Seminars in Cancer Biology 20 (2010) 246–253

Contents lists available at ScienceDirect

Seminars in Cancer Biology

journal homepage: www.elsevier.com/locate/semcancer

Review

Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer

M.T. Romanish[a,b], C.J. Cohen[a,b], D.L. Mager[a,b,*]

[a] Terry Fox Laboratory, British Columbia Cancer Agency, 675 West 10th Avenue, Vancouver, BC, Canada V5Z 1L3
[b] Dept. of Medical Genetics, Univ. of British Columbia, Vancouver, BC, Canada

ARTICLE INFO

Keywords:
Endogenous retrovirus
Cancer
Epigenetics
Genome instability
DNA methylation

ABSTRACT

Malignancy results from a complex combination of genetic and epigenetic changes, the full effects of which are still largely unknown. Here we summarize current knowledge of the origin, retrotranspositional activity, epigenetic state, and transcription of human endogenous retroviruses (HERVs), and then discuss the potential effects of their deregulation in cancer. Evidence suggests that cancer-associated epigenetic changes most likely underlie potential HERV-mediated effects on genome and transcriptome instability and may play a role in malignancy. Despite our currently limited understanding of the importance of HERVs or other transposable elements in cancer development, we believe that the emerging era of high-throughput sequencing of cancer genomes, epigenomes, and transcriptomes will provide unprecedented opportunities to investigate these roles in the future.

1. Introduction

Remarkably, almost half of the human genome comprises transposable, or transposed, elements (TEs), which fall into three main categories: LTR (long terminal repeat) retrotransposons, non-LTR retrotransposons (LINEs and SINES) and DNA transposons [1]. The LTR retrotransposons comprise endogenous retroviruses (ERVs) and other sequences with LTR structures, referred to hereafter as human endogenous retroviruses (HERVs), and their activity throughout primate evolution has resulted in ~8% of the human genome as identifiably HERV derived [1]. Although the structure, function, and impact of HERVs on the human genome has been studied in detail, their potential involvement in malignancy is only beginning to be appreciated. Here we compare the expression, promoter activity, and epigenetic regulation of HERVs in normal cells and during malignancy and discuss possible mechanisms by which these elements could contribute to oncogenesis.

2. Origin and structure of HERVs

For several vertebrate species, the concurrent existence of endogenous and exogenous forms of very similar retroviruses provides evidence that ERVs originated from germ cell infections by exogenous retroviruses during the course of evolution [2,3]. The time point at which a retrovirus first entered the genome will subsequently determine in which species that ERV family is present today. Ancient elements, such as members of the HERV-L family, inserted more than 80 million years ago and are common to most mammals, whereas younger elements such as HERV-K are primate- or even human-specific [4]. Similar to proviruses of their exogenous counterparts, a typical HERV is flanked by the transcriptional regulatory signal-containing LTRs that bound the viral genes. Invariably, an autonomous HERV encodes the *gag*, *pro*, and *pol* genes, and the occasional existence of an often-mutated *env* gene hearkens to a time of extracellular 'life'. Indeed, adaptation to an intracellular life-cycle as retrotransposons, as has been shown for some mouse ERV families [5,6], may be a prerequisite for amplification to significant copy numbers. However, the most recent, human-specific HERV-K insertions have reportedly resulted from infection by unique, yet related, HERV-K viruses, rather than by intracellular retrotransposition [7]. There also exist families of LTR retrotransposons in the genomes of humans and other mammals that may never have encoded viral genes [8] and which presumably amplified using the retrotransposition machinery of autonomous elements.

Exogenous retroviruses are classified based on virion structure and sequence [9], whereas sequence relationships alone are used to group HERVs into three general classes: class I (gamma

Abbreviations: TE, transposable element; LINE, long interspersed nuclear element; SINE, short interspersed nuclear element; (H)ERV, (human) endogenous retrovirus; LTR, long terminal repeat; HML, human mammary tumor virus-like; ORF, open reading frame; DNMT, DNA methyltransferase.

* Corresponding author at: Terry Fox Laboratory, BC Cancer Agency, 675 West 10th Avenue, Vancouver, BC, Canada V5Z 1L3. Tel.: +1 604 675 8139; fax: +1 604 877 0712.
E-mail addresses: mromanis@bccrc.ca (M.T. Romanish), ccohen@bccrc.ca (C.J. Cohen), dmager@bccrc.ca (D.L. Mager).

1044-579X/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.semcancer.2010.05.005

retroviruses), class II (beta retroviruses), and class III (spuma retroviruses) [10]. There are between 50 and 200 different families of HERVs depending on the criteria used to delineate their interrelatedness [10–12]. The current convention of naming HERVs utilizes the single letter amino acid code corresponding to the tRNA primer that is used to reverse transcribe its genome [13], but efforts are ongoing to establish an improved nomenclature system [14]. Class I HERVs are represented by the most families and largest genomic fraction [10,15], and in fact outnumber class II elements by a factor of 10 [1]; a trend that is reversed in the sequenced mouse genome [16]. However, the class II group contains the potentially active HERV-K elements, along with several other older families [17]. Lastly, the class III elements are the oldest recognizable HERV elements, and may represent the single largest class depending on whether or not the non-autonomous mammalian apparent LTR retrotransposons (MaLRs) are included [1]. For further general background on HERVs, interested readers can consult several recent reviews [4,17–19].

3. Lack of HERV mutagenic activity

Unlike in mouse [20], ERVs in human no longer pose a significant threat as insertional mutagens, since only one family, HERV-K, is thought to encode copies still capable of retrotransposition, and no disease-causing insertions have been reported. Indeed, nearly 90% of HERV elements exist as recombined solitary LTRs [1,10], and the remainder have accumulated inactivating mutations over the course of million years of evolution. In sharp contrast, LINEs and SINEs exhibit considerable levels of insertional polymorphism in human populations, and some new insertions of these elements cause disease [21,22]. Current estimates indicate that one in approximately 25 births experience a novel L1 or *Alu* integration [21], but only 13 HERV-K elements exhibiting variable population frequencies have been identified in human [18,23]. Nonetheless, the potential reactivation of HERV-K elements resulting in mutagenic retrotranspositional events during malignancy cannot be discounted and is discussed further in Section 8.

4. HERV expression and LTR promoter activity in normal cells

Dozens of intact viral open reading frames (ORFs) exist in the human genome and represent a variety of HERV families [15,24]. Coding competent *gag* or *pol* genes are the most common, and these numbers increase almost two-fold when taking into account ORFs that can be corrected by a single nucleotide change [24]. Furthermore, 16 coding competent *env* genes have been discovered in the human genome [15]. The conservation of these ORFs suggests their possible domestication, i.e. they may convey a beneficial function in the host. While this is uncertain in the majority of cases, domestication of two human *env* genes has been demonstrated and these have a likely role in syncytiotrophoblast formation [25]. Strikingly, similar but independent domestication events of ERV *env* genes have occurred in mouse [26], sheep [27], and rabbits [28], suggesting that ERVs may have been central in the radiation of placental mammals.

Although a very small proportion of HERVs can potentially encode protein(s), a plethora of data indicate that these and other defective copies are still actively transcribed, pointing to a retained transcriptional regulatory function for many LTRs. Of particular interest is the tissue-specificity of such gene expression. One analysis of HERV-K (HML-2) expressed sequence tags (ESTs) found that stem cells, germ cells and neuronal cells are most permissive to LTR activity of this family [29]. Another report addressing overall class II *pol* expression by qRT-PCR identified high levels of transcription in brain, testis, kidney, fetal liver, and adrenal gland, but not uterus, placenta or muscle [30]. Furthermore, a microarray-based analysis of *pol* domain transcription for 20 different HERV families across a panel of human tissues revealed that a minority of the assayed families are broadly expressed [31]. In this study, rather, the majority of HERV families display distinct tissue tropism, with thyroid gland, skin, uterus, and cervix demonstrated as the most permissive tissues to HERV *pol* transcription [31]. Consistent with *pol* microarray data, HERV-E *env* transcripts are also detected in most tissues screened by RT-PCR, with the exception of heart, liver, lung, and muscle [32]. A similar pattern was also demonstrated for the 16 coding competent *env* genes in the human genome [15], with the exception of HERV-R *env*, which is expressed in all twenty tissues tested. Another study substantiated the expression dynamics observed for three of these *env* copies, and noted that the *env* gene expressed from a HERV-P element exhibits high expression in the brain, lung, testis, thymus, and uterus [33]. Unsurprisingly, those *env* genes with a demonstrated function in human placenta, HERV-W, -FRD, and -R, are most highly expressed in that tissue, but testis is the only tissue exhibiting expression of all 16 coding-competent *env* genes [15]. These data indicate that particular HERV LTRs express their associated retroviral genes in cell- or developmental-specific contexts, although the significance of this remains elusive. Moreover, in most cases it is unknown how many copies of a family are transcriptionally active. Unfortunately this question is difficult to address except for old HERV families that have diverged sufficiently for transcript sequences to be uniquely mapped to a particular genomic locus.

In addition to their native function as promoters of retroviral genes, some LTRs have been exapted as promoters of nearby cellular genes [34–37]. An initial estimate calculated a 0.7% frequency of LTR promoter adoption by cellular genes, which translates to approximately 200 expected examples [38]. This approximation is supported by bioinformatics analysis of CAGE and PET libraries [35] and our own more recent analysis in which we found 158 cases of LTRs overlapping with 10 or more ESTs [36]. Interestingly, LTR exaptation events fall into three main classes; those that specifically function to augment transcription of the associated gene in a particular tissue (which often occurs in the placenta), those that confer widespread non-specific transcription, and those that have become converted as the main gene promoter [36,39]. The nature of expression from LTR promoters is similar to the observed transcriptional activity from HERV families described above. Again it is uncertain whether low levels of broadly transcribed sequences truly represent functional events, or are merely a consequence of residual LTR promoter activity. Since particular HERV families present unique combinations of transcription factor binding sites, perhaps this renders them permissive to augmenting transcription in specific tissues or developmental contexts in some cases [40–43].

5. Epigenetic control of ERV expression in normal tissues

We have discussed the known expression profiles of HERVs at both the RNA and protein levels. However, in most cases HERVs (and indeed most TEs) are transcriptionally silenced by epigenetic mechanisms. Transcriptionally active loci, whether of domesticated retroviral proteins or LTR-derived chimeric transcripts, are rare when compared to the total number of HERV elements. This fact is particularly pertinent when we consider the deregulation of this transcriptional repression during cancer development in Sections 6 and 7.

Studies in mice have shown DNA methylation of TEs normally equates to transcriptional repression, apart from a transitional period during early embryogenesis when a wave of global demethylation reactivates transcription of many ERVs [44,45]. Sev-