We are IntechOpen, the first native scientific publisher of Open Access books

3,350
Open access books available

108,000
International authors and editors

1.7 M
Downloads

151
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Influence of Endogenous Viral Sequences on Gene Expression

Kozue Sofuku and Tomoyuki Honda

Abstract

Endogenous viral elements (EVEs) are the heritable sequences present in eukaryotic genomes that have originated from viral nucleotide sequences. EVEs are subdivided into two groups, according to the presence or absence of long terminal repeats (LTRs). EVEs with LTRs are called endogenous retroviruses (ERVs), and they account for approximately 8% of the human genome. EVEs without LTRs seem to be related to non-reverse-transcribing RNA and DNA viruses, and recent studies have revealed that numerous vertebrate genomes contain these non-LTR EVEs. Such EVEs are proposed to play essential roles in gene expression. EVEs can regulate gene expression as cis-regulatory DNA and RNA elements. EVE-derived non-coding RNAs and/or proteins can also influence cell transcriptomes in trans. To maintain cell integrity, cells epigenetically silence the expression of most EVEs, making these elements generally biochemically inert. These epigenetic alterations around the EVE loci can also affect host transcriptomes. Here, we highlight the current knowledge available on the regulatory activities of ERVs and non-retroviral EVEs, especially the EVEs derived from bornaviruses, which are known as endogenous bornavirus-like elements (EBLs). Better knowledge of this area will improve our understanding of gene regulation and also the co-evolution of viruses and their hosts.

Keywords: endogenous viral sequences, retroviruses, bornavirus, long terminal repeats, co-evolution, genome

1. Introduction

Various viruses appear to have left heritable sequences originated from viral nucleotide sequences, called endogenous viral elements (EVEs), in eukaryotic genomes. EVEs are distinguished by the presence or absence of long terminal repeats (LTRs). EVEs with LTRs
are called endogenous retroviruses (ERVs). The LTRs contain cis-regulatory sequences and RNA polymerase II (Pol II) promoters [1]. ERVs are formed by the integration of ancient retroviruses into the host genome during infection, and they account for around 8% of the human genome contents. Some ERV-derived genes that have been co-opted by the host play essential roles in biological processes, such as placentation in humans [2, 3]. On the other hand, recent studies have revealed that numerous vertebrate genomes also contain non-LTR EVEs, EVEs that have no LTRs [4–6]. Among these non-LTR EVEs, the bornavirus-derived EVEs (endogenous bornavirus-like elements (EBLs)), which have been relatively well studied, have provided clues about the biological significance of non-LTR EVEs in mammals [4, 7–11]. EBLs are the DNA sequences in vertebrate genomes (i.e., primates, rodents, and afrotherians) that are formed by the long interspersed nuclear element-1 (LINE-1)-mediated integration of viral sequences of an ancient non-retroviral RNA virus, bornavirus [4]. LINE-1, a host retrotransponson, encodes two proteins, ORF1p and ORF2p, which form LINE-1 ribonucleoprotein (RNP) together with LINE-1 RNA [12, 13]. ORF2p is known as endonuclease and reverse transcriptase in the LINE-1 retrotransposition, which is also used for retrotransposition of viral mRNAs of non-retroviral RNA viruses, thereby producing non-LTR EVEs. EBLs derived from the N, M, G, and L genes of bornaviruses, which are designated as EBLN, EBLM, EBLG, and EBLL, respectively, have been reported so far [14]. Although EBLs do not contain any cognate promoter sequences derived from bornavirus sequences, some EBLs are thought to influence gene expression.

EVEs use various mechanisms to regulate gene expression. First, genomic EVEs can regulate gene expression as cis-regulatory DNA elements. Second, EVEs produce non-coding RNAs and/or proteins that influence nearby genes and/or the global transcriptome in trans. Third, alterations in the epigenetic environment around the EVEs can also affect the transcriptome. In this review, we provide a brief overview of the regulatory activities (e.g., promoter activity and epigenetic regulation) of ERVs and EBLs in the context of gene expression regulation.

2. The influence of ERVs on gene expression

The exogenous retroviral genome contains the following genes: gag, which encodes the gene encoding retroviral structural proteins, pol, which encodes reverse transcriptase, protease, ribonuclease and integrase, and env, which encodes the envelope protein. The retrovirus viral genome also contains a primer-binding site (pbs) and the packaging signal (Ψ), both of which are important to the viral life cycle (Figure 1A). The reverse transcriptase encoded in the pol gene synthesizes viral DNA (proviral DNA) from the viral RNA, and the proviral DNA is then inserted into the host’s genome which, when inherited in germ-line cells, become ERVs (Figure 1B). ERVs have been co-opted with the host and play essential roles in gene expression (Figure 2).
2.1. Gene regulation by ERVs as regulatory DNAs

The LTRs of human ERVs (HERVs) have strong Pol II regulatory sequences [15, 16] and contain abundant transcription factor binding sites that function as promoters for HERV expression [17]. Although the full-length HERV is considered to have two LTRs, up to 85% of HERVs have undergone recombinatorial deletion [18], making most HERV loci solo LTRs. Solo LTRs can still serve as promoters in both the sense and antisense orientations and influence gene expression [19, 20].
For example, **IL2RB** and **NOS3** are genes whose expression in the placenta is solely related to the presence of LTR promoters [19]. Stem cell-specific LTR-derived promoters, such as mouse ERVK and human ERV1, control the expression of nuclear transcripts [21], whose expression is associated with maintenance of pluripotency. MER39 (an ERV1 class member) constitutes the promoter for human endometrial Prl [22]. MER41, another HERV, works as a *cis*-regulatory sequence of AIM2 (a non-self DNA sensor), thereby regulating inflammatory responses [23]. The ERV-9 LTR is located near the 5′ end of the locus control region, around 40–70 kb upstream of the human fetal γ- and adult β-globin genes. LTR deletion was found to drastically suppress the β-globin gene and reactivate the γ-globin gene through a competitive mechanism involving globin gene switching [24]. Some lineage-specific ERVs, such as LTR19B and MER41, have dispersed numerous IFN-inducible enhancers in human genomes, thereby shaping the evolution of the transcriptional network underlying the interferon (IFN) response [23]. The expression of very long intergenic RNAs (vlincRNAs), which also control pluripotency, is driven by HERV LTR [25], suggesting a role for HERV LTRs in regulating the expression of not only protein-coding genes but also long non-coding RNAs (IncRNAs) [26].
2.2. Gene regulation by ERV proteins

The expression products of HERVs can also affect the physiological functioning and development of the host’s tissues. For example, HERV-W (ERVWE1), HERV-FRD, and ERV-3 are three HERVs whose intact env genes are expressed as proteins in the human placenta [27–30]. HERV proteins play important roles in the proper formation of the placenta and are involved in the suppression of fetal tissue rejection [27, 31, 32]. The transmembrane envelope proteins of HERV-K, which modulate the expression of numerous cytokines, provide an example of gene expression regulation by a HERV protein [33]. HERVs may also be linked to a strategy used for inhibiting exogenous virus replication. For example, Friend virus susceptibility 1 (Fv1), a mouse gene that originated from the gag gene of an ancient retrovirus, is known to restrict murine leukemia virus (MLV) at a stage after entry but before integration and formation of the provirus, thereby inhibiting viral replication [34, 35].

2.3. Gene regulation by HERV-driven lncRNAs

lincRNA-RoR is a large intergenic non-coding RNA driven by HERV-H [36]. lincRNA-RoR modulates reprogramming and is indeed expressed at much higher levels in the embryonic stem cell line, H1-hESC, and human-induced pluripotent stem cells than in any other tissue or cell line [36, 37]. Knockdown of lincRNA-RoR affects the expression of other stem cell factors such as KLF4, SOX2, and NANOG [38, 39], resulting in an exit from pluripotent state [37]. Together with vlincRNAs [25], HERV-driven lncRNAs can influence the transcriptome of the genes involved in pluripotency.

2.4. Gene regulation by epigenetic modification of ERVs

In addition to the abovementioned roles, LTRs are important sites for epigenetic modifications that restrict HERV in the human genome. DNA methylation, which is carried out by DNA methyltransferases, histone methylation, and histone deacetylation are the major host mechanisms used for gene silencing [40, 41]. Indeed, HERVs are heavily methylated in normal tissues [42]. By contrast, histone deacetylation alone is not sufficient to repress HERV expression. Rather, histone deacetylation in combination with other epigenetic modifications, particularly DNA methylation, is required for sufficient silencing of HERVs [43]. Furthermore, histone demethylation, which is carried out by lysine-specific histone demethylases (KDMs), also silences HERV expression [44, 45]. All these epigenetic alterations to ERV loci can affect the expression of nearby genes. For example, MuERV-L/MERVL, a mouse ERV, is repressed by a KDM1A-mediated epigenetic modification [45]. Some zygotic genome activation (ZGA) genes use an LTR of MERVL as a promoter or contain an MERVL element within 5 kb of their transcriptional start sites [45]. These ERV-linked ZGA genes become de-repressed in KDM1A mutant cells, which coincide with an expanded cell fate potential [45]. Thus, KDM1A recruitment to the MERVL LTRs seems to alter the chromatin structure around the loci, which in turn suppresses the expression of ERV-linked ZGA genes during early mammalian embryonic development.
2.5. Possible links between ERVs and human diseases

The recent studies on ERVs have revealed possible interactions between ERVs and their hosts with the potential to contribute to the development of diseases such as cancer and neurologic diseases. For example, the HERV expression is upregulated in various types of cancers [46–48]. Many HERV LTR regions, such as LTR10 and MER61, have a near-perfect p53 DNA binding site [49]. The tumor suppressor protein p53 is a sequence-specific transcription factor, which regulates genes of diverse biological pathways [50]. Thus, ERVs may regulate carcinogenesis via the p53 pathway. CSF1R gene, an oncogene, is activated by a demethylated MaLR LTR [51]. LTR-driven CSF1R is expressed aberrantly in anaplastic large cell lymphoma [51], suggesting that ERV LTRs may also directly contribute to tumor growth via activation of oncogenes. HERVs have also involved in neurological and psychiatric diseases. For example, the expression levels of HERV-H are significantly higher in patients with attention deficit hyperactivity disorder (ADHD) compared with healthy controls [52]. Furthermore, the HERV-W env mRNA expression is selectively upregulated in brain tissue from patients with multiple sclerosis compared with controls [53]. Although links between the upregulation of ERVs and these diseases are reported, the contribution of upregulated ERVs to the disease development is still unclear and further studies are clearly required for demonstrating it.

3. The influence of nonretroviral EVEs on gene expression

EBLs are the only nonretroviral RNA virus-derived EVEs found in the human genome. EBLs seem to be generated from bornavirus mRNA in a LINE1-dependent manner (Figure 1C and D). Thus, they are a unique form of a processed pseudogene, which is derived from the sequences of an exogenous virus but not endogenous sequences, and they evidence the mechanism of retrotransposon-mediated RNA-to-DNA information flow from the virus to the host [4]. In the human genome, seven EBLNs (hsEBLN-1 to hsEBLN-7) and one EBLG have been identified to date [4–6]. All seven hsEBLNs are expressed as RNAs in at least one tissue, suggesting the possibility of a biological function for these EBLs [9].

3.1. Gene regulation by EBLN RNAs

hsEBLN-1 is one of the most studied EBLs in the human genome. Because no natural selection of hsEBLN-1 and its orthologues is detected [54], hsEBLN-1 is thought to function as a DNA element or non-coding RNA, or even to have lost its function (Figure 3). He et al. reported that 1067 and 2004 genes are up- and downregulated, respectively, after knockdown of hsEBLN-1 RNA in human oligodendroglia cells [55]. The top 10 most upregulated genes were PI3, RND3, BLZF1, SOD2, EPGN, SBSN, INSIG1, OSMR, CREB3L2, and MSMO1, and the top 10 most downregulated genes were KRTAP2–4, FLRT2, DIDO1, FAT4, ESCO2, ZNF804A, SUV420H1, ZC3H4, YAE1D1, and NCOA5. Gene ontology revealed that hsEBLN-1 may regulate the expression of genes related to the cell cycle, the mitogen-activated protein kinase pathway, p53 signaling, and apoptosis [55].
Unlike ERVs, EBLs are thought not to be transposable themselves. Nevertheless, the hsEBLN-1 locus is silenced by several epigenetic blocks, dominantly histone deacetylation and DNA methylation, similar to the case of human immunodeficiency virus (HIV) provirus silencing [9, 56, 57]. This contrasts with the silencing mechanism of ERVs because, as described above, DNA methylation but not histone deacetylation plays a major role [58]. Thus, the silencing mechanisms for the hsEBLN-1 locus might be more similar to those of exogenous retroviruses than to those of ERVs. This epigenetic alteration around hsEBLN integration may affect the epigenetic status of its neighboring loci and, consequently, the expression of nearby genes. Histone deacetylase (HDAC) inhibitor treatment did not affect transcription of the COMMD3 gene in mouse and rat cells, which have no EBLN sequence at the locus syntenic to the hsEBLN-1 locus, whereas the treatment led to decreased transcription of the COMMD3 gene in human and monkey cells, which have the EBLN sequence at the locus. COMMD3 belongs to the copper metabolism gene MURR1 domain-containing (COMMD) family. COMMD proteins have a structurally conserved COMM domain, and they are all able to interact with different NF-κB subunits [59]. Because one of the central roles of NF-κB is induction of proinflammatory mediators like cytokines,

Figure 3. Influence of EBLs on gene expression. (1) An EBL-derived lncRNA regulates the expression of a neighboring gene. piRNAs derived from EBLs may have the potential to provide antiviral responses against bornaviruses. (2) An EBL protein inhibits bornavirus transcription. Likewise, some EBL proteins may possibly regulate host gene expression. (3) Alteration of the epigenetic environment to restrain the EBL expression may influence the expression of neighboring genes. The ochre broken line circle (Ac) indicates deacetylation of the EBL locus.
chemokines, and adhesion molecules, EBLN-1 may regulate immune responses indirectly through the COMMD3-NF-κB pathway [59, 60]. Moreover, suppression of the hsEBLN-1 RNA induced by HDAC inhibitor treatment using siRNA against hsEBLN-1 RNA eliminated the HDAC inhibitor-induced downregulation of COMMD3 gene expression. Thus, hsEBLN-1 RNA may function as a lncRNA that scaffolds transcriptional repressors of the COMMD3 gene around the locus, thereby downregulating its expression.

Several EBLN-derived small RNAs in mouse and rat are annotated as PIWI-interacting RNAs (piRNAs) in the GenBank database [61]. piRNAs are 25–33 nucleotides in length, are found in diverse organisms such as flies, fish, and mammals [62], and protect germ-line cells from transposons [62]. piRNA clusters are transcribed as long single-stranded precursor RNAs derived from the piRNA clusters in the host genome, which are further processed into small mature piRNAs. Mature piRNAs guide Argonaute proteins, such as PIWI and MIWI proteins, to complementary target sequences. Argonaute proteins cleave the target RNAs, suppressing their expression. piRNAs are also known to epigenetically silence the target gene loci. All piRNAs derived from EBLNs are antisense relative to the proposed ancient bornaviral nucleoprotein mRNA [61]. These observations offer a possible role for the EBLN-derived piRNA-like RNAs in interfering with bornavirus mRNAs [61].

3.2. Gene regulation by EBLN proteins

Among the human EBLNs, hsEBLN-1 and hsEBLN-2 have maintained long open reading frames with the potential to code for proteins of 366 and 225 amino acids, respectively. Indeed, some studies have reported that hsEBLN-1 proteins were detected in particular cell lines [63]. Moreover, Kobayashi et al. reported that EBLNs encode functional proteins in afrotherians [10]. Therefore, it is still possible that EBLN proteins regulate gene expression in trans. Furthermore, EBLNs may potentially inhibit the replication of related exogenous viruses, similarly to certain ERVs. EBLN from the thirteen-lined ground squirrel (Ictidomys tridecemlineatus) genome, named itEBLN, is associated with bornavirus RNPs and inhibits bornavirus polymerase activity [7].

4. Conclusions

The researches on gene regulation by EVEs have provided us with important knowledge about the evolution of regulatory sequences in the genome [5, 64]. Although integrated viral sequences are usually eliminated from the host genome, some eventually reach fixation and form EVEs. Such EVEs are not merely genetic parasites; rather, they introduce useful genetic novelties to the genome. In this article, we briefly reviewed two types of EVEs, ERVs and the non-LTR EVEs, EBLs. ERVs provide novel regulatory sequences and sites for epigenetic regulation. Transcripts derived from ERVs can also function as lncRNAs or protein-coding mRNAs, which may regulate gene expression. In particular, ERV-related transcripts are often associated with pluripotency. EBLs might also function as regulatory DNA elements such as promoters and enhancers. They are transcribed in one tissue at least, suggesting that
EBL transcripts may function as lncRNAs or protein-coding mRNAs. Consistently, we have shown the evidence for the roles of EBL transcripts as lncRNA molecules in gene expression. In particular, several EBLs are associated with antiviral responses against related viruses. Additionally, both ERVs and EBLs regulate not only host gene expression, but related viral gene expression also. Further extensive studies on EVEs will augment our understanding of their biological significance in gene expression and their involvement in the co-evolution of viruses and mammals.

Acknowledgements

The preparation of this article was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI grant numbers JP24115709, JP25115508, JP25860336 (TH), MEXT KAKENHI grant number 15K08496 (TH), and grants from the Takeda Science Foundation, Senri Life Science Foundation, and The Shimizu Foundation for Immunology and Neuroscience grant for 2015 (TH). We thank Sandra Cheesman for editing a draft of this manuscript.

Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>EVE</td>
<td>endogenous viral element</td>
</tr>
<tr>
<td>LTR</td>
<td>long terminal repeat</td>
</tr>
<tr>
<td>ERV</td>
<td>endogenous retrovirus</td>
</tr>
<tr>
<td>Pol II</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>EBL</td>
<td>endogenous bornavirus-like element</td>
</tr>
<tr>
<td>LINE-1</td>
<td>long interspersed nuclear element-1</td>
</tr>
<tr>
<td>RNP</td>
<td>ribonucleoprotein complex</td>
</tr>
<tr>
<td>HERV</td>
<td>human endogenous retrovirus</td>
</tr>
<tr>
<td>lncRNA</td>
<td>long non-coding RNA</td>
</tr>
<tr>
<td>MLV</td>
<td>murine leukemia virus</td>
</tr>
<tr>
<td>KDM</td>
<td>lysine-specific histone demethylase</td>
</tr>
<tr>
<td>ZGA</td>
<td>zygotic genome activation</td>
</tr>
<tr>
<td>ADHD</td>
<td>attention-deficit hyperactivity disorder</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
</tbody>
</table>
HDAC  histone deacetylase
siRNA  small interfering RNA
piRNA  PIWI-interacting RNA

Author details
Kozue Sofuku and Tomoyuki Honda*
*Address all correspondence to: thonda@virus.med.osaka-u.ac.jp
Division of Virology, Department of Microbiology and Immunology, Graduate School of Medicine, Osaka University, Osaka, Japan

References


[34] Stoye JP. Fv1, the mouse retrovirus resistance gene. Revue Scientifique et Technique. 1998;17:269-277


