

ENDOGENOUS RETROVIRUSES AS GENETIC MODULES THAT SHAPE THE GENOME REGULATORY NETWORKS DURING EVOLUTION

Mikola Popov^{1,2}, Tetyana Kolotova¹, Maria Davidenko¹

¹ SI «Mechnikov Institute of Microbiology and Immunology of National Academy of Medical Sciences of Ukraine», Pushkinska St. 14, Kharkiv, 61057, Ukraine, e-mail: imiamn@ukr.net

² V. N. Karazin Kharkiv National University, 6 Svobody Sq., Kharkiv, 61022, Ukraine, e-mail: med@karazin.ua

Endogenous retroviruses (ERV) are the descendants of exogenous retroviruses that integrated into the germ cells genome, fixed and became inheritable. ERVs have evolved transcriptional enhancers and promoters that allow their replication in a wide range of tissue. Because ERVs comprise the regulatory elements it could be assume that ERVs capable to shape and reshape genomic regulatory networks by inserting their promoters and enhancers in new genomic loci upon retrotransposition. Thus retrotransposition events can build new regulatory regions and lead to a new pattern of gene activation in the cell.

In this review we summarize evidence which revealed that ERVs provide a plethora of novel gene regulatory elements, including tissue specific promoters and enhancers for protein-coding genes or long noncoding RNAs in a wide range of cell types. The accumulated findings support the hypothesis that the ERVs have rewired the gene regulatory networks and act as a major source of genomic regulatory innovation during evolution.

KEY WORDS: endogenous retroviruses, enhancer, promoter, lncRNA, regulatory networks, R-operon, mobile elements, genome evolution

ЕНДОГЕННІ РЕТРОВІРУСИ ЯК ГЕНЕТИЧНІ МОДУЛІ, ЩО ФОРМУЮТЬ РЕГУЛЯТОРНІ МЕРЕЖІ ВПРОДОВЖ ЕВОЛЮЦІЇ

Попов М. М.^{1,2}, Колотова Т. Ю.¹, Давиденко М. Б.¹

¹ ДУ «Інститут мікробіології та імунології ім. І. І. Мечникова НАН України», вул. Пушкінська, 14, м. Харків, 61057, Україна

² Харківський національний університет імені В. Н. Каразіна, пл. Свободи, 6, м. Харків, 61022, Україна

Ендогенні ретровіруси (ERV) є нащадками екзогенних ретровірусів, які впровадилися в геном статевих клітин і стали передаватися у спадок. Впродовж еволюції ERVs ретровіруси придбали енхансери і промотори, які дозволили їм експресуватися у різних тканинах. Оскільки ERVs мають регуляторні елементи, то можна припустити, що вони здатні формувати і переформувати регуляторні мережі генома за допомогою впровадження промоторів і енхансерів у нові локуси генома при ретротранспозиції.

У огляді зібрані дані, які підтверджують роль ERVs ретровірусів у формуванні ряду енхансерів і промоторів генів, що кодують білки, і генів, які кодують молекули довгих некодуємих РНК у різних тканинах. У огляді також підсумовуванні докази ролі ERVs у переформатуванні регуляторних мереж. Ці дані підтверджують припущення, згідно якому ERVs ретровіруси сприяють виникненню нових регуляторних програм і геномних інновацій впродовж еволюції.

КЛЮЧОВІ СЛОВА: ендогенні ретровіруси, енхансери, промотори, регуляторні програми, гени довгої некодуємої РНК, R- оперон, мобільні елементи, еволюція генома

ЭНДОГЕННЫЕ РЕТРОВИРУСЫ КАК ГЕНЕТИЧЕСКИЕ МОДУЛИ, ФОРМИРУЮЩИЕ РЕГУЛЯТОРНЫЕ СЕТИ В ТЕЧЕНИЕ ЭВОЛЮЦИИ

Попов Н. Н.^{1,2}, Колотова Т. Ю.¹, Давиденко М. Б.¹

¹ ГУ «Институт микробиологии и иммунологии им. И. И. Мечникова НАН Украины», ул. Пушкинская, 14, г. Харьков, 61057, Украина

² Харьковский национальный университет имени В. Н. Каразина, пл. Свободы, 6, г. Харьков, 61022, Украина

Эндогенные ретровирусы (ERV) являются потомками экзогенных ретровирусов, которые внедрились в геном половых клеток и стали передаваться по наследству. В течение эволюции ERV ретровирусы приобрели энхансеры и промоторы, которые позволили им экспрессироваться в различных тканях. Поскольку ERV обладают регуляторными элементами, то можно предположить, что они способны форматировать и переформатировать регуляторные сети генома с помощью внедрения промоторов и энхансеров в новые локусы генома при ретротранспозиции.

В обзоре собраны данные, которые подтверждают роль ERV ретровирусов в формировании ряда энхансеров и промоторов генов, кодирующих белки, и генов, кодирующих молекулы длинных некодирующих РНК, в различных тканях. В обзоре также суммированы доказательства роли ERV в переформатировании регуляторных сетей. Эти данные подтверждают предположение, согласно которому ERV ретровирусы способствуют возникновению новых регуляторных программ и геномных инноваций в течение эволюции.

КЛЮЧЕВЫЕ СЛОВА: эндогенные ретровирусы, энхансеры, промоторы, регуляторные программы, гены длинной некодирующей РНК, R-оперон, мобильные элементы, эволюция генома

INTRODUCTION

It becomes more and more obvious that animals and plants live in symbiosis with microorganisms. And not only with bacteria and the simplest, but also with viruses. Understanding the possibility of symbiotic relations with viruses came at the very last time, before that they were considered only as parasites. The reason for the one-sided assessment of the role of viruses is the difficulty in studying the laws underlying the symbiosis, the secrecy of such laws from the eyes of researchers. However, as the relationship between viruses and animals has been studied, it has come to realize that parasitic relationships leading to the development of pathologies are an exception rather a rule in the relationship between viruses and organisms. In most cases, co-operative relationships are established between them [1].

Integration of viral genomes, including genomes of RNA viruses, into the host genomes occurs with an unexpectedly high frequency [2]. Representatives of a number of RNA and DNA viruses were found in the vertebrate and human genome, among them Ebola virus, filoviruses, coronaviruses, circoviruses, hepadnaviruses and parvoviruses [3–7]. Integration mechanisms are the introduction of retroviruses using reverse transcriptase, sometimes accompanied by

recombination with other viral sequences [8], and also, possibly, insertion into the genome during the DNA breaks repairing using the mechanism of joining non-homologous ends [9].

Viruses participate along with bacteria in the horizontal transfer of genetic material, including mobile elements between organisms, which play a big role in adaptation of organisms and evolution to the external environment [10].

Viruses don't only carry mobile elements, but can also contribute to the emergence of new mobile elements. Thus, new DNA transposons can arise as a result of merges of transposons and DNA-containing viruses. The mechanism of such a fusion is recombination.

Exogenous retroviruses, while integrating into the genome during evolution, transformed into one of the fractions of mobile elements, called endogenous retroviruses (ERV).

Mobile elements can change the genome both actively and passively. ERV, transferring genes and cis-regulatory elements, change the genome actively. Mobile elements contribute passively to ectopic recombination and, accordingly, to the occurrence of duplications, deletions or karyotypic rearrangements [11].

Changing the genome, mobile elements can facilitate and accelerate evolution [12–13]. The presence of mobile elements in the genomes is a certain risk. It reduces somewhat the current

adaptation of the organism, but gives advantages in the future. Species containing many mobile elements in the genome acquire a certain evolutionary potential, and if the environment, the habits of the animal or the habitat change, in case of an animal colliding with some challenges, they have a tool for rapid genome change and response to challenges.

Indeed, ERVs played a huge role in the formation of vertebrate genomes, and primarily in the formation of regulatory elements of the genome and the reformatting of genetic regulatory programs in the formation of species, which will be discussed in detail in the article. Emergence of the placenta in mammals is the example of evolutionary innovation, in the origin of which endogenous retroviruses played an important role [13].

Thus, mobile elements formed by retroviruses or acquired due to horizontal transport with the help of viruses are a tool of rapid evolutionary changes or saltation that contradict the dominant ideas about gradual changes in the genome.

FORMATION OF VERTEBRATE GENOMES OF BY ENDOGENOUS RETROVIRUSES

Mobile elements have largely formed a fraction of the repeating sequences of the genome. In vertebrate genomes, the proportion of the repeating sequences is from 1.2 % in primitive fish to 38 % in reptiles [11]. The genome of birds contains the relatively low number of repeats about 6–12 %. At least one third of the genome in mammals is formed by repeating elements, and the genome of some primates is half composed of repeating sequences [11]. According to modern estimates, 69 % of the human genome is repetitive genetic elements [14].

In general, at least in vertebrates, the relationship between the number of repeating sequences and the complexity of the organism is traced. The more evolutionarily developed the species are, the more their genome contains non-coding proteins of genes, including repetitive DNA [11].

Repeating sequences of vertebrates are largely formed by mobile elements, including ERVs. During the evolution of vertebrates, exogenous retroviruses were implanted many times in germinal cells and transmitted to the offspring, becoming ERVs [15]. All placental mammals contain endogenous retroviruses.

They played a large role in the formation of the placenta of mammals as was noted above [13].

Currently, it has been possible to observe the process of endogenization of retroviruses in Australian koalas. This species was transported under threat of extinction to the islands in the early 20th century. On the islands, koalas were infected with marmoset retroviruses and gibbon leukaemia virus. Many of them died, but some survived and gained stability. Retroviruses were introduced into germinal cells in surviving koalas and viruses were transmitted to offspring [16].

ERVs compose 8 % of the human genome [17]. In comparison, all sequences that encode proteins make up about 1.5 % of the human genome [17]. Complete ERVs encode a group-specific protein Gag, Pro protease, Pol Polymerase and occasionally envelope protein. Long terminal repeats (LTRs) flank ERVs on both sides. LTR elements are necessary for replication of the retrovirus and contain cis-regulatory sequences which transcription factors specifically bind with, as well as promoters which transcription starts from. The length of the long terminal repeats is about 1000 bp. Recombination between 5' and 3' LTR sequences of endogenous retroviruses results in the formation of single LTRs, 577,000 of which were detected in the human genome [18]. 90 % of ERVs are single LTR elements.

The role of retroviruses in genome formation is determined by the fact that they contain a wide variety of regulatory elements. ERVs are potential sources of enhancers, alternative promoters, splice sites [19], and sites for polyadenylation [20]. So about 40 transcription factors can bind to LTR elements and regulate the transcription of human ERV type K (HERVK) [21]. The presence of regulatory elements creates a potential for tissue-specific expression of the retroviruses themselves, as well as for reformatting the expression of genes of the host genome.

Today, it is not known whether the ERVs contain retroviruses cis-regulatory elements that allow them to interact with trans-regulatory factors of the cell, prior to insertion into the genome, or they acquire them after the introduction. At least in some cases it was possible to show that cis-regulatory elements existed in retroviruses before they were introduced into the genome. So, elements of mice containing the regulatory modules RLTR9B2, RLTR9D and RLTR9E inherited

these modules and the ability to regulate gene expression from progenitor retroviruses prior to their introduction into the genome [22].

TISSUE-SPECIFIC ACTIVATION OF ENDOGENOUS RETROVIRUSES

For a long time it was believed that mobile elements are epigenetically suppressed and therefore cannot play an active role in the regulation of gene expression. However, a number of data allowed overcoming this error.

First, mobile elements and endogenous retroviruses among them form tissue specifically DNase I sensitive regions. These regions differ of about 100 times based on sensitivity to DNase I. DNase I sensitive regions are chromatin regions with an open chromatin configuration, that is functionally active regions. Scientists have constructed a map of DNase I sensitive regions of the genomes of a number of human cell lines [23]. About 2.9 million DNase I sensitive regions in total were found in the human genome. Approximately 3 % of the DNase sensitive regions are located in the start site region of the transcription of the gene-encoding genes, but lies within 2.5 kb from the point at which the 5 % of transcriptions start. The remaining 95 % of DNase I sensitive regions are located at a great distance from the starting transcription areas in introns and in intergenic regions. The formation of DNase I sensitivity of sites located at large distances is largely tissue-specific.

44 % of DNase I sensitive regions are located in mobile elements. Moreover, if we consider primate-specific sensitive regions of DNase I sensitive regions, then this value reaches 63 % [24]. DNase I sensitive regions are mostly concentrated in long terminal repeats of LTR endogenous retroviruses. Mapping of DNase sensitive sites in normal, embryonic and cancer cells has shown that up to 80% of ERVs in the human genome form tissue specifically DNase I sensitive regions with an open chromatin structure [24]. Tissue-specificity is determined by cis-regulatory sequences of LTR elements. The formation of an open chromatin structure by LTR elements is often associated with the expression of neighbouring genetic loci [24].

Secondly, it was previously thought that mobile elements inactivity is to a large extent due to the hypermethylation of their DNA sequences. However, the study of methylation of 928 subfamilies of mobile elements in

embryonic and terminally differentiated human tissues showed that the DNA of these elements is tissue-specific and specific with respect to the kind of mobile element being hypomethylated [25]. Certain classes of ERVs were mostly tissue specifically hypomethylated. We studied the hypomethylation of mobile elements in only 4 types of cells. About 10 % of the studied TE subfamilies are hypomethylated in these tissues. However, if we study more cells, it is likely that a significantly greater percentage of mobile elements are tissue-specific hypomethylated.

A significant part of the genes among located close to the hypomethylated tissue-specific mobile elements is composed of genes encoding proteins necessary for this type of tissue, and gene expression correlates with the hypomethylation of nearby mobile elements. Moreover, hypomethylation is accompanied by the acquiring of a typical epigenetic marker of enhancers by these areas. Many of the hypomethylated mobile elements do have enhancer activity, which is detected by the reporter method. Also, many of these sequences have binding sites to transcription factors that are specific to the respective tissues. Therefore, hypomethylated sequences of mobile elements can potentially function as enhancers. Nevertheless, this is not proven.

Thirdly, an LTR element contains cis elements, and often even clusters of cis-elements which tissue regulatory transcription factors binding with. The total number of DNA fragments came from human endogenous retroviruses is estimated to be 717,778. Approximately (~15 %) of the 110,000 fragments contain at least one binding site with a transcription factor [26]. According to recent estimates, the human genome contains 794,972 binding sites with 97 transcription factors [27].

On the average, about 20% of binding sites with 26 regulatory transcription factors are located in mobile elements, mainly in LTR elements in human and mouse genomes. Some of the mobile elements formed 5 %, and the other 40 % of the overall number of all binding sites with a certain transcription factor [28]. Binding sites with transcription factors have an open chromatin structure in the LTR elements, and therefore DNase sensitivity, they are hypomethylated and contain epigenetic modifications of histones typical for enhancers. On the average, 66 % of the binding sites of transcription factors with mobile elements are formed tissue specifically [28].

Three main clusters can be distinguished among LTR elements of human ERV (HERV) based on binding patterns with transcription factors [27]. The first class includes LTR elements associated with transcription factors of pluripotency (SOX2, POU5F1 and NANOG). For example, evolutionarily young LTR7 elements transcribed actively in pluripotent cells and they are enriched with binding sites with SOX2, POU5F1 and KLF4 transcription factors of pluripotency. The second class includes LTR elements that bind to factors expressed in the embryonic ectoderm and embryonic mesoderm (GATA4 / 6, SOX17 and FOXA1 / 2), and the third class elements bind to hematopoietic transcription factors (SPI1 (PU1), GATA1 / 2 and TAL1). There is also a group of HERV containing a binding site with CTCF factors, i.e. cis sequences capable of forming insulators and topological domains [27].

Binding sites with transcription factors are not only tissue-specific, but also species-specific. Up to 25 % of all binding sites in embryonic stem cells of humans and mice with key transcription factors of pluripotency OCT4 and NANOG came from mobile elements specific to these species, including HERV [29].

More than 98 % of 132,197 0 binding sites with 26 transcription factors localized in the mobile elements of the human genome, are absent in the genome of mice [28]. At the same time, there are also conservative binding sites. In all likelihood, species-specificity of cis-regulatory elements arises from the introduction and amplification of specific mobile elements, which occurred after the separation of the two lines leading to the appearance of human and mice. Another interesting phenomenon is that there is an expansion of species-specific binding sites with transcription factors in both genomes.

Finally, ERVs can not only contain epigenetic markers of active chromatin and bind tissue-specifically to transcription factors, but can also be expressed tissue-specifically and even in response to environmental conditions. Tissue-specificity of ERVs expression is confirmed by data obtained from the ENCODE program, as well as from the number of other studies [30]. Thousands of retroviral sequences are specifically activated in cells, especially embryonic cells, cancer cells, as well as in response to various stimuli [31].

ERVs of the human genome are expressed in oocytes, zygotes, 2–8-cell embryos, morulae and blastocysts, as well as in embryonic stem cells. To the greatest extent they are expressed at the stage of development of oocytes to 4-cell embryos [32]. The expression of ERVs is reduced from the stage of the 8-cell embryo. ERVs are expressed stage specifically, and also specific for differentiated cell populations that arise among blastocysts. The majority of ERV elements expressed during the listed stages of embryonic development are not activated in the tissues of an adult organism. Specific cis-regulatory sequences of LTR elements determine the stage specificity of expression of ERV.

In B lymphocytes, expression of endogenous human and mouse retroviruses is activated by stimulation of cell proliferation *in vitro* and *in vivo*, as well as in chronic diseases, including B cell lymphoma. A small number of LTR elements, which are constantly activated in stimulated B cells, are detected, with one of the Xmv45 retroviruses being expressed in a larger amount than the rest of them combined. The expression of a large number ERVs is at the same time activated during B cell transformation [33].

Thus, LTR elements acquire DNAase I sensitivity in a tissue-specific manner, become hypomethylated, acquire epigenetic histone modifications typical for transcriptionally active loci and enhancers, bind to transcription factors and even are transcribed [24–25, 28]. Together, these data indicate the biochemical activity of cis-regulatory elements of ERVs and suggest that they are the source of regulatory sequences of the genome.

However, these data are insufficient to state that cis-regulatory elements of ERVs really control the expression of protein-coding genes. For example, cis-regulating binding sites can serve as a buffer for transcription factors, landing sites from which transcription factors begin to scan DNA in search of an attachment site. At the same time, species-specific ERVs capable of tissue-specific activation are an excellent tool for creating new regulatory elements and genetic programs in the evolution process. Experimental data confirm the role of endogenous retroviruses in the formation of alternative promoters and enhancers, as well as genes of long non-coding RNAs and the role in the reformatting of regulatory networks [29–33].

ENDOGENOUS RETROVIRUSES AS A SOURCE OF REGULATORY ELEMENTS AND NON-CODING RNA GENES

When cis elements of the LTR intrude close to the genes, the sequences of ERVs retroviruses can form alternative promoters, thereby increasing the number of gene isoforms, while the gene can acquire new tissue-specific expression [34–38]. At the moment, a small number of cases of the alternative promoters' formation by ERVs have been well studied.

A classic example is the acquisition of the ability to express in the human salivary glands a gene that encodes the enzyme amylase. It acquired this ability due to the insertion of the LTR element and the formation of an LTR alternative promoter [34].

The human gene encoding the B3GALT5 metabolism factor is expressed in many differentiated cells, but the primate-specific alternative promoter formed by LTR element is used in cells of the large intestine [35].

Prolactin is not produced by the uterus of a number of mammals, such as rabbits, dogs, pigs and armadillos. At the same time, it is produced during pregnancy by the uterus of the primates, mice and elephants. Regulation of the expression of the gene encoding the precursor of prolactin in the uterus has evolved in mentioned mammals. The alternative gene promoter contains the DNA transposon in humans and spider monkeys, as well as the MER39 retrovirus, and the mouse alternative promoter originated from the MER77 endogenous retrovirus [36].

The NAIP gene encodes an inhibitory apoptosis of neuronal proteins. The NAIP gene is regulated by a variety of promoters both in humans and in mice that do not coincide between them. LTR elements of ERVs in the human genome have formed an alternative gene promoter that allows the gene to be expressed in testicles [37]. Rodents contain several copies of the NAIP gene. The main constitutive promoter of these genes is formed by LTR elements of the ORR1E ERV. In addition, the MT-C ERVs formed a minor promoter of two copies of the NAIP gene.

The gene that encodes the erythroid transcription of Pu. 1 mice factor has an alternative promoter. The promoter is formed by the LTR element of the ORR1A0 retrovirus located in the intron of the Pu.1 gene [38]. A

chimeric transcript of Pu.2 is formed when expressed from this promoter. It induces erythroid differentiation in vitro.

Thus, cis-regulatory sequences of LTR elements of ERVs initiate the synthesis of transcripts from gene-encoding genes in addition to its own synthesis [34, 36, 38]. Transcripts formed are chimeric RNA molecules by structure. Chimeric transcripts contain sequences of retroviruses at the 5' end. The rest of the transcripts are identical to the sequences of the corresponding genes.

ERV are potentially able to quickly reformat regulatory networks possessing the ability to move in the genome, and on the other hand, being introduced near the genes and supplying them with alternative promoters. Indeed, ERVs in the evolution participated species-specific and tissue-specific in the formation of regulatory programs, which will be discussed below.

At the same time, ERVs retroviruses can intrude at large distances from genes encoding proteins and form distal regulatory elements – enhancers.

LTR9 element located at a distance of 40–70 kb upstream of human gamma and beta globin genes forms an enhancer that activates the expression of the β -globin gene in transgenic mice [39]. Moreover, even hypermethylated ERV9 LTR possesses enhancer activity since in vivo deletion of LTR by CRISPR-cas9 method suppresses gene activity more than 50 % [40].

MaLR LTR is an enhancer that controls the expression of the proopiomelanocortin Pomc gene in the pituitary and hypothalamus of mammalian [41]. And the enhancer, which results from the MaLR of LTR element, provides about 80 % of the Pomc gene expression [42].

The tissue-specific enhancer hsERVPRODHD formed by the ERV controls the transcription of the PRODHD gene in the hippocampus. The gene encodes proline dehydrogenase and, apparently, participates in the synthesis of neurotransmitters in the central nervous system [43]. Expression of the gene is necessary for the normal functioning of the central nervous system. The enhancer activity of hsERVPRODHD is manifested in the hypomethylated state during the attachment of the transcription factor SOX2.

The enhancer can be formed by co-opting the regulatory sequences of several mobile elements [44].

So, for example, the region originating from AmnSINE1 of the non-autonomous retrotransposon does not have enhancer activity, but nevertheless preserves its conservatism in the platypus and human genome. However, the integration of the DNA of the transposon allows the acquisition of a binding site with the transcription factor Msx1 after the single-pass divergence. The endogenous retrovirus MER117 is introduced at the next stage after the masculine divergence, resulting in the formation of a modern enhancer [44].

There are reasons to believe that many enhancers have been formed in this way. Indeed, in the human genome 54 (8.6 %) out of 626 conservative AmnSINE1 loci are associated with other evolutionarily conserved mobile elements, among which LTR-containing ERVs and DNA transposons. Such sites are potential enhancers, however, this has not been proved yet.

Data obtained from the ENCODE project showed that at least 75 % of the genome is transcribed with RNA formation, despite the fact that the protein-encoding DNA sequences make up only 1.5 % [45]. These data were obtained by studying of 15 cell lines. Therefore, the data is understated and in fact the percentage can be even higher. As already discussed, most of the genome repeating sequences, including ERV, are actively transcribed [46].

A significant portion of the transcribed RNA is formed by long non-coding RNA (lncRNA). lncRNAs are molecules whose length is above 200 bp. According to the NONCODE database, there were 96,308 genes encoding lncRNA and 172216 lncRNA transcripts in 2018 [47].

The genes encoding lncRNA and mRNA are similar in size and structure. The transcription of lncRNA starts from promoters that contain binding sites with transcription factors and epigenetic markers typical for transcriptionally active genes [45]. The lncRNAs are mainly transcribed by RNA polymerase II. The lncRNA molecules contain a cap at the 5' end and are polyadenylated at the 3' end and are characterized by alternative splicing [45].

The tissue-specificity of expression is more characteristic for lncRNA genes than for protein-encoding genes. lncRNA genes are expressed not only tissue-specifically, but, apparently, each cell contains its unique set of lncRNA molecules [48].

lncRNAs are not a homogeneous class of molecules, but a mixture of molecules with different biochemical mechanisms of action and function. The ability to interact complementarily with DNA, modular organization and alternative splicing allow lncRNA molecules to function as address epigenetic modulators, which, deliver epigenetic information to the right place in response to external actions and the existing metabolic situation in the cell.

Mobile elements play an important role in the occurrence of long non-coding RNA in vertebrates from fish to humans [49, 50]. ERVs played the main role in the formation of lncRNA genes in mice and humans, and DNA transposon played such role in the genome of zebra fish. Mobile elements formed largely the primate genes encoding lncRNA as well as. Thus, mobile elements were detected in 83 % of 9241 lncRNA molecules and accounted for 42 % of the total sequence of all human lncRNAs [49]. ERVs appear to be the main factor contributing to the emergence of new lncRNA genes due to their ability to species-specific incorporation into genomes and spread in them, as well as due to regulatory sequences presence in them. Genes of human lncRNA are enriched by the sequences of ERV1, ERVL-MaLR, ERVL and ERVK retroviruses.

ERV insertion could diversify the already existing lncRNA genes in some cases during the evolution. At the same time, in other cases, new lncRNAs appeared as a result of the ERV insertion. Apparently, the presence of cis-regulatory elements and tissue-specific transcription were the main properties that allowed retroviruses to form new lncRNA genes during evolution [50].

Indeed, ERVs are located mainly in the region of 5' ends of lncRNA transcripts in a sense orientation, i. e. in a position that allows LTR elements to initiate transcription and regulate it. For example, transcription of the lncRNARoR gene in human embryonic stem cells is controlled by the cis-regulatory sequences of the LTR7/HERVH element that bind to OCT4, NANOG, and SOX2 transcription factors [50].

A new class of lncRNA has recently been discovered, which is called chromatin enriched RNA (cheRNA). Almost all genes encoding completely cheRNA are formed by retrotransposons, including ERV.

Most cheRNA molecules interact with RNA polymerase II and remain bound to chromatin by transcription or stopping transcription [51].

In general, cheRNA genes are expressed tissue-specifically. Proximity to the genes encoding the expressed cheRNA cells in this type of cell is most accurately combined with the transcriptional activity of the protein-encoding genes. Moreover, such proximity is more often associated with the transcriptional activity of protein-encoding genes than the expression of long non-coding RNAs of other classes and even the transcription of enhancers in this type of cells. Deletion of several cheRNA molecules resulted in a significant, about 75 % reduction in the transcription of a nearby lying gene. Together, these data indicate that cheRNA acts as a transcriptional activator. However, it is not clear if the transcription of these loci itself causes such an effect, or synthesized RNA molecules are needed.

A number of data suggest that a rapid species-specific formation of new lncRNA genes occurs due to mobile elements during evolution [50, 52]. A significant portion of the genes of human lncRNA arose recently, apparently due to the activity of mobile elements. Indeed, 40 % of lncRNA containing the mobile elements are specific for primates [50].

Very interesting data were obtained during the study of tomatoes. Comparison of the lncRNAs of two tomato species *Solanum lycopersicum* and *Solanum pimpinellifolium* showed that a small part 6.7 %, (24 of 353) of lncRNA molecules appeared to be common for both species [52]. And only less than 0.4 % of lncRNAs are common for all sequenced genomes of tomatoes and potatoes. Apparently, the appearance of lncRNA genes is associated with mobile elements in the genomes of two species of tomato, since 85 % of *Lycopersicon*-specific lncRNA molecules contain mobile elements.

Thus, ERVs can regulate gene activity, not only by forming enhancers and alternative promoters, but also by species-specifically creating genes of lncRNA. lncRNA promoters are formed thanks to retroviruses, as well as tissue-specific regulatory networks controlling the expression of lncRNA genes. lncRNA, regulate the transcription of protein-coding genes in turn, including the transcription of adjacent genes. And, finally, mobile elements,

which ERVs are among, can cause rapid evolutionary changes of lncRNA genes.

REFORMATTING REGULATORY NETWORKS BY ERVS

Changes in genetic regulatory programs underlie phenotypic differences between species and within the species. However, the mechanisms of changes in regulatory networks in evolution are poorly understood. Extremely intriguing is the regulatory networks feature to change quickly and consistently. In order for the genome to acquire a certain set of regulatory elements for coordinated regulation, a number of corresponding single mutations must appear in those regulatory regions of the genome that regulate the corresponding genes. So that is necessary that a multiplicity of identical mutations arose in a variety of regions of the genome in a short evolutionary time interval. At the same time, several implementations of mobile elements containing regulatory elements in several regions of the genome are needed to form a new regulatory network.

LTR elements of ERVs contain not only cis-regulatory elements, but also cis-regulatory modules, which significantly expands their potentialities in the formation of new regulatory networks [22]. Cis-regulatory modules are a set of cis-regulatory sequences binding transcription factors that co-regulate the activity of target genes. Thus, ERVs containing single cis-regulatory elements and clusters of binding sites with transcription factors are a good natural tool for rapid and consistent changes in regulatory chains.

One can imagine at least three ways in which ERVs form and reformat regulatory networks. First, they can distribute alternative promoters in the genome. Secondly, they can contribute to the acquisition of new enhancers by multiple genes. And thirdly, they can form networks of tissue-specific and species-specific expressed lncRNA genes.

The spread of alternative promoters by ERVs is confirmed by a variety of data. Synthesis of about 6–30 % transcripts containing cap at the 5' end isolated from various embryonic and differentiated cells of mice and humans starts with mobile elements [53]. These transcripts are mainly tissue-specific.

Expression of a number of ERVs is activated in mature oocytes and at the stage of the double cellular embryo of mice, but with further

development of the embryo their expression is suppressed. Chimeric transcripts are synthesized when LTR elements are activated in mature ovum, their transcription starts from alternative promoters formed by MaLR and ERVK retrovirus families [54]. It was possible to detect more than 500 chimeric transcripts, in the formation of which 307 protein coding genes take part, using sequencing RNA sequences of cells isolated from a population of mouse embryonic stem cells and corresponding to cells of a 2-cell embryo. Synthesis of chimeric transcripts begins with LTR elements of the retroviruses of the MERVL family and extends beyond the retrovirus, including genes [55].

When the mouse 2–4-cell embryos 259 transcript were sequenced, the relationship between activation of LTR elements of ERVs and a temporary and strong increase in the transcription of adjacent genes were confirmed. LTR elements regulating significantly the transcription of neighbouring genes have been enriched by binding sites with home box containing transcription factors [56]. As it is known, home box-containing transcription factors are key regulators of morphogenesis in embryonic development.

However, the question arises whether chimeric transcripts has the function. In, firstly, the fact that more than 90 out of the 626 chimeric transcripts synthesized in mouse embryonic stem cells retain the open reading frame proves their functional significance [55].

Secondly, genes encoding critical for the differentiation of early embryonic cells transcription factors GATA and TEAD, use LTR as alternative promoters [55].

The promoter of the Dicer1 gene is the promoter containing the CpGIsland in most cells, and the promoter in mouse oocytes is the LTR element of the ERVs – MaLR [57]. Deletion of an alternative promoter reduces the expression of the Dicer1 gene in oocytes and causes infertility [58].

Thirdly, cultured mouse embryonic cells that express transcripts from LTR sequences and cells that do not express transcripts from LTR elements have different phenotypes [55].

Regulatory networks that are altered by ERV-derived enhancers are found in cells of tissues associated with sexual reproduction, in embryonic stem cells and on the early stages of embryogenesis, in erythroblasts and in

terminally differentiated liver cells and the immune system.

The cis-regulatory element of the MaLR retroviruses binds to the Tbx6 transcription factor. Expression of at least four genes, whose enhancers are formed by MaLR LTR, significantly decreases in mice deficient in Tbx6 transcription factor [59]. Tbx6 regulates gene expression during early embryogenesis.

Mice placenta cells used ERVs for species-specific reformatting of the regulation of gene expression. A class of retroviruses RLTR13D5 was detected by revealing the profile of epigenetic markers and binding sites with the transcription factors of stem cells of trophoblasts in mice and rats. It forms a significant part of active in the placenta enhancers [60]. These ERVs retroviruses contain binding sites with transcription factors Eomes, Cdx2 and Elf5, which play key role in trophoblast regulatory networks. Enhancers made by retroviruses have formed a network consisting of hundreds of elements, and control the synthesis of one-third of all placental specific transcripts [60].

However, the RLTR13 family of ERV, which formed a regulatory network of trophoblast stem cells, is specific for mice. Even in rats, most mouse enhancers formed by the ERVs of the RLTR13 family are absent.

LTR elements played a significant role in the formation of new regulatory networks that evolved in monkeys in the liver.

Indeed, 77.1 % of the cis-regulatory elements specific for monkeys and virtually all the cis-regulatory elements those are specific for the human genome, cis-regulatory elements of the liver overlap with retrotransposons [61]. Regulatory activity of a number of elements containing retrotransposons was confirmed by studying the activity of synthesized consensus sequences in cultured liver cells HepG2 using a luciferase reporter assay.

The LTR elements of ERVs and the SVA retrotransposons containing retroviral elements contributed to the greatest extent to the formation of regulatory programs that appeared at the last stages of the evolution of regulatory programs. At the same time, only 16.0 % of evolutionarily conserved cis-regulatory sequences contain mobile elements. Thus, cortical regulatory programs that ensure the identity of liver cells persist throughout the evolution of primates, while peripheral

regulatory programs are rapidly evolving with the help of ERV.

The formation of a line-specific and tissue-specific regulatory network activated by interferon is the most rigorously demonstrated example of the participation of ERVs in the reformatting of regulatory networks by the formation of new enhancers.

Researchers used the CRISPR-Cas9 method for deletion part of ERVs specific for primates of the MER41 family containing cis elements which transcription factors activated by interferon bind to [62]. The genes were no longer regulated by interferon as a result of the deletion, their expression decreased, which manifested among other things in the alteration of various phenotypic signs, among which decreased inflammatory response in response to infection. Thus it was possible to show that the human genome contains 962 ERVs of the MER41 family, which bind to the transcription factors STAT1 and IRF1 in at least one type of cells. Together, these data show that the MER41 elements have formed a regulatory network controlled by interferon.

MER41 contains a tandem sequence that binds to the transcription factor STAT1. The same sequence contains MER41-related retroviruses of lemurous, bats, carnivores and artiodactyls. Therefore, it can be assumed that related ERVs form interferon-induced enhancers in different mammalian species. Indeed, the consensus sequence of MER41-like LTR elements of dogs and cows shows activity in the luciferase reporter assay in the HeLa cell line in response to induction by interferon.

Mice do not have a MER41 family of retroviruses, but endogenous gamma-retroviruses RLTR30B specific for mice also formed enhancers controlled by interferon. Methods of bioinformatics allowed revealing the connection between RLTR30B elements and genome loci containing the immune response genes. Consequently, two different families of endogenous retroviruses formed convergent interferon-regulated immune response programs in two mammalian species: humans and mice.

Next ERV have formed regulatory networks consisting of lncRNA genes.

LTR elements of mice ERVs control the transcription of lncRNA genes in the post-mitotic phase of the cell cycle of spermatocytes and round spermatids [63]. LTR elements have formed tissue-specific lncRNA promoters, thus

creating a regulatory network that allows them to regulate the expression of lncRNA genes in spermatogenesis. Interestingly, a small part of the transcript elements initiated with ERVs encodes the open reading frames, which allow the peptides synthesis.

LTR elements of the ERV1 form regulatory sequences that control the tissue-specific expression of lncRNA in human testes [64].

The synthesis of transcripts starts from a number of HERV expressed in oocytes, zygotes, 2–8 cell embryos, morulae and blastocysts, as well as in embryonic stem cells, and extends beyond the retrovirus. Then, RNA molecules are formed as a result of splicing containing exons of non-retroviral nature. For example, 95 % of the elements of the MLT2A1 family of retroviruses encode such RNA molecules. Most of the non-retroviral exons are unannotated and, apparently, form a non-coding lncRNA [32].

The lncRNA genes containing human H retroviruses of the HERVH family are specifically expressed in embryonic stem cells and induced pluripotent stem cells. Expression of HERVH containing lncRNA genes is necessary to maintain pluripotency in cells [49]. 127 lncRNA genes are transcribed in the embryonic stem cells, containing the LTR7 elements of the HERVH in the sense orientation near the sites of the beginning of transcription. LTR elements of HERVH containing lncPHK genes bind with OCT4 and NANOG pluripotency transcription factors [49].

Another method by which endogenous retroviruses can form regulatory networks is found in erythroblasts, which ERV9 human retroviruses have formed a regulatory network in, creating both enhancers, and lncRNA genes [65].

lncRNA transcribed from LTR retrotransposons of ERV9 activates the transcription of key erythroid genes and modulates erythropoiesis *ex vivo*. Theoretically, ERV9 lncRNA can regulate the transcription of key erythropoiesis genes by acting in cis or in trans, diffusing from the site of synthesis to the target gene, which can be located on another chromosome. To understand the mechanism of action of ERV9 lncRNA, the synthesized transcripts were analysed before and after the global deletion or locus-specific deletion of ERV9 lncRNA in human erythroblasts containing ~ 4000 copies of ERV9 LTR and in mice erythroblasts containing one

transgenic copy of the primate specific ERV9 LTR in the locus, which encodes the beta-haemoglobin gene.

As a result, it was shown that ERV9 lncRNA, synthesized from the ERV9 LTR element, which controls the transcription of the beta-globin gene, remains associated with LTR and interacts with transcription factors and polymerase II, forming an enhancer complex. The enhancer complex interacts with the downstream gene promoter, thereby activating transcription.

In ERV9 erythroblasts, lncRNA is transcribed from many of 4000 copies of ERV9 retrovirus, stabilizes the enhancer complex and activates the transcription of a number of genes in cis, including key erythropoiesis genes, including haemoglobin genes, as well as genes encoding the transcription factors of erythropoiesis KLF1 and CCNDBP1.

So, the regulatory networks of early embryonic tissues, pluripotent embryonic cells, liver cells, erythroblasts, interferon-induced gamma genes are reformatted species-specifically by endogenous retroviruses. These data support the hypothesis that ERVs are used during evolution for species-specific reformatting of regulatory programs. Moreover, ERVs are convergent used by various species to form regulatory networks induced by gamma interferon. Examples of convergent use of ERVs for the formation of promoters were given earlier. These examples of completely amazing convergence remain a mystery of evolution. Is the use of mobile elements for similar regulatory programs creation in different species of animals due to chance? Or as Shapiro writes in his article: the elements responding to some as-yet-to-be-defined regulatory process that guides the adaptive integration of newly established regulatory signals? To answer this question, it is necessary to calculate the probability according to which such a regulatory network can form in evolution, based on the assumption of chance. On the other hand, there is data currently accumulating according to which the introduction of mobile elements is not a random process, but occurs localized in time and space [11]. It is interesting that another type of retrotransposon, long LINE intersperse elements, are activated in neurons during differentiation [66]. The sites for introducing LINE elements in neurons are not accidental. They are predominantly localized in enhancers

actively transcribed in the neurons genes [67]. Moreover, it is assumed that LINE elements are implanted into double DNA ruptures that are formed in genes and actively transcribed in terminally differentiated neurons. In other words, the localization of retrotransposon introduction sites is determined by the functional activity of neurons [68].

HYPOTHESIS OF R-OPERON

It is well known that homologous DNA sequences have the ability to recognize and interact with each other. The interaction between homologous sequences will inevitably affect the location of DNA molecules in the core space. Therefore, a model has been proposed, according to which repeated homologous elements of the genome form and/or stabilize the specific spatial structure of both interphase chromatin and mitotic chromosomes [69–70].

According to the model, when interphase chromatin is laid, an association occurs between the homologous mobile elements resulting in the formation of homologous pairs, which then form a repetition assembly (RA). The organization of interphase chromatin is the basis for formation of mitotic chromosomes. More and more homologous pairs establish contact with each other during the laying of mitotic chromosomes. The chromatin filaments gradually become denser, forming mitotic chromosomes accordingly. Thus, repeating elements form the skeleton of interphase chromatin and mitotic chromosomes.

However, the question arises whether the localization of repetitions, mainly formed by mobile elements, is primary. Whether it is the driver of the process of laying interphase chromosomes or just a consequence of chromatin packing.

The following data are further indication of the assumption of the primacy of the interaction between homologous mobile elements

Firstly, it is well established that chromatin has the ability to localize in loci with homologous DNA sequences [71–73].

Secondly, it has been studied the interchromosomal contacts of various families of repeats in the genomes of human embryonic stem cells, drosophilas and mice, and also in three human cell lines to test this hypothesis and built a global picture of the spatial organization of repeats in chromosomes [72]. The degree of localization of repeats formed by

different families of mobile elements, including LTR-containing ERV, DNA transposons, short interspersed elements (SINE) and long interspersed elements (LINE) was quantitatively evaluated.

All families of mobile elements contained subfamilies prone to localization in nuclear space. That is, the formation of clusters in three-dimensional space turned out to be a common feature of mobile elements genomes of different organisms.

It was possible to show the conservatism of the organization of synthetic blocks in the nuclear space of mice and humans. Synthetic regions containing a similar set of mobile elements form similar spatial contacts in the genomes of mice and humans.

The most frequently collocated mobile elements in space are evolutionarily more ancient and contain binding sites with transcription factors. The presence of cis-regulatory sequences in the localizing elements indicates the possibility of regulating the formation of three-dimensional contacts between mobile elements by transcription factors, and, consequently, by the environmental conditions and the metabolism of the cell.

Together, these data suggest that contacts between retrotransposons are not a passive consequence of chromatin packing, but actively influence the architecture of interphase chromatin. These data are in good agreement with the hypothesis of the R-operon [73]. According to this hypothesis, the eukaryotic genome forms structurally functional domains called organized repeats of operons (R-operons) in the nucleus by means of homologous interactions between mobile elements. Each R-operon consists of associated mobile elements and adjacent gene-encoding protein cells.

Representations of repetitive R-operons significantly expand the possibilities of specific regulation of genes and cooperation between genes. What does it happen due to?

Firstly, genes located at large distances in the linear genome in R-operons are in contact with each other in the three-dimensional nucleus space and thus can be regulated by a set of those cis-regulatory elements and the transcription factors that bind to and regulate each of these genes. Indeed, it has been shown that enhancers can regulate the work of genes in trans while approaching genes in space [74–76]. Thus, the R-operons formed by repetition

assemblies are structures that enable cooperative gene regulation by a set of those cis-elements that control the activity of each of the genes that are part of regulon.

Secondly, genes possessing the same cis-elements can coexist with different repetitions and be part of various R-operons accordingly.

Thirdly, the transcriptional domains formed in such way are dynamic. Due to the dissociation of repeating assemblies, old R-operons can disappear, and new domains with a different combination of repeats and genes due to the association of repetitions can form.

The dynamics of the formation of homologous sequences pairs can be influenced by the concentration of ions in the cell, the expression of proteins necessary for homologous pairing, the pattern of epigenetic modifications of histones and mobile elements, and the activity of mobile elements regulated by transcription factors.

Transcription factors specifically interacting with regulatory elements of ERVs can facilitate the establishment of contacts between mobile elements that is confirmed experimentally. Thus, HERV possessing binding sites with transcriptional regulators such as NANOG and OCT4 are localized in human embryonic stem cells. However, localization disappears in cells in which the expression of transcription factors is suppressed [74].

Transcription itself can lead to the establishment of contacts between repetitive sequences. A number of experimental data confirms the assumption about the existence of transcription factories – nuclei regions in which transcription occurs and which can contain up to hundreds of simultaneously operating RNA polymerase molecules [77]. But many ERVs are transcribed in a cell and accordingly can be localized in transcription factories. In the process of differentiation and in response to external conditions, the pattern of transcribed ERVs can change, and accordingly the contacts between them will change.

Indeed, according to a number of studies, the expression of retroviruses in cells of even one type varies considerably. This is due to epigenetic mechanisms [78–79]. Thus, R-operons will be formed not only tissue-specifically, but even specifically for each cell.

The regulation of association and dissociation of repetition assemblies is a mechanism of coordinated changes in gene activity in response to changes in cell

metabolism and promotes the emergence of new combinations of expressed genes in response to changing environmental conditions. Therefore co-expression of genes coordinated with the help of R-operons expands essentially the possibility for emergence of new temporary cooperation between genes dependent on the specific context that has developed in the cell.

CONCLUSION

Summarizing the data, we can note the following. ERVs contain regulatory elements or clusters of elements, by means of which they can be tissue-activated and transcribed. ERVs participated in the reformatting of regulatory networks and in the creation of gene-specific non-coding RNA forming binding sites with

transcription factors and spreading them inside the genome. During the evolution, different ERVs were injecting into the genomes of different species, but in a number of cases they were used by genomes to solve similar problems and reformat similar regulatory programs. Therefore, it can be assumed that the participation of ERVs in the formation of regulatory networks obeys certain laws and is not completely random.

In cells, due to homologous interactions, ERVs can form regulatory R-operons, which provide an additional way for the disintegration of old and emerging new gene associations in response to cell-building conditions and provide a level of genome plasticity that was previously difficult to imagine.

REFERENCES

1. Roossinck M. J. Move over bacteria! Viruses make their mark as mutualistic microbial symbionts // *J. Virol.* 2015. Vol. 89. P. 6532–6535.
2. Belyi V. A., Levine A. J., Skalka A. M. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes // *PLoSPathog.* 2010. Vol. 6. P. e1001030.
3. Belyi V. A., Levine A. J., Skalka A. M. Sequences from ancestral single-stranded DNA viruses in vertebrate genomes: The Parvoviridae and Circoviridae are more than 40 to 50 million years old // *J. Virol.* 2010. Vol. 84. P. 12458–12462.
4. Horie M., Honda T., Suzuki Y., Kobayashi Y., Daito T., Oshida T. [et al.] Endogenous non-retroviral RNA virus elements in mammalian genomes // *Nature.* 2010. Vol. 463. P. 84–87.
5. Gilbert C., Meik J. M., Dashevsky D., Card D. C., Castoe T. A., Schaack S. Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes // *Proc. Biol. Sci.* 2014. Vol. 281. P. 20141122.
6. Taylor D. J., Leach R. W., Bruenn J. Filoviruses are ancient and integrated into mammalian genomes // *BMC Evol. Biol.* 2010. Vol. 10. P. 193.
7. Feschotte C., Gilbert C. Endogenous viruses: Insights into viral evolution and impact on host biology // *Nat. Rev. Genet.* 2012. Vol. 13. P. 283–296.
8. Geuking M. B., Weber J., Dewannieux M., Gorelik E., Heidmann T., Hengartner H., Zinkernagel R. M., Hangartner L. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration // *Science.* 2009. Vol. 323. P. 393–396.
9. Bill C. A., Summers J. Genomic DNA double-strand breaks are targets for hepadnaviral DNA integration // *Proc. Natl. Acad. Sci. USA.* 2004. Vol. 101. P. 11135–11140.
10. Gilbert C., Cordaux R. Viruses as vectors of horizontal transfer of genetic material in eukaryotes // *Curr. Opin. Virol.* 2017. Vol. 25. P. 16–22.
11. Shapiro J. A. Living Organisms Author Their Read-Write Genomes in Evolution // *Biology (Basel).* 2017. Vol. 6(4). P. pii: E42
12. Oliver K. R., Greene W. K. Transposable elements and viruses as factors in adaptation and evolution: an expansion and strengthening of the TE-Thrust hypothesis // *Ecol. Evol.* 2012. Vol. 2, N.11. P. 2912–2933.
13. Oliver K. R., Greene W. K. Mobile DNA and the TE-Thrust hypothesis: Supporting evidence from the primates // *Mob. DNA.* 2011. Vol. 2. P. 8.
14. De Koning A. P., Gu W., Castoe T. A., Batzer M. A., Pollock D. D. Repetitive elements may comprise over two-thirds of the human genome // *PLoS Genet.* 2011. Vol. 7. P. e1002384.
15. Rowe H. M., Trono D. Dynamic control of endogenous retroviruses during development // *Virology.* 2011. Vol. 411. P. 273–287.
16. Tarlinton R. E., Meers J., Young P. R. Retroviral invasion of the koala genome // *Nature.* 2006. Vol. 442. P. 79–81.
17. Lander E. S., Linton L. M., Birren B., Nusbaum C., Zody M. C., Baldwin J. et al. Initial sequencing and analysis of the human genome // *Nature.* 2001. Vol. 409. P. 860–921.

18. Friedli M., Trono D. The developmental control of transposable elements and the evolution of higher species // *Annu. Rev. Cell. Dev. Biol.* 2015. Vol. 31. P. 429-451.
19. Kapitonov V. V., Jurka J. The long terminal repeat of an endogenous retrovirus induces alternative splicing and encodes an additional carboxy-terminal sequence in the human leptin receptor // *J. Mol. Evol.* 1999. Vol. 48. P. 248 – 251.
20. Mager D. L., Hunter D. G., Schertzer M., Freeman J. D. Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3) // *Genomics.* 1999. Vol. 59. P. 255–263.
21. Manghera M., Douville R. N. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? // *Retrovirology.* 2013. Vol.10. P. 16.
22. Sundaram V., Choudhary M. N., Pehrsson E., Xing X., Fiore C., Pandey M., Maricque B., Udawatta M., Ngo D., Chen Y., Paguntalan A., Ray T., Hughes A., Cohen B. A., Wang T. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus // *Nat. Commun.* 2017. Vol. 8. P.4550.
23. Thurman R. E., Rynes E., Humbert R., Vierstra J., Maurano M. T., Haugen E., Sheffield N. C., Stergachis A. B., Wang H., Vernot B., et al. The accessible chromatin landscape of the human genome // *Nature.* 2012. Vol. 489. P. 75– 82.
24. Jacques P. É., Jeyakani J., Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements // *PLoS Genet.* 2013. Vol. 9, N 5. P. e1003504.
25. Xie M., Hong C., Zhang B., Lowdon R.F., Xing X., Li D., Zhou X., Lee H. J., Maire C. L., Ligon K. L., Gascard P., Sigaroudinia M., Tlsty T. D., Kadlecsek T., Weiss A., O'Geen H., Farnham P. J., Madden P. A., Mungall A. J., Tam A., Kamoh B., Cho S., Moore R., Hirst M., Marra M. A., Costello J. F., Wang T. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape // *Nat. Genet.* 2013. Vol.45. P. 836-841.
26. Garazha A., Ivanova A., Suntsova M., Malakhova G., Roumiantsev S., Zhavoronkov A., Buzdin A. New bioinformatic tool for quick identification of functionally relevant endogenous retroviral inserts in human genome // *Cell Cycle.* 2015. Vol. 14. P. 1476–1484.
27. Ito J., Sugimoto R., Nakaoka H., Yamada S., Kimura T., Hayano T. et al. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses // *PLoS Genet.* 2017. Vol. 13. P. e1006883.
28. Sundaram V., Cheng Y., Ma Z., Li D., Xing X., Edge P., Snyder M. P., Wang T. Widespread contribution of transposable elements to the innovation of gene regulatory networks // *Genome Research.* 2014. Vol. 24. P. 1963–1976.
29. Kunarso G., Chia N. Y., Jeyakani J., Hwang C., Lu X., Chan Y. S., Ng H. H., Bourque G. Transposable elements have rewired the core regulatory network of human embryonic stem cells // *Nat. Genet.* 2010. Vol. 42 (7). P. 631–634.
30. Djebali S., Davis C. A., Merkel A., Dobin A., Lassmann T., Mortazavi A., et al. Landscape of transcription in human cells // *Nature.* 2012. Vol. 489. P. 101–108.
31. Haase K., Mosch A., Frishman D. Differential expression analysis of human endogenous retroviruses based on ENCODE RNA-seq data // *BMC Med. Genomics.* 2015. Vol.8. P.71.
32. Goke, J., Lu X., Chan Y. S., Ng H. H., Ly L. H., Sachs F., Szczerbinska I. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells // *Cell Stem Cell.* 2015. Vol. 16. P. 135–141.
33. Attig J., Young G. R., Stoye J. P., Kassiotis G. Physiological and Pathological Transcriptional Activation of Endogenous Retroelements Assessed by RNA-Sequencing of B Lymphocytes // *Front. Microbiol.* 2017. Vol.8. P.2489.
34. Ting C. N., Rosenberg M. P., Snow C. M., Samuelson L. C., Meisler M. H. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene // *Genes Dev.* 1992. Vol. 6. P. 1457–1465.
35. Dunn C. A., Medstrand P., Mager D. L. An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon // *Proc. Natl. Acad. Sci. U. S. A.* 2003. Vol.100. P.12841–12846.
36. Emera D., Casola C.; Lynch V. J., Wildman D. E., Agnew D., Wagner G. P. Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements // *Mol. Biol. Evol.* 2012. Vol.29. P. 239–247.
37. Romanish M. T., Lock W. M., van de Lagemaat L. N., Dunn C. A., Mager D. L. Repeated recruitment of LTR retrotransposons as promoters by the anti-apoptotic locus NAIP during mammalian evolution // *PLoS Genet.* 2007. Vol. 3. P. e10.
38. Upton K. R., Faulkner G. J. Blood from ‘junk’: the LTR chimeric transcript Pu.2 promotes erythropoiesis // *Mob. DNA.* 2014. Vol. 5. P. 15.

39. Pi W., Zhu X., Wu M., Wang Y., Fulzele S., Eroglu A., Ling J., Tuan D. Long-range function of an intergenicroretrotransposon // *Proc. Natl. Acad. Sci. U S A.* 2010. Vol. 107 (29). P. 12992–12997.
40. Hu T., Zhu X., Pi W., Yu M., Shi H., Tuan D. Hypermethylated LTR retrotransposon exhibits enhancer activity // *Epigenetics.* 2017. Vol. 12. P. 226–237.
41. Franchini L. F., López-Leal R., Nasif S., Beati P., Gelman D. M., Low M. J., de Souza F. J., Rubinstein M. Convergent evolution of two mammalian neuronal enhancers by sequential exaptation of unrelated retrotransposons // *Proc. Natl. Acad. Sci. U S A.* 2011. Vol. 108.P.15270–15275.
42. Lam D. D., de Souza F. S., Nasif S., Yamashita M., López-Leal R., Otero-Corchon V., Meece K., Sampath H., Mercer A. J., Wardlaw S. L., Rubinstein M., Low M. J. Partially redundant enhancers cooperatively maintain Mammalian pomc expression above a critical functional threshold // *PLoS Genet.* 2015. Vol. 11. P. e1004935.
43. Suntsova M., Gogvadze E. V., Salozhin S., Gaifullin N., Eroshkin F., Dmitriev S. E., et al. Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene *PRODH* // *Proc. Natl. Acad. Sci. U S A.* 2013. Vol. 110. P. 19472–19477.
44. Nishihara H., Kobayashi N., Kimura-Yoshida C., Yan K., Bormuth O., Ding Q., Nakanishi A., Sasaki T., Hirakawa M., Sumiyama K., Furuta Y., Tarabykin V., Matsuo I., Okada N. Coordinately Co-opted Multiple Transposable Elements Constitute an Enhancer for *wnt5a* Expression in the Mammalian Secondary Palate // *PLoS Genet.* 2016. Vol. 12. P. e1006380.
45. Derrien T., Johnson R., Bussotti G., Tanzer A., Djebali S., Tilgner H., Guernec G., Martin D., Merkel A., Knowles D. G., et al. The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression // *Genome Res.* 2012. Vol. 22. P. 1775–1789.
46. Pennisi E. ENCODE project writes eulogy for junk DNA // *Science.* 2012. Vol. 337. P. 1159.
47. Fang S., Zhang L., Guo J., Niu Y., Wu Y., Li H., Zhao L., Li X., Teng X., Sun X., Sun L., Zhang M. Q., Chen R., Zhao Y. NONCODEV5: a comprehensive annotation database for long non-coding RNAs // *Nucleic Acids Res.* 2018. Vol. 46, D1. P. D308-D314.
48. Liu S.J., Nowakowski T. J., Pollen A. A., Lui J. H., Horlbeck M. A., Attenello F. J., He D., Weissman J. S., Kriegstein A. R., Diaz A. A., Lim D.A., et al. Single-cell analysis of long non-coding RNAs in the developing human neocortex // *Genome Biol.* 2016. Vol. 17.P. 67.
49. Kelley D., Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs // *Genome Biol.* 2012. Vol. 13. P. R107.
50. Kapusta A., Kronenberg Z., Lynch V. J., Zhuo X., Ramsay L., Bourque G., Yandell M., Feschotte C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs // *PLoS Genetics* 9. 2013. P. e1003470.
51. Werner M. A., Sullivan R. N., Shah R. D., Nadadur A. T., Grzybowski V., Galat I. P., Moskowitz A. J. Ruthenburg. Chromatin-enriched lncRNAs can act as cell-type specific activators of proximal gene transcription // *Nat. Struct. Mol. Biol.* 2017. Vol. 24. P. 596–603.
52. Wang X., Ai G., Zhang C., Cui L., Wang J., Li H., Zhang J., Ye Z. Expression and diversification analysis reveals transposable elements play important roles in the origin of Lycopersicon-specific lncRNAs in tomato // *New Phytol.* 2016. Vol. 209. P. 1442–1455.
53. Faulkner G. J., Kimura Y., Daub C. O., Wani S., Plessy C., Irvine K. M., Schroder K., Cloonan N., Steptoe A. L., Lassmann T., Waki K., Hornig N., Arakawa T., Takahashi H., Kawai J., Forrest A. R. R., Suzuki H., Hayashizaki Y., Hume D. A., Orlando V., Grimmond S. M., Carninci P. The regulated retrotransposon transcriptome of mammalian cells // *Nat. Genet.* 2009. Vol. 41. P. 563–571.
54. Peaston A. E., Evsikov A. V., Graber J. H., de Vries W. N., Holbrook A. E., Solter D., Knowles B. B. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos // *Dev Cell.* 2004. Vol. 7. P. 597–606.
55. Macfarlan T. S., Gifford W. D., Driscoll S., Lettieri K., Rowe H. M., Bonanomi D., Firth A., Singer O., Trono D., Pfaff S. L. Embryonic stem cell potency fluctuates with endogenous retrovirus activity // *Nature.* 2012. Vol. 487. P. 57–63.
56. Ge S. X. Exploratory bioinformatics investigation reveals importance of “junk” DNA in early embryo development // *BMC Genom.* 2017. Vol. 18. P. 200–489.
57. Flemr M., Malik R., Franke V., Nejeplinska J., Sedlacek R., Vlahovicek K., Svoboda P. A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes // *Cell.* 2013. Vol. 155. P. 807–816.
58. Fuentes D. R., Swigut T., Wysocka J. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation // *Elife.* 2018. Vol.7. P. e35989.
59. Yasuhiko Y., Hirabayashi Y., Ono R. LTRs of Endogenous Retroviruses as a Source of Tbx6 Binding Sites // *Front Chem.* 2017. Vol. 5.P. 34.

60. Chuong E., Rumi M. A., Soares M. J., Baker J. C. Endogenous retroviruses function as species-specific enhancer elements in the placenta // *Nat. Genet.* 2013. Vol. 45. P. 325–329.
61. Trizzino M., Park Y., Holsbach-Beltrame M., Aracena K., Mika K., Caliskan M., Perry G. H., Lynch V. J., Brown C. D. Transposable elements are the primary source of novelty in primate gene regulation // *Genome Res.* 2017. Vol. 27. P. 1623–1633.
62. Chuong E.B., Elde N.C., Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses // *Science.* 2016. Vol. 351(6277). P.1083–1087.
63. Davis M. P., Carrieri C., Saini H. K., van Dongen S., Leonardi T., Bussotti G., Monahan J. M., Auchynnikava T., Bitetti A., Rappsilber J., Allshire R. C., Shkumatava A., O'Carroll D., Enright A. J. Transposon-driven transcription is a conserved feature of vertebrate spermatogenesis and transcript evolution // *EMBO Rep.* 2017. Vol. 18. P. 1231–1247.
64. Chishima T., Iwakiri J., Hamada M. Identification of Transposable Elements Contributing to Tissue-Specific Expression of Long Non-Coding RNAs // *Genes (Basel).* 2018. Vol.9. P. E23.
65. Hu T., Pi W., Zhu X., Yu M., Ha H., Shi H., Choi J. H., Tuan D. Long non-coding RNAs transcribed by ERV-9 LTR retrotransposon act in cis to modulate long-range LTR enhancer function // *Nucleic Acids Res.* 2017. Vol.45. P.4479-4492.
66. Muotri A.R. Chu V. T., Marchetto M. C., Deng W., Moran J. V., Gage F. H. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition // *Nature.* 2005. Vol. 435. P. 903–910.
67. Upton K. R., Gerhardt D. J., Jesuadian J. S., Richardson S. R., Sánchez-Luque F. J., Bodea G. O., Ewing A. D., Salvador-Palomeque C., van der Knaap M. S., Brennan P. M., Vanderver A., Faulkner G. J. Ubiquitous L1 mosaicism in hippocampal neurons // *Cell.* 2015. Vol. 161. P. 228–239.
68. Newman A. G., Bessa P., Tarabykin V., Singh P. B. Activity-DEpendent Transposition // *EMBO Rep.* 2017. Vol. 18. P. 346–348.
69. Tang S.-J. A Model of DNA Repeat-Assembled Mitotic Chromosomal Skeleton // *Genes.* 2011. Vol. 2. P. 661–670.
70. Tang S. J. A Model of Repetitive-DNA-Organized Chromatin Network of Interphase Chromosomes // *Genes.* 2012. Vol. 3. P. 167–175.
71. Cournac A., Koszul R., Mozziconacci J. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements // *Nucleic Acids Res.* 2016. Vol. 44. P. 245–255.
72. Nishikawa J., Ohshima T. Selective association between nucleosomes with identical DNA sequences // *Nucleic Acids Res.* 2013. Vol. 41. P. 1544–1554.
73. Tang S.-J. The R-Operon: A Model of Repetitive DNA-Organized Transcriptional Compartmentation of Eukaryotic Chromosomes for Coordinated Gene Expression// *Genes (Basel).* 2016. Vol. 7. P. E16.
74. Patel B., Kang Y., Cui K., Litt M., Riberio M. S., et al. Aberrant TAL1 activation is mediated by an interchromosomal interaction in human T-cell acute lymphoblastic leukemia // *Leukemia.* 2014. Vol. 28. P. 349–361.
75. Markenscoff-Papadimitriou E., Allen W. E., Colquitt B. M., Goh T., Murphy K. K., et al. Enhancer interaction networks as a means for singular olfactory receptor expression // *Cell.* 2014. Vol. 159. P. 543–557.
76. Williams A., Spilianakis C. G., Flavell R. A. Interchromosomal association and gene regulation in trans // *Trends Genet.* 2010. Vol. 26. P. 188–197.
77. Li G., Ruan X., Auerbach R., Sandhu K., Zheng M., Wang P., Poh H., Goh Y., Lim J., Zhang J. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation // *Cell.* 2012. Vol. 148. P. 84–98.
78. LoM. Y., Rival-Gervier S., Pasceri P., Ellis J. Rapid transcriptional pulsing dynamics of high expressing retroviral transgenes in embryonic stem cells // *PLoS One.* 2012. Vol. 7.P. e37130.
79. Schlesinger S., Meshorer E., Goff S. P. Asynchronous transcriptional silencing of individual retroviral genomes in embryonic cells // *Retrovirology.* 2014. Vol. 11. P. 31.