Effects of Retroviruses on Host Genome Function

Patric Jern and John M. Coffin

Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111; email: Patric.Jern@tufts.edu, John.Coffin@tufts.edu

Annu. Rev. Genet. 2008. 42:20.1–20.23

The *Annual Review of Genetics* is online at genet.annualreviews.org

This article's doi: 10.1146/annurev.genet.42.110807.091501

0066-4197/08/1201-0001$20.00

Key Words

Human Endogenous Retrovirus, LTR, transcription, recombination, methylation

Abstract

For millions of years, retroviral infections have challenged vertebrates, occasionally leading to germline integration and inheritance as ERVs, genetic parasites whose remnants today constitute some 7% to 8% of the human genome. Although they have had significant evolutionary side effects, it is useful to view ERVs as fossil representatives of retroviruses extant at the time of their insertion into the germline, not as direct players in the evolutionary process itself. Expression of particular ERVs is associated with several positive physiological functions as well as certain diseases, although their roles in human disease as etiological agents, possible contributing factors, or disease markers—well demonstrated in animal models—remain to be established. Here we discuss ERV contributions to host genome structure and function, including their ability to mediate recombination, and physiological effects on the host transcriptome resulting from their integration, expression, and other events.

ERV: endogenous retrovirus

HERV: human ERV

LTR: long terminal repeat

ALV: avian leukosis virus

MLV: murine leukemia virus

INTRODUCTION

Retroviruses, found in all mammals and a wide range of other vertebrates, provide unique opportunities for the study of the biology and evolution of virus-host relationships. Occasionally, infection of a germline cell by a retrovirus may lead to an integrated provirus that is passed to the offspring and inherited as a Mendelian gene; this is known as an endogenous retrovirus (ERV) (113). Human endogenous retroviruses (HERVs) constitute about 7%–8% of the human genome (17, 45). Many, but not all, HERVs have defects in some or all of their genes. ~~However,~~ despite millions of years since integration into the genome of a human ancestor, some HERV genes still have open reading frames (ORFs) and thus the possibility of protein expression. Expression of HERVs has been associated with several positive physiological functions as well as certain diseases, although their role as etiological agent, possible contributing factor, or ~~a~~ disease marker remains to be established. Although infectious virus resulting from ERV expression can be found in some animal species, and some HERVs exhibit insertional polymorphism, indicating recent acquisition, or perhaps still active members, no active human ERV has yet been found. Recently, however, an infectious representative of the most recently acquired HERV-K(HML2) group was reconstituted based on the consensus sequence of a number of different proviruses and found to be infectious in vitro (29), providing a new tool for continued investigation of ERVs and their potential effects on their host's cellular and genomic functions.

RETROVIRUSES

Genome Organization

A retrovirus consists of an encapsidated dimer of positive-sense single-stranded RNA, enclosed in a capsid, which in turn is enclosed in a lipid bilayer envelope. The retrovirus' life cycle differs from that of other organisms in that it includes transformation of genetic material from RNA to DNA, integration of that DNA into the host genome to form the provirus, transcription of the provirus to form genome and messenger RNA, translation and processing of virion proteins, and finally closure of the replication cycle by budding of virions from the cell surface. A typical replication-competent provirus is about 7–11 kb in size and consists mainly of the coding regions for *gag*, *pro*, *pol*, and *env*, flanked on both 5′- and 3′-ends by long terminal repeats (LTRs) formed during reverse transcription (**Figure 1**). Each of the LTRs is composed of the unique U3 and U5 regions separated by a segment (R) repeated at each end of genome RNA. U3 may vary in length and contains binding sites for different cellular transcription factors for enhancing and promoting proviral transcription. Multiple studies have shown that transcription factor binding sites and other important LTR motifs like the TATA-box coupled with a GC/GT-box specifying transcriptional initiation and the AATAAA signaling polyadenylation and 3′ end formation have remained functional in many HERV LTRs (70). The order of the structural genes (*gag-pro-pol-env*) and the arrangement of their major cleavage products are completely conserved among all retroviruses (**Figure 1**) and are necessary for virion proteins to be expressed in the proper relative amounts, to interact in a specified order, and to guide each other into position in order for correct virion assembly (22). Env is translated from a spliced subgenomic RNA and later cleaved into a trimer of SU (surface) and TM (transmembrane) subunits (**Figure 1**).

Replication and Host Defense

The binding of SU to a cellular receptor forces TM into the vicinity of the cell membrane, thus enabling its rearrangement into a fusion-competent form. Some groups of retroviruses, including the alpharetrovirus avian leukosis virus (ALV) and the gammaretrovirus murine leukemia virus (MLV), can be divided into subgroups based on their use of completely different cell surface proteins as receptors (18). Following fusion, the virion core is released into the cytoplasm (**Figure 2**), and the retroviral

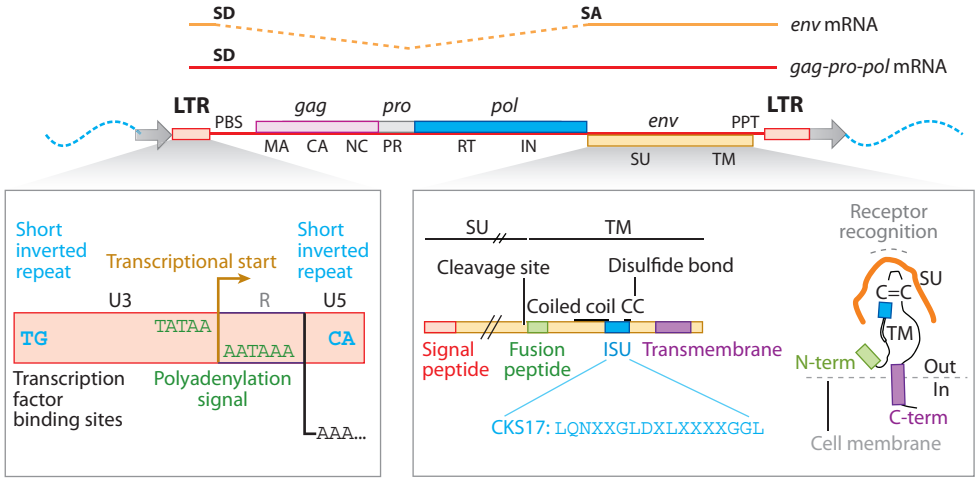

Figure 1

Provirus structure. Large arrows indicate 4–6-bp target site duplications formed during integration of the viral DNA. Simple retrovirus mRNAs are shown above. Abbreviations: PBS, primer binding site; ISU or CKS17, immunosuppressive domain; SD, splice donor; SA, splice acceptor, ppt, polypurine tract. Viral genes (proteins): *gag* (MA, matrix; CA, capsid; NC, nucleocapsid); *pro* (PR, protease); *pol* (RT, reverse transcriptase; IN, integrase); *env* (SU, surface protein; TM, transmembrane protein).

RNA genome undergoes reverse transcription into double-stranded DNA within a structure derived from the virion core to form the preintegration complex (PIC) including the retroviral DNA and IN, probably along with some cellular factors. ~~Although~~ the full functions of PICs are not yet entirely described and much work remains, ~~the~~ DNA is subsequently transported to the nucleus and integrated, using the virus-encoded IN, into the chromosomal DNA. Located immediately downstream of the U5 region in the genome is an 18-nucleotide-long primer binding site (PBS), complementary to the 3′-sequence of a host transfer RNA (tRNA), which is used as a primer for initiation of reverse transcription. Transcription from the provirus starts at the 5′ U3-R junction and the 3′R-U5 junction provides the site of 3′ polyadenylation (**Figure 1**). The major splice donor site (SD) downstream of the PBS is used for generation of subgenomic mRNAs, including *env*. Following translation, the Gag and Gag-Pro-Pol polyproteins localize to the cell membrane into which the Env protein is inserted. Assembly occurs by budding of the complex of unprocessed polyproteins and a dimer of the progeny genome RNA. After budding from the cell membrane, the virion matures as the polyproteins are cleaved into functional subunits.

To counter the threat imposed by infecting retroviruses, an array of host defense strategies has evolved. Some of these strategies involve a block at the level of entry. In mice and chickens, Env protein expressed from endogenous proviruses can prevent the cell surface receptor from interacting with MLV (66) or ALV (reviewed in Reference 126) of the same subgroup. In these species as well, there is considerable polymorphism in susceptibility to MLV or ALV infection due to point mutations in the genes encoding cell surface receptors themselves, leading to an evolutionary arms race that results in the appearance of the different subgroups of infecting viruses.

Receptor blocking is also known in sheep, where endogenous JSRV (enJSRV) interferes with the entry of exogenous (infectious, XRV) JSRV (98). Additionally, some enJSRVs can prevent replication by expression in the same cell as exogenous JSRV, due to dominant lethal mutations in *gag* that prevent hybrid capsids from exiting the cell (98).

XRV: exogenous (infectious) retrovirus

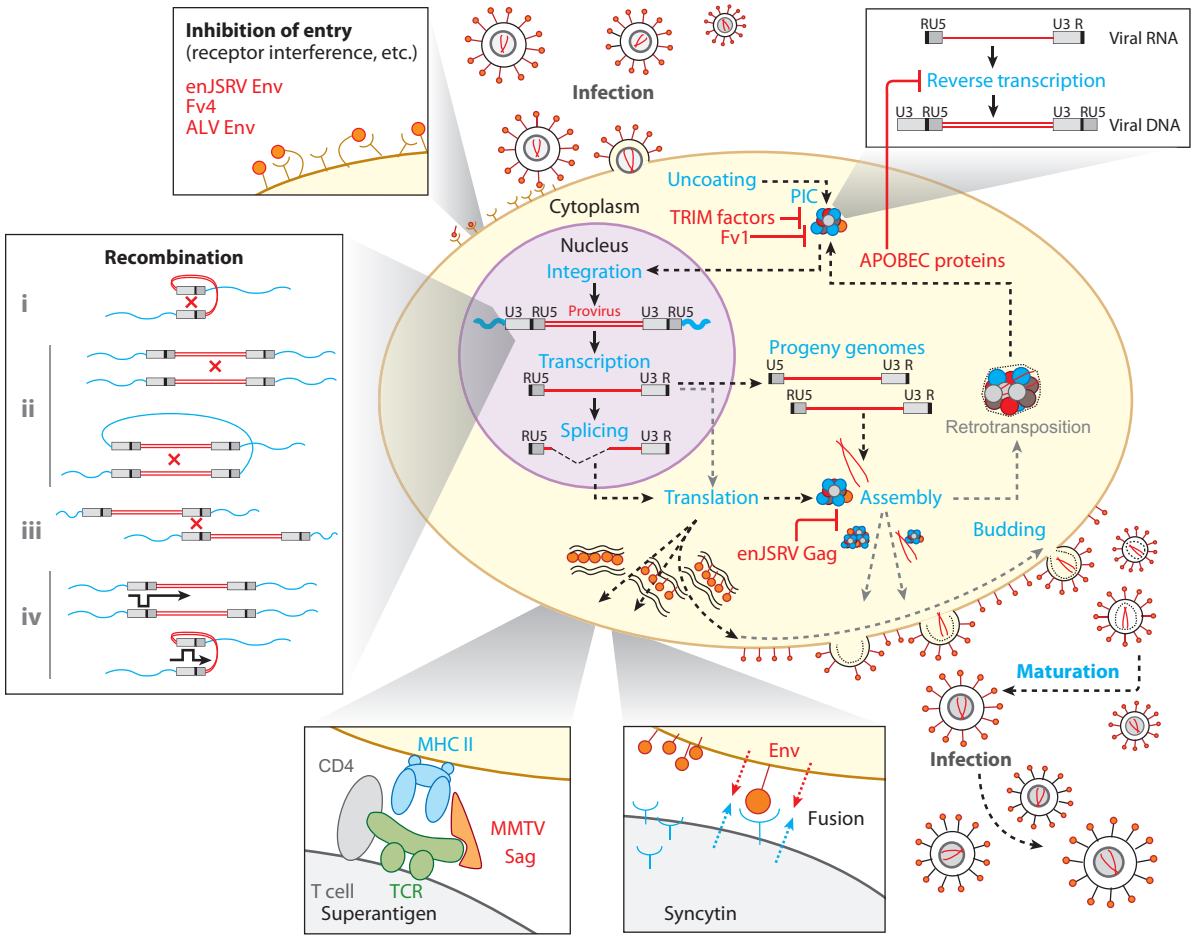

Figure 2

Retrovirus life cycle, host cell interaction, and host retrovirus-derived proteins. Host inhibitory factors are indicated in red lettering. Details shown in the boxes are discussed in the text.

A completely different mechanism of resistance is provided by endogenous MMTV proviruses, which, in common with their exogenous counterparts, express a gene, *sag*, encoding a superantigen capable of interacting with V_β chains on the surface of T cells (**Figure 2**) (1). Expression of *sag* on B cells following exogenous infection stimulates T cells to secrete cytokines that promote division of the infected cells and, hence, spread of the virus. Expression of *sag* from endogenous proviruses causes deletion of the T cells expressing the cognate V_β chain during early development, reducing efficiency of spread after exogenous infection with the same virus. As with *env* genes, the variety of V_β types recognized by *sag* genes of different MMTV strains reflects an evolutionary arms race between virus and host.

MMTV: mouse mammary tumor virus

In the case of MLV, the *Fv1* gene, derived from the *gag* gene of a different Murine ERV (112), blocks infection in a virus- and strain-specific manner, determined by a single amino acid residue in the CA region of Gag, which specifies sensitivity to restriction by the allele found in B-type (e.g., BALB/c) vs N-type (e.g., NIH Swiss) mice. Fv1 is a cytoplasmic protein that confers a restriction on replication after reverse transcription and before nuclear import

and integration (**Figure 2**). More recently, an analogous innate defense mechanism affecting HIV-1 infection in some primates was identified as a protein known as TRIM5α, a member of the large tripartite motif family of host proteins. TRIM5α of Old World rhesus monkeys restricts HIV-1 infection (117) at a stage after entry, but prior to reverse transcription. Similarity of the mechanisms of Fv1 and TRIM5α restriction is demonstrated by the fact that that TRIM5α from humans, rhesus macaques, and African green monkeys also restricts N-tropic (but not B-tropic) MLV (129).

An additional innate cell restriction mechanism against retrovirus replication in human cells is imposed by the cytidine deaminase APOBEC3G (and also APOBEC3F), promoting G to A mutations by deamination of cytidine to deoxyuracil during minus-strand DNA synthesis (13, 41), resulting in G to A changes in the plus (sense) strand of the provirus. In mice, the single ABOBEC3 variant restricts retrotransposition of Env-deficient mouse MusD and IAP elements (35) as well as infection and in vivo spread of MMTV (95). APOBEC3 may also have been an actor in the silencing of nonecotropic endogenous MLVs (52).

ENDOGENOUS RETROVIRUSES

Endogenization

Endogenous retroviruses (ERVs) are genetic elements that reside as proviruses in their host's genome, presenting the only known "fossil" record of an infectious agent. Although retroviruses usually infect somatic cells, occasionally a retrovirus infects a germline cell and the acquired provirus can then be passed to the offspring and inherited as a normal Mendelian gene (113). Some 7%–8% of the human genome is of retroviral and retrotransposon origin (17, 45). No infectious or autonomously retrotransposing HERVs have yet been observed despite the presence of open reading frames in a few of them.

Recently acquired endogenous proviruses, including ones found in chickens, mice, cats, and some primates, may retain the ability to give rise to infectious virus [although their expression is generally greatly limited by CpG methylation (102)]. Thus they may continue to be transmitted both vertically as a provirus to its host's offspring or horizontally, by infecting somatic cells, in coexistence or competition with their exogenous infectious counterparts. With time, proviruses may become fixed in the host genome, with subsequent selection for those that are least harmful or even beneficial. It has been estimated that, for most HERVs, this process has been going on for at least the past 100 million years (Mya), with an apparent peak in numbers around 30–45 Mya around the time of the split between the Old and New World monkey lineages [recently reviewed in (6)]. This process is currently observed in some species, such as MLV and MMTV in mice [reviewed in (18)], KoRV in Koala (118) and JSRV in domesticated sheep (5), but it has not yet been shown with HERVs, although there are some hints [see section on ERV Polymorphism, below]. Nevertheless, the nonhuman proviruses provide valuable models to study the transitional states of endogenization to understand how HERVs once became fixed.

IAP: intracisternal A-type particles

Endogenization: nonlethal retrovirus integration into germline cell and subsequent inheritance as a Mendelian gene

Classification and Distribution

At present, there is no well-established or accepted standard for naming and classifying all ERVs. For HERVs, traditionally the tRNA complementary to the PBS (**Figure 1**) has been used for this purpose (68). Thus, members of the HERV-H group contain a PBS complementary to histidine-tRNA, and most members of the HERV-K(HML2) group have a Lysine-tRNA PBS. This classification is, however, unreliable as proviruses of the same phylogenetic groups may display differences in PBS (49, 52), and otherwise unrelated proviruses may use the same tRNA as primer. The situation is even more chaotic in other species, and an accurate and usable system of classification and nomenclature is badly needed. The current RepBase nomenclature (55) is based on nucleotide identity to machine-generated consensus sequences of repetitive elements, but it does not apply well to retroviral sequences, where ~~studies of~~

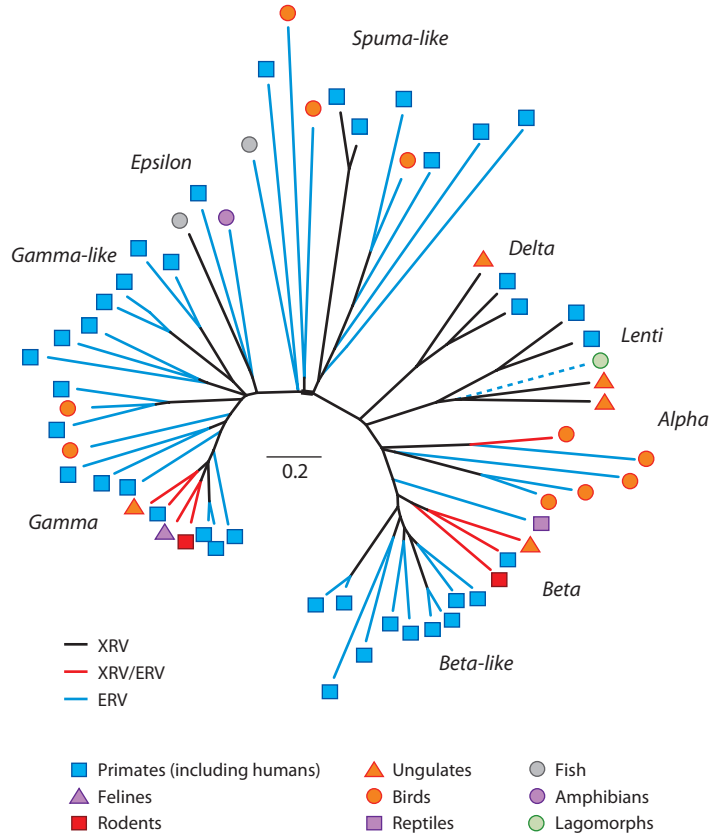

Figure 3

The seven retroviral genera: alpha-, beta-, gamma-, delta-, epsilon-, lenti-, and spuma-like retroviruses and their intermediate groups. Shown is an unrooted tree based on Pol sequences. The large various host species are indicated with symbols next to each taxonomic unit. Black branches indicate viruses known only in exogenous infectious forms (XRV); red branches indicate viruses present in both XRV and endogenous (ERV) forms; and blue branches indicate ERVs. Modified from Jern (50).

phylogenetic (**Figure 3**) and related comparisons have proved to be more useful for classification of ERVs (50, 53, 63).

Endogenous retroviruses, in the form of either infectious virus or proviruses, have been reported and characterized from most vertebrates (40) and have, to date, been found to represent all retroviral genera except deltaretroviruses (the group that includes HTLV and BLV) (**Figure 3**). The increasing availability of genome sequences from different species highlights the need to revise current nomenclature to suit the inclusion of the numerous novel ERVs encountered, and to more rationally classify those that are already known. A recent computer-aided analysis of a subset of the current species genomes made available through genome sequencing projects shows a variation in ERV types and number, with over 3500 more or less complete HERV proviruses in the human genome and even more in other mammals, up to 8000 in mouse, for example (110).

The abundance and diversity of ERVs among species and their integration patterns within a species are useful indications of evolutionary selection and host-ERV dynamics. For example, there is evidence that acquisition of a specific allele of TRIM5α may have protected human ancestors from infection with an extinct virus (PtERV1) found in the genomes of chimpanzees and gorillas, but absent from the human genome (56). Another computer-based analysis of the human genome sequence has shown that, relative to the distribution of integration sites of replicating viruses, integration sites for all classes of LTR elements are underrepresented within and in the vicinity of genes (86), and it has further been noted that ERVs of most families are less likely to be found in introns than in intergenic regions (121). Furthermore, those that are found within introns tend to be in the opposite transcriptional orientation from that of their host gene. These effects become more pronounced with greater age of the provirus, providing clear evidence for selection against proviruses that may have deleterious effects on gene expression, reducing their probability of fixation. A model for retention of intronic ERVs integrated in antisense chromosomal gene transcription ~~ERV in introns~~ has been proposed. In this model, potential cryptic splice sites introduced by an ERV are blocked from normal cellular mRNA splicing (121).

ERVs AND GENOMIC EVOLUTION

Endogenous Proviruses and Host-Virus Relationships

With time, virus infections in a species tend toward a relatively benign host-virus

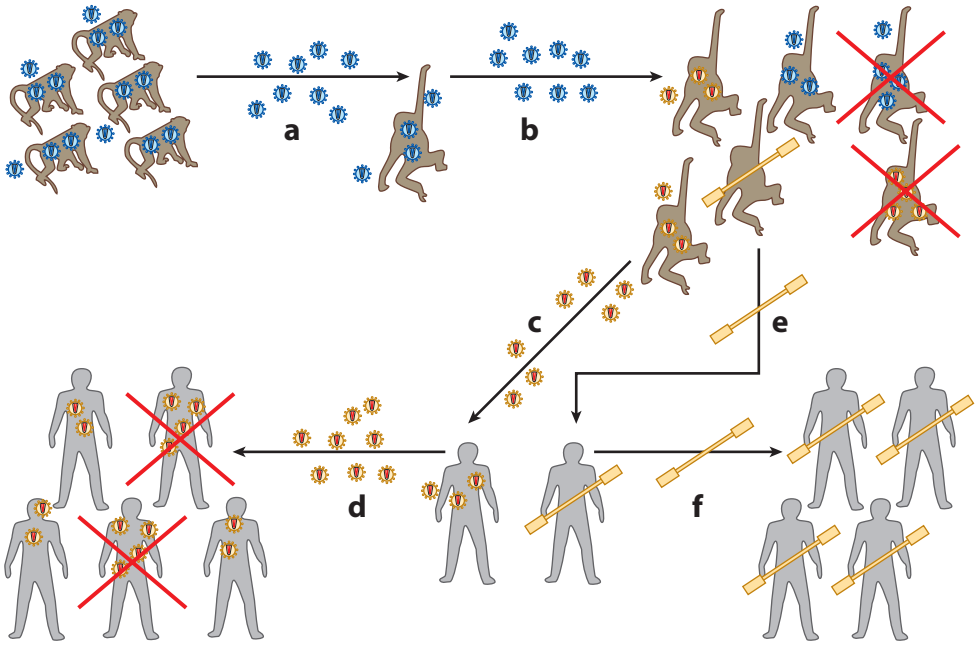

Figure 4

Host-virus and host-endogenous provirus relationship. Three species are pictured together with various types of retrovirus transmissions. Transmission of virus species (*a*), followed by spread in species (*b*) leading to selection within species (*b*, noted by red Xs), and followed by continued spread of virus (*c*) to start the cycle over again (*d*). Additionally, a retrovirus infecting a germline cell may become fixed in the population and spread through generations. Although the virus may become extinct, even after speciation events it can still be detected in descendant species as an ERV (*e* and *f*).

relationship (**Figure 4**) that generally allows the virus to infect and spread with minimal harm to its host, except to promote transmission. Transmission to another species is often accompanied by increased pathogenic effects in a species-specific manner. HIV-1 and -2 provide examples of this effect, as related SIVs do not harm their African primate hosts significantly but can be quite virulent in other species, including humans. Transmission of the virus to a new species may thus increase its pathogenic effects and lead to extinction of the virus in the new species. In some cases, viruses with reduced pathogenicity may arise and continue to spread within the new species, concomitant with selection of variant hosts that can resist the infection or its pathogenic consequences.

If a retrovirus infects a germ cell and becomes endogenous (**Figure 4**), its expression may interfere with other infecting exogenous retroviruses [see section on Replication and Host Defense, above] leading to a selective advantage for animals that have acquired the endogenous provirus and contributing to the extinction of the exogenous counterpart within that species. Once fixed, the endogenous provirus will continue to be passively transmitted to subsequent generations and ultimately become a fossilized record that allows us to study earlier infections of species during their evolution. In a few cases, fixation of proviruses has been promoted by selection for other beneficial effects, such as trophoblast fusion (discussed below). Given the relative rarity of such benefits, and the very large number of proviruses, most integrated proviruses likely have very little selective consequence.

ERV Polymorphism

As compared to some other mammals, humans exhibit very little polymorphism in provirus content from one individual to the next, and most proviruses are found in the same location in chimpanzees as well, implying an age of more than 5 million years. The most recently active group is the MMTV-related HERV-K(HML2), and the discovery of an almost intact member (83), as well as the observation of a number of proviruses in the same group that are unique to humans (44), increased the interest in polymorphism of HERVs, and raised expectations that an active, infectious, member might be found.

Two principal mechanisms lead to polymorphism of a provirus among individuals. Homologous recombination between two LTRs (**Figure 2**), excluding proviral DNA and leaving a solo LTR at the locus (44, 113), represents most of the polymorphisms (44). For most groups, fixed solo LTRs vastly outnumber their cognate ancestral proviruses (113). They appear to form more frequently relatively soon after integration, probably due to a higher recombination rate in young integrations without accumulated mutations (10). About a half-dozen proviruses of this group also display insertional polymorphism, where both a provirus and its allelic preintegration site can be found (44), implying relatively recent integration. All the polymorphic proviruses are found in widely distributed human groups, implying that their integration preceded human radiation out of Africa (44, 74, 78), and must have been well over 10,000 years ago. The existence of additional polymorphic proviruses supports the hypothesis of a recently or even still active HERV-K(HML2) allele (8), although such a transpositionally active or infectious element remains to be identified.

Infection and Reinfection

The distribution of endogenous proviruses in the genomes of mammalian and avian species implies that the process of endogenization of a given group of viruses, once initiated, has continued for very long periods of time. The earliest proviral representatives of HERV-K(HML2), for example, are found in all Old World primates, implying that the group has existed for more than 25 million years (85), yet the most recent exemplars are only a few tens of thousands of years old. After initial entry of the founder of the group into the germline, the mechanisms of retrotransposition and reinfection that provide for the long survival and dispersal of proviruses of a single type throughout the genome are unclear. There are two extreme possibilities. On the one hand, expression of a provirus leads to occasional retrotransposition into new sites within a germline cell. This sort of intracellular spread must occur with all non-LTR elements, as well as some elements, such as MusD and IAPs of mice, which are originally derived from retroviruses, but lack *env* genes and are unable to give rise to infectious virus (76). On the other hand, continuity may be provided entirely by replication and transmission as exogenous virus, with occasional infection of germline cells occurring in viremic animals. For most endogenous proviruses, the correct explanation is probably somewhere in between, with endogenous proviruses occasionally giving rise to infectious virus, which can then spread through the individual, and perhaps among individuals—and occasionally across species—and subsequently infect a germline cell. In the only animal model for this process, a high level of viremia resulting from expression and replication of an endogenous MLV in pregnant mice of the correct genetic background led to frequent infection of the germline of female offspring during active oogenesis late in fetal development (reviewed in Referene 106). The necessity for rounds of viral replication between integration events is implied by the low ratios of nonsynonymous to synonymous mutations in *env* genes of HERV-K(HML2) proviruses, consistent with purifying selection and a continuing need for functional Env proteins (9).

Once present in the germline, a provirus behaves like any other piece of chromosomal

DNA and is subject to the same rules of evolution—mutation, selection, recombination, etc.—that govern the rest of the genome.

Even if defective, as is the case with most or all ERVs in any given species, endogenous proviruses can contribute to replication of other ERVs or to related exogenous viruses by complementation or recombination. For example, some strains of inbred mice carry two endogenous ecotropic (EMV) proviruses, both defective, yet they often become viremic at an early age with replication-competent recombinants between them (47). Because the mechanism of retroviral recombination requires copackaging of genomes into an infectious virion, the two proviruses must also be capable of complementing one another, each provirus providing functional proteins to make up for the defect in the other. In a more extreme case, high-leukemic strains of mice, such as AKR, are viremic at birth, with an MLV that subsequently undergoes recombination with at least two other endogenous MLVs to give rise to the virus that eventually causes the leukemia (116). Similar types of recombination can also involve infecting exogenous viruses. For example, subgroup J ALV, the cause of some serious outbreaks of disease in commercial poultry, is a recombinant in which an exogenous ALV has acquired the *env* gene of an old, and mostly defective, endogenous provirus (12). Also, "patch repair" of defective MLV mutants by localized recombination with proviruses in the mouse cells on which they are grown is a well-known phenomenon (89). At this time, no examples of such recombination events involving HERVs are known.

Complementation by ERVs of genetic defects in infecting exogenous viruses or other endogenous viruses has also been reported. For example, some endogenous ALVs can express functional Env, but not Gag-Pro-Pol products. Cells containing such proviruses allow production of infectious virus following infection with a common strains of Rous sarcoma virus containing deletions of *env* (119). In humans the spread of a subset of HERV-K(HML2) proviruses, known as type 1, which carry a 292-bp deletion in *env*, must have been accomplished by complementation with functional Env protein expressed from another provirus (28). Proviruses capable of complementing disrupted proviruses *in trans*, as proposed for the proliferation of some HERV-H (75), have been referred to as midwife elements (49). The small group of HERV-Fc elements (11) within the larger HERV-H-like group possibly has had "midwife" properties as the single-copy HERV-Fc1 is, despite a great age implied by divergent LTRs (5.7% different), almost intact in *gag* and *pol*, and intact in *pro* and *env* (49).

Midwife element: scarce but relatively complete element that assists amplification of related elements by providing proteins *in trans*

Horizontal Transfer

Unlike many DNA viruses that establish long-term relationships with their hosts, effective cross-species transmission of retroviruses has been relatively common, at least over evolutionary time, and endogenous proviruses provide a good record of this process (102). Among galliform birds, recent endogenous ALVs are found in chickens and pheasants, but are completely absent from closely related species, including turkey and quail. All species in this group contain more ancient proviruses, reflecting infection of a common ancestor (23). In primates, the PtERV1 elements (51, 100, 130) are found in the genomes of chimpanzees together with gorillas, baboons, and macaques, but not in humans, implying recent introgression into the ancestors of some species. Further, phylogenetic trees of PtERV *gag* and *env* differ from generally accepted primate species trees, indicating horizontal transfers. Such transfers have been described for BaEV-related viruses, which have spread among African primates, and also to cats, and even to Australia in recent evolutionary time (118, 122). The BaEV-related Koala retroviruses (KoRV) is presently found in a transitional state between infectious and fixed endogenous proviruses. Additionally, phylogenetic studies have shown signs of possible trans-species transfers of other MLV-like gammaretroviruses to several vertebrate species (82). Thus, cross-species transmissions of virus derived from or closely related to endogenous viruses of one species may have occurred more

or less frequently through various routes and, when endogenized, they have contributed to the genomic evolution of their new host.

ERV-RELATED HOST EFFECTS

ERV-Derived Proteins

In addition to offering protection against exogenous virus infection, some endogenous viral gene products have been coopted for other important physiological functions. The best-characterized of these proteins are known as syncytins for their apparent role in placental development. The human protein syncytin-1, which is the product of the *env* gene of the well-described HERV-W provirus, is expressed in trophoblasts (77), cells that form the outer layer of the placenta, where it mediates cell fusion and syncytium formation (**Figure 2**) (88). Screening of the human genome for possible functional *env* genes, followed by cloning and expression of 16 other fusogenic ERV Env candidates, led to the identification of the *env* gene of HERV-FRD, which has the same properties as syncytin-1, and its product was named syncytin-2 (14). Phylogenetic analysis of the *env* genes encoding both syncytin-1 and -2 shows that both have been subjected to strong purifying selection during primate evolution consistent with the proposed role in placentation (15, 81). Indeed, syncytin-2 isolated from both New and Old World primates has retained its ability to encode a functional Env protein, despite the very long time (>40 million years) that must have passed since its integration, in contrast to the severe damage accumulated in the other viral genes during this time (62) More recently, additional fusogenic, placentally expressed, murine endogenous Env proteins named Syncytin A and B (34) have been described, as has an endogenous JSRV-related provirus of sheep that encodes an Env protein with similar properties and expression (97). In the latter case, direct evidence for an important role in development has been obtained by the demonstration of placental development defects following inhibition of *env* expression (32). These five proviruses are not closely related to one another, are at different integration sites, and therefore must have been independently acquired in three different mammalian orders. Human and murine syncytins are encoded by gammaretrovirus-like proviruses, and JSRV is a betaretrovirus (**Figure 3**). Their cooption for a common physiological role represents a remarkable example of convergent evolution. Whether fusogenic ERV-derived proteins are involved in other normal host functions beyond placental development remains to be explored.

Some retroviral Env proteins include an immunosuppressive domain within TM (**Figure 1**) (21). Introduction ~~into cancer cells~~ of an infectious murine retrovirus *env* expression vector presenting this domain can, in a mouse model, promote tumor growth by allowing escape from immune surveillance (80). Recently, it was also shown that the placentally expressed ERVs, human syncytin-2 and mouse syncytin-B have similar immunosuppressive properties (81), as do some of the abundant HERV-H Env proteins (25, 79), raising the possibility that their expression may play a similar role in human cancer progression. The interesting idea that the immunosuppressive function of ERV Env proteins may also play a role in protection of the developing fetus from rejection by the maternal immune response (115) awaits a critical experimental test.

Expression of an endogenous provirus is, by itself, insufficient evidence to establish such a physiologic role. The *env* gene of the well-studied ERV3 (or HERV-R) is expressed at a high level in several fetal tissues, particularly in the developing adrenal gland (2), and has been proposed to contribute to cellular differentiation as well as placental development. However, a polymorphism homozygous in 1% of the Caucasian population results in a premature stop codon in this gene (26). Thus, unless the truncated protein is sufficient for its normal function, it is hard to conceive a primary function for ERV3 Env during fetal development. Expression of HERV-K(HML2) at high levels is common in human placenta, as well as certain malignancies (particularly germ-cell, breast,

and prostate cancer), often leading to production of normal-looking, but noninfectious virions; however, the physiologic or pathologic significance of this observation awaits a genetic test.

In addition to the common retroviral genes (**Figure 1**), both exogenous and endogenous MMTVs encode a superantigen (*sag*), a cell surface protein presented by major histocompatibility complex (MHC) (**Figure 2**), required for transmission and pathogenesis (1). As well as blocking effective spread of exogenously transmitted virus, expression of Sag from endogenous proviruses leads to depletion of large subsets of T cells, which can lead to altered resistance to other pathogens, such as polyoma virus (73).

Sag: superantigen

ERVs and Disease

Ever since the discovery of pathogenic effects, especially cancer, of MLV, RSV, MMTV and other infectious retroviruses, as well as some of their endogenous counterparts, in well-studied animal models (see Reference 18), the role of HERVs as causal or secondary factors contributing to human disease including cancers and various ~~neuro~~degenerative disorders has been debated. Many attempts to link ERVs and disease have focused on transcription of endogenous proviruses associated with disease states. Without sufficient genetic support such observations created the field of "rumorvirology" (123), characterized by highly overinterpreted conclusions.

The mouse has been a particularly useful model for studies of ERVs, their dynamics, and phenotypic effects on their host. Studies of endogenous MLVs and IAPs have identified associated genetic disorders in mice including hairless (*hr*), dilute (*d*), and agouti (*A*) (48, 92, 114). In these cases, causality could be established because of relatively frequent spontaneous recombination events that generate solo LTRs and simultaneously restore the normal phenotype. Indeed, from the frequency of *d/+* revertants appearing in colonies of *d/d* mice, the rate of solo LTR formation at this locus could be estimated at about 4.5×10^{-6} events per meiotic generation (109). Genetic proof of a causal role for endogenous MLV and MMTV in cancer is provided by the invariable presence of clonal proviruses in the tumors derived from the ERVs by replication in the host; in the former case, this is accompanied by several recombination events to generate chimeric viruses with multiple endogenous parents. These new proviruses are often integrated in the vicinity of known protooncogenes whose altered expression is intimately involved in oncogenesis. In the absence of direct genetic proof such as that offered by novel clonal integration of proviruses in cancer cells, causality of ERVs for nonmalignant diseases can be difficult to establish. For example, the presence of Env proteins of endogenous xenotropic MLV in immune complexes in a mouse model of lupus erythematosus (46) was believed to provide evidence for causality, until it was shown that mice bred to lack the relevant provirus could still exhibit the disease (24).

Genetic proof establishing a connection between a HERV and disease has been much harder to obtain. With the exception of lymphoma associated with use of a gene therapy vector (64) and some cell lines derived from HTLV-associated tumors (20), activation of a protooncogene by a provirus has not been reliably observed in any human cancer. A role for exogenous infection by two endogenous-derived murine viruses has been proposed for human cancers. One group has reported evidence for MMTV infection in human breast cancer after cross-species transfer (87), but neither conclusive evidence nor confirmation from other groups has been forthcoming. Better evidence, including reconstruction of infectious virus, has been obtained for the association of xenotropic MLV with a small sample of human prostate cancers (31), but the infected cells are stromal, not tumor, cells.

Although statistically significant upregulation of transcription of HERVs relative to control tissue is repeatedly observed in some cancers (16, 123), no novel proviruses have been reported in these cases, nor has any polymorphic provirus been genetically linked to

disease susceptibility (93, 123). Particular attention has been on the MMTV-related HERV-K(HML2) group. Spliced *env* transcripts have been detected in human breast cancer, but not in healthy controls (125), and particles derived from the polymorphic HERV-K108 and K113 proviruses containing mature Gag and Env proteins have been isolated from human melanomas (94). The altered expression observed is most likely a consequence, not a cause, of the transformation event, perhaps related to the altered transcriptional milieu of the cancer cell. Given that there are likely to be as-yet undetected polymorphic proviruses of this group present at low frequency in the human population (44, 93), and that functional HERV-K(HML2) can be reconstituted from the consensus of the most recent human proviruses (29), a continued investigation into this tantalizing association is warranted. Reactivation of HERV-K(HML2) has been observed in HIV-infected patients (39), consistent with the idea that some or all of these proviruses are susceptible to transcriptional activation by change in the state of the host cell.

Bioinformatic approaches, including *in silico* analyses of expression by scanning databases for reported expressed sequence tags (ESTs), have proved useful as complements to the laboratory experimental data. An extensive study matched HERV proviruses representing the distantly related genera (gamma- and beta-like retroviruses) to ESTs, detected more frequently in cancer tissues than in normal tissues (111). ERV association with cancers is not only a human (HERVs) or mouse (MLV and MMTV) phenomenon, but is also found in other mammals; e.g., lung tumors in sheep caused by the Jaaksiekte retrovirus (JSRV) (128). However, in most cases the variation in retroviral genomic portions between species (110) creates a suspicion that transcriptional upregulation may be largely secondary effects of the disease, as shown for several retroviral sequences in a murine model of cancer cachexia (91).

Although the ERV-cancer connection has been given most attention and has yielded many interesting results, more obscure in the ERV-disease context are numerous reports that have attempted to connect HERVs to neurodegenerative disorders such as multiple sclerosis (MS) and schizophrenia. Multiple sclerosis–associated retrovirus (MSRV) was found to be expressed in cell lines and plasma from MS patients and was characterized as HERV-W (99). HERV-W Env expression was found to be upregulated in demyelinating brain tissues of MS patients (4). Involvement of another provirus, HERV-H, in MS has also been suggested (90). Thus, two distinct retroviruses have been erratically associated with MS. An explanation could be that several proviral loci are activated. It is also possible that the change in HERV expression may be the result, and not the cause, of inflammatory disease within the brain and increased macrophage activity (54).

An increase in HERV-W RNA expression has been shown in monozygotic twin pairs discordant for schizophrenia as well as in schizophrenia patients compared to those of healthy controls (27). However, other retrovirus transcripts were also found, albeit in lower amounts (59). Although low expression was detected, these experiments were strengthened by the use of owl monkey kidney cells, which are from an Old World monkey that does not have many of the more recent HERV integrations. Further, the HERV-W receptor is the transporter protein for glutamate, an important brain signaling molecule (69), increasing the possibility of a connection to neurodegenerative disorders. However, elevated HERV-W RNA expression was not detected in brain tissue using real-time PCR, but HERV-H RNA expression was significantly higher than controls (38). This issue clearly needs further exploration before a definitive result can be proclaimed.

Promoters and Enhancers

Retroviral elements that integrate in the vicinity of genes may influence normal genome functions in their host (**Table 1**; **Figure 5**). The numerous potential binding sites for transcription factors in the proviral LTRs may influence

Table 1 Examples of mammalian ERV and XRV effects on the transcriptome

Effect[1]	Function	Provirus/solo LTR	Examples	Reference
A	Promoter	ERV9 LTR	ZNF80 (zinc finger protein)	(30)
B	Alternate promoter	HERV-E	APOCI (apolipoprotein CI)	(84)
C	Bidirectional promoter	HERV-L	DSCR4 and DSCR8 (Down syndrome critical region)	(33)
D	Promoter, intergenic splicing	HERV-H	PLA2L (phospholipase A2-like)	(37)
E	Tissue-specific alternate promoter	HERV-P	NAIP (neuronal apoptosis inhibitory protein)	(105)
F	Promoter	ERV3/HERV-R	H-PLK (human provirus linked Krüppel gene)	(60)
G	Exonization	MLV	Dilute (d) coat color	(48)
H	Promoter, enhancer	ERV9 LTR	β-globin locus	(72)
I	Tissue-specific enhancer	HERV-E	Amy1 (salivary amylase)	(107)
J	Tissue-specific regulation	HERV-E	PTN (pleiotropin)	(108)
K	Promoter, enhancer	MLV	*Evi-1* proto-oncogene	(7)
L	Poly-A	HERV-K(HML2) LTR	LEPR (leptin receptor)	(58)
M	Alternative splicing, Poly-A	MLV	Hr (hairless)	(114)

[1] See **Figure 5**.

transcriptional activity of nearby genes (70). Alteration of expression of genes most commonly results from LTRs found upstream of genes in antisense orientation or downstream in sense orientation (106). A telling example is the expression of amylase in the human parotid glands, where integration of HERV-E in reverse orientation upstream of a copy of the pancreatic amylase gene promotes its expression and release into saliva (107). The specificity of its LTR for salivary expression (70) implies that the ancestral virus was normally transmitted in saliva. Similarly, bidirectional promoter activity from LTRs has also been observed in the large HERV-H group, which had strong promoter activities in several cell lines (36) and for HERV-L/ERV1 (33). In human malignant trophoblasts, HERV-E integrated into the growth factor gene pleiotropin (PTN) has generated cell type–specific promoter activity (108). LTR promoters can further enhance the transcription from a native promoter (**Figure 5**). An example is the presence of an HERV-E LTR that increases the native promoter activity and expression of apolipoprotein C-I (84). Such LTR promoter and enhancer functions can influence native promoters over a very long range; distances up to 100 kb have been observed (7, 127).

In a bioinformatic study of the human genome, it was shown that all classes of LTR elements were underrepresented within and in the vicinity of genes (86). Such a distribution is not observed in recently integrated proviruses (19), implying that it results from selection—probably because of the potential of proviruses to influence transcription of nearby genes. Solo LTRs, which are much more abundant than their cognate proviral counterparts, can retain promoter activity (30, 72) and sometimes (but not always) lead to polyadenylation of spliced chromosomal transcripts (**Table 1**; **Figure 5**) (58). It can thus be concluded that ERVs, solo LTRs, and other transposable elements have had a major impact on the evolution of gene families in mammals (120).

Given the large number of ERVs and related elements in the vertebrate genome, as well as the potential of their LTRs to provide strong enhancer and promoter elements, one might expect them to wreak transcriptional havoc, with a very large fraction of the RNA synthetic effort of an organism devoted to their expression. Also, such expression might be expected

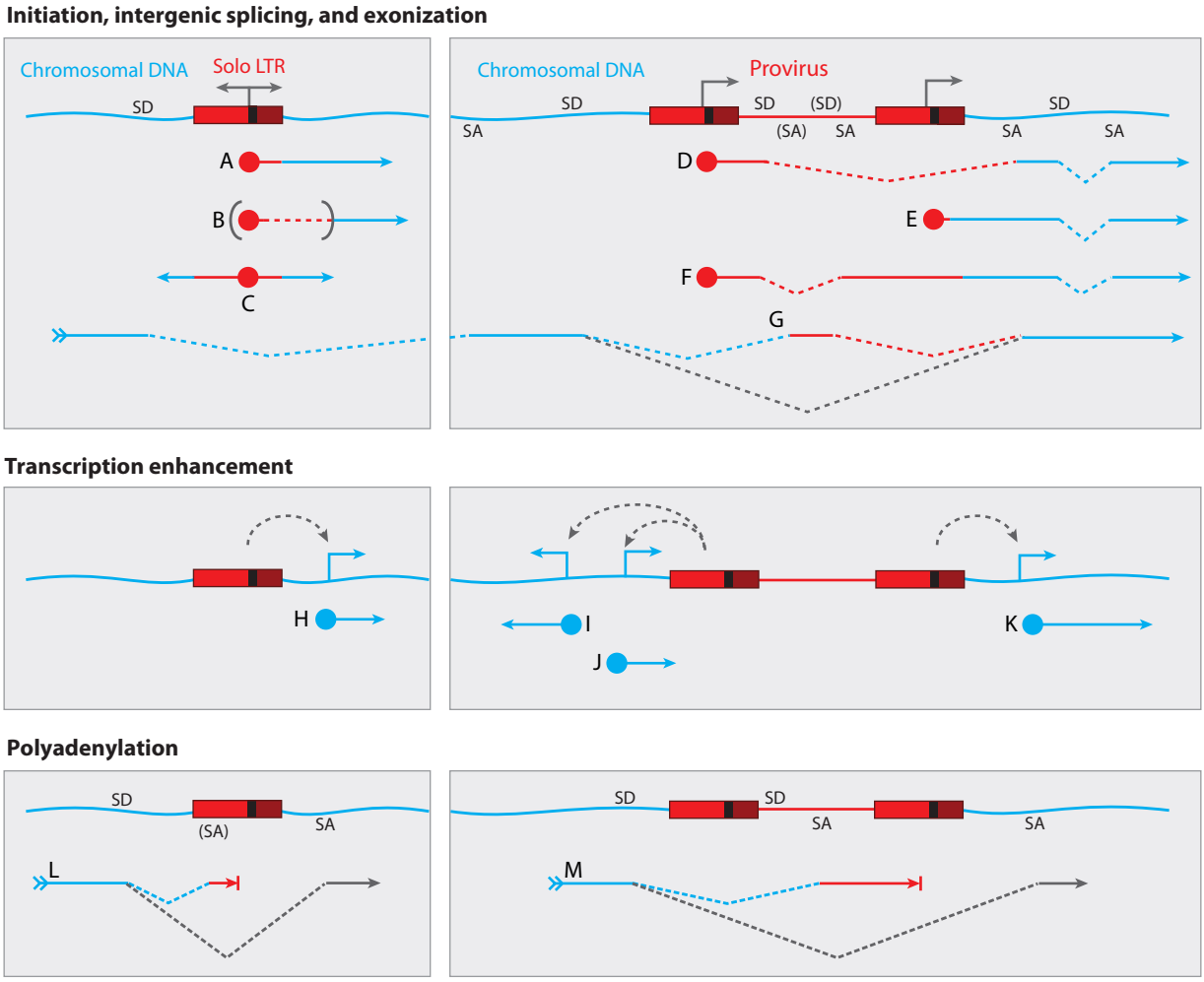

Figure 5

Proviral and solo LTR effects on the chromosomal transcriptome. LTRs can promote transcription of native chromosomal genes and also enhance transcription from native promoters. Transcription initiated in a provirus can lead to intergenic splicing with downstream native genes and intronic ERVs can introduce exonization from the normal genomic transcripts. Lastly intronic LTRs can under some conditions act as alternative polyadenylation signals and cause premature termination of the native gene transcript. Arrows indicate direction of transcription; filled circles indicate transcription start sites; and dotted lines indicate sequences removed by splicing. Viral-derived sequences are shown in red. SD, splice donor site; SA, splice acceptor site. Examples for A-M are given in **Table 1**.

to lead to release of the potentially pathogenic viruses encoded by some proviruses and subsequent reinfection of the host. Nevertheless, despite the large numbers of proviruses, most tissues in an organism do not express high levels of ERV transcripts or replicating viruses. Although some proviruses are probably defective for transcription, and others are highly tissue-specific, the principal overriding control is at the level of CpG methylation. Methylation of a large fraction of genomic DNA, including proviruses, occurs during early development (101) and persists thereafter unless reversed by specific developmental signals or other events such as DNA repair. Indeed, it has been argued that a primary function of the methylation

machinery is to render ERV elements harmless (102). Perversely, in so doing, methylation may also greatly reduce the selective disadvantage conferred by endogenous proviruses and thus promote their accumulation over evolutionary time.

Variation in gene expression due to differential methylation can even extend across generations in some cases. An example of this kind of inherited epigenetic state in mice is transcription originating in an IAP element inserted upstream of the *agouti* (*A*) gene locus, causing ectopic expression of Agouti protein, resulting in yellow fur, obesity, diabetes, and increased susceptibility to tumors (92). Agouti expression can result in variegated phenotypes ranging from yellow fur to wild-type agouti and intermediate phenotypes, correlating with the level of IAP methylation. The mottled fur colors may thus derive from stochastic and incomplete methylation silencing of IAP expression during early embryogenesis, resulting in a mosaic pattern (127). In the human genome the three proviral HERV-E LTRs, including HERV-E.PTN (**Table 1**; **Figure 5**), which function as an additional tissue-specific promoters in the placenta (70), have variable, but generally reduced levels of methylation compared to peripheral blood leukocytes (103). Indeed, many HERVs are primarily transcribed in placenta compared to other tissues, indicating stronger LTR activities in this tissue (70).

The role of methylation in protection against somatic effects of transposable element expression recently gained support with the finding that mice with heterozygous knockdown of the maintenance CpG methyltransferase (*Dnmt1*) showed a high frequency of thymic lymphomas. These tumors are characterized by increased expression of *Notch1* due to novel intronic IAP integrations resulting in 5′-truncated *Notch1* transcripts (42). Increased ERV transcription has also been reported in patients infected with HIV-1 (39). Whether this effect is related to methylation or some other epigenetic mechanism remains to be examined.

Alternative and Intergenic Splicing

Integrated retroviral elements can also affect gene expression by providing alternative and aberrant sites for splicing of transcripts of native cellular genes. The leptin obesity hormone receptor (LEPR, **Table 1**) exists in two variants that differ in size due to alternative splicing into a HERV-K LTR (**Figure 5**) (58). Intergenic splicing can also occur following expression of a gene driven by an upstream LTR, as demonstrated by a HERV-H provirus, whose 5′LTR initiates transcription of a phospholipase A2-related gene encoding a digestive enzyme normally expressed in the pancreas. The aberrant transcript is translated into the PLA2L (PLA2-like) protein, which is expressed in human teratocarcinoma cells (37). It was later shown that the last two thirds of PLA2L were derived from the human orthologue of mouse Otoconin-90 (PLA2L/OC90), a major protein in the otoconia of the inner ear, which are vital for the sense of gravity (124). The transcript is the product of intergenic splicing between a HERV-H element and two downstream genes normally independently expressed from different promoters, initiated in the HERV-H LTR, and spliced from the major viral splice donor downstream into HHLA1 (HERV-H LTR-associated gene), followed by a second splice into PLA2L/OC90 (65).

As noted above, HERVs are less likely to be found in introns than in intergenic regions, and transcriptionally active intronic HERVs are more frequently found in the antisense orientation relative to the transcriptional direction of the enclosing gene (121). The increasing bias against sense orientation with increasing age of the elements is consistent with the model that ERV transcripts may serve as antisense protectors of cryptic chromosomal intronic splice sites. Briefly, if an ERV is integrated into an intron, in the sense orientation, it may introduce new canonical splice sites as previously identified for MLV in the dilute (*d*) coat color locus of mice (**Table 1**; **Figure 5**) (48), and can thus interfere with normal gene function. Usually, such effects will be deleterious, and the

Gene conversion: nonreciprocal recombination in which genetic information is copied from one allele or repeat element to another

Ectopic recombination: atypical genetic rearrangement that occurs between similar DNA segments in chromosomes

ERV will be selected against. ERVs integrated in the antisense orientation within introns are less likely to have functional splice or poly(A) sites, and even if present, access to them may be blocked by the synthesis of transcripts originating in the LTR.

Shaping of the Genome

Beyond effects on gene expression, endogenous proviruses have also played significant roles in the organization of the host genome. The most prominent mechanisms involve recombination between identical sequences either within an element or between related elements. Recombination can occur in several ways (**Figure 2**). (*i*) Coding regions of an integrated provirus may become lost after homologous recombination between the two LTRs, leaving a solo LTR at the locus (44, 53, 113). Solo LTRs are present 10 to 100 times more frequently than their cognate ancestral proviruses (113). In the case of the *dilute* provirus in mice, solo LTR formation causes a readily detectable difference in coat color, and it has been possible to estimate the rate of this event at about 4.5×10^{-6} per generation (109). Recent studies have shown that recombination-mediated solo LTR formation occurs more rapidly soon after integration than after mutations have accumulated in the proviral LTRs (10), and that the persistence of proviruses is dependent on the recombination rate and tolerance in the host's genome (61). (*ii*) Homologous recombination between two proviruses in the same orientation on the same chromosome results in loss of viral and genetic sequence between recombination sites. If they are in opposite orientation, the result is an inversion of the intervening chromosomal region. (*iii*) Recombination between 3′ and 5′ LTRs of a given provirus on sister chromatids results in a tandem provirus (two proviruses flanked by LTRs while sharing one LTR) on one chromatid, and a solo LTR on the other. (*iv*) Gene conversion results in nonreciprocal exchange of sequences without proviral loss in such a way that all or a portion of one proviral sequence is converted to the sequence of the other (71).

Such recombination events are not unique to ERVs; any repeated sequences of the same size and distribution should be subject to exactly the same mechanisms of rearrangement. However, unique properties of retrovirus replication make ERVs powerful and sensitive tools for revealing and quantifying such events. These properties include the identity of the LTRs at the time of integration, the creation of short duplications of host DNA on either side of the provirus, the enormous number of potential sites of integration in the host genome, and the lack of disruption of flanking host sequence. Application of these facts to phylogenetic analyses of HERV-K(HML2) LTRs in the human genome has made it possible to detect ectopic recombination in more than 16% of them, corresponding to large chromosomal rearrangements occurring and being fixed at a rate of once per provirus per 80 My (45). Such a crossover between two HERV-I loci on the Y chromosome appears to be a cause of inherited male infertility due to loss of the 792-kb fragment that contains the Azoospermia factor, AZFa (57). Recombination events mediated by ERV proviruses may also have provided useful genomic plasticity. The density of repetitive elements including HERVs in some gene loci, such as the human MHC classes I and II genes, compared to other gene regions that are more or less free of ERV integrations, contributes to plasticity of these gene clusters and their resulting immunohaplotypes (3, 67).

Given the moderate set of examples of ERV-mediated recombination events in crucial genomic regions, it seems probable that they have had a profound overall effect in shuffling of genomic regions, exons, and regulatory information into new contexts and thereby altering the dynamic functions of the host genome.

CONCLUDING REMARKS

Retroviruses are unique among infectious agents in their ability to establish themselves as inherited DNA elements in the form of ERVs, and unique among inherited DNA elements in their potential for transmission from

individual to individual and species to species as infectious agents. As whole animal genome sequences become increasingly available, the wide range of evolutionary phenomenology related to ERVs is being revealed. Clearly, retroviruses are very old: Even the most ancient of proviruses bear all the features associated with modern retroviruses, and they have most likely existed since—and perhaps well before—the dawn of vertebrate evolution. Given the very large numbers of ERVs, represented both as nearly full-length and fragmentary proviruses—estimated at some 100,000 in the human genome (96) of which solo LTRs constitute the large majority, and more or less complete HERVs are estimated at some 3500 (96, 110)—their role in shaping the genome must have been very large, and it is likely that we have only scratched the surface with the examples discussed in this review.

The evolutionary forces that have led to the impressive accumulation of these elements in germline DNA are only poorly understood. Clearly, there is a balance between positive, negative, and neutral selective influences. On the positive side is the expression of viral gene products as useful new genes. On the negative side is the potential for gene disruption or misexpression resulting from ERV integration, as well as the potential for somatic spread of replicating virus leading to pathogenic consequences. The apparently increasing paucity of proviruses integrated within genes with increasing evolutionary age is a sign of negative selection.

The large number of provirus-derived sequences in all animal genomes argues strongly that their most important source is neutral or nearly neutral accumulation resulting from the infection of germline cells with viruses replicating in the host at the time. Although a few retrovirus-like elements, such as IAPs and MusD, have clearly devolved from viruses into intracellular retrotransposable elements (104), the large majority of animal LTR elements appear to be fossilized viruses whose DNA was inserted into the germline following infection. There are three possible origins for the viruses that become ERVs: They are derived directly from other ERVs in the same individual; they are derived from exogenous viruses enzootic in the host species; or (as in modern-day koalas) they are derived from enzootic viruses recently transmitted from another host species. For the reasons stated above, we think it most useful to view ERVs from the standpoint of fossil representatives of retroviruses extant at the time of their insertion into the germline, rather than their role as direct players in the evolutionary process itself. With a few exceptions, the evolutionary forces of most importance in shaping the genomes of ERVs are most likely to be those acting through somatic replication of the virus. Similarly, the evolution of inhibitory host genes—from receptor mutations to inhibitory genes like APOBEC3, TRIM5, and Fv1—is most likely to have been driven by selective pressure exerted by somatic replication of the cognate viruses, not by effects of germline integration. Although the important, interesting, and varied effects of endogenous provirus integration on the genome of all vertebrates can long outlive the viruses that gave rise to them, full understanding and appreciation of the evolutionary processes involved demands that we always keep the virus in mind.

SUMMARY POINTS

1. Over most or all of their evolutionary history, mammalian genomes have encountered infecting retroviruses. Some of these have remained as genetic parasites, remnants of which constitute some 8% of the human genome today.
2. ERV-mediated recombination events have had profound effects in the shaping of the host's genome, and new ERV integrations introduce added variation to the host transcriptomes.

3. Expression of ERVs has been associated with several positive physiological functions as well as certain diseases. Although their roles as an etiological agents, possible contributing factors, or markers of disease have been well established in experimental animals, they remain to be established in humans.
4. Although ERVs have clearly played important roles in evolution, it most useful to view them as fossil representatives of retroviruses extant at the time of their insertion into the germline, and their evolutionary roles as secondary to virological events.

ACKNOWLEDGMENTS

P.J. is a recipient of a postdoctoral fellowship from the Wenner-Gren Foundation. J.M.C. is a Research Professor of the American Cancer Society, with support from the George W. Kirby Foundation, and was supported by grant R37 CA 089441 from the National Cancer Institute. Space constraints made it impossible to cite all relevant publications in this review; our sincere apologies and appreciation to all colleagues whose important work is not cited.

LITERATURE CITED

1. Acha-Orbea H, MacDonald HR. 1995. Superantigens of mouse mammary tumor virus. *Annu. Rev. Immunol.* 13:459–86
2. Andersson AC, Venables PJ, Tonjes RR, Scherer J, Eriksson L, Larsson E. 2002. Developmental expression of HERV-R (ERV3) and HERV-K in human tissue. *Virology* 297:220–25
3. Andersson G, Svensson AC, Setterblad N, Rask L. 1998. Retroelements in the human MHC class II region. *Trends Genet.* 14:109–14
4. Antony JM, Van Marle G, Opii W, Butterfield DA, Mallet F, et al. 2004. Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination. *Nat. Neurosci.* 7:1088–95
5. Arnaud F, Caporale M, Varela M, Biek R, Chessa B, et al. 2007. A paradigm for virus-host coevolution: sequential counter-adaptations between endogenous and exogenous retroviruses. *PLoS Pathog.* 3:e170
6. Bannert N, Kurth R. 2006. The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genomics Hum. Genet.* 7:149–73
7. Bartholomew C, Ihle JN. 1991. Retroviral insertions 90 kilobases proximal to the Evi-1 myeloid transforming gene activate transcription from the normal promoter. *Mol. Cell Biol.* 11:1820–28
8. Belshaw R, Katzourakis A, Paces J, Burt A, Tristem M. 2005. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol. Biol. Evol.* 22:814–17
9. Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, et al. 2004. Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. USA* 101:4894–99
10. Belshaw R, Watson J, Katzourakis A, Howe A, Woolven-Allen J, et al. 2007. Rate of recombinational deletion among human endogenous retroviruses. *J. Virol.* 81:9437–42
11. Benit L, Calteau A, Heidmann T. 2003. Characterization of the low-copy HERV-Fc family: evidence for recent integrations in primates of elements with coding envelope genes. *Virology* 312:159–68
12. Benson SJ, Ruis BL, Fadly AM, Conklin KF. 1998. The unique envelope gene of the subgroup J avian leukosis virus derives from ev/J proviruses, a novel family of avian endogenous viruses. *J. Virol.* 72:10157–64
13. Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH. 2004. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Curr. Biol.* 14:1392–96

14. Blaise S, de Parseval N, Benit L, Heidmann T. 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. USA* 100:13013–18
15. Blaise S, de Parseval N, Heidmann T. 2005. Functional characterization of two newly identified Human Endogenous Retrovirus coding envelope genes. *Retrovirology* 2:19
16. Blomberg J, Ushameckis D, Jern P. 2005. Evolutionary aspects of human endogenous retroviral sequences (HERVs) and disease. See Ref. 117a, pp. 227–62
17. Bock M, Stoye JP. 2000. Endogenous retroviruses and the human germline. *Curr. Opin. Genet. Dev.* 10:651–55
18. Boeke JD, Stoye JP. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements. See Ref. 22a, pp. 343–436
19. Buzdin A, Ustyugova S, Khodosevich K, Mamedov I, Lebedev Y, et al. 2003. Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. *Genomics* 81:149–56
20. Chi KD, McPhee RA, Wagner AS, Dietz JJ, Pantazis P, Goustin AS. 1997. Integration of proviral DNA into the PDGF beta-receptor gene in HTLV-I-infected T-cells results in a novel tyrosine kinase product with transforming activity. *Oncogene* 15:1051–57
21. Cianciolo GJ, Copeland TD, Oroszlan S, Snyderman R. 1985. Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins. *Science* 230:453–55
22. Coffin JM. 1994. Structure and Classification of Retroviruses. In *The Retroviridae*, ed. J Levy, pp. 19–49. New York: Plenum
22a. Coffin JM, Hughes SH, Varmus HE, eds. 1997. *Retroviruses*. New York: Cold Spring Harbor Lab. Press
23. Coffin JM, Tsichlis PN, Conklin KF, Senior A, Robinson HL. 1983. Genomes of endogenous and exogenous avian retroviruses. *Virology* 126:51–72
24. Datta SK, Owen FL, Womack JE, Riblet RJ. 1982. Analysis of recombinant inbred lines derived from "autoimmune" (NZB) and "high leukemia" (C58) strains: independent multigenic systems control B cell hyperactivity, retrovirus expression, and autoimmunity. *J. Immunol.* 129:1539–44
25. de Parseval N, Casella J, Gressin L, Heidmann T. 2001. Characterization of the three HERV-H proviruses with an open envelope reading frame encompassing the immunosuppressive domain and evolutionary history in primates. *Virology* 279:558–69
26. de Parseval N, Heidmann T. 1998. Physiological knockout of the envelope gene of the single-copy ERV-3 human endogenous retrovirus in a fraction of the Caucasian population. *J. Virol.* 72:3442–45
27. Deb-Rinker P, Klempan TA, O'Reilly RL, Torrey EF, Singh SM. 1999. Molecular characterization of a MSRV-like sequence identified by RDA from monozygotic twin pairs discordant for schizophrenia. *Genomics* 61:133–44
28. Dewannieux M, Blaise S, Heidmann T. 2005. Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses. *J. Virol.* 79:15573–77
29. Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, et al. 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.* 16:1548–56
30. Di Cristofano A, Strazullo M, Longo L, La Mantia G. 1995. Characterization and genomic mapping of the ZNF80 locus: Expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Res.* 23:2823–30
31. Dong B, Kim S, Hong S, Das Gupta J, Malathi K, et al. 2007. An infectious retrovirus susceptible to an IFN antiviral pathway from human prostate tumors. *Proc. Natl. Acad. Sci. USA* 104:1655–60
32. Dunlap KA, Palmarini M, Varela M, Burghardt RC, Hayashi K, et al. 2006. Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proc. Natl. Acad. Sci. USA* 103:14390–95
33. Dunn CA, Romanish MT, Gutierrez LE, van de Lagemaat LN, Mager DL. 2006. Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene* 366:335–42
34. Dupressoir A, Marceau G, Vernochet C, Benit L, Kanellopoulos C, et al. 2005. Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc. Natl. Acad. Sci. USA* 102:725–30
35. Esnault C, Heidmann O, Delebecque F, Dewannieux M, Ribet D, et al. 2005. APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* 433:430–33

36. Feuchter A, Mager D. 1990. Functional heterogeneity of a large family of human LTR-like promoters and enhancers. *Nucleic Acids Res.* 18:1261–70
37. Feuchter-Murthy AE, Freeman JD, Mager DL. 1993. Splicing of a human endogenous retrovirus to a novel phospholipase A2 related gene. *Nucleic Acids Res.* 21:135–43
38. Forsman A, Yun Z, Hu L, Uzhameckis D, Jern P, Blomberg J. 2005. Development of broadly targeted human endogenous gammaretroviral pol-based real time PCRs quantitation of RNA expression in human tissues. *J. Virol. Methods* 129:16–30
39. Garrison KE, Jones RB, Meiklejohn DA, Anwar N, Ndhlovu LC, et al. 2007. T cell responses to human endogenous retroviruses in HIV-1 infection. *PLoS Pathog.* 3:e165
40. Gifford R, Tristem M. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* 26:291–315
41. Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, et al. 2003. DNA deamination mediates innate immunity to retroviral infection. *Cell* 113:803–9
42. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A. 2008. Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice. *Oncogene* 27:404–8
43. Hughes JF, Coffin JM. 2001. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat. Genet.* 29:487–89
44. Hughes JF, Coffin JM. 2004. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc. Natl. Acad. Sci. USA* 101:1668–72
45. Int. Human Genome Sequencing Consort. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860–921
46. Izui S, Elder JH, McConahey PJ, Dixon FJ. 1981. Identification of retroviral gp70 and antigp70 antibodies involved in circulating immune complexes in NZB X NZW mice. *J. Exp. Med.* 153:1151–60
47. Jenkins NA, Copeland NG. 1985. High frequency germline acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. *Cell* 43:811–19
48. Jenkins NA, Copeland NG, Taylor BA, Lee BK. 1981. Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. *Nature* 293:370–74
49. Jern P, Sperber GO, Blomberg J. 2004. Definition and variation of human endogenous retrovirus H. *Virology* 327:93–110
50. Jern P, Sperber GO, Blomberg J. 2005. Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology* 2:50
51. Jern P, Sperber GO, Blomberg J. 2006. Divergent patterns of recent retroviral integrations in the human and chimpanzee genomes: probable transmissions between other primates and chimpanzees. *J. Virol.* 80:1367–75
52. Jern P, Stoye JP, Coffin JM. 2007. Role of APOBEC3 in genetic diversity among endogenous murine leukemia viruses. *PLoS Genet.* 3:2014–22
53. Johnson WE, Coffin JM. 1999. Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. USA* 96:10254–60
54. Johnston JB, Silva C, Holden J, Warren KG, Clark AW, Power C. 2001. Monocyte activation and differentiation augment human endogenous retrovirus expression: implications for inflammatory brain diseases. *Ann. Neurol.* 50:434–42
55. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110:462–67
56. Kaiser SM, Malik HS, Emerman M. 2007. Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein. *Science* 316:1756–58
57. Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH. 2000. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum. Mol. Genet.* 9:2563–72
58. Kapitonov VV, Jurka J. 1999. 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.* 48:248–51

59. Karlsson H, Bachmann S, Schroder J, McArthur J, Torrey EF, Yolken RH. 2001. Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proc. Natl. Acad. Sci. USA* 98:4634–39
60. Kato N, Shimotohno K, VanLeeuwen D, Cohen M. 1990. Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger protein related to Kruppel. *Mol. Cell Biol.* 10:4401–5
61. Katzourakis A, Pereira V, Tristem M. 2007. Effects of recombination rate on human endogenous retrovirus fixation and persistence. *J. Virol.* 81:10712–17
62. Katzourakis A, Rambaut A, Pybus OG. 2005. The evolutionary dynamics of endogenous retroviruses. *Trends Microbiol.* 13:463–68
63. Katzourakis A, Tristem M. 2005. Phylogeny of human endogenous and exogenous retroviruses. See Ref. 117a, pp. 186–203
64. Kohn DB, Sadelain M, Glorioso JC. 2003. Occurrence of leukaemia following gene therapy of X-linked SCID. *Nat. Rev. Cancer* 3:477–88
65. Kowalski PE, Freeman JD, Mager DL. 1999. Intergenic splicing between a HERV-H endogenous retrovirus and two adjacent human genes. *Genomics* 57:371–79
66. Kozak CA, Gromet NJ, Ikeda H, Buckler CE. 1984. A unique sequence related to the ecotropic murine leukemia virus is associated with the Fv-4 resistance gene. *Proc. Natl. Acad. Sci. USA* 81:834–37
67. Kulski JK, Anzai T, Inoko H. 2005. ERVK9, transposons and the evolution of MHC class I duplicons within the alpha-block of the human and chimpanzee. *Cytogenet. Genome Res.* 110:181–92
68. Larsson E, Kato N, Cohen M. 1989. Human endogenous proviruses. *Curr. Top. Microbiol. Immunol.* 148:115–32
69. Lavillette D, Marin M, Ruggieri A, Mallet F, Cosset FL, Kabat D. 2002. The envelope glycoprotein of human endogenous retrovirus type W uses a divergent family of amino acid transporters/cell surface receptors. *J. Virol.* 76:6442–52
70. Leib-Mösch C, Seifarth W, Schön U. 2005. Influence of human endogenous retroviruses on cellular gene expression. See Ref. 117a, pp. 123–43
71. Li WH. 1997. *Molecular Evolution*. Sunderland, MA:
72. Long Q, Bengra C, Li C, Kutlar F, Tuan D. 1998. A long terminal repeat of the human endogenous retrovirus ERV-9 is located in the 5′ boundary area of the human beta-globin locus control region. *Genomics* 54:542–55
73. Lukacher AE, Ma Y, Carroll JP, Abromson-Leeman SR, Laning JC, et al. 1995. Susceptibility to tumors induced by polyoma virus is conferred by an endogenous mouse mammary tumor virus superantigen. *J. Exp. Med.* 181:1683–92
74. Macfarlane C, Simmonds P. 2004. Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. *J. Mol. Evol.* 59:642–56
75. Mager DL, Freeman JD. 1995. HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. *Virology* 213:395–404
76. Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. 2006. Retroviral elements and their hosts: insertional mutagenesis in the mouse germline. *PLoS Genet.* 2:e2
77. Mallet F, Bouton O, Prudhomme S, Cheynet V, Oriol G, et al. 2004. The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology. *Proc. Natl. Acad. Sci. USA* 101:1731–36
78. Mamedov I, Lebedev Y, Hunsmann G, Khusnutdinova E, Sverdlov E. 2004. A rare event of insertion polymorphism of a HERV-K LTR in the human genome. *Genomics* 84:596–99
79. Mangeney M, de Parseval N, Thomas G, Heidmann T. 2001. The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties. *J. Gen. Virol.* 82:2515–18
80. Mangeney M, Heidmann T. 1998. Tumor cells expressing a retroviral envelope escape immune rejection in vivo. *Proc. Natl. Acad. Sci. USA* 95:14920–25
81. Mangeney M, Renard M, Schlecht-Louf G, Bouallaga I, Heidmann O, et al. 2007. Placental syncytins: genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl. Acad. Sci. USA* 104:20534–39
82. Martin J, Herniou E, Cook J, Waugh O'Neill R, Tristem M. 1997. Human endogenous retrovirus type I–related viruses have an apparently widespread distribution within vertebrates. *J. Virol.* 71:437–43

83. Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, Meese E. 1999. An almost-intact human endogenous retrovirus K on human chromosome 7. *Nat. Genet.* 21:257–58
84. Medstrand P, Landry JR, Mager DL. 2001. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J. Biol. Chem.* 276:1896–903
85. Medstrand P, Mager DL. 1998. Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* 72:9782–87
86. Medstrand P, van de Lagemaat LN, Mager DL. 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* 12:1483–95
87. Melana SM, Nepomnaschy I, Sakalian M, Abbott A, Hasa J, et al. 2007. Characterization of viral particles isolated from primary cultures of human breast cancer cells. *Cancer Res.* 67:8960–65
88. Mi S, Lee X, Li X, Veldman GM, Finnerty H, et al. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403:785–89
89. Mikkelsen JG, Lund AH, Kristensen KD, Duch M, Sorensen MS, et al. 1996. A preferred region for recombinational patch repair in the 5′ untranslated region of primer binding site-impaired murine leukemia virus vectors. *J. Virol.* 70:1439–47
90. Moller-Larsen A, Christensen T. 1998. Isolation of a retrovirus from multiple sclerosis patients in self-generated Iodixanol gradients. *J. Virol. Methods* 73:151–61
91. Monitto CL, Berkowitz D, Lee KM, Pin S, Li D, et al. 2001. Differential gene expression in a murine model of cancer cachexia. *Am. J. Physiol. Endocrinol. Metab.* 281:E289–97
92. Morgan HD, Sutherland HG, Martin DI, Whitelaw E. 1999. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23:314–18
93. Moyes D, Griffiths DJ, Venables PJ. 2007. Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. *Trends Genet.* 23:326–33
94. Muster T, Waltenberger A, Grassauer A, Hirschl S, Caucig P, et al. 2003. An endogenous retrovirus derived from human melanoma cells. *Cancer Res.* 63:8735–41
95. Okeoma CM, Lovsin N, Peterlin BM, Ross SR. 2007. APOBEC3 inhibits mouse mammary tumour virus replication in vivo. *Nature* 445:927–30
96. Paces J, Pavlicek A, Paces V. 2002. HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.* 30:205–6
97. Palmarini M, Gray CA, Carpenter K, Fan H, Bazer FW, Spencer TE. 2001. Expression of endogenous betaretroviruses in the ovine uterus: effects of neonatal age, estrous cycle, pregnancy, and progesterone. *J. Virol.* 75:11319–27
98. Palmarini M, Mura M, Spencer TE. 2004. Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation. *J. Gen. Virol.* 85:1–13
99. Perron H, Garson JA, Bedin F, Beseme F, Paranhos-Baccala G, et al. 1997. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. The Collaborative Research Group on Multiple Sclerosis. *Proc. Natl. Acad. Sci. USA* 94:7583–88
100. Polavarapu N, Bowen NJ, McDonald JF. 2006. Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biol.* 7:R51
101. Reik W. 2007. Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447:425–32
102. Reiss D, Mager DL. 2007. Stochastic epigenetic silencing of retrotransposons: Does stability come with age? *Gene* 390:130–35
103. Reiss D, Zhang Y, Mager DL. 2007. Widely variable endogenous retroviral methylation levels in human placenta. *Nucleic Acids Res.* 35:4743–54
104. Ribet D, Harper F, Dupressoir A, Dewannieux M, Pierron G, Heidmann T. 2008. An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* 18:597–609
105. Romanish MT, Lock WM, van de Lagemaat LN, Dunn CA, Mager DL. 2007. Repeated recruitment of LTR retrotransposons as promoters by the antiapoptotic locus NAIP during mammalian evolution. *PLoS Genet.* 3:e10
106. Rosenberg N, Jolicoeur P. 1997. Retroviral pathogenesis. See Ref. 22a, pp. 475–586

107. Samuelson LC, Wiebauer K, Snow CM, Meisler MH. 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol. Cell Biol.* 10:2513–20
108. Schulte AM, Lai S, Kurtz A, Czubayko F, Riegel AT, Wellstein A. 1996. Human trophoblast and choriocarcinoma expression of the growth factor pleiotrophin attributable to germ-line insertion of an endogenous retrovirus. *Proc. Natl. Acad. Sci. USA* 93:14759–64
109. Seperack PK, Strobel MC, Corrow DJ, Jenkins NA, Copeland NG. 1988. Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice. *Proc. Natl. Acad. Sci. USA* 85:189–92
110. Sperber GO, Airola T, Jern P, Blomberg J. 2007. Automated recognition of retroviral sequences in genomic data–RetroTector. *Nucleic Acids Res.* 35:4964–76
111. Stauffer Y, Theiler G, Sperisen P, Lebedev Y, Jongeneel CV. 2004. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immun.* 4:2
112. Stoye JP. 1998. Fv1, the mouse retrovirus resistance gene. *Rev. Sci. Tech.* 17:269–77
113. Stoye JP. 2001. Endogenous retroviruses: still active after all these years? *Curr. Biol.* 11:R914–16
114. Stoye JP, Coffin JM. 1988. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. *J. Virol.* 62:168–75
115. Stoye JP, Coffin JM. 2000. Reproductive biology: a provirus put to work. *Nature* 403:715–17
116. Stoye JP, Moroni C, Coffin JM. 1991. Virological events leading to spontaneous AKR thymomas. *J. Virol.* 65:1273–85
117. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. 2004. The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. *Nature* 427:848–53
117a. Sverdlov ED, ed. 2005. *Retroviruses and Primate Genome Evolution*. Georgetown, TX: Eurekah.com/ Landes Biosci.
118. Tarlinton RE, Meers J, Young PR. 2006. Retroviral invasion of the koala genome. *Nature* 442:79–81
119. Tsichlis PN, Coffin JM. 1980. Recombinants between endogenous and exogenous avian tumor viruses: role of the C region and other portions of the genome in the control of replication and transformation. *J. Virol.* 33:238–49
120. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* 19:530–36
121. van de Lagemaat LN, Medstrand P, Mager DL. 2006. Multiple effects govern endogenous retrovirus survival patterns in human gene introns. *Genome Biol.* 7:R86
122. Van Der Kuyl AC, Dekker JT, Goudsmit J. 1996. Baboon endogenous virus evolution and ecology. *Trends Microbiol.* 4:455–59
123. Voisset C, Weiss RA, Griffiths DJ. 2008. Human RNA "rumor" viruses: the search for novel human retroviruses in chronic disease. *Microbiol. Mol. Biol. Rev.* 72:157–96
124. Wang Y, Kowalski PE, Thalmann I, Ornitz DM, Mager DL, Thalmann R. 1998. Otoconin-90, the mammalian otoconial matrix protein, contains two domains of homology to secretory phospholipase A2. *Proc. Natl. Acad. Sci. USA* 95:15345–50
125. Wang-Johanning F, Frost AR, Jian B, Epp L, Lu DW, Johanning GL. 2003. Quantitation of HERV-K env gene expression and splicing in human breast cancer. *Oncogene* 22:1528–35
126. Weiss RA. 1993. Cellular receptors and viral glycoproteins involved in retrovirus entry. In *The Retroviridae*, ed. JA Levy, pp. 1–108. New York: Plenum
127. Whitelaw E, Martin DI. 2001. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet.* 27:361–65
128. Wootton SK, Halbert CL, Miller AD. 2005. Sheep retrovirus structural protein induces lung tumours. *Nature* 434:904–7
129. Yap MW, Nisole S, Lynch C, Stoye JP. 2004. Trim5alpha protein restricts both HIV-1 and murine leukemia virus. *Proc. Natl. Acad. Sci. USA* 101:10786–91
130. Yohn CT, Jiang Z, McGrath SD, Hayden KE, Khaitovich P, et al. 2005. Lineage-specific Expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLoS Biol.* 3:1–11