

REVIEW

Open Access



Transposable elements as essential elements in the control of gene expression

Alemu Gebrie^{1*}

Abstract

Interspersed repetitions called transposable elements (TEs), commonly referred to as mobile elements, make up a significant portion of the genomes of higher animals. TEs contribute in controlling the expression of genes locally and even far away at the transcriptional and post-transcriptional levels, which is one of their significant functional effects on gene function and genome evolution. There are different mechanisms through which TEs control the expression of genes. First, TEs offer cis-regulatory regions in the genome with their inherent regulatory features for their own expression, making them potential factors for controlling the expression of the host genes. Promoter and enhancer elements contain cis-regulatory sites generated from TE, which function as binding sites for a variety of trans-acting factors. Second, a significant portion of miRNAs and long non-coding RNAs (lncRNAs) have been shown to have TEs that encode for regulatory RNAs, revealing the TE origin of these RNAs. Furthermore, it was shown that TE sequences are essential for these RNAs' regulatory actions, which include binding to the target mRNA. By being a member of cis-regulatory and regulatory RNA sequences, TEs therefore play essential regulatory roles. Additionally, it has been suggested that TE-derived regulatory RNAs and cis-regulatory regions both contribute to the evolutionary novelty of gene regulation. Additionally, these regulatory systems arising from TE frequently have tissue-specific functions. The objective of this review is to discuss TE-mediated gene regulation, with a particular emphasis on the processes, contributions of various TE types, differential roles of various tissue types, based mostly on recent studies on humans.

Keywords Transposable elements, Mobile elements, Gene expression

Introduction

Transposable elements (TEs), also called mobile elements, are DNA fragments that may move about inside a host genome and typically make new copies of themselves while they do so. They are present across all forms of life, accounting for 50% of the mammalian genome [1–3]. TEs are present in the genomes of bacteria, plants and mammals, and are divided in two major classes known as Class I retrotransposons and Class II DNA transposons

[4], and these two groups vary from one another in terms of the way they transpose. Class II TEs are less common (3.5%) in the human genome and are regarded as DNA fossils because no family of DNA transposons is still active today [5]. The development of genomics and large-scale functional tests has revealed new knowledge on the many functions of TEs [6].

While DNA transposons move commonly by a cut-and-paste mechanism, retrotransposons do so by a copy-and-paste fashion [7]. The transcription of class I retrotransposons results in an intermediate RNA molecule that may be reverse-transcribed into DNA using reverse transcriptase to create a new copy of the retrotransposon in the genome. On the other hand, Class II DNA transposons produce an enzyme called transposase that separates the parental sequence from the genome

*Correspondence:

Alemu Gebrie
alemugebrie2@gmail.com

¹ Department of Biomedical Sciences, School of Medicine, Debre Markos University, Debre Markos, Ethiopia



before mediating its reintegration into another region of the genome [4, 8].

Retrotransposons come in a variety of forms, such as non-LTR retrotransposons and endogenous retroviruses (ERVs), which are distinguished by the presence of long terminal repeats (LTRs). Long nuclear elements (LINEs), short-interspersed elements (SINEs), and SVAs are further classifications for non-LTR retrotransposons [4, 9]. LINEs make up the majority of non-LTR retrotransposons in the human genome, accounting for 20.4% of it, followed by SINEs (13.1%), LTRs (9.1%), and SVAs (0.1%) [10, 11].

The consequences of TE insertions on host gene expression might be beneficial or harmful, like any mutational process. Regardless of their transposition competency, TE regulatory sequences can be co-opted for host regulatory activities. Evidence gives new thoughts on TE mobility and regulatory potential and serves as a vital resource for population history and disease genetics research [12]. Mechanistically, TEs can influence gene expression either transcriptionally [13], post-transcriptionally [14], or at the step of translation [2, 15] through their encoded products which include both proteins and non-coding RNAs (ncRNAs). More complex than originally thought, the mechanisms by which TEs affect host gene-regulatory networks include: the addition of TFBSs, promoters, and enhancers, alteration of 3D chromatin organization, production of regulatory ncRNAs, co-option/exaptation/domestication of TE-derived coding sequences as new transcriptional effector proteins, and collateral consequences of TE silencing mechanisms [2]. The objective of this review is to discuss TE-mediated gene regulation, with a particular emphasis on the mechanisms, contributions of various TE types, and differential roles of various tissue types, based mostly on recent studies on humans.

The roles of transposable elements in the human genome and cell

The evolution of genetic information, as well as DNA duplication, stability, and gene expression, are just a few of the numerous facets of DNA function that TEs may affect. The discovery of TEs' involvement in genome evolution and gene function has altered the previously held belief that TEs are junk, parasitic, colonizing, or selfish DNA [16]. New genes with crucial host functions can be produced as a result of TEs [17, 18].

According to a number of studies, TEs play an important role in regulating stem cell characteristics, the epithelial to mesenchymal transition, inflammation, and adaptive characteristics such as elevated gene expression, enhanced gene replication, stress tolerance, and aging [19, 20]. It is shown that the transcriptional control

of stress-response genes in *Drosophila melanogaster* is influenced by a variety of families of transposable elements [21]. Also, transposable elements have a significant role in synaptic plasticity, cognition, and tissue development and morphogenesis [22, 23]. In cancer and other inflammatory disorders, the expression of transposon elements triggers a cytokine response and causes the recruitment and infiltration of immune cells [24–28]. In addition to their roles in genomic instability and the trans-regulation of human genes, human endogenous retroviruses have been linked to both the activation and downregulation of the host immune system [29]. A recent study has established how lineage-specific TEs can promote evolutionary turnover and divergence of innate immune regulatory networks and reveals a novel function for B2 SINEs as inducible enhancer elements that affect immunity in mice [30].

Transposons could be altered to incorporate a reporter gene that, when randomly inserted into the bacterial chromosome, can fuse to a gene on the chromosome [31]. This kind of transposon library screening for reporter expression under various situations enables the identification of fusions that are appropriate to stress conditions or a particular therapy. A genome-wide picture of the bacterial regulatory network organization may be obtained from the characterization of these fusions [31]. In addition, a study demonstrated that the expression of retrotransposon is clearly related to aging in *Drosophila* [32].

The negative roles of transposable elements in the human genome

Through processes dependent on and independent of transposition, TEs can lead to genomic/epigenomic instability, which may result in different disease conditions, cell death or the development of cancer [20, 26, 33, 34]. The insertion of TEs into the genome's coding regions has the potential to cause missense or non-sense mutations as well as frameshift mutations linked to premature termination. For instance, when *Alu* elements are inserted into mRNA's exonic regions, the open reading frame (ORF) of that specific coding region is altered, which has an impact on gene expression [35]. As *Alu* elements and *LINE-1* can introduce novel splice sites inside an intron, resulting in alternative splicing events that compromise transcriptional integrity, the insertion of TEs into intronic regions can also have negative consequences [35–37]. Additionally, some research has indicated that the insertion of TEs into the 5' or 3' regions of genes may impair favorable gene expression [37]. As a result, TE insertions' cumulative effects on gene

expression have been linked to a variety of disease conditions, including cancer and genetic disorders [33, 38].

The genomic sequence, chromatin, and nuclear contexts are only a few examples of the factors that interact during TE integration. This variety in insertion-site distribution and evolutionary strategies is explained by these factors [39]. If TEs' capacity to migrate across the genome is not correctly managed, it might be harmful to the host. Long believed to be a mutagen when directed at protein-coding genes, the adverse consequences of mobile element activation were thought to be damaging by triggering chromosomal breakage [40]. TEs typically induce gene disruption and significant genomic abnormalities, including inversions, deletions, and duplications, as a result of their inherent mobility throughout the genome [26].

However, TE transposition may also happen during germline development and, less commonly, in somatic cells [41, 42]. Different instances from various animals and transposon classes have shown the harmful impacts of germline transpositions throughout time. For instance, the P-element DNA transposon (Class II) in *Drosophila*, which is the source of dysgenic characteristics, and LINE1 (L1; Class I) insertions in human haemophilia A can be taken as examples [43, 44]. Since a growing amount of data could link the somatic transposition of TE with harmful biological consequences, somatic transpositions have also drawn a lot of interest. The mariner-Mos1 element of *Drosophila* (Class II transposon), which may be transferred during the *Drosophila* life cycle and adversely influence behavioral activities and embryonic survival [45], is one of the most intriguing examples of this.

There is a plethora of evidence that somatic TE insertions can upregulate oncogenes and lead to genomic rearrangements, which in turn promote different types of cancer [34, 46–48]. Cancer, including ovarian [49], colorectal [50], and Fanconi anemia [51] cancer, to mention a few, is likely one of the clinical conditions that has been most thoroughly studied and associated to a new wave of transpositions in somatic cells. Transposable elements can break free from epigenetic silencing, as has been demonstrated in the majority of cancers. In these cancers, lower methylation levels (hypomethylation) and the dysregulated chromatin modification of L1 retrotransposons lead to their integration into novel sites (insertional mutagenesis), a process that has been observed in pancreatic ductal adenocarcinoma [52] and esophageal squamous cell carcinoma [53].

Through control of cancer-related cellular processes, alternative splicing increases the incidence and progression of several cancer types [54, 55]. The occurrence of alternative splicing events in cancer can be caused by

TEs with the genetic capacity to hop to other sections of the genome [56]. By adjusting various mechanisms, including exonization, providing splicing donor/acceptor sites, alternative regulatory sequences or stop codons, driving exon disruption or epigenetic regulation, TEs can integrate into the genome, primarily in the intronic regions, and induce cancer-specific alternative splicing [57]. Additionally, TEs have the ability to create microRNAs (miRNAs), which regulate the number of transcripts by inhibiting translation or promoting transcript destruction at the post-transcriptional level. Notably, TE insertion alters the whole process of gene expression before and after transcription in cancer cells, favoring the growth of the disease [56].

It is now known that TEs are also linked to the development of other brain illnesses, such as autism and schizophrenia [58–61], in addition to cancer. An investigation found 10 polymorphic TE insertions that are plausible candidates for causative roles in neurologic and psychiatric illnesses, on par with other variants [62]. Whole-genome sequencing of schizophrenia patients' brains revealed specific L1 insertions that were preferentially localized to synapse- and schizophrenia-related genes. This raised the possibility that the hyperactivation of L1 retrotransposons plays a role in the susceptibility and pathophysiology of schizophrenia [63, 64]. Transposons have also been suggested to have a part in the pathogenesis of some neurological changes in Fragile X as well as aberrant social behaviors [65].

Additionally, neurodegenerative disorders including Alzheimer's disease have been linked to the pathophysiology of TEs' abnormal activation and mobilization [66]. The differential expression of numerous TEs and the prevalence of neurofibrillary tangles in post-mortem human brains have been linked in an intriguing way by Guo et al. [66], suggesting a connection between TE activation and genomic instability in Tau-mediated AD processes.

Additionally, cortical spreading depression (CSD), an evolutionarily conserved phenomenon that involves a slow, self-propagating depolarization wave that is linked to the spontaneous depression of electrical neuronal activity, has been hypothesized as a neuroprotective mechanism that can silence TEs mobilization through epigenetic mechanisms [67]. Genome destabilization could be resisted by the effects of DNA methylation-mediated epigenetic control of LINE sequence silencing in conjunction with histone modifications in CSD-induced tolerance [68, 69]. This could stop disabling phenomena like senescence.

Furthermore, genes for antibiotic resistance that are transcriptionally silent are activated by transposable elements. The reservoir of transcriptionally inactive genetic

material found in bacterial genomes can be triggered by a variety of transposon-related recombination events, which encourage the creation of new drug-resistant bacterial strains [70].

Mechanisms of transposable elements silencing

To prevent the harm brought on by the mobilization of TEs, many protective systems have developed within organisms throughout time [71]. These systems—which are frequently lost in cancer cells—include DNA methylation [72], heterochromatin formation [73], histone alterations [74], and mRNA editing [72]. The majority of TEs in somatic cells are silenced by one of these processes, DNA methylation [75] (Fig. 1).

The interactions between the TEs and a large number of non-coding RNAs are the basis of one well-known germline process [76]. The majority of research on regulatory RNAs has focused on the PIWI-interacting RNAs (piRNAs), which interact with TEs at various levels [77–79]. These RNAs have the ability to silence genes at the transcriptional level or modify the accessibility of proteins to the DNA needed for transcription through epigenetic alterations [80]. They are linked to the RNA-induced silencing complex and have the ability to degrade TE transcripts at the post-transcriptional level by producing double-stranded RNAs (dsRNAs) that can be broken down into small-interfering RNAs (siRNAs). In general, piRNAs function in the gonads to shield the germ-line genomes of both males and females from transposable elements [81]. By altering

the histone 3 lysine 9 (H3K9) methylation status, piRNAs suppress TE expression through a mechanism that is consistent across species [82]. Transposon transcription was shown to be enhanced in *Drosophila melanogaster* brains when the Histone 3 lysine 36 methylation mechanism was disrupted [83]. In particular, MIWI2 protein/piRNAs complex recognizes nascent transcripts originating from full-length LINES and calls for a histone methyltransferase, which deposits the histone 3 trimethyl lysine 9 (H3K9me3) mark on LINE repeats in the germ-line genome [82]. PiRNAs also formed complexes with H1/H3K9me3 and heterochromatin protein 1a (HP1a) in an ovarian somatic cell line, altering chromatin accessibility and influencing TE transcription [84]. Additionally, there is evidence that TEs in *Drosophila* somatic cells are silenced as a result of piRNA [85].

Other molecular processes have also developed, such as TE packing into transcriptionally quiet heterochromatin and TE distribution in areas of low gene density [86]. The targeted accumulation of repressive histone alterations silences TEs. In a variety of cells, tissues, species, and biological situations, DNA methylation has primarily been seen as a technique for preventing transposon movement and maintaining genomic integrity. The fact that TE methylation analyses are frequently used as surrogates for global DNA methylation analyses, given that TEs make up such a sizable portion of the mammalian genome, reflects the close relationship between TEs and their silencing by DNA methylation [87, 88].

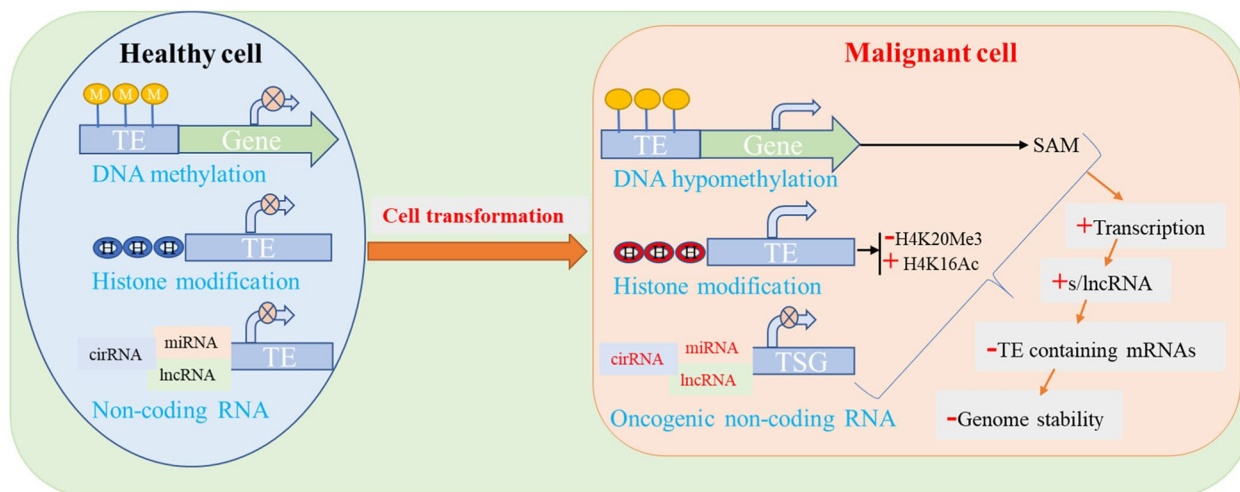


Fig. 1 TEs are regulated in both healthy and cancerous cells. Epigenetic changes such as DNA methylation, histone modification, and non-coding RNA (eg cirRNA, miRNA, and lncRNA) inhibit the function of TEs in healthy cells (left panel). During cellular transformation, hypomethylation with increased S-Adenosyl methionine (SAM), various histone modifications (like methylation and acetylation), and oncogenic non-coding RNAs, which inhibit the expression of tumor suppressor genes (TSGs), all contribute to the loss of repressive signals and the uncontrolled production of TEs in cancer cells (right panel). DNA breakdown, mutations, and genomic instability result from all these (arrows indicate the increased activity, cross circle indicates inhibition; (+) sign indicates increment, (-) sign indicates decrement, and cross sign indicates inhibition) [26]

In addition, a recent investigation has shown for the first time that *Drosophila* Fragile X protein is necessary for transposon inactivation in the larval and adult brains of *Drosophila* "loss of function" *dFmr1* mutants [65]. Results from a study also show that the RNA helicase MOV10 inhibits LINE1 retrotransposition in mice in a dosage-dependent way [89].

Transposable elements in the control of gene expression

If not properly regulated, TE mobilization, expression, and insertion can have detrimental implications on cell physiology. They can also often have a role in regulating gene expression and modifying genomic structure [90]. A ubiquitous mobilization of the L1, Alu, and SINE-R/VNTR/Alu (SVA) transposons, for example, can modify the gene regulatory networks of several types of neurons, notably those in the hippocampus [91, 92]. This is demonstrated in adult brain tissue. Similar to this, several investigations have also shown that L1 retro-transposition in neural precursor cells has a functional purpose [42, 93]. It has been demonstrated, using multi-omic profiling, that L1-promoters are dynamically active in both the developing and adult human brain [94]. Numerous of these transcripts are co-opted as regulatory RNAs or chimera transcripts, and L1s produce hundreds of these developmentally regulated and cell-type-specific transcripts.

One human-specific transcript expressed only during brain development is LINC01876, an L1-derived lncRNA. L1s are implicated in human-specific developmental processes as a result of decreased size of cerebral organoids and premature differentiation of neural progenitors caused by CRISPRi-silencing of LINC01876. Therefore, it has been demonstrated that L1-derived transcripts offer a previously unrecognized layer of transcriptome complexity that is unique to humans and primates and contributes to the functional diversity of the human brain [94–96]. Given that TEs can play a dual and contradictory function in the proper differentiation and development of neuronal mosaicism and in the start of neurological illness, the manifestation of TEs in the brain is symbolic of this "double-edged sword" phenomenon.

These are intriguing illustrations of how TE evolution has included both mechanisms to prevent these invasive sequences from having a negative impact on genome function and systems to allow them to play an active and beneficial part in it. It is becoming apparent that TEs are a crucial component of the genome’s regulatory toolbox [97]. RNA translation, alternative splicing, and gene transcription are just a few biological processes for which repetitive sequences have shown promise as regulators [98]. More and more evidence is mounting that TEs can play a crucial role in regulating gene expression in a variety of mechanisms (Fig. 2).

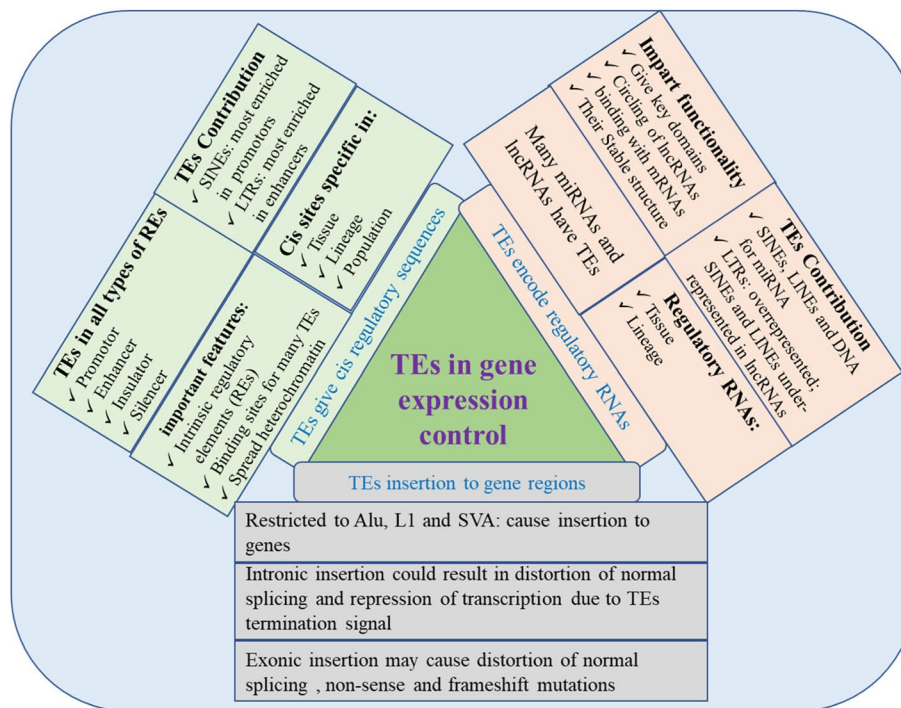


Fig. 2 Different mechanisms that TEs influence gene expression regulation

The role of TEs in epigenetic gene expression control

Only recently has the role of TEs in 3D genome architecture been studied. TEs have an impact on 3D chromatin architecture with a direct effect on the folding of chromosomes [2, 99]. By serving as insulator elements, TEs can potentially affect the structure of the host chromatin [100–102].

The distribution of TEs has historically been thought to have evolved concurrently with the mechanisms to regulate their expression, and the relationship between TEs and epigenetic modifications has frequently been viewed as consisting of silencing TEs by DNA methylation and histone modification [103]. Transposable elements are also becoming a significant source of epigenetic markers that can affect gene expression, which raises the possibility that TE insertion plays a role in directing epigenetic changes to a particular locus. The cause-and-effect relationship between TE mobilization and distribution and the epigenetic control of gene expression is still a difficult problem. Active intragenic TE insertions preferentially occur into genes in the antisense direction (from 3' to 5') during evolution. This prevents sequence invasion during demethylation waves in the genome, which happen, for example, during the development of germ cells and early embryogenesis [14, 104].

These overlapping sense/antisense transcripts mute the TEs by entering an endo-siRNA pathway that is controlled by DICER and Argonaute 2 (AGO2) and is triggered by global demethylation, which raises the levels of repressive histone marks [104]. It is yet unknown what chemical mechanism causes this correlation between the suppression of TEs and the rise in histone repressive marks [104]. KZFP/KAP1 (Krüppel associated box (KRAB) zinc finger protein/KRAB-associated protein 1) complex plays a crucial role in maintaining heterochromatin, with DNA methylation marks at TEs shielding the loci from TET-mediated demethylation, according to new research in naive murine embryonic stem cells prior to implantation [105]. Ecco et al. [106] discovered that two KRAB/ZFP (Kinc Finger Protein) family members control TE targets by histone-based processes in differentiated tissues, which are not necessarily associated with the DNA methylation state of the loci. Additionally, it has been demonstrated that ZFP92 controls the transcription of particular genes in different tissues by the repression of particular TEs [107].

The work demonstrates that the interactions between the TEs and their KRAB-ZFP controllers affect the expression of neighboring genes. It has been shown that primate-specific ERVs serve as docking sites for the co-repressor protein KAP1 (also known as TRIM28) to produce local heterochromatin in human brain progenitor cells, making this connection even more obvious there

[108]. KAP1 binds to the ERVs and represses them, which controls the expression of nearby genes crucial for brain development [108]. The interactions of the transcriptional regulators human silencing hub (HUSH) and microorchidia family CW-type zinc finger 2 (MORC2) with evolutionarily young full-length L1s situated in the transcriptionally permissive euchromatic region, which promotes the deposition of histone H3K9me3, a specific mark for transcriptional silencing, are another example of the regulation of neighboring genes by TEs. A reduction in mRNA expression and potential effects on the RNA polymerase II (POL II) elongation rate might result from this MORC2/HUSH-bound L1 specific impact spreading to nearby genes [109].

Certain kinds of TEs, particularly younger LINES, have been discovered to affect chromatin accessibility in the livers of several inbred mouse strains, serving as a source of chromatin diversity. This demonstrates the ability of TEs to control tissue-specific genes, which may lead to phenotypic variability among populations [110]. Transposable elements can actively reorganize the chromatin structure to regulate gene expression over a lengthy period of time. About 10% of TE families have been discovered to be enriched in active genomic areas generally and across various organs. While L1 LINES and ERV LTRs are the most often enriched TE classes in the repressed areas targeted with the H3K9me3 epigenetic mark, SINES and DNA transposons are the most frequently enriched classes in the active chromatin regions [111].

Intriguingly, open euchromatin areas show the strongest epigenetic impact of TEs. By comparing the epigenomes of two *D. melanogaster* strains, it has been shown, for example, that the enrichment of repressive epigenetic marks around euchromatic TEs is caused by the presence of TEs rather than by the preferential insertion of TEs into genomic regions already enriched with repressive epigenetic marks. This pattern explained why TE-flanking alleles had lower transcript levels and greater histone 3 dimethyl lysine 9 (H3K9me2) enrichment than similar alleles without neighboring TE insertions [112]. Similar to this, the analysis of epigenetic marks in flies with and without Bari-Jheh, a natural transposon that affects the expression of nearby genes, revealed significant differences in histone 3 trimethyl lysine 4 (H3K4me3), H3K9me3, and histone 3 trimethyl lysine 27 (H3K27me3) histone marking in relation to oxidative stress conditions, highlighting that this TE element influences gene expression by affecting the local chromatin state. These illustrations imply that the gene expression of neighboring genes is significantly influenced by the various TE distributions seen in the germlines of various organisms/strains and species [113] (Fig. 3A).

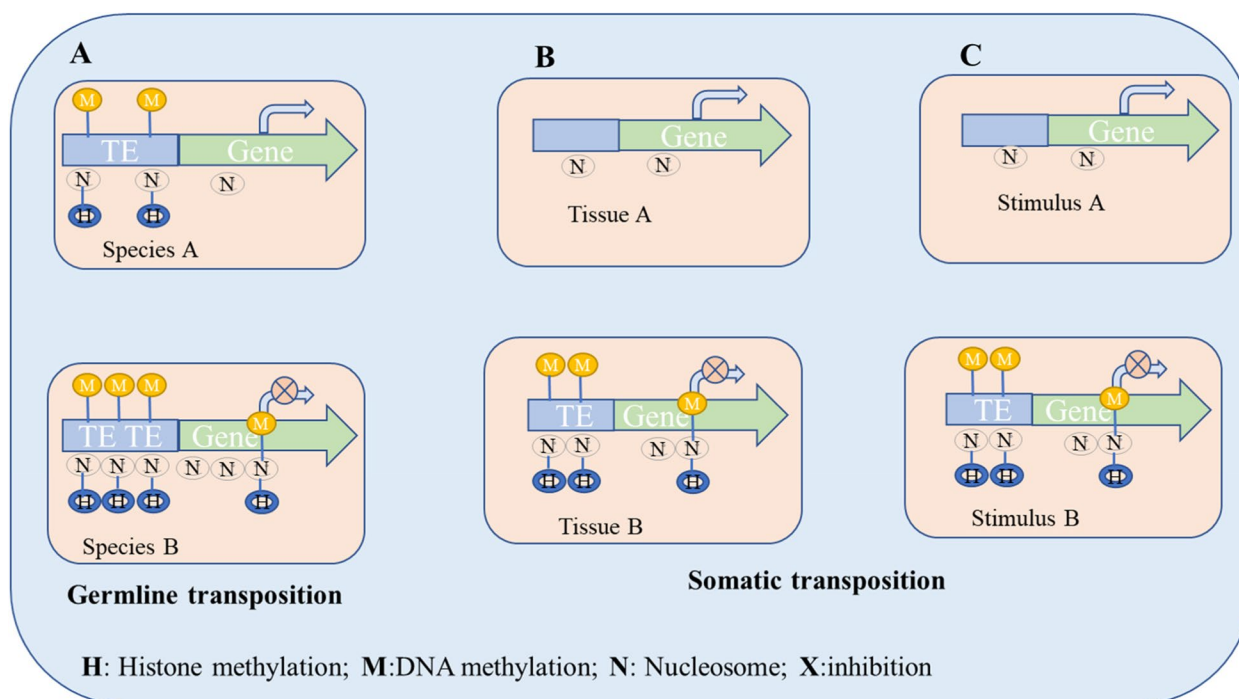


Fig. 3 The consequence of TE distribution on the epigenetic control of gene expression due to changes in methylation of histones and DNA across species, tissues of the same organisms, and stimuli or circumstances. **A** The epigenetic control of a particular gene is altered by the varied distribution of TEs in evolution. TE element controls gene expression by influencing the local chromatin state due to changes in methylation of histones and DNA. **B** The expression of a particular gene is impacted by the differential redistribution of TEs in various cells and tissues of the same organism during development. **C** The expression of a certain gene is influenced by the relocalization of TEs sequence in the same cell following a particular stimulus or circumstance

The hypothesis that TEs have been co-opted and that their distributions have co-evolved with the control of gene expression is supported by the intriguing fact that varied TE distributions resulting from somatic transpositions impacting gene expression are also relevant in the same individual [114]. In actuality, TE enrichment differs between tissues, and TEs have binding sites for tissue-specific master transcription regulators [111]. The fact that integration is only permitted in open chromatin areas explains one aspect of TE targeting. The vicinity of neuronal genes is where somatic LINE insertions are abundant in mammalian brains.

The non-random and targeted tissue-specific distribution of TEs might be viewed as a way to genetically fix a landmark, which could result in an epigenetic regulation of neighboring gene expression, if we consider that TEs can be the target of epigenetic marks (Fig. 3B). The discovery that various environmental conditions cause L1 transposition through various basic helix-loop-helix PER-ARNT-SIM (bHLH/PAS) proteins raises the prospect of L1 insertions being targeted differently under various forms of stress [115]. Experimental data suggest that TE insertions are targeted in

ways that go beyond the ostensible mechanistic need for accessible chromatin [116].

A temporal and functional hierarchy of transcriptional and epigenomic alterations in response to stress is established in *Arabidopsis thaliana* by the increased DNA methylation that silences TEs near environmental-induced genes [117]. It is feasible to suggest that, in response to particular stimuli, the mobilization and insertion of TEs may also be regulated in adult tissues and post-mitotic cells to drive the epigenetic regulation of particular genes (Fig. 3C).

Transposable elements in long-range regulation

Different families of TEs have developed many binding sites for transcription factors during the course of evolution, resulting in various transcriptome landscapes [14]. The ENCODE (Encyclopedia of DNA Elements) data comprises roughly 2 million transcription factors binding sites (TFBSs) that coincide with putatively regulation-competent human retrotransposons. For example, these retrotransposons (44% SINES, 33% LINES, and 23% LR/ERVs) are situated in a 5-kb gene promoter neighborhood [118]. According to the findings, SINES are more

common than LINE-derived transcription factor binding sites (TFBSs) outside of a 5-kb region close to the transcription start site, but the opposite is true within that region [118].

In addition, while it has long been hypothesized that the repeat sequences that TEs disperse across genomes serve as a source of TFBSs that encourage the emergence of new gene regulatory networks [2], it has only recently become clear that the proteins that TEs encode themselves offer complementary pathways to achieve this result. It was suggested that the process of transposase capture may be a recurring idea in the formation of transcription factors by the discovery that well-characterized transcription factors, such as the paired box (PAX) proteins, feature DNA-binding domains that appear to have arisen from transposases [119].

Additionally, the pathways most significantly influenced by the various retrotransposon distributions have been connected to crucial procedures such as cell stress and immunological responses, ribosome biogenesis, chromatin remodeling, DNA replication, mitotic spindle organization, and cell cycle advancement [118]. The discovery that an evolutionary conserved genomic region called AS3 9, made up of three TEs inserted side by side, serves as a distal enhancer for *wnt5a* expression during the morphogenesis of the mammalian secondary palate was made by Nishihara et al. [120].

Functional analyses have demonstrated that the Amn-SINE1, X6b DNA, and MER117 retrotransposons were co-opted by a retroposition/transposition mechanism during the evolution of mammals. This co-option resulted in the acquisition of a specific *Msx1* protein binding site within the X6b DNA sequence, which together with *Wnt5a* is involved in palatogenesis. According to this study, the great variety of numerous cis-regulatory elements (CREs) may have evolved as a result of the combination of several TEs that were all present in the same DNA segment [120].

TEs' role as cis-regulatory elements in the genome

It is believed that the majority of CREs newly evolved during primate evolution are directly derived from TEs [121, 122]. Transposable elements frequently contribute to cis-regulatory elements, tissue-specific expression, and alternative promoters in zebrafish, according to epigenomic analysis [123]. In mammalian genomes, transposable elements are a significant source of various cis-regulatory sequences (Fig. 2). According to some studies, 20% of the CREs found in the human genome may have been taken from TEs [124, 125]. By offering binding sites for trans-acting factors, TEs significantly contribute to all cis-regulatory regions (promoters, enhancers, silencers, and insulators) in the human genome [122]. TEs serve as

a reservoir for a variety of regulatory functions and are crucial to the evolution of many regulatory components. They either offer substitute enhancers and promoters or change the activity of the current promoters [126, 127].

It has been well established that TEs may adapt to regulatory elements in the human genome and take on non-TE activities [128, 129]. The transcriptional activity of TE-derived sequences in regulatory elements has been empirically verified in several investigations [127, 130, 131]. According to one study, out of the 35,007 promoters, 75% were identified to have TE-derived sequences, with some promoters possessing as many as ten TEs [132]. However, only 6.8% of the TFBSs in promoters were found to be TE-derived, according to the study.

Studies have shown that TFs bind to TEs and that these proteins contain TF-binding sequence motifs [125, 132, 133]. In the human genome, TFBSs do not just happen to exist across TEs at random. A TF's binding sites are more likely to include copies of particular TE families [133].

Different TE types contribute differently to the regulatory elements in the human genome [134]. While L1s were shown to be least prevalent in the regulatory areas, Alu elements were revealed to contribute the most to all varieties of regulatory regions. Additionally, in gene-surrounding regions, SINE-derived TFBSs outnumber LINE-derived TFBSs, but the inverse is true for regions outside the gene neighborhood [118]. Additionally, contrary to LINES, SINEs are more common in promoters than in non-promoter areas [132].

It is well known that TE-derived regulatory sequences regulate the expression of a large number of genes in the human genome. By adopting a reporter gene expression strategy or by finding alternative transcripts which start at TE sequences, certain research that concentrated on particular genes were able to uncover TE-derived regulatory elements. For instance, TE-derived regulatory elements control the expression of the Proopiomelanocortin (POMC), Colony Stimulating Factor 1 Receptor, Fc Epsilon Receptor Ig, CD8 genes, and many others [127, 130, 135–137].

Numerous TEs that provide cis-regulatory sequences frequently work in a tissue-specific manner and are crucial for the differential expression of genes in various tissues. Human tissues differ in the epigenetic state of TEs, which affects the profile of TE regulatory functions in various tissue types [138]. One way that TEs are thought to contribute to evolutionary innovation in gene regulation is through their tissue-specificity. Studies that concentrated on certain genes have shown that TEs are adapted to tissue-specific regulatory sequences. For instance, an LTR retroelement offers a neural enhancer for the immunological and POMC genes, and it was shown that Alu sequences give T cell promoters and

enhancers for the FCER1 and CD8 genes, respectively [127, 130, 136, 139].

In stem cells, TEs exhibit remarkable cis-regulatory functions. Early germ cells, which share many transcriptional characteristics, rely on a largely overlapping collection of transcription factors, and have widely permissive chromatin landscapes that may further enhance TE activation and have transcriptional activity of TEs that is typically higher [140]. Early in development, young TE families—often LTR elements with embryonic TFBSs in their ancestral sequence—display extremely particular transcriptional patterns [141, 142].

In somatic cells, TEs support cis-regulatory gene networks through the following mechanisms: overlap between the cis-regulatory programs of somatic cells and stem cells, retroviral hijacking of transcription factors expressed in different types of immune cells, or gain of somatic regulatory activity through TE sequence mutations that take place after genomic insertion [139, 143, 144].

Additionally, TE-derived regulatory sites frequently are species/lineage-specific and add innovation and variety to speciation. Future thorough analyses including all regulatory element types across a wide range of species ought to offer more information [127, 145].

However, it has been discovered that mobile element insertion polymorphisms are the most common structural variations in the human genome. Alu elements, L1s, and SVAs are the three groups of retrotransposons that are predominantly in charge of producing human TE polymorphisms [146–149]. The average difference between the two haploid human genomes of the same person is

thought to be around 1000 TE insertions [6]. There have not been many studies connecting human polymorphic TEs with variations in gene expression between groups, though.

mRNA decay and splicing

Transposable elements can influence the activity of snc/lncRNAs: microRNAs, circular RNAs, as well as the stability of mRNA through non-sense-mediated decay [14]. Given that a sizable portion of these ncRNAs have their roots in TEs, there is a strong relationship between TEs and regulatory RNAs. The Alu sequences, a subfamily of SINES elements, are frequently found in introns or the 3'-untranslated region (UTR) of mature and pre-mRNAs [150]. Alu elements included in the sequences of mRNA and lncRNAs may contribute to Staufen-mediated degradation (SMD). Alu elements found in the 3'-UTR of SMD targets and Alu elements found in cytoplasmic and polyadenylated lncRNAs have been shown to imperfectly pair to produce STAU1 binding sites [151]. These locations can interact with the STAU protein, causing a downregulated mRNA expression profile that initiates SMD [151] (Fig. 4).

Lineage-specific 3'-UTR SINES have a special role in the convergence of gene expression patterns between species, as shown by the ability of these SINES to control the amounts of mRNAs by directing SMD from orthologous genes in many species (including human) [152]. The control of mRNA quantity and alternative splicing can also be influenced by sequences produced from transposable elements and localized in RNA transcripts. TEs inserted into introns are detected by the splicing

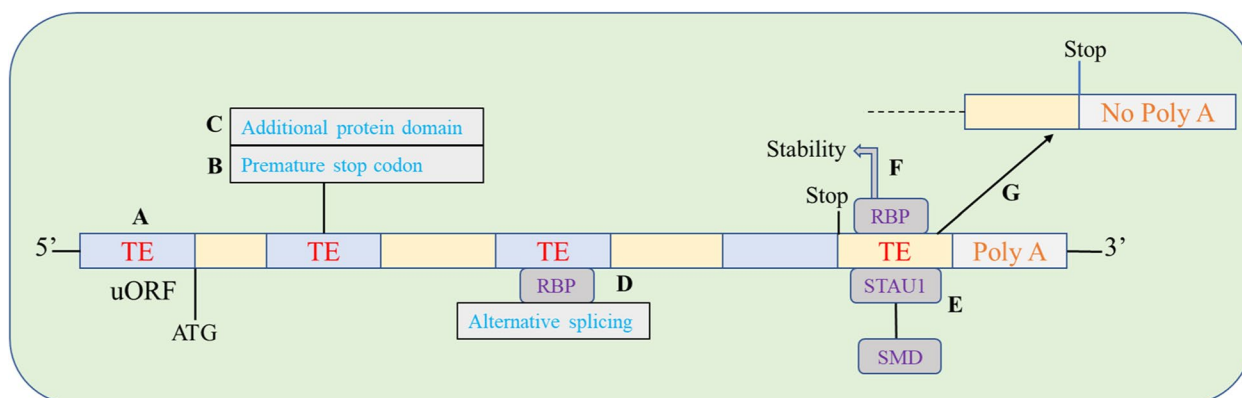


Fig. 4 Regulation of gene expression by TE-dependent post-transcription. **A** The presence of an upstream ORF that contributes in the control of the main ORF translation is brought about by a TE inside the 5'-UTR; **B** The existence of an extra domain inside the encoded protein is determined by the exonization of a TE and consequently its translation; **C** The presence of a premature stop codon due to the exonization of a TE inside the coding region of an mRNA can lead to the Nonsense-mediated Decay process; **D** The alternative splicing process can be impacted by TE sequence interactions with RNA-binding proteins (RBP); **E** An mRNA's 3'-UTR contains TE sequences that might cause STAU-mediated degradation; **F** acts as a docking point for RBP important in maintaining RNA stability, such as the HuR protein; **G** or causes the production of a shorter, poly-A tailless mRNA, which results in translational repression

machinery and recruited into RNA transcripts as exons in a process known as exonization [14]. TEs have a lot of splice-donor and splice-acceptor sites, which help in alternative splicing. They can thus interact with a wide variety of RNA binding proteins (RBPs) that have particular TE binding sites that they like to attach to, such as the Human antigen R (HuR) or Fused in Sarcoma proteins, which prefer to bind to U-rich motifs [152]. However, the disappearance of TE binding sites for several RBPs suggests that these sites influence transcript quantity and splicing in a manner comparable to that observed in gene-binding sites found in non-repetitive sequences [152]. Additionally, it has been noted that in a small number of instances, the impact of RBP binding might vary according to the particular TE family bound. If it is not bound to an Alu element in a U-rich region, the RBP HuR, for instance, confers transcript stability [153].

Additionally, the Alu-containing RNAs might be able to organize themselves into stable structural domains, which would probably result in new biological activities [154]. By modulating the levels of many nonsense-mediated RNA decay switch exons (NSEs), pseudo-exons produced by the activation of cryptic splice sites that serve as a buffer to prevent Alu-mediated NSE activation, Kralovicova et al. [155] have demonstrated that an intronic transposed element highly similar to medium reiterated frequency repeat family 51 can affect gene expression.

Transposable elements contribute to noncoding RNAs (ncRNAs)

Noncoding RNAs, which are often longer than 200 nucleotides and are not translated into proteins, include long non-coding RNAs and small non-coding RNAs [156, 157]. There are several types of long non-coding RNAs, each with a unique genomic location in regard to genes and exons [158]. These include Long intervening/intergenic noncoding RNAs which do not overlap protein-coding genes, intronic ncRNAs, and sense and antisense lncRNAs. On the other hand, small ncRNAs consist of a variety of RNAs, including piwiRNAs, microRNAs, and small nucleolar RNAs [157, 159].

There is no doubt that TEs have significantly influenced regulatory RNAs (miRNAs and lncRNAs) [160, 161] (Fig. 2). Certain TE families' palindromic sequences play critical roles in the hairpin structure of miRNAs, and various TEs are connected to various miRNA families. An investigation on the origins of human miRNAs from transposons revealed that miRNAs are most frequently produced from LINE elements (108 miRNAs) and SINE elements (94 miRNAs), and less frequently from DNA transposons (64 miRNAs) and LTR-containing retroelements (53 miRNAs) [162]. Additionally, mature miRNAs that are not hairpins

contain TE sequences. The fact that TEs are found in promoters, introns, and exons of lncRNA genes emphasizes the role that TEs play in the production of lncRNAs [163, 164].

Compared to protein-coding genes, long non-coding RNAs (lncRNAs) have a substantially greater density of TE-derived sequences, with TEs making up around 30% of the total lncRNA sequence in human and mouse [165, 166]. TEs offer significant cues for the synthesis of lncRNAs, just as they do for protein-coding genes. For instance, LTR-derived promoters control the expression of 10% of human lncRNAs [165]. Additionally, many ways by which TE sequences produced as a component of lncRNA species might alter gene expression [167, 168]. Numerous lncRNAs produced from TE have been linked to embryonic development. The transcriptome of pluripotent stem cells has been found to be 30% more complicated than that of differentiated cells, likely as a result of the prominent transcriptional activation of TEs [169]. Of note, some are significantly expressed in these cells. The variety of molecular methods that may be used to analyze TE-derived ncRNAs has skyrocketed in recent years. As a result, it can be anticipated that there will be plenty of intriguing and novel roles for ncRNAs originating from TEs still to be uncovered.

The circular RNAs (circRNA), a novel family of short noncoding RNAs with gene control activities, are another class of small RNAs implicated in TE regulation. Recent mammalian investigations have demonstrated that transposons with the capacity to drive circRNA synthesis via reverse complementary pairing are abundant in the flanking regions of circRNAs [170, 171]. A variety of Alu pairings are found in human introns when multiple circRNAs derived from the same gene locus are present, pointing to a potential role for pairing competition in the development of alternative circularizations [172]. LINE1-like elements and their reverse complementary pairs (LLERCPs) are highly abundant in the flanking regions of circRNAs, according to recent research in maize that sequenced circRNA [173]. It is interesting to note that when LLERCP transcription rises, circRNA accumulation changes and linear transcript levels fall [173].

The expression of a different class of long non-coding RNAs known as cis-natural antisense transcripts (cis-NATs) can be influenced by TEs in various ways, according to Jung et al. [174]. First, depending on the TE, different promoters can be used to transcribe NATs that come from TEs. They can also be exonized by TEs, with the newly generated exon complementing the exon of a gene encoding a sense protein [174]. Thus, NATs may use dsRNA formations to engage in RNA interference or adenosine deaminase acting on RNA (ADAR) pathways to mediate the production of sense transcripts.

The results of several studies demonstrate that TEs have a role in how regulatory RNAs work, including, but not limited to, assisting circularize lncRNAs, binding regulatory RNA to target mRNAs, and creating stable secondary structures for regulatory RNAs [175, 176]. Different sncRNA and lncRNA types receive functional features from TE-derived sequences, which makes them crucial for regulatory RNA activities [170, 177].

The lineage and tissue specificity of TE-derived regulatory RNAs has been shown in several investigations. Because of their tendency to be less conserved and more lineage-specific, TEs are thought to be a significant source of the lineage-specificity of regulatory RNAs [163, 167, 178–180]. Also, TE-enriched regulatory RNAs can be tissue-specific, according to research. For instance, in the work by Kang et al., it was discovered that a total of 29 human lncRNAs had tissue-specific expression, of which 20 were lncRNAs originating from TE [174]. Additionally, TE sequences are present in 9 out of the 11 lncRNAs that have been shown to be expressed in cancer cell lines, indicating that these lncRNAs play a role in the development of cancer [166, 181, 182].

SINEs and LINEs make up the majority of TEs that contribute to the sequence of lncRNA. SINEs and LINEs are underrepresented in lncRNAs compared to the whole genome, while LTRs are overrepresented. In conclusion, the distribution of TEs in the introns of lncRNA genes is essentially similar to that of the entire genome, but LINEs are underrepresented in exonic and promoter areas while LTRs are overrepresented in lncRNA exons and promoters compared to protein-coding genes [166, 179, 183].

Transposable elements and protein translation

Retrotransposons have developed alongside genes, inserting into various locations along the gene bodies and resulting in a wide range of outcomes (Fig. 2). The expression of many genes' proteins is affected by transposable element insertions into mRNAs' 3'-UTRs or 5'-UTRs in a variety of ways. According to Kitano et al.'s research [184], TEs are also involved in the translational control of several genes through the use of upstream open reading frames (uORFs). One study's findings point to a transposon as the source of uORFs and reveal a novel function for transposable elements in influencing protein abundance and phenotypic variety by altering translation rate [185]. Canonical ORFs located downstream of the uORFs in eukaryotic mRNAs can act as cis-acting elements to inhibit or promote translation through the use of the major families of retrotransposons, such as LINEs and SINEs.

Kitano et al. [184] used the human RefSeq mRNA sequence database to show that 10% of human uORFs are produced and controlled by TEs located in the 5'-UTR

of mRNAs. Although earlier research has shown that retrotransposons function as translational regulators, it is still unclear how DNA transposons affect the translation of the protein-host. The genomes of both plants and animals include a large number of MITEs, or miniature inverted-repeat transposable elements. Their presence in the 3'-UTRs of rice mRNAs has been shown to have a regulatory effect via a translational repression mechanism. In rice, the Ghd2 gene, a member of the CCT gene family (CONSTANS (CO), CO-Such, and TIMING OF CAB1), controls crucial agronomic variables like grain quantity, plant height, and heading date [186]. The Dicer-like 3a (OsDCL3a) pathway, which may produce shorter mRNA without a poly-A tail by processing the nascent MITE transcripts, is the method by which the MITEs inhibit Ghd2's translation [186]. It is still unknown how MITEs suppress mRNA translation and which phase of mRNA translation is impeded by them [186]. TEs have been linked to the creation of novel alternative mRNA splicing isoforms when found in the coding region of genes.

This can be viewed as a transitional stage in the evolution of new genes. Examples include the lamina-associated polypeptide 2alpha (LAP2alpha) domain-containing splice isoforms of the mammalian thymopoietin (TMPO) and zinc finger protein 451 (ZNF451) genes, which are both related to the first ORF from a retrotransposon-like *Dictyostelium* intermediate repeat sequence 1. The canonical protein and a brand-new non-canonical protein isoform are both produced by both mRNAs. Particularly, during evolution, the LAP2a specific isoform of TMPO was co-opted to support a new and significant role in the cell [54, 187].

In a process known as domestication, the evolutionary insertion of TEs in gene coding areas has also resulted in chimeras. An excellent example of retrotransposon domestication is the activity regulated cytoskeleton associated Protein (Arc), which most research indicates is descended from a vertebrate lineage of Ty3/gypsy retrotransposons [187]. Particularly important for learning and memory is the cellular immediate-early gene Arc, whose mRNA is found near the synaptic junction. Arc possesses an internal ribosomal entry site that permits cap-independent translation, which is an intriguing way in which the control of Arc mRNA mirrors that of viral RNA [188].

The protein structure of the Arc subdomains has been demonstrated by crystallography to create a bi-lobar architecture akin to the capsid domain of the human immunodeficiency virus gag protein [188]. These results imply that Gag-containing components have been repurposed by evolution to mediate intercellular communication in the neurological system [188].

Therapeutic potential of TEs

Transposable elements might serve as therapeutic targets for a variety of complicated diseases, including malignancies and CNS-related genetic disorders. Transposable elements are capable of sequence insertion or deletion, enabling precise control of gene expression and modifications to pathophysiological pathways. The creation of specialized oligonucleotide sequences aimed at these sequences may be made possible by the identification of certain TEs involved in the etiology of the diseases. The potential of TEs as gene therapy tools has been examined in a number of preclinical investigations using various models of human disorders [26]. It is noteworthy that gene therapy employing TE-based vectors has shown to be a promising approach for the treatment of many hereditary and acquired human disorders. In addition, many DNA-transposon-based vectors were modified for gene therapy procedures by taking use of qualities like its potential for integration and non-viral nature [189].

Current clinical studies frequently target TEs or benefit from TE biology. Clinical studies using checkpoint inhibitor treatment for immune signaling against renal, ovarian, colorectal, and melanoma malignancies that include TE signaling pathways are currently being conducted [190].

In relation to TEs, both humoral and cell-mediated immunity have been investigated. Several malignancies, including ovarian and melanoma patients as well as teratocarcinoma cell lines, have been linked to anti-ERV-K antibodies [191]. Adaptive immune activity to target TEs as new therapeutic targets was found to be aided in cancer patients by T-cell-mediated and autologous humoral response.

Overexpression of transposon elements in different human diseases is due to demethylation of the TE loci [192]. The TE transcript mechanism, however, is occasionally independent of DNA methylation [193]. This raises the possibility of additional TE regulatory mechanism for non-coding RNAs and histone alterations. Human disorders are significantly influenced by RNA modification [194]. It has been revealed that transposon RNA M(6)A underwent one of its modifications [195]. Transposons may have a role in certain human disease mechanisms, by making use of attractive targets for treatments. In addition to RNA changes, one may look at the uncharacterized DNA modifications of TEs for additional study. One such is m6dA, which is found in the human genome at certain locations, is linked to enhanced transcription activity, and has been implicated in cancer [196, 197]. The link between TE loci and biomarkers raised in disease states would be an intriguing area for further research.

Limitation of the review

This review has a limitation in that it non-specifically addresses the role of transposable elements in the regulation of gene expression. It is more descriptive since it is a narrative review rather than a systematic review and/or meta-analysis, which are supported by statistical analyses and which can objectively answer a particular subject. Therefore, this review presents the authors' own perspectives on a more general topic.

Conclusion and perspective

There are a number of recent discoveries that support the increasingly clear active involvement of TEs in genome function, highlighting their impact on the control of gene expression. In addition to providing ready-to-use TFBSs or undergoing mutations to generate binding motifs for TFs, TEs have inherent regulatory mechanisms for controlling their own expression. Many genes' regulatory elements contain TE sequences, which are involved in both short- and long-range regulation of gene expression. By actively taking part in the production of regulatory RNAs, TEs also contribute to the control of genes. There is still much to learn about the function of transposable elements in gene regulation and their therapeutic potential.

Abbreviations

bHLH	Basic helix-loop-helix
CD8	Cluster of differentiation 8
CircRNA	Circular RNAs
cis-NATs	Cis-natural antisense transcripts
CREs	Cis-regulatory elements
CSD	Cortical spreading depression
H3K9	Histone 3 lysine 9
HP1a	Heterochromatin protein 1a
HuR	Human antigen R
HUSH	Regulators human silencing hub
KRAB	Krüppel associated box
LINEs	Long interspersed nuclear elements
LLERCPs	LINE1-like elements and their reverse complementary pairs
LTRs	Long terminal repeats
miRNAs	MicroRNAs
NSEs	Nonsense-mediated RNA decay switch exons
piRNAs	P-element Induced WImpy testis (PIWI)-interacting RNAs
POMC	Proopiomelanocortin
RBPs	RNA binding proteins
s/lncRNAs	Short/Long non-coding Ribonucleic acids
SINEs	Short-interspersed nuclear elements
siRNAs	Small-interfering RNAs
SMD	Staufen-mediated degradation
SVAs	SINEs Variable Number of Tandem Repeats Alu
TEs	Transposable elements
TFBSs	Transcription factor binding sites
UORFs	Upstream open reading frames
UTR	Untranslated region

Acknowledgements

Not Applicable.

Authors' contributions

AG: Conception of research review, literature review, interpretation and write up of the manuscript.

Funding

No fund has been obtained for this study.

Availability of data and materials

Not Applicable.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 27 April 2023 Accepted: 8 August 2023

Published online: 18 August 2023

References

- Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. *Annu Rev Genet.* 2020;54:539.
- Fueyo R, Judd J, Feschotte C, Wysocki J. Roles of transposable elements in the regulation of mammalian transcription. *Nat Rev Mol Cell Biol.* 2022;23(7):481–97.
- Hoyt SJ, Storer JM, Hartley GA, Grady PG, Gershman A, de Lima LG, Limouse C, Halabian R, Wojenski L, Rodriguez M. From telomere to telomere: the transcriptional and epigenetic state of human repeat elements. *Science.* 2022;376(6588):eabk3112.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet.* 2007;8(12):973–82.
- Pace JK, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res.* 2007;17(4):422–32.
- Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS. Ten things you should know about transposable elements. *Genome Biol.* 2018;19(1):1–12.
- Lodish H, Berk A, Kaiser CA, Kaiser C, Krieger M, Scott MP, Bretscher A, Ploegh H, Matsudaira P. *Molecular cell biology.* 8th ed. New York: W.H. Freeman; 2016.
- Gifford WD, Pfaff SL, Macfarlan TS. Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol.* 2013;23(5):218–26.
- Lee S-I, Kim N-S. Transposable elements and genome size variations in plants. *Genomics Inf.* 2014;12(3):87–97.
- Tang W, Mun S, Joshi A, Han K, Liang P. Mobile elements contribute to the uniqueness of human genome with 15,000 human-specific insertions and 14 Mbp sequence increase. *DNA Res.* 2018;25(5):521–33.
- Lancet D, Consortium HGS. Initial sequencing and analysis of the human genome. *Nature.* 2001;409(6822):860–921.
- Autio MI, Bin Amin T, Perrin A, Wong JY, Foo RS-Y, Prabhakar S. Transposable elements that have recently been mobile in the human genome. *BMC Genomics.* 2021;22(1):1–17.
- Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet.* 2017;18(2):71–86.
- Drongitis D, Aniello F, Fucci L, Donizetti A. Roles of transposable elements in the different layers of gene expression regulation. *Int J Mol Sci.* 2019;20(22):5755.
- Fort V, Khelifi G, Hussein SM. Long non-coding RNAs and transposable elements: a functional relationship. *Biochim Biophys Acta (BBA)-Mol Cell Res.* 2021;1868(1):118837.
- Ayarpadikannan S, Kim H-S. The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics Inf.* 2014;12(3):98–104.
- Elisaphenko EA, Kolesnikov NN, Shevchenko AI, Rogozin IB, Nesterova TB, Brockdorff N, Zakian SM. A dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS ONE.* 2008;3(6): e2521.
- Mi S, Lee X, Li X-p, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang X-Y, Edouard P, Howes S. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature.* 2000;403(6771):785–9.
- Schneider BK, Sun S, Lee M, Li W, Skvir N, Neretti N, Vijg J, Secombe J. Expression of retrotransposons contributes to aging in *Drosophila*. *Genetics.* 2023;224(2):iyad073.
- Hsu P-S, Yu S-H, Tsai Y-T, Chang J-Y, Tsai L-K, Ye C-H, Song N-Y, Yau L-C, Lin S-P. More than causing (epi) genomic instability: emerging physiological implications of transposable element modulation. *J Biomed Sci.* 2021;28(1):1–14.
- Villanueva-Cañas JL, Horvath V, Aguilera L, González J. Diverse families of transposable elements affect the transcriptional regulation of stress-response genes in *Drosophila melanogaster*. *Nucleic Acids Res.* 2019;47(13):6842–57.
- Miao B, Fu S, Lyu C, Gontarz P, Wang T, Zhang B. Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biol.* 2020;21(1):1–25.
- Nikolaïenko O, Patil S, Eriksen MS, Bramham CR. Arc protein: a flexible hub for synaptic plasticity and cognition. *Seminars in cell & developmental biology.* 2018;77:33–42.
- Kong Y, Rose CM, Cass AA, Williams AG, Darwish M, Lianoglou S, Haverty PM, Tong A-J, Blanchette C, Albert ML. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat Commun.* 2019;10(1):5228.
- Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS. Ten things you should know about transposable elements. *Genome Biol.* 2018;19:1–12.
- Bhat A, Ghatage T, Bhan S, Lahane GP, Dhar A, Kumar R, Pandita RK, Bhat KM, Ramos KS, Pandita TK. Role of transposable elements in genome stability: implications for health and disease. *Int J Mol Sci.* 2022;23(14):7802.
- Zheng R, Dunlap M, Lyu J, Gonzalez-Figueroa C, Bobkov G, Harvey SE, Chan TW, Quinones-Valdez G, Choudhury M, Vuong A et al. LINE-associated cryptic splicing induces dsRNA-mediated interferon response and tumor immunity. *bioRxiv.* posted February 24, 2023.
- Carrillo D, Reggiardo RE, Lim J, Mantalas G, Peddu V, Kim DH. Transposable element RNA dysregulation in mutant KRAS(G12C) 3D lung cancer spheroids. *bioRxiv.* posted February 28, 2023.
- Grandi N, Tramontano E. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front Immunol.* 2018;9:2039.
- Horton I, Kelly CJ, Dziulko A, Simpson DM, Chuong EB. Mouse B2 SINE elements function as IFN-inducible enhancers. *Elife.* 2023;12:e82617.
- Figueroa-Bossi N, Balbontin R, Bossi L. Use of transposable reporters in the analysis of bacterial regulatory networks. *Cold Spring Harb Protoc.* 2023.
- Schneider BK, Sun S, Lee M, Li W, Skvir N, Neretti N, Vijg J, Secombe J. Expression of retrotransposons contributes to aging in *Drosophila*. *Genetics.* 2023;224(2):iyad073.
- Konkel MK, Batzer MA. A mobile threat to genome stability: The impact of non-LTR retrotransposons upon the human genome. *Seminars in cancer biology.* 2010;20(4):211–21.
- Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, Zhang D, Li D, Xing X, Kim S. Transposable elements drive widespread expression of oncogenes in human cancers. *Nat Genet.* 2019;51(4):611–7.
- Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet.* 2009;10(10):691–703.
- Belancio VP, Hedges DJ, Deininger P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.* 2006;34(5):1512–21.

37. Lev-Maor G, Ram O, Kim E, Sela N, Goren A, Levanon EY, Ast G. Intronic Alu s influence alternative splicing. *PLoS Genet.* 2008;4(9): e1000204.
38. Shah N, Liang Y. Chimeric transcripts of transposable elements and genes are a source of tumor-specific antigens. *Nat Genet.* 2023;55(4):538–9.
39. Sultana T, Zamborlini A, Cristofari G, Lesage P. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet.* 2017;18(5):292–308.
40. Belancio VP, Deininger PL, Roy-Engel AM. LINE dancing in the human genome: transposable elements and disease. *Genome Med.* 2009;1(10):1–8.
41. Haig D. Transposable elements: self-seekers of the germline, team-players of the soma. *BioEssays.* 2016;38(11):1158–66.
42. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature.* 2005;435(7044):903–910.
43. Kazazian HH, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature.* 1988;332(6160):164–6.
44. Bingham PM, Kidwell MG, Rubin GM. The molecular basis of PM hybrid dysgenesis: the role of the P element, a P-strain-specific transposon family. *Cell.* 1982;29(3):995–1004.
45. Pereira CM, Stoffel TJ, Callegari-Jacques SM, Hua-Van A, Capy P, Loreto EL. The somatic mobilization of transposable element mariner-Mos1 during the *Drosophila* lifespan and its biological consequences. *Gene.* 2018;679:65–72.
46. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ III, Lohr JG, Harris CC, Ding L, Wilson RK. Landscape of somatic retrotransposition in human cancers. *Science.* 2012;337(6097):967–71.
47. Tubio JM, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, Gundem G, Pipinikas CP, Zamora J, Raine K. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science.* 2014;345(6196):1251343.
48. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, Santamarina M, Ju YS, Temes J, Garcia-Souto D. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nat Genet.* 2020;52(3):306–19.
49. Tang Z, Steranka JP, Ma S, Grivainis M, Rodić N, Huang CRL, Shih I-M, Wang T-L, Boeke JD, Fenyő D. Human transposon insertion profiling: analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. *Proc Natl Acad Sci.* 2017;114(5):E733–40.
50. Pitkänen E, Cajuso T, Katainen R, Kaasinen E, Välimäki N, Palin K, Taipale J, Aaltonen LA, Kilpivaara O. Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer. *Oncotarget.* 2014;5(3):853.
51. Brégnard C, Guerra J, Déjardin S, Passalacqua F, Benkirane M, Laguette N. Upregulated LINE-1 activity in the Fanconi anemia cancer susceptibility syndrome leads to spontaneous pro-inflammatory cytokine production. *EBioMedicine.* 2016;8:184–94.
52. Rodić N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, Sharma R, Kohutek ZA, Huang CR, Ahn D, Mita P. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat Med.* 2015;21(9):1060–4.
53. Doucet-O'Hare TT, Sharma R, Rodić N, Anders RA, Burns KH, Kazazian HH Jr. Somatic LINE-1 insertions in normal esophagus undergo clonal expansion in esophageal squamous cell carcinoma. *Hum Mutat.* 2016;37(9):942–54.
54. Abascal F, Tress ML, Valencia A. Alternative splicing and co-option of transposable elements: the case of TMPO/LAP2a and ZNF451 in mammals. *Bioinformatics.* 2015;31(14):2257–61.
55. Bernard A, Boidot R, Végran F. Alternative splicing in cancer and immune cells. *Cancers.* 2022;14(7):1726.
56. Kim WR, Park EG, Lee YJ, Bae WH, Lee DH, Kim H-S. Integration of TE induces cancer specific alternative splicing events. *Int J Mol Sci.* 2022;23(18):10918.
57. Krasileva KV. The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes. *Curr Opin Plant Biol.* 2019;48:18–25.
58. Doyle GA, Crist RC, Karatas ET, Hammond MJ, Ewing AD, Ferraro TN, Hahn C-G, Berrettini WH. Analysis of LINE-1 elements in DNA from postmortem brains of individuals with schizophrenia. *Neuropsychopharmacology.* 2017;42(13):2602–11.
59. Shpyleva S, Melnyk S, Pavliv O, Pogribny I, Jill James S. Overexpression of LINE-1 retrotransposons in autism brain. *Mol Neurobiol.* 2018;55(2):1740–9.
60. Alesi V, Genovese S, Lepri FR, Catino G, Lodo S, Orlando V, Di Tommaso S, Morgia A, Martucci L, Di Donato M et al. Deep Intronic LINE-1 insertions in NF1: expanding the spectrum of neurofibromatosis type 1-associated rearrangements. *Biomolecules.* 2023;13(5):725.
61. DeRosa H, Smith A, Geist L, Cheng A, Hunter RG, Kentner AC. Maternal immune activation alters placental histone-3 lysine-9 tri-methylation, offspring sensorimotor processing, and hypothalamic transposable element expression in a sex-specific manner. *Neurobiol Stress.* 2023;24:100538.
62. Ahn HW, Worman ZF, Lechinska A, Payer LM, Wang T, Malik N, Li W, Burns KH, Nath A, Levin HL. Retrotransposon insertions associated with risk of neurologic and psychiatric diseases. *EMBO Rep.* 2023;24(1): e55197.
63. Jahangir M, Li L, Zhou J-S, Lang B, Wang X-P. L1 Retrotransposons: a potential endogenous regulator for schizophrenia. *Front Genet.* 2022;13: 878508.
64. Bundo M, Toyoshima M, Okada Y, Akamatsu W, Ueda J, Nemoto-Miyauchi T, Sunaga F, Toritsuka M, Ikawa D, Kakita A. Increased L1 retrotransposition in the neuronal genome in schizophrenia. *Neuron.* 2014;81(2):306–13.
65. De Donno MD, Puricella A, D'Attis S, Specchia V, Bozzetti MP. Expression of transposable elements in the brain of the *Drosophila melanogaster* model for fragile X syndrome. *Genes (Basel).* 2023;14(5):1060.
66. Guo C, Jeong H-H, Hsieh Y-C, Klein H-U, Bennett DA, De Jager PL, Liu Z, Shulman JM. Tau activates transposable elements in Alzheimer's disease. *Cell Rep.* 2018;23(10):2874–80.
67. Stetler RA, Leak RK, Gan Y, Li P, Zhang F, Hu X, Jing Z, Chen J, Zigmond MJ, Gao Y. Preconditioning provides neuroprotection in models of CNS disease: paradigms and clinical significance. *Prog Neurobiol.* 2014;114:58–83.
68. Rana G, Donizetti A, Virelli G, Piscopo M, Viggiano E, De Luca B, Fucci L. Cortical spreading depression differentially affects lysine methylation of H3 histone at neuroprotective genes and retrotransposon sequences. *Brain Res.* 2012;1467:113–9.
69. Drongitis D, Rainone S, Piscopo M, Viggiano E, Viggiano A, De Luca B, Fucci L, Donizetti A. Epigenetics and cortical spreading depression: changes of DNA methylation level at retrotransposon sequences. *Mol Biol Rep.* 2016;43(8):755–60.
70. Lipszyc A, Szuplewska M, Bartosik D. How do transposable elements activate expression of transcriptionally silent antibiotic resistance genes? *Int J Mol Sci.* 2022;23(15):8063.
71. Muñoz-Lopez M, Vilar-Astasio R, Tristan-Ramos P, Lopez-Ruiz C, Garcia-Pérez JL. Study of transposable elements and their genomic impact. *Transposons Retrotransposons.* 2016;1400:1–19.
72. Reik W. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature.* 2007;447(7143):425–32.
73. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, De Klein A, Wessels L, De Laat W. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature.* 2008;453(7197):948–51.
74. Martens JH, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* 2005;24(4):800–12.
75. Göke J, Lu X, Chan Y-S, Ng H-H, Ly L-H, Sachs F, Szczerbinska I. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell.* 2015;16(2):135–41.
76. Hadjiargyrou M, Delihans N. The intertwining of transposable elements and non-coding RNAs. *Int J Mol Sci.* 2013;14(7):13307–28.
77. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet.* 2019;20(2):89–108.
78. Anzelon TA, Chowdhury S, Hughes SM, Xiao Y, Lander GC, MacRae IJ. Structural basis for piRNA targeting. *Nature.* 2021;597(7875):285–9.
79. Yamashiro H, Siomi MC. PIWI-interacting RNA in *Drosophila*: biogenesis, transposon regulation, and beyond. *Chem Rev.* 2017;118(8):4404–21.

80. Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Tóth KF. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* 2013;27(4):390–9.
81. Clark JP, Lau NC. Piwi Proteins and piRNAs step onto the systems biology stage. *Syst Biol RNA Binding Proteins.* 2014;825:159–97.
82. Pezic D, Manakov SA, Sachidanandam R, Aravin AA. piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev.* 2014;28(13):1410–28.
83. Lindehell H, Schwartz YB, Larsson J. Methylation of lysine 36 on histone H3 is required to control transposon activities in somatic cells. *Life Sci Alliance.* 2023;6(8):e202201832.
84. Iwasaki YW, Murano K, Ishizu H, Shibuya A, Iyoda Y, Siomi MC, Siomi H, Saito K. Piwi modulates chromatin accessibility by regulating multiple factors including histone H1 to repress transposons. *Mol Cell.* 2016;63(3):408–19.
85. Malone CD, Brenneke J, Dus M, Stark A, McCombie WR, Sachidanandam R, Hannon GJ. Specialized piRNA pathways act in germline and somatic tissues of the *Drosophila* ovary. *Cell.* 2009;137(3):522–35.
86. Lippman Z, Gendrel A-V, Black M, Vaughn MW, Dedhia N, Richard McCombie W, Lavine K, Mittal V, May B, Kasschau KD. Role of transposable elements in heterochromatin and epigenetic control. *Nature.* 2004;430(6998):471–6.
87. Yang G, Lee Y-H, Jiang Y, Shi X, Kertbundit S, Hall TC. A two-edged role for the transposable element *Kid* in the rice *ubiquitin2* promoter. *Plant Cell.* 2005;17(5):1559–68.
88. Baba Y, Yagi T, Sawayama H, Hiyoshi Y, Ishimoto T, Iwatsuki M, Miyamoto Y, Yoshida N, Baba H. Long interspersed element-1 methylation level as a prognostic biomarker in gastrointestinal cancers. *Digestion.* 2018;97(1):26–30.
89. Guan Y, Gao H, Leu NA, Vourekas A, Alexiou P, Maragkakis M, Kang Z, Mourelatos Z, Liang G, Wang PJ. The MOV10 RNA helicase is a dosage-dependent host restriction factor for LINE1 retrotransposition in mice. *PLoS Genet.* 2023;19(5): e1010566.
90. Gasparotto E, Burattin FV, Di Gioia V, Panepuccia M, Ranzani V, Marasca F, Bodega B. Transposable elements co-option in genome evolution and gene regulation. *Int J Mol Sci.* 2023;24(3):2610.
91. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature.* 2011;479(7374):534–7.
92. Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sánchez-Luque FJ, Bodea GO, Ewing AD, Salvador-Palomeque C, Van Der Knaap MS, Brennan PM. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell.* 2015;161(2):228–39.
93. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O'Shea KS, Moran JV, Gage FH. L1 retrotransposition in human neural progenitor cells. *Nature.* 2009;460(7259):1127–31.
94. Garza R, Atacho D, Adami A, Gerdes P, Vinod M, Hsieh P, Karlsson O, Horvath V, Johansson PA, Pandiloski N. L1 retrotransposons drive human neuronal transcriptome complexity and functional diversification. *bioRxiv.* Posted March 06, 2023..531072.
95. Muotri AR, Marchetto MC, Coufal NG, Oefner R, Yeo G, Nakashima K, Gage FH. L1 retrotransposition in neurons is modulated by MeCP2. *Nature.* 2010;468(7322):443–6.
96. Thomas CA, Paquola AC, Muotri AR. LINE-1 retrotransposition in the nervous system. *Annu Rev Cell Dev Biol.* 2012;28:555–73.
97. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet.* 2007;8(4):272–85.
98. de Souza FS, Franchini LF, Rubinstein M. Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol.* 2013;30(6):1239–51.
99. Choudhary MNK, Quaid K, Xing X, Schmidt H, Wang T. Widespread contribution of transposable elements to the rewiring of mammalian 3D genomes. *Nat Commun.* 2023;14(1):634.
100. Kirkland JG, Raab JR, Kamakaka RT. TFIIIC bound DNA elements in nuclear organization and insulation. *Biochim Biophys Acta (BBA)-Gene Regul Mech.* 2013;1829(3–4):418–424.
101. Ichiyanagi T, Katoh H, Mori Y, Hirafuku K, Boyboy BA, Kawase M, Ichiyanagi K. B2 SINE copies serve as a transposable boundary of DNA methylation and histone modifications in the mouse. *Mol Biol Evol.* 2021;38(6):2380–95.
102. Lunyak VV, Prefontaine GG, Núñez E, Cramer T, Ju B-G, Ohgi KA, Hutt K, Roy R, García-Díaz A, Zhu X. Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science.* 2007;317(5835):248–51.
103. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.* 2009;19(8):1419–28.
104. Berrens RV, Andrews S, Spensberger D, Santos F, Dean W, Gould P, Sharif J, Olova N, Chandra T, Koseki H. An endosRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell Stem Cell.* 2017;21(5):694–703. e697.
105. Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. *Epigenet Chromatin.* 2018;11(1):1–18.
106. Ecco G, Cassano M, Kaulzaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D. Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues. *Dev Cell.* 2016;36(6):611–23.
107. Osipovich AB, Dudek KD, Trinh LT, Kim LH, Shrestha S, Cartailier JP, Magnuson MA. ZFP92, a KRAB domain zinc finger protein enriched in pancreatic islets, binds to B1/Alu SINE transposable elements and regulates retroelements and genes. *PLoS Genet.* 2023;19(5): e1010729.
108. Brattås PL, Jönsson ME, Fasching L, Wahlstedt JN, Shahsavani M, Falk R, Falk A, Jern P, Parmar M, Jakobsson J. TRIM28 controls a gene regulatory network based on endogenous retroviruses in human neural progenitor cells. *Cell Rep.* 2017;18(1):1–11.
109. Liu N, Lee CH, Swigut T, Grow E, Gu B, Bassik MC, Wysocka J. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature.* 2018;553(7687):228–32.
110. Du J, Leung A, Trac C, Lee M, Parks BW, Lusic AJ, Natarajan R, Schones DE. Chromatin variation associated with liver metabolism is mediated by transposable elements. *Epigenet Chromatin.* 2016;9(1):1–16.
111. Trizzino M, Kapusta A, Brown CD. Transposable elements generate regulatory novelty in a tissue-specific fashion. *BMC Genomics.* 2018;19(1):1–12.
112. Lee YCG, Karpen GH. Pervasive epigenetic effects of *Drosophila* euchromatic transposable elements impact their evolution. *Elife.* 2017;6: e25762.
113. Guio L, Vieira C, González J. Stress affects the epigenetic marks added by natural transposable element insertions in *Drosophila melanogaster*. *Sci Rep.* 2018;8(1):1–10.
114. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet.* 2012;46(1):21–42.
115. Ishizaka Y, Okudaira N, Tamura M, Iijima K, Shimura M, Goto M, Okamura T. Modes of retrotransposition of long interspersed element-1 by environmental factors. *Front Microbiol.* 2012;3:191.
116. Mourier T, Nielsen LP, Hansen AJ, Willerslev E. Transposable elements in cancer as a by-product of stress-induced evolvability. *Front Genet.* 2014;5:156.
117. Secco D, Wang C, Shou H, Schultz MD, Chiarenza S, Nussaume L, Ecker JR, Whelan J, Lister R. Stress induced gene expression drives transient DNA methylation changes at adjacent repetitive elements. *Elife.* 2015;4:e09343.
118. Nikitin D, Penzar D, Garazha A, Sorokin M, Tkachev V, Borisov N, Poltorak A, Prassolov V, Buzdin AA. Profiling of human molecular pathways affected by retrotransposons at the level of regulation by transcription factor proteins. *Front Immunol.* 2018;9:30.
119. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet.* 2008;9(5):397–405.
120. Nishihara H, Kobayashi N, Kimura-Yoshida C, Yan K, Bormuth O, Ding Q, Nakanishi A, Sasaki T, Hirakawa M, Sumiyama K. Coordinately co-opted multiple transposable elements constitute an enhancer for *wnt5a* expression in the mammalian secondary palate. *PLoS Genet.* 2016;12(10): e1006380.

121. Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* 2017;27(10):1623–33.
122. Jacques P-É, Jeyakani J, Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet.* 2013;9(5): e1003504.
123. Lee HJ, Hou Y, Maeng JH, Shah NM, Chen Y, Lawson HA, Yang H, Yue F, Wang T. Epigenomic analysis reveals prevalent contribution of transposable elements to cis-regulatory elements, tissue-specific expression, and alternative promoters in zebrafish. *Genome Res.* 2022; 32(7):1424–36.
124. Lowe CB, Haussler D. 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. 2012;7(8):e43128.
125. Sundaram V, Wysocka J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. *Philos Trans R Soc B.* 2020;375(1795):20190347.
126. Conley AB, Piriyaoponga J, Jordan IK. Retroviral promoters in the human genome. *Bioinformatics.* 2008;24(14):1563–7.
127. Franchini LF, López-Leal R, Nasif S, Beati P, Gelman DM, Low MJ, de Souza FJ, Rubinstein M. Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retroposons. *Proc Natl Acad Sci.* 2011;108(37):15270–5.
128. Brosius J, Gould SJ. On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA". *Proc Natl Acad Sci.* 1992;89(22):10706–10.
129. Gould SJ, Vrba ES. Exaptation—a missing term in the science of form. *Paleobiology.* 1982;8(1):4–15.
130. Brini AT, Lee GM, Kinet J-P. Involvement of Alu sequences in the cell-specific regulation of transcription of the gamma chain of Fc and T cell receptors. *J Biol Chem.* 1993;268(2):1355–61.
131. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, James Kent W, Haussler D. A distal enhancer and an ultra-conserved exon are derived from a novel retroposon. *Nature.* 2006;441(7089):87–90.
132. Kellner M, Makalowski W. Transposable elements significantly contributed to the core promoters in the human genome. *Sci China Life Sci.* 2019;62(4):489–97.
133. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* 2014;24(12):1963–76.
134. Zeng L, Pederson SM, Cao D, Qu Z, Hu Z, Adelson DL, Wei C. Genome-wide analysis of the association of transposable elements with gene regulation suggests that alu elements have the largest overall regulatory impact. *J Comput Biol.* 2018;25(6):551–62.
135. Samuelson LC, Wiebauer K, Snow C, Meisler MH. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol Cell Biol.* 1990;10(6):2513–20.
136. Hambor J, Mennone J, Coon M, Hanke J, Kavathas P. Identification and characterization of an Alu-containing, T-cell-specific enhancer located in the last intron of the human CD8 alpha gene. *Mol Cell Biol.* 1993;13(11):7056–70.
137. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, Köchert K, Bouhleh MA, Richter J, Soler E. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med.* 2010;16(5):571–9.
138. Pehrsson EC, Choudhary MN, Sundaram V, Wang T. The epigenomic landscape of transposable elements across normal human development and anatomy. *Nat Commun.* 2019;10(1):1–16.
139. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science.* 2016;351(6277):1083–7.
140. Rodriguez-Terrones D, Torres-Padilla M-E. Nimble and ready to mingle: transposon outbursts of early development. *Trends Genet.* 2018;34(10):806–20.
141. Flehr M, Malik R, Franke V, Nejepsinska J, Sedlacek R, Vlahovicek K, Svoboda P. A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. *Cell.* 2013;155(4):807–16.
142. Sakashita A, Maezawa S, Takahashi K, Alavattam KG, Yukawa M, Hu Y-C, Kojima S, Parrish NF, Barski A, Pavlicev M. Endogenous retroviruses drive species-specific germline transcriptomes in mammals. *Nat Struct Mol Biol.* 2020;27(10):967–77.
143. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A. Proteomics. Tissue-based map of the human proteome. *Science (New York, NY).* 2015;347(6220):1260419–1260419.
144. Simonti CN, Pavličev M, Capra JA. Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. *Mol Biol Evol.* 2017;34(11):2856–69.
145. Rayan NA, Del Rosario RC, Prabhakar S. Massive contribution of transposable elements to mammalian regulatory sequences. *Seminars in cell & developmental biology.* 2016;57:51–6.
146. Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526(7571):68.
147. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA. SVA elements: a hominid-specific retroposon family. *J Mol Biol.* 2005;354(4):994–1007.
148. Batzer MA, Deininger PL. A human-specific subfamily of Alu sequences. *Genomics.* 1991;9(3):481–7.
149. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci.* 2003;100(9):5280–5.
150. Daniel C, Lagergren J, Öhman M. RNA editing of non-coding RNA and its role in gene regulation. *Biochimie.* 2015;117:22–7.
151. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature.* 2011;470(7333):284–8.
152. Lucas BA, Lavi E, Shiue L, Cho H, Katzman S, Miyoshi K, Siomi MC, Carmel L, Ares M Jr, Maquat LE. Evidence for convergent evolution of SINE-directed Staufen-mediated mRNA decay. *Proc Natl Acad Sci.* 2018;115(5):968–73.
153. Kelley DR, Hendrickson DG, Tenen D, Rinn JL. Transposable elements modulate human RNA abundance and splicing via specific RNA-protein interactions. *Genome Biol.* 2014;15(12):1–16.
154. Kim EZ, Wespiser AR, Caffrey DR. The domain structure and distribution of Alu elements in long noncoding RNAs and mRNAs. *RNA.* 2016;22(2):254–64.
155. Kralovicova J, Moreno PM, Cross NC, Pêgo AP, Vorechovsky I. Antisense oligonucleotides modulating activation of a nonsense-mediated RNA decay switch exon in the ATM gene. *Nucleic Acid Therapeut.* 2016;26(6):392–400.
156. Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. *Nat Rev Genet.* 2009;10(3):155–9.
157. Ma L, Bajic VB, Zhang Z. On the classification of long non-coding RNAs. *RNA Biol.* 2013;10(6):924–33.
158. Fernandes JC, Acuña SM, Aoki JI, Floeter-Winter LM, Muxel SM. Long non-coding RNAs in the regulation of gene expression: physiology and disease. *Non-coding RNA.* 2019;5(1):17.
159. Ransohoff JD, Wei Y, Khavari PA. The functions and unique features of long intergenic non-coding RNA. *Nat Rev Mol Cell Biol.* 2018;19(3):143–57.
160. Azlan A, Dzaki N, Azzam G. Argonaute: the executor of small RNA function. *J Genet Genomics.* 2016;43(8):481–94.
161. Piriyaoponga J, Jordan IK. A family of human microRNA genes from miniature inverted-repeat transposable elements. *PLoS ONE.* 2007;2(2): e203.
162. Mustafin RN, Khusnutdinova E. Perspective for studying the relationship of miRNAs with transposable elements. *Curr Issues Mol Biol.* 2023;45(4):3122–45.
163. Yuan Z, Sun X, Jiang D, Ding Y, Lu Z, Gong L, Liu H, Xie J. Origin and evolution of a placental-specific microRNA family in the human genome. *BMC Evol Biol.* 2010;10(1):1–12.
164. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 2011;25(18):1915–27.
165. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, Yandell M, Feschotte C. Transposable elements are major contributors to

- the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* 2013;9(4): e1003470.
166. Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* 2012;13(11):1–14.
 167. Johnson R, Guigó R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA.* 2014;20(7):959–76.
 168. Kapusta A, Feschotte C. Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications. *Trends Genet.* 2014;30(10):439–52.
 169. Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, Bonetti A, Voineagu I, Bertin N, Kratz A. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet.* 2014;46(6):558–66.
 170. Jeck WR, Sorrentino JA, Wang K, Slevin MK, Burd CE, Liu J, Marzluff WF, Sharpless NE. Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA.* 2013;19(2):141–57.
 171. Zhang X-O, Wang H-B, Zhang Y, Lu X, Chen L-L, Yang L. Complementary sequence-mediated exon circularization. *Cell.* 2014;159(1):134–47.
 172. Chen L-L, Yang L. Regulation of circRNA biogenesis. *RNA Biol.* 2015;12(4):381–8.
 173. Chen L, Zhang P, Fan Y, Lu Q, Li Q, Yan J, Muehlbauer GJ, Schnable PS, Dai M, Li L. Circular RNAs mediated by transposons are associated with transcriptomic and phenotypic variation in maize. *New Phytol.* 2018;217(3):1292–306.
 174. Jung J, Lee S, Cho H-S, Park K, Ryu J-W, Jung M, Kim J, Kim H, Kim D-S. Bioinformatic analysis of regulation of natural antisense transcripts by transposable elements in human mRNA. *Genomics.* 2019;111(2):159–66.
 175. Shin C, Nam J-W, Farh KK-H, Chiang HR, Shkumatava A, Bartel DP. Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell.* 2010;38(6):789–802.
 176. Spengler RM, Oakley CK, Davidson BL. Functional microRNAs and target sites are created by lineage-specific transposition. *Hum Mol Genet.* 2014;23(7):1783–93.
 177. Gong C, Maquat LE. “Alu” strious long ncRNAs and their roles in shortening mRNA half-lives. *Cell Cycle.* 2011;10(12):1882–3.
 178. Priyapongsa J, Mariño-Ramírez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. *Genetics.* 2007;176(2):1323–37.
 179. Qin S, Jin P, Zhou X, Chen L, Ma F. The role of transposable elements in the origin and evolution of microRNAs in human. *PLoS ONE.* 2015;10(6): e0131365.
 180. Mattick JS. Deconstructing the dogma: a new view of the evolution and genetic programming of complex organisms. *Ann N Y Acad Sci.* 2009;1178(1):29–46.
 181. Kang D, Kim Y-J, Hong K, Han K. TE composition of human long non-coding RNAs and their expression patterns in human tissues. *Genes Genomics.* 2015;37(1):87–95.
 182. Chishima T, Iwakiri J, Hamada M. Identification of transposable elements contributing to tissue-specific expression of long non-coding RNAs. *Genes.* 2018;9(1):23.
 183. Kannan S, Chernikova D, Rogozin IB, Poliakov E, Managadze D, Koonin EV, Milanese L. Transposable element insertions in long intergenic non-coding RNA genes. *Front Bioeng Biotech.* 2015;3:71.
 184. Kitano S, Kurasawa H, Aizawa Y. Transposable elements shape the human proteome landscape via formation of cis-acting upstream open reading frames. *Genes Cells.* 2018;23(4):274–84.
 185. Chen G, Wang R, Jiang Y, Dong X, Xu J, Xu Q, Kan Q, Luo Z, Springer NM, Li Q. A novel active transposon creates allelic variation through altered translation rate to influence protein abundance. *Nucleic Acids Res.* 2023;51(2):595–609.
 186. Shen J, Liu J, Xie K, Xing F, Xiong F, Xiao J, Li X, Xiong L. Translational repression by a miniature inverted-repeat transposable element in the 3′ untranslated region. *Nat Commun.* 2017;8(1):1–10.
 187. Volff JN. Cellular genes derived from Gypsy/Ty3 retrotransposons in mammalian genomes. *Ann N Y Acad Sci.* 2009;1178(1):233–43.
 188. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, Yoder N, Belnap DM, Erlendsson S, Morado DR. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell.* 2018;172(1–2):275–288. e218.
 189. Hudecek M, Izsvák Z, Johnen S, Renner M, Thumann G, Ivics Z. Going non-viral: the sleeping beauty transposon system breaks on through to the clinical side. *Crit Rev Biochem Mol Biol.* 2017;52(4):355–80.
 190. Taylor K, Yau HL, Chakravarthy A, Wang B, Shen SY, Ettayebi I, Ishak CA, Bedard PL, Razak AA, Hansen AR. An open-label, phase II multicohort study of an oral hypomethylating agent CC-486 and durvalumab in advanced solid tumors. *J Immunother Cancer.* 2020;8(2):e000883.
 191. Hahn S, Ugurel S, Hanschmann K-M, Strobel H, Tondera C, Schadendorf D, Löwer J, Löwer R. Serological response to human endogenous retrovirus K in melanoma patients correlates with survival probability. *AIDS Res Hum Retroviruses.* 2008;24(5):717–23.
 192. Anwar SL, Wulaningsih W, Lehmann U. Transposable elements in human cancer: causes and consequences of deregulation. *Int J Mol Sci.* 2017;18(5):974.
 193. Zamudio N, Barau J, Teissandier A, Walter M, Borsos M, Servant N, Bourc’his D. DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes Dev.* 2015;29(12):1256–70.
 194. Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. *RNA.* 2017;23(12):1754–69.
 195. Hwang S-Y, Jung H, Mun S, Lee S, Park K, Baek SC, Moon HC, Kim H, Kim B, Choi Y. L1 retrotransposons exploit RNA m6A modification as an evolutionary driving force. *Nat Commun.* 2021;12(1):1–14.
 196. Pacini CE, Bradshaw CR, Garrett NJ, Koziol MJ. Characteristics and homogeneity of N6-methylation in human genomes. *Sci Rep.* 2019;9(1):1–12.
 197. Koziol MJ, Bradshaw CR, Allen GE, Costa AS, Frezza C. Identification of methylated deoxyadenosines in genomic DNA by dA6m DNA immunoprecipitation. *Bio-Protoc.* 2016;6(21):e1990–e1990.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

