* **84**: 559–67.
126. **Kashi Y, King D, Soller M.** 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet* **13**: 74–8.
127. **Santi L, Wang Y, Stile MR,** *et al.* 2003. The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene Bkn3. *Plant J* **34**: 813–26.
128. **Sangwan I, O'Brian MR.** 2002. Identification of a soybean protein that interacts with GAGA element dinucleotide repeat DNA. *Plant Physiol* **129**: 1788–94.
129. **Bailey AD, Pavelitz T, Weiner AM.** 1998. The microsatellite sequence (CT)n x (GA)n promotes stable chromosomal integration of large tandem arrays of functional human U2 small nuclear RNA genes. *Mol Cell Biol* **18**: 2262–71.
130. **Gause M, Schaaf CA, Dorsett D.** 2008. Cohesin and CTCF: cooperating to control chromosome conformation? *Bioessays* **30**: 715–8.
131. **Prasanth KV, Spector DL.** 2007. Eukaryotic regulatory RNAs: an answer to the 'genome complexity' conundrum. *Genes Dev* **21**: 11–42.
132. **Pheasant M, Mattick JS.** 2007. Raising the estimate of functional human sequences. *Genome Res* **17**: 1245–53.
133. **Lavrov SA, Kibanov MV.** 2007. Noncoding RNAs and chromatin structure. *Biochemistry (Mosc)* **72**: 1422–38.
134. **Alekseyenko AA, Peng S, Larschan E,** *et al.* 2008. A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* **134**: 599–609.
135. **Brouwer JR, Willemsen R, Oostra BA.** 2009. Microsatellite repeat instability and neurological disease. *Bioessays* **31**: 71–83.
136. **Guo X, Mrazek J.** 2008. Long simple sequence repeats in host-adapted pathogens localize near genes encoding antigens, housekeeping genes, and pseudogenes. *J Mol Evol* **67**: 497–509.
137. **Huda A, Marino-Ramirez L, Landsman D,** *et al.* 2009. Repetitive DNA elements, nucleosome binding and human gene expression. *Gene* **436**: 12–22.
138. **Lawson MJ, Zhang L.** 2008. Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5′-UTR region. *Gene* **407**: 54–62.
139. **Thierry A, Bouchier C, Dujon B,** *et al.* 2008. Megasatellites: a peculiar class of giant minisatellites in genes involved in cell adhesion and pathogenicity in *Candida glabrata*. *Nucleic Acids Res* **36**: 5970–82.
140. **Polavarapu N, Marino-Ramirez L, Landsman D,** *et al.* 2008. Evolutionary rates and patterns for human transcription factor binding sites derived from repetitive DNA. *BMC Genomics* **9**: 226.