Transposable elements as genetic accelerators of evolution: contribution to genome size, gene regulatory network rewiring and morphological innovation

Hidenori Nishihara*

School of Life Science and Technology, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8501, Japan

In the current era, as a growing number of genome sequence assemblies have been reported in animals, an in-depth analysis of transposable elements (TEs) is one of the most fundamental and essential studies for evolutionary genomics. Although TEs have, in general, been regarded as non-functional junk/selfish DNA, parasitic elements or harmful mutagens, studies have revealed that TEs have had a substantial and sometimes beneficial impact on host genomes in several ways. First, TEs are themselves diverse and thus provide lineage-specific characteristics to the genomes. Second, because TEs constitute a substantial fraction of animal genomes, they are a major contributing factor to evolutionary changes in genome size and composition. Third, host organisms have co-opted many repetitive sequences as genes, *cis*-regulatory elements and chromatin domain boundaries, which alter gene regulatory networks and in addition are partly involved in morphological evolution, as has been well documented in mammals. Here, I review the impact of TEs on various aspects of the genome, such as genome size and diversity in animals, as well as the evolution of gene networks and genome architecture in mammals. Given that a number of TE families probably remain to be discovered in many non-model organisms, unknown TEs may have contributed to gene networks in a much wider variety of animals than considered previously.

Key words: animal evolution, co-option, genome size, retrotransposons, transposable elements

INTRODUCTION

Approximately half of the human genome is composed of transposable elements (TEs) (Fig. 1A; International Human Genome Sequencing Consortium, 2001). TEs are mobile elements that can mobilize their sequences with or without proliferation within the genome. Although different kinds of TEs are found in various organisms, all major TEs can be classified into two main categories, retrotransposons and DNA transposons (Fig. 1B; Kazazian, 2004). Retrotransposons, also known as Class I TEs, propagate their copy sequences in the host genome by reverse transcription of their RNAs (retrotransposition; a copy-and-paste mechanism). LTR-retrotransposons and LINEs (long interspersed elements) are major classes of

retrotransposons (Deininger and Batzer, 2002). LTR-retrotransposons have long terminal repeats (LTRs) consisting of several hundred base pairs that flank an internal coding region of the elements. LINEs, also called non-LTR-retrotransposons, encode a reverse transcriptase that recognizes the 3' end sequence of the LINE RNA during retrotransposition (Moran et al., 1996; Kajikawa and Okada, 2002). DNA transposons, also known as Class II TEs, encode a transposase flanked by short terminal inverted repeats (TIRs or IRs). The transposase recognizes TIRs and mobilizes the elements (transposition; a cut-and-paste mechanism). Meanwhile, many non-autonomous TEs, which do not encode enzymes for transposition, have been identified. The non-autonomous TEs have a specific sequence that can be recognized by enzymes encoded by partner autonomous elements, which enables them to be transposed (Fig. 1B). For example, non-autonomous elements that have LTRs and no coding region, called LARDs (large retrotransposon derivatives)
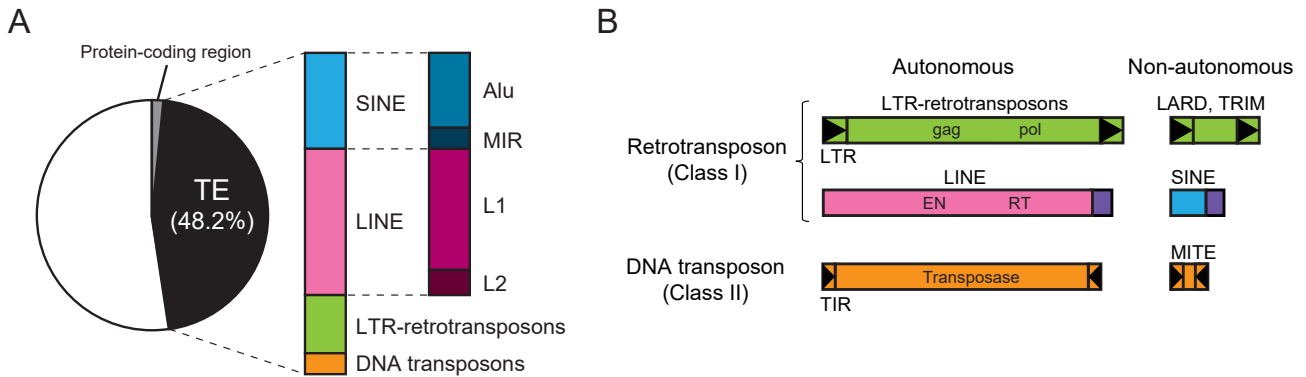
Fig. 1. Fraction of TEs in the human genome and classification of TEs. (A) Proportion of protein-coding sequences (gray), TEs (black), and other DNA (white) in the human genome. Major components of SINEs are *Alu* and MIR, whereas those of LINEs are L1 and L2. (B) Basic classifications of eukaryotic TEs. LTRs and TIRs are represented by boxes with triangles. The 3' end sequences shared between LINEs and SINEs are represented by purple boxes. EN, endonuclease; RT, reverse transcriptase.

or TRIMs (terminal-repeat retrotransposons in miniature), can replicate using enzymes encoded by corresponding autonomous elements (Havecker et al., 2004). SINEs (short interspersed elements) are frequently observed in the genomes of multicellular organisms (Okada, 1991) and share 3' end sequences with partner LINEs so that the 3' sequence of the SINE RNA can be recognized *in trans* by the reverse transcriptase encoded by the partner LINEs during retrotransposition (Ohshima et al., 1996; Kajikawa and Okada, 2002). In the case of DNA transposons, MITEs (miniature inverted-repeat transposable elements) are non-autonomous elements that possess TIRs with no coding region (Wessler et al., 1995). Thus, all TEs have been hierarchically classified based on their mobilization mechanism, domain structure and consensus sequences (Kojima, 2018, 2019).

In general, TEs move throughout the host genome without providing a beneficial effect and have been traditionally regarded as selfish DNA, genetic parasites, or a part of "junk DNA" (Orgel and Crick, 1980), with some exceptions in bacteria (MacHattie and Jackowski, 1977). Indeed, almost all TE insertion events are considered neutral mutations because random nucleotide substitutions accumulate in the TE sequences after transposition (Lunter et al., 2006; Cordaux and Batzer, 2009). As a negative effect, TE insertions can occasionally cause or become a cause of genetic defects (Hancks and Kazazian, 2016). In contrast, given that TEs occupy a large fraction of the genome, it was proposed long ago that they have a beneficial function such as gene regulation (Britten and Davidson, 1969, 1971). Studies have indeed revealed a number of functional elements derived from TEs.

This paper is an overview of TEs as a component of and a driving force for the diversification of genomes in animals. In addition, several cases of TE exaptation (Brosius and Gould, 1992), in which these elements acquire functions that have an advantageous effect on the host in mammals, are reviewed. Finally, I discuss the possibility that TE exaptations have frequently occurred and thus have contributed to the evolution of many genetic and morphological features in various animal lineages.

TRANSPOSABLE ELEMENTS AS A DETERMINANT OF GENOMIC CHARACTERISTICS

There are several hundred kinds of TE families in humans, and over 1,200 of their consensus sequences are available in Repbase (Bao et al., 2015; Kojima, 2018). A large proportion of TEs consist of Class I elements (retrotransposons), such as SINEs, LINEs and LTR-retrotransposons, in the human genome (Fig. 1A). Among them, LINEs make up the largest fraction, and L1, in particular, is one of the most active TEs in the human genome (Moran et al., 1996). Also, *Alu* is an active SINE family that can retrotranspose using enzymes encoded by L1 (Dewannieux et al., 2003), and no less than 10% of the human genome is composed of a million copies of *Alu* elements (International Human Genome Sequencing Consortium, 2001). Among the many kinds of human TEs, only some of the families, such as L1 and *Alu*, are currently active (Kojima, 2018), whereas the vast majority of TE families are no longer active and can be observed only through their ancient remnants in the human genome. The L1 family is thought to have been active over 100 million years ago, and such insertions have been detected among distantly related species that diverged around that time (Churakov et al., 2009; Nishihara et al., 2009). Almost no TEs that were inserted before the last common ancestor of amniotes (310 million years ago) have been detected in the human genome other than TE sequences that are evolutionarily conserved. This observation suggests the possibility that most non-coding sequences, other than TEs, are remnants of TEs that can no longer be detected

(Brosius, 2014).

During evolution, new TE families emerge and become transpositionally active. Many lineage-specific families of SINEs and LTR-retrotransposons have emerged during mammalian evolution, resulting in a large difference in TE composition among species. As shown in Fig. 2, over 100 SINE families have been identified in vertebrates to date (Ohshima and Okada, 2005; Nishihara and Okada, 2008; Vassetzky and Kramerov, 2013; Bao et al., 2015). For example, a number of CAN SINE (SINEC) sequences are observed in the dog genome, whereas the opossum harbors Opo-1, marsupial Mar1, and Mar3 (Fig. 2A). These SINE families have emerged independently, and their distribution is restricted in each clade at the order/family level in mammals (Fig. 2; Nishihara and Okada, 2008). Thus, although the overall prevalence of SINEs is similar among the human, dog and opossum genomes (>10% of each genome), the specific sequences vary considerably. In birds, only a few active SINE families are known in spite of the availability of nearly complete genome sequences for dozens of species. Bird genomes seem to harbor fewer kinds of TEs compared with other vertebrates, and TEs account for only 4–10% of avian genomes (Zhang et al., 2014). In other vertebrates, it is generally observed that multiple SINE families coexist in the genome (Fig. 2B), which may reflect the higher diversity of LINE families in these species (Nikaido et al., 2013). Thus, it is no exaggeration to say that TEs contribute to the genetic identity of the primary sequences of host genomes.

TRANSPOSABLE ELEMENTS, GENOMIC DIVERSITY AND GENOME SIZE IN ANIMALS

Genome sizes of mammals range from 2 to 3.5 Gbp, and ~30–50% of these genomes are composed of TEs. If the human genome (3 Gbp) contained no TEs, it would be half its current size. Notably, the proportion of TEs is roughly correlated with the genome size in mammals (Fig. 3A). This suggests that the main factor for the increase in genome size is amplification of retrotransposons, which is consistent with a report that changes in genome size can be explained by a balance between the dynamic gain and loss of DNA through TE expansion and large segmental deletions, respectively (Kapusta et al., 2017). More importantly, given the distribution of many kinds of SINE families within each clade (Fig. 2), they may have a large impact on both genome diversification and genome size.

Furthermore, among animals, the genome size and proportion of TEs are diverse, $\sim 10^8$–10^{10} bp and 0.1–62%, respectively. Correlation between the proportion of TEs and the host genome size is also widely observed (Fig. 3B), and TEs may have been the most important determinant of genome size in these species. However, it is important to note that in some species repeat content may be underestimated because of a lack of sufficient characterization and accurate manual curation of TEs (Platt et al., 2016). Positive correlations between TE fractions and genome sizes are also observed in other eukaryotes (Kidwell, 2002; Tenaillon et al., 2010; Elliott and Gregory, 2015). An interesting recent finding is that species-specific amplification of SINEs caused genome expansions in larvaceans, a subgroup of tunicates (Naville et al., 2019). Thus, TEs have largely contributed to an increase in both the size and diversity of genomes during animal evolution.

PLEIOTROPIC IMPACTS OF TE CO-OPTIONS ON GENE EVOLUTION

In general, TEs have no essential function for host survival, and most cases of their insertion are considered to be neutral mutations (Lunter et al., 2006). However, some TE insertion events do provide either negative or positive effects on genetic function (Fig. 4A). For example, although transposition of TEs is repressed by epigenetic mechanisms in most somatic cells, accidental *de novo* TE insertions (mostly *Alu* and L1) within protein-coding exons can cause disorders, such as cancers (Helman et al., 2014; Hancks and Kazazian, 2016). Even when they are located outside exons, TEs can cause unequal homologous recombination between copies, occasionally resulting in genomic deletions associated with human diseases (Deininger and Batzer, 1999; Sen et al., 2006). In contrast, some TEs have been co-opted/exapted and utilized as exons. For instance, >100 copies of *Alu* and MIR (mammalian-wide interspersed repeat) elements have been identified as exonized in the human genome (Fig. 4A(a); Lev-Maor et al., 2003; Krull et al., 2005, 2007). *Alu* and other TEs in 3' UTRs of genes can also provide a polyadenylation signal, whereas those located in introns can result in truncated RNAs (Fig. 4A(b); Lee et al., 2008; Chen et al., 2009).

Epigenetic regulation of TE sequences can affect the expression level of neighboring genes (Fig. 4A(c)). For example, whereas rodent B1 SINEs are generally hypermethylated, those proximal to promoters are hypomethylated (Ichiyanagi et al., 2011). Given the enrichment of B1 in the promoters of genes with testis-specific expression, these SINEs may be involved in the regulation of their associated genes (Ichiyanagi et al., 2011). In addition, some TEs act as promoters for the transcription of proximal exons and have contributed to transcriptional changes in the host (Faulkner et al., 2009; Emera and Wagner, 2012). Interestingly, L1 possesses an antisense promoter activity in the 5' UTR (in human) or ORF1 (in mouse) and contributes to the transcription of many genes (Speek, 2001; Li et al., 2014). Furthermore, the human L1 5' UTR itself is also translated, and thousands
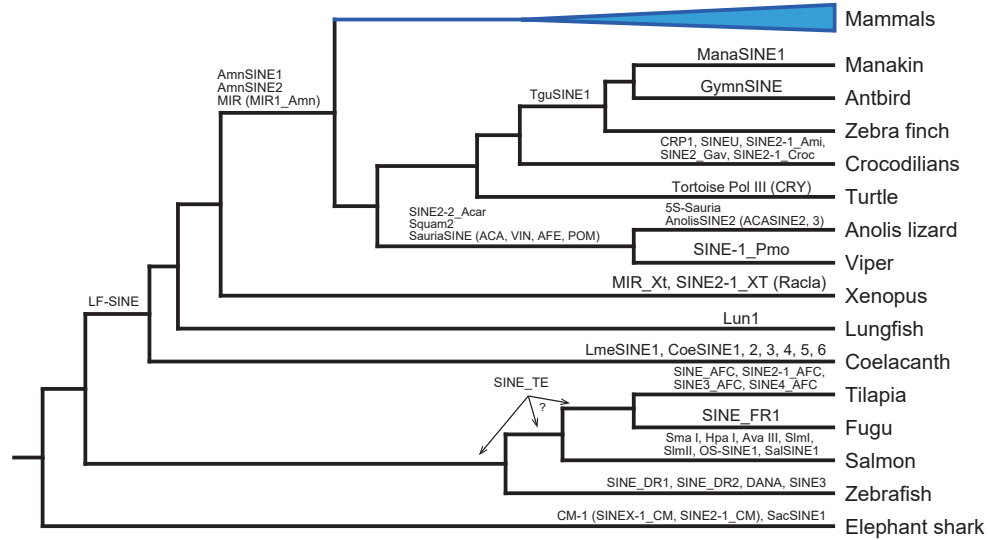
A
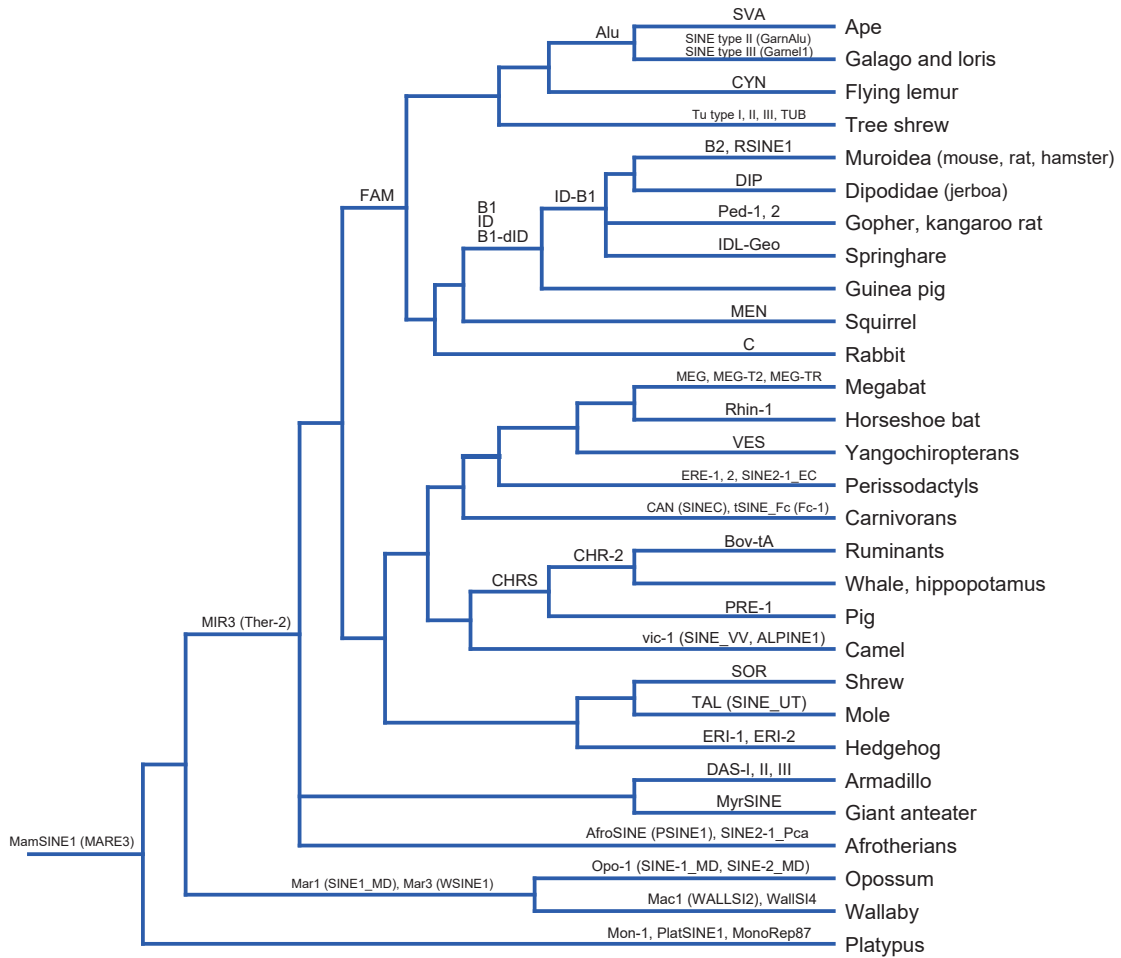
B

Fig. 2.   Distribution of SINE families along with the phylogeny of mammals (A) and non-mammalian vertebrates (B).   Synonyms are shown in parentheses.
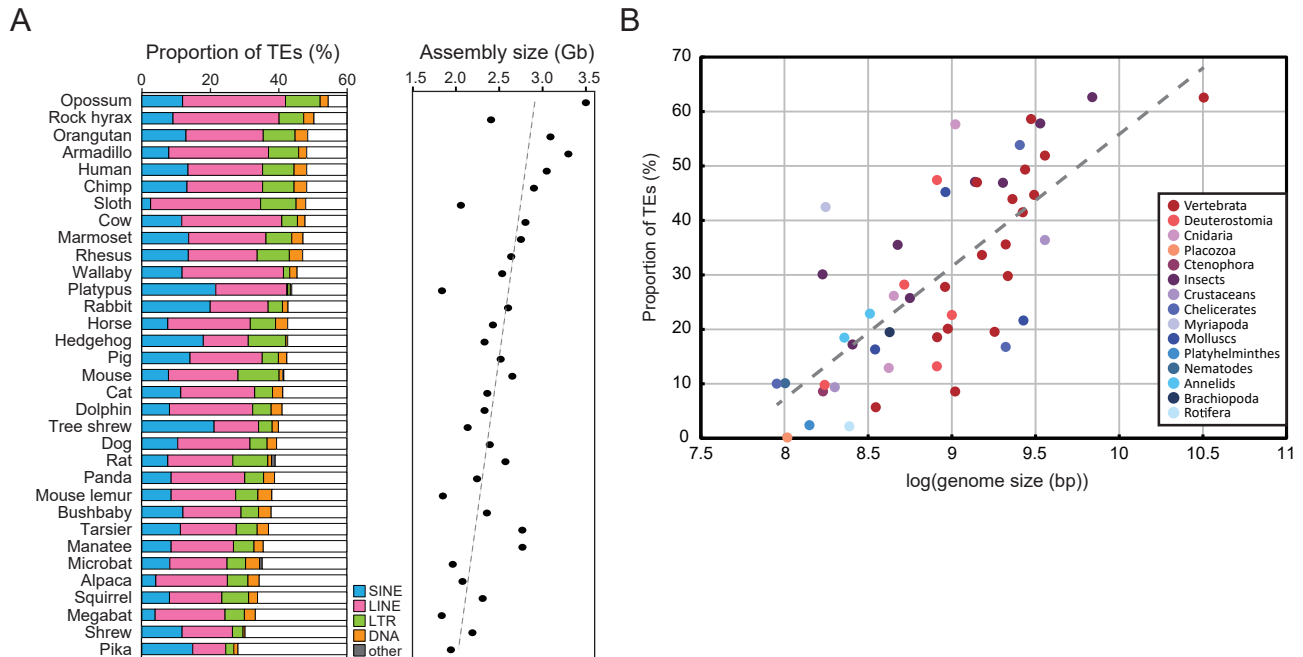
Fig. 3. TEs and genome sizes in animals. (A) The proportion of each class of TEs in mammalian genomes and the genome assembly sizes. The species are arranged in descending order of their TE content. TE content data were obtained from the RepeatMasker website (http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html). (B) Correlation between genome size and the proportion of TEs in animals. The data were retrieved from the original reports of genome sequencing projects. Vertebrata: human, mouse, opossum, platypus, chicken, alligator, painted turtle, lizard, frog, axolotl, coelacanth, *Takifugu*, Nile tilapia, Atlantic salmon, zebrafish, spotted gar and elephant shark; Deuterostomia: lamprey, lancelet, tunicate, sea urchin and acorn worm; Cnidaria: coral, *Hydra* and starlet sea anemone; Placozoa: *Trichoplax*; Ctenophora: comb jelly; Insects: mountain pine beetle, silkworm, honey bee, fruit fly, locust, dampwood termite, *Blattella germanica* and American cockroach; Crustaceans: amphipod and water flea; Chelicerates: tick, spider mite and velvet spider; Myriapoda: centipede; Molluscs: owl limpet, snail and octopus; Platyhelminthes: tapeworm; Nematodes: *Caenorhabditis elegans*; Annelids: freshwater leech and marine polychaete; Brachiopoda: brachiopod; Rotifera: rotifer.

of such potential ORFs are present in the human genome (Denli et al., 2015). Although the importance of their biological functions remains unclear, it is likely that TEs such as L1 have large effects on the transcriptome and proteome in humans. Remarkably, L1 is retrotranspositionally active in neuronal cells, unlike in other somatic cells (Muotri et al., 2005; Faulkner and Garcia-Perez, 2017). The genome mosaicism caused by L1 retrotranspositions may provide differences in somatic cell transcriptomes or proteomes and affect neuronal plasticity in humans (Singer et al., 2010).

Some ORFs encoded by TEs such as LTR-retrotransposons have been exapted to serve as protein-coding genes (domestication) that are responsible for mammalian morphogenesis. *Peg10* (*Sirh1*) and *Peg11* (*Sirh2/Rtl1*) are both paternally expressed imprinted genes derived from LTR-retrotransposons domesticated in a common ancestor of Theria and Eutheria, and were revealed to be responsible for the formation and functional maintenance of the placenta, respectively (Ono et al., 2006; Sekita et al., 2008). Another important example is the mammalian syncytin genes derived from endogenous retroviruses (ERVs). ERVs are a kind of TE and are considered to have originated evolutionarily from LTR-

retrotransposons (Xiong and Eickbush, 1990). Human syncytin genes are involved in trophoblastic cell fusion during placental morphogenesis (Mi et al., 2000; Malassiné et al., 2007). Remarkably, syncytin genes found in various mammals, such as mice and cows, arose independently in their respective lineages via domestication of ERVs. Thus, LTR-retrotransposons have contributed to both common and lineage-specific features in placental formation. In contrast, Arc proteins derived from *gag* of LTR-retrotransposons were revealed to form virus-like capsids in mammalian neurons (Pastuzyn et al., 2018). As these proteins can mediate intercellular transfer of *Arc* mRNA, it is possible that *Arc* is involved in the intercellular communication of genetic information via mRNA transfer between neurons in the brain (Pastuzyn et al., 2018). In addition, in other animals, LTR-retrotransposon sequences can play a variety of important biological roles (for details, see Horie, 2019).

CIS-REGULATORY EVOLUTION ACCELERATED BY TRANSPOSABLE ELEMENTS

TEs have had a great impact on *cis*-regulatory evolution. In this section, I review studies about TE contribu-
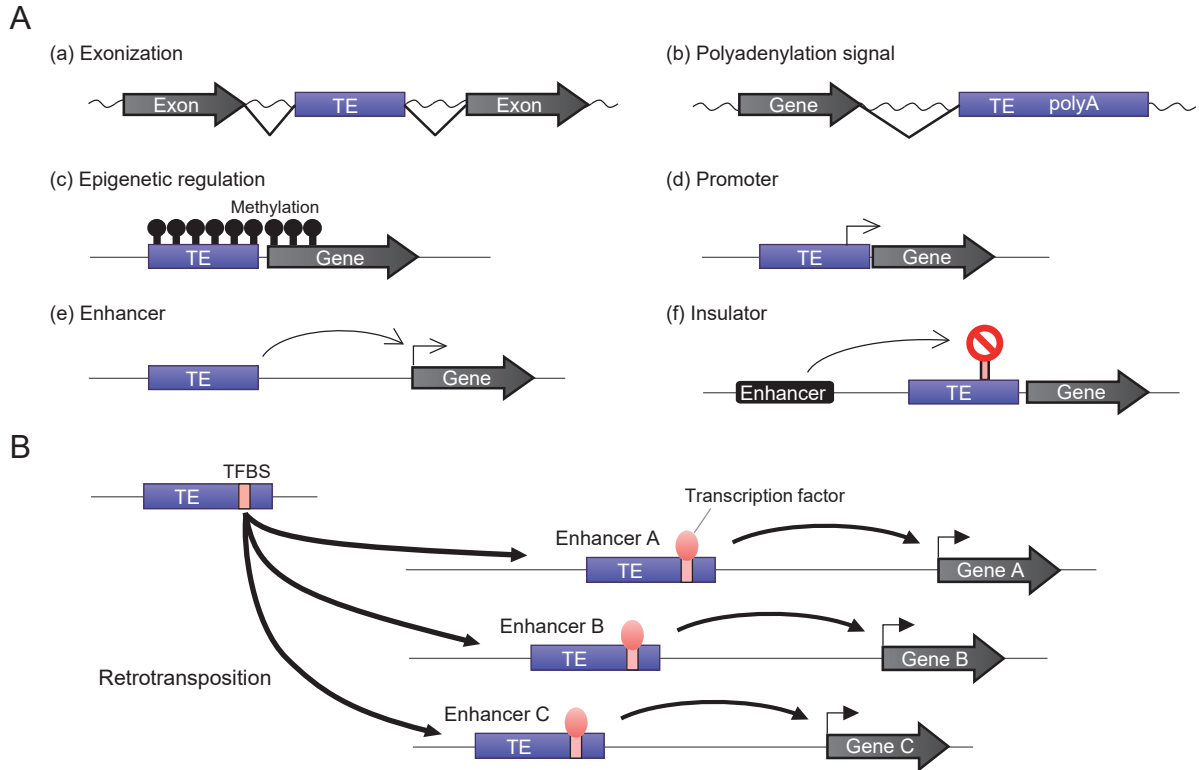
A

B

Fig. 4. The evolutionary impact of TEs on genes and *cis*-regulatory elements. (A) Some TEs provide protein-coding exons (a), poly-adenylation signals (b) or epigenetic regulation for gene expression (c). As *cis*-regulatory elements, some TEs can serve as promoters (d), enhancers (e) or insulators (f). (B) Retrotransposons can increase and spread potential transcription factor binding sites (TFBSs) that can be source sequences of *cis*-regulatory elements.

tions in mammals to 1) the evolution of a novel enhancer involved in morphological innovation, 2) an increase in binding sites of CTCF, a chromatin architectural protein, through retrotransposition leading to an expansion of lineage-specific chromatin boundaries, 3) an increase in binding sites for transcription factors leading to an expansion of potential *cis*-regulatory elements, and 4) early embryonic development.

After TEs become integrated into the host genome, some may acquire biological functions (exaptation or cooption), such as promoters and enhancers, and become involved in the regulation of proximal genes (Fig. 4A(d, e)). There are hundreds of thousands of evolutionarily conserved non-coding elements (CNEs), which are putative functional elements (Bejerano et al., 2004; Siepel et al., 2005). Some CNEs overlap with TEs (Bejerano et al., 2006; Nishihara et al., 2006; Gentles et al., 2007; Lowe et al., 2007; Mikkelsen et al., 2007), suggesting that some TEs are involved in *cis*-regulatory alterations and possibly in morphological changes during evolution. Indeed, several copies of TEs serve as enhancers or promoters (Bejerano et al., 2006; Santangelo et al., 2007; Sasaki et al., 2008; Franchini et al., 2011; Emera and Wagner, 2012). For example, >100 elements of AmnSINE1, an anciently inactivated SINE family in amniotes, were found to

overlap with CNEs (Nishihara et al., 2006; Sasaki et al., 2008; Hirakawa et al., 2009), and some of them serve as distal enhancers for genes involved in mammalian morphogenesis. One CNE derived from AmnSINE1 is an enhancer for *fgf8* expression in the diencephalon (Sasaki et al., 2008; Nakanishi et al., 2012), while another acts as a distal enhancer of *satb2* in the neocortex and may be involved in the formation of the corpus callosum, a eutherian-specific brain structure (Tashiro et al., 2011). Furthermore, a CNE derived from AmnSINE1 and two DNA transposons (X6b_DNA and MER117) act together as a distal enhancer of *wnt5a* during palatogenesis (Fig. 5A; Nishihara et al., 2016a). Wnt5a is responsible for complete closure of the secondary palate, which is a mammal-specific morphological feature (He et al., 2008; Angielczyk, 2009). As a closed secondary palate is required to allow neonates to suckle successfully (Erkan et al., 2013), acquisition of this feature is considered one of the most innovative morphological changes in a common ancestor of mammals. In addition to AmnSINE1, other TEs have also been broadly involved in *cis*-regulatory evolution in vertebrates, such as a neuronal enhancer derived from LF-SINEs (Bejerano et al., 2006), two mammalian neuronal enhancers that independently evolved from distinct TEs (Franchini et al., 2011), and an MER39 element that
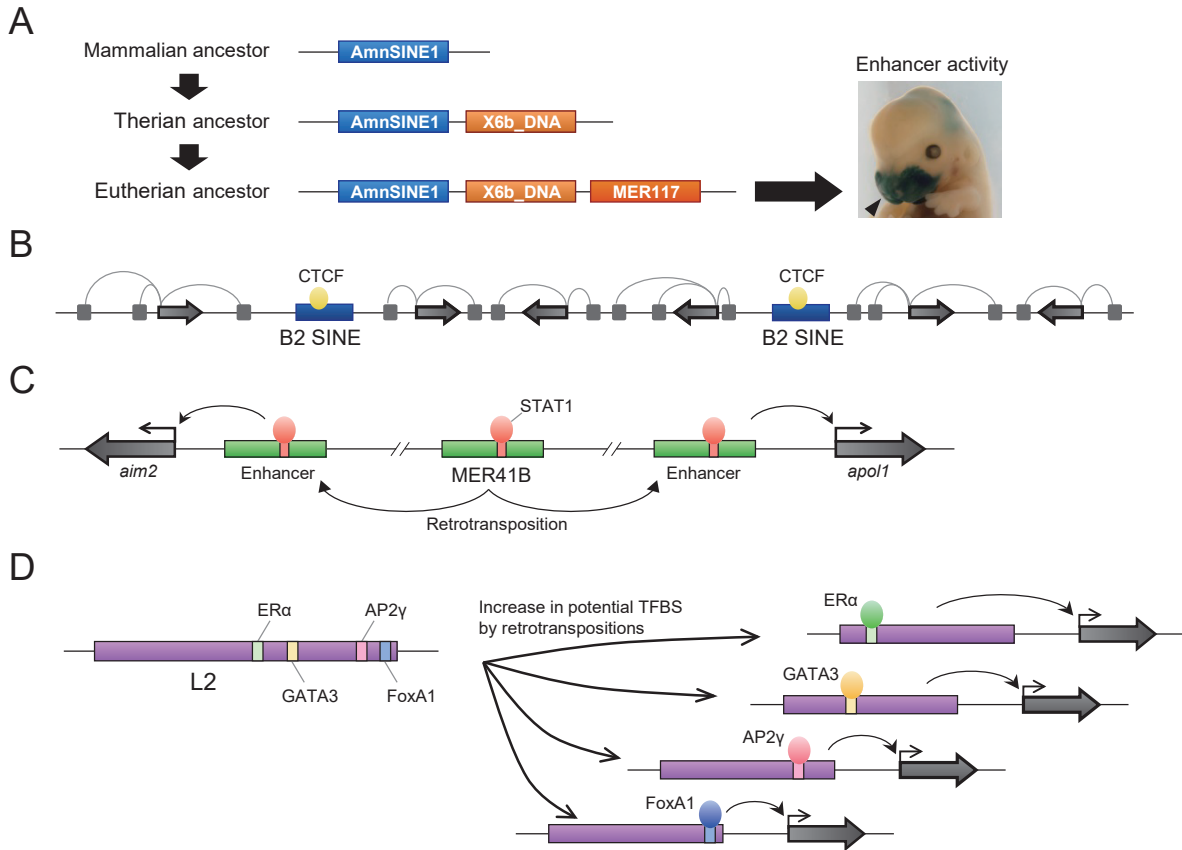
Fig. 5. Four examples of TE exaptation. (A) Three TEs, AmnSINE1, X6b_DNA and MER117, have transposed stepwise during mammalian evolution and coordinately act as a distal enhancer for *wnt5a* expression in the frontonasal region (arrowhead) and the secondary palate in mice. (B) Rodent B2 SINEs contain CTCF binding sites, which may contribute to TAD formation. (C) MER41Bs contain a binding site for STAT1 and have contributed to *cis*-regulatory evolution involved in innate immunity through transcriptional control of related genes such as *aim2* and *apol1*. (D) L2s include binding sites for four transcription factors (ERα, FoxA1, GATA3 and AP2γ) and have provided many regulatory elements during mammary gland evolution.

acts as a prolactin gene promoter (Emera and Wagner, 2012). Thus, many exaptation events have occurred during evolution, and alterations to gene expression patterns by TEs may have contributed to morphogenetic evolution (Okada et al., 2010).

Some TEs act as insulators in mammals (Fig. 4A(f); Lynch et al., 2011; Wang et al., 2015). For example, murine B2 SINEs possess a CTCF binding site in their consensus sequence (Lunyak et al., 2007; Schmidt et al., 2012). Because CTCF binding sites act as insulators that show an enhancer-blocking property, this finding suggests that B2 has increased the number of potential insulator elements in the genome (Fig. 5B). CTCF proteins are highly enriched at the boundaries of topologically associating domains (TADs) and are involved in the demarcation of transcriptionally active/inactive domains (Dixon et al., 2012). B2 is also enriched in TAD boundary regions (Dixon et al., 2012), and an increase in CTCF binding sites as a result of B2 retrotransposition during evolution might have influenced genome organization in the nucleus. Because the distribution of the B2 family is restricted to a subset of rodents (Fig. 2A), the spread of the binding sites by this SINE occurred only in this group. Notably, other mammalian SINEs such as CAN (SINEC) in carnivores and Mar1 in marsupials also possess CTCF binding sites (Schmidt et al., 2012). Not only SINEs but also ERVs are responsible for the demarcation of TADs. For example, *de novo* insertion of HERV-H elements introduces new TAD boundaries, and deletion of HERV-H elements causes elimination of TAD boundaries in human pluripotent stem cells (Zhang et al., 2019). Thus, a wide variety of TEs may be involved in the lineage-specific dynamics of higher-order chromatin structure and gene regulatory networks via expansion of potential chromatin boundaries.

Furthermore, some TE families possess binding sites for transcription factors in their consensus sequences (e.g., Bourque et al., 2008; Kunarso et al., 2010; Chuong et al., 2016; Nishihara, 2019). Various LTR-retrotransposons have contributed to an increase in the number of binding sites for various transcription factors such as Oct4 (Jacques et al., 2013; Sundaram et al., 2017), Sox2

(Bourque et al., 2008; Jacques et al., 2013; Sundaram et al., 2017) and Elf5 (Chuong et al., 2013). For example, MER41B is an LTR-retrotransposon family and possesses a STAT1 binding site in its consensus sequence (Chuong et al., 2016). Hundreds of MER41B copies are bound by STAT1 in human cells, and some of these indeed serve as enhancers of proximal interferon-stimulated genes, such as *aim2* and *apol1* (Fig. 5C). MER41 elements, which have expanded independently in multiple mammalian clades, may have contributed to lineage-specific *cis*-regulatory changes involved in innate immunity (Chuong et al., 2016).

These findings support the Britten–Davidson model in which TEs can generate many source sequences for regulatory elements via transposition (Fig. 4B; Britten and Davidson, 1969, 1971; Sundaram and Wang, 2018). Furthermore, it has been proposed that TEs have altered gene regulatory programs, which could lead to the modification of developmental systems (Britten and Davidson, 1971). This proposal can be verified by identifying TEs that possess binding site(s) for a master regulator of a gene regulatory network responsible for morphogenesis, because, according to this model, TEs could increase the number of developmental enhancers having the same or similar functions, leading to an expansion of downstream genes of the regulator (Fig. 4B). For example, initial development of the mammary gland, a morphological feature characteristic of mammals, requires binding of ERα and related pioneer factors (FoxA1, GATA3 and AP2γ) to many enhancers and promoters to induce a dramatic change in the expression of downstream genes (Manavathi et al., 2014). Dozens of kinds of human TEs possess binding motifs for these four transcription factors, and indeed thousands of TE copies are bound by these factors and exhibit enhancer-specific chromatin states (Nishihara, 2019). These TEs, such as MIR and L2, may have had a substantial impact on *cis*-regulatory evolution associated with mammary gland development. Notably, L2 elements, which are ancient mammalian LINEs, most likely increased and spread the binding sites of all four factors and thus contributed extensively to the production of potential enhancers (Fig. 5D). In a similar way, it is possible that a large number of TEs have been involved in an expansion of gene regulatory networks leading to morphological innovation.

Recent studies have revealed that TEs are also involved in early embryogenesis in mammals. One type of ERV (HERVs) is highly expressed in human embryonic stem (ES) cells (Wang et al., 2014). Long non-coding RNAs derived from HERVs determine ES cell identity (Lu et al., 2014) and regulate pluripotency in early embryonic development by serving as a key component of the regulatory network (Durruthy-Durruthy et al., 2016). Also, HERVs, as well as other evolutionarily young TEs, serve as enhancers in human ES cells and significantly con-tribute to genome activation during human early embryogenesis (Lu et al., 2014; Pontis et al., 2019). In contrast, transcription of L1 is also activated in the early mouse embryo. Silencing of L1 leads to a decrease in chromatin accessibility and to developmental delay, indicating that activation of L1 regulates chromatin opening and is essential for normal embryonic development (Jachowicz et al., 2017). In mouse gonocytes of a certain stage, TEs such as L1 are enriched in transiently accessible genomic regions, and transcriptional upregulation of evolutionarily young L1 elements occurs in these regions (Yamanaka et al., 2019). This result is suggestive of a role for the TEs in chromatin accessibility also in the germline. Thus, TEs such as ERVs and L1 are deeply involved in the early stage of development in mammals.

REPETITIVE SEQUENCES AND NUCLEAR ARCHITECTURE

The genomic distribution of TEs may be associated with nuclear architecture. LINEs are enriched in AT-rich and gene-poor regions in the genome and in heterochromatin, which corresponds to G-bands of chromosomes (Korenberg and Rykowski, 1988; International Human Genome Sequencing Consortium, 2001). In nuclei, LINEs are enriched in constitutive lamina-associated domains (cLADs) and the nuclear periphery wherein gene expression is largely repressed (Ichiyanagi et al., 2011; Meuleman et al., 2013; Solovei et al., 2016). In contrast, SINEs are enriched in gene-rich euchromatin regions, which correspond to R-bands of chromosomes. In nuclei, SINEs are depleted in cLADs but are observed in the nuclear interior (Ichiyanagi et al., 2011; Meuleman et al., 2013; Solovei et al., 2016). It remains unclear what molecular mechanism causes this mutually exclusive distribution of LINEs and SINEs in the genome. It may be a result of natural selection if SINEs have a regulatory function related to gene expression (Ichiyanagi, 2013).

In addition to TEs, other repetitive sequences make up a portion of the human genome, such as satellites, minisatellites and microsatellites (short tandem repeats; STRs). RepeatMasker annotation indicates that simple repeats occupy 1.5% of the human genome. Most of the simple repeats are found in non-coding sequences, and some of them can be used as genetic markers for population genetics because they represent length polymorphisms among individuals (Rosenberg et al., 2002). Recent studies have revealed that a portion of STRs act as regulatory elements to control gene expression in humans (Gymrek et al., 2016; Hannan, 2018). Remarkably, some satellite repeats are involved in the formation of TADs and larger domains. For example, the inactive X chromosome in mice can be separated into two large mega-domains, and the DXZ4 macrosatellite region located at the boundary of these domains is essential for their formation

(Giorgetti et al., 2016). Also, dozens of STRs are associated with diseases, and the disease-associated STRs are enriched at TAD boundaries. Patients with fragile X syndrome with the mutation-length STR exhibit loss of CTCF occupancy, disruption of the TAD boundary, and a reduced expression of the associated gene FMR1 (Sun et al., 2018). Thus, it is likely that not only TEs but also many satellite repeats have contributed to the evolution of the *cis*-regulatory network and three-dimensional organization of the genome in the mammalian cell nucleus.

One interesting finding under this topic is that in nocturnal mammals, heterochromatin localizes to the central region of the nucleus in rod cells and functions like a lens to send light efficiently to the outer segments, whereas it is distributed mainly in the periphery of the nucleus in rod cells of diurnal mammals (Solovei et al., 2009). In mice, constitutive heterochromatin, which can be marked by major satellite repeats (MSRs), localizes to the central region of the nucleus in rod cells (Solovei et al., 2009). The rod cells of owl monkeys, the only genus of nocturnal/cathemeral simian primates, show a spherical heterochromatin block in the central region of the nucleus, and a primary component of the heterochromatin region is the OwlRep, a megasatellite DNA that has expanded specifically in the owl monkey lineage (Koga et al., 2017; Nishihara et al., 2018). Thus, because megasatellites as a whole can be a major component of heterochromatin, lineage-specific expansion of satellite repeats might have had an impact on the nuclear architecture associated with nocturnal adaptation in this lineage.

TRANSPOSABLE ELEMENTS AS ACCELERATORS OF FUNCTIONAL COMPLEXITY DURING ANIMAL EVOLUTION

Together, TEs exhibit a large diversity in their transpositional machinery, structure and sequences (Fig. 1), and >1,200 consensus sequences of TEs have been reported at the subfamily level in humans. Although TEs have generally been regarded as junk/selfish DNA, it has been reported primarily in mammals that a subset of TEs has pivotal or assistive roles in gene regulation leading to morphogenesis (Figs. 4 and 5). However, it remains largely unknown how many of the total 4.5 million copies of TEs in the human genome are involved in gene regulation and morphogenesis. Apart from mammals, some TEs also have positive or negative effects on gene expression and contribute to morphological changes in non-mammalian vertebrates (Santos et al., 2014) and invertebrates (van't Hof et al., 2016). It is expected that there are, in a variety of animals, diverse, lineage-specific exapted TEs that have contributed not only to genome size (Fig. 3) but also to morphological innovation.

Because SINEs exhibit lineage-specific distribution at the mammalian order/family level (Fig. 2), it is reasonable

that some of them have provided lineage-specific functions. One of the remarkable features of SINEs is that several SINE families in different species share a highly similar central sequence and therefore can be grouped together as a SINE superfamily (Ogiwara et al., 2002), in contrast to the more typical SINEs that have different sequences in their central regions. SINE superfamilies such as CORE-SINEs (Gilbert and Labuda, 1999, 2000; Munemasa et al., 2008), V-SINEs (Ogiwara et al., 2002), DeuSINEs (Nishihara et al., 2006) and MetaSINEs (Nishihara et al., 2016b) are widely distributed in animals, but the functions of the central sequences of the SINE superfamilies remain unknown. AmnSINE1, some copies of which have been exapted in mammals as described above (Fig. 5A), is a member of the DeuSINE superfamily (Nishihara et al., 2006), and the central sequences of ancient MIRs, a member of the CORE-SINE superfamily, were a source of binding sites for multiple transcription factors (Fig. 4B; Bourque et al., 2008; Rohrmoser et al., 2018; Nishihara, 2019). These findings suggest the possibility that the central sequences of the SINE superfamily members have provided functional source sequences to the genome through retrotransposition and thus may have been frequently exapted in other animals.

To fully understand the TE exaptation landscape in animals, the following three approaches will be required. First, because various types of TEs have been reported and hierarchically classified in mammals (Kojima, 2018, 2019), it can be easily expected that in other animals, hundreds of kinds of TEs may also constitute a large fraction of the genome. An accurate annotation of TEs will be necessary to understand their variety and coverage in animal genomes (Platt et al., 2016), although such studies lag behind the recent exponential increase in whole-genome sequence assemblies. Second, genome-wide identification of functional elements using epigenetic studies such as ChIP-seq and Hi-C and epigenetic editing technologies (Hilton et al., 2015; Jachowicz et al., 2017), as well as genome engineering techniques such as the CRISPR-Cas9 system (Ran et al., 2013), will be applied widely to various non-model organisms in the future. ChIP-seq analyses enabled the identification of a number of TE-derived functional elements that have led to large-scale alteration of *cis*-regulatory networks (Fig. 5B–5D). Notably, new methods have recently been developed for epigenetic and chromatin interaction analyses in a single cell or a small number of cells in mammals (Nagano et al., 2013; Zhang et al., 2016; Harada et al., 2019). Application of such new technologies for identification of exapted TEs should unveil many cases of lineage-specific and tissue-specific acquisition/modification of gene regulatory networks driven by TE expansion. Third, multidisciplinary studies, including computational genomics, chromosomal analyses, cell biology technology and developmental studies, should be

a powerful approach for revealing the multifaceted contributions of a number of TEs to the evolution of a wide variety of animals. In particular, studies about nuclear organization and chromosome territory as well as cell biological approaches may reveal a possible crosstalk between the organization of TEs and gene regulation in the nucleus. Thus, an omic atlas of TE exaptations, as revealed via such approaches, will provide a new perspective on the molecular mechanisms for the acquisition and evolution of functional elements that can lead to the emergence of morphological innovations.

REFERENCES

Angielczyk, K. D. (2009) *Dimetrodon* is not a dinosaur: using tree thinking to understand the ancient relatives of mammals and their evolution. Evo. Edu. Outreach **2**, 257–271.

Bao, W., Kojima, K. K., and Kohany, O. (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob. DNA **6**, 11.

Bejerano, G., Lowe, C. B., Ahituv, N., King, B., Siepel, A., Salama, S. R., Rubin, E. M., Kent, W. J., and Haussler, D. (2006) A distal enhancer and an ultraconserved exon are derived from a novel retroposon. Nature **441**, 87–90.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., and Haussler, D. (2004) Ultraconserved elements in the human genome. Science **304**, 1321–1325.

Bourque, G., Leong, B., Vega, V. B., Chen, X., Lee, Y. L., Srinivasan, K. G., Chew, J.-L., Ruan, Y., Wei, C.-L., Ng, H. H., et al. (2008) Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. **18**, 1752–1762.

Britten, R. J., and Davidson, E. H. (1969) Gene regulation for higher cells: a theory. Science **165**, 349–357.

Britten, R. J., and Davidson, E. H. (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Q. Rev. Biol. **46**, 111–138.

Brosius, J. (2014) The persistent contributions of RNA to eukaryotic gen(om)e architecture and cellular function. Cold Spring Harb. Perspect. Biol. **6**, a016089.

Brosius, J., and Gould, S. J. (1992) On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". Proc. Natl. Acad. Sci. USA **89**, 10706–10710.

Chen, C., Ara, T., and Gautheret, D. (2009) Using *Alu* elements as polyadenylation sites: a case of retroposon exaptation. Mol. Biol. Evol. **26**, 327–334.

Chuong, E. B., Elde, N. C., and Feschotte, C. (2016) Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science **351**, 1083–1087.

Chuong, E. B., Rumi, M. A. K., Soares, M. J., and Baker, J. C. (2013) Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat. Genet. **45**, 325–329.

Churakov, G., Kriegs, J. O., Baertsch, R., Zemann, A., Brosius, J., and Schmitz, J. (2009) Mosaic retroposon insertion patterns in placental mammals. Genome Res. **19**, 868–875.

Cordaux, R., and Batzer, M. A. (2009) The impact of retrotransposons on human genome evolution. Nat. Rev. Genet. **10**, 691–703.

Deininger, P. L., and Batzer, M. A. (1999) Alu repeats and human disease. Mol. Genet. Metab. **67**, 183–193.

Deininger, P. L., and Batzer, M. A. (2002) Mammalian retroelements. Genome Res. **12**, 1455–1465.

Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C. N., Diedrich, J. K., Aslanian, A., Ma, J., Moresco, J. J., et al. (2015) Primate-specific ORF0 contributes to retrotransposon-mediated diversity. Cell **163**, 583–593.

Dewannieux, M., Esnault, C., and Heidmann, T. (2003) LINE-mediated retrotransposition of marked Alu sequences. Nat. Genet. **35**, 41–48.

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485**, 376–380.

Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E. J., Davila, J., Mall, M., Wong, W. H., Wysocka, J., et al. (2016) The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. Nat. Genet. **48**, 44–52.

Elliott, T. A., and Gregory, T. R. (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. Philos. Trans. R. Soc. Lond. B Biol. Sci. **370**, 20140331.

Emera, D., and Wagner, G. P. (2012) Transformation of a transposon into a derived prolactin promoter with function during human pregnancy. Proc. Natl. Acad. Sci. USA **109**, 11246–11251.

Erkan, M., Karaçay, S., Atay, A., and Günay, Y. (2013) A modified feeding plate for a newborn with cleft palate. Cleft Palate Craniofac. J. **50**, 109–112.

Faulkner, G. J., and Garcia-Perez, J. L. (2017) L1 mosaicism in mammals: extent, effects, and evolution. Trends Genet. **33**, 802–816.

Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., Schroder, K., Cloonan, N., Steptoe, A. L., Lassmann, T., et al. (2009) The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. **41**, 563–571.

Franchini, L. F., López-Leal, R., Nasif, S., Beati, P., Gelman, D. M., Low, M. J., de Souza, F. J. S., and Rubinstein, M. (2011) Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. Proc. Natl. Acad. Sci. USA **108**, 15270–15275.

Gentles, A. J., Wakefield, M. J., Kohany, O., Gu, W., Batzer, M. A., Pollock, D. D., and Jurka, J. (2007) Evolutionary dynamics of transposable elements in the short-tailed opossum *Monodelphis domestica*. Genome Res. **17**, 992–1004.

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

Gilbert, N., and Labuda, D. (2000) Evolutionary inventions and continuity of CORE-SINEs in mammals. J. Mol. Biol. **298**, 365–377.

Giorgetti, L., Lajoie, B. R., Carter, A. C., Attia, M., Zhan, Y., Xu, J., Chen, C. J., Kaplan, N., Chang, H. Y., Heard, E., et al. (2016) Structural organization of the inactive X chromosome in the mouse. Nature **535**, 575–579.

Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M. J., Price, A. L., Pritchard, J. K., Sharp, A. J., et al. (2016) Abundant contribution of short tandem repeats to gene expression variation in humans. Nat.

Genet. **48**, 22–29.

Hancks, D. C., and Kazazian, H. H., Jr. (2016) Roles for retrotransposon insertions in human disease. Mob. DNA **7**, 9.

Hannan, A. J. (2018) Tandem repeats mediating genetic plasticity in health and disease. Nat. Rev. Genet. **19**, 286–298.

Harada, A., Maehara, K., Handa, T., Arimura, Y., Nogami, J., Hayashi-Takanaka, Y., Shirahige, K., Kurumizaka, H., Kimura, H., and Ohkawa, Y. (2019) A chromatin integration labelling method enables epigenomic profiling with lower input. Nat. Cell Biol. **21**, 287–296.

Havecker, E. R., Gao, X., and Voytas, D. F. (2004) The diversity of LTR retrotransposons. Genome Biol. **5**, 225.

He, F., Xiong, W., Yu, X., Espinoza-Lewis, R., Liu, C., Gu, S., Nishita, M., Suzuki, K., Yamada, G., Minami, Y., et al. (2008) Wnt5a regulates directional cell migration and cell proliferation via Ror2-mediated noncanonical pathway in mammalian palate development. Development **135**, 3871–3879.

Helman, E., Lawrence, M. S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M. (2014) Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res. **24**, 1053–1063.

Hilton, I. B., D'Ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. **33**, 510–517.

Hirakawa, M., Nishihara, H., Kanehisa, M., and Okada, N. (2009) Characterization and evolutionary landscape of AmnSINE1 in Amniota genomes. Gene **441**, 100–110.

Horie, M. (2019) Interactions among eukaryotes, retrotransposons and riboviruses: endogenous riboviral elements in eukaryotic genomes. Genes Genet. Syst. this issue.

Ichiyanagi, K. (2013) Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. Genes Genet. Syst. **88**, 19–29.

Ichiyanagi, K., Li, Y., Watanabe, T., Ichiyanagi, T., Fukuda, K., Kitayama, J., Yamamoto, Y., Kuramochi-Miyagawa, S., Nakano, T., Yabuta, Y., et al. (2011) Locus- and domain-dependent control of DNA methylation at mouse B1 retrotransposons during male germ cell development. Genome Res. **21**, 2058–2066.

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature **409**, 860–921.

Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J., and Torres-Padilla, M.-E. (2017) LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat. Genet. **49**, 1502–1510.

Jacques, P.-É., Jeyakani, J., and Bourque, G. (2013) The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. **9**, e1003504.

Kajikawa, M., and Okada, N. (2002) LINEs mobilize SINEs in the eel through a shared 3' sequence. Cell **111**, 433–444.

Kapusta, A., Suh, A., and Feschotte, C. (2017) Dynamics of genome size evolution in birds and mammals. Proc. Natl. Acad. Sci. USA **114**, E1460–E1469.

Kazazian, H. H., Jr. (2004) Mobile elements: drivers of genome evolution. Science **303**, 1626–1632.

Kidwell, M. G. (2002) Transposable elements and the evolution of genome size in eukaryotes. Genetica **115**, 49–63.

Koga, A., Tanabe, H., Hirai, Y., Imai, H., Imamura, M., Oishi, T., Stanyon, R., and Hirai, H. (2017) Co-opted megasatellite DNA drives evolution of secondary night vision in Azara's owl monkey. Genome Biol. Evol. **9**, 1963–1970.

Kojima, K. K. (2018) Human transposable elements in Repbase:

genomic footprints from fish to humans. Mob. DNA **9**, 2.

Kojima, K. K. (2019) Structural and sequence diversity of eukaryotic transposable elements. Genes Genet. Syst. this issue.

Korenberg, J. R., and Rykowski, M. C. (1988) Human genome organization: Alu, LINES, and the molecular structure of metaphase chromosome bands. Cell **53**, 391–400.

Krull, M., Brosius, J., and Schmitz, J. (2005) Alu-SINE exonization: en route to protein-coding function. Mol. Biol. Evol. **22**, 1702–1711.

Krull, M., Petrusma, M., Makalowski, W., Brosius, J., and Schmitz, J. (2007) Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). Genome Res. **17**, 1139–1145.

Kunarso, G., Chia, N.-Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y.-S., Ng, H.-H., and Bourque, G. (2010) Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat. Genet. **42**, 631–634.

Lee, J. Y., Ji, Z., and Tian, B. (2008) Phylogenetic analysis of mRNA polyadenylation sites reveals a role of transposable elements in evolution of the 3'-end of genes. Nucleic Acids Res. **36**, 5581–5590.

Lev-Maor, G., Sorek, R., Shomron, N., and Ast, G. (2003) The birth of an alternatively spliced exon: 3' splice-site selection in *Alu* exons. Science **300**, 1288–1291.

Li, J., Kannan, M., Trivett, A. L., Liao, H., Wu, X., Akagi, K., and Symer, D. E. (2014) An antisense promoter in mouse L1 retrotransposon open reading frame-1 initiates expression of diverse fusion transcripts and limits retrotransposition. Nucleic Acids Res. **42**, 4546–4562.

Lowe, C. B., Bejerano, G., and Haussler, D. (2007) Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc. Natl. Acad. Sci. USA **104**, 8005–8010.

Lu, X., Sachs, F., Ramsay, L., Jacques, P.-É., Göke, J., Bourque, G., and Ng, H.-H. (2014) The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol. **21**, 423–425.

Lunter, G., Ponting, C. P., and Hein, J. (2006) Genome-wide identification of human functional DNA using a neutral indel model. PLoS Comput. Biol. **2**, e5.

Lunyak, V. V., Prefontaine, G. G., Núñez, E., Cramer, T., Ju, B.-G., Ohgi, K. A., Hutt, K., Roy, R., García-Díaz, A., Zhu, X., et al. (2007) Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. Science **317**, 248–251.

Lynch, V. J., Leclerc, R. D., May, G., and Wagner, G. P. (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. **43**, 1154–1159.

MacHattie, L. A., and Jackowski, J. B. (1977) Physical structure and deletion effects of the chloramphenicol resistance element Tn9 in phage lambda. *In*: DNA, Insertion Elements, Plasmids and Episomes. (eds.: Bukhari, A. I., Shapiro, J. A., and Adhya, S. L.), pp. 219–228. Cold Spring Harbor Laboratory, New York.

Malassiné, A., Blaise, S., Handschuh, K., Lalucque, H., Dupressoir, A., Evain-Brion, D., and Heidmann, T. (2007) Expression of the fusogenic HERV-FRD Env glycoprotein (syncytin 2) in human placenta is restricted to villous cytotrophoblastic cells. Placenta **28**, 185–191.

Manavathi, B., Samanthapudi, V. S. K., and Gajulapalli, V. N. R. (2014) Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development. Front. Cell Dev. Biol. **2**, 34.

Meuleman, W., Peric-Hupkes, D., Kind, J., Beaudry, J.-B., Pagie, L., Kellis, M., Reinders, M., Wessels, L., and van Steensel, B. (2013) Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. Genome Res. **23**, 270–280.

Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., Tang, X. Y., Edouard, P., Howes, S., et al. (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature **403**, 785–789.

Mikkelsen, T. S., Wakefield, M. J., Aken, B., Amemiya, C. T., Chang, J. L., Duke, S., Garber, M., Gentles, A. J., Goodstadt, L., Heger, A., et al. (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature **447**, 167–177.

Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., and Kazazian, H. H., Jr. (1996) High frequency retrotransposition in cultured mammalian cells. Cell **87**, 917–927.

Munemasa, M., Nikaido, M., Nishihara, H., Donnellan, S., Austin, C. C., and Okada, N. (2008) Newly discovered young CORE-SINEs in marsupial genomes. Gene **407**, 176–185.

Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., and Gage, F. H. (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature **435**, 903–910.

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., and Fraser, P. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature **502**, 59–64.

Nakanishi, A., Kobayashi, N., Suzuki-Hirano, A., Nishihara, H., Sasaki, T., Hirakawa, M., Sumiyama, K., Shimogori, T., and Okada, N. (2012) A SINE-derived element constitutes a unique modular enhancer for mammalian diencephalic *Fgf8*. PLoS One **7**, e43785.

Naville, M., Henriet, S., Warren, I., Sumic, S., Reeve, M., Volff, J. N., and Chourrout, D. (2019) Massive changes of genome size driven by expansions of non-autonomous transposable elements. Curr. Biol. **29**, 1161–1168.

Nikaido, M., Noguchi, H., Nishihara, H., Toyoda, A., Suzuki, Y., Kajitani, R., Suzuki, H., Okuno, M., Aibara, M., Ngatunga, B. P., et al. (2013) Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome Res. **23**, 1740–1748.

Nishihara, H. (2019) Retrotransposons spread potential *cis*-regulatory elements during mammary gland evolution. Nucleic Acids Res. **47**, 11551–11562.

Nishihara, H., Kobayashi, N., Kimura-Yoshida, C., Yan, K., Bormuth, O., Ding, Q., Nakanishi, A., Sasaki, T., Hirakawa, M., Sumiyama, K., et al. (2016a) Coordinately co-opted multiple transposable elements constitute an enhancer for *wnt5a* expression in the mammalian secondary palate. PLoS Genet. **12**, e1006380.

Nishihara, H., Maruyama, S., and Okada, N. (2009) Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. Proc. Natl. Acad. Sci. USA **106**, 5235–5240.

Nishihara, H., and Okada, N. (2008) Retroposons: genetic footprints on the evolutionary paths of life. Methods Mol. Biol. **422**, 201–225.

Nishihara, H., Plazzi, F., Passamonti, M., and Okada, N. (2016b) MetaSINEs: broad distribution of a novel SINE superfamily in animals. Genome Biol. Evol. **8**, 528–539.

Nishihara, H., Smit, A. F., and Okada, N. (2006) Functional non-coding sequences derived from SINEs in the mammalian genome. Genome Res. **16**, 864–874.

Nishihara, H., Stanyon, R., Kusumi, J., Hirai, H., and Koga, A. (2018) Evolutionary origin of OwlRep, a megasatellite DNA associated with adaptation of owl monkeys to nocturnal lifestyle. Genome Biol. Evol. **10**, 157–165.

Ogiwara, I., Miya, M., Ohshima, K., and Okada, N. (2002) V-SINEs: a new superfamily of vertebrate SINEs that are widespread in vertebrate genomes and retain a strongly conserved segment within each repetitive unit. Genome Res. **12**, 316–324.

Ohshima, K., Hamada, M., Terai, Y., and Okada, N. (1996) The 3' ends of tRNA-derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. Mol. Cell. Biol. **16**, 3756–3764.

Ohshima, K., and Okada, N. (2005) SINEs and LINEs: symbionts of eukaryotic genomes with a common tail. Cytogenet. Genome Res. **110**, 475–490.

Okada, N. (1991) SINEs. Curr. Opin. Genet. Dev. **1**, 498–504.

Okada, N., Sasaki, T., Shimogori, T., and Nishihara, H. (2010) Emergence of mammals by emergency: exaptation. Genes Cells **15**, 801–812.

Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T., Suzuki-Migishima, R., Ogonuki, N., Miki, H., et al. (2006) Deletion of *Peg10*, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat. Genet. **38**, 101–106.

Orgel, L. E., and Crick, F. H. (1980) Selfish DNA: the ultimate parasite. Nature **284**, 604–607.

Pastuzyn, E. D., Day, C. E., Kearns, R. B., Kyrke-Smith, M., Taibi, A. V., McCormick, J., Yoder, N., Belnap, D. M., Erlendsson, S., Morado, D. R., et al. (2018) The neuronal gene *Arc* encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. Cell **172**, 275–288.

Platt, R. N., 2nd, Blanco-Berdugo, L., and Ray, D. A. (2016) Accurate transposable element annotation is vital when analyzing new genome assemblies. Genome Biol. Evol. **8**, 403–410.

Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A., Theunissen, T. W., Jaenisch, R., and 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.

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. **8**, 2281–2308.

Rohrmoser, M., Kluge, M., Yahia, Y., Gruber-Eber, A., Maqbool, M. A., Forné, I., Krebs, S., Blum, H., Greifenberg, A. K., Geyer, M., et al. (2018) MIR sequences recruit zinc finger protein ZNF768 to expressed genes. Nucleic Acids Res. **47**, 700–715.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., and Feldman, M. W. (2002) Genetic structure of human populations. Science **298**, 2381–2385.

Santangelo, A. M., de Souza, F. S. J., Franchini, L. F., Bumaschny, V. F., Low, M. J., and Rubinstein, M. (2007) Ancient exaptation of a CORE-SINE retroposon into a highly conserved mammalian neuronal enhancer of the proopiomelanocortin gene. PLoS Genet. **3**, 1813–1826.

Santos, M. E., Braasch, I., Boileau, N., Meyer, B. S., Sauteur, L., Böhne, A., Belting, H.-G., Affolter, M., and Salzburger, W. (2014) The evolution of cichlid fish egg-spots is linked with a *cis*-regulatory change. Nat. Commun. **5**, 5149.

Sasaki, T., Nishihara, H., Hirakawa, M., Fujimura, K., Tanaka, M., Kokubo, N., Kimura-Yoshida, C., Matsuo, I., Sumiyama, K., Saitou, N., et al. (2008) Possible involvement of SINEs

in mammalian-specific brain formation. Proc. Natl. Acad. Sci. USA **105**, 4220–4225.

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G. D., Marshall, A., Flicek, P., and Odom, D. T. (2012) Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell **148**, 335–348.

Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., Hino, T., Suzuki-Migishima, R., Kohda, T., Ogura, A., et al. (2008) Role of retrotransposon-derived imprinted gene, *Rtl1*, in the feto-maternal interface of mouse placenta. Nat. Genet. **40**, 243–248.

Sen, S. K., Han, K., Wang, J., Lee, J., Wang, H., Callinan, P. A., Dyer, M., Cordaux, R., Liang, P., and Batzer, M. A. (2006) Human genomic deletions mediated by recombination between *Alu* elements. Am. J. Hum. Genet. **79**, 41–53.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., et al. (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. **15**, 1034–1050.

Singer, T., McConnell, M. J., Marchetto, M. C. N., Coufal, N. G., and Gage, F. H. (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends Neurosci. **33**, 345–354.

Solovei, I., Kreysing, M., Lanctôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J., and Joffe, B. (2009) Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. Cell **137**, 356–368.

Solovei, I., Thanisch, K., and Feodorova, Y. (2016) How to rule the nucleus: *divide et impera*. Curr. Opin. Cell Biol. **40**, 47–59.

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

Sun, J. H., Zhou, L., Emerson, D. J., Phyo, S. A., Titus, K. R., Gong, W., Gilgenast, T. G., Beagan, J. A., Davidson, B. L., Tassone, F., et al. (2018) Disease-associated short tandem repeats co-localize with chromatin domain boundaries. Cell **175**, 224–238.

Sundaram, V., Choudhary, M. N., Pehrsson, E., Xing, X., Fiore, C., Pandey, M., Maricque, B., Udawatta, M., Ngo, D., Chen, Y., et al. (2017) Functional *cis*-regulatory modules encoded by mouse-specific endogenous retrovirus. Nat. Commun. **8**, 14550.

Sundaram, V., and Wang, T. (2018) Transposable element mediated innovation in gene regulatory landscapes of cells: re-Visiting the "gene-battery" model. Bioessays **40**, 1700155.

Tashiro, K., Teissier, A., Kobayashi, N., Nakanishi, A., Sasaki, T., Yan, K., Tarabykin, V., Vigier, L., Sumiyama, K., Hirakawa,

M., et al. (2011) A mammalian conserved element derived from SINE displays enhancer properties recapitulating Satb2 expression in early-born callosal projection neurons. PLoS One **6**, e28497.

Tenaillon, M. I., Hollister, J. D., and Gaut, B. S. (2010) A triptych of the evolution of plant transposable elements. Trends Plant Sci. **15**, 471–478.

van't Hof, A. E., Campagne, P., Rigden, D. J., Yung, C. J., Lingley, J., Quail, M. A., Hall, N., Darby, A. C., and Saccheri, I. J. (2016) The industrial melanism mutation in British peppered moths is a transposable element. Nature **534**, 102–105.

Vassetzky, N. S., and Kramerov, D. A. (2013) SINEBase: a database and tool for SINE analysis. Nucleic Acids Res. **41**, D83–D89.

Wang, J., Vicente-García, C., Seruggia, D., Moltó, E., Fernandez-Miñán, A., Neto, A., Lee, E., Gómez-Skarmeta, J. L., Montoliu, L., Lunyak, V. V., et al. (2015) MIR retrotransposon sequences provide insulators to the human genome. Proc. Natl. Acad. Sci. USA **112**, E4428–E4437.

Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Raskó, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N. V., Schumann, G. G., et al. (2014) Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature **516**, 405–409.

Wessler, S. R., Bureau, T. E., and White, S. E. (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr. Opin. Genet. Dev. **5**, 814–821.

Xiong, Y., and Eickbush, T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9**, 3353–3362.

Yamanaka, S., Nishihara, H., Toh, H., Nagai, L. A. E., Hashimoto, K., Park, S.-J., Shibuya, A., Suzuki, A. M., Tanaka, Y., Nakai, K., et al. (2019) Broad heterochromatic domains open in gonocyte development prior to *de novo* DNA methylation. Dev. Cell **51**, 21–34.

Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X., Ming, J., Wu, X., Zhang, Y., Xu, Q., et al. (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. Nature **537**, 553–557.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwold, M. J., Meredith, R. W., et al. (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. Science **346**, 1311–1320.

Zhang, Y., Li, T., Preissl, S., Amaral, M. L., Grinstein, J. D., Farah, E. N., Destici, E., Qiu, Y., Hu, R., Lee, A. Y., et al. (2019) Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. Nat. Genet. **51**, 1380–1388.