Heliyon 7 (2021) e06029

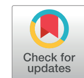
Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Review article

Transposable elements, contributors in the evolution of organisms (from an arms race to a source of raw materials)



Fatemeh Pourrajab [a,b,*], Seyedhossein Hekmatimoghaddam [c,d,**]

[a] Nutrition and Food Security Research Center, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
[b] Department of Biochemistry and Molecular Biology, School of Medicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
[c] Department of Advanced Medical Sciences and Technologies, School of Paramedicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran
[d] Department of Laboratory Sciences, School of Paramedicine, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

ARTICLE INFO

Keywords:
Transposable elements
Genomes
Species evolution
Non-coding RNAs

ABSTRACT

There is a concept proposing that the primitive lineages of prokaryotes, eukaryotes, and viruses emerged from the primordial pool of primitive genetic elements. In this genetic pool, transposable elements (TEs) became a source of raw material for primitive genomes, tools of genetic innovation, and ancestors of modern genes (e.g. ncRNAs, tRNAs, and rRNAs). TEs contributed directly to the genome evolution of three forms of life on the earth. TEs now appear as tools that were used to giving rise to sexual dimorphism and sex determination, lineage-specific expression of genes and tissue differentiation and finally genome stability and lifespan determination.

1. Introduction

Data represent a concept proposing that the primitive lineages of prokaryotes, eukaryotes, and viruses emerged from the primordial pool of primitive genetic elements, the ancestors of both cellular and viral genes [1, 2, 3]. In this pool, the emergence of transposable elements (TEs) and their substantial genetic diversity antedates the advent of full-fledged genomes, allowing for extensive gene mixing at this early stage of evolution. Herein, there is evidence describing stem-loop hairpin RNAs and palindrome-forming sequences as the origin of TEs, as well as, the ancestors of current and ancient parasitic and non-parasitic elements. Now, TEs are putatively viewed as a raw material of primitive genomes, tools of genetic innovation, and ancestors of modern genes (e.g. ncRNAs, tRNAs, and rRNAs) (Figure 1). It can be assumed from the literature that interactions in the host–TE system, in the evolutionary processes, led to the evolution of increasing organizational complexity in life such as the emergence of unicellular and multi-cellular organisms [4, 5, 6, 7]. Thus at the early stage of life evolution, varieties of TEs were existed and co-opted to mitigate evolutionary conflicts in the host genome, as an unexpectedly requisite and inevitable evolutionary process for cellular function and protein formation. Cells needed TEs as defense systems that protect their genomes against the proliferation of infectious or invasive agents and stress-full conditions "as an arms race" [8, 9, 10, 11]. Thereafter, most copies of TEs became inactive in genomes, since their transpositions produce only detrimental effects on organisms [4, 12, 13]. Data briefly propose a new, coherent scenario for viruses and cellular life forms as "The ancient Virus World and evolution of cells" that is best compatible with comparative-genomic analysis and naturally linked to models of cellular evolution (Figure 1). Under this concept, the principal lineages of viruses and related selfish agents emerged from the primordial pool of primitive genetic elements, the ancestors of both cellular and viral genes. Thus, the numerous gene exchanges and acquisitions attained at the later stages of evolution, where modern viruses and other cellular life were inferred to descend from primary elements that belonged to the primordial genetic pool [3, 14] (see Table 1).

2. Transposable elements (TEs) & genome evolution

Accordingly, "The history of TEs and their role in genomes provides one of the best examples of how scientific concepts in biology emerge and then evolve into new concepts". Now, TEs emerge to play major roles in shaping genomes in evolution and the speciation [3, 4, 11, 12]. TEs are present and act as the primary contributors to the bulk and structure of the genomic DNA in all forms of life (e.g. ~ 90% and ~45–69% of plant and human genomes, respectively) [15, 16, 17]. A body of works has indicated that TEs constitute large fractions of genomes and might act as

* Corresponding author.
** Corresponding author.
E-mail addresses: mina_poorrajab@yahoo.com (F. Pourrajab), shhekmati@yahoo.com (S. Hekmatimoghaddam).

https://doi.org/10.1016/j.heliyon.2021.e06029
Received 13 August 2020; Received in revised form 8 October 2020; Accepted 13 January 2021
2405-8440/© 2021 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

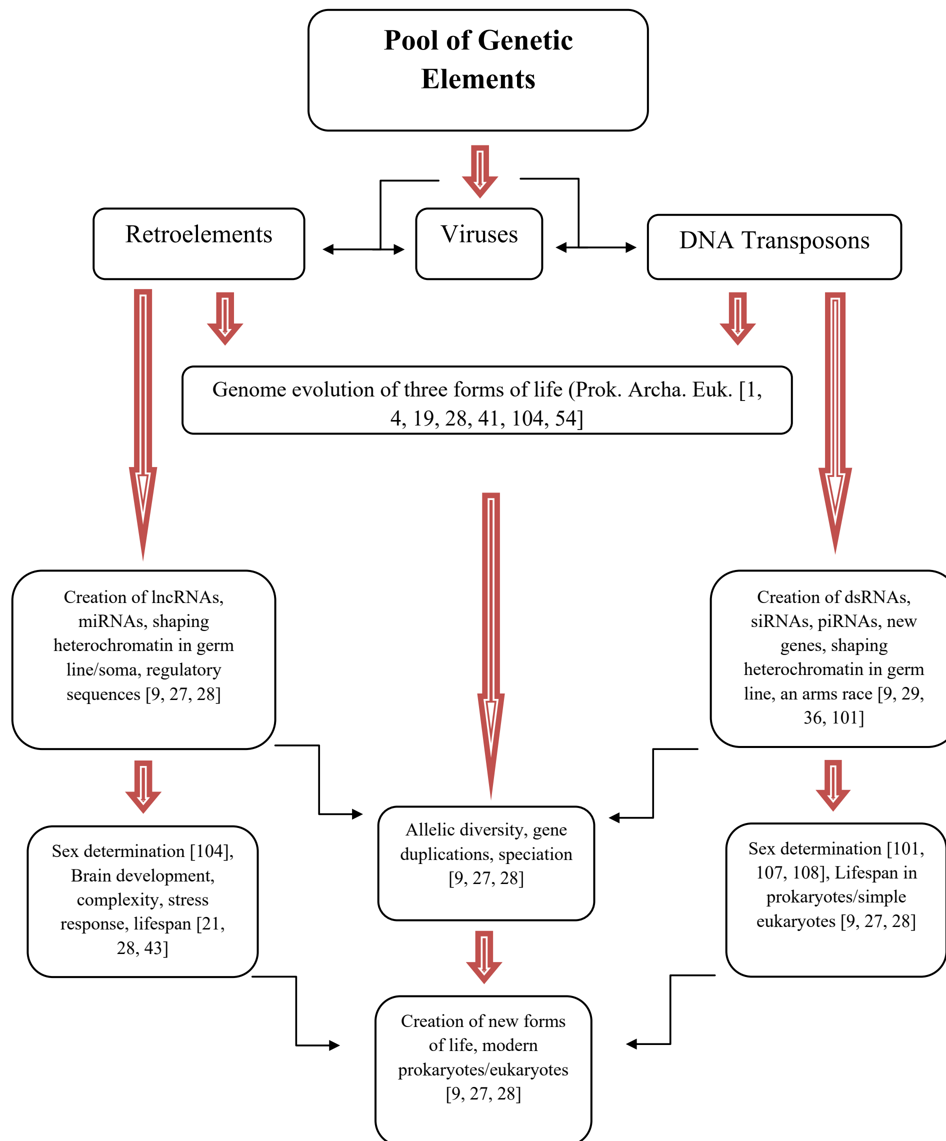


Figure 1. Origin and evolution of viruses and three forms of life on the earth (Prokaryotes, Archaea and Eukaryotes) according to the concept of primordial pool of primitive genetic elements, as a source of raw material participating to form ancestors of modern genes and genomes.

Table 1. Some small RNAs originated from or enriched in TEs, are mentioned here.

Small RNA	TE origin/Enriched element	Organism/Identified cluster/location
hsa-mir-548 family	*MITEs* (DNA-type elements)	Primate-specific miRNA family/Human Chr 6, 8, X, ...
C19MC	*Alu* elements	Primate-specific miRNA clusters/Chr 19
miR-371/372/373[*]	*Alu* elements	Human/*Alu*-enriched cluster/Chr 19
miR-513/506/507/508/509/510/514	*Alu* elements	Primate-specific/*Alu*-enriched miR506-514 cluster/Chr X
miR-892c/890/888/892a/892b/891b/891a	*Alu* elements	Primates/*Alu*-enriched miR-888 cluster/Chr X

Abbreviations: *Miniature inverted-repeat transposable* elements (*MITEs*); chromosome 19 miRNA cluster (C19MC); Chromosome 19 (Chr 19); Chromosome X (Chr X).
[*] Homologous to murine *miR-290/291a/291b/292/293/295* on Chr 7.

powerful facilitators of genome evolution to generate genetic novelties in both an active mode and a passive mode. A model of early evolution has predicted that the emergence of TEs in conjunction with main classes of viruses antedates “the advent of full-fledged cells” (Figure 1). The model has proposed a pre-cellular evolution scenario, under which the selfish genetic elements (TEs) evolved prior to typical cells (before bacteria and archaea had been arrived at the scene). The primordial gene pool was formed in a network of inorganic compartments, allowing for extensive gene mixing which led to substantial genetic diversity, and became ancestral to viruses which led to the early evolution of the RNA interference mechanism against virus-like agents [1, 2, 3, 6, 8]. Reports also highlight the role of TEs in the initial steps of differentiation and evolution of sex chromosomes. There is a strong link between TE accumulation and the emergence of the sex-determination (SD) locus [18, 19].

Herein from the literature, it is hypothesized that TEs existed in genomes and shaped the "Last universal common ancestor" first ancestors,

arose once, and then contribute to the evolution of kingdoms (viruses and cellular life forms: prokaryotic and then eukaryotic cells, multi-cellular organisms, and eusocial animals) [15, 16, 17, 18, 19].

TEs are now appeared to be major players in genome evolution, SD, tissue differentiation, and lifespan expansion [3, 6, 15, 19].

TEs create two distinguished classes; class I elements, or retrotransposons which using reverse transcriptase to copy themselves and divided into Long Terminal Repeat (LTR, e.g. the *human endogenous retroviruses* (*HERVs*) and non-LTR elements [15, 18]. Retroviruses are largely restricted to vertebrates, in particular mammals [8]. It is estimated that retroelement-derived sequences make up over 50% of mammalian genomes (mostly non-LTR elements) and over 75% of some plant genomes, e.g. maize [1, 8]. Even though retroelements are present in prokaryotes, but they are extremely abundant in eukaryotic species and show more enormous variation in them [2].

Class II elements include DNA transposons and can mobilize in genomes by the "cut and paste" mechanism. The two classes subdivide further into super-families and then into families based on the transposition mechanism, sequence similarities, and structural relationships [15, 18].

Differential accumulation and specific combinations of different types of TEs have contributed to the evolution of organisms. The most variation (~57%) exists in flowering plants wherein the lineage of Ty3/gypsy LTR-retrotransposons and driven elements play major roles [20]. However, the specific integrations of TEs with certain patterns of insertion preferences in the genomes indicate that the evolutionary events of TEs and their insertions occurred before the emergence of eukaryotic groups [15, 19, 21]. For example, cis-regulatory elements in the genome contain specific integrations of TEs in promoter regions, in sex-determination regions (SDRs) on XY.ZW and in X inactivation center (XIC) [19, 21, 22, 23, 24]. Theorizing, TEs were opted and inserted as 'domesticated agents' in the genome to serve cellular function beneficial to the host organism for example for adaptation in conflicts of evolutionary events [9, 10, 11].

Once inserted, TEs fixed strictly in populations and their abundance became controlled by RNA interfering (RNAi) mechanisms, as well as, by inhibitor proteins (e.g., KP repressor of *Drosophila P* element and IS50 in bacteria). There is a balance between the forces of transposition (increasing element abundance) and the action of purifying natural selection at the host level, which then removing individuals with high copy number [25, 26, 27]. There is no convincing evidence for the horizontal transfer of TEs between eukaryotic groups [9].

In the crosses between specific lineages of animals, for instance, *Drosophila melanogaster* (*D. melanogaster)*, genetic mutations, and sterility occur, all of which are a result of the mobilization of specific TEs and the emergence of hybrid dysgenesis phenomenon. The deleterious effect of the mobilization of the *P* and *I* elements in *Drosophila* is a result of crosses [3, 4]. To prevent the deleterious effect of TEs mobilization RNAi strategy was opted in evolution to maintain genomes largely constant throughout life [28]. Most importantly, germ cells prevent TEs activation and keep the genome unchanged through generations by effectively silencing them via the effective RNAi pathway Piwi-piRNA besides using other RNAi systems [25, 29].

2.1. TEs and evolution of viruses

Accordingly, viruses and selfish genetic elements (TEs) are predominant entities in the biosphere, concerning both physical abundance and genetic diversity [1]. According to the theory of "The ancient Virus World and evolution of cells", there was an ancient pool of genes and genomes, where the evolution of viruses and cells formed distinctly and retained their identity continuously throughout the whole history of life. Under this new and coherent scenario for the evolution of viruses and cellular life forms, TEs might be ancestors of both viruses and cellular life, evolved before the typical modern cells (Figure 1). There was an emergence of primordial viral ancestors under the emergence of cellular genomes within networks of inorganic compartments [2, 3, 5]. TEs and viruses share common features in their genome structures and biochemical abilities; however, evolution has put the RNAi mechanism to tightly inhibit TE mobilization and transposition in the host genome [1, 2, 4]. Additionally, short stem-loop hairpin-forming palindromic sequences are present at the origins of all genomes, TEs, and the forms of parasitism [1, 2, 5]. Speculating that under certain conditions and at multiple times, DNA transposons and retro-virus-like elements were arisen independently in the kingdoms of the primordial pool of primitive genetic elements, and their horizontal transfer led to the origination of viruses with structural and replicative gene modules along with additional acquisitions of diverse genes (Figure 1) [1, 3, 4, 30]. The most remarkable aspect of the evolution of viruses by TEs is that they can be tractable in host TEs, at least in the central features [1, 8]. The similarity that exists between TEs and viruses in their encoding sequences, e.g. reverse transcriptase-like sequences, clearly reveal the primitive link between TEs and viruses [3, 6, 14]. Postulating, the great majority of eukaryotic RNA viruses originated from the ancestor of the picorna-like viruses which assembled from the primordial gene pool "the hypothetical primordial RNA world" [1, 8].

On the other side, self-synthesizing transposons are explained as a primary stage in the evolution of viruses [5, 6, 14]. Such an example is the Polinton/Maverick family of self-synthesizing transposons widespread in eukaryotes and abundant in the genomes of some protists. Comparative genomic analysis of polintons, polinton-like viruses (PLV), and the other viruses with double-stranded (ds) DNA genomes infecting eukaryotes and prokaryotes have exhibited that the polintons could be the ancestors of a broad range of eukaryotic viruses including adenoviruses and members of the order "Megavirales" as well as linear cytoplasmic plasmids. Polintons are proposed to be able to alternate between the transposon and viral lifestyles. Host-like TEs are now appeared to have made a major contribution to the evolution of all classes of viruses as well as the hosts [1, 2, 14].

The classes of viruses emerge dramatically different between prokaryotes and eukaryotes. Most importantly, the host ranges of virus super-families are exclusively limited to host elements with relatively close evolutionary origins and close integration of sequences [1, 30]. Retroviruses are present among all placental mammals, are largely restricted to vertebrates, and are particularly abundant in mammals. The narrow host ranges of RNA viruses, limited to animals and plants, imply relatively co-evolutionary origin [1, 8]. Herein, most viruses that can infect prokaryotic cells possess double-stranded (ds) DNA genomes with a substantial minority of single-stranded (ss) DNA viruses. In contrast, in eukaryotes, RNA viruses account for the majority of the virome diversity although ssDNA and dsDNA viruses are common as well. These are clues for the origins of major classes of prokaryotic/eukaryotic viruses and in particular, their likely roots in de novo assembled from TEs [1, 2].

2.2. TEs in the evolution of eukaryotes

One of the direct contributions of TEs in the evolution of eukaryotes is as a source of raw material used for the assembly of new genes and functions of genomes. Evidence indicates that TEs are one of the primary determinants of genome size differences among all forms of life (Figure 2A) [9,28,31]. It is estimated that TEs occupy between 37- 55% of mammalian genomes (mostly *Short interspersed nuclear elements; SINEs*) [32]. In plants, TEs render the plant genome sizes to a wide range of variation spanning several orders of magnitude (e.g. 85-65% in maize and rice, respectively). In the animal kingdom, the proportion of TEs may vary up to 77% in the frog *Pelophylax esculentus* with a relatively big genome of 5.5–7.8 Gb. The giant size of the salamander genomes, ranging from 14-74 Gb, is corresponded to the massive numbers of TEs (>78%) especially dominated by LTRs [7, 18, 33, 34]. The percentage of TEs in the genome of invertebrates such as the fruit fly *D. melanogaster* (15–22%) and the worm *Caenorhabditis elegans (C. elegans)* (12%), differed dramatically (Figure 2A) [7, 18].

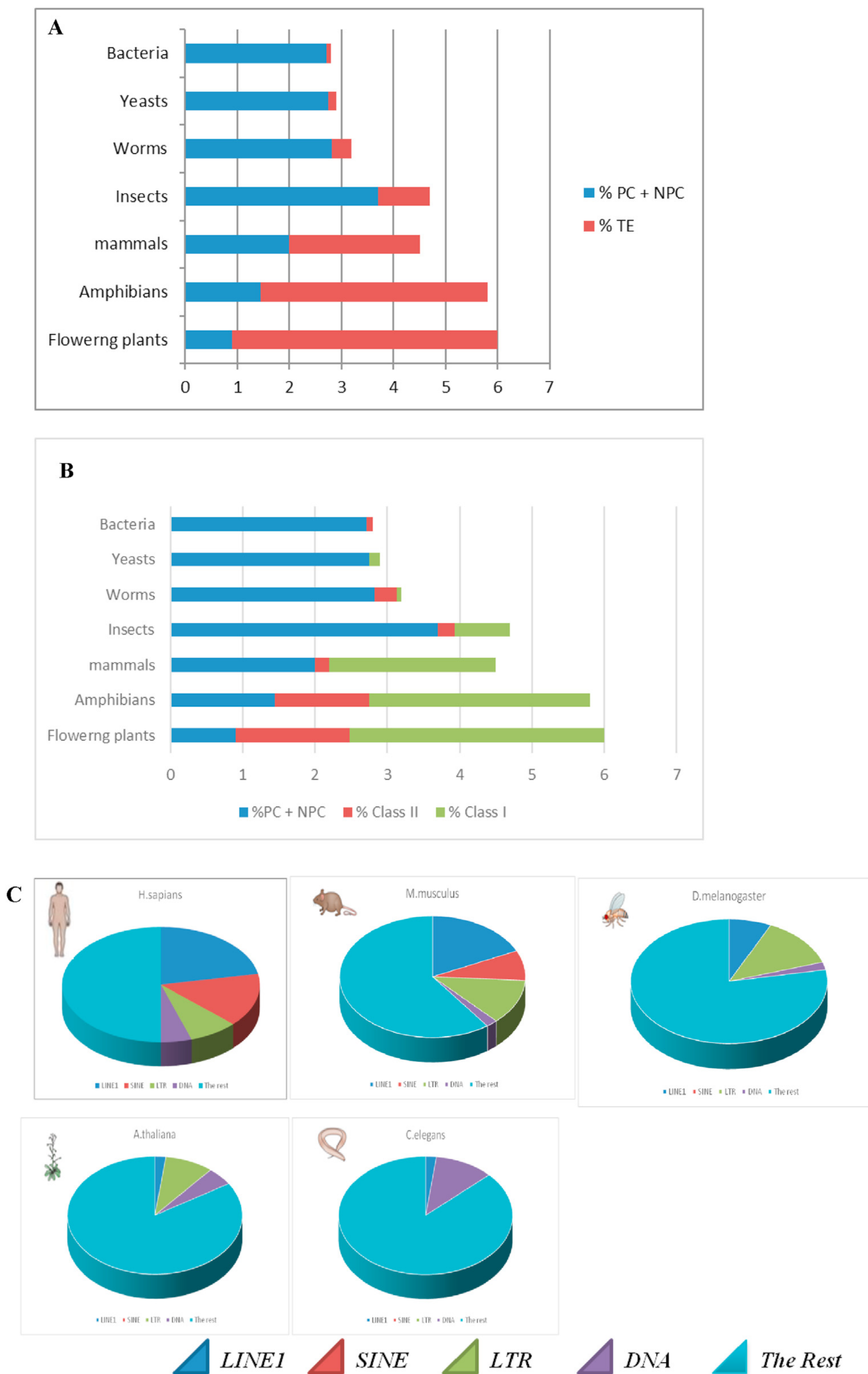
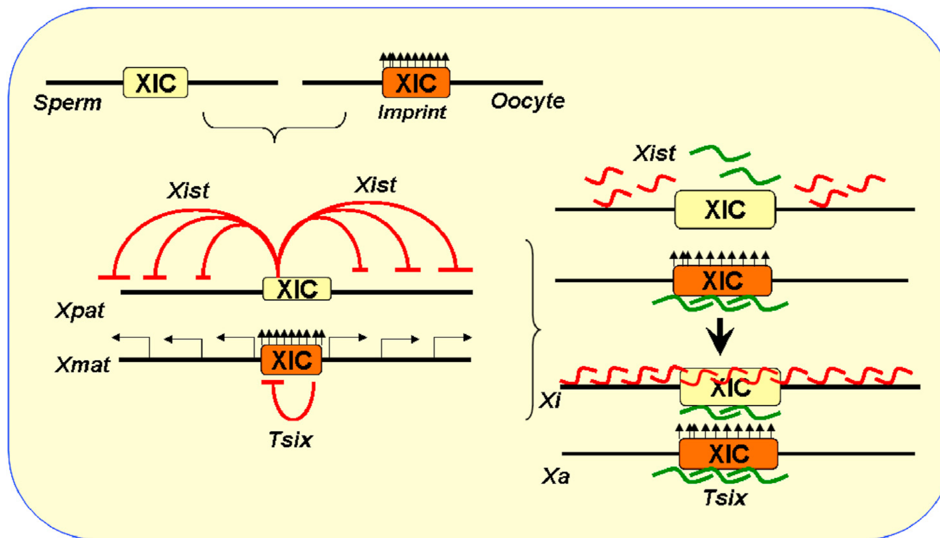


Figure 2. A) The graph tries to illustrating the relative C-values (the haploid genome size) of the genomes among different kingdoms, and the relative proportions of (protein + non-protein coding sequences, (PC)) and transposable elements (TEs). B) The relative proportions of PC, retrotransposons (class I) and DNA transposons (class II) in diverse eukaryotic genomes and bacteria [9, 39]. C) Variation and complexity of species is at least due TE compositions, simply exemplified here. Indeed, the genome size variations observed in multi-cellular eukaryotes are due to retrotransposons, mainly LTRs.

Eukaryotic genomes typically are much larger than prokaryotic genomes due to various accumulations of retroelements in their genomes (Figure 2A and B) [20]. Retroelements are the major elements in increasing the genome size of multi-cellular eukaryotes such as maize and the rice *Oryza australiensis* (Figure 2B) [9].

Notable, DNA transposons are widely abundant in the genome of prokaryotes and single-celled eukaryotes that might be evolved at the early stage of evolution of primordial genomes (Figure 2B) [35, 36, 37]. In contrast to prokaryotes and archaea, diverse eukaryotic genomes are replete with retroelements of different varieties. It is supposed that there is a relationship between TE profile of a genome and the organism development complexity (Figure 2B) [6, 16].

Believing, the most striking differences between eukaryotic species are TEs types of genomes and their abundance. As an example, the human



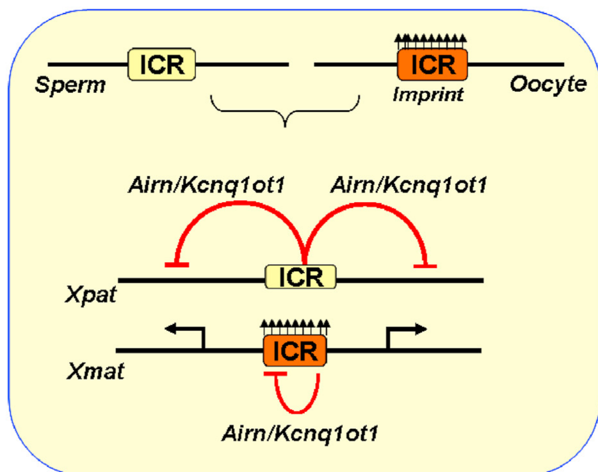


Figure 3. Examples of TE families act as functional domains in the lncRNA transcripts, or master-regulatory regions associated with tissue specificity function of lncRNAs. In mammals, the lncRNAs *Xist/Tsix* **(A)** and *Igf2r/Airn, Kcnq1, Snprn/Ube3A* **(B)** and their master regulatory regions are evolutionary derived from TEs. Both X inactivation center (XIC) and imprinting control region (ICR) are composed of TEs-derived cis-acting elements methylated in the oocyte vs the sperm. **A)** XIC is maternally methylated but paternally un-methylated. The imprints of XIC prevent X to express lncRNA *Xist* which silencing genes in cis, while the paternal un-methylated XIC facilitates the early embryonic expression of both lncRNAs *Xist/Tsix* by its promoter being organized in active epigenetic form. *Tsix* is an antisense transcript of *Xist*. Both *Xist/Tsix* counterparts are paradoxically essential in the mutually exclusive choice of X activation (Xa) and X inactivation (Xi). In female ES cells, in the pre- X chromosome inactivation (XCI) state, *Tsix* RNA is expressed from both Xs and spanning a 40-kb region overlapping *Tsix* and *Xist* loci and keep them in an open chromatin state. At the onset of cell differentiation, the Xi-elect loses *Tsix* expression because of active *Xist* expression and its negative effect on the 40-kb domain to losing euchromatic marks. **B)** Similarly, maternally imprinted gene clusters (e.g. *Igf2r/Airn*, *Kcnq1*, *Snprn/Ube3A*) are harboring lncRNAs originating within or near ICRs and inducing imprinting in cis, similar to *Xist/Tsix* in XCI. Maternally methylated ICRs harbors the promoter for lncRNAs such as *Kcnq1ot1*, *Snprn*, and *Airn*, then un-methylated paternal allele (resembling *Xist*) is expressed and silence genes in cis. In the *Snrpn* locus, *Ube3a* is expressed from the maternal allele (resembling *Tsix*).

genome comprises more than 50% TEs, mostly LINEs and SINEs, while the genome of the nematode *C. elegans* has nearly 12% DNA TEs (Figure 2C) [4, 9]. Variation of species is at least due to TE compositions, simply exemplified by the four species of single-celled eukaryotes *Entamoeba.* The genomes of *Entamoeba invadens (E. invadens)* and *Entamoeba moshkovskii* (*E. moshkovskii)* host many families of DNA transposons and are distantly related to fungi, while the genomes of *E. histolytica* and *E. dispar* contain virtually no DNA transposons but instead colonized by several lineages of non-LTR retrotransposons and are distantly related to animals. *Entamoeba* genomes, however, are composed of the same relative proportion of TEs (5%–7%) [9].

In the zebrafish (*Danio rerio*) and in the nematode (*C. elegans*) genomes, respectively more than 75% and 95% of TEs are estimated to be DNA transposons, whereas, in the fruit fly (*D. melanogaster*), ~90% are retrotransposons where LTRs are the most abundant TEs (>50%) (Figure 2C) [38, 39]. Flowering plants have got the most variation in the genome size, and thereafter exists amphibians and insects that show the most variations in the genome size. However, mammals exhibit less variability in the genome size (Figure 2A) [18].

Moreover, TEs are also involved in SD (e.g. in animals) and accumulating in the SD locus. They also influence the kind of mating system (e.g. self-crosses or out-crosses in plants), where *Arabidopsis* species is as an instance [19, 24, 31]. There is a dramatic variation in TE copy number and composition in different fish taxa, ranging from 55% in the zebrafish (*Danio rerio*) to only 6% in the green spotted pufferfish (Tetraodon nigroviridis). All major types of eukaryotic TEs and an overall higher TE diversity than other vertebrates are present in fishes wherein cause fishes comprising the five main lineages of vertebrates including; jawless, cartilaginous, ray-finned, and lobe-finned fishes [19, 34]. In species, TE contents and their density profiles are completely different and unique to each species [19]. Data display that the increase in the TE constituent associates with invasiveness and widespreadness of species and more adaptation to various circumstances [4, 6, 7]. Evidence exhibit that TE content of genomes associates with the evolution of organism complexity scales. For example, in bacteria and archaea as well as single-celled eukaryotes as the simplest life that appeared on the earth, their respective genomes consist of a high percentage of protein-coding (PC) transcripts than metazoans and multi-cellular eukaryotes (emerging from a

significant increase in non-PC regions of the genome and reduction in PC percentage of the genome).

Notable, the increase in non-PC regions of the genome associates with an expansion in the classes of non-coding RNAs (ncRNAs), including both small (for example, piRNA, miRNA) and long (for example, lncRNA) families.

The development of molecular regulatory systems and the rapid evolution of the primate brain with the acquisition of higher-order cognition were rendered via ncRNAs [28, 40, 41]. Evidence from the literature leads to the assumption that TEs were not originated from duplication or diverging events arbitrarily and spread through the genome evenly. But instead, they lead to assuming that TEs were arisen de novo ("originated from stem-loop hairpin RNAs and palindrome-forming sequences, emerged in the primordial pool of primitive genetic elements, the ancestors of both cellular and viral genes") and evolved a species-specific pattern in the genomes (Figure 1) [1, 2, 3, 9, 18, 19, 27, 41].

2.2.1. TEs as adaptive tools in genomic shocks of eukaryotes

Accordingly, a genome content of TEs is a kind of genetic adaptive tool that causing individual adaption (Figure 2A), e.g. to temperate climates and leading to population heterogeneity [21, 43]. TEs have contributed to genetic diversity and have had beneficial effects on evolutionary innovations [18, 44]. Studies report TE activation as a tool for genetic adaptation in a variety of domesticated plants and animals [15, 43]. Accordingly, TEs act as effectors in response to genomic shocks. For example in plants under some circumstances, TEs become active and increase in copy number, to countermeasure against stress [45]. Among the processes that may cause activation of TEs, domestication, polyploidy, inter-species and inter-generic hybridization can be mentioned [6]. An instance of somatic transposition that emerged in inter-species hybridization is the *Drosophila* hybrid dysgenesis phenomenon, in which crosses between specific lines of *D. melanogaster* led to various genetic changes, including sterility and increased-mutation and recombination rates. These effects associate with transposition and the mobilization of specific TEs: *P* elements (for the P/M system) and I elements (for the I/R system) [19].

DNA methylation and RNAi machinery are mechanisms essential for the epigenetic silencing of TEs, but environmental changes can lead to physiological, and therefore, epigenetic stress, which disrupts the tight control of TEs by these two mechanisms [21]. Following abiotic or biotic stress conditions, somatic TEs become active in plants such as; in temperature, nitrate starvation, wounding, etc. The transcription of specific transposons or retrotransposons may be induced by a temperature-dependent DNA methylation mechanism. For example, in the *Antirrhinum majus* "https://en.wikipedia.org/wiki/Species" species, the transcription of the *Tam3* transposon becomes active if exposed to a 10 °C temperature. The transcription of the *Tnt1* retrotransposon is induced by low temperatures in tobacco and tomatoes or by a fungal attack in tobacco. Also by several biotic and abiotic stresses including UV light, wounding, salicylic acid, and fungal attack, a *Ty-1 Copia LTR*-retrotransposon becomes active in oats [18].

Some TEs contain "Heat shock protein" heat-shock-like promoters and their rate of transcription increases if the cell is subjected to stress, besides increasing the mutation rate under these conditions, which might be beneficial to the cell [15]. In *Arabidopsis thaliana*, winter cold triggers epigenetic silencing of the floral repressor *FLOWERING LOCUS C (FLC)*. Cold weather causes a large increase in the antisense lncRNA *COOLAIR* which silencing sense FLC transcription and promotes Polycomb occupancy. Additionally, lncRNA *COLD AIR* (Cold Associated Intronic Noncoding RNA) helps to vernalization-mediated epigenetic repression of *FLC*. Both *COOLAIR* and *COLD AIR* act in sense/antisense manner and appear as epigenetic regulators to serve spatial and temporal specificity. The ncRNA- Polycomb repressive complex 2 (PRC2) relationship is an evolutionarily conserved mechanism in plants and mammals in gene repression [46].

Or in three sunflower species following their hybridization, *Ty3/gypsy-like LTR* retrotransposons became independently active. During rice domestication, the DNA transposon mPing increases its copy number by 40 per plant in the generation. Bursts of various TEs have been detected in several genotypes from a small marginal population of a wild relative of cultivated wheat [6].

In mosquitoes, organophosphorus insecticide resistance is primarily due to the overproduction of non-specific esterases, which sequester the insecticide before it reaches its target molecule acetylcholinesterase. Overproduction of carboxylesterases occurs in many resistant pest species where predominantly is caused by TE-derived gene amplification. A *long interspersed nuclear element* (*LINE*), downstream of the allele (Est-locus) looks responsible for the amplification process [18].

In insecticide-resistant *D. melanogaster*, the presence of an LTR-retrotransposon in the 5′ end of the *Cyp6g1* gene causing its overexpression and leading to the resistance to a variety of insecticide classes. The *Cyp6g1* gene is encoding the metabolic enzyme CYP6G1 that detoxifying insecticides [18]. In *Drosophila* populations, changes in the environmental temperature during development, disrupt the epigenetic silencing of TEs, and lead to their transposition [36].

TEs act as effectors in response to genomic shocks for instance, in *Drosophila,* telomere erosion could activate the mobilization of telomeric retroelements via a DNA-damage signaling pathway that will eventually restore telomere function by retroelements addition to the ends [47]. *Drosophila* telomeres were formed by repeats of two non-LTR retrotransposons, *HeT-A* and *TART*, which transpose only to chromosome ends in response to genomic shocks in telomere erosion [48, 49].

In human and large long-lived species, the structure of the telomere-specific elements indicates a TE ancestor recruited to perform the cellular function of telomeres with the help of telomerase. Also, the telomerase RT gene (*TERT*) is clustered with retroelements and all are located near the telomeres, a mechanism whereby the telomerase gene is regulated by retroelements and is located very close to the telomeres [30, 47, 50].

Notably, there may be a low level of somatic retrotransposition in humans, and primates wherein leading to short-lived animals [53, 54]. Literature account TE activation as the main mechanism involved in aging: a primary consequence of TE activation in the genome causes its disintegration. Herein, activation of TEs may occur in human somatic cells whereby causes genome disintegration which would be accompanied by aging [29, 36, 51, 55]. Somatic retrotransposition in other short-lived organisms such as *Drosophila* and maize (whose genome is doubled by retrotransposition), is conferring half of the observed mutations [15, 17]. Additionally, the immortality of certain immortal organisms including hydras and another Cnidarian, the jellyfish *T. nutricula*, is proposing to be due to the tight control of TE activity in their genome. There is a system in somatic cells (the Piwi-piRNA pathway) of these eukaryotes operating to maintain TEs inactive and prevent their mobilization, despite subjected to the same external DNA damaging factors as all other eukaryotic organisms, and that their DNA repair systems are not known to exhibit extraordinary effectiveness [29, 36, 55]. In germ cells from all organisms, TEs transposition is tightly prevented by mechanisms effectively silencing them, including the Piwi-piRNA strong system and RNAi [29].

2.2.1.1. TEs in advanced evolution of primates. Mostly, *LINEs* and *SINEs* (e.g. *L1* and *Alu* families) shape the primate genome landscape [40, 54, 56]. In the formation of primates, a mass of *SINEs* was inserted in their genomes (estimated ~74–93.5 * 10^3 new insertions). Many SINEs are expressed under conditions of stress [6, 15, 21]. Between families, *Alu* elements were opted as primate-specific, and the major cause for the length difference of certain genomic regions among primates and development of molecular pathways in the primate brain [12, 25, 41, 57]. In particular, different subfamilies of *Alu* elements with sequence variation accumulated in any given species with lineage-specific insertion and fixation, which cannot be removed by deletion processes. There

were species-specific numbers of Alu inserted and fixed in any given species [15, 58]. Evolution of the human genome is exceptional among all of the primates and millions of animals due to specific insertions of *Alu* elements [59], the most frequent HYPERLINK "https://en.wikipedia.org/wiki/Transposable_element" \o "Transposable element" TEs in the human genome, where particularly contributing to transcendency of human brain and disease more predispositions [15, 25, 28]. Estimating, *Alu* elements constitute ~42% of all detected TEs, and ~19% of the whole genome size, in human lineage [18]. Distribution of different *Alu* elements within one chromosome and between different chromosomes is uneven but not random. For example in human chromosomes 14, 16, and 21, *Alu* clusters concentrate in centromeric regions, whereas in chromosomes 4, 19, 20, X or Y clusters exist near the genes controlling metabolic, transport, and signaling processes [12, 41, 42]. *Alu* elements are relatively rich in CpG residues, which appear to be responsible for approximately 25% of all of the methylation in the human genome [15]. The *L1–Alu* pairs were opted dominant in the human genome [17, 58, 60]. *L1* family represent the only remaining mobile *LINE* family in human, constituting >17–20% of the genome, however, *Alu* yet are more abundant in copy number than *L1s* due to their 20-fold smaller element size [60]. *Alu* elements are extremely prevalent within ncRNAs, and the most frequent nuclear transcripts (hnRNAs) containing primarily *Alu* sequences [15, 28]. The *Alu* inserts in refmRNA collection show the great differences between humans and chimpanzees. For instance, genetic changes resulted from *Alu* insertions include; human *CMAH* loss of function and loss of Sia N-glycolylneuraminic acid (Neu5Gc) synthesis thereby accumulation of precursor N-acetylneuraminic acid (Neu5Ac); increased expression of alpha 2-6–linked Sias (likely because of changed expression of *ST6GALI*); and multiple changes in *SIGLEC* genes encoding Sia-recognizing Ig-like lectins (Siglecs). The hydroxylase gene is intact in all nonhuman primates, whereas the same region in the human genome is replaced by an *AluY* element that inactivates this gene in humans [59, 61]. Especially, *Alu* insertion contributed to the formation of human-specific new genes FLJ33706 and *microcephalin (MCPH1)* whose mRNA and protein expression specified in the human brain [59]. In modern human-model evolution, new *Alu* elements; *AluYa5, AluYb8, AluYc1*, etc. were inserted in the genome, where conferring the advanced evolution of modern-human brain size (being related to 4 million years ago) [59]. Brain evolution is a great process in the creation of primates and humans. From an evolutionary perspective, humans are unique in the advance of evolution and the number of Alu elements [28, 58], which addressed to a high speed of human-lineage evolution and a striking increase in the brain size (tripled in mass) whose function changed enormously [59]. The difference between the human and chimpanzee genomes is estimated ~4% [61], wherein compare to lower primates, there are ~5,530 (12 new *Alu*) and 1,642 (5 new *Alu*) new insertions in their genomes, respectively. Postulated, each insertion has great potential power on genome evolution due to a wide range of designed mechanisms [59]. The frequency of *Alu* elements in humans is three times more than in chimpanzees. This greater content seems to be a cause of differences between human and chimpanzee in phenotypic traits such as; relative brain size, age at first reproduction, longevity, sperm count, declarative memory, the theory of mind, HIV progression to AIDS (rare in apes), viral infections such hepatitis B/C, influenza and incidence of carcinomas (rare in apes) [61]. In the human genome, there is a strong correlation between the density of *Alu* elements and clustering of *miRNA* genes (e.g. chromosomes 19 and X which each encoding a high number of primate-specific miRNAs, and are accumulated unusually with a high number of *Alu* elements), however, chromosome Y was Kept poor from *Alu* elements [15].

3. All roads lead to rome: diversification and birth of species-specific RNAs

Approximately 1.5% of mouse and human genomes encode protein information, wherein ~60%–80% are transcribed into RNA, respectively [62]. Up to 98% of human RNA transcripts represent ncRNAs in cells [54]. Postulating, the degree of organism complexity among species correlated with the genome contents of TEs and proportion of each transcribed into ncRNAs than with the number of protein-coding genes, since protein diversification retained limited and protein machinery remained largely constant in the evolution. Herein, ncRNAs confer divergence and complexity needed in the evolution of life formed on the earth [28, 41, 63]. In evolution, ncRNAs derived from TEs were presented in all three forms of life; archaea, bacteria, and eukaryote genomes. Plants and animals possess all forms of ncRNAs (siRNAs, miRNAs, piRNAs, and lncRNAs) [28, 52, 64]. Genomic TE landscapes mirrored in the TE content of their lncRNA, miRNA, and RNAi repertoires [22]. So, TEs opted as genetic tools leading to lineage speciation by the birth of non-conserved and lineage-specific RNAs [25, 45]. In primates, there are many non-conserved and lineage-specific lncRNAs and miRNAs. *lncRNA-RoR* and *Xist* (*X-inactive specific transcript;* lncRNAs) involved in genome reprogramming of human pluripotent stem cells and X inactivation, respectively, derived from an assemblage of specific TEs [26, 65, 66]. In particular, ncRNAs (specifically miRNAs/lncRNAs) derived from *Alu* elements are non-conserved and primate-specific, wherein participating in brain development and higher cognitive abilities in primates [12, 41, 54, 56].

Together, data indicate that each TE-derived miRNA subfamily or lncRNA has a unique evolutionary history and its emergence, quite independent and different from other members in the same family or cluster [66].

3.1. RNA interference (RNAi) in fixation of TEs

In all forms of life, small ncRNAs (sncRNAs) through RNAi machinery function to maintain a steady-state of efficient repression of TEs, pre/post-transcriptionally. The RNAi machinery is an efficient and potent mechanism that evolved for fixation and transcriptional silencing of TEs [25, 67, 68]. RNAi includes miRNAs, "Piwi-interacting RNA" piRNAs, and "SiRNA" siRNAs, which "Gene silencing" silence genomic TEs through "Epigenetics" epigenetic mechanisms; pre-transcriptionally by "DNA methylation" DNA methylation or post-transcriptionally through TEs transcripts after they have been transcribed, such that no transposition of TEs have been defined to be occurred in human and wild-type plants through generations [15, 40].

In humans, for example, TEs-derived transcripts with cis-encoded sense/antisense patterns are formed to coordinate the formation of RNAi machinery [54, 70]. TE cis-regulatory regions are found ~~in the~~ upstream of the LTR of retrotransposons, whereby represent motifs that facilitate host recognition by RNAi and its defense action against transposition with subsequent TE silencing [40]. In somatic PGCs, TEs-containing clusters (e.g., X-linked flamenco locus) are transcribed to produce piRNAs that are almost exclusively antisense to TEs [72]. In human sperm, sncRNAs have unique compositions and functions that effectively masking repetitive TEs and protect the genome integrity from invasive elements [71]. If a TE escapes suppression and becomes activated host defense through the RNAi machinery re-suppresses it in a self-reinforcing loop for the re-establishment of silencing.

In plants, as another example, TEs dual-code siRNAs and miRNAs (long double-stranded and short hairpin (stem-loop) RNA structures, characteristic of siRNAs and miRNAs, respectively). Plants and animals possess these sncRNAs, miRNAs, and siRNAs, mostly in defense against transgenes and viruses [28, 40, 69]. Moreover, there is a family of sncRNAs, the piwi-interacting RNAs (piRNAs), defined functioning particularly in simple eukaryotes with immortality, germ-lines, and the brain of animals as a strong defense against transposon activity [28, 29, 52, 64].

In *C. elegans,* RNAi is derived from dsRNAs from the terminal inverted repeat (*TIR*) sequences found at the ends of *Tc1* elements. The *TIR-derived* single-stranded siRNAs are essential to silence TEs in the germline as well as, in defense against viral infection and somatic TE mobilization [69].

In mammalian germ cells, RNAi pathways are participating to produce phase-specific gene expression patterns [73, 74]. Genome-wide methylation dynamic changes occur in germ cells of mammals during gametogenesis in a sex- and the sequence-specific manner by miRNA-, siRNA- and lncRNA-depended mechanism [75].

In mammals, all sncRNAs are abundantly present in both female and male germ cells but transiently replaced by miRNAs and piRNAs in spermatozoa and endo-siRNAs in oocytes and zygotes [76, 77]. Expression profiles of miRNA clusters show similar patterns between PGC and spermatogonia cells and between oocytes and zygotes [73, 76]. The oocyte has enormous reproductive potential, but limited time for lifespan since depleted from piRNAs. However, the only cell having the potential to rescue the oocyte is fertilizing spermatozoon with its sncRNAs.

The embryo developmental potential and early zygotic gene activation are each dependent on paternally derived sncRNA cargo. The male gamete carries several sncRNAs, to the oocyte during fertilization that may all participate in successful embryogenesis. Herein, it is expected that sncRNAs produced through the oocyte lncRNA interactions with sperm sncRNAs, governing oocyte-to-zygote programs, participating as key developmental processes encompassing pronuclear formation, DNA repair, orchestrating oocyte activation, the transition from maternal to embryonic gene control, and the establishment of imprints in early embryos [77, 78, 79, 80]. Genome-wide methylation dynamic changes occur in germ cells of mammals during gametogenesis in a sex- and the sequence-specific manner by miRNA-, siRNA- and lncRNA-depended mechanism. Mammalian germ cells reestablish large-scale de novo DNA methylation and genomic imprints specific for each oocyte and sperm. These steps are required for normal germline differentiation and embryonic development (Figure 3) [75,77]. There are 2 types of methylation patterns in the oocytes: (i) methylation across the transcribed regions, which might be required for the establishment of maternal methylation imprints and normal embryogenesis, and (ii) retroviral methylation, which might be essential for silencing retrotransposons and normal oogenesis [22, 75]. Oocytes use two types of methylation mechanisms: Dnmt3L-dependent methylation, which is required for maternal methylation imprinting (sex-specific), and Dnmt3L-independent methylation, which is siRNA- and lncRNA-depended and essential for endogenous retroviral DNA silencing [75, 81]. After zygote formation, maternal lncRNAs containing antisense TEs become downregulated by paternal sncRNAs (miRNAs and piRNAs) that explaining the piRNA-mediated silencing in the zygote and the specificity of *Alu-lncRNAs* [27, 82].

3.2. TEs in the origin of microRNAs

Until now, researchers have identified a large number of miRNAs genes in humans, rhesus, and mice that are derived from TEs (which mostly overlap with repeats *LINEs, SINEs* and *LTRs*) [12, 52]. Both miRNAs and lncRNAs show great diversity in both copy numbers and sequence variations that are addressed to their specific-TEs constitutions [26, 66]. Conserved miRNAs/lncRNAs have functional orthologs in multiple species, while non-conserved one is species-specific and a key regulator in lineage speciation [52]. Human miRNAs derived from *MITEs*, *Alu*, and *L1s* are non-conserved miRNAs and human-specific. *Miniature inverted-repeat transposable* elements (*MITEs*) are DNA-type elements contributing to miRNA genes (e.g. *hsa-mir-548*) [52,70,83]. Primate-specific miRNAs were formed generally in clusters and enriched with *Alu* elements [41, 81]. They co-evolved in specific chromosomal locations, whereby presented more advanced regulatory networks in primates in comparison to other animals. Clustered miRNAs are from the same families [84]. Two primate-specific miRNA clusters enriched in *Alu* elements have been identified until now. The largest one discovered so far is the *Alu*-enriched chromosome 19 miRNA cluster (C19MC) where *Alu* elements were distributed over the whole clusters [41, 66, 85]. Human C19MC consists of 46 genes encoding 59 different mature miRNAs. Intriguingly, C19MC expression exclusively occurs during early embryogenesis from the paternal allele and become silent later on during development [41]. The C19MC preferentially contribute to developmental novelties in human, chimpanzee, orangutan, rhesus monkey, and marmoset [65]. C19MC expression mainly occurs in the reproductive system, in embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), but not in adult tissues. Sixteen miRNAs of C19MC share the same “AAGUGC” seed sequence with ESC *miR-302/-372* family [84, 85]. There is a dual relationship between *Alu* elements and miRNAs in C19MC whereby creating anti-sense miRNAs to easily target *Alu* elements. Evidence shows co-evolution between human-specific miRNAs and *Alu* repeats in the genome [81]. Another *Alu*-enriched cluster on Chr 19 is *miR-371/372/373**, and *miR-373*, homologs to murine *miR-290/291a/291b/292/293/295* on Chr 7. To shift from self-renewal to differentiation, this set of miRNAs derived from *Alu* must be switched off. In the case of humans, these miRNAs are epigenetically silenced by TEs-derived siRNAs [64, 85]. The second identified one is then mapped to chromosome X and known as cluster *miR-506-514*, preferentially expressed in testis, and is essential for male sex maturation in primates [66]. The cluster encodes miRNAs with seven different seeds; *miR-513/506/507/508/509/510/514*, each containing multiple copies in primates but not in rodents or dogs. These miRNAs are preferentially expressed in male germ cells at the time of spermatogenesis. There is a strong correlation between this miRNA expression and male sexual maturation in primates [66, 76, 86]. The *miR-506-514* cluster has the greatest lineage-specific diversity in terms of both copy numbers and sequence variations. For example, each primate species possesses a particular variety of multiple copies of the evolutionary-novel *miR-513* subfamily [66]. While the *miR-513* different-copy numbers occurred in all primates, *miR-514* amplifications occurred only in human and chimpanzee genomes. As for TEs, evidence refuses the divergence of new miRNAs through tandem duplication of an existing gene followed by substitution on nucleotides. TEs opted as functional elements in lineage divergence of X-liked miRNAs that occurred independently in species evolution [41, 66].

Another novel miRNA cluster discovered in primates is located in human Xq27.3 and consists of six distinct miRNAs including *miR-890/888/892a/892b*. All six miRNAs were well conserved among primate species but unidentifiable in other mammalian species (including mouse, rat, cat, dog, horse, cow, opossum, and platypus). Like as two previous clusters, data reveal a strong correlation between changes in the expression of these miRNAs and male sexual maturation, that suggesting regulatory roles of this cluster in testis development and spermatogenesis [86, 87]. For example, *miR-890/888/892a/892b* is controlling sperm maturity and male fertility [41]. The X-linked miRNA clusters were detected to be expressed specifically in testis during spermatogenesis and meiotic sex chromosome inactivation (MSCI) [73, 88]. While sex chromosomal genes undergo epigenetic silencing in spermatocytes during meiotic prophase I by MSCI, X-linked miRNA clusters escape silencing to participate in MSCI itself. Hypothesizing, there was an election for clustering of miRNAs on the X chromosome, whereby participating in male maturation and spermatogenesis. MSCI is not confined to mammals, metazoans as diverse as the fruit fly, grasshopper, the nematode worm, and chickens also demonstrate MSCI as rule in the evolution which is driven by TEs-derived sncRNAs [73, 89].

Post-fertilization, high abundant oocyte lncRNAs scavenge sperm miRNAs and lead to the demethylation of the highly methylated paternal genome [71]. Post-fertilization, sperm miRNAs tethered by maternal lncRNAs cause targeted degradation of stored maternal RNA transcripts with established function in chromatin remodeling [76, 79]. Some of the paternal miRNAs that may affect the early development are including *hsa-mir-34c/375/252/25* and *hsa-miR-148a.* The last one, *hsa-miR-148a* particularly down-regulates *DNA methyltransferase 3b* (*DNMT3b*) by recognizing an evolutionarily conserved coding sequence [71, 90]. The most abundant miRNAs detected in human spermatozoa are epi-miRNAs that repressing the epigenetic machinery: *hsa-miR-140/21/152/148a*. For example, *miR-152* together with *hsa-miR-148a*, directly targets

DNMT1. In oocyte-to-embryo transition and genomic reprogramming, there is a temporary shift in DNA methylation, as well as, in RNAi pathway, from miRNAs to siRNAs through TEs-related lnc-RNAs. Degradation of oocyte lncRNAs by sperm sncRNAs seems to give rise to trans-acting short interfering RNAs [79, 80].

3.3. TE and evolution of lncRNAs

Long non-coding RNAs (lncRNAs) are longer than 200 nucleotides and often polyadenylated, but they are devoid of evident open reading frames (ORFs) [63]. In evolution, they added further complexity to animals. They are present only in invertebrates, vertebrates, plants, and about one-third of them are primate-specific. Herein, a functional theme in lncRNAs formation is in the evolution of primate brains [28, 41]. In particular, *Alu* family was opted for primate-specific lncRNAs mainly expressed in the brain, testis, and ovary. Likewise, *human endogenous retrovirus subfamily H* (*HERVH*) was opted as functional elements of lncRNAs expressed in human embryonic stem cells (hESCs), as required to maintain hESC identity [27, 54, 82]. TE families inserted in the lncRNA transcripts associates with tissue specificity function of lncRNAs, as functional domains or regulatory sequences of lncRNAs. There is substantial inter-species and intra-lineage variation in the lncRNA landscapes wherein reflecting differences in the coverage and types of TEs embedded in the genomes [27, 82].

Generally, lncRNAs are non-conserved through the evolution as exemplified by mouse and human *Xist/XIST* transcripts that exhibiting only 49 % sequence identity [54]. Most of the human lncRNAs that are functionally characterized do not have identifiable orthologs in non-primate species, except for *Xist* and *Cyrano* [27]. Data display that particular families of TEs were opted and inserted non-randomly in lncRNA sequences [41, 54] and lncRNAs were arising de novo from TEs insertions. For instance zebrafish and humans lncRNAs show extensive sequence similarity, where related exclusively to the shared repetitive elements. It displays that lncRNA genes, instead of originating from duplication or mutation events, arose de novo from TEs insertions [41]. Once lncRNAs formed, TEs fixed and substantially contributed to the functional diversification in species evolution [16, 26, 41]. A high level of TE insertions in lncRNA sequences was the rule rather than the exception compatible with lncRNA activities [26]. Perhaps the most compelling example comes again from *XIST* (human *Xist*) enriched in repeats, whose TE content has increased in the human lineage [26, 41, 63]. Indeed, TEs became tools for diversification and birth of new lineage-specific lncRNAs for instance, *HERVH* in the birth of new human-specific lncRNAs [26, 27]. Or, *human antisense lncRNA 1* (*Uchl1-as1*) is essentially needed for the stress induction of *ubiquitin carboxy-terminal hydrolase L1* (*UCH-L1*) that is exclusively expressed in neurons and cells of the diffuse neuroendocrine system. *UCH-1* expression increases the available pool of ubiquitin to be tagged onto proteins destined to be degraded by the proteasome, and required for normal "Synapse" synaptic and "Cognitive" cognitive function in humans. Neurodegenerative disorders in humans "Parkinson's disease" Parkinson's disease (PD) and "Alzheimer's disease" Alzheimer's disease (AD) are closely relevant to *UCH-L1* and *Uchl1-as1* expression [12]. There is a primate-specific lncRNA *(lncND)* gene, which controls the neuronal signaling pathway and stands for brain development in primates. In the new world primates, *lncND* has 16 new *Alu* insertions (specific *miR-143-3p-recognition elements* (*MREs*)) at the 5′-region which tethering *miR-143* and indirectly up-regulate the expression of miR-target genes (e.g. miR-143 target genes *Notch-1/2*) involved in the advanced development of the brain in the new world primates. However, *lncND* in genomes of the old world monkeys and apes lack conserved *Alu* insertions (16 MERs) inserted in the new world monkeys [41].

3.3.1. LncRNAs & coordination in gene silencing or imprinting

X chromosome inactivation (XCI) and genomic imprinting in mammals are examples of the coordinate activation of lncRNAs. Both processes are controlled by cis-acting master control regions evolutionary derived by TEs, X inactivation center (XIC), and imprinting control region (ICR), respectively (Figure 3) [22, 91]. In mammals, XIC makes a region ~500-kb encoding a cluster of lncRNAs including *Xist* and *Tsix* (Figure 3A) [24, 91]. In mammals, imprinting of the gene clusters is controlled by ICRs, CpG-rich cis-regulatory elements (CpG islands) marked by DNA methylation most often on the maternal alleles. Most ICRs are methylated in the female germline during oocyte growth [22, 23]. Thereby, both XIC and ICR are composed of TEs-derived cis-acting elements methylated in the oocyte vs the sperm. Selective expression of *Xist* in female individuals, but not in males, relates to un-methylated XIC received from the male gamete [22, 91]. The maternally imprinted gene clusters (e.g. *Igf2r/Airn*, *Kcnq1*, *Snprn/Ube3A*) are harboring lncRNAs whose promoters originating within or near ICRs. The most common mechanism used for imprinting these clusters relies on the expression of a lncRNA in cis and exploits much of what has been identified for *Xist/Tsix* in XCI (Figure 3A) [23].

There are at least several hundred other sense–antisense pairs detected within mammalian genomes, which acting similar to *Xist–Tsix* pair, as well as, *Igf2r/Airn*, *Kcnq1/Kcnq1ot1*, *Snprn/Ube3A*, *Igf2/H19*, *Dlk1/Gtl2* cluster genes (Figure 3B) [23, 91]. Maternally methylated ICRs harbor the promoter for lncRNAs such as *Kcnq1ot1*, *Snprn*, and *Airn* [23]. For the *Igf2r* domain, transcription of the *Airn* lncRNA is governed by a promoter within the ICR and is expressed from the unmethylated paternal allele (resembling *Xist*). In somatic cells, transcription of *Airn* over the *Igf2r* promoter prevents *Igf2r* expression by recruiting enzymes that confer repressive modifications to silence genes in cis. Similarly in the *Kcnq1* cluster, the ICR contains the promoter of the *Kcnq1ot1* lncRNA. On the paternal allele, the ICR is unmethylated, where allowing *Kcnq1ot1* expression and leads to silencing the paternal allele of the linked genes in cis (resembling *Xist*). In the *Snrpn* locus, *Ube3a* is expressed from the maternal allele (resembling *Tsix*) [22, 23]. All of these lncRNAs arose in clusters containing repetitive elements and similarly contained a poly-A tail and function in sense/antisense manner. As an instance, maternally imprinted *anti-sense lncRNA Airn* leads to maternal expression of the *IGf2r/Sic22a2/Sic22a3* gene cluster, in sense (Figure 3B) [22, 23].

Cis-acting regions in both XCI and genomic imprinting are mostly methylated by the female germline during oocyte growth, but a few of them, including the ICRs for the *H19/Igf2* and *Gtl2/Dlk1* clusters, are methylated on the paternal allele before birth in the male germline [23, 91].

Accordingly, high *L1s*-density in the human X chromosome positively correlates with the ability of *XIST* to spread across the Xi. Evidence show coordinated long-range control in cis by lncRNAs in genomic imprinting and imprinted X inactivation [22, 23, 62].

XIST (the human variant) is a 17-kb transcript, expressed by XIC solely from Xi, contains several repeats derived from TEs, and making functional elements in X silencing. For instance, repeat A and F derived from ERVB5 and DNA transposon, respectively. Their mutations result in the abrogation of *XIST*-mediated silencing [26, 62]. Repeat A and F is necessary for the recruitment of PCR2 repressor complexes and the transcription factor YY1, respectively [26, 54]. YY1 is a “bivalent” protein capable of binding both *Xist* RNA and DNA only the nucleation site on inactive X (Xi) (not on active X (Xa)) and tethering Xist to its site of synthesis on the Xi. *Xist* RNA coats only the Xi, where is expressed. On the other side, *Tsix* a 40-kb antisense transcript acts in cis at the *Xist* promoter, to repress *Xist* expression by silencing the *Xist* promoter. *Tsix* serves as a potent antagonist of *Xist* expression [23, 91].

Herein, *Xist* is produced exclusively from the Xi since the maternal allele imprints prevent XIC to express *Xist*, while both parental alleles in XIC can express *Tsix* (Figure 3A) [22, 23]. *Xist* covers the Xi and initiates chromosome-wide silencing as it accumulates and blankets the X in cis only when Tsix is present. The cross-talk of *Tsix-Xist* is essential for the precise choice of Xi. Once chosen, the Xi-elect becomes distinct from Xa by *Xist* marking it in a strictly cis-limited fashion [62]. In coordination with *Tsix* and with each other as well, the *Xist*-PRC2 complexes load onto

the Xi "nucleation center'' within *Xist*'s exon 1 [22]. Hear, the counterparts *Xist/Tsix* are paradoxically essential in the mutually exclusive choice of Xa and Xi. In female ES cells, in the pre-XCI state, *Tsix* RNA is expressed from both Xs and establishes the 40-kb region overlapping *Tsix* and *Xist* loci which may preserve it in an open chromatin state. At the onset of cell differentiation, the Xi-elect loses *Tsix* expression as a result of repressive marks by *Xist* and its negative effect on the 40-kb domain to losing euchromatic marks. The repressive marks on Xi prevent *Tsix* expression [62].

Accordingly, the RNAi system is involved in the initiation of XCI and in recruiting *Xist* to Xi in female cells. Reports show that *Xist/Tsix* form long double-strand RNAs (dsRNAs) then processed into small RNAs, called xiRNAs in a Dicer-dependent manner [62, 92]. This is a parental imprinting pattern and xiRNAs that select Xa and Xi in an exclusive manner whereby ncRNAs of the XIC act in cis; i.e., on the chromosome that synthesizes them (Figure 2A) [23]. In Xa-elect, *Tsix* RNA expression maintains the 40-kb *Tsix/Xist* locus in cis in a euchromatic configuration, while recruits xiRNAs and Dnmt3a to the *Xist* promoter to methylate and lock in the silent state of *Xist*. Dnmt3a methylation is mediated by small RNAs created from long *Xist/Tsix* dsRNA via the RNAi pathway like RNA-directed DNA methylation (RDDM) and transcriptional gene silencing (TGS) in plants and yeast, respectively [23, 62, 92]. Studies support the role of sncRNAs in the mechanism(s) responsible for initiating and/or maintaining MSCI in the male germline where about 80% of X-linked miRNA genes have been shown to escape MSCI process. MSCI-escaping miRNAs play a similar role in the inactivation of XCI in male germ cells and the inactivation of paternal X [92]. Sperm miRNAs participate to methylate paternal genome and actively to mark paternal genes, specifically have a close relation to phenotypic and genetic properties in species [71, 74]. Similarly, end-siRNAs originating from overlapping long duplexes exist in oocytes where recruiting DNA methylation at the promoter sites and, at the same time, maintaining the steady-state repression of the locus [23, 62, 92].

4. Lifespan and mechanism by TEs in aging

The lifespan of species in nature has amazing diversities, differences ~150,000-fold of magnitude between the shortest- and longest-lived species. The lifespan spectrum contains exceptionally long-lived species, such as the bowhead whale (lives >200 y, the longest-lived mammal) and at the other end, the African turquoise killifish (*Nothobranchius furzeri*; *N. furzeri*), as a short-lived animal [93]. T. killifish (turquoise killifish) strains are highly inbred, thereby all autosomes and X chromosomes are nearly homozygous. Each strain in its separate captivity has its distinct lifespan (expanding 4–6 months). This lifespan variability is addressed to intra-species polymorphisms of Y chromosomes. Each strain has its distinct Y chromosome different from other strains. The genomic region associated with lifespan is supposed to be on the sex chromosome, close to the SDR. T. killifish genomes contain a repeat content of 64.6%, comprising 42.1% dispersed retrotransposons and 22.5% tandem repeats [94, 95]. The long lifespan of some animals such as the elephant lineage and bowhead whales relates to tightly preserving their genome stability, powerful DNA repair, and very low incidence of diseases. In the bowhead whale, there are duplications of genes associated with proteostasis, DNA damage repair, proteasome regulation, and ubiquitination associating with reduced cancer risk and aging [96, 97, 98]. These gene duplications were derived by TEs conferring phenotypic innovations in lineage evolution. Like humans, repeat

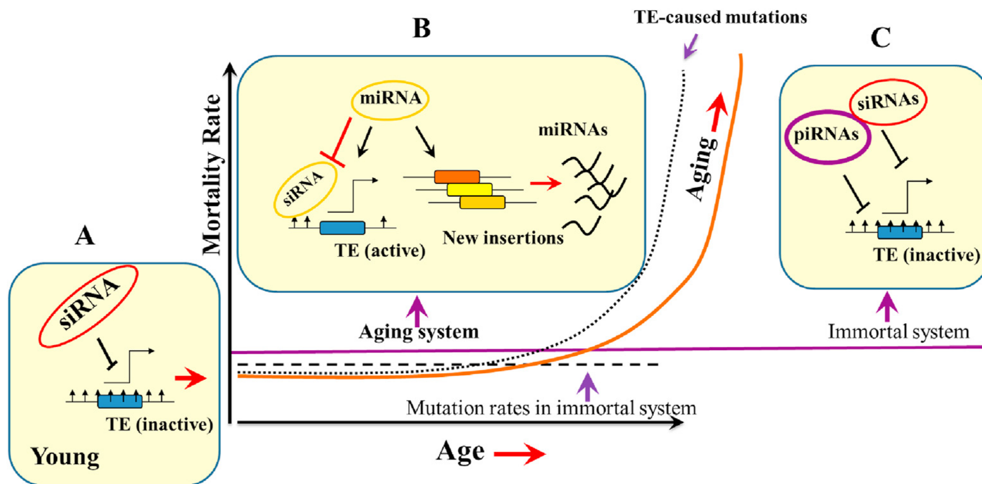


Figure 4. Repression of transposable-elements (TEs), as anti-aging and anti-cancer mechanism. A) (Left) In young animals, TEs are repressed both transcriptionally and post-transcriptionally by RNAi system (preferentially by siRNAs). B) As animals age, miRNA system replace siRNAs, whose biogenesis competing with siRNA biogenesis wherein the genetic surveillance mechanisms break down and become less effective and leading to loss of heterochromatin structure, and reteroelements become progressively active during the lifespan. Thus, during adulthood, the gradual release of TEs can generate considerable levels of molecular damage that overwhelm the capacity of the cellular maintenance and DNA repair systems. C) In immortal system like as germ line, somatic activity of piRNA pathway is an evolutionary tool conferring longevity and regeneration capacity in the immortal system. The Piwi-piRNA system effectively silences TEs and due to the lack of TEs mutagenesis, the germ line genome remains constantly stable (the horizontal violet line). In aging systems, in addition to constant mutations, TEs also generate damages at an increasing rate in somatic cells throughout the lifespan (dotted gray curve). In these cells, the Piwi-piRNA pathway is largely inactive and a competition between siRNA and miRNA pathway allows self-replicating TEs to accumulate exponentially. When the level of damages passes a critical threshold, the saturated repair machinery cannot eliminate all of them, leading to a significant amount of cell death and aging (orange curve) [29, 35, 53, 55, 68, 101].

sequences make up 41% of the bowhead genome, most of which belong to *LINEs*, such as *L1*, however, the bowhead genome is virtually devoid of *SINEs*. Herein, protective molecular adaptations by TEs were linked to genes associated with DNA repair, cell-cycle regulation, cancer, and aging [38, 96, 97]. Also, the genome of elephant lineage includes at least 20 copy numbers of *the p53* gene evolved by TEs. These duplications encode at least 20 isoforms of p53 and correlate with the evolution of increased body size and an enhanced DNA damage response in long-lived elephants. Interestingly, the p53 copies were flanked by *RTE-type* non-LTR retrotransposons (*RTE-LA*), whereby p53 gene duplications evolved, potently fixed, transcribed, and translated in elephant tissues. Among the mechanisms input in large body-sized animals to resolve for long life, there is a decrease in the copy number of oncogenes, an increase in the copy number of tumor suppressor genes, reduced metabolic rates whereby leading to decreased free radical production, reduced retroviral activity and load, increased immune surveillance [77, 97, 98]. Here, there are several examples of TE-derived gene duplications in primates such as *RNASE1*, a pancreatic ribonuclease gene, in leaf-eating monkeys that contributed to adaptative changes in diet and digestive physiology, a duplication of *GLUD1* in hominoids that subsequently acquired brain-specific functions [38, 97].

At the genomic level, aging associates with increased activity of TEs and related transposition which generally conferring genome instability in aged animals (Figure 4A and B) [29,55,98]. TEs activity has an important role in the lifespan of cells. In short-lived animals such as mouse strains, TEs do not tightly fix and possess high activity whereby leading to genome instability. Herein, *C. elegans* old worms show markedly increased expression of transposons and related-transcript levels in their cells. TEs play a dual role in cells through inducing RNAi machinery (piRNA/miRNAs/siRNAs pathway) whereby on one side induces anti-aging function and on the other side is giving rise cells to aging/-senescence by translocation and genome disintegration (Figure 4A and B) [29, 36, 77]. Somatic activation of TEs during aging leads to increased expression of miRNAs and aged-related disorders such as a neuronal decline in the brain and cardiac disorders in the cardiovascular system (Figure 4B) [99, 100]. In humans, somatic transposition of LINEs (e.g. *L1s*) increases with aging, in particular in the adult human brain. Expression of *L1* elements has been detected in senescence where SIRT6 and DNMTs failed to repress *L1* transcription into inactive heterochromatin [29, 64, 99]. Moreover, caloric restriction causes an increase in Sir2 activity and repression of chromatin regions related to TEs [35, 98]. Suppression of *Alu* elements by RNAi system in aged adult stem cells can reverse the senescent phenotype and reinstall the cells' self-renewing properties. Even more, factors causing molecular damage do not influence the rate of aging to great extent, as either the damage is repaired, or the damaged cell is eliminated from the tissue through cell loss [29, 99]. It is noteworthy that *L1s* activation in pathological conditions leads to progression to cancer; which raises the activation of piRNA pathway and the ectopic expression of PIWI proteins is occurred in several types of cancer**.** As a new concept, somatic TEs activity in ESCs was a mechanism conferring phenotypic heterogeneity and new combinations in the population of individuals. TEs shape the eukaryotic proteome landscape via the formation of cis-acting elements upstream of open reading frames, as major regulators of gene expression and protein translation [22, 54, 56].

4.1. TEs silencing by piRNA pathway: the road to immortality

The earliest animals such as single-celled, eukaryotic protozoa (*Paramecium* and *Tetrahymena*) show great longevity and potential immortality since they possess somatic piRNA pathway (Figure 4C) with several characteristics relevant to aging, including asexual clone immortality, regeneration, and the ability to cycle between dedifferentiation and differentiation, all of which contribute to the longevity [36, 101]. Reports represent the Piwi-piRNA pathway as a feature shared by non-aging (potentially immortal) biological systems such as the germline, certain organisms from 'lower' eukaryotic taxa (e.g. *Planaria* and *Cnidaria),* and immortal stem cells [45, 55]. These immortal systems can reproduce clonally that is the progeny can actually 'regenerate' from somatic cells of the parental body [55, 96]. The *Cnidarians* (e,g, *hydras,* the jellyfish *Turritopsis nutricula)* and *planarians,* for example as the simplest multi-cellular animals, have abilities to reorganize, rejuvenate and regenerate their lost body parts whereby can produce great longevity and potential immortality [29, 55]. The immortality and regeneration capacity in these immortal systems are referred to Piwi-piRNA pathway and the tight control of TE activity in their genome (Figure 4). Similarly, germ cells do not age since effectively silence TEs and prevent their activation by using the Piwi-piRNA system [6, 36]. These eukaryotic systems are immortal since operating to maintain their genome stability, despite subjected to the same external DNA damaging factors as all other eukaryotic organisms, and that their DNA repair systems are not known to exhibit extraordinary effectiveness [9, 55]. They are subjected to ionizing radiation, harmful factors generated by their metabolism, or high temperature and oxidative stress (damaging DNA)- but yet these immortal systems do not show any signs of aging [29, 36, 99]. Evidence briefly brings into account the tight connection between TEs mobilization and aging and anti-aging function of RNAi pathway [35, 55, 90]. The somatic activity of piRNA pathway is an evolutionary tool conferring longevity and regeneration capacity in immortal systems [29, 35, 99]. Accordingly, in soma, retroelements become progressively active during the lifespan [35, 55, 100], wherein enforcing endo-siRNAs or miRNAs machinery to come to function and contribute to TE repression. However, they are not as efficient as the piRNA pathway in silencing TEs due to two factors: they repress TE transcripts only when TE is transcribed and processed, and has a reduced capacity to pack silenced TEs into heterochromatin. Additionally, siRNA/miRNA RNAi machinery competes in sharing proteins [44, 67, 68]. Strongly, piRNAs are single-stranded sense and antisense TEs-derived transcripts generated by Piwi-mediated cleavage, a process that is essential and initiated in the zygote by paternally derived piRNAs [44]. In animals, piRNAs are self-amplifying by a designated-loop mechanism named the ping-pong cycle. The Piwi/piRNA system is required in the establishment rather than in the maintenance of DNA methylation patterns in primordial germ cells (PGCs) [44, 99]. Thereby, aging somatic cells are less capable to preserve their heterochromatin structure and progressively lose it, which gradually leads to transcriptional activation of repressed TEs. Thus, during adulthood, the gradual release of TEs can generate considerable levels of molecular damage that overwhelm the capacity of the cellular maintenance and DNA repair systems, including autophagy, the ubiquitin-proteasome system, molecular chaperones, and the distinct DNA repair pathways (Figure 4). The repair and maintenance (cell cleaning) mechanisms are likely to be equally effective in the soma and germline in eliminating damages produced by metabolic and environmental factors**.** The genome of somatic cells progressively accumulates mutations, mostly genomic rearrangements, as the individual ages; whereas the integrity of genetic material in germline cells remains stable, from generation to generation. Mutations that disrupt the piRNA biogenesis pathway in mouse and fish cause germline-specific cell death and sterility. The majority of piRNAs is clustered in discrete genomic loci that are active specifically in germ cells [29, 36, 55].

Mammalian germ cells reestablish large-scale de novo DNA methylation and genomic imprints specific for each oocyte and sperm. These steps are required for normal germline differentiation and embryonic development [75]. In zygotes, mainly at the 2-cell stage, genome reconstruction and zygotic gene activation (ZGA) occur by numerous bidirectional maternally inherited lncRNAs associating with the activation of their cognate genes in zygotes [102]. In animal evolution, short lifespan co-evolved with sex evolution [94], since single-celled organisms or multi-cellular organisms that able to arise from a single cell (asexually) can reproduce potentially immortal clones as large populations of totipotent/pluripotent stem cells; whereby giving rise to animals with both great regenerative powers and potentially very long lives [36]. In the T. killifish, lifespan seems to be co-evolved with the SD

evolution. The genomic region associated with a different lifespan between strains is on the sex chromosome, close to SDR [94, 95].

5. Evolution of sex chromosomes and TEs

The sex chromosomes are among the most diverse genetic systems in all of biology. Among metazoans, there are two SD systems XY and ZW. For example, in mammalians, inheritance of XX determines the female sex and inheritance of XY specifies the male sex, whereas in birds, females are ZW and males are ZZ. Mammalian X carries over three times more genes than the Y does while the chicken Z carries over ten times more than the W. Sex chromosomes display enormous diversity in morphology, gene content, and specific molecular mechanisms involved in SD, all of which used to deduce multiple and independent origins of sex chromosomes. There is a lack of correlation between the evolutionary age and the stages of differentiation and the degree of TEs accumulation in sex chromosomes [18, 19, 103]. Despite the independent origins of sex chromosomes, they all share unique features shaped by a common set of TEs. All of the unique features of sex chromosomes are addressed to TEs; the evolution of SD locus, heterochromatinization, recombination suppression, morphological differentiation of sex chromosomes, and dosage compensation [25, 103].

TEs are accounted as a source of raw materials in the early evolution of sex chromosomes and as active drivers of SD [19]. The simplest example is in two plant species *Arabidopsis thaliana (A. thaliana)* and its closely related congener *Arabidopsis lyrata (A. lyrata)*. The genome of *A. lyrata* has a two-to threefold higher TE copy number than *A. thaliana*. Importantly, two species differ in the mating system. *A. lyrata* is an obligate outcrosser, whereas *A. thaliana* is predominantly a selfer. The difference in the mating system shows the potential of TEs in the evolution of sex and the efficacy of TEs to be selfers or outcrosses [31]. The sex chromosomes (XY/ZW) have substantially massive contents of specific families of TEs, much higher than autosomes [19, 25, 103].

The enrichment of TEs in Y (W) and X (Z) chromosomes conferred the emergence of SD loci [103, 104]. These loci were concentrated with highly sex-specific genes (e.g. SRX and SRY), and formed from massive content of TEs that became silenced and heterochromatic [19, 103].

The repeat density of the sex-determining regions on the Y/W (SRY) and X/Z (SRX) chromosomes that are identifying as sex-linked genes is higher than in the rest of the chromosome [19, 103]. TEs vary extensively in their effects on sex chromosomes; some TEs are almost completely absent on the Y and strongly accumulate on X chromosomes [25, 88]. There is a massive accumulation of *L1-Alu* pairs on the human X where functioning in dosage compensation. In the inactivation process, *L1s* are serving as "way stations'' for the spread of the inactivation signal [24, 105] and as Repeat Insertion Domains of LncRNA (RIDL) in Xist which promoting the spread of Xist along X, the agent of X-chromosome inactivation [17, 22, 24].

TEs formed unique patterns on X- and Y-linked zinc finger genes (ZFX & ZFY), the SD sequences evolved in mammalian sex chromosomes. Humans contain a massive accumulation of *Alu* elements in ZFX & ZFY. Recombination became suppressed in ZFX & ZFY due to determine the sex type in mammals. In humans, both genes are ubiquitously expressed in adult tissues, and ZFX is not subject to X inactivation [19, 104].

The fish species have a variety of SD systems; sex chromosomes (XY and ZW), with different SD mechanisms and temperature-dependent SD [19].

In mammals, the sex type is determined by the presence or absence of the Y chromosome, which encodes the SRY gene necessary for testis development. In contrast to a previous theory that the genes situated on the Y chromosome are disposable and in various stages of decay, as well as, lack of functional constraint, the Y chromosome was recently found to contain several housekeeping genes and genes that are expressed in the testes. Thus, the Y chromosome is not devoid of functional constraint as previously supposed and TEs show a more unique abundance in Y regions [19, 104].

In contrast to the previous theory of gradual accumulation of TEs on the sex chromosomes and their gradual evolution to miRNAs and divergence, no miRNAs exist in Y-chromosomes. The gradual accumulation of TEs overtime would lead to deriving miRNAs and lncRNAs, while we observed the opposite on the Y chromosome and on the autosomal chromosomes of species beyond mammals (having higher densities of miRNAs on the autosomes) [25, 41, 88].

Mammalians have a higher density of miRNAs on X chromosomes, while in lower species than mammalians such as mosquito and fruitfly species densities of miRNAs across the X chromosome are lower than autosomes [41, 88].

In contrast, the mammalian Y chromosome carries no miRNAs and a low density of divergent *Alu* elements instead is replete with lncRNA-encoding genes**.** Sex-linked miRNAs are non-conserved and species-specific, which represent a potentially important source of novel functionalities during evolution [44, 88]. In humans, *L1-Alu* pairs were opted to accumulate on the X**.** *Alu*-derived miRNAs are expressed under stress conditions of the genome [88, 105].

5.1. Sex determination and TEs-derived ncRNAs" all roads lead to rome"

TEs copy numbers influence the genome size as well as, SD and mating system [19, 31]. The proposed mechanism is that TE insertions may be subject to differential gene expression between sexes similar to dosage compensation that happened in mammal X chromosome [31, 72]. However, TEs do these roles through ncRNAs that are involved in the regulatory networks governing SD and sex-specific gene expressions in PGCs, during the development of postnatal reproductive organs [106]. For example, the primate-specific X-linked clusters (e.g. miR-506-514 cluster, derived from DNA transposon) are preferentially expressed in testis, and are essential for male sex maturation in primates [66, 86]. The ncRNAs function not only in mammals, in sex determination and differentiation, but also in other species they are involved [31]. There is an example for SD in the WZ system in silkworm *Bombyx mori (B. mori)*, possibly analogous to those that exist in birds and reptiles. In this system, a single female-specific piRNA (named *Fem*) encoded by the W chromosome, is the primary determinant of sex in the silkworm. The full-length sense transposons expressed from the W are piRNA precursors that complex with the piwi-like protein Siwi [107, 108]. The piRNA-precursor *Fem* is derived from reiterative elements in the SDR, complexes with Siwi and targets the mRNA of the protein-coding gene Masculinizer (*Masc*). Targeting *Masc* leads to the production of a female-specific splice variant of *B. mori doublesex* (*Bmdsx*) that acting at the downstream of SD cascade. Silencing of the *Masc* messenger RNA by the *Fem* piRNA is required for the production of female-specific isoforms of Bmdsx in female embryos. The Masc protein controls both dosage compensation and masculinization in male embryos [107]. *Fem* piRNA forms a complex with Siwi (silkworm Piwi), which cleaves *Masc,* a Z-linked mRNA encoding a protein responsible for male differentiation and sexual maturation [108]. In the silkworm, TEs almost fully occupy the W chromosome with no functional protein-coding gene. *Female-enriched* piRNAs are defined as the only female-determination factor and transcripts produced from the sex-determining region of the W chromosome [107, 108].

Instead of genetic SD (GSD), there is another strategy in reptile for SD, wherein temperature determined sex (temperature sex determination; TSD) under the control of a specific gene [109]. Extreme temperatures can override GSD in a species with heteromorphic sex chromosomes (a ZZ/ZW system) and that an evolutionary transition from GSD to TSD can occur in one generation with loss of the W chromosome. Cold-inducible RNA-binding protein (CIRBP) is a TSD gene whose expression is induced during specification of gonad fate at female-producing temperatures. Induced expression of the CIRBP allele A in embryos exposed to a female-producing temperature leads to the development of ovaries. Clutches with higher *CIRBP* expression produce more females, while clutches with lower *CIRBP* expression produce more males. CIRBP plays a crucial role through its function in mRNA processing, RNA export, and translation. CIRBP expression reminds TEs roles in the stress-induced

expression of genes, where some TEs contain "Heat shock protein" heat-shock-like promoters that increasing their expression in temperature stress [15, 109]. Human *CIRBP* is located on chromosome 19 versus its paralogs including X-linked RNA-binding motif protein (RBMX), and the Y-linked RNA-binding motif protein family (RBMYs), are located on the sex chromosomes and expressed exclusively in the testis of normal people. As well, *CIRBP* is in the same family as *Sex-lethal* and *Transformer*, which are sex-determining genes in fruit flies [109].

About half of the transcriptional units on the Y chromosome of mammals encode lncRNAs, that are preferentially or exclusively expressed in the testis and involved in spermatogenesis and SD. A major source of testis-specific lncRNAs is antisense-TEs transcripts as precursors for piRNAs. For example; two lncRNAs Nct1 and Nct2 that are exclusively expressed in testis and spermatocytes, are precursors of piRNAs. In contrast to X, the mammalian Y-chromosome does not contain any miRNAs and the majority of miRNAs expressed in the testis are X-encoded and involved in MSCI of spermatogenesis [66].

6. Conclusions

Herein, TEs can be viewed as ancestors of current and ancient parasitic and non-parasitic genetic elements, as well as, as a raw material of primitive genomes. TEs were opted as tools for genetic innovation and for increasing organizational complexity in the life on the earth, such as the emergence of unicellular and multi-cellular organisms. TEs participated in the birth of non-conserved and lineage-specific RNAs which leading to the lineage speciation on the earth. In primates, there are many non-conserved and lineage-specific lncRNAs and miRNAs derived from primate-specific TEs. In particular, *Alu* elements that are primate-specific TEs led to the development of specific molecular pathways in the primate brains. Beside lineage speciation, TEs have appeared to play major roles in the development of sex chromosomes and lifespan determination.

Declarations

Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Data will be made available on request.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] E.V. Koonin, V.V. Dolja, M. Krupovic, Origins and evolution of viruses of eukaryotes: the ultimate modularity, Virology 479–80 (2015) 2–25.
[2] M. Krupovic, E.V. Koonin, Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems, Curr. Opin. Microbiol. 31 (2016) 25–33.
[3] E.V. Koonin, T.G. Senkevich, V.V. Dolja, The ancient Virus World and evolution of cells, Biol. Direct 1 (29) (2006) 1–27.
[4] C. Biémont, A brief history of the status of transposable elements: from junk DNA to major players in evolution, Genetics 186 (4) (2010) 1085–1093.
[5] H. Seligmann, D. Raoult, Unifying view of stem–loop hairpin RNA as origin of current and ancient parasitic and non-parasitic RNAs, including in giant viruses, Curr. Opin. Microbiol. 31 (2016) 1–8.
[6] E.V. Koonin, Viruses and mobile elements as drivers of evolutionary transitions, Phil. Trans. R. Soc. B 371 (2016) 2–13.
[7] N.V. Fedoroff, Transposable elements, epigenetics, and genome evolution, Science 338 (2012) 758–767.
[8] K.R. Oliver, GreeneWK, Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-Thrust hypothesis, Ecol. Evol. 2 (11) (2012 Nov) 2912–2933.
[9] C. Feschotte, E.J. Pritham, DNA transposons and the evolution of eukaryotic genomes, Annu. Rev. Genet. 41 (2007) 331–368.
[10] C.1 Vieira, C. Nardon, C. Arpin, D. Lepetit, C. Biémont, Evolution of genome size in Drosophila. is the invader's genome being invaded by transposable elements? Mol. Biol. Evol. 19 (7) (2002) 1154–1161.
[11] D. Jangam, C. Feschotte, E. Betrán, Transposable element domestication as an adaptation to evolutionary conflicts, Trends Genet. 33 (11) (2017) 817–831.
[12] M. Hadjiargyrou, N. Delihas, The intertwining of transposable elements and non-coding RNAs, Int. J. Mol. Sci. 14 (7) (2013) 13307–13328.
[13] R.E. Mills, E.A. Bennett, R.C. Iskow, S.E. Devine, Which transposable elements are active in the human genome? Trends Genet. 23 (4) (2007) 183–191.
[14] M. Krupovic, E.V. Koonin, Self-synthesizing transposons: unexpected key players in the evolution of viruses and defense systems, Curr. Opin. Microbiol. 31 (2016) 25–33.
[15] P. Deininger, Alu elements: know the SINEs, Genome Biol. 12 (12) (2011) 236.
[16] S. Kannan, D. Chernikova, I.B. Rogozin, E. Poliakov, D. Managadze, E.V. Koonin, L. Milanes, Transposable element insertions in long intergenic non-coding rna genes, Front. Bioeng. Biotechnol. 3 (71) (2015) 1–9.
[17] A.F. Smit, Interspersed repeats and other mementos of transposable elements in mammalian genomes, Curr. Opin. Genet. Dev. 9 (6) (1999) 657–663.
[18] B. Chénais, A. Caruso, S. Hiard, N. Casse, The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments, Gene 509 (2012) 7–15.
[19] D. Chalopin, J.N. Volff, D. Galiana, J.L. Anderson, M. Schartl, Transposable elements and early evolution of sex chromosomes in fish. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology, Chromosome Res. 23 (3) (2015) 545–560.
[20] J. Macas, P. Novák, J. Pellicer, J. Čížková, A. Koblížková, P. Neumann, I. Fuková, J. Doležel, L.J. Kelly, I.J. Leitch, In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae, PloS One 10 (11) (2015), e0143424.
[21] A. Belyayev, Bursts of transposable elements as an evolutionary driving force, J. Evol. Biol. 27 (12) (2014) 2573–2584.
[22] S.K. Kota, R. Feil, Epigenetic transitions in germ cell development and meiosis, Dev. Cell 19 (5) (2010) 675–686.
[23] J.T. Lee, M.S. Bartolomei, X-inactivation, imprinting, and long noncoding RNAs in health and disease, Cell 152 (6) (2013) 1308–1323.
[24] J.M. Calabrese, T. Magnuson, Roles of long non-coding RNAs in X-chromosome inactivation, Mol. Biol. Long Non-coding RNAs (2013) 69–94.
[25] S.F. Li, G.J. Zhang, J.H. Yuan, C.L. Deng, W.J. Gao, Repetitive sequences and epigenetic modification: inseparable partners play important roles in the evolution of plant sex chromosomes, Planta 243 (5) (2016) 1083–1095.
[26] R. Johnson, R. Guigó, The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs, RNA 20 (7) (2014) 959–976.
[27] A. Kapusta, Z. Kronenberg, V.J. Lynch, X. Zhuo, L. Ramsay, G. Bourque, M. Yandell, C. Feschotte, Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs, PLoS Genet. 9 (4) (2013), e1003470.
[28] G. Barry, Integrating the roles of long and small non-coding RNA in brain function and disease, Mol. Psychiatr. 19 (4) (2014) 410–416.
[29] Á. Sturm, Z. Ivics, T. Vellai, The mechanism of ageing: primary role of transposable elements in genome disintegration, Cell. Mol. Life Sci. 72 (10) (2015) 1839–1847.
[30] A.J. Flavell, Retroelements, reverse transcriptase and evolution, Comp. Biochem. Physiol. 2 (1) (1995) 3–15.
[31] J.D. Hollister, L.M. Smith, Y.L. Guo, F. Ott, D. Weilugel, B.S. Gaut, Transposable elements and small RNAs contribute to gene expression divergence between Arabidopsis thaliana and Arabidopsis lyrata, Proc. Natl. Acad. Sci. U. S. A. 108 (6) (2011) 2322–2327.
[32] J. Bouckenheimer, P. Fauque, C.H. Lecellier, C. Bruno, T. Commes, J.M. Lemaître, J. De Vos, S. Assou, Differential long non-coding RNA expression profiles in human oocytes and cumulus cells, Sci. Rep. 8 (1) (2018) 2202.
[33] M.J. Madison-Villar, C.1 Sun, N.C. Lau, M.L. Settles, R.L. Mueller, Small RNAs from a big genome: the piRNA pathway and transposable elements in the salamander species Desmognathus fuscus, J. Mol. Evol. 83 (3-4) (2016) 126–136.
[34] C.G. Sotero-Caio, R.N. Platt, A. Suh, D.A. Ray, Genome Biol evol. Evolution and diversity of transposable elements in vertebrate genomes, Gen. Biol. Evol. 9 (1) (2017) 161–177.
[35] W.C. Orr, Tightening the connection between transposable element mobilization and aging, Proc. Natl. Acad. Sci. U. S. A. 113 (40) (2016) 11069–1106170. 1.
[36] R.S. Petralia, M.P. Mattson, P.J. Yao, Aging and longevity in the simplest animals and the quest for immortality, Ageing Res. Rev. 16 (2014) 66–82.
[37] K. Mochizuki, Developmentally programmed, RNA-directed genome rearrangement in Tetrahymena, Dev. Growth Differ. 54 (1) (2012) 108–119.
[38] N. Sela, E. Kim, G. Ast, 70. Tehseaerc hrole of transposable elements in the evolution of non-mammalian vertebrates and invertebrates, Genome Biol. 11 (6) (2010) R59.

[39] Ö. Deniz, J.M. Frost, M.R. Branco, Regulation of transposable elements by DNA modifications, Nat. Rev. Genet. 20 (7) (2019) 417–431.
[40] A. Bousios, B.S. Gaut, Mechanistic and evolutionary questions about epigenetic conflicts between transposable elements and their plant hosts, Curr. Opin. Plant Biol. 30 (2016) 123–133.
[41] H.M. Awan, A. Shah, F. Rashid, G. Shan, Primate-specific long non-coding RNAs and MicroRNAs, Dev. Reprod. Biol. 15 (3) (2017) 187–195.
[42] O.E. Mustafina, The possible roles of human Alu elements in aging, Front. Genet. 4 (2013) 96.
[43] J.1 González, K. Lenkov, M. Lipatov, J.M. Macpherson, D.A. Petrov, High rate of recent transposable element-induced adaptation in Drosophila melanogaster, PLoS Biol. 6 (10) (2008) e251.
[44] D.J. Obbard, D.J. Finnegan, RNA interference: endogenous siRNAs derived from transposable elements, Curr. Biol. 18 (13) (2008 8) R561–R563.
[45] M. Nosaka, J. Itoh, Y. Nagato, A. Ono, A. Ishiwata, Y. Sato, Role of transposon-derived small RNAs in the interplay between genomes and parasitic DNA in rice, PLoS Genet. 8 (9) (2012), e1002953.
[46] K.C.1 Wang, H.Y. Chang, Molecular mechanisms of long noncoding RNAs, Mol. Cell 43 (6) (2011) 904–914.
[47] J.P. Abad, B. De Pablos, K. Osoegawa, P.J. De Jong, A. Martín-Gallardo, A. Villasante, TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres, Mol. Biol. Evol. 21 (9) (2004) 1620–1624.
[48] S. Rashkova, S.E. Karam, R. Kellum, M.L. Pardue, Gag proteins of the two Drosophila telomeric retrotransposons are targeted to chromosome ends, J. Cell Biol. 159 (3) (2002) 397–402.
[49] M.L. Pardue, O.N. Danilevskaya, K.L. Traverse, K. Lowenhaupt, Evolutionary links between telomeres and transposable elements, Genetica 100 (1-3) (1997) 73–84.
[50] H.C. Kopera, J.B. Moldovan, T.A. Morrish, J.L. Garcia-Perez, J.V. Moran, Similarities between long interspersed element-1 (LINE-1) reverse transcriptase and telomerase, Proc. Natl. Acad. Sci. U. S. A. 108 (51) (2011) 20345–20350.
[51] J.W. Shay, Role of telomeres and telomerase in aging and cancer, Canc. Discov. 6 (6) (2016) 584–593.
[52] J. Piriyapongsa, L. Mariño-Ramírez, I.K. Jordan, Origin and evolution of human microRNAs from transposable elements, Genetics 176 (2) (2007) 1323–1337.
[53] R. Rebollo, Y. Zhang, D.L. Mager, Transposable elements: not as quiet as a mouse, Genome Biol. 13 (6) (2012) 159.
[54] S. Ganesh, P. Svoboda, Retrotransposon-associated long non-coding RNAs in mice and men, Pflügers Archiv 468 (6) (2016) 1049–1060.
[55] Á. Sturm, A. Perczel, Z. Ivics, T. Vellai, The Piwi-piRNA pathway: road to immortality, Aging Cell 16 (5) (2017) 906–911.
[56] S. Kitano, H. Kurasawa, Y. Aizawa, Transposable elements shape the human proteome landscape via formation of cis-acting upstream open reading frames, Gene Cell. 23 (4) (2018) 274–284.
[57] A.D. Bailey, C.K. Shen, Sequential insertion of Alu family repeats into specific genomic sites of higher primates, Proc. Natl. Acad. Sci. U. S. A. 90 (15) (1993) 7205–7209.
[58] S. Ayarpadikannan, H.S. Kim, The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases, Genomics Inform. 12 (3) (2014) 98–104.
[59] R.J. Britten, Transposable element insertions have strongly affected human evolution, Proc. Natl. Acad. Sci. U. S. A. 107 (46) (2010) 19945–19948.
[60] S.J. Klein, R.J. O'Neill, Transposable elements: genome innovation, chromosome diversity, and centromere conflict, Chromosome Res. 26 (1-2) (2018) 5–23.
[61] A. Varki, T.K. Altheide, Comparing the human and chimpanzee genomes: searching for needles in a haystack, Genome Res. 15 (12) (2005) 1746–1758.
[62] J.T. Lee, Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome, Genes Dev. 23 (16) (2009) 1831–1842.
[63] A. Fatica, I. Bozzoni, Long non-coding RNAs: new players in cell differentiation and development, Nat. Rev. Genet. 15 (1) (2014) 7–21.
[64] Y.R. Fujii, RNA genes: retroelements and virally retroposable microRNAs in human embryonic stem cells, Open Virol. J. 4 (2010) 63–75.
[65] R. Zhang, Y.Q. Wang, B. Su, Molecular evolution of a primate-specific microRNA family, Mol. Biol. Evol. 25 (7) (2008) 1493–1502.
[66] Z. Sun, Y. Zhang, R. Zhang, X. Qi, B. Su, Functional divergence of the rapidly evolving miR-513 subfamily in primates, BMC Evol. Biol. 13 (2013) 255.
[67] M. Mirkovic-Hö sle, K. Förstemann, Transposon defense by endo-siRNAs, piRNAs and somatic pilRNAs in Drosophila: contributions of loqs-PD and R2D2, PloS One 9 (1) (2014), e84994.
[68] R. Shalgi, Y. Pilpel, M. Oren, Repression of transposable-elements-a microRNA anti-cancer defense mechanism? Trends Genet. 26 (6) (2010) 253–259.
[69] J. Piriyapongsa, I.K. Jordan, Dual coding of siRNAs and miRNAs by plant transposable elements, RNA 14 (5) (2008) 814–821.
[70] J. Chen, M. Sun, L.D. Hurst, G.G. Carmichael, J.D. Rowley, Genome-wide analysis of coordinate expression and evolution of human cis-encoded sense-antisense transcripts, Trends Genet. 21 (6) (2005) 326–329.
[71] S.A. Krawetz, A. Kruger, C. Lalancette, R. Tagett, E. Anton, S. Draghici, M.P. Diamond, A survey of small RNAs in human sperm, Hum. Reprod. 26 (12) (2011) 3401–3412.
[72] S. Shpiz, S. Ryazansky, I. Olovnikov, Y. Abramov, A. Kalmykova, Euchromatic transposon insertions trigger production of novel Pi- and endo-siRNAs at the target sites in the Drosophila germline, PLoS Genet. 10 (2) (2014), e1004138.
[73] N. Kotaja, MicroRNAs and spermatogenesis, Fertil. Steril. 101 (6) (2014) 1552–1562.
[74] S. Hilz, A.J. Modzelewski, P.E. Cohen, A. Grimson, The roles of microRNAs and siRNAs in mammalian Spermatogenesis, Development 143 (17) (2016) 3061–3073.
[75] H. Kobayashi, T. Sakurai, M. Imai, N. Takahashi, A. Fukuda, O. Yayoi, S. Sato, K. Nakabayashi, K. Hata, Y. Sotomaru, Y. Suzuki, T. Kono, Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks, PLoS Genet. 8 (1) (2012), e1002440.
[76] J. García-López, L. Alonso, D.B. Cárdenas, H. Artaza-Alvarez, D. Hourcade Jde, S. Martínez, M.A. Brieño-Enríquez, J. Del Mazo, Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization, RNA 21 (5) (2015) 946–962.
[77] F. Pourrajab, A. Vakili Zarch, S. Hekmatimoghaddam, M.R. Zare-Khormizi, The master switchers in the aging of cardiovascular system, reverse senescence by microRNA signatures; as highly conserved molecules, Prog. Biophys. Mol. Biol. 119 (2) (2015) 111–128.
[78] J.H. Martin, E.G. Bromfield, R.J. Aitken, B. Nixon, Biochemical alterations in the oocyte in support of early embryonic development, Cell. Mol. Life Sci. 74 (3) (2017) 469–485.
[79] R. Karlic, S. Ganesh, V. Franke, E. Svobodova, J. Urbanova, Y. Suzuki, F. Aoki, K. Vlahovicek, P. Svoboda, Long non-coding RNA exchange during the oocyte-to-embryo transition in mice, DNA Res. 24 (2) (2017) 129–141.
[80] T.1 Watanabe, Y. Totoki, A. Toyoda, M. Kaneda, S. Kuramochi-Miyagawa, Y. Obata, H. Chiba, Y. Kohara, T. Kono, T. Nakano, M.A. Surani, Y. Sakaki, H. Sasaki, Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes, Nature 453 (7194) (2008) 539–543.
[81] S. Lehnert, P. Van Loo, P.J. Thilakarathne, P. Marynen, G. Verbeke, C. Frans, Schuit. Evidence for Co-evolution between human MicroRNAs and alu-repeats, PloS One slu4 (2) (2009), e4456.
[82] T. Chishima, J. Iwakiri, M. Hamada, Identification of transposable elements contributing to tissue-specific expression of long non-coding RNAs, Genes 9 (23) (2018) 1–14.
[83] J. Piriyapongsa, I.K. Jordan, A family of human MicroRNA genes from miniature inverted-repeat transposable elements, PloS One 2 (2) (2007), e203.
[84] P.N. Nguyen, C.J. Huang, S. Sugii, S.K. Cheong, K.B. Choo, Selective activation of miRNAs of the primate-specific chromosome 19 miRNA cluster (C19MC) in cancer and stem cells and possible contribution to regulation of apoptosis, J. Biomed. Sci. 24 (1) (2017) 20.
[85] M. Noguer-Dance, S. Abu-Amero, M. Al-Khtib, A. Lefèvre, P. Coullin, G.E. Moore, J. Cavaillé, The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta, Hum. Mol. Genet. 19 (18) (2010) 3566–3582.
[86] R. Zhang, Y. Peng, W. Wang, B. Su, Rapid evolution of an X-linked microRNA cluster in primates, Genome Res. 17 (5) (2007) 612–617.
[87] J. Li, Y. Liu, D. Dong, Z. Zhang, Evolution of an X-linked primate-specific micro RNA cluster, Mol. Biol. Evol. 27 (3) (2010) 671–683.
[88] X. Guo, B. Su, Z. Zhou, J. Sha, Rapid evolution of mammalian X-linked testis microRNAs, BMC Genom. 10 (2009) 97.
[89] R. Song, S. Ro, J.D. Michaels, C. Park, J.R. McCarrey, W. Yan, Many X-linked microRNAs escape meiotic sex chromosome inactivation, Nat. Genet. 41 (4) (2009) 488–493.
[90] N. Ghasemzadeh, F. Pourrajab, A. Dehghani Firoozabadi, S. Hekmatimoghaddam, F. Haghiralsadat, Ectopic microRNAs used to preserve human mesenchymal stem cell potency and epigenetics, EXCLI J 17 (2018) 576–589.
[91] J.E. Froberg, L. Yang, J.T. Lee, Guided by RNAs: X-inactivation as a model for lncRNA function, J. Mol. Biol. 425 (19) (2013) 3698–3706.
[92] W. Yan, J.R. McCarrey, Sex chromosome inactivation in the male, Epigenetics 4 (7) (2009) 452–456.
[93] I. Harel, B.A. Benayoun, B. Machado, P.P. Singh, C.K. Hu, M.F. Pech, D.R. Valenzano, E. Zhang, S.C. Sharp, S.E. Artandi, A. Brunet, A Platform for rapid exploration of aging and diseases in a naturally short-lived vertebrate, Cell 160 (5) (2015) 1013–1026.
[94] D.R. Valenzano, B.A. Benayoun, P.P. Singh, E. Zhang, P.D. Etter, C.K. Hu, M. Clément-Ziza, D. Willemsen, R. Cui, I. Harel, B.E. Machado, M.C. Yee, S.C. Sharp, C.D. Bustamante, A. Beyer, E.A. Johnson, A. Brunet, The african turquoise killifish genome provides insights into evolution and genetic architecture of lifespan, Cell 163 (2015) 1539–1554.
[95] K. Reichwald, A. Petzold, P. Koch, B.R. Downie, N. Hartmann, S. Pietsch, M. Baumgart, D. Chalopin, M. Felder, M. Bens, A. Sahm, K. Szafranski, S. Taudien, M. Groth, I. Arisi, A. Weise, S.S. Bhatt, V. Sharma, J.M. Kraus, F. Schmid, S. Priebe, T. Liehr, M. Görlach, M.E. Than, M. Hiller, H.A. Kestler, J.N. Volff, M. Schartl, A. Cellerino, C. Englert, M. Platzer, Insights into sex chromosome evolution and aging from the genome of a short-lived fish, Cell 163 (6) (2015 Dec 3) 1527–1538.
[96] M. Keane, J. Semeiks, A.E. Webb, Y.I. Li, V. Quesada, T. Craig, L.B. Madsen, S. van Dam, D. Brawand, P.I. Marques, P. Michalak, L. Kang, J. Bhak, H.S. Yim, N.V. Grishin, N.H. Nielsen, M.P. Heide-Jørgensen, E.M. Oziolor, C.W. Matson, G.M. Church, G.W. Stuart, J.C. Patton, J.C. George, R. Suydam, K. Larsen, C. López-Otín, M.J. O'Connell, J.W. Bickham, B. Thomsen, J.P. de Magalhães, Insights into the evolution of longevity from the bowhead whale genome, Cell Rep. 10 (1) (2015) 112–122.
[97] M. Sulak, L. Fong, K. Mika, S. Chigurupati, L. Yon, N.P. Mongan, R.D. Emes, V.J. Lynch, TP53 copy number expansion is associated with the evolution of

increased body size and an enhanced DNA damage response in elephants, Elife 5 (2016), e11994.

[98] S. Hekmatimoghaddam, A. Dehghani Firoozabadi, M.R. Zare-Khormizi, F. Pourrajab, Sirt1 and Parp1 as epigenome safeguards and microRNAs as SASP-associated signals, in cellular senescence and aging, Ageing Res. Rev. 40 (2017) 120–141.

[99] K. Saito, M.C. Siomi, Small RNA-mediated quiescence of transposable elements in animals, Dev. Cell 19 (5) (2010) 687–697.

[100] F. Pourrajab, F. Torkian Velashani, M. Khanaghaei, S. Hekmatimoghaddam, M. Rahaie, M.R. Zare-Khormizi, Comparison of miRNA signature versus conventional biomarkers before and after off-pump coronary artery bypass graft, J. Pharmaceut. Biomed. Anal. 134 (2017) 11–17.

[101] K. Mochizuki, Developmentally programmed, RNA-directed genome rearrangement in Tetrahymena, Dev. Growth Differ. 54 (1) (2012) 108–119.

[102] N. Hamazaki, M. Uesaka, K. Nakashima, K. Agata, T. Imamura, Gene activation-associated long noncoding RNAs function in mouse preimplantation development, Development 142 (5) (2015) 910–920.

[103] T. Ezaz, J. Deakin, Repetitive sequence and sex chromosome evolution in vertebrates, Adv. Evol. Biol. 1 (1) (2014) 1–9.

[104] R. Erlandsson, J.F. Wilson, S. Pääbo, Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans, Mol. Biol. Evol. 17 (5) (2000 May) 804–812.

[105] G. Abrusán, J. Giordano, P.E. Warburton, Analysis of transposon interruptions suggests selection for L1 elements on the X chromosome, PLoS Genet. 4 (8) (2008), e1000172.

[106] L. McFarlane, D. Wilhelm, Non-coding RNAs in mammalian sexual development, Sex Dev. 3 (2009) 302–316.

[107] T. Kiuchi, H. Koga, M. Kawamoto, K. Shoji, H. Sakai, Y. Arai, G. Ishihara, S. Kawaoka, S. Sugano, T. Shimada, Y. Suzuki, M.G. Suzuki, S. Katsuma, A single female-specific piRNA is the primary determiner of sex in the silkworm, Nature 509 (7502) (2014) 633–636.

[108] S.1 Katsuma, M. Kawamoto, T. Kiuchi, Guardian small RNAs and sex determination, RNA Biol. 11 (10) (2014) 1238–1242.

[109] A.L. Schroeder, K.J. Metzger, A. Miller, T. Rhen, A novel candidate gene for temperature-dependent sex determination in the common snapping turtle, Genetics 203 (1) (2016) 557–571.