

Review



Cite this article: Sundaram V, Wysocka J. 2020 Transposable elements as a potent source of diverse *cis*-regulatory sequences in mammalian genomes. *Phil. Trans. R. Soc. B* **375**: 20190347. <http://dx.doi.org/10.1098/rstb.2019.0347>

Accepted: 23 October 2019

One contribution of 15 to a discussion meeting issue ‘Crossroads between transposons and gene regulation’.

Subject Areas:
genomics, genetics, evolution

Keywords:
transcription factor binding, transposons, enhancers, hourglass model and gene regulation

Authors for correspondence:
Vasavi Sundaram
e-mail: vsundaram@ebi.ac.uk
Joanna Wysocka
e-mail: wysocka@stanford.edu

†These authors contributed equally to this work.

Transposable elements as a potent source of diverse *cis*-regulatory sequences in mammalian genomes

Vasavi Sundaram^{1,†} and Joanna Wysocka^{2,3,4,†}

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK

²Department of Chemical and Systems Biology, ³Department of Developmental Biology, and ⁴Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, USA

ID VS, 0000-0001-6571-7027; JW, 0000-0002-6909-6544

Eukaryotic gene regulation is mediated by *cis*-regulatory elements, which are embedded within the vast non-coding genomic space and recognized by the transcription factors in a sequence- and context-dependent manner. A large proportion of eukaryotic genomes, including at least half of the human genome, are composed of transposable elements (TEs), which in their ancestral form carried their own *cis*-regulatory sequences able to exploit the host *trans* environment to promote TE transcription and facilitate transposition. Although not all present-day TE copies have retained this regulatory function, the pre-existing regulatory potential of TEs can provide a rich source of *cis*-regulatory innovation for the host. Here, we review recent evidence documenting diverse contributions of TE sequences to gene regulation by functioning as enhancers, promoters, silencers and boundary elements. We discuss how TE-derived enhancer sequences can rapidly facilitate changes in existing gene regulatory networks and mediate species- and cell-type-specific regulatory innovations, and we postulate a unique contribution of TEs to species-specific gene expression divergence in pluripotency and early embryogenesis. With advances in genome-wide technologies and analyses, systematic investigation of TEs' *cis*-regulatory potential is now possible and our understanding of the biological impact of genomic TEs is increasing.

This article is part of a discussion meeting issue ‘Crossroads between transposons and gene regulation’.

1. Introduction

Following Barbara McClintock's foundational maize kernel experiments in which transposable elements (TEs) were first discovered [1], Davidson & Britten [2] postulated a possible role of repetitive sequences and TEs in gene regulation. These prescient early predictions have now found abundant experimental support. While early work documented individual instances of TE co-option for regulatory functions, systematic analysis of TE contribution to gene regulation was for a long time limited by technological challenges. The repetitiveness of TE sequences across a genome made it difficult to precisely identify mapping locations of next-generation sequencing (NGS) data from TEs and precluded systematic perturbative experiments. However, significant advances in computational software and longer sequencing reads now facilitate mapping of individual TEs in the genome. The vast majority of TEs have lost their ability to actively transpose (fewer than 1% of human TEs are capable of transposition, reviewed in [3]), and over the course of evolutionary time have accumulated mutations that increase the uniqueness of each TE and enable higher recovery of mapped NGS reads. Therefore, large-scale epigenomic and transcriptomic studies can now include the previously ignored approximately 50% (and

perhaps up to approx. 70% [4]) of the genome derived from TEs, allowing scientists to systematically identify candidate TEs that escape epigenetic silencing and gain regulatory signatures. Additionally, high-throughput genomic technologies (such as massively parallel reporter assays or CRISPR editing/activation/interference) for quantifying regulatory potential of specific sequences, deleting genomic regions or perturbing their function now enable us to determine functional roles of TEs in gene regulation and genome organization. These technological advancements have led to a recent explosion of studies documenting TE contribution to gene regulation via various *cis*-regulatory modes, which we discuss here with a focus on mammalian gene regulation.

2. Transcription factor binding sites contributed by transposable elements

Transcription factors (TF) are proteins that regulate gene expression by binding to DNA at specific sequence motifs. Chromatin immunoprecipitation (ChIP) maps of genome-wide TF occupancy demonstrate abundant binding of diverse TFs (including STAT1 [5], TP53 [6], OCT4, NANOG and CTCF [7–9]) within TE sequences. Large-scale genome-wide analyses of 26 TFs by using sequencing (ChIP-seq) in human and mouse cells [10] revealed that up to 40% (minimum: 2%, average: 20%) of TF-binding events are derived from TEs [11]. TE-derived TF-binding sites were widespread—every TF assayed had some fraction of binding events in TEs. Moreover, waves of expansions of various TE families at different evolutionary times appear to coincide with the expansion of target repertoire for specific TFs [12]. A majority of the TEs that were bound by the given TF also contained DNA sequence motifs for the TF, suggesting direct, sequence-dependent TF binding to TEs. Transcription factors with more binding locations in the genome also had higher fractions of binding to TEs [6]. This increase in the number of TF-binding locations mediated by TEs could be associated with an increase in the TF's gene target repertoire and broader impact on gene expression, although the latter notion remains to be functionally assessed [13]. This supports the idea that TEs could contribute to coordinated regulation in the genome by spreading DNA substrate for TF binding across the genome [2].

TE-derived TF-binding sites are non-randomly distributed across TE subfamilies. Specific TE subfamilies enrich for the binding of distinct TFs [11]. Based on 26 TFs assayed by ChIP-seq in human and mouse, there are 710 TE subfamilies that enrich for binding of at least one TF. In the genome, TE subfamilies have several hundred (up to a few thousands) of individual insertions, most of which accumulate substitutions through genetic drift (figure 1*a*). It has been widely suggested that TEs form the substrate for TF-binding site evolution and innovation [14]. However, even for TE subfamilies enriched for specific TF motifs, not every genomic copy of a TE can be effectively bound by those TFs for several reasons. First, most TEs are epigenetically silenced and thus can be accessed by TFs only under conditions in which this repression can be overcome by either the high affinity of a TF for the TE-derived binding motif, its ability to function as a pioneer factor that can directly bind nucleosomal DNA, cooperativity with other TFs or global or local epigenetic changes [15–18] (figure 1*b*). Second, a certain fraction of elements within a TE subfamily

might contain a given TF-binding motif, while many others might have lost it via genetic drift [6,11] (figure 1*a*). Alternatively, uneven distribution of TF binding within a TE subfamily may arise, because only a few actively transposing TE copies acquired the motif (figure 1*a*), which they then distributed throughout the genome. For example, recent work [19] on the *Mus caroli* genome revealed a unique variant of CTCF-binding sites (motif variation: 18C→T) contributed by the species-specific expansion of rodent B2 SINE (B2_Mm1; Short INterspersed Elements—Class I, non-autonomous TEs). Finally, individual insertions within a TE subfamily may acquire TF-binding activity via mutations that complete an already present suboptimal motif (figure 1*a*). All aforementioned mechanisms contribute to the observed non-uniform binding of TFs within TE subfamilies.

The majority of TE-derived TF-binding events are species-specific and are driven by the expansion of different TE subfamilies at different evolutionary timepoints [6,8,9,11]. Species-specific effects of TEs on gene regulation are demonstrated for various TFs, including pluripotency factors (OCT4, NANOG), CTCF, tumour-suppressor TP53, STAT1 (involved in immune response [20]) and pregnancy-associated factors regulated by cAMP [21]. TEs are widely involved in placental development in mammals including the exaptation of *syncytin* [22–25] and the contribution of placental enhancers [26]. A DNA transposon (Class II TEs), MER20, is associated with approximately 13% of the differentially expressed endometrial genes across mammals and is thought to have re-modelled the placental gene regulatory network by recruiting cAMP signalling to endometrial cells. MER20 contains binding sites for various TFs, which also acquire chromatin signatures associated with functional regulatory elements [21]. In addition, ancient mammalian TEs are enriched for binding sites of TFs associated with hormone responsiveness and involved in specifying the endometrial cellular identity [26]. TE-derived TF-binding sites could also provide robustness to gene regulation programmes by increasing the redundancy in the network for TF binding. Alternatively, recent analyses of cancer samples reveal that TF-binding sites in TEs can also drive tumour formation and prognosis [27]. Taken together, there is broad evidence for TEs being bound by TFs and impacting gene regulation in direct and indirect ways.

3. Diversity of *cis*-regulatory modules encoded by transposable elements

Cis-regulatory modules represent clusters of TF binding, and many TE sequences are bound by multiple TFs. For example, rodent endogenous retroviruses (ERVs), including a mouse-specific RLTR13D5, contribute species-specific enhancers to placental development in trophoblast stem cells (TSCs) and are bound by critical trophoblast TFs, including Eomes, Cdx2 and Elf5 [28]. Similarly, pluripotency factors co-bind certain mouse-specific LTRs (i.e. long terminal repeats—Class I, autonomous TEs) and are capable of enhancing gene expression in mouse embryonic stem cells (ESCs) [29]. Specifically, RLTR9B2, RLTR9D and RLTR9E subfamilies enrich for Esrrb, Klf4 and Sox2-binding motifs, in a specific organization and orientation. The approximation of the ancestral TE state (i.e. the RepBase sequence [30], which is based on a consensus-based approximation of the ancestral

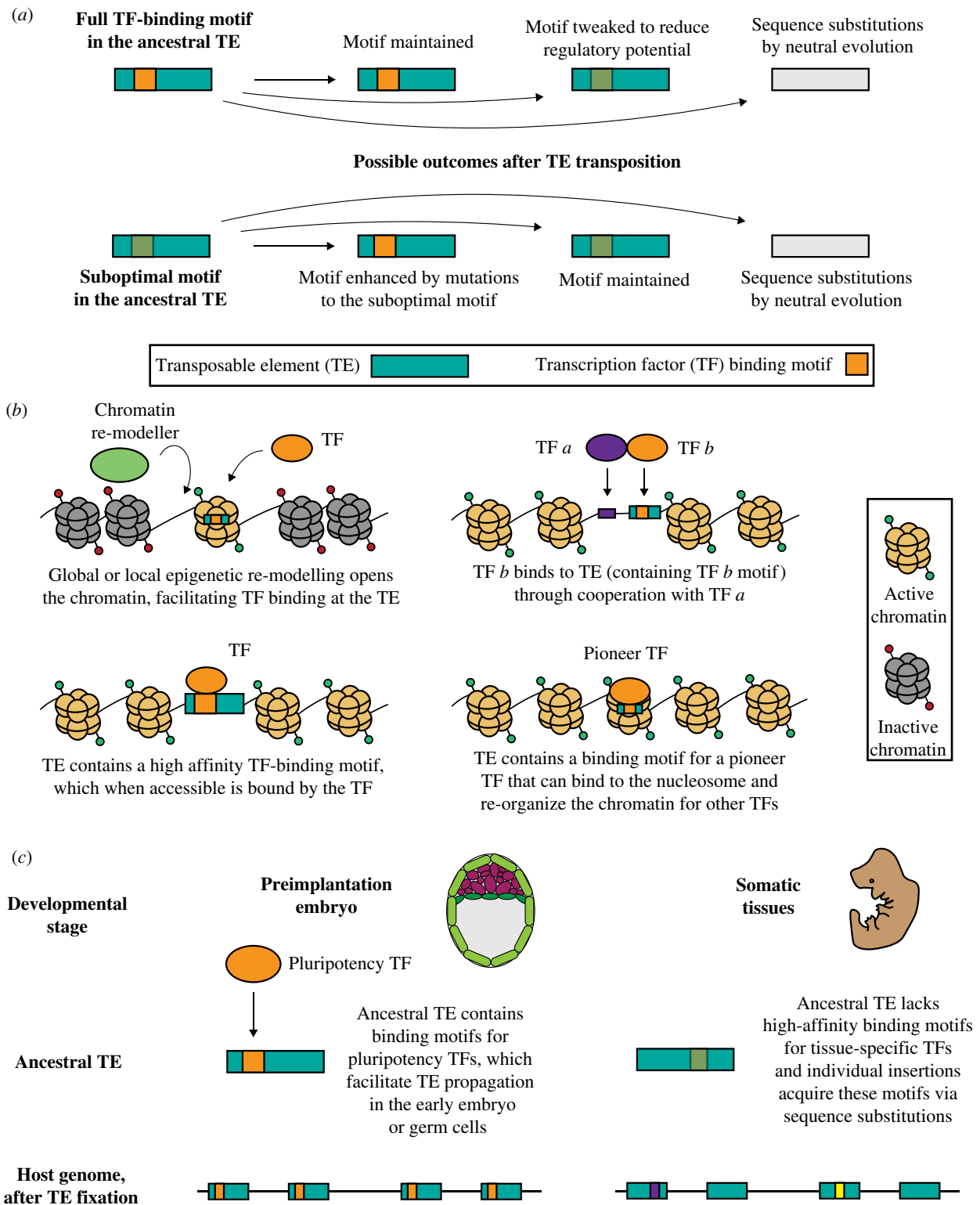


Figure 1. TF binding to TEs. (a) There are various possible outcomes from transposition of the ancestral TE (teal rectangles) that leads to variation in TF-binding motifs (orange motifs) observed in present-day TEs. When ancestral TEs contain functional TF-binding motifs (upper panel), they can spread these motifs across the genome, which might be co-opted and maintained, modified or lost by neutral substitution. Alternatively, ancestral TEs might serve as a substrate for the evolution of new or enhanced TF-binding motifs (lower panel). (b) TF binding is not only dependent on sequence but also on chromatin context. (Clockwise from top-left) TF binding can occur after chromatin re-modelling, through cooperation with another TF, through the binding of a pioneer TF to the nucleosome or through direct binding to a strong motif. (c) Differences in the TFs that bind TEs correspond to developmental stages. Preimplantation embryos express pluripotency TFs that can bind to the ancestral TE and also permit en masse TE entry into the genome. Alternatively, in somatic tissues, TEs might not have somatic TF-binding motifs but could evolve them via neutral substitutions. Ancestral TEs can contain suboptimal TF-binding motifs that become a bona fide binding site for TFs in somatic tissues, through a few nucleotide modifications.

sequence) for these subfamilies also contains the same *cis*-regulatory grammar (arrangement and orientation of motifs), suggesting that the ancestral TEs spread these *cis*-regulatory binding modules (figure 1c) across the mouse genome, after the mouse–rat split less than 12–14 million

years ago. Functional validation (using massively parallel reporter assays, MPRA [31]) of the modularity of these binding motifs reveals synergy (rather than merely additivity) to enhance gene expression in the ancestral TEs [13], as opposed to acquiring the regulatory potential via de novo mutations

over evolutionary times [12]. Similarly, in human naive pluripotent cells, evolutionarily young, ape-specific LTR5HS and SVA elements are bound and regulated by pluripotency TFs OCT4 and KLF4 [32,33].

Early embryos express totipotency/pluripotency factors and have lower epigenomic restriction which provides a window of opportunity for transcriptional reactivation of TEs (figure 1c; [34–37]). Indeed, preimplantation embryos have widespread TE transcription, mediated by permissive chromatin and pluripotency TFs that bind to TEs. Once somatic gene expression programmes are switched on, epigenomic restriction increases and reduces TE activity. It is likely that ancestral TEs lack complete and optimized binding motifs for an entire repertoire of TFs that would be required for overcoming epigenomic restriction in a given somatic tissue, and instead in some instances gain these motifs via neutral substitution (figure 1c). In agreement with this notion, only a fraction of TEs within a specific class show active epigenomic states and TF-binding activity in somatic tissues, while most are silenced. Nonetheless, recent analyses of epigenomic states of TEs across 127 human cell types and tissues [10,38,39] revealed that TEs compose one-quarter of the regulatory epigenome (defined as active regulatory and transcribed chromHMM states, DNase hypersensitivity peaks and H3K27ac peaks) [40]. One emerging theme in these studies is that motif-rich LTRs of ERVs are particularly fruitful substrates for evolving new regulatory elements for the host, with many LTRs acquiring inducible or cell type-specific regulatory potential. Across the 127 human cell and tissue samples, almost half the TEs have a signature of putative active regulatory state in at least one of the analysed cell types. Interestingly, although TEs represent a significant fraction of regulatory and transcribed chromatin states across all epigenomes, in a given epigenome TEs are relatively depleted in active states, as most are silenced [40,41].

Cis-regulatory elements such as enhancers and promoters (figure 2a) are densely populated clusters of TF-binding motifs which typically require occupancy of multiple TFs for their activation (reviewed in [42]). TEs that escape silencing and are bound by multiple TFs are much more likely to acquire features of the bona fide enhancer or promoter elements and in turn exert a functional impact on gene regulation. For example, the aforementioned mouse and human LTR elements bound by multiple pluripotency TFs are enriched for chromatin signatures associated with enhancers including H3K27ac and H3K4me1. More broadly, epigenomic surveys of different cell types and organisms found that a substantial portion of open chromatin regions are derived from TEs, including 63% of primate-specific regions [43]. Overall, TEs represent a potent source of context-specific regulators in the genome.

4. Transposable elements as a substrate for evolving new enhancers

Enhancers are canonically defined as non-coding DNA sequences that act to drive transcription independently of their relative distance, location or orientation to their cognate promoter (figure 2b). The ability of enhancers to act at a distance and in a modular and spatio-temporally restricted manner allows a gene to be regulated by multiple distal enhancers. This facilitates enormous combinatorial complexity of

gene expression repertoires and allows the emergence of diverse phenotypic outcomes from the fixed set of genes [42,44–46]. As determinants of tissue-specific gene expression, enhancers can be considered as information integration hubs (reviewed in [47]) where genomic (i.e. sequence) information is read in the context of a specific cell type (by the lineage-specific TFs), signalling environment (by TF effectors of signalling pathways) and chromatin state (permissive or restrictive for TF access), and together is translated into precise spatio-temporal control of transcription. Over evolutionary times, enhancers are rapidly evolving regulatory elements compared with promoters and gene expression [48]. Indeed, developmental enhancers are fertile targets for evolutionary change, as they both are tissue-selective (allowing modulation of target gene transcription in a subset of tissues without affecting other pleiotropic gene functions) and commonly exist in groups of redundant or partially redundant elements (facilitating accumulation of genetic variation by buffering the risk of lethality). Thus, it comes as no surprise that TEs, which harbour TF-binding sites and can transpose to disseminate them across the genome, emerged as a major substrate for evolving tissue- and species-specific enhancers.

TE-derived enhancers are involved in context-specific gene regulation [47], affecting various biological cellular states, including pregnancy-related gene expression [21], early embryonic stages in humans [49] and innate immunity [20]. In the latter case, functional contribution of MER41 to STAT1-mediated interferon- γ -responsive enhancers and gene expression has been directly demonstrated [20]. Contributions of TEs to regulatory landscapes are particularly pronounced when species-specific enhancers are considered, including through waves of repeat expansions [12]. For example, the majority of ape-specific and human-specific liver enhancers overlap TEs (77% of ape-specific enhancers compared with 16% of evolutionarily conserved enhancers) [50]. Quantitative comparisons of enhancer landscapes between chimpanzee and human cranial neural crest cells (CNCCs) showed that species-specific changes in enhancer activity can be explained by the underlying regulatory sequence divergence and is linked to expression differences in genes involved in development of craniofacial structures [51]. Even when just orthologous human and chimpanzee sequences are considered (excluding species-specific TE insertions from the analysis), nearly half of the species-specific enhancers in CNCCs are derived from TEs and enriched for specific ERV and L1 families. Thus, TE sequences contribute towards enhancer evolution between closely related species and likely drive phenotypic differences too.

5. Transposable element origin of an hourglass developmental divergence?

Despite the key contribution of TEs to species-specific enhancers, in most mouse and human tissues TEs enriched for enhancer signatures are mostly 'older' than non-regulatory genomic TEs [40,41]. This suggests that mutations that occurred over evolutionary times allowed individual insertions within specific TE subclasses to acquire mutations facilitating gain of enhancer function (figure 1a). Indeed, uterine and liver enhancers have often evolved from ancient TE sequences [26,48]. Similarly, enhancer activity divergence between human and chimpanzee CNCCs commonly arises from just a few single-nucleotide mutations in older TE insertions [51]. Interestingly,

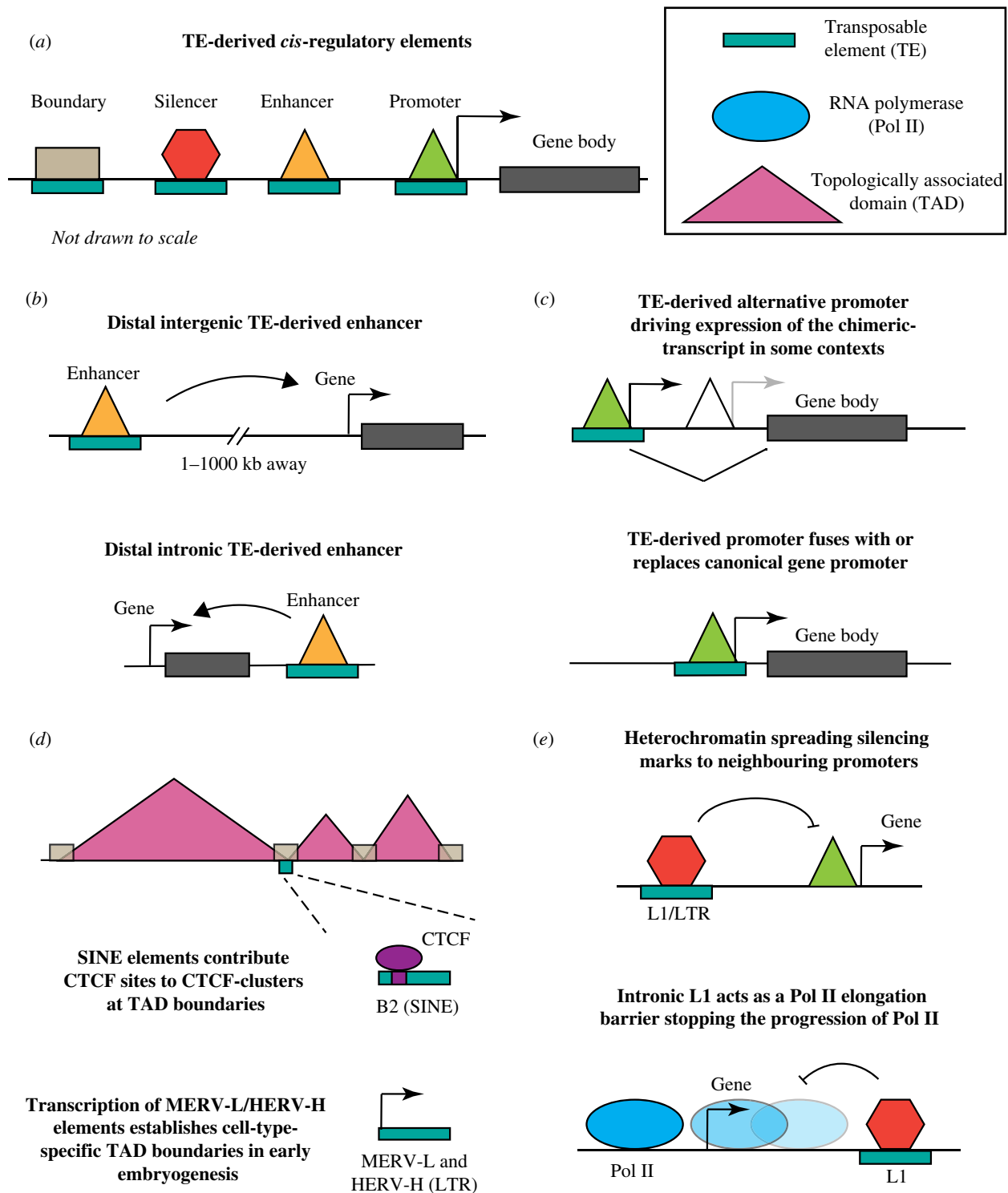


Figure 2. TE-derived *cis*-regulatory elements. (a) Here, we review TEs' role in gene expression via enhancers, promoters, boundary elements and silencers. (b) TE-derived enhancers act distally as either intergenic (upper panel) or intronic (lower panel). (c) TEs contribute promoters either as alternative promoters driving chimeric transcripts (upper panel) or as a replacement for the canonical promoter (lower panel). (d) TE-derived boundary elements contribute to topologically associated domains (TADs) by providing CTCF-binding sites (upper panel) and also maintain TADs by TE transcription (lower panel). (e) TEs can also act as a silencer by spreading heterochromatin (upper panel) or stalling Pol II elongation (lower panel).

this picture changes in preimplantation embryos and pluripotent cells, where evolutionarily young LTR elements, such as ape-specific LTR5HS in humans and murine-specific RLTR9B2, RLTR9D, RLTR9E and RLTR13D6 in the mouse, appear to have been broadly co-opted for enhancer function, with a large subset of individual elements within the given class showing the occupancy of pluripotency TFs (figure 1c) and chromatin signatures of active enhancers [29,52].

We speculate that such broad activity of evolutionarily young LTRs in naive pluripotency may not only result from

permissive epigenetic state associated with global DNA hypomethylation, but may be a consequence of the fact that these ancient retroviruses must have been able to replicate in germ cells or preimplantation embryo cells in order to persist through vertical transmission (figure 1c). Thus, LTRs of retroviruses that successfully endogenized have likely been optimized for directing expression in early embryo/germ cells to begin with, and accumulating mutations over evolutionary times may decrease, rather than increase, their regulatory activity. One such example of a TE is HERV-H

whose expression in human ESCs is not only abundant but also a pluripotency marker [53] and is mediated by pluripotency TFs [54,55]. Maladaptive insertions of such elements can be selected out of the population prior to implantation, whereas those that persist either are neutral or provide some adaptive benefits for the host. Interestingly, an hourglass model of developmental constraint postulates high regulatory divergence between closely related species in early and late development, and greater constraint at the phylotypic stage. It is tempting to speculate that the broad utilization of young (and often highly species-specific) TEs for regulation in the early embryo provides a major contribution to the species-specific differences in gene expression that have been observed during early embryogenesis [56–58].

Importantly, studies that perturb the function of these naive pluripotency-associated LTRs via CRISPR interference (CRISPRi) or LTR editing have found support for their contribution to long-range gene regulation [49,52]. Interestingly, however, effects of LTR5HS inactivation on gene expression in human cells are much more pervasive and stronger in magnitude than those seen for RLTR13D6 elements in mouse cells. Nearly 300 human genes, spanning a wide range of distances and distinct orientations with respect to the LTR, show significant dependency on the LTR5HS activity [49]. By contrast, only a minority RLTR13D5/RLTR13D6 elements seem to have a detectable effect on gene regulation in mouse ESCs [52]. Although a more robust CRISPRi approach used in the human study could potentially account for these disparate observations, it is also possible that genuine biological differences in inherent enhancer strength of the LTRs, distinct redundancies within regulatory landscapes or differences in selective pressures may account for diverse magnitude of LTR contributions to early embryonic gene regulation in apes and rodents. Regardless, these observations underscore the importance of functional assessment of regulatory contributions of specific TEs—or indeed, any candidate enhancer sequences—via perturbative experiments.

Over the recent years, since more systematic perturbative experiments of candidate non-coding sequences have become possible, we have learned that mammalian regulatory landscapes are extremely complex and often contain redundant enhancers and that the impact of individual putative enhancer deletion/silencing on expression of target genes is dependent on inherent enhancer strength, number and strength of other active enhancers acting on the same gene in a given cell type, and spatial constraints reflected in the frequency of contact with the target promoter(s) [59–63]. Furthermore, while relationships among enhancers within a regulatory landscape are often additive [64], examples of other relationships, including synergistic, competitive and hierarchical, have been described in the literature (reviewed in [42]). We should expect that TE-derived enhancers will participate in these complex relationships and thus their activity (or lack thereof) should be interpreted in a broader context of a given regulatory landscape.

6. Transposable element-derived promoters drive context-specific gene expression

Transcriptional promoters, like enhancers, consist of multiple TF-binding sites. While enhancers can act over very large

genomic distances, promoters are immediately proximal to gene transcription start sites and contain *cis*-signals that allow efficient assembly of the Pol II preinitiation complex [45] (figure 2c). Although many promoters are bidirectional, most show a strong bias in the direction of transcription (of long stable transcripts), again in contrast with most enhancers, which give rise to bidirectional transcripts termed eRNAs (typically short and unstable) [45]. There is currently much debate in the field as to which functional features distinguish enhancers from promoters, and there appears to be a continuum, with a subset of elements able to function as both. In that context, it is interesting that LTR elements, which, by definition, are retroviral promoters, can also act as long-range enhancers for the host genes, independently of the transcriptional directionality or orientation of the LTR relative to the target promoter [49].

Nonetheless, LTRs in ERV sequences provide ready-made promoters in human and mouse [55,65–67]. TE-derived promoters [68] can either fuse with/replace a canonical gene promoter or serve as an alternative promoter either upstream or downstream from the canonical transcription start site (figure 2c). TE-derived alternative promoters can drive expression of chimeric transcripts, which may vary in levels when compared with those driven by the canonical promoter (figure 2c; reviewed in [69]). In some cases, such transcripts can also differ in their coding potential, giving rise to either chimeric or truncated proteins, and thus facilitating evolutionary innovation at the protein level [69]. Around 18% of human transcription start sites are defined by CAGE-seq overlap TEs [70], and tend to be in gene-dense regions often functioning as alternative promoters [71]. The first well-characterized example of TE-derived promoters was found in viable yellow agouti (A^{vy}) mice, where the A^{vy} allele contains an IAP (intracisternal A particle; LTR) element in the 5' end of the *agouti* (*A*) allele [72]. Variable DNA methylation levels across individuals at the A^{vy} promoter result in a variable coat colour in each mouse, in addition to other traits [73], and can also be influenced by maternal diet [74,75]. In the totipotent two-cell stage (2C) of mouse embryos and in 2C-like cells which spontaneously arise in mouse ESC cultures, MERVL elements are bound by TF Dux and serve as alternative promoters, giving rise to chimeric transcripts of the so-called 2C genes [76,77]. The human genome also encodes context-specific TE-derived promoters. For example, more than a thousand human genes have promoters derived from TEs [78,79]. LTRs among these TE-derived promoters enrich for cell-specific gene regulation. Epigenetic modification by deleting DNA methyltransferases in human neuronal progenitor cells results in specific activation of young, hominoid-specific LINE-1 (L1) elements [80], which can then serve as alternative promoters for neuronal protein-coding genes.

LINE-L1 (i.e. Long Interspersed Elements; Class I autonomous TEs) elements are also capable of bidirectional transcription: L1s have two promoters in their 5' UTR—one sense and one antisense [81]. L1 antisense promoters produce many chimeric transcripts from adjacent genes [82]. Disease conditions that alter epigenetic landscape can promote aberrant activation of promoters contained within TEs. Recent work analysing transcripts in thousands of cancer samples (from The Cancer Genome Atlas—TCGA—Program) demonstrates are epigenetic reactivation of cryptic promoters in TEs that can drive gene expression in cancer, often referred to

as onco-exaptation ([27,83]; examples are listed in [69]). TE-derived onco-exaptation events are not only responsible for a significant fraction of the target oncogene's expression in the examined samples but were also associated with poor patient survival. Taken together, TE-derived gene promoters can impact both developmental and diseased transcription states.

7. Role of transposable elements in three-dimensional genome architecture

In addition to sequences that promote or enhance gene expression, TEs also contribute to the maintenance of genome architecture. Three-dimensional chromatin organization not only determines the packaging of DNA in the cell, but also influences gene regulation by demarcating regulatory neighbourhoods (figure 2*d*). Measurements of chromosome contacts (from a genome-wide chromosome conformation capture assay—Hi-C, as reviewed in [84]) identified self-associating domains known as topologically associating domains (TADs; [85,86]). TAD boundaries are enriched for binding of CTCF, a zinc finger protein that mediates the formation of structural chromatin loops by restraining cohesin-mediated loop extrusion [86–88]. SINE elements are enriched for CTCF binding [9,11], and are over-represented at TAD boundaries [85]. However, TAD boundaries are typically deeply conserved in evolution [89], suggesting that species-specific rewiring of TADs by TEs is not common. Instead, it has been suggested that rodent SINEs can contribute to the maintenance of clustered CTCF sites at TAD boundaries, which in turn promotes the maintenance of genome organization [90]. Clusters of CTCF sites are over-represented at TAD boundaries and likely provide buffering against turnover of individual sites to stabilize TADs and chromatin loops. Along with maintaining chromatin contact loops through TE-derived CTCF sites, TEs can also contribute to the establishment of species-specific chromatin loops by depositing novel anchor CTCF motifs [91].

Although most TAD-boundary elements are cell-type invariant, evolutionarily conserved and enriched for CTCF motifs, a small subset of TAD boundaries appears to escape this generality [92]. Indeed, a recent report highlights pluripotent stem-cell-specific TAD boundaries that are detected in human ESCs, but weakened during differentiation to cardiomyocytes [93]. These cell-type-specific boundaries are devoid of clusters of CTCF sites, and instead enriched in highly transcribed HERV-H elements, a primate-specific class of LTR TEs with previously described roles in pluripotency [53,54]. The establishment of TAD boundaries is dependent on the transcription of HERV-H (figure 2*d*), as directed silencing of the HERV-H can eliminate the insulation and impact the gene regulatory landscape. Similarly, in mouse 2C cells, MERVL elements are not only the main source of promoters for directing expression of the early 2C genes, but also facilitate the formation of chromatin domain boundaries in a manner that coincides with their transcriptional upregulation [94]. Intriguingly, these new observations again highlight a particularly prevalent role for TEs in genome regulation and organization during early development.

8. Collateral effects of transposable element silencing on gene regulation

Although TEs provide a rich substrate for evolving diverse *cis*-regulatory elements for the host, most genomic TEs are epigenetically silenced to prevent mutagenic effects of their transposition. These silencing events often occur within transcriptional gene units, as a large subset of TEs (estimated at approx. 60% in the human and mouse genomes) is located in introns [95]. For genes that are active, this can lead to conflicts between polymerase passage through the gene and heterochromatin formation at intronic or promoter-proximal TEs, and in turn can affect host gene expression (figure 2*e*). One feature of heterochromatin is its ability to spread beyond the initial recruitment site via mechanisms that often involve coupling of the epigenetic writers and readers (reviewed in [96]). For example, in tetrapod vertebrates, KRAB zinc finger proteins mediate the repression of specific TE subclasses via DNA-sequence recognition and recruitment of a repressive complex including H3K9me3 methyltransferase SETDB1 and the reader of H3K9me3, HP1 [97,98]. KRAB-domain-mediated silencing typically spreads to several, and in some cases even tens of kilobases away from the initial recruitment site, which can lead to the repression at neighbouring gene promoters [99,100]. Similarly, DNA methylation can spread from TEs to nearby promoters, although some studies suggest that such events are relatively rare and actively counteracted by the host [100,101].

Interestingly, TE silencing events can impact gene expression even over large genomic distances from promoters. For example, evolutionarily younger, full-length LINE-1 (L1) elements located within introns of active genes are preferentially targeted for repression and H3K9me3-deposition by the human silencing hub (HUSH) complex [102,103]. From the genome defence perspective, such elements—encompassing transposition-competent L1PA1 (L1Hs)—need to be suppressed as they pose the highest threat to genomic integrity, especially when embedded within a transcriptionally permissive environment. However, these intronic silencing events create heterochromatic islands of at least 7 kb length (the length of a full-length L1) that elongating Pol II must pass through, and genic H3K9me3 as well as L1 insertions have both been shown to decrease Pol II elongation rates [104,105]. Consequently, transcript levels of active genes containing evolutionarily young full-length L1s are downregulated, often subtly, in an L1 sequence-dependent and HUSH complex-dependent manner [102]. These L1s can be thought of as genetically determined but epigenetically regulated silencer elements that can quantitatively modulate expression levels of hundreds of human genes at large distances from gene promoters. Thus, the host genome needs to balance the need for young intragenic TE suppression with the collateral effects that such silencing can have on gene expression.

9. Concluding remarks

Evolutionary differences across species are thought to be largely driven by changes in gene expression, mediated by divergence in *cis*-regulatory elements [106,107]. Recent progress in the field revealed that a substantial portion of mammalian *cis*-regulatory sequences is derived from TEs. These TE-derived *cis*-regulatory elements are often cell type- and species/clade-specific and can contribute to gene

expression regulation through many diverse mechanisms, which we have reviewed here.

The prevalence of TE utilization for regulatory functions may differ between cell types and developmental stages. Indeed, TEs seem to play an outsized role during mammalian pre- and peri-implantation development, where whole subclasses of TEs (such as MERVL in mice, or HERV-K and HERV-H in humans) function in host gene regulation as alternative promoters, enhancers or boundary elements. This widespread utilization of TEs during early development is likely facilitated both by the global epigenomic de-repression during this time of embryogenesis and by the fact that to successfully propagate through vertical transmission, ancestral TEs must have been able to direct transcription of their own sequences in toti/pluripotent cells or germ cells. Consequently, many evolutionarily young TEs, which are typically species- or clade-specific, can bind toti- and pluripotency-associated TFs. We speculate that, at least in mammals, transcriptional divergence during early development, as postulated by the hourglass model, is largely mediated by the massive co-option of TEs for host gene regulation.

By contrast, in somatic tissues, epigenomic control is restored and TEs are largely repressed via a multitude of chromatin-silencing mechanisms. However, different fractions of

TEs, often representing older insertions, escape the epigenomic control by accumulating sequence changes that make them compatible with cell-type- and context-specific functions. Since we have only begun to systematically assess the function of TEs in gene control, it is likely that we are still vastly underestimating their impact as well as the diversity of mechanisms by which TEs can influence transcription, post-transcriptional gene regulation, genome organization and evolutionary divergence. Nevertheless, our understanding of TEs has come a long way from the notion of 'junk' DNA. What persists is Barbara McClintock's early vision of TEs as 'controlling elements' [108].

Data accessibility. This article has no additional data.

Authors' contributions. Both authors contributed to all aspects of the manuscript.

Competing interests. We have no competing interests to declare.

Funding. V.S. was supported by a fellowship from the EMBL Interdisciplinary Postdoc (EIPD) programme under Marie Skłodowska-Curie actions COFUND (grant no. 664726). This work was supported by NIH R35 GM131757 and Howard Hughes Medical Institute grants to J.W.

Acknowledgements. V.S. thanks Paul Flicek for helpful discussions and comments on the manuscript, and Miguel Branco for helpful discussions during various stages of drafting this manuscript.

References

- McClintock B. 1950 The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA* **36**, 344–355. (doi:10.1073/pnas.36.6.344)
- Davidson EH, Britten RJ. 1979 Regulation of gene expression: possible role of repetitive sequences. *Science* **204**, 1052–1059. (doi:10.1126/science.451548)
- Mills RE, Bennett EA, Iskow RC, Devine SE. 2007 Which transposable elements are active in the human genome? *Trends Genet.* **23**, 183–191. (doi:10.1016/j.tig.2007.02.006)
- de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011 Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384. (doi:10.1371/journal.pgen.1002384)
- Schmid CD, Bucher P. 2010 MER41 repeat sequences contain inducible STAT1 binding sites. *PLoS ONE* **5**, 11425. (doi:10.1371/journal.pone.0011425)
- Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D. 2007 Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl Acad. Sci. USA* **104**, 18 613–18 618. (doi:10.1073/pnas.0703637104)
- Bourque G *et al.* 2008 Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762. (doi:10.1101/gr.080663.108)
- Kunarsow G, Chia N-Y, Jeyakani J, Hwang C, Lu X, Chan Y-S, Ng H-H, Bourque G. 2010 Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631–634. (doi:10.1038/ng.600)
- Schmidt D *et al.* 2012 Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335–348. (doi:10.1016/j.cell.2011.11.058)
- Yue F *et al.* 2014 A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364. (doi:10.1038/nature13992)
- Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014 Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976. (doi:10.1101/gr.168872.113)
- Marnetto D, Mantica F, Molineri I, Grassi E, Pesando I, Provero P. 2018 Evolutionary rewiring of human regulatory networks by waves of genome expansion. *Am. J. Hum. Genet.* **102**, 207–218. (doi:10.1016/j.ajhg.2017.12.014)
- Sundaram V, Wang T. 2018 Transposable element mediated innovation in gene regulatory landscapes of cells: re-visiting the 'gene-battery' model. *Bioessays* **40**, 1700155. (doi:10.1002/bies.201700155)
- Bourque G. 2009 Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr. Opin. Genet. Dev.* **19**, 607–612. (doi:10.1016/j.gde.2009.10.013)
- Deniz Ö, Frost JM, Branco MR. 2019 Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **20**, 417–431. (doi:10.1038/s41576-019-0106-6)
- Sanchez-Luque FJ *et al.* 2019 LINE-1 evasion of epigenetic repression in humans. *Mol. Cell* **75**, 590–604. (doi:10.1016/j.molcel.2019.05.024)
- Yin Y *et al.* 2017 Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239. (doi:10.1126/science.aaj2239)
- Lea AJ *et al.* 2018 Genome-wide quantification of the effects of DNA methylation on human gene regulation. *eLife* **7**, e37513. (doi:10.7554/eLife.37513)
- Thybert D *et al.* 2018 Repeat associated mechanisms of genome evolution and function revealed by the *Mus caroli* and *Mus pahari* genomes. *Genome Res.* **28**, 448–459. (doi:10.1101/gr.234096.117)
- Chuong EB, Elde NC, Feschotte C. 2016 Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **351**, 1083–1087. (doi:10.1126/science.aad5497)
- Lynch VJ, Leclerc RD, May G, Wagner GP. 2011 Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat. Genet.* **43**, 1154–1159. (doi:10.1038/ng.917)
- Emera D, Wagner GP. 2012 Transposable element recruitments in the mammalian placenta: impacts and mechanisms. *Brief. Funct. Genomics* **11**, 267–276. (doi:10.1093/bfpg/els013)
- Cornelis G *et al.* 2013 Captured retroviral envelope syncytin gene associated with the unique placental structure of higher ruminants. *Proc. Natl Acad. Sci. USA* **110**, E828–E837. (doi:10.1073/pnas.1215787110)
- Dupressoir A, Lavalie C, Heidmann T. 2012 From ancestral infectious retroviruses to bona fide cellular genes: role of the captured *syncytins* in

- placentation. *Placenta* **33**, 663–671. (doi:10.1016/j.placenta.2012.05.005)
25. Mangeney M, Renard M, Schlecht-Louf G, Bouallaga I, Heidmann O, Letzelter C, Richaud A, Ducos B, Heidmann T. 2007 Placental syncytiotrophoblasts: genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl Acad. Sci. USA* **104**, 20 534–20 539. (doi:10.1073/pnas.0707873105)
 26. Lynch VJ *et al.* 2015 Ancient transposable elements transformed the uterine regulatory landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep.* **10**, 551–562. (doi:10.1016/j.celrep.2014.12.052)
 27. Jang HS *et al.* 2019 Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **51**, 920. (doi:10.1038/s41588-019-0416-9)
 28. Chuong EB, Rumi MAK, Soares MJ, Baker JC. 2013 Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* **45**, 325–329. (doi:10.1038/ng.2553)
 29. Sundaram V *et al.* 2017 Functional *cis*-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat. Commun.* **8**, 14550. (doi:10.1038/ncomms14550)
 30. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005 Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467. (doi:10.1159/000084979)
 31. Kwasniewski JC, Fiore C, Chaudhari HG, Cohen BA. 2014 High-throughput functional testing of ENCODE segmentation predictions. *Genome Res.* **24**, 1595–1602. (doi:10.1101/gr.173518.114)
 32. Grow EJ *et al.* 2015 Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225. (doi:10.1038/nature14308)
 33. Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. 2019 Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* **24**, 724–735. (doi:10.1016/j.stem.2019.03.012)
 34. Surani MA, Hayashi K, Hajkova P. 2007 Genetic and epigenetic regulators of pluripotency. *Cell* **128**, 747–762. (doi:10.1016/j.cell.2007.02.010)
 35. Biechele S, Lin CJ, Rinaudo PF, Ramalho-Santos M. 2015 Unwind and transcribe: chromatin reprogramming in the early mammalian embryo. *Curr. Opin. Genet. Dev.* **34**, 17–23. (doi:10.1016/j.gde.2015.06.003)
 36. Eckersley-Maslin MA, Alda-Catalinas C, Reik W. 2018 Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mol. Cell Biol.* **19**, 436–450. (doi:10.1038/s41580-018-0008-z)
 37. Theunissen TW, Jaenisch R. 2017 Mechanisms of gene regulation in human embryos and pluripotent stem cells. *Development* **144**, 4496–4509. (doi:10.1242/dev.157404)
 38. Consortium RE *et al.* 2015 Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330. (doi:10.1038/nature14248)
 39. Gerstein MB *et al.* 2012 Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100. (doi:10.1038/nature11245)
 40. Pehrsson EC, Choudhary MN, Sundaram V, Wang T. 2019 The epigenomic landscape of transposable elements across normal human development and anatomy. *Nat. Commun.* **10**, 5640. (doi:10.1038/s41467-019-13555-x)
 41. Simonti CN, Pavličev M, Capra JA. 2017 Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. *Mol. Biol. Evol.* **34**, 2856–2869. (doi:10.1093/molbev/msx219)
 42. Long HK, Prescott SL, Wysocka J. 2016 Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170–1187. (doi:10.1016/j.cell.2016.09.018)
 43. Jacques P-É, Jeyakani J, Bourque G. 2013 The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* **9**, e1003504. (doi:10.1371/journal.pgen.1003504)
 44. Schaffner W. 2015 Enhancers, enhancers—from their discovery to today's universe of transcription enhancers. *Biol. Chem.* **396**, 311–327. (doi:10.1515/hsz-2014-0303)
 45. Kim TK, Shiekhhattar R. 2015 Architectural and functional commonalities between enhancers and promoters. *Cell* **162**, 948–959. (doi:10.1016/j.cell.2015.08.008)
 46. Xie M *et al.* 2013 DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat. Genet.* **45**, 836–841. (doi:10.1038/ng.2649)
 47. Buecker C, Wysocka J. 2012 Enhancers as information integration hubs in development: lessons from genomics. *Trends Genet.* **28**, 276–284. (doi:10.1016/j.tig.2012.02.008)
 48. Villar D *et al.* 2015 Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566. (doi:10.1016/j.cell.2015.01.006)
 49. Fuentes DR, Swigut T, Wysocka J. 2018 Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *eLife* **7**, e35989. (doi:10.7554/eLife.35989)
 50. Trizzino M, Park YS, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD. 2017 Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* **27**, 1623–1633. (doi:10.1101/gr.218149.116)
 51. Prescott SL, Srinivasan R, Marchetto M, Grishina I, Narvaiza I, Selleri L, Gage F, Swigut T, Wysocka J. 2015 Enhancer divergence and *cis*-regulatory evolution in the human and chimp neural crest. *Cell* **163**, 68–83. (doi:10.1016/j.cell.2015.08.036)
 52. Todd CD, Deniz Ö, Taylor D, Branco MR. 2019 Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *eLife* **8**, e44344. (doi:10.7554/eLife.44344)
 53. Santoni FA, Guerra J, Luban J. 2012 HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9**, 111. (doi:10.1186/1742-4690-9-111)
 54. Ohnuki M *et al.* 2014 Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc. Natl Acad. Sci. USA* **111**, 12 426–12 431. (doi:10.1073/pnas.1413299111)
 55. Wang J *et al.* 2014 Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409. (doi:10.1038/nature13804)
 56. Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010 Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814. (doi:10.1038/nature09634)
 57. Irie N, Kuratani S. 2011 Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nat. Commun.* **2**, 248. (doi:10.1038/ncomms1248)
 58. Kuijk E, Geijsen N, Cuppen E. 2015 Pluripotency in the light of the developmental hourglass. *Biol. Rev.* **90**, 428–443. (doi:10.1111/brv.12117)
 59. Osterwalder M *et al.* 2018 Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243. (doi:10.1038/nature25461)
 60. Hay D *et al.* 2016 Genetic dissection of the α -globin super-enhancer *in vivo*. *Nat. Genet.* **48**, 895–903. (doi:10.1038/ng.3605)
 61. Gasperini M *et al.* 2019 A genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390. (doi:10.1016/j.cell.2018.11.029)
 62. Fulco CP *et al.* 2019 Activity-by-contact model of enhancer specificity from thousands of CRISPR perturbations. *bioRxiv*, 529990. (doi:10.1101/529990)
 63. Bahr C *et al.* 2018 A *Myc* enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **373**, 25. (doi:10.1038/nature25193)
 64. Berthelot C, Villar D, Horvath J, Odom DT, Flicek P. 2017 Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nat. Ecol. Evol.* **2**, 152–163. (doi:10.1038/s41559-017-0377-2)
 65. Dunn CA, van de Lagemaat LN, Baillie GJ, Mager DL. 2005 Endogenous retrovirus long terminal repeats as ready-to-use mobile promoters: the case of primate β 3GAL-T5. *Gene* **364**, 2–12. (doi:10.1016/j.gene.2005.05.045)
 66. Conley AB, Piriyaopongsa J, King Jordan I. 2008 Retroviral promoters in the human genome. *Bioinformatics* **24**, 1563–1567. (doi:10.1093/bioinformatics/btn243)
 67. Cohen CJ, Lock WM, Mager DL. 2009 Endogenous retroviral LTRs as promoters for human genes: a

- critical assessment. *Gene* **448**, 105–114. (doi:10.1016/j.gene.2009.06.020)
68. Jordan IK, Rogozin IB, Glazko GV, Koonin EV. 2003 Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.* **19**, 68–72. (doi:10.1016/S0168-9525(02)00006-9)
69. Babaian A, Mager DL. 2016 Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* **7**, 24. (doi:10.1186/s13100-016-0080-x)
70. Djebali S *et al.* 2012 Landscape of transcription in human cells. *Nature* **489**, 101–108. (doi:10.1038/nature11233)
71. Faulkner GJ *et al.* 2009 The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563–571. (doi:10.1038/ng.368)
72. Present HV, Vrieling H, Miller KA, Wolff GL, Barsh GS. 1994 Neomorphic *agouti* mutations in obese yellow mice. *Nat. Genet.* **8**, 59–65. (doi:10.1038/ng0994-59)
73. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999 Epigenetic inheritance at the *agouti* locus in the mouse. *Nat. Genet.* **23**, 314–318. (doi:10.1038/15490)
74. Waterland RA, Jirtle RL. 2003 Transposable elements: targets for early nutritional effects on epigenetic gene regulation. *Mol. Cell. Biol.* **23**, 5293–5300. (doi:10.1128/MCB.23.15.5293-5300.2003)
75. Whitelaw E, Martin DI. 2001 Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* **27**, 361–365. (doi:10.1038/86850)
76. Macfarlan TS *et al.* 2012 Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63. (doi:10.1038/nature11244)
77. Hendrickson PG *et al.* 2017 Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934. (doi:10.1038/ng.3844)
78. Huda A, Bowen NJ, Conley AB, Jordan IK. 2011 Epigenetic regulation of transposable element derived human gene promoters. *Gene* **47**, 39–48. (doi:10.1016/j.gene.2010.12.010)
79. Kellner M, Makalowski W. 2019 Transposable elements significantly contributed to the core promoters in the human genome. *Sci. China Life Sci.* **62**, 489–497. (doi:10.1007/s11427-018-9449-0)
80. Jönsson ME *et al.* 2019 Activation of neuronal genes via LINE-1 elements upon global DNA demethylation in human neural progenitors. *Nat. Commun.* **10**, 3182. (doi:10.1038/s41467-019-11150-8)
81. Redik K, Speek M. 2006 L1 antisense promoter drives tissue-specific transcription of human genes. *J. Biomed. Biotechnol.* **2006**, 71753. (doi:10.1155/JBB/2006/71753)
82. Nigumann P, Redik K, Mätlik K, Speek M. 2002 Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* **79**, 628–634. (doi:10.1006/geno.2002.6758)
83. Brocks D *et al.* 2017 DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat. Genet.* **49**, 1052–1060. (doi:10.1038/ng.3889)
84. Schmitt AD, Hu M, Ren B. 2016 Genome-wide mapping and analysis of chromosome architecture. *Nat. Rev. Mol. Cell Biol.* **17**, 743–755. (doi:10.1038/nrm.2016.104)
85. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012 Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380. (doi:10.1038/nature11082)
86. Rao SSP *et al.* 2014 A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680. (doi:10.1016/j.cell.2014.11.021)
87. Lieberman-Aiden E *et al.* 2009 Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293. (doi:10.1126/science.1181369)
88. Rao SSP *et al.* 2017 Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320. (doi:10.1016/j.cell.2017.09.026)
89. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DÂT, Tanay A, Hadjur S. 2015 Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* **10**, 1297–1309. (doi:10.1016/j.celrep.2015.02.004)
90. Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, Flicek P. 2020 Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* **21**, 5. (doi:10.1186/s13059-019-1894-x)
91. Choudhary MN, Friedman RZ, Wang JT, Jang HS, Zhuo X, Wang T. 2018 Co-opted transposons help perpetuate conserved higher-order chromosomal structures. *bioRxiv*, 485342. (doi:10.1101/485342)
92. Eres IE, Luo K, Hsiao CJ, Blake LE, Gilad Y. 2019 Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLoS Genet.* **15**, e1008278. (doi:10.1371/journal.pgen.1008278)
93. Zhang Y *et al.* 2019 Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **51**, 1380–1388. (doi:10.1038/s41588-019-0479-7)
94. Kruse K, Diaz N, Enriquez-Gasca R, Gaume X, Torres-Padilla M-E, Vaquerizas JM. 2019 Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. *bioRxiv*, 523712. (doi:10.1101/523712)
95. Sela N, Mersch B, Gal-Mark N, Lev-Maor G, Hotz-Wagenblatt A, Ast G. 2007 Comparative analysis of transposed element insertion within human and mouse genomes reveals *Alu*'s unique role in shaping the human transcriptome. *Genome Biol.* **8**, R127. (doi:10.1186/gb-2007-8-6-r127)
96. Allshire RC, Madhani HD. 2017 Ten principles of heterochromatin formation and function. *Nat. Rev. Mol. Cell Biol.* **19**, 229–244. (doi:10.1038/nrm.2017.119)
97. Ecco G, Imbeault M, Trono D. 2017 KRAB zinc finger proteins. *Development* **144**, 2719–2729. (doi:10.1242/dev.132605)
98. Yang P, Wang Y, Macfarlan TS. 2017 The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet.* **33**, 871–881. (doi:10.1016/j.tig.2017.08.006)
99. Groner AC, Meylan S, Ciuffi A, Zangger N, Ambrosini G. 2010 KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet.* **6**, 1000869. (doi:10.1371/journal.pgen.1000869)
100. Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K. 2011 Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet.* **7**, 1002301. (doi:10.1371/journal.pgen.1002301)
101. Rebollo R, Miceli-Royer K, Zhang Y, Farivar S, Gagnier L, Mager DL. 2012 Epigenetic interplay between mouse endogenous retroviruses and host genes. *Genome Biol.* **13**, R89. (doi:10.1186/gb-2012-13-10-r89)
102. Liu N, Lee CH, Swigut T, Grow E, Gu B, Bassic MC, Wysocka J. 2018 Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature* **553**, 228–232. (doi:10.1038/nature25179)
103. Robbez-Masson L *et al.* 2018 The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. *Genome Res.* **28**, 836–845. (doi:10.1101/gr.228171.117)
104. Saint-André V, Batsché E, Rachez C, Muchardt C. 2011 Histone H3 lysine 9 trimethylation and HP1 γ favor inclusion of alternative exons. *Nat. Struct. Mol. Biol.* **18**, 337–344. (doi:10.1038/nsmb.1995)
105. Han JS, Szak ST, Boeke JD. 2004 Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429**, 268–274. (doi:10.1038/nature02536)
106. King M-C, Wilson AC. 1975 Evolution at two levels in humans and chimpanzees. *Science* **188**, 4184. (doi:10.1126/science.1090005)
107. Wray GA. 2007 The evolutionary significance of cis-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216. (doi:10.1038/nrg2063)
108. McClintock B. 1956 Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol.* **21**, 197–216. (doi:10.1101/SQB.1956.021.01.017)