ISSN 1062-3604, Russian Journal of Developmental Biology, 2018, Vol. 49, No. 6, pp. 339–355. © Pleiades Publishing, Inc., 2018.
Original Russian Text © R.N. Mustafin, 2018, published in Ontogenez, 2018, Vol. 49, No. 4S, pp. S3–S20.

REVIEWS

Functional Dualism of Transposon Transcripts in Evolution of Eukaryotic Genomes

R. N. Mustafin[a, b, *]

[a] *Bashkir State Medical University, Ufa, 450008 Russia*
[b] *Bashkir State University, Ufa, 450076 Russia*
**e-mail: ruji79@mail.ru*

Received August 11, 2018; in final form, October 30, 2018

Abstract—The ability of transposons to unite genes separated by their insertions encoding common biological processes into regulatory networks contributed, simultaneously with the complication of eukaryotes, to their evolutionary success by forming new universal systems. By means of these systems, including DNA methylation, histone modifications, the relationship of telomeres with transposons, splicing regulation, and RNA interference, global distribution of transposons in the genomes was accompanied by the emergence of their structural innovations, dynamic regulatory sequences, and protein-coding genes. The mobile elements contributed to the evolution of protein-coding genes by their duplication, as well as exonization, and domestication of the transposons themselves. The resulting new genes contain transposon sequences involved in their management by means of regulatory networks and noncoding RNAs also originating from the mobile elements. A strategy wherein the translation of noncoding RNA genes contributed to the selection of the obtained polypeptides as functional cellular proteins was developed during evolution. At the same time, noncoding RNAs are also processed into molecules involved in the regulatory processes independently or as a part of the protein complexes. The duality of functions was inherent to all noncoding RNAs whose nonrandom decay/processing leads to the formation of molecules that have a regulatory effect on the transposons and protein-coding genes. A strategy wherein primary transposon transcripts interact with different systems of their processing (arisen to protect the hosts from transposons), forming functional RNA molecules translated into the peptides, was developed in the evolution of eukaryotes. The transposons are universal sources for these strategies; this explains their global distribution in eukaryotic genomes and domestication in the system of "double search" for targets for functional interaction of noncoding RNAs and processed products of their translation. In addition to splicing, primary transcripts of some protein-coding genes can also be processed in functional noncoding RNAs involved in common biological reactions with the gene protein product. This substantiates the associations of multifactorial diseases with the gene SNP since they can cause inactivation of RNA domains. It was suggested that functional dualism of the transposon transcripts could be an important condition of the emergence of life, while the mobile elements are one of fundamental properties of living.

Keywords: protein-coding genes, histones, noncoding RNAs, methylation, peptides, transposons
DOI: 10.1134/S1062360418070019

ABBREVIATIONS

PCGs, protein-coding genes.

snRNAs, small nuclear RNAs (a class of small ncRNAs that are found in the nuclei of eukaryotes and are involved in the regulation of splicing, transcription factors, and in maintaining telomere integrity).

snoRNAs, small nucleolar RNAs (a class of small ncRNAs that are involved in methylation and pseudouridylation of transfer RNAs, small nuclear RNAs, and ribosomal RNAs.

ncRNAs, noncoding RNAs.

NSs, nucleotide sequences.

rRNAs, ribosomal RNAs.

TFs, transcription factors (the proteins that bind to specific DNA sequences and control mRNA transcription on DNA matrix).

tRNAs, transfer RNAs.

ERVs, endogenous retroviruses.

IGSs, intergenic spacers.

LINEs, long interspersed nuclear elements.

lncRNAs, long noncoding RNAs.

LTRs, long terminal repeats.

RT, reverse transcriptase.

sdRNAs, sno-derived RNAs.

SINEs, short interspersed nuclear elements.

SNPs, single nucleotide polymorphisms.

SR, sarcoplasmic reticulum.

TE, transposable elements.

tRFs, tRNA-derived RNA fragments.

tsRNAs, tRNA-derived small RNAs.

UTR, untranslated region.

INTRODUCTION

Transposons (transposable elements, TEs) are genetic objects that can change their localization in the genome (Startek et al., 2017). They differ in the structure and movement mechanism and are divided into classes, subclasses (for DNA transposons), orders, superfamilies, families, subfamilies. The class I retrotransposons (retroTEs) move by the "copy–paste" mechanism, and the class II DNA transposons (DNA-TE) by the "cut–paste" (TIR and Crypton orders) and "rolling circle" (Helitron and Maverick orders) mechanisms. The division into orders is conducted based on the common element organization, peculiarities of moving enzymes, and mechanisms of incorporation into the genome (Fig. 1) (Wicker et al., 2007). The most common TE classification was realized in Repbase (http://www.girinst.org/repbase/). The class I TE transposition is performed through an intermediate RNA link, when the initial sequence remains in the same place (Startek et al., 2017). All retro-TEs have two common genes: *gag* and *pol*. The *gag* genes encode group-specific antigen (GAG) proteins that form virus-like particles (VLPs). The *pol* genes encode the enzymes reverse transcriptase (RT), integrase (IN), and protease (PR) (Alzohairy et al., 2013). Both TE classes are found in prokaryotes and eukaryotes; however, retroelements in the latter were the most common, occupying from 3 to 85% of the genome nucleotide sequences (NSs) in different members (Startek et al., 2017). TEs are even contained in the genomes of giant viruses involved in their functioning (Filee, 2018). It was demonstrated that TEs as a part of pandoraviruses contribute to their evolutionary transformations (Zhang et al., 2018).

The exchange of auxiliary set genes using horizontal transfer is an important condition for the survival of bacteria and archaea. Therefore, TEs in prokaryotes also play a key role in evolution contributing to the emergence of new genes due to the transition of auxiliary genes to basic (Ravin and Shestakov, 2013) by means of retroelements. For example, the bacterium TEs usually transfer antibiotic resistance genes (Babakhani and Oloomi, 2018). Retrons were found in *E. coli* and *M. xanthus* in 1989 and DGR retroelements in the *Bordetella* genus in 2002. Despite reverse transcriptase coding, the latter do not show mobility but provide protection from bacteriophages. In addition, clustered regularly interspaced short palindromic repeats (CRISPR) and Abi, also encoding RT and protecting them from alien sequences, were found in the genomes of bacteria (Toro and Nisa-Martinez, 2014). More than 50% of RT bacteria are encoded by the group II introns that are divided into several main groups: A, B, C, D, E, F, G, CL1/2 (chloroplast-like), and ML (mitochondrion-like) (Toro et al., 2018). The classification of TEs, plasmids, and phages of prokaryotes is presented in the ACLAME database (http://aclame.ulb.ac.be) (Leplae et al., 2004). However, despite the presence of TEs and complex protection systems, prokaryotic genomes are much simpler than eukaryotic and contain much fewer TEs, which is reflected in their sizes. The emergence of a number of conservative mechanisms contributing to the use of new TE insertions for variability and adaptation in evolution is a reason for complications in the structure and functioning of eukaryotic genomes. Horizontal gene transfer, which can also be global like endogenous TE distribution in eukaryotic genomes, is more important in the variability of prokaryotes. For example, two enterobacteria (*Escherichia coli* and *Salmonella*) are separated by approximately 70% of their genes; similar divergence level is observed between the genomes of two *E. coli* strains (laboratory K12 strain and pathogenic O157-H7 that differ by 30% of their genomes). Most differences are explained by prophages. Similarly, the main differences between the *Listeria monocytogenes* and *Listeria innocua* genomes correspond to prophages integrated into the latter (Leplae et al., 2004).

Since the activity of TEs (like other forms of mutations) can be deleterious in terms of phenotypic effects, understanding of the wide spread of transposons in eukaryotic genomes requires analyzing the meaning of evolutionary accumulation of TEs. In this respect, inactivation of TEs due to the accumulation of mutations by them, silencing of TE and their domestication by the hosts, and horizontal TE transfer are of significance (Song and Schaack, 2018). However, TEs are key components of the adaptive evolution of eukaryotes (Barry, 2018). Despite the development and perfection of the mechanisms to combat selfish sequences, the distribution of TEs in eukaryotic genomes is characterized by a global scale, while TE protection systems contributed to the emergence of fundamental differences from prokaryotic genomes. These differences include nucleosome DNA packaging with the possibility of its management by means of histone modification and DNA methylation, linear structure of the chromosomes with telomeres at the ends, global distribution of introns and retrogens, and use of noncoding RNAs (ncRNAs) in managing all biological processes. Eukaryotes evolved from archaea, many of which inhabit aggressive conditions (Shestakov, 2017; Henikoff and Smith, 2015) that contribute to selecting the organisms with the development of the maximum possible number of biomolecules required for the appearance of new adaptive properties for survival. This could be a reason for TE distribution and the development of universal systems of their usage by the

Class I MGE—retrotransposons

Order	Superfamily	Structural organization
LTR	*Copia*	GAG AP INT RT RH
	Gypsy	GAG AP RT RH INT
	Bel-Pao	GAG AP RT RH INT
	Retrovirus	GAG AP RT RH INT ENV
	ERV	GAG AP RT RH INT ENV
DIRS	*DIRS*	GAG AP RT RH YR
	Ngaro	GAG AP RT RH YR
	VIPER	GAG AP RT RH YR
PLE	*Penelope*	RT EN
LINE	*R2*	RT EN
	RTE	APE RT
	Jockey	ORF1 APE RT
	LI	ORF1 APE RT
	I	ORF1 APE RT RH
SINE	*tRNA*	
	7SL	
	5S	

Class II MGE—DNA transposons

Order	Superfamily	Structural organization
TIR	*Tc1/mairner*	TRANSPOSASE
	hAT	
	Mutator	
	Merlin	
	Transib	
	P	
	PiggyBac	
	PIF-Harbinger	TRANSPOSASE ORF2
	CACTA	TRANSPOSASE ORF2
Crypton	*Crypton*	YR
Helitron	*Helitron*	RPA Y2HEL
Maverick	*Maverick*	C-INT ATP CYP POL B

Fig. 1. Transposon classification according to Wicker. LTR, long terminal repeats; GAG, capsid protein; AP, aspartic proteinase; INT, integrase; RT, reverse transcriptase; RH, RNase H; TIR, terminal inverted repeats; ENV, envelope protein; YR, tyrosine recombinase; EN, endonucleases; APE, apurinic endonucleases; ORF, open reading frame; RPA, replication protein A; HEL, helicase; ATP, packaging ATPase; CYP, cysteine protease; POLB, DNA polymerase B.

host. The ability of TEs to change localization within the genome leads to the emergence of new structural and functional activities that contribute to the evolution of genomes and speciation (Alzohairy et al., 2013).

In eukaryotes, endogenous distribution of TEs in the evolution led to the use of their NSs by the hosts on a global scale. TEs became key sources of ncRNAs (Borchert et al., 2011; Li et al., 2011; Yuan et al., 2011; Roberts et al., 2013; Gim et al., 2014; Johnson and Guigo, 2014; Kapusta and Feschotte, 2014; Lorenzetti et al., 2016; Cho, 2018), protein-coding genes (PCGs), including the transcription factors (TFs) (Zdobnov et al., 2005; Volff, 2006; Feschotte, 2008; Sinzelle et al., 2009; Alzohairy et al., 2013; Garavis et al., 2013; Duan et al., 2017; Schrader and Schmitz, 2018), as well as TF binding sites and other regulatory structures (de Souza et al., 2013; Jaques et al., 2013; Mak et al., 2014; Ito et al., 2017). In addition, TEs can be a source of satellite tandem arrays in the genomes (McGurk and Barbash, 2018). At the same time, TEs change (due to the accumulation of mutations) beyond recognition and become integral conservative structures and genes. It is assumed that TEs constitute a much larger part of eukaryotic genomes than identified by standard research methods (Goerner-Potvin and Bourque, 2018). For example, it is considered that TEs occupy approximately 45% of all human DNA; however, the analysis by means of specific oligonucleotides recognizing TE fragments demonstrated that their portion is much higher and occupies more than 60% of the genome (de Koning et al., 2011). At present, the identification methods for "hidden" TEs in the genomes of different species are being actively developed to determine more accurately their role in the development of regulatory NSs and PCGs (Goerner-Potvin and Bourque, 2018). Even taking into account the imperfections of the detection methods, according to modern data, TEs became sources of a significant part of NSs of the genomes located in intergenic regions and in PCG introns, in eukaryotes

in the evolution. For example, TEs in the Drosophila genome make up to 22, up to 40 in mouse, and up to 90% in plants (Yuan et al., 2011). Due to global TE distribution, the sizes of eukaryotic genomes differ by 70 000 times from 2.3 (*Saccharomyces cerevisiae*) to 150 000 million base pairs (*Paris japonica*). For comparison, the minimal prokaryotic genome size in the *Micoplasma genitalium* is less than the maximal in the *Myxococcus xanthus* by just 16 times (Patrushev and Minkevich, 2007). Despite the fact that the portion of NSs encoding the proteins in eukaryotic genomes is from a part to several percent of all DNA, TEs were the most important sources of their emergence and evolution by duplicating the existing genes (Huang et al., 2009; Sakai et al., 2011; Grandi et al., 2015; Tan et al., 2016; Zhu et al., 2016; Kubiak and Makalowska, 2017; Cerbin and Jiang, 2018), exonization (Wang et al., 2012; Elkon et al., 2013; Tajnik et al., 2015), and domestication of TEs themselves (Zdobnov et al., 2005; Volff et al., 2006; Feschotte, 2008; Sinzelle et al., 2009; Alzohairy et al., 2013; Garavis et al., 2013; Duan et al., 2017; Schrader and Schmitz, 2018).

EMERGENCE OF NEW PROTEINS USING TRANSPOSONS

TEs contribute to the host gene duplication by insertion of processed RNA products of the genes by means of reverse transcriptase (Cerbin and Jiang, 2018). At the same time, the resulting retrocopies are flanked by transposon NSs, which contributes to their management by regulatory networks with TEs (Schrader and Schmitz, 2018). For example, retrocopies in mammals contain endonuclease cleavage site TTTT/AA and poly(A) tail specific for LINE-1 (long interspersed nuclear elements) in flanking regions. In Drosophila, retrogens can be flanked by long terminal repeats (LTRs) (Tan et al., 2016). In plants (including Arabidopsis, rice, and many other species), many retrogens were also identified, some of which have functionality (for example, the *Sun* gene in tomatoes, the *CYP98A8* and *CYP98A9* genes in Arabidopsis). The plant retrocopies are mainly flanked by LTRs, since LTR-containing TEs mainly contribute to their formation, although L1-like retroelements are also involved in this process (Zhu et al., 2016). Intron-free retrogens can acquire new introns, the preservation of which also indicates the inclusion of retrocopies in the functioning of regulatory networks (both as sources of transcription and translation products). Due to this, retrogens can even functionally substitute the parental gene or develop a new function (Kubiak and Makalowska, 2017). Thus, TEs contribute to protein diversity in the evolution due to retrotransposition of the existing PCGs, the preservation of which contributes to finding interrelations by the genes in developing new regulatory networks in the evolution. It is assumed that the number of retrogens is much greater than that determined by standard identification methods due to 5'-truncation, although approximately 10% of all PCGs contain one and more retrocopies even taking this into account (Grandi et al., 2015). For example, more than 1000 retrogens are contained in the human genome that are transcribed and show biological activity (Huang et al., 2009). In rice, the expression of at least 66% of detected retrocopies was determined; however, they were with lower values than their initial genes but with tissue-specific correlation of their levels (Sakai et al., 2011).

The domestication of TE genes was also a common way of the emergence of new PCGs in the evolution (Joly-Lopez and Bureau, 2018; Schrader and Schmitz, 2018). Telomerase, which is conservative for most eukaryotes and originated from retroelement RT, is the most striking example (Garavis et al., 2013). Telomerase has significant similarities with RT of L1 elements (Kopera et al., 2011), while the L1 is mainly transposed in the telomere region (helping to preserve their function) in the cell lines of Chinese hamster ovaries with damaged telomere function (Morrish et al., 2007). The loss of the telomerase gene in Drosophila in the evolution was successfully substituted by retro-TEs HeT-A, TART, and TAHRE (Casacuberta, 2017). Many genes originating from TEs (especially from transposase) were detected in animals, plants, and yeasts (Feschotte, 2008). In vertebrates, 1000 genes originating from retro-TEs were detected (Zdobnov et al., 2005). And newly formed PCGs can encode TFs due to the use of DNA-binding TE domains and carry out the most important functions in managing the genome work. At the same time, many TF binding sites also come from the transposon NSs, indicating the global role of TEs in the development of the gene regulatory networks (Feschotte, 2008; de Souza et al., 2013; Jaques et al., 2013; Mak et al., 2014; Ito et al., 2017). The ability of selfish elements for self-regulation (conserved in the evolution of the genomes of their hosts) indicates the existence of the "TEs–ncRNAs–peptides" strategy (Fig. 2), in which the transposon transcripts were selected for functional suitability simultaneously with their protein products, since ncRNAs are able to translate into peptides (Zhang et al., 2013; Anderson et al., 2015; Lauressergues et al., 2015; Couzigou et al., 2016; Lv et al., 2016; Nelson et al., 2016), while TEs are the most important sources of ncRNAs (Borchert et al., 2011; Li et al., 2011; Yuan et al., 2011; Roberts et al., 2013; Gim et al., 2014; Johnson and Guigo, 2014; Kapusta and Feschotte, 2014; Lorenzetti et al., 2016). In addition, PCGs originating from TEs perform the most important cellular functions (Sinzelle et al., 2009). Besides domestication for regulating the genome work, TEs became sources of the proteins involved in chromatin remodeling and cellular functions (Joly-Lopez and Bureau, 2018). It indicates a key role of TEs in the development of distinc-

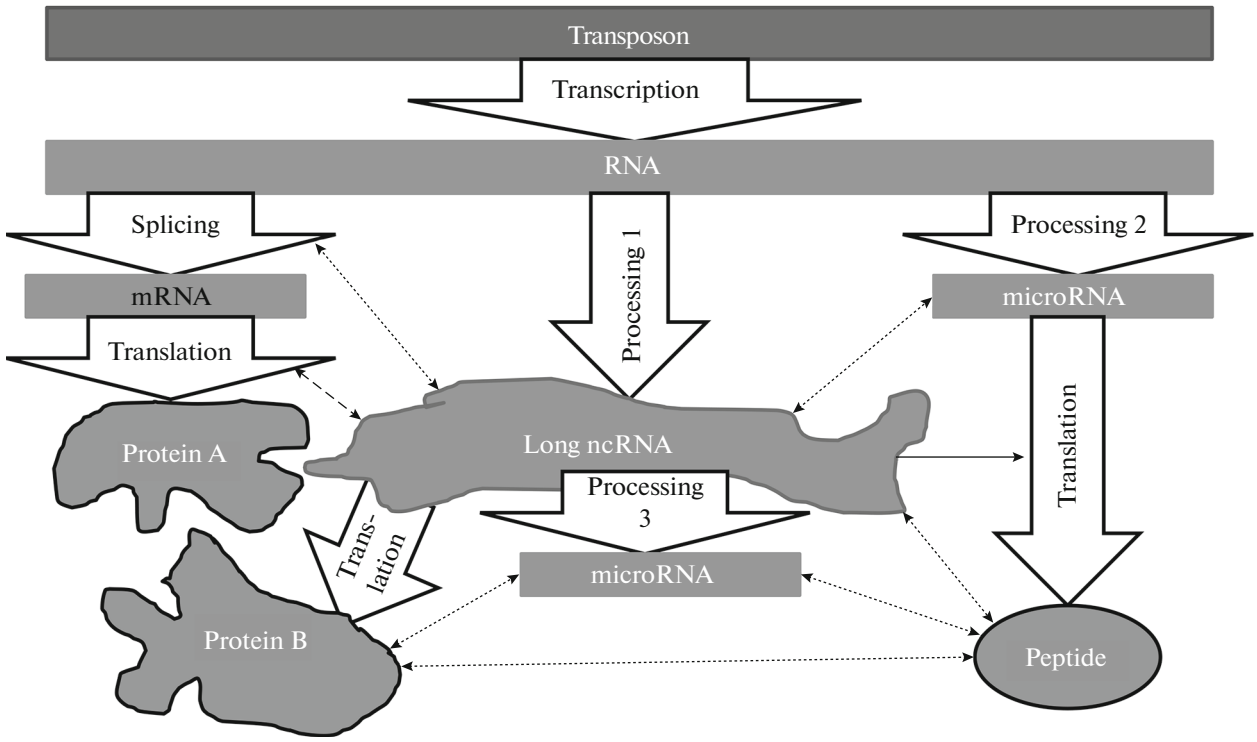

Fig. 2. Scheme of a "double search" for targets in the "transposons–noncoding RNAs–(poly)peptides" strategy in emergence of new protein-coding genes in evolution.

tive features of eukaryotic chromosomes. For example, the BEAF-32 insulator binding the chromatin with nuclear matrix and centromere-binding Abp1 proteins (Feschotte, 2008), as well as HDP1/2 histone acetyltransferase complex proteins in Arabidopsis (Duan et al. 2017), are formed from transposase. In humans, insulators originating from mammalian-wide interspersed repetitive (MIR) element family TEs, which recruit the chromatin modification enzymes to change the regulatory gene networks, were detected (Wang et al., 2015). In plants, the *far1*, *fhy3*, and *frs* genes originated by domestication of DNA-TE (Alzohairy et al., 2013). Many new PCGs originated in evolution during domestication of different TE genes. The RAG genes in vertebrates; *Daysleeper* in Drosophila; *Tram, Buster1-3*, *Zbed4*, and *P52rlPK* in mammals; *Cenp-B* in all eukaryotes; *Metnase* and *Pgbd* in humans and mice; *harbi1* in fishes, frogs, and mammals; *Daysleeper* in Arabidopsis; and *Fob1p* in yeasts were formed from transposase. The *Gin-1* and *Fob1p* genes originated from integrase. The *Mart*, *Ma*, *Fv1*, *PEG10*, *Rtl1*, and *MyEF-3* genes in mammals originated from the *Gag*-genes. The *Syncytin-1*, *-2*, *-A*, *-B* genes originated from retrovirus envelope genes (Volff, 2006; Alzohairy et al., 2013). The protein products of domesticated TE genes manage the chromatin structure (the Cenp-B, BEAF-32, HIM-17 proteins), are involved in apoptosis (THAP0, THAP1, E93), and control the cell cycle (the family of THAP, LIN-36, LIN-15B proteins) (Sinzelle et al., 2008). Large-scale structural analysis of TE proteome allowed to predict the structures of hundreds of proteins from a representative set of DNA-TE and LINEs for general structural characteristics of the proteins that originated from TEs in the evolution. The ORF1 and Gag proteins of retroelements contain a large number of structural changes and, despite their low conservatism, preserve them in the evolution. DNA-TE proteins were the most ancient and contain the folds that already existed when the first cellular organisms emerged (Abrusan et al., 2013). According to recent studies, the last universal common ancestor of the cellular life appeared approximately 3.9 billion years ago. Eubacteria and archaea appeared approximately 3.4 billion years ago, while modern eukaryotes originated much later (less than 1.84 billion years ago) (Betts et al., 2018). DNA-TE proteins have lower contact order than randomly selected control proteins, which allows them to fold rapidly and to avoid aggregation (Abrusan et al., 2013).

TE insertions in introns and intergenic regions contribute to exonization and alternative splicing, which plays an important role in protein evolution and gaining new functional domains by them. In this respect, short interspersed nuclear elements (SINEs) and LINEs containing latent splicing signals (that explains their distribution in introns) show a significant activity. For example, up to 5% alternative exons and 3'-untranslated regions (UTRs) in the human genome

originated by exonization of intergenic *Alu*-elements (Tajnik et al., 2015). Epigenetic regulation of intragenic TEs inserted in introns frequently contribute to a change in the gene regulation, phenotypic expression, and genomic evolution (Saze, 2018). More than 50% genes in humans are characterized by the presence of alternative 3'UTRs contributing to their tissue- and stage-specific regulation (Elkon et al., 2013). The content of transposon NSs in 3'UTRs of the genes has an evolutionary significance for the selection of optimal regulatory networks of the gene expression management, since TEs are sources of ncRNAs, the epigenetic effect of which extends to the genes, in 3'UTRs (and introns) of which TEs or their remains are contained. TEs also contribute to the emergence of new introns using their NSs for alternative splicing and exonization (Wang et al., 2012). It is assumed that spliceosomal introns typical for eukaryotes originated from the group II introns that have the capability for retrotransposition (Novikova and Belfort, 2017) and encode their own RT (Toro et al., 2018). At the same time, global distribution of introns mainly occurred by transpositions, TE insertions, double DNA break reparation (in which TEs are also involved) (Yenerall and Zhou, 2012), which is explained by the presence of acceptor and donor splicing sites in TE NSs (Belancio et al., 2010). In the evolution, TEs were the sources of the main spliceosome component (Prp8), which was formed from RT with lost catalytic activity (Arkhipova, 2018). The role of TEs in the emergence of universal genome regulation systems explains the preservation of the phenomenon of the gene mosaic structure (conservative for eukaryotes) in the evolution. Data on the presence of interrelations of many ncRNAs with introns and clear synergic effect are a confirmation. Moreover, a part of introns are able themselves to process in functional ncRNA molecules after the splicing (Rearick et al., 2011), indicating the important role of TEs in providing functional dualism of the gene transcripts required for the emergence of a maximal number of regulatory interrelations as a source of dynamic and plastic variability and adaptation in natural selection, and this is one of the explanations for the intron distribution and preservation in the evolution.

EMERGENCE OF REGULATORY NETWORK USING TRANSPOSONS

The preservation of new genes containing TE sequences in the evolution inevitably leads to the fact that the expression of a new protein is managed by a new regulatory network with a large number of interrelations and control by epigenetic factors. That is, a dual control system arises (Fig. 3). On the one hand, TEs in the evolution are sources of both TFs and TF binding sites (Feschotte, 2008; Ito et al., 2017). On the other hand, new proteins emerging from TEs contain regions binding to ncRNAs of a transposon origin, since TEs are important sources of siRNAs, microRNAs, and long noncoding RNAs (lncRNAs) (Borchert et al., 2011; Li et al., 2011; Yuan et al., 2011; Roberts et al., 2013; Gim et al., 2014; Johnson and Guigo, 2014; Kapusta and Feschotte, 2014; Lorenzetti et al., 2016). Due to this, complex regulatory interrelations between epigenetic factors and TFs are created, which provides coordinated successive stages of the gene expression management in the cell divisions. This is the key to the cell differentiation regulation in ontogenesis with simultaneous dynamism of this process due to the mobility and stress sensitivity inherent to TEs (Feng et al., 2013; Wheeler, 2013; Mustafin and Khusnutdinova, 2018), causing the possibility of the evolution of eukaryotes. The splicing machine is also important in these regulatory systems in eukaryotes, which explains a global distribution of spliceosomal introns that interact with many different microRNAs, snoRNAs, siRNAs, piRNAs, and lncRNAs when managing the gene expression (Fig. 4) (Rearick et al., 2011). Thus, piRNAs in Drosophila regulate the transposon pre-mRNA splicing both in somatic and germ cells (Barry, 2018). Despite the origin from the group II introns (Novikova and Belfort, 2017), only spliceosomal introns became widespread in eukaryotic genomes. This indicates their important role in the gene expression regulation, while the relationship with ncRNAs and TEs indicates the existence of a conservative facilitation system between TEs, splicing, RNA interference (RNAi), histones, and telomeres for eukaryotes. At the same time, introns can be lost during the evolution and mainly spread by transpositions and by means of TEs (Yenerall and Zhou, 2012). The involvement of TEs in the distribution of introns assumes the presence of their NSs in introns, which contributes to the gene expression management not only at the spliceosome level, but also transcriptionally (due to the effect of ncRNAs of the transposon origin). In addition, TEs in introns contribute to the formation of alternative splicing variants (Feschotte, 2008).

The global TE distribution in eukaryotic genomes contributes to the use of their NSs in gene regulation as TF binding sites, promoters, and enhancers (de Souza et al., 2013; Sahebi et al., 2018). TEs can change the expression of not only adjacent but also the genes not related to insertions, change the chromatin modification, be sources of *cis*-acting regulatory elements, and can insert in the existing enhancers that regulate the transcription (Sahebi et al., 2018). At the same time, TEs are associated with species-specific changes of the gene expression in higher vertebrates, indicating their role in the speciation (Zeng et al., 2018) and tissue-specific genome work regulation (Trizzino et al., 2018). A pronounced correlation between TE activity and speciation (Ricci et al., 2018), as well as the role of TEs in the emergence of genetic defects in hybrids preventing the species crossing (Serrato-

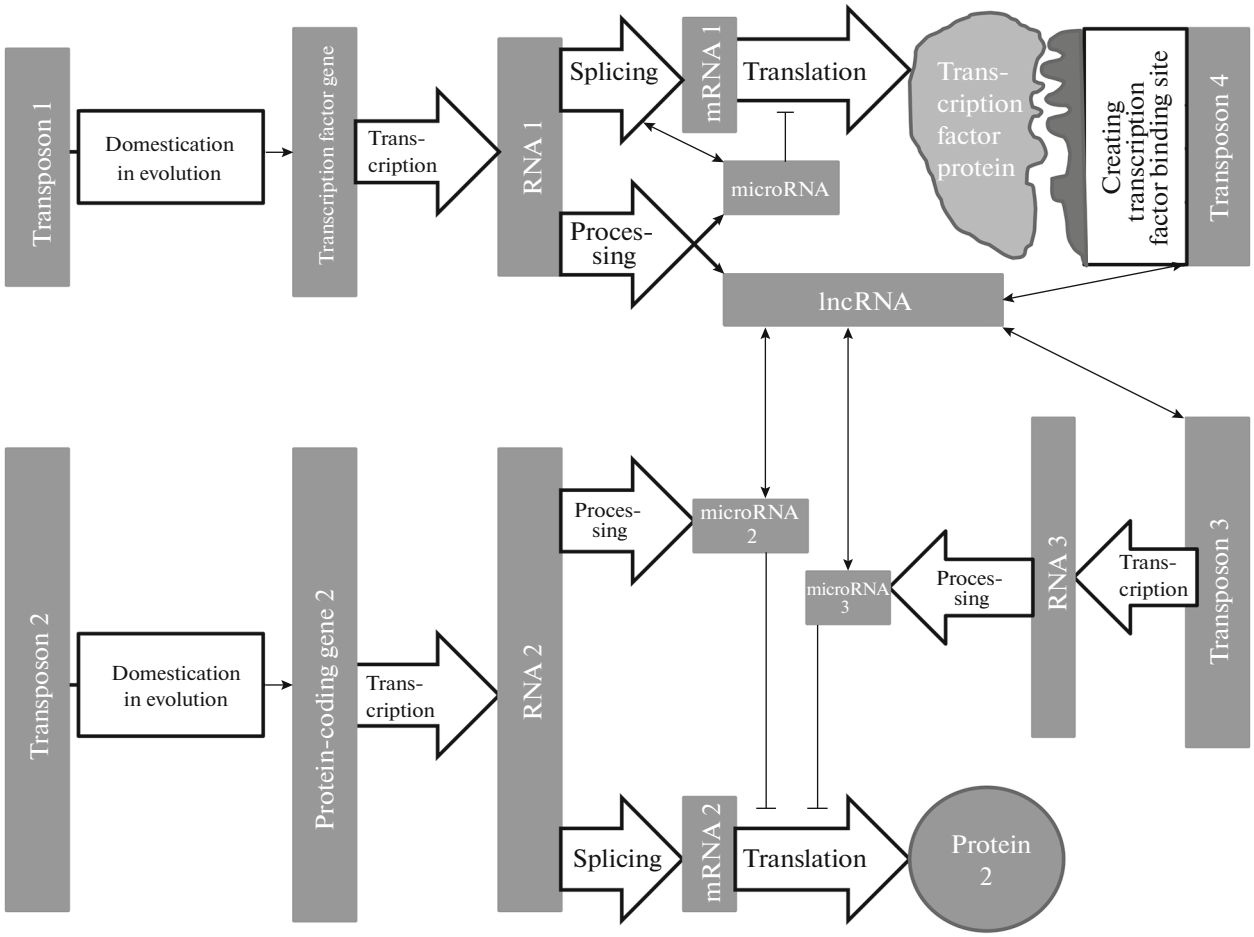

Fig. 3. Role of transposons in creating regulatory networks of genomes with double gene expression control system by means of noncoding RNAs (ncRNAs) and transcription factors.

Capuchina and Matute, 2018), were detected. Unlike changes in the genes themselves, a setup of regulatory networks of the gene expression plays a more significant role in morphological evolution. For example, when studying TF binding sites in the members of 29 different mammalian species, it was detected that the fragments of TE sequences are used by the hosts as regulatory structures on a global scale (Lowe and Haussler, 2012). During the comparative analysis of 29 mammalian species, 280000 regulatory structures originating from TEs, including those that have the properties of enhancers, were identified (de Souza et al., 2013). Besides insulators (Wang et al., 2015), the enhancers playing an important role in *cis*-regulation of PCG expression (Jjingo et al., 2014) are also formed from MIR in the human genome. For multicellular organisms, most tissue-specific activated regulators of the genes are derived from TEs, indicating the important role of transposons in eukaryotic evolution. It is interesting that up to 20% transcriptome is initiated from retro-TEs in early embryonic development, while multiple TE movements during the evolution caused a redistribution of TF binding sites performing tissue-specific regulation (Mak et al., 2014). The number of active regulatory elements of transposon origin managing the host gene expression has a colossal scale. For example, approximately 800000 TF binding sites that originated only from LTR-containing TEs were found in the human genome. Many of these regulatory elements are characterized by a tissue-specific character of activation (Ito et al., 2017). The family of retroviruses, which gave birth to LTR5HS, spread in the germ line of hominids approximately 15 million years ago creating similar NSs in the human genome. As a result, approximately 90% of these elements were identified by means of only 12 specific RNAs. Despite the fact that LTR5HS are evolutionary young TEs in the human genome, their ability to function as enhancers to control the work of the genes was registered (Judd and Feschotte, 2018). At the same time, simultaneously with the use of domesticated TE genes in ontogenesis, cooptation of the transposon NSs occurred for regulatory interrelations of these genes (Mustafin and Khusnutdinova, 2018). Domestication of the retrovirus *env*

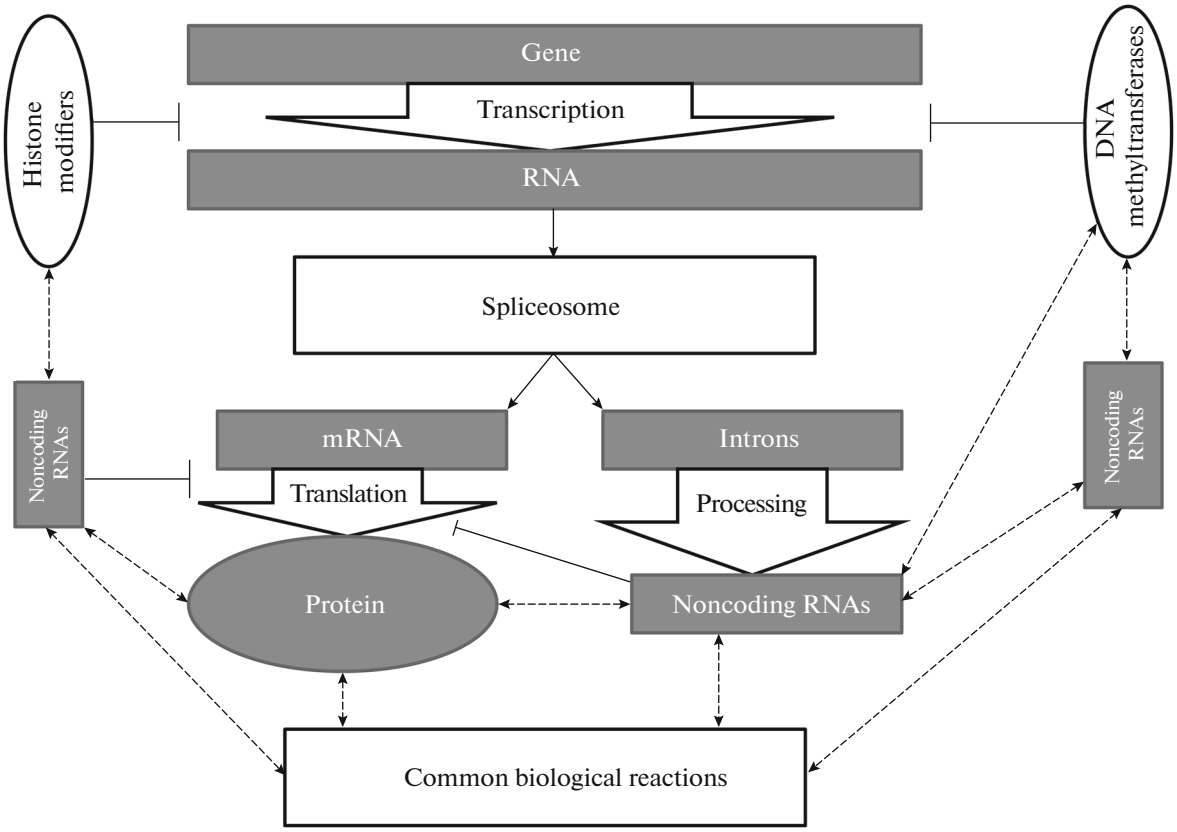

Fig. 4. Relationship of splicing with other regulatory systems of eukaryotes.

gene (Dupressoir et al., 2012) contributing to the development of the placenta can be an example, while the expansion of the RLTR13D5 and MER20 transposons was required for the expression of the genes in endometrium interaction with the product of this gene (Chuong et al., 2013). It was found that approximately half of the functional regulatory sequences in the human genome has a transposon origin (Jacques et al., 2013). Hundreds of thousands of enhancers of different animal and plant species that originated from TEs were detected, indicating the universality of the principle of cooptation of the transposon regulatory sequences by eukaryotic genomes (de Souza et al., 2013). Global scales of the cooptation of the transposon NSs both for the gene expression regulation in ontogenesis and for the emergence of new PCGs can be estimated on the example of TE interrelation with telomeres and histones, the conservatism of which for all eukaryotes indicates a key significance of TEs as universal systems of the genome modeling in the evolution.

INTERRELATION OF TRANSPOSONS WITH TELOMERES AND HISTONES

In natural selection, the development of telomeres at the chromosome ends occurred due to TEs, since telomerase originated from the retroelement RT in the evolution (Garavis et al., 2013). This property is a typical difference of eukaryotes from bacteria and archaea. Although some members of bacteria can have a linear DNA structure, they contain no telomeres at the chromosome ends, and ring nucleoids are typical for most of them (Ravin and Shestakov, 2013). In Drosophila, telomerase was lost in the evolution, and the HeT-A, TART, and TAHRE retroelements are used to form telomeres (Casacuberta, 2017). A pronounced increase in telomere size without telomerase was found in neoplasms with mutations in the *ATRX* gene, the product of which remodels nucleosomes due to the interaction with the H3 tail (Henikoff and Smith, 2015). In addition, the Atrx binds to retroelement regions called intracisternal A particle (IAP) and contributes to their silencing (contributing to the formation of heterochromatin) (Sadic et al., 2015). The role of the protein remodeling nucleosomes in the alternative telomere elongation and TE regulation indicates the presence of interrelations between TEs, telomeres, and histones allowing germinative TE insertions (preserved in the evolution) to change the structure and number of the chromosomes. Indeed, telomeric sequences are found in interstitial chromosome parts emerging due to telomere fusion during the chromosome number reduction (Badaeva and Salina, 2013). The role of the chro-

matin remodeling in these regions indicates complex interrelations with TEs that became sources of telomeres in the evolution (Garavis et al., 2013).

In bacteria and archaea, the HU protein is involved in DNA packaging, while histones are involved in some archaea and all eukaryotes. It is assumed that octameric nucleosomes of eukaryotes evolved from simpler tetrameric nucleosomes of archaea. But unlike archaeal, eukaryotic histones contain tails, modification of which contributes to epigenetic managing the gene expression (Henikoff and Smith, 2015). However, this mechanism most likely initially emerged as a protection system for silencing of TEs introduced in the genome. This indicates an important role of TEs in histone evolution. It can be assumed that the emergence of the octameric nucleosome complex and the conservative system of histone proteins in eukaryotes occurred simultaneously with the adaptation of genomes to multiple TER insertions and the creation of other universal systems, such as RNAi, splicing machine, and telomeres. Selection of the most optimal histone sets for the suitability could occur with TEs, indicating some peculiarities. For example, the histone genes are located as tandem repeats, while their mRNAs contain no polyadenylated sites ending with specific secondary "stem–loop" structures. These structures interact with the SLBP protein for subsequent metabolic stages (Marzluff and Koreski, 2017). The spatial organization of the histone mRNAs shows the functionality like lncRNAs, the domains of which are mainly constructed from TE sequences (Kapusta and Feschotte, 2014). That is, the histone gene transcription products show functional dualism (spatial RNA domains are able to interact with the proteins simultaneously with the ability to translation). This property allows one to assume the role of TEs in the histone gene evolution, since the transposon NSs play a key role in the development of secondary and tertiary structures of RNA molecules that are capable of being involved in different biological reactions (Johnson and Gujigo 2014), while the functional dualism of TE transcripts is their typical feature. This is due to the effect of protective host systems performing the transposon RNA processing and the properties developed in evolution of the obtained processing products to the functionality in the interests of the host for gaining adaptive properties by it due to the involvement of ncRNAs in different biological reactions either independently (as ribozymes or riboswitches) or as a part of RNPs or RISC.

It is possible that functional domains of RNA histones are constructed with TEs, which indicates their close relationship in the evolution, since TEs contributed to the emergence of the most important cellular proteins by exonization, duplication, and domestication, creating the basis for the emergence of lines with new adaptive properties. The emergence of telomerase from the retroelement reverse transcriptase (Garavis et al., 2013) and centromere-binding Cenp-B proteins from the *pogo* TEs of all eukaryotes (Volff, 2006) and domestication of the *env* retrovirus envelope gene by mammalian genomes using it to create the combined placental layer on the uterus surface are examples. This gene (called syncytin) plays an important role in mammalian embryonic development (Sinzelle et al., 2009). In addition to cooptation for creating new proteins, TEs were actively used in evolution as sources of regulatory sequences (Fig. 5). It is interesting that the size range of conservative and nonconservative regions of PCG introns has maxima close to the length of nucleosome and dinucleosome DNA. This indicates that the intron length is determined by functional load (Vinogradov, 2011). In addition, this confirms the role of TEs in the formation of the conservative histone complex by a "search" for possible variants in evolution with TEs, since introns are obliged to mobile elements for their origin (Yenerall and Zhou, 2012; Novikova and Belfort, 2017). This example characterizes the universality of a principle of "double search" for structural and functional relationships of nucleic acid sequences with the spatial configuration of domains of secondary and tertiary DNA, RNA, and protein structures. That is, the host interaction with TEs during the natural selection of eukaryotes with TEs occurred due to universal patterns of transposon self-regulation, which allowed the evolution of their hosts. TEs are both sources of regulatory sequences and domesticated proteins interacting with them. TEs spread in eukaryotic genomes, but they are involved at the same time in evolution of the systems fighting TEs (nucleosome evolution). The universality of self-regulation by the products of their own expression contributed to the evolution of the genome structure and function. For example, centromeres emerged due to TEs and telomeres (Garavis et al., 2013; Klein and O'Neill, 2018; McGurk and Barbash, 2018), while centromere-binding eukaryotic Cenp-B proteins originated from the *pogo* TEs in the evolution (Volff, 2006). In addition, TEs are important sources of microRNAs (Borchert et al., 2011; Li et al., 2011; Yuan et al., 2011; Roberts et al., 2013; Gim et al., 2014; Lorenzetti et al., 2016) that cause silencing of their own TEs and other genes containing homologous NSs, while the products of pri-microRNA translation stimulate the expression of their own microRNAs (Couzigou et al., 2015; Lauressergues et al., 2015; Couzigou et al., 2016; Lv et al., 2016).

TRANSPOSONS–ncRNAs–PEPTIDES STRATEGY IN EMERGENCE OF NEW GENES

In the evolution of eukaryotes, TEs were not only sources of PCGs and regulatory sequences but also sources of ncRNAs, including siRNAs, microRNAs (Borchert et al., 2011; Li et al., 2011; Yuan et al., 2011;

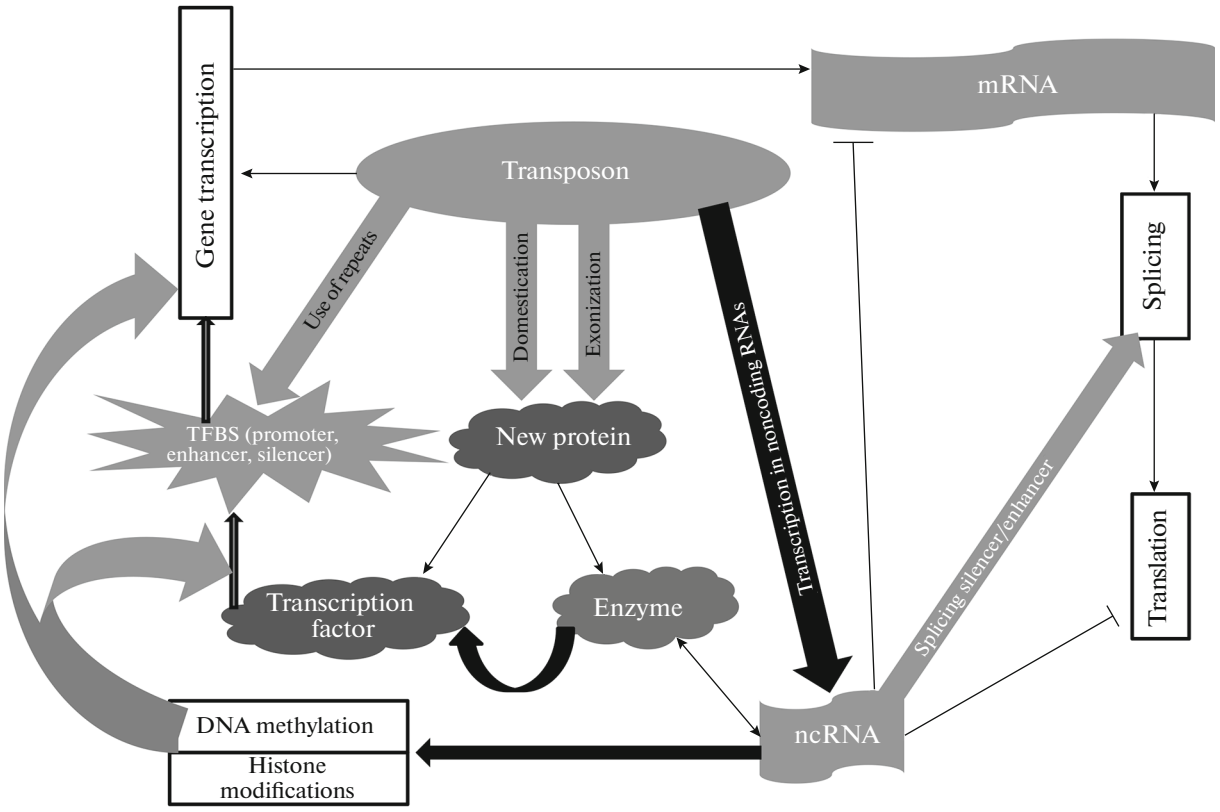

Fig. 5. Relationship of transposons, noncoding RNAs, transcription factors, and regulatory structures of eukaryotic genomes.

Roberts et al., 2013; Gim et al., 2014; Lorenzetti et al., 2016; Cho, 2018) and lncRNAs (Johnson and Guigo, 2014; Kapusta and Feschotte, 2014). At the same time, most of the eukaryotic genomes are formed from TE sequences significantly exceeding the portion identified by standard methods (de Koning et al., 2011; Goerner-Potvin and Bourque, 2018). Since most ncRNAs are transcribed from intergenic and intronic regions where TEs are located, it can be assumed that transposons are the main sources of ncRNAs. Indeed, thousands of microRNAs originating from TEs in animals were detected in different works (Borchert et al., 2011; Yuan et al., 2011; Roberts et al., 2013; Gim et al., 2014). In plants, most microRNAs are homologous to TEs, which indicates their origin from TEs (Li et al., 2011; Lorenzetti et al., 2016; Cho, 2018). TEs have an important effect on evolution due to their role in the emergence of lncRNAs. LncRNAs are transcription regulators due to the interaction with TFs and other DNA-binding proteins, RNA polymerase II, and histone-modifying complexes. LncRNAs can perform their functions as independent active molecules as riboswitches or ribozymes as well as a part of RNP (Long et al., 2017). Like proteins, secondary structures of lncRNA molecules have a modular organization and contain discrete domains that are mainly created due to TEs. Approximately 83% of lncRNA domains contain at least one TE fragment (Johnson and Guigo, 2014), while many mature lncRNAs consist completely of TEs (Kapusta and Feschotte, 2014). LINE1 retrotransposons have lncRNA-like function and are involved in the gene expression regulation to manage the self-renewal of embryonic stem cells and preimplantation development (Honson and Macfarlan, 2018). Moreover, TEs can be used directly as lncRNA genes (Lu et al., 2014), indicating a primary property of TEs for functional dualism of their transcripts. That is, the transposon transcripts are able to be processed both as functional RNA molecules with a domain structure and in ribosome-interacting RNAs, whose translation products are involved in different biological reactions. This "TEs–ncRNAs–(poly)peptides" strategy is a source for the "search" for optimal pathways of the gene network regulation in the evolution and can be a universal property of the construction of dynamic biological systems. Moreover, the fact that lncRNAs are also able to translate into peptides was an amazing discovery (Zhang et al., 2013; Anderson et al., 2015; Lv et al., 2016; Nelson et al., 2016). More surprising is that, in addition to the translation into peptides, lncRNAs are capable of further processing with the formation of miRNAs (Guo et al., 2014; Lv et al., 2018). At the same time, it was found that premicroRNAs are also able to be translated with the for-

mation of functional peptides (Couzigou et al., 2015; Lauressergues et al., 2015; Couzigou et al., 2016; Lv et al., 2016). That is, regulatory systems of eukaryotic genomes are much more complex than prevailing concepts in genetics. The study of new data allows us to come to the innovative conclusion that TEs could be an ancient universal system providing the self-reproduction of living matter at the origin of life, since conservative ncRNAs, including tRNAs (transfer RNAs), rRNAs (ribosomal RNAs), snRNAs (small nuclear RNAs), and snoRNAs (small nucleolar RNAs) are capable of nonrandom processing (like TE transcripts) using RNA products for the interference of the transposon sequences (Ender et al., 2008; Jacob et al., 2012; Li et al., 2012; Kumar et al., 2014; Venkatesh et al., 2016; Martinez et al., 2017).

The functionality itself simultaneously of both RNA molecules and their protein products can indicate the origin from TEs that is important for determining the evolutionary history of PCGs. Indeed, phylogenetic analyses allowed the detection in 2006 of the formation of new PCGs due to the selection of different variants of lncRNA expression (Levine et al., 2006). In 2008, the same pattern in the occurrence of new PCGs was found on the *S. cerevisiae* (Cai et al., 2008). In 2012, 24 different PCGs were identified that emerged from lncRNAs by their selection during the "search" for variants of molecular and biochemical interactions (Xie et al., 2012). In 2014, it was proven based on the study of six animal species (*Mus musculus*, *Danio rerio*, *D. melanogaster*, *A. thaliana*, *S. cerevisiae*, *Homo sapiens*) that the emergence of PCGs de novo is common and is a source of innovations in evolution. It was found that most lncRNAs of all six species interact with ribosomes and show similar coding potential with evolutionary young PCGs (Ruiz-Orera et al., 2014). In addition, the translation of lncRNAs into peptides (that are involved in different biological reactions like PCGs) was significantly confirmed in animals. For example, skeletal muscle-specific lncRNA expression, which is translated into the peptide mioregulin (MLN), was detected in mice. MLN shows structural and functional similarity with phospholamban and sarcolipin that control the muscle relaxation due to the regulation of the absorption of calcium ions in sarcoplasmic reticulum (SR) (Anderson et al., 2015). The functioning of another peptide also generated by lncRNA translation was found in SR. This peptide was called dwarf open reading frame (DWORF). It enhances the activity of sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA) displacing its inhibitors phospholamban, sarcolipin, and MNL in cardiomyocytes and skeletal muscles (Nelson et al., 2016). In plants, the peptides generated by lncRNA translation are involved in different biological processes: IPS1 regulates phosphate absorption, ENOD40 is involved in symbiotic relationship with bacteria, COLDAIR and COOLAIR control the flowering time (Zhang et al., 2013).

It was found that, besides lncRNAs, pre-microRNAs are also capable of being translated into functional peptides in eukaryotes. Formed in this case miPEPs have a regulatory effect on PCGs (Lv et al., 2016). Simultaneously with the ability to process into mature microRNAs, pre-microRNAs contain small open reading frames (smORF) with the translation into miPEPs that are able to enhance the transcription of related microRNAs (Couzigou et al., 2015). Functional activity of miPEPs is proven, and some of them (miPEP172c) are already successfully used in agriculture to optimize the plant agronomic properties. For example, due to enhancing the expression of miR172c, miPEP172c contributes to the formation of the soybean root system nodules, which increases the yield by stimulating a symbiosis with rhizobia (Couzigou et al., 2016). In Arabidopsis, miPEP165a was detected, while at least one of 50 pre-microRNAs contain smORF required from pre-microRNA translation into the peptide. In the *Medicago truncatula*, pre-miR171b was found (Lauressergues et al., 2015). The presence of similar properties in pre-microRNAs of animals and fungi is not excluded; this is a subject for further studies.

It is interesting that, besides PCGs originating from lncRNA genes (Levine et al., 2006; Cai et al., 2008; Xie et al., 2012; Ruiz-Orera et al., 2014), the genes of conservative ncRNAs (tRNAs, rRNAs, snRNAs, snoRNAs) have a functional dualism (Li et al., 2012; Kumar et al., 2014; Venkatesh et al., 2016; Martinez et al., 2017). This allows us to assume that the mobile elements could be primary universal structures in the evolution (at the origin of life) that are able to increase the size of biopolymers due to transposition and be sources for the emergence of a variety of other functional molecules by transcription and translation. TEs could be involved in the emergence and evolution of the genes of all known ncRNAs. It was demonstrated that constitutively expressed tRNAs, rRNAs, snoRNAs, and snRNAs mainly produce small 5' and 3' terminal fragments. Like microRNA processing, these terminal fragments are generated in an asymmetrical way with preferential support of 5' or 3' ends. These fragments were more common than microRNAs and, like microRNAs, they are processed to be involved in a variety of biological processes that are conservative for distantly related species. It is proven that the fragments that originated from snoRNAs function as microRNAs with the length from 18 nucleotides and more (Li et al., 2012).

Three fragment types called tRNA derived RNA fragments (tRF) are generated from tRNA: obtained from the 3'-end of mature tRNAs (tRF-3s), from the extreme 5'-end of mature tRNAs (tRF-5s), and from the 3'-fragment of tRNA precursor (tRF-1s). In humans, tRF-5s and tRF-3s are bound by Argonaute family

proteins functioning like microRNAs and causing their silencing of target RNAs. tRF molecules are generated not by chance but they are precisely formed regardless of DROSHA and DICER fragments present in all living organisms (from bacteria to human). Despite the fact that tRF are more evolutionarily conservative than microRNAs and are present in the same abundance in the cells, common nomenclature and unique identifiers still do not exist for them (Kumar et al., 2014). tRF molecules are nonrandom products of asymmetric tRNA processing with the length of 18–26 nucleotides. These ncRNAs are produced in large amounts in virus-infected or stressed cells. In addition, tRF are accumulated in large amounts in wild type male gametes not exposed to stress in flowering and nonflowering plants. In Arabidopsis, tRF are processed by means of Dicer-like1 and connect with AGO1 like microRNAs. The formed tRF-AGO1 complex affects specific targets and cleaves mRNAs of transcriptionally active TEs (Martinez et al., 2017). It is interesting that tRF-3s has a high complementarity to endogenous retroviral (ERV) sequences in the human genome. The role of tRF-3s in ERV expression regulation is assumed by RNA interference. It was demonstrated that tRNAs and rRNAs are exposed to stress-induced cleavage with the formation of stable RNA products (this mechanism is conservative from yeasts to human). Stress-induced tRNA decay causes the formation of the products with a length from 30 to 50 nucleotides called stress-induced tRNA-derived RNAs (sitRNAs), tRNA-derived stress-induced RNAs (tiRNAs), and tRNA halves. One more class (tRNA-derived small RNAs, tsRNAs, functioning like microRNAs) was also found. Listed tsRNAs are processed by means of Dicer or RNAse-Z depending on tRF localization both in mature tRNA and in its precursor (Li et al., 2012). The fact that microRNAs can also be generated from preexisting tRNAs (besides listed tRNA fragment classes) became an amazing discovery. For example, 20 microRNAs that originated from tRNAs that share sequences with tRF were detected (Venkatesh et al., 2016). Data on the role of nonrandom products of tRNA cleavage in ERV silencing (Li et al., 2012) can indicate common mechanisms of the emergence of TEs and tRNA genes at early stages of the evolution of living beings since these structures were selected as universal systems of biopolymer self-reproduction. The detection of functional relationships of TEs with tRNAs can be a confirmation. Thus, TE Ty3 in yeasts recognize tRNA genes by the interaction of their integrase with RNA polymerase III TFIIIB subunit involved in tRNA transcription. In the amoebozoan *Dictyostelium discoideum*, ribonuclease encoded by retroTE DGLT-A interacts with RNA polymerase III TFIIIC subunit (Kling et al., 2018).

Besides tRNAs, asymmetrical processing (specific decay) from the 3' or 5' ends of snRNAs, snoRNAs, and rRNAs leads to the formation of RNA fragments with the length of approximately 20 nucleotides (Li et al., 2012). It was also detected that long intergenic spacers (IGS) located between rRNA genes are not inert but show a polyfunctionality required for correct cell functioning. Due to the induction of different ncRNA transcripts from IGS, the cell is able both to regulate rRNA synthesis and to bind large amount of proteins, thus modulating important molecular networks. Unlike coding rRNAs, IGS are characterized by a high level of variability both by NSs and by the length, which are 2.5 kb in yeasts, 5.1 kb in Drosophila, and approximately 30 kb in the *Xenopus laevis,* chickens, mice, and primates. At the same time, changes in NSs are caused by insertions of a large number of retroTEs (Jacob et al., 2012). Small RNAs originated from snoRNAs (playing an important role in rRNA maturation) were also found. For example, RNA with the length from 20 to 25 nucleotides, which stably binds to Ago proteins, is processed from ACA45 snoRNA. The processing does not depend on Drosha/DGCR8 complex but requires Dicer. These small ncRNAs that originated from snoRNAs function like microRNAs with the target effect on specific mRNAs (Ender et al., 2008). These microRNA-like RNAs called sno-derived RNAs (sdRNAs) with evolutionary conservative sizes and localization originate from a wide range of snoRNA loci in animals (human, mouse, chicken, Drosophila), Arabidopsis, and fission yeasts. snoRNAs direct RNA modification localizing in nucleoli and Cajal bodies in eukaryotes. RNA pathway components are associated with these structures. In animals, sdRNA with the length of 20–24 nucleotides are generated from the 3' end of H/ACA snoRNA. sdRNAs with bimodal size approximately 17–19 and more than 27 nucleotides mainly from the 5' end originate from C/D snoRNA. sdRNAs bind to AGO7 in Arabidopsis and to Ago1 in fission yeasts with typical 5' nucleotide displacements. That is, the relationship between RNA silencing and snoRNA-mediated RNA processing systems exists in eukaryotes, while sdRNA is an ancient class of small ncRNAs (Taft et al., 2009). Thus, all known ncRNAs are characterized by the ability for processing to new ncRNAs, the association of which with specific proteins allows them to function for the silencing of TE and PCG transcripts providing a dynamic homeostasis of eukaryotic cells (Fig. 6). Functional dualism is typical for TE transcripts caused by the effect of the processing systems that emerged to protect the host from TE distribution. Therefore, functional dualism and the possibility of processing of all known most ancient ncRNAs with their use for TE RNA interference indicates the role of mobile elements in the emergence of these ncRNAs in evolution at the origin of living beings. The universal TE ability for self-reproduction and self-regulation by the processed products of their own transcription

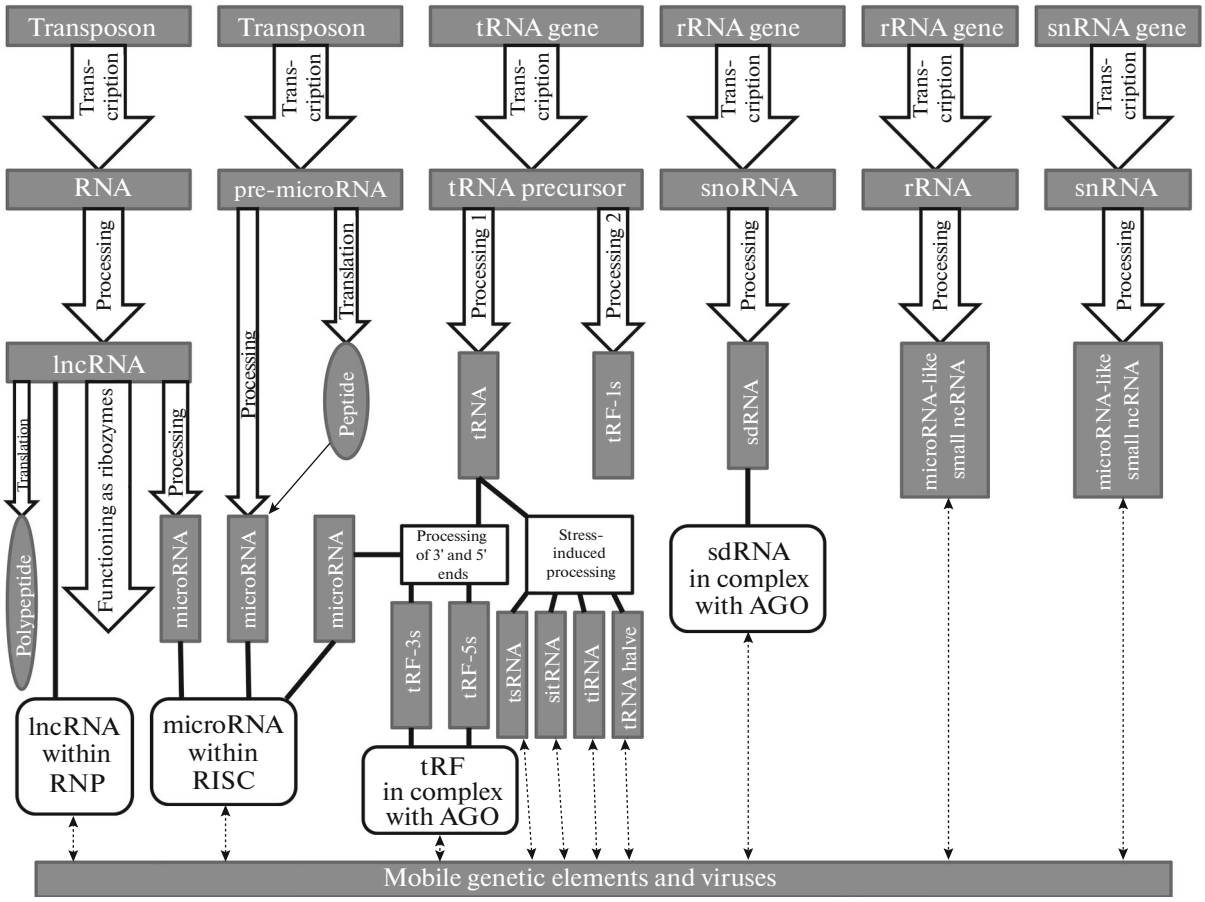

Fig. 6. Functional dualism of noncoding RNAs. LncRNA, long noncoding RNA.

allows us to assume that TEs are a fundamental property of living beings required for the self-reproduction of biopolymers at the origin of life on Earth.

CONCLUSIONS

Physical separation by insertions of the transposons in DNA structure of the genes involved in common biological reactions (constituting an operon in prokaryotes) in eukaryotes was compensated by the development of complex regulatory networks. These networks consist of TEs, originating from them regulatory sequences, proteins, and ncRNAs that provide more dynamic gene work management in common biochemical reactions. Complex molecular relationships formed in this case between different biological processes became more labile providing the possibility of better adaptation and complication of the organisms with the appearance of new cellular functions and intercellular relationships. The preservation of germinative TE insertions in evolution due to their adaptive significance (the formation of new regulatory sequences and new PCG domains) was accompanied by a change in the chromosome structure and number. This is caused by the origin of telomeres from TEs, their dynamic relationships with TEs in the evolution and ontogenesis, as well as the presence of a complex system of interactions between TEs, histones, and telomeres. The relationship of TEs with histones is caused by their specific modification along with a change in DNA methylation under the influence of siRNAs and microRNAs originating from TEs. Thus, the functioning of complex systems between TEs, telomeres, histones, DNA methylation, splicing, and RNAi systems (that provide a dynamic regulation of the use of inserted TEs for the host's needs in evolution and ontogenesis) is typical for eukaryotes. The relationship of these systems provided the possibility of the emergence of new PCGs from ncRNAs originating from TEs. At the same time, the proteins newly generated in the evolution are characterized by mutual regulation with RNA molecules from which they were translated, since these RNAs show the functionality either independently or as a part of RNP (for lncRNAs) or RISC (for microRNAs). The "TEs–ncRNAs–peptides" strategy in the evolution of genomes indicates a potential functionality of most mRNAs translated into the proteins. On the one hand, this could become a key to

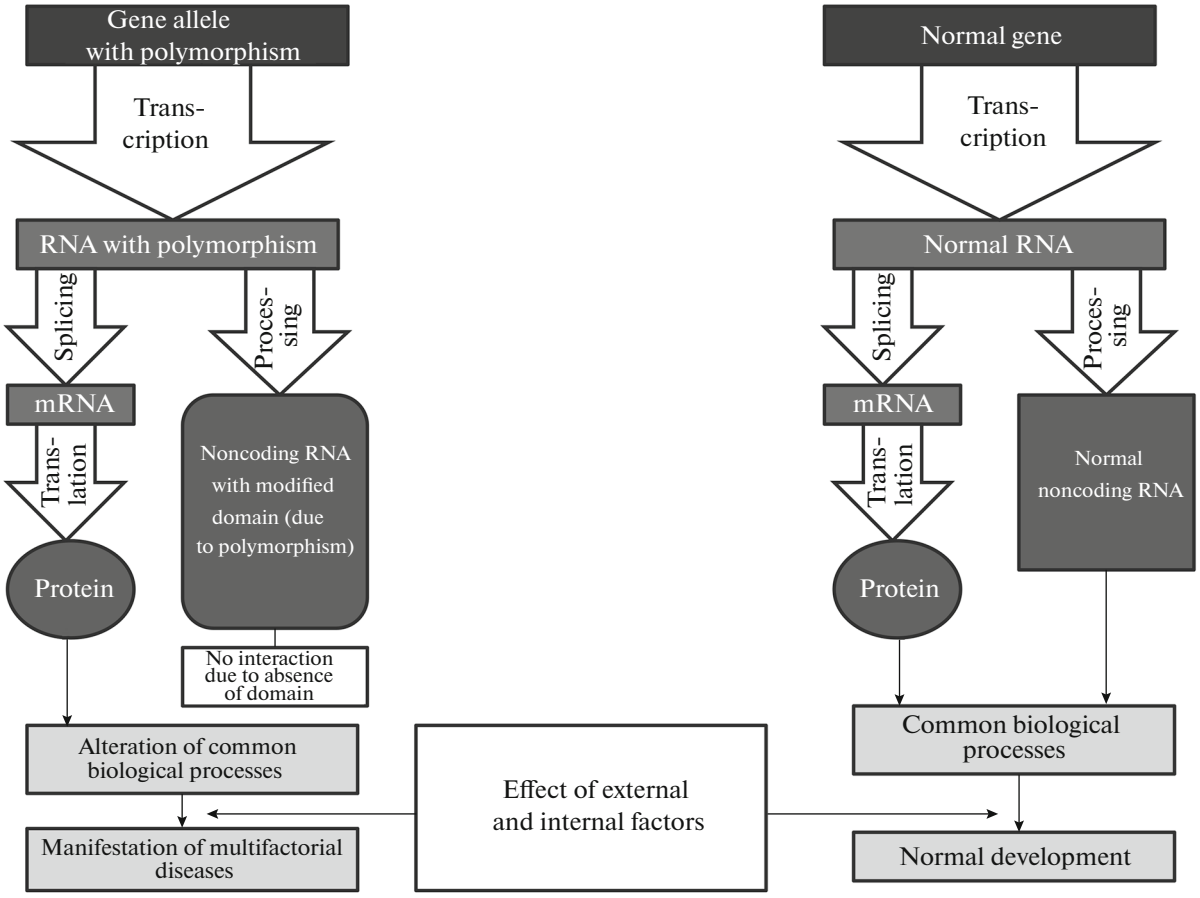

Fig. 7. Scheme of single nucleotide polymorphism (SNP) effect on development of multifactorial disease associated with it.

explaining the role of SNPs causing no change in the amino acid sequences of the proteins but playing a role in the genesis of multifactorial diseases (Fig. 7). On the other hand, the detection of cellular protein mRNA functionality can indicate their emergence in the evolution with the involvement of TEs with the possible use of "TEs–ncRNAs–peptides" strategy. It can be assumed that this strategy originated primarily at the origin of life on the Earth as a universal property, due to which there was a "search" for most optimal interactions between spatial structures of RNAs and proteins with DNA and RNA nucleotide sequences (mainly repeats).

REFERENCES

Abrusan, G., Zhang, Y., and Szilagyi, A., Structure prediction and analysis of DNA transposon and LINE retrotransposons proteins, *J. Biol. Chem*., 2013, vol. 288, pp. 16127–16138.

Anderson, D.M., Anderson, K.M., Cang, C.L., et al., A micropeptide encoded by a putative long noncoding RNA regulates muscle performance, *Cell*, 2015, vol. 160, pp. 595–606.

Alzohairy, A.M., Gyulai, G., Jansen, R.K., and Bahieldin, A., Transposable elements domesticated and neofunctionalized by eukaryotic genomes, *Plasmid*, 2013, vol. 69, no. 1, pp. 1–15.

Arkhipova, I.R., Neutral theory, transposable elements, and eukaryotic genome evolution, *Mol. Biol. Evol.*, 2018, vol. 35, no. 6, pp. 1332–1337.

Babakhani, S. and Oloomi, M., Transposons: the agents of antibiotic resistance in bacteria, *J. Basic Microbiol.*, 2018. doi 10.1002/jobm.201800204

Badaeva, E.D. and Salina, E.A., Genome structure and chromosome analysis of the plants, *Vavilovskii Zh. Genet. Sel.*, 2013, vol. 17, pp. 1017–1042.

Barry, G., Small RNAs and transposable elements are key components in the control of adaptive evolution in eukaryotes, *BioEssays*, 2018, vol. 40, no. 8, p. e1800070.

Belancio, V.P., Roy-Engel, A.M., and Deininger, P.L., All y'all need to know 'bout retroelements in cancer, *Semin. Cancer Biol.*, 2010, vol. 20, pp. 200–210.

Betts, H.C., Puttick, M.N., Clark, J.W., et al., Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin, *Nat. Ecol. Evol.*, 2018. doi 10.1038/s41559-018-0644-x

Borchert, G.M., Holton, N.W., Williams, J.D., et al., Comprehensive analysis of microRNA genomic loci

identifies pervasive repetitive-element origins, *Mobile Genet. Elem.*, 2011, vol. 1, no. 1, pp. 8–17.

Cai, J., Zhao, R., Jiang, H., and Wang, W., De novo origination of a new protein-coding gene in *Saccharomyces cerevisiae, Genetics*, 2008, vol. 179, pp. 487–496.

Casacuberta, E., Drosophila: retrotransposons making up telomeres, *Viruses*, 2017, vol. 9, no. 7, p. 192.

Cerbin, S. and Jiang, N., Duplication of host genes by transposable elements, *Curr. Opin. Genet. Dev.*, 2018, vol. 49, pp. 63–69.

Cho, J., Transposon-derived non-coding RNAs and their function in plants, *Front. Plant. Sci.*, 2018, vol. 9, p. 600.

Chuong, E.B., Rumi, M.A., Soares, M.J., and Baker, J.C., Endogenous retroviruses function as species-specific enhancer elements in the placenta, *Nat. Genet.*, 2013, vol. 45, no. 3, pp. 325–329.

Couzigou, J.M., Lauressergues, D., Becard, G., and Comier, J.P., miRNA-encoded peptides (miPEPs): a new tool to analyze the role of miRNAs in plant biology, *RNA Biol.*, 2015, vol. 12, pp. 1178–1180.

Couzigou, J.M., Andre, O., Cuillotin, B., et al., Use of microRNA-encoded peptide miPEP172c to stimulate nodulation in soybean, *New Phytol.*, 2016, vol. 211, pp. 379–381.

de Koning, A.P., Gu, W., Castoe, T.A., et al., Repetitive elements may comprise over two-thirds of the human genome, *Plos Genet.*, 2011, vol. 7, no. 12, p. e1002384.

de Souza, F.S., Franchini, L.F., and Rubinstein, M., Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong, *Mol. Biol. Evol.*, 2013, vol. 30, no. 6, pp. 1239–1251.

Duan, C.G., Wang, X., Xie, S., et al., A pair of transposon-derived proteins function in a histone acetyltransferase complex for active DNA demethylation, *Cell Res.*, 2017, vol. 27, no. 2, pp. 226–240.

Dupressoir, A., Lavialle, C., and Heidmann, T., From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation, *Placenta*, 2012, vol. 33, pp. 663–671.

Elkon, R., Ugalde, A.P., and Agami, R., Alternative cleavage and polyadenylation: extent, regulation and function, *Nat. Rev. Genet.*, 2013, vol. 14, no. 7, pp. 496–506.

Ender, C., Krek, A., Friedlander, M.R., et al., Human snoRNA with microRNA-like functions, *Mol. Cell*, 2008, vol. 32, no. 4, pp. 519–528.

Feng, G., Leem, Y.E., and Levin, H.L., Transposon integration expression of stress response genes, *Nucleic Acids Res.*, 2013, vol. 41, no. 2, pp. 775–789.

Feschotte, C., Transposable elements and the evolution of regulatory networks, *Nat. Rev. Genet.*, 2008, vol. 9, no. 5, pp. 397–405.

Filee, J., Giant viruses and their mobile genetic elements: the molecular symbiosis hypothesis, *Curr. Opin. Virol.*, 2018, vol. 33, pp. 81–88.

Garavis, M., Gonzalez, C., and Villasante, A., On the origin of the eukaryotic chromosome: the role of noncanonical DNA structures in telomere evolution, *Genome Biol. Evol.*, 2013, vol. 5, pp. 1142–1150.

Gim, J., Ha, H., Ahn, K., et al., Genome-wide identification and classification of microRNAs derived from repetitive elements, *Genomic Inf.*, 2014, vol. 12, no. 4, pp. 261–267.

Grandi, F.C., Rosser, J.M., Newkirk, S.J., et al., Retrotransposition creates sloping shores: a graded influence of hypomethylated CpG islands on flanking CpG sites, *Genome Res.*, 2015, vol. 25, no. 8, pp. 1135–1146.

Goerner-Potvin, P. and Bourque, G., Computational tools to unmask transposable elements, *Nat. Rev. Genet.*, 2018, vol. 19, no. 11, pp. 688–704.

Guo, L., Zhao, Y., Yang, S., et al., An integrated evolutionary analysis of miRNA-lncRNA in mammals, *Mol. Biol. Rep.*, 2014, vol. 41, pp. 201–207.

Henikoff, S. and Smith, M.M., Histone variants and epigenetics, *Cold Spring Harbor Perspect. Biol.*, 2015, vol. 7, no. 1, p. a019364. doi 10.1101/cshperspect.a019364

Honson, D.D. and Macfarlan, T.S., A lncRNA-like role for LINE1s in development, *Dev. Cell*, 2018, vol. 46, no. 2, pp. 132–134.

Huang, C.J., Lin, W.Y., Chang, C.M., and Choo, K.B., Transcription of the rat testis-specific *Rtdpoz-T1* and *-T2* retrogenes during embryo development: co-transcription and frequent exonisation of transposable element sequences, *BMC Mol. Biol.*, 2009, vol. 10, pp. 74–88.

Ito, J., Suqimoto, R., Nakaoka, H., et al., Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses, *PLoS Genet.*, 2017, vol. 13, no. 7, p. e1006883.

Jacob, M.D., Audas, T.E., Mullineux, S.T., and Lee, S., Where no RNA polymerase has gone before: novel functional transcripts derived from the ribosomal Intergenic spacer, *Nucleus*, 2012, vol. 3, no. 4, pp. 315–319.

Jacques, P.E., Jeyakani, J., and Bourgue, G., The majority of primate-specific regulatory sequences are derived from transposable elements, *PLoS Genet.*, 2013, vol. 9, no. 5, p. e1003504.

Jjingo, D., Conley, A.B., Wang, J., et al., Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression, *Mobile DNA*, 2014, vol. 5, pp. 5–14.

Johnson, R. and Guigo, R., The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs, *RNA*, 2014, vol. 20, no. 7, pp. 959–976.

Joly-Lopez, Z. and Bureau, T.E., Exaptation of transposable element coding sequences, *Curr. Opin. Genet. Dev.*, 2018, vol. 49, pp. 34–42.

Judd, J. and Feschotte, C., Gene expression: transposons take remote control, *eLife*, 2018, vol. 7, p. e40921.

Kapusta, A. and Feschotte, C., Volatile evolution of long noncoding RNA repertoires: mechanisms and biological implications, *Trends Genet.*, 2014, vol. 30, no. 10, pp. 439–452.

Klein, S.J. and O'Neill, R.J., Transposable elements: genome innovation, chromosome diversity, and centromere conflict, *Chromosome Res.*, 2018, vol. 26, nos. 1–2, pp. 5–23.

Kling E., Spaller T., Schiefner J. et al., Convergent evolution of integration site selection upstream of tRNA genes by yeast and amoeba retrotransposons, *Nucleic Acids Res.*, 2018, vol. 46, pp. 7250–7260.

Kopera, H.C., Moldovan, J.B., Morrish, T.A., et al., Similarities between long interspersed element-1 (LINE-1)

reverse transcriptase and telomerase, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, vol. 108, pp. 20345–20350.

Kubiak, M.R. and Makalowska, I., Protein-coding genes' retrocopies and their functions, *Viruses*, 2017, vol. 9, p. 80.

Kumar, P., Anaya, J., Mudunuri, S.B., and Dutta, A., Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets, *BMC Biol.*, 2014, vol. 12, p. 78.

Lauressergues, D., Couzigou, J.M., Clemente, H.S., et al., Primary transcripts of microRNAs encode regulatory peptides, *Nature*, 2015, vol. 520, no. 7545, pp. 90–93.

Leplae, R., Hebrant, A., Wodak, S.J., and Toussaint, A., ACLAME: a classification of mobile genetic elements, *Nucleic Acids Res.*, 2004, vol. 32, pp. 45–49.

Levine, M.T., Jones, C.D., Kern, A.D., et al., Novel genes derived from noncoding DNA in Drosophila melanogaster are frequently X-linked and exhibit testis-biased expression, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, vol. 103, pp. 9935–9939.

Li, Y., Li, C., Xia, J., and Jin, Y., Domestication of transposable elements into microRNA genes in plants, *PLoS One*, 2011, vol. 6, p. e19212.

Li, Z., Ender, C., Meister, G., et al., Extensive terminal and asymmertric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs, *Nucleic Acids Res.*, 2012, vol. 40, no. 14, pp. 6787–6799.

Long, Y., Wang, X., Youmans, D.T., and Cech, T.R., How do lncRNAs regulate transcription? *Sci. Adv.*, 2017, vol. 3, no. 9, p. eaao2110.

Lorenzetti, A.P., de Antonio, G.Y., Paschoal, A.R., and Domingues, D.S., Plant TE-MIR DB: a database for transposable element-related microRNAs in plant genomes, *Funct. Integr. Genomics*, 2016, vol. 16, pp. 235–242.

Lowe, C.B. and Haussler, D., 29 mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome, *PLoS One*, 2012, vol. 7, no. 8, p. e43128.

Lu, X., Sachs, F., Ramsay, L., et al., The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity, *Nat. Struct. Mol. Biol.*, 2014, vol. 21, no. 4, pp. 423–425.

Lv, S., Pan, L., and Wang, G., Commentary: primary transcripts of microRNAs encode regulatory peptides, *Front. Plant Sci.*, 2016, vol. 7, p. 1436.

Lv, J., Wang, L., Zhang, J., et al., Long noncoding RNA H19-derived miR-675 aggravates restenosis by targeting PTEN, *Biochem. Biophys. Res. Commun.*, 2018, vol. 497, pp. 1154–1161.

Martinez, G., Choudury, S.G., and Slotkin, R.K., tRNA-derived small RNAs target transposable element transcripts, *Nucleic Acids Res.*, 2017, vol. 45, no. 9, pp. 5142–5152.

Marzluff, W.F. and Koreski, K.P., Birth and death of histone mRNAs, *Trends Genet.*, 2017, vol. 33, no. 10, pp. 745–759.

Mak, K.S., Burdach, J., Norton, L.J., et al., Repression of chimeric transcripts emanating from endogenous retrotransposons by a sequence-specific transcription factor, *Genome Biol.*, 2014, vol. 15, no. 4, p. R58.

McGurk, M.P. and Barbash, D.A., Double insertion of transposable elements provides a substrate for the evolution of satellite DNA, *Genome Res.*, 2018, vol. 28, no. 5, pp. 714–725.

Morrish, T.A., Garcia-Perez, J.L., Stamato, T.D., et al., Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres, *Natrure*, 2007, vol. 446, no. 7132, pp. 208–212.

Mustafin, R.N. and Khusnutdinova, E.K., The role of transposons in epigenetic regulation of ontogenesis, *Russ. J. Dev. Biol.*, 2018, vol. 49, no. 2, pp. 61–78.

Nelson, B.R., Makarewich, C.A., Anderson, D.M., et al., A peptide encoded by a transcript annotated as long noncoding RNA enhances SERCA activity in muscle, *Science*, 2016, vol. 351, pp. 271–275.

Novikova, O. and Belfort, M., Mobile group II introns as ancestral eukaryotic elements, *Trends Genet.*, 2017, vol. 33, no. 11, pp. 773–783. doi 10.1016/j.tig2017.07.009

Patrushev, L.I. and Minkevich, I.G., The problem of genome size of eukaryotes, *Usp. Biol. Khim.*, 2007, vol. 47, pp. 293–370.

Ravin, N.V. and Shestakov, S.V., Genome of prokaryotes, *Vavilovskii Zh. Genet. Sel.*, 2013, vol. 17, no. 4-2, pp. 972–984.

Rearick, D., Prakash, A., McSweeny, A., et al., Critical association of ncRNA with introns, *Nucleic Acids Res.*, 2011, vol. 39, no. 6, pp. 2357–2366.

Ricci, M., Peona, V., Guichard, E., et al., Transposable elements activity is positively related to rate of speciation in mammals, *J. Mol. Evol.*, 2018, vol. 86, pp. 303–310.

Roberts, J.T., Cooper, E.A., and Favreau, C.J., Formation from transposable element insertions and noncoding RNA mutations, *Mobile Genet. Elem.*, 2013, vol. 1, p. e27755.

Ruiz-Orera, J., Messeguer, X., Subirana, J.A., and Alba, M.M., Long non-coding RNAs as a source of new peptides, *eLife*, 2014, vol. 3, p. e03523. doi 10.7554/eLife.03523

Sadic, D., Schmid,t K., Groh, S., et al., Atrx promotes heterochromatin formation at retrotransposons, *EMBO Rep.*, 2015, vol. 16, no. 7, pp. 836–850.

Sakai, H., Mizuno, H., Kawahara, Y., et al., Retrogenes in rice (*Oryza sativa* L. ssp. *japonica*) exhibit correlated expression with their source genes, *Genome Biol. Evol.*, 2011, vol. 3, pp. 1357–1368.

Saze, H., Epigenetic regulation of intragenic transposable elements: a two-edged sword, *J. Biochem.*, 2018, vol. 164, pp. 323–328.

Schrader, L. and Schmitz, J., The impact of transposable elements in adaptive evolution, *Mol. Ecol.*, 2018. doi 10.111/mec.14794

Serrato-Capuchina, A. and Matute, D.R., The role of transposable elements in speciation, *Genes* (Basel), 2018, vol. 9, no. 5, p. E254.

Shestakov, S.V., The role of archaea in the origin of eukaryotes, *Ekol. Genet.*, 2017, vol. 15, no. 4, pp. 52–59.

Sinzelle, L., Izsvak, Z., and Ivics, Z., Molecular domestivation of transposable elements: from detrimental parasites to useful host genes, *Cell Mol. Life Sci.*, 2009, vol. 66, pp. 1073–1093.

Song, M.J. and Schaack, S., Evolutionary conflict between mobile DNA and host genomes, *Am. Nat.*, 2018, vol. 192, no. 2, pp. 263–273.

Startek, M.P., Nogly, J., Gromadka, A., et al., Inferring transposons activity chronology by TRANScendence—TEs database and de-novo mining tool, *BMC Bioinf.*, 2017, vol. 18, p. 422.

Taft, R.J., Glazov, E.A., Lassmann, T., et al., Small RNAs derived from snoRNAs, *RNA*, 2009, vol. 15, pp. 1233—1240.

Tan, S., Cardoso-Moreira, M., Shi, W., et al., LTR-mediated retroposition as a mechanism of RNA-based duplication in metazoans, *Genome Res.*, 2016, vol. 26, no. 12, pp. 1663—1675.

Tajnik, M., Vigilante, A., Braun, S., et al., Inergenic *Alu* exonisation facilitates the evolution of tissue-specific transcript ends, *Nucleic Acids Res.*, 2015, vol. 43, no. 21, pp. 10492—10505.

Toro, N. and Nisa-Martinez, R., Comprehensive phylogenetic analysis of bacterial reverse transcriptases, *PLoS One*, 2014, vol. 9, no. 11, p. e114083.

Toro, N., Martínez-Abarca, F., González-Delgado, A., and Mestre, M.R., On the origin and evolutionary relationships of the reverse transcriptases associated with type III CRISPR-Cas systems, *Front. Mircrobiol.*, 2018, vol. 9, p. 1317.

Trizzino, M., Kapusta, A., and Brown, C.D., Transposable elements generate regulatory novelty in a tissue-specific fashion, *BMC Genomics*, 2018, vol. 19, no. 1, p. 468.

Venkatesh, T., Suresh, P.S., and Tsutsumi, R., tRFs: miRNAs in disguise, *Gene*, 2016, vol. 579, pp. 133—138.

Vinogradov, A.E., Base functions of genome structure of eukaryotes, *Extended Abstract of Doctoral (Biol.) Dissertation*, St. Petersburg, 2011.

Volff, J.N., Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes, *BioEssays*, 2006, vol. 28, pp. 913—922.

Wang, D., Su, Y., Wang, X., et al., Transposon-derived and satellite-derived repetitive sequences play distinct functional roles in mammalian intron size expansion, *Evol. Bioinform. Online*, 2012, vol. 8, pp. 301—319.

Wang, J., Vicente-Garcia, C., Seruqqia, D., et al., MIR retrotransposons sequences provide insulators to the human genome, *Proc. Natl. Acad. Sci. U.S.A.*, 2015, vol. 112, pp. 4428—4437.

Wheeler, B.S., Small RNAs, big impact: small RNA pathways in transposon control and their effect on the host stress response, *Chromosome Res.*, 2013, vol. 21, pp. 587—600.

Wicker, T., Sabot, F., Hua-Van, A., et al., A unified classification system for eukaryotic transposable elements, *Nat. Rev. Genet.*, 2007, vol. 8, no. 12, pp. 973—982.

Xie, C., Zhang, Y.E., Chen, J.Y., et al., Hominoid-specific de novo protein-coding genes originating from long non-coding RNAs, *PLoS Genet.*, 2012, vol. 8, p. e1002942.

Yenerall, P. and Zhou, L., Identifying the mechanisms of intron gain: progress and trends, *Biol. Direct*, 2012, vol. 7, p. 29. doi 10.1186/1745-6150-7-29

Yuan, Z., Sun, X., Liu, H., and Xie, J., MicroRNA genes derived from repetitive elements and expanded by segmental duplication events in mammalian genomes, *PLoS One*, 2011, vol. 6, p. e17666.

Zdobnov, E.M., Campillos, M., Harrington, E.D., et al., Protein coding potential of retroviruses and other transposable elements in vertebrate genomes, *Nucleic Acids Res.*, 2005, vol. 33, pp. 946—954.

Zhang, J., Mujahid, H., Hou, Y., et al., Plant long ncRNAs: a new frontier for gene regulatory control, *Am. J. Plant Sci.*, 2013, vol. 4, pp. 1038—1045.

Zeng, L., Pederson, S.M., Kortschak, R.D., and Adelson, D.L., Transposable elements and gene expression during the evolution of amniotes, *Mobile DNA*, 2018, vol. 9, p. 17.

Zhang, H.H., Zhou, Q.Z., Wang, P.L., et al., Unexpected invasion of miniature inverted-repeat transposable elements in viral genomes, *Mobile DNA*, 2018, vol. 9, p. 19.

Zhu, Z., Tan, S., Zhang, Y., and Zhang, Y.E., LINE-1-like retrotransposons contribute to RNA-based gene duplication in dicots, *Sci. Rep.*, 2016, vol. 6, art. ID 24755.

Translated by A. Barkhash