

Minireview

Endogenous retroviruses in the human genome sequence

David J Griffiths

Address: Wohl Virion Centre, Department of Immunology and Molecular Pathology, Windeyer Institute of Medical Sciences, University College London, Cleveland Street, London W1T 4JF, UK. E-mail: d.j.griffiths@ucl.ac.uk

Published: 5 June 2001

Genome Biology 2001, **2(6)**:reviews1017.1–1017.5

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/2/6/reviews/1017>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

Abstract

The human genome contains many endogenous retroviral sequences, and these have been suggested to play important roles in a number of physiological and pathological processes. Can the draft human genome sequences help us to define the role of these elements more closely?

One of the many striking findings to come from the sequencing of the human genome is that some 45% of our DNA is composed of transposable elements such as LINE and Alu retroelements and DNA transposons [1-3]. Around 8% of the genome is derived from sequences with similarity to infectious retroviruses, which can be easily recognized because all infectious retroviruses contain at least three genes, including *gag* (encoding structural proteins), *pol* (viral enzymes), and *env* (surface envelope proteins), as well as long terminal repeats (LTRs; see Figure 1). The existence of human endogenous retroviruses (HERVs) has been known for many years [4], but their abundance in the genome was not predicted by earlier studies. HERVs represent the remnants of ancestral retroviral infections that became fixed in the germline DNA. Subsequent retrotransposition events amplified these sequences to a high load within the genome. The drafts of the human genome have provided a wealth of information about the abundance and distribution of HERVs, and several new subtypes have been identified [5]. This sequence information can now be used for the design of novel experimental strategies to investigate the biological functions of HERVs. This article briefly reviews the evolution and abundance of HERVs and the available evidence for their function in both normal and pathological processes.

HERVs have been grouped into three broad classes - I, II and III - on the basis of sequence similarity to different genera of infectious retroviruses. Each class has a number of subgroups, many of which are named according to an older system of HERV nomenclature based on the specificity of the tRNA primer-binding site (Figure 1). Class I HERVs are

related to gammaretroviruses such as murine leukemia virus (MLV); class I includes HERV-W and HERV-H, among many other subgroups. Class II HERVs are related to betaretroviruses such as mouse mammary tumor virus and include several types of HERV-K element. Class III HERVs are distantly related to spumaretroviruses and include HERV-L and HERV-S.

Like other transposable elements, HERVs are thought to have played an important role in the evolution of mammalian genomes, and the human genome sequence has already been of use in phylogenetic studies of HERVs. By analyzing HERV integration sites, the evolution of these elements has been tracked through the primate lineage. Measurement of the divergence of LTR sequences has also been used as a 'molecular clock' to estimate the age of HERVs (given that the LTRs are identical at the time of integration) [5]. Class I and class III HERVs are the oldest groups and are present throughout the primate lineage, while class II includes HERVs that have been active most recently. Many class II loci are restricted to chimpanzees and humans and a few proviruses of the HERV-K(HLM-2) subgroup are human-specific [6], indicating that these viruses have been active within the last 5 million years.

Cellular functions of HERVs

Although HERVs have retained some similarity to their exogenous counterparts, they have acquired many mutations over the course of evolutionary time so that, with a few exceptions, they are now defective and incapable of producing

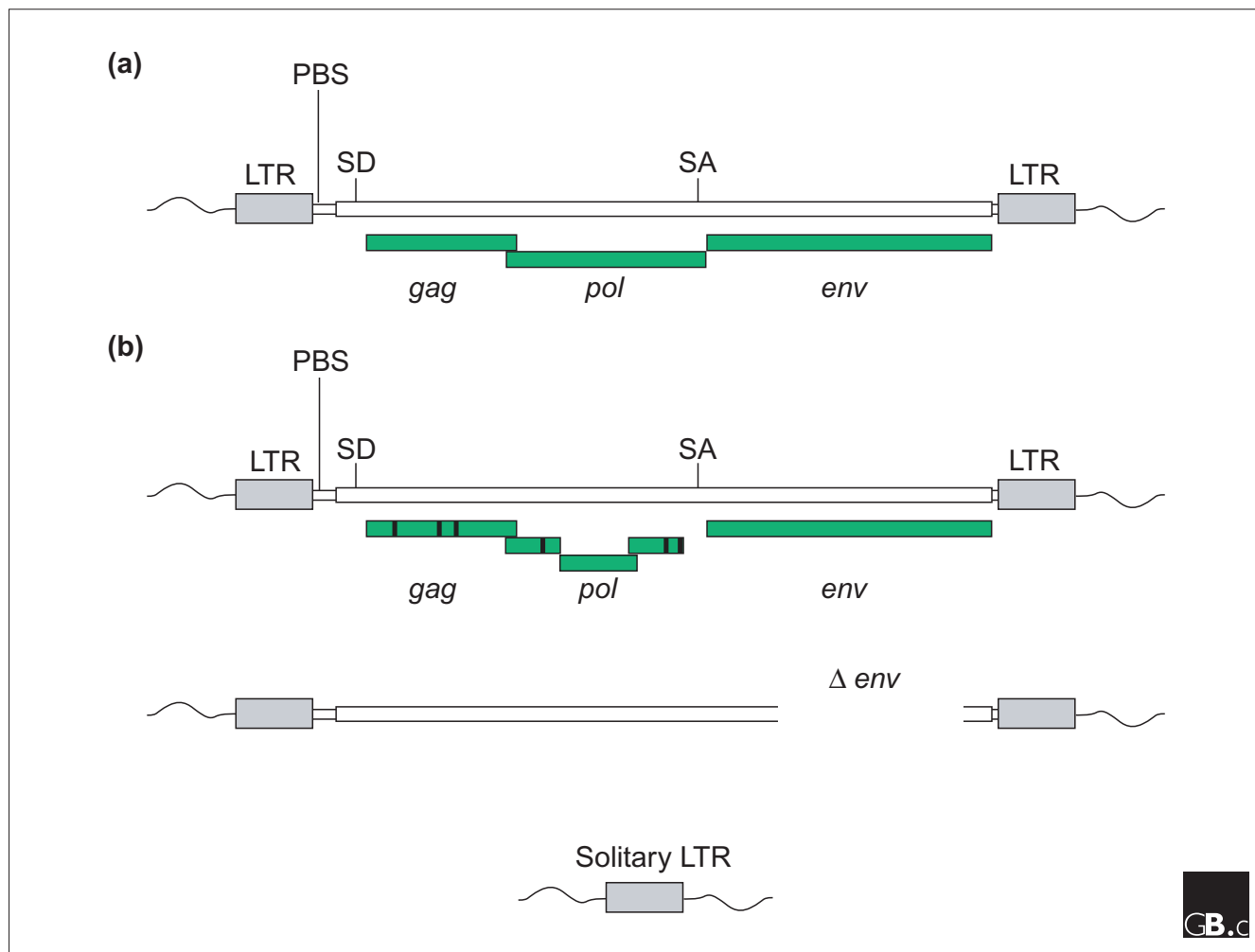


Figure 1

Structure of retroviral proviruses. **(a)** Infectious retroviruses have at least three genes: *gag*, which encodes the structural proteins of the viral core; *pol*, which encodes the viral enzymes, including reverse transcriptase; and *env*, which encodes the surface glycoproteins of the viral envelope. Viral protein expression is controlled by promoter and enhancer elements and polyadenylation signals in the long terminal repeats (LTRs), which are generated during reverse transcription. Other regulatory elements are also present in the viral genome, including splice donor (SD) and acceptor (SA) sites (for *env* expression) and a primer-binding site (PBS) for a specific tRNA molecule used to initiate reverse transcription. The tRNA specificity varies among different retroviruses and has been used to classify endogenous retroviruses in the human genome. **(b)** Human endogenous retroviruses (HERVs) have a similar structure to the proviruses of infectious retroviruses but typically contain many inactivating mutations including point mutations (dark bands), frameshifts and deletions (particularly in *env*). Frequently, the entire central portion has been lost by homologous recombination, leaving behind a 'solitary LTR'. Although almost all HERVs are defective, the LTRs may still be active, and transcription of HERVs is common particularly in fetal tissue and in inflammatory disease and cancer. In a few cases, coding competence has been retained for *env* even when adjacent viral genes are heavily mutated, suggesting that selective pressures have maintained these open reading frames because they serve a cellular function.

protein (Figure 1b). Analysis of the draft human genome has so far found only three HERV proviruses with complete open reading frames for *gag*, *pol* and *env* (the three essential viral genes) [1], and at least one of these HERVs is mutated at a critical residue in the reverse transcriptase domain of *pol* [7]. This is in contrast to the situation in some other species, such as pigs and mice, in which a few endogenous retroviruses have retained the capacity for infectious transmission [8]. Because of the activity of endogenous viruses in animals,

there remains a great deal of interest in identifying biologically functional HERVs, and specific candidates may be detected by further analysis of the human genome sequence.

The best example of a HERV with a known function is HERV-W. The envelope proteins of this HERV are thought to mediate fusion of trophoblasts, an essential step during formation of the placenta [9]. A role in membrane fusion is not surprising since this is the role of the viral Env protein during

retroviral infection following binding to a cell surface receptor. Interestingly, trophoblast fusion by HERV-W Env appears to be independent of a specific receptor molecule. A different HERV (ERV-3) had previously been suggested to provide the trophoblast fusion function but was later ruled out by the discovery of individuals who are homozygous for an inactivating mutation [10]. Sequence comparisons in different individuals may yet reveal such polymorphisms for HERV-W.

Another function proposed for HERVs is in determining resistance to viral infection. In mice, resistance to infection by MLV is controlled by a Gag-like protein encoded by *Fv1*, an otherwise defective endogenous retrovirus related to human HERV-L [11]. The molecular basis of this restriction is not yet known, but it has been suggested that the Fv1 protein interacts with the incoming core of the MLV viral particle in a dominant-negative manner, thereby inhibiting infection. Cell lines from other species, including humans, also have an intracellular restriction to MLV infection [12], but the mouse and human restrictions occur at different stages of viral entry. Although the inhibitory factor in humans has yet to be cloned, it is possible that endogenous retroviral Gag proteins may provide a general mechanism for controlling retroviral infection in mammals.

HERVs and disease

HERVs have frequently been proposed as etiological cofactors in chronic diseases such as cancer, autoimmunity and neurological disease [13]. Unfortunately, despite intense effort from many groups, there remains little direct evidence to support these claims, and moreover some studies have served only to muddy the waters for others. One particular difficulty has been picking out the coding-competent subset of HERVs from the large background 'noise' of defective elements. The clinical heterogeneity of many of the associated diseases, such as lupus erythematosus, rheumatoid arthritis and multiple sclerosis, has also been a problem, since HERVs may be involved in specific subtypes of a particular disease and such subtypes may not be recognized by current diagnostic criteria. The availability of the human genome sequence will facilitate the identification of those HERV loci most likely to be involved in disease. In addition, a deeper understanding of the genetic basis of these diseases should lead to a more precise definition of disease subtypes. In turn, this may clarify the part played by HERVs.

Much of the evidence that links HERVs to disease comes from the detection of expressed retroviral sequences in patient tissue by degenerately primed PCR. For example, HERV-W was first identified in a search for retroviruses in people with multiple sclerosis [14], and the same HERV was recently detected in both cerebrospinal fluid and brain tissue from patients with schizophrenia [15]. The significance of HERV RNA expression in studies such as these remains unclear, because disease causation cannot be

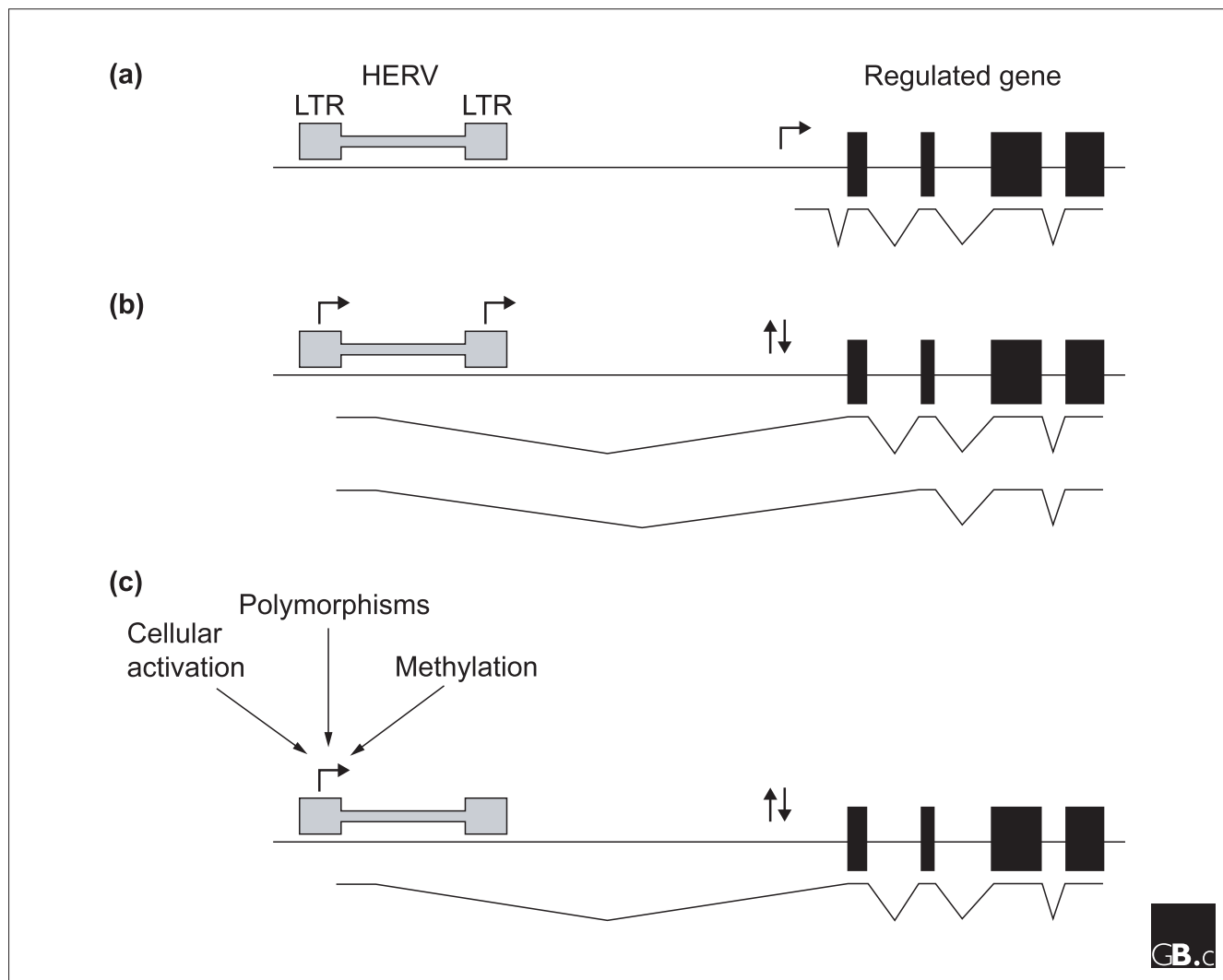
proved simply by the detection of virus expression, particularly for a ubiquitous sequence such as a HERV. In addition, although HERV RNA expression is known to be increased in several autoimmune diseases and cancers, there is usually activation of a range of class I and class II HERVs rather than expression of a single provirus. In general, further validation of disease association for HERVs has not been described. An exception is the HERV-K(HML-2) subgroup, which has been implicated in germ-cell tumors. As noted above, this subgroup includes the youngest and most active HERVs, and specific attention has focused on the ability of these elements to form virus-like particles in teratocarcinoma-derived cell lines [16]. Some of these particles are able to bud from the cell surface, although it is doubtful whether they are infectious. In addition, patients with germ-cell tumors frequently have antibodies to HERV-K Gag and Env proteins [17]. Recent work has shown that a regulatory protein produced by HERV-K(HML-2) can bind the transcription factor PLZF (promyelocytic leukemia zinc finger protein), which is required for spermatogenesis [18]. Impairment of spermatogenesis is associated with increased frequency of germ-cell tumors, and thus perturbation of PLZF function could provide a mechanism for the involvement of HERV-K in tumorigenesis.

Clearly, many questions remain. Are HERV-K particles produced from a single intact provirus or does trans-complementation of proteins from several loci lead to the production of composite particles? Are any of these particles infectious and do they contribute to tumor formation? The human genome sequence may be able to help to address these issues by identifying those copies of HERV-K(HML-2) that are most likely to contribute to particle production. Comparison of these loci between patients with germ-cell tumors and controls may then reveal differences which could be the focus of further research.

HERVs as regulators of gene expression

Most studies on the pathological potential of HERVs have looked for expression of HERV RNA or protein, on the assumption that disease symptoms result from inflammatory or autoimmune reactions to HERV proteins. The effect of HERVs in disease may instead be at the level of cellular gene transcription, however, since it is well known that enhancer and promoter elements in retroviral LTRs can influence the transcription of neighboring genes (Figure 2a,b). This can result in transcriptional activation or gene silencing and in changes in tissue specificity of expression [19,20]. Data mining of the human genome sequence has already been used to identify two HERV-E LTRs that act as alternative promoters for cellular genes [21].

The activity of HERV LTRs may be modulated by several factors (Figure 2c). Differential methylation of LTR promoters has recently been proposed as a mechanism for mediating

**Figure 2**

HERVs as regulators of host gene expression. HERV LTRs can influence expression of flanking host genes (host exons are indicated by black boxes and HERV is indicated in grey). **(a)** The normal transcript from the cellular promoter. **(b)** An alternative transcript from the HERV LTR (for example, see [21]). Either the 5' or the 3' LTR can be used, but transcripts from the 5' LTR may use the HERV splice donor site and could potentially generate alternative transcripts. Expression of the cellular gene may be either activated or silenced by this mechanism. **(c)** The activity of the HERV LTR may be modulated by other factors such as methylation [22], polymorphisms [23] and the activation state of the cell.

phenotypic variation [22]. LTR polymorphisms could also explain some of the differences in HERV expression between individuals, since small changes in LTR sequence can have large effects on promoter function [23]. Furthermore, factors that affect the general level of cell activation will also influence LTR activity. Whether HERV-driven modulation of gene expression is involved in disease etiology remains to be determined, but the increased detection of HERV transcripts in diseases such as cancer and autoimmunity confirms that LTR activity is altered in these conditions. A search of the genome for HERVs and other transposon promoters close to candidate disease genes may produce useful information. Analysis of the selected

HERV LTRs for polymorphisms might then reveal important differences between people with and without particular diseases.

The human genome sequence has already provided a great deal of useful information for studies on the evolution of HERVs and their role in shaping the genome. The issue of a biological function for HERVs is more difficult to address, but the genome sequence can be exploited to identify those HERV loci most likely to be capable of producing proteins or viral particles. Use of additional types of information, including information about consensus sequences of repetitive families, which is regularly updated at Repbase [24] and

about gene expression, from EST databases such as dbEST [25], will also be necessary, together with a comparison of HERVs from a number of individuals with different diseases and in different populations. This analysis may well require the development of more sensitive search algorithms for detecting HERVs in sequence data. The challenge will then be to use this information to design incisive experiments to determine the pathological or physiological roles of HERVs.

References

- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, et al.: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, et al.: **The sequence of the human genome.** *Science* 2001, **291**:1304-1351.
- Li WH, Gu Z, Wang H, Nekrutenko A: **Evolutionary analyses of the human genome.** *Nature* 2001, **409**:847-849.
- Boeke JD, Stoye JP: **Retrotransposons, endogenous retroviruses, and the evolution of retroelements.** In *Retroviruses*. Edited by Coffin JM, Hughes SH and Varmus HE. Cold Spring Harbor; Cold Spring Harbor Laboratory Press: 1997; 343-435.
- Tristem M: **Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database.** *J Virol* 2000, **74**:3715-3730.
- Medstrand P, Mager DL: **Human-specific integrations of the HERV-K endogenous retrovirus family.** *J Virol* 1998, **72**:9782-9787.
- Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, Meese E: **An almost-intact human endogenous retrovirus K on human chromosome 7.** *Nat Genet* 1999, **21**:257-258.
- Patience C, Takeuchi Y, Weiss RA: **Infection of human cells by an endogenous retrovirus of pigs.** *Nat Med* 1997, **3**:282-286.
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al.: **Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.** *Nature* 2000, **403**:785-789.
- de Parseval N, Heidmann T: **Physiological knockout of the envelope gene of the single-copy ERV-3 human endogenous retrovirus in a fraction of the Caucasian population.** *J Virol* 1998, **72**:3442-3445.
- Best S, Le Tissier P, Towers G, Stoye JP: **Positional cloning of the mouse retrovirus restriction gene *Fv1*.** *Nature* 1996, **382**:826-829.
- Towers G, Bock M, Martin S, Takeuchi Y, Stoye JP, Danos O: **A conserved mechanism of retrovirus restriction in mammals.** *Proc Natl Acad Sci USA* 2000, **97**:12295-12299.
- Lower R: **The pathogenic potential of endogenous retroviruses: facts and fantasies.** *Trends Microbiol* 1999, **7**:350-356.
- Perron H, Garson JA, Bedin F, Beseme F, Paranhos-Baccala G, Komurian-Pradel F, Mallet F, Tuke PW, Voisset C, Blond JL, et al.: **Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis.** *Proc Natl Acad Sci USA* 1997, **94**:7583-7588.
- Karlsson H, Bachmann S, Schroder J, McArthur J, Torrey EF, Yolken RH: **Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia.** *Proc Natl Acad Sci USA* 2001, **98**:4634-4639.
- Bieda K, Hoffmann A, Boller K: **Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines.** *J Gen Virol* 2001, **82**:591-596.
- Boller K, Janssen O, Schuldes H, Tonjes RR, Kurth R: **Characterization of the antibody response specific for the human endogenous retrovirus HTDV/HERV-K.** *J Virol* 1997, **71**:4581-4588.
- Boese A, Sauter M, Galli U, Best B, Herbst H, Mayer J, Kremmer E, Roemer K, Mueller-Lantzsch N: **Human endogenous retrovirus protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc finger protein.** *Oncogene* 2000, **19**:4328-4336.
- Schulte AM, Lai S, Kurtz A, Czubayko F, Riegel AT, Wellstein A: **Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus.** *Proc Natl Acad Sci USA* 1996, **93**:14759-14764.
- Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH: **Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene.** *Genes Dev* 1992, **6**:1457-1465.
- Kowalski PE, Freeman JD, Mager DL: **Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes.** *Genomics* 1999, **57**:371-379.
- Whitelaw E, Martin DI: **Retrotransposons as epigenetic mediators of phenotypic variation in mammals.** *Nat Genet* 2001, **27**:361-365.
- Schon U, Seifarth W, Baust C, Hohenadl C, Erfle V, Leib-Mosch C: **Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats.** *Virology* 2001, **279**:280-291.
- Repbase Update** [http://www.girinst.org/Repbase_Update.html]
- Expressed Sequence Tags Database** [<http://www.ncbi.nlm.nih.gov/dbEST/index.html>]