Rockefeller University
Digital Commons @ RU

Student Theses and Dissertations

2010

Reconstitution and Characterization of Human Endogenous Retrovirus-K

Young Nam Lee

Follow this and additional works at: http://digitalcommons.rockefeller.edu/student_theses_and_dissertations

 Part of the Life Sciences Commons

Recommended Citation

Lee, Young Nam, "Reconstitution and Characterization of Human Endogenous Retrovirus-K" (2010). *Student Theses and Dissertations.* Paper 78.

This Thesis is brought to you for free and open access by Digital Commons @ RU. It has been accepted for inclusion in Student Theses and Dissertations by an authorized administrator of Digital Commons @ RU. For more information, please contact mcsweej@mail.rockefeller.edu.

RECONSTITUTION AND CHARACTERIZATION OF HUMAN ENDOGENOUS RETROVIRUS-K

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Young Nam Lee

June 2010

© Copyright by Young Nam Lee 2010

RECONSTITUTION AND CHARACTERIZATION OF HUMAN ENDOGENOUS RETROVIRUS-K

Young Nam Lee, Ph.D.

The Rockefeller University 2010

Retroviruses are a family of clinically significant and scientifically fascinating viruses that infect a wide array of organisms from all vertebrate classes. The two hallmark events in the life cycle of retroviruses are the reverse transcription of the single stranded RNA (ssRNA) genome generating a double stranded DNA (dsDNA) and the integration of this dsDNA into the host genome. Because integration is irreversible and the infected cells are usually difficult to target for elimination in the host, the infection is generally permanent. HIV-1, the most important and well-studied member of all retroviruses, is the causative agent of acquired immune deficiency syndrome (AIDS) for which no vaccine or cure is known. Since recognition of the AIDS epidemic, around 25 million people have died from HIV-1 related causes, including 2 million in 2007. Currently, 33 million people are believed to be living with the virus, with most of these people living in sub-Saharan Africa, where 67% of all infected people reside and 75% of AIDS deaths occurred in 2007.

When retroviruses infect germ cells or germ cell progenitors, the virus can become endogenized. These viruses, called endogenous retroviruses (ERV), make up more than 8% of the human genome. The integrated virus will be present in the genome

of all cells of the individual derived from the infected germ cell, and be passed on to progeny in a Mendelian manner to following generations. Both chance and the insertion's effect on the fitness of the host can determine the allelic frequency in the population. Hence, elements which produce large quantities of viral proteins and progeny or elements that insert into a necessary gene will likely reduce the fitness of the host and as an allele will be negatively selected in the host population.

Currently, there is no known replication competent HERV, as most proviruses are filled with deletions and premature stop codons. However, one family of Class II HERV, HERV-K(HML-2), seems to have been replicating until recently. The HERV-K(HML-2) family includes human specific members and elements that are polymorphic in the human population, suggesting replication since the divergence of humans from chimpanzees 6 million years ago and potentially more recently as well.

In this body of work, the problem of the lack of a replication competent virus sequence is circumvented by deducing a consensus sequence from the youngest set of HERV proviruses. Named HERV-K_{CON}, we find that many of its components are functional individually and together enable infection of target cells in a single-cycle infection system. Using this system, we have characterized the previously unknown aspects of HERV-K(HML-2) life cycle, such as location of assembly and budding, dependency on cell replication, and more extensively, its integration site preference. HERV-K_{CON}'s interaction with current anti-retroviral host proteins is accessed, and evidence of the same interaction occurring in vivo is presented in the context of APOBEC3G.

I would like to dedicate this work to my family. My parents, who started with very little, have always provided me with a supportive and loving environment, and instilled in me the value of education and hard work. My sister has always been kind, and despite our divergent professional and personal paths, still represents in many ways what I strive toward. They enabled everything I have done.

ACKNOWLEDGMENTS

I would like to acknowledge the people my laboratories, past and present, for their scientific and personal support. Thank you for making late nights and weekends fun, in and out of lab.

I would also like to acknowledge my friends. Thank you for your patience, your company, and your support. You made the unbearable moments pass and happier moments last.

I also appreciate the help and support of my committee, Dr. Charles Rice, Dr. Jack Lenz, and Dr. Nina Papavasiliou, and my collaborators, Dr. Mike Malim (UCL), Dr. Fredrick Bushman (University of Pennsylvania), and Dr. Troy Brady (University of Pennsylvania).

The Dean's office members and the administrative staff of Rockefeller University and ADARC have been tremendously helpful over the years in multiple aspects of my life. Thank you.

Lastly, I would like to acknowledge the city. I love New York. It is the only home I have known in my adult life. I pray that I will be back one day.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS

Brady T.L., **Lee Y.N.**, Ronen K., Malani N., Berry C.C., Bieniasz P.D., Bushman, F.D. (2009) Integration target site selection by a resurrected human endogenous retrovirus. Genes and Development. 23(5):633-42.

Jouvenet N., Neil S.J., Zhadina M., Zang T., Kratovac Z., **Lee Y.N.**, McNatt M., Hatziioannou T., Bieniasz P.D. (2009) Broad-spectrum inhibition of retroviral and filoviral particle release by tetherin. Journal of Virology. 83:1837-44.

Lee Y.N., Malim M.H., and Bieniasz P.D. (2008) Hypermutation of an ancient human retrovirus by APOBEC3G. Journal of Virology. 82:8762-70.

Lee Y.N. and Bieniasz P.D. (2007) Reconstitution of an infectious human endogenous retrovirus. PLoS Pathogens. 3:e10.

I. INTRODUCTION

1. Retroviridae

Retroviruses are a family of clinically significant and scientifically fascinating viruses that infect a wide array of organisms from all vertebrate classes [reviewed in (Goff, 2007)]. The two hallmark events in the life cycle of retroviruses are the reverse transcription of the single stranded RNA (ssRNA) genome generating a double stranded DNA (dsDNA) and the integration of this dsDNA into the host genome (Baltimore, 1970; Temin and Mizutani, 1970). Because integration is irreversible and the infected cells are usually difficult to target for elimination in the host, the infection is generally permanent.

The classification of retroviruses is based on phylogenetic analysis using the highly conserved RT sequence (Figure 1) [reviewed in (Goff, 2007)]. Alpharetroviruses are avian retroviruses that are often oncogenic, such as Rous sarcoma virus (RSV). Betaretroviruses include protoptype members mouse mammary tumor virus (MMTV) and Mason Pfizer monkey virus (MPMV), which both assemble at a perinuclear region, then are transported to the plasma membrane for release (Chopra and Mason, 1970; Jensen et al., 1970; Rhee and Hunter, 1990a). Gammaretroviruses has the largest number of members and also is the only genus that infects multiple vertebrate classes (Gifford and Tristem, 2003; Goff, 2007). The most widely studied member is the group of murine leukemia viruses (MLV), which can be divided into subgroups based on their receptor usage (Goff, 2007). Deltaretroviruses infect primates and cattle, and contain multiple accessory proteins (Goff, 2007). Human T-lymphotropic virus 1 (HTLV-1), the prototype member that was the first human retrovirus discovered, causes T cell leukemia and lymphoma in humans (Poiesz et al., 1980; Yoshida et al., 1982). Epsilonretroviruses

infect fish and reptiles and have large and complex genomes up to 13kb in length, with the walleye dermal sarcoma virus as the prototype (Martineau et al., 1992; Martineau et al., 1990). Spumaretroviruses, as exemplified by chimpanzee foamy virus, are often highly cytopathic in cell culture, but are not associated with any disease in host (Linial, 1999). Lentiviruses have a unique cone shaped core morphology that distinguishes them in electron microscopy from other retroviruses (Goff, 2007). The name is derived from the slow infection and disease progression in the infected hosts from the Latin word *lentus*, which means "slow" (Desrosiers, 2007). The group can be roughly divided into primate and nonprimate members, with the highly related human immunodeficiency viruses (HIVs) and simian immunodeficiency viruses (SIVs) making up the primate group.

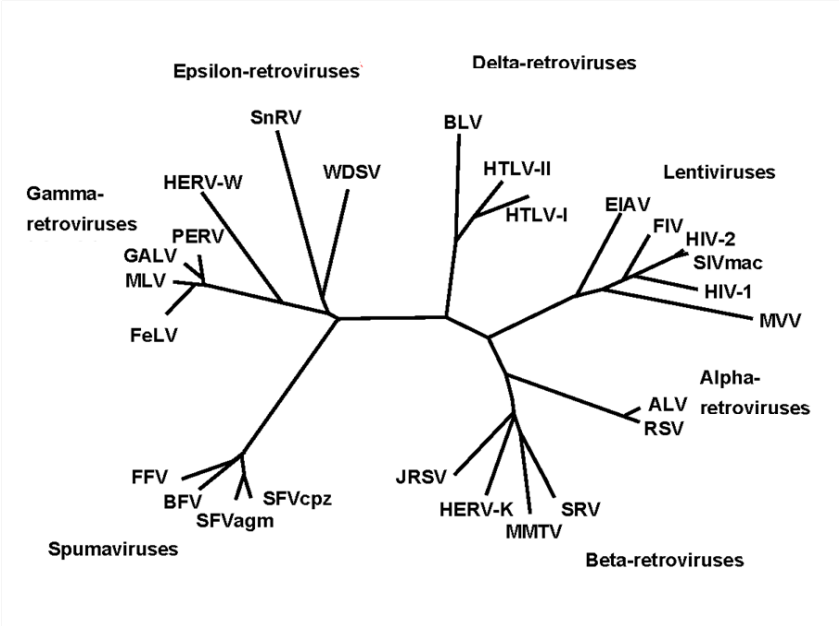


Figure 1 Phylogenetic analysis of Retrovirus family

An unrooted neighbor-joining phylogenetic tree based on the amino acid sequence of RT protein. SnRV, snake retrovirus. WDSV, walleye dermal sarcoma virus. PERV, porcine endogenous retrovirus. GALV, gibbon ape leukemia virus. FeLV, feline leukemia virus. FFV, feline foam virus. BFV, bovine foamy virus. SFV, simian foamy virus. SRV, squirrel monkey retrovirus. BLV, bovine leukemia virus. Other viruses are discussed in the text. This figure is taken from Weiss (Weiss, 2006).

HIV-1, the most important and well-studied member of all retroviruses, is the causative agent of acquired immune deficiency syndrome (AIDS) for which no vaccine or cure is known. In 1981, Centers for Disease Control and Prevention reported five cases of *Pneumocystis carinii* pneumonia in young, homosexual men, a disease that normally affects only severely immunosupressed patients (CDC, 1981). The causative agent was identified by two separate groups in 1983 and 1984 from patients exhibiting AIDS or pre-AIDS symptoms (Barre-Sinoussi et al., 1983; Popovic et al., 1984). Since recognition of the AIDS epidemic, around 25 million people have died from HIV-1 related causes, including 2 million in 2007 (UNAIDS, 2008). Currently, 33 million people are believed to be living with the virus, with most of these people living in sub-Saharan Africa, where 67% of all infected people reside and 75% of AIDS deaths occurred in 2007 (UNAIDS, 2008).

Retrovirus genomes range between 7kb to 13kb in length, but the basic genomic features are common to all retroviruses (Figure 2). Two copies of the long terminal repeat (LTR) that flank the ends of the viral genome are identical at the time of integration. They contain important cis elements such as the promoter sequence, TATA box, and polyA signal. The repeat nature of LTR is important in the process of reverse transcription, and ensures that the entire viral sequence is copied to following virus generations. Between the LTRs are the three protein coding open reading frames (ORFs), *gag*, *pol*, and *env*. The *gag* ORF encodes the structural protein domains that target Gag to a membrane (matrix, MA), form the virus core (capsid, CA), and encapsidate the viral genome (nucleocapsid, NC), and is sufficient to form virus-like particles. The *pol* ORF encodes the protease, responsible for cleaving the polyproteins into individual

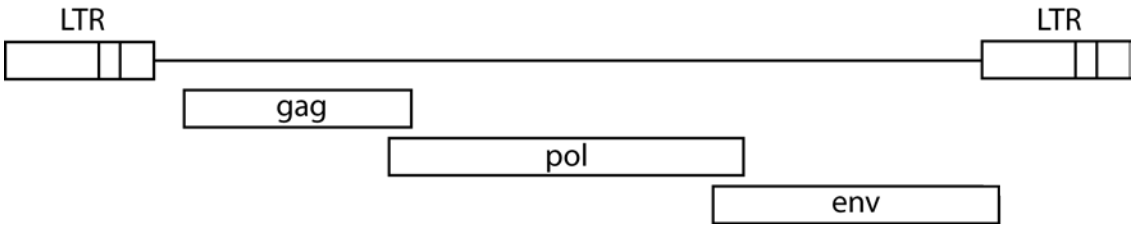


Figure 2 Basic genomic layout of retroviruses

All retroviruses have the same basic genomic layout. LTR, long terminal repeats, includes three regions U3, R, and U5. The solid line between the two LTRs represents the genomic sequence. Three main ORFs *gag*, *pol*, and *env* are depicted by the open boxes between the two LTRs. No accessory proteins are depicted.

components during maturation, reverse transcriptase (RT), and integrase (IN) proteins. However, betaretroviruses and deltaretroviruses encode protease in a separate ORF from Pol. The *env* ORF encodes for the Env protein, necessary for interaction with host receptor and virus entry. Some viruses encode additional proteins, collectively called accessory proteins, which are usually products of different splice events.

HIV-1 life cycle, shown in Figure 3, has been studied in great detail and serves as an example for all retroviruses. The infection begins with the interaction between the virus Env and a host cell membrane protein or proteins that is recognized and bound by Env, also called a receptor. Most viruses use one receptor protein, such as MMTV, which uses transferrin receptor (Tfr-1) as its receptor (Ross et al., 2002), but primate lentiviruses utilize CD4, and an additional coreceptor of either CCR5 or CXCR4 (Alkhatib et al., 1996; Choe et al., 1996; Dalgleish et al., 1984; Deng et al., 1996; Doranz et al., 1996; Feng et al., 1996; Maddon et al., 1986). For most retroviruses, the interaction between Env and its receptor leads to the fusion of the virus membrane and the host cell plasma membrane and the entry of the viral core into the cell cytoplasm (Marsh and Helenius, 1989; McClure et al., 1990; Stein et al., 1987). A few exceptions of this method of entry

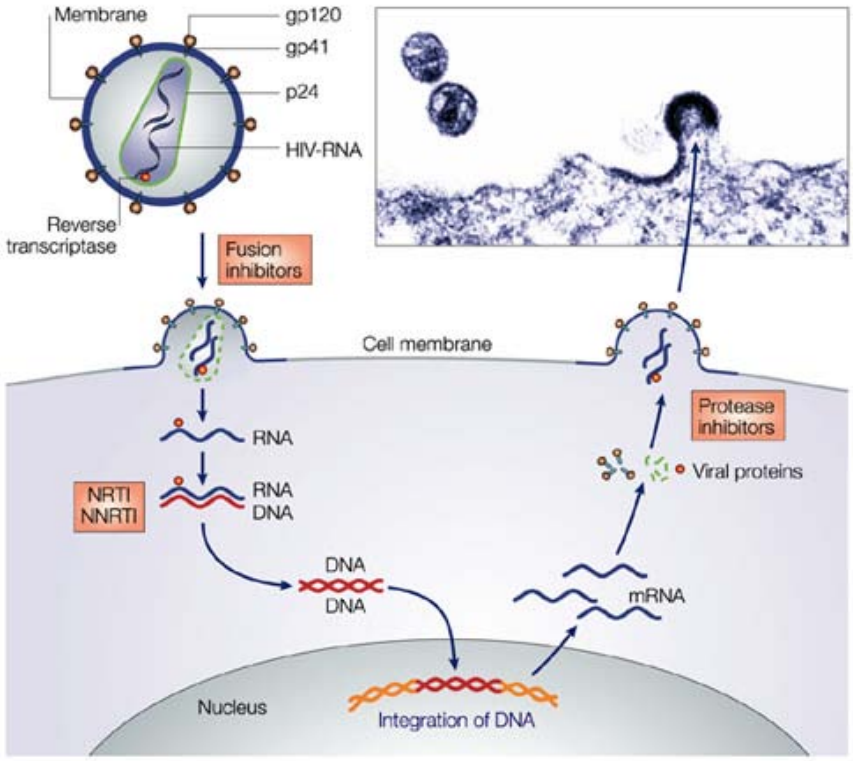


Figure 3 Life cycle of HIV-1

A simplified life cycle of HIV-1 is depicted. gp120 and gp41 are the transmembrane and surface subunits of HIV-1 Env. p24 is HIV-1 CA. The three classes of antiretroviral drugs, fusion inhibitors, RT inhibitors, and protease inhibitors, are depicted at the location of their action. NRTI, nucleoside RT inhibitor. NNRTI, non-nucleoside RT inhibitor. This figure is taken from Simon et al (Simon and Ho, 2003).

are MMTV, some strains of MLV, and subgroup A of avian leukosis virus (ALV), which are thought to be dependent on pH-dependent conformational change of Env in endosomes prior to fusion of the host and virus membranes (McClure et al., 1990; Mothes et al., 2000; Ross et al., 2002). Recently, data suggesting an endocytosed route of entry for HIV-1 as also been published (Miyauchi et al., 2009).

The fusion of the two membranes leads to the entry of the virus core into the cytoplasm. The core is composed of a shell of CA, and within the shell two copies of the viral genome and the associated virus proteins such as NC, RT, and IN. Different viruses form different core shapes according to genera, which are distinguishable by electron

microscopy. The disassembly of the virus core, a process named uncoating, is not well understood, but permits reverse transcription to begin in most retroviruses. What triggers reverse transcription is unknown, but is likely linked to a proper disassembly of the core, as mutations that disrupt the stability but not the formation of the core impairs reverse transcription (Forshey et al., 2002). It has also been postulated to be the exposure to high concentrations of dNTPs in the cytoplasm during uncoating (Goff, 2007). Foamy viruses are an exception, where reverse transcription begins as early as assembly in the producer cell, and virus particles contain partially or completely reverse transcribed virus genomes (Moebes et al., 1997; Yu et al., 1999).

The reverse transcription complex (RTC) is a large ribonucleoprotein structure composed of the two strands of ssRNA coated with NC, and also associated with RT, IN, CA, and Vpr, an accessory protein, in HIV-1 in which reverse transcription takes place (Bukrinsky et al., 1993; Fassati and Goff, 2001; Heinzinger et al., 1994). In MLV, the RTC is composed minimally of the viral genome, CA, RT, and IN (Fassati and Goff, 1999). The two ssRNA are used as template for the synthesis of the final dsDNA viral genome by RT in a series of discrete and highly ordered steps depicted in Figure 4 (Gilboa et al., 1979a; Gilboa et al., 1979b).

After reverse transcription, the dsDNA and associated proteins are called preintegration complex (PIC), a large complex that contains elements necessary for nuclear entry and integration. In addition to the viral dsDNA, MLV PIC is known to contain CA, and likely also contains RT and IN proteins (Bowerman et al., 1989). HIV PIC contains the dsDNA genome, Vpr, IN, and RT, but no longer contains CA and NC,

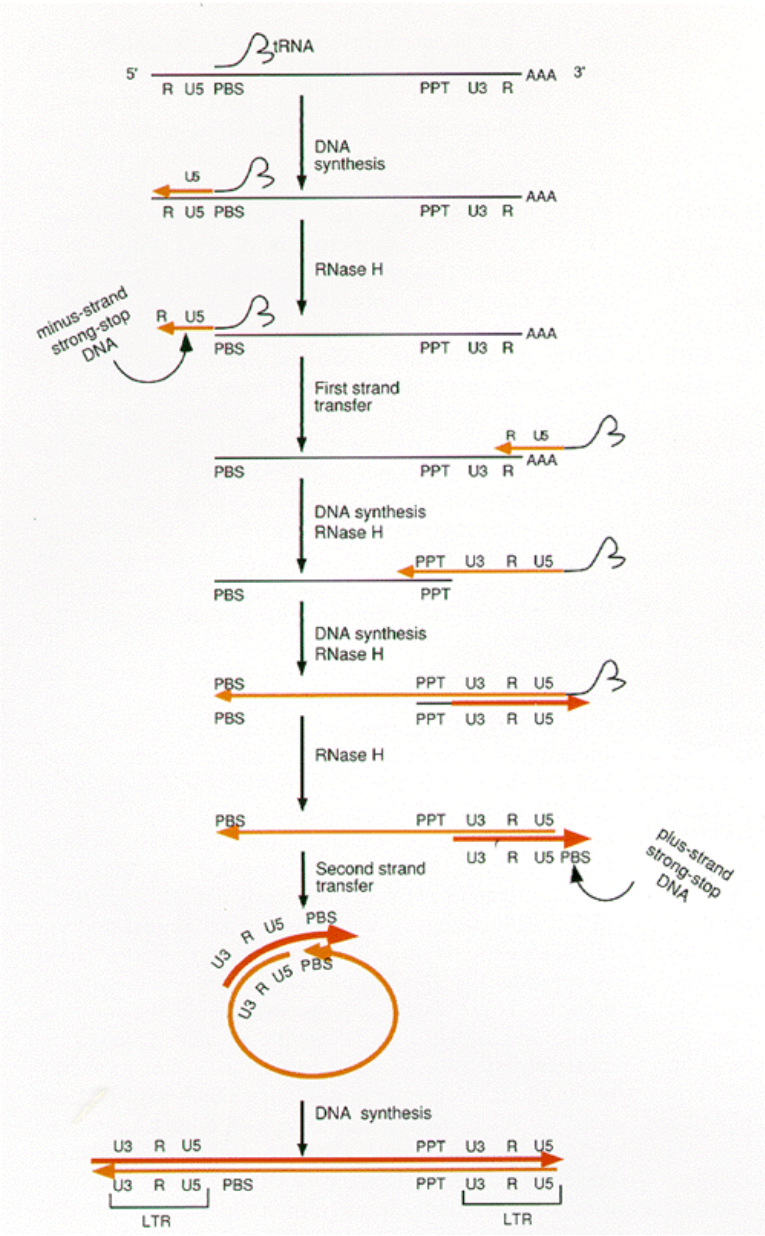


Figure 4 Reverse transcription of retroviruses

The black line represents RNA, light and dark orange lines represent minus and plus strands of DNA, respectively. Reverse transcription is primed by 3' end of tRNA bound to the primer binding site (PBS) to begin the synthesis of minus-strand strong stop DNA. Once RT reaches the 5' end of RNA, the DNA is degraded by the RNase-H portion of RT. In first strand transfer, the minus-strand DNA anneals to the 3' end of RNA by the repeated (R) sequence to resume minus strand DNA synthesis, along with RNaseH digestion of the RNA in RNA:DNA hybrid sequence. A polypurine tract (PPT) in the RNA genome that is relatively resistant to RNaseH primes the synthesis of plus strand strong stop DNA, which continues until the PBS of the minus strand. RNaseH removes the tRNA, exposing the PBS in the plus strand that is complementary to PBS in the minus strand. In second strand transfer, the PBS in the minus and plus strands anneal. Each strand serves as the template for the other for completion of reverse transcription. This figure is taken from Telesnitsky et al (Telesnitsky and Goff, 1997).

which likely has disassociated from the complex since reverse transcription (Bukrinsky et al., 1993; Heinzinger et al., 1994; Miller et al., 1997; Vodicka et al., 1998).

Retroviral PICs access the host genome for integration via different methods. MLV requires mitosis for the breakdown of the nuclear membrane to access the host genome, as they do not possess a means to cross the nuclear membrane (Lewis et al., 1992; Lewis and Emerman, 1994). In contrast, lentiviruses can infect nondividing cells, due to the ability of lentiviral PICs to actively cross the intact nuclear membrane in an ATP dependent manner (Lewis et al., 1992; Lewis and Emerman, 1994; Weinberg et al., 1991). An exception to these two methods is RSV, which can infect nonreplicating cells at low efficacy, although it infects dividing cells much more efficiently (Hatziioannou and Goff, 2001).

After the nuclear import of PIC, the dsDNA is integrated into the host genomic DNA via 3' end processing, strand transfer, and gap repair (Figure 5). In 3' end processing, the blunt dsDNA viral genome is processed by IN whereby the two terminal nucleotides at the 3' strand are removed (Brown et al., 1989; Roth et al., 1989). The resulting 3' OH groups are used to transfer the viral DNA into the host genome via attack of the phosphodiester bonds in the host DNA in strand transfer (Brown et al., 1989; Engelman et al., 1991; Fujiwara and Mizuuchi, 1988). The 5' ends are then repaired by unknown factors (Roe et al., 1997), leaving a repeat of 4-6 nucleotides from the host DNA flanking the integrated virus where the strand transfer occurred. As there is no mechanism to target the integrated virus sequence and excise it from the genome, integration is irreversible.

The integrated virus sequence, now called a provirus, contains promoter

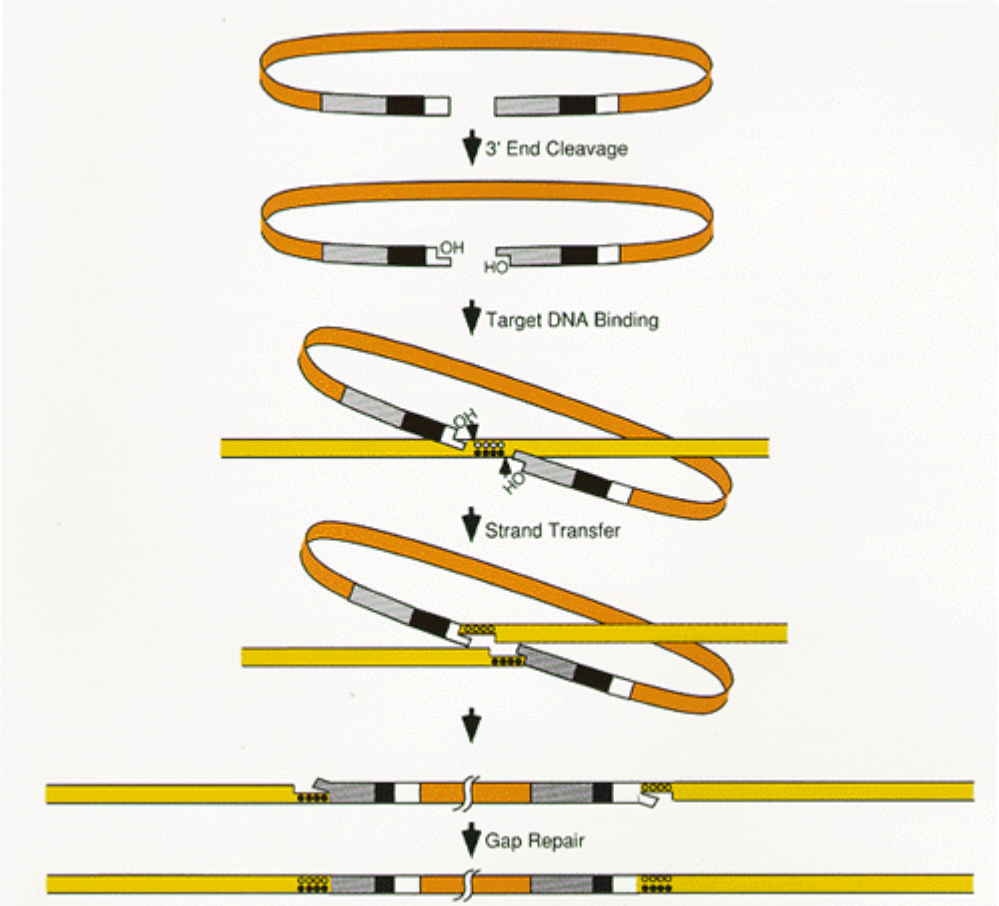


Figure 5 Retrovirus integration

The orange line represents the viral cDNA, while the yellow line represents the genomic DNA. The U3, R, and U5 regions in the LTR are represented as grey, black and white boxes. In 3' end processing, IN removes two nucleotides from the 3′ ends of the viral DNA, exposing 3′ hydroxyl groups. In strand transfer, IN joins the 3′ ends of viral DNA to the genomic DNA. DNA repair enzymes fill the gaps and repairs the break. The resulting provirus is flanked by repeated segments of the target DNA.

sequences in the U3 region of 5' LTR to host RNA polymerase II system for transcription. In HIV-1 and HTLV-1, accessory proteins Tat and Tax promote the elongation of the transcript and induce higher levels of transcript, respectively (Cann et al., 1985; Felber et al., 1985; Kao et al., 1987). Like most RNA polymerase II transcripts, the viral transcript is capped at the 5' end, processed, and polyadenylated at the polyadenylation site in the 3' LTR (Goff, 2007). The full length transcript is also spliced to yield transcripts for Env and accessory proteins.

There are two means by which retroviruses export partially spliced or unspliced transcripts from the nucleus to the cytoplasm, such as the full length genomic transcript. Many retroviruses, such as HIV-1, HTLV-1, and MMTV, encode for an accessory protein named Rev, Rex, and Rem, respectively, which bind to full length transcripts and export them to the cytoplasm in a CRM-1 dependent manner (Bogerd et al., 1998; Felber et al., 1989; Hammarskjold et al., 1989; Indik et al., 2005b; Inoue et al., 1987; Mertz et al., 2005b; Neville et al., 1997). In contrast, unspliced MPMV transcripts contain a constitutive export element (CTE) near the 3' end of the genomic RNA that enable transcript export without additional viral proteins (Bray et al., 1994). RNA with CTE are exported in a TAP/NXF1-mediated pathway (Braun et al., 2001; Gruter et al., 1998; Guzik et al., 2001; Katahira et al., 1999).

Viral proteins are translated from exported transcripts as polyproteins. The three main ORFs, *gag*, *pol*, and *env*, encode for the structural proteins, enzymatic proteins, and the Env protein. In betaretroviruses and deltaretroviruses, the protease protein, which is located 5' in the pol ORF in other viruses, is a separate ORF from the rest of *pol*. *Pol* or *prot-pol* ORFs are translated as an extension of *gag* via translational read-through, where the stop codon in Gag is suppressed by the ribosome, which inserts a $tRNA^{gln}$ at the stop codon a fraction of the time (Yoshinaka et al., 1985), or translational frameshifting where the ribosome slips near the end of *gag* ORF and continues translation in a different reading frame (Jacks and Varmus, 1985). Hence, Gag, Gag-Prot, Gag-Prot-Pol, or Gag-Pol polyproteins are translated depending on the virus.

The translated proteins and full length viral genome assemble to produce new virus particles. Most retroviruses assemble at the plasma membrane via MA in the form

of either a stretch of basic residues, named membrane-binding domain (M domain) (Verderame et al., 1996; Zhou et al., 1994), and/or myristoylation of MA (Henderson et al., 1983). At the plasma membrane, Gag is thought to multimerize by the effects of the interaction domain (I domain) located at the amino terminus of NC (Campbell and Vogt, 1995; Carriere et al., 1995; Hansen and Barklis, 1995). The binding of RNA by NC is thought to promote Gag multimerization by acting as a structural scaffold where Gag proteins can interact (Campbell and Vogt, 1995). Betaretroviruses and spumaretroviruses are exceptional, as the particles first assemble in a pericentriolar region in the cytoplasm instead of the plasma membrane, then are transported to the plama membrane for release (Rhee et al., 1990; Yu et al., 2006). The RNA genome is packaged into the particle usually via the packaging (Ψ) sequence located in the untransted region between the 5' LTR and *gag* ORF, which is either near or overlaps with sequences important for viral RNA dimerization and splicing and binds to the Cis-His motifs in NC [reviewed in (D'Souza and Summers, 2005)].

The final release of virus particle from the host cell is accomplished via the late domain (L domain), which are found in various locations in the Gag protein of different viruses [reviewed in (Bieniasz, 2006)]. PT/SAP, PPXY, and YPXL motifs have been identified thus far as L domains, and each helps recruit a network of proteins called endosomal sorting complexes required for transport (ESCRT) that are normally involved in vacuolar protein sorting pathway. HIV-1 contains the PTAP motif which is recognized by Tsg101 (Garrus et al., 2001; Martin-Serrano et al., 2001; VerPlank et al., 2001). RSV, MPMV, HTLV-1, and MLV contain the PPPY motif which is recognized by Nedd4 family members (Blot et al., 2004; Kikonyogo et al., 2001; Martin-Serrano et al., 2005;

Sakurai et al., 2004; Yasuda et al., 2002). Equine infectious anemia virus (EIAV) contains the YPDL motif, which binds to Alix (Martin-Serrano et al., 2003; Strack et al., 2003). Viruses with mutations in L domains fail to separate from the host cell membrane remain tethered by a membraneous stalk (Gottlinger et al., 1991; Wills et al., 1994; Xiang et al., 1996), indicating that the ESCRT complexes may help the final scission of the lipid bilayer between the virus and the cell. After the separation of the virus particle and the host cell, virus protease cleaves the polyproteins into the individual components, which reassemble in the virus to form a new, mature virus particle.

The methods by which retroviruses cause diseases are diverse. Oncogenic retroviruses were first recognized in 1911 by Peyton Rous by the discovery that transfer of cell-free supernatant from chicken sarcoma can cause sarcoma in a new host (Rous, 1910, 1911). This discovery was followed by the discovery of viral agent in mice called Bittner's agent transmitted by breast milk that causes breast cancer (Bittner, 1936, 1942; Lyons and Moore, 1962; Moore, 1963), as well as a retrovirus that caused "spontaneous leukemia" in AKR mice (Gross, 1951). These agents, RSV, MMTV, and MLV respectively, are the first retroviruses described and can cause cancer via different methods. RSV and MLV are examples of transducing oncogenic retroviruses. RSV carries the *v-src* gene, an avian oncogene which can transform infected cells. MLV is also capable of transferring viral oncogenes, such as *v-abl* (Ableson MLV), *v-ras* (Harvey MLV), and *v-mos* (Moloney murine sarcoma virus) (Goff et al., 1980; Shih et al., 1979; Van Beveren et al., 1981). The discovery that viral oncogenes actually derive from the host opened up a new field of molecular cancer research (Flint et al., 2004; Hanafusa et al., 1977; Stehelin et al., 1976). Cellular counterparts of these genes are now known as

proto-oncogenes, and often are involved in cell cycle or death, hence any changes to their expression or activity may induce abnormal states, such as cancer. Non-transducing oncogenic retroviruses do not carry oncogenic genes, but may induce cancer via insertional mutagenesis; hence, all retroviruses are potentially non-transducing oncogenic retroviruses. A clinical example of viral insertional mutagenesis causing cancer is the five gene therapy patients who were treated for SCID-X, an X-linked disorder caused by a lack of a functional γc cytokine receptor, which results in the lack of T and NK lymphocytes (Hacein-Bey-Abina et al., 2008; Hacein-Bey-Abina et al., 2003; Howe et al., 2008). In two separate studies, 19 of 20 patients developed functional T and NK lymphocytes, but five developed leukemia. All five patients displayed an over-abundance of cell clones with insertion in or near proto-oncogenes, as well as additional chromosomal irregularities in three patients.

Non-cancer related immunodeficiency is a second means by which retroviruses commonly cause human disease. Primate lentiviruses infect $CD4^+$ cells, hence can infect various immune cells such as T cells, macrophages, and dendritic cells (Dalgleish et al., 1984; Klatzmann et al., 1984). HIV-1 infection is characterized by a slow decline of $CD4^+$ T cells over years, which eventually results in immunodeficiency [reviewed in (Simon and Ho, 2003; Stevenson, 2003)]. Whether the $CD4^+$ cells die from virus infection or immune reaction against the infection is still unclear. Nonprimate lentiviruses often infect macrophages or dendritic cells, and can cause immunodeficiencies (feline immunodeficiency virus, bovine immunodeficiency virus), arthritis-like diseases (caprine arthritis encephalitis virus, Visna Maedi virus), and anemia (EIAV).

2. Integration site preference of retroviruses

Many of these diseases are due to the integration of the viral genome, which accounts for the persistence of infection and insertional mutagenesis. Integrase is the viral protein that catalyses the integration of the viral sequence into the host genome, one of the hallmarks of retrovirus infection. An interesting and important aspect of integrase function is how it determines where in the genome to integrate the viral genome. Studies have shown that most retroviruses have a unique preference for the location of integration. This preference correlates strongly with genomic markers such as transcription units or promoter regions rather than a specific nucleotide sequence. For example, HIV-1 has a two-fold preference to integrate into active transcription units over matched random controls, a feature which is also independent from the target cell type used (Barr et al., 2006; Mitchell et al., 2004; Schröder et al., 2002a). This preference may be dependent on the association between IN protein and the host protein LEDGF/p75, which rescues IN from proteosomal degradation in the cytoplasm, enhances DNA binding of IN, and tethers IN to host chromatin (Busschots et al., 2005; Cherepanov et al., 2003; Llano et al., 2004a; Llano et al., 2004b; Maertens et al., 2003). LEDGF also interacts with IN proteins of other lentiviruses, namely HIV-2, EIAV, BIV, VMV, FIV, and EIAV, and may help determine their preferences as well (Busschots et al., 2005; Cherepanov, 2007; Llano et al., 2004b).

Conversely, the IN proteins of nonlentiviruses, such as RSV (alpharetrovirus), MPMV (betaretrovirus), MLV, feline leukemia virus (FeLV, gammaretroviruses), and HTLV-1 (deltaretrovirus) do not interact with LEDGF (Busschots et al., 2005; Cherepanov, 2007; Llano et al., 2004b). HTLV-1 and MMTV have been shown to have

no integration site preference (Derse et al., 2007; Faschinger et al., 2008). MLV has a strong integration preference for promoter regions and CpG islands (Hematti et al., 2004; Mitchell et al., 2004; Wu et al., 2003). Lastly, avian sarcoma-leukosis virus (ASLV) has a largely random integration with minor preference for genes (Barr et al., 2005; Mitchell et al., 2004). What determines this difference, whether it be other host proteins or intrinsic differences in the IN proteins, is unknown.

3. Endogenous retroviruses

When retroviruses infect germ cells or germ cell progenitors, the virus can become endogenized [reviewed in (Bannert and Kurth, 2004; Gifford and Tristem, 2003; Lower et al., 1996)]. These viruses are called endogenous retroviruses (ERV). The provirus will be present in the genome of all cells of the individual derived from the infected germ cell, and be passed on to progeny in a Mendelian manner to following generations. Both chance and the insertion's effect on the fitness of the host can determine the allelic frequency in the population. Hence, elements which produce large quantities of viral proteins and progeny or elements that insert into a necessary gene will likely reduce the fitness of the host and as an allele will be negatively selected in the host population.

Endogenous retroviruses exist as either a provirus or a solo LTR. Over time, the two LTR sequences in a provirus can homologously recombine, deleting all internal sequence and leaving behind only a solo-LTR that is a composite of the two original LTRs. How long an element remains as a provirus until its LTRs recombine is unknown, but likely variable and stochastic. One study has shown that one nucleotide difference

may reduce the likelihood of recombination by ten-fold (Belshaw et al., 2007). In humans, around 85% of all human ERVs (HERVs) exist in the solo-LTR form (Lander et al., 2001). Theoretically, exogenous retroviruses can also form solo-LTRs, but this potential is probably of little importance in the face of the pathogenesis of the full length virus. A preintegration site describes a state within the host population where the integration has not occurred. Potentially, all sequence in a host is a preintegration site, but the term is applied as a relative state to an integrated element rather than to signify the integration potential at that location.

As discussed earlier, integration can be detrimental via insertional mutagenesis of genes, but integration into introns can provide alternate splice sites and premature polyA signals. While most likely harmful, these insertions can provide new transcript isoforms (Kapitonov and Jurka, 1999; Mager et al., 1999). Also, as an LTR includes promoter and enhancer sequences, integration near a gene can alter transcription levels or the tissue specificity with which a given gene is expressed (Dunn et al., 2003; Landry et al., 2002; Medstrand et al., 2001; Medstrand et al., 2002). At a genomic scale, ERVs can induce genomic rearrangements between highly similar elements, such as ecotopic recombination (Hughes and Coffin, 2001, 2005).

Endogenous retroviruses are present in the genomes of all vertebrates except Agnathas, and are closely related to all genera of retroviruses except deltaretroviruses (Herniou et al., 1998; Jern et al., 2005; Martin et al., 1997) . The naming of HERVs has been erratic, but many are named after the primer binding site and the tRNA most likely used to prime reverse transcription (Bannert and Kurth, 2004). Thus, HERV-K indicates a HERV that likely uses tRNA-lysine as its RT primer.

Most ERVs align closely with beta-, gamma-, and spumaretroviruses, which are otherwise named as Class I, II, and III, respectively. In humans, there are around 72 HERV Class I families, such as HERV-H, HERV-W, and HERV-FRD, which intermix with known exogenous gammaretroviruses in phylogenetic analysis (Jern et al., 2005; Lander et al., 2001; Tristem, 2000). HERV-W and HERV-FRD include rare examples of endogenous retroviral sequence being co-opted by the host. One locus each of HERV-W and HERV-FRD proviruses have maintained expression and fusogenic capacity of the Env protein in trophoblast cells in the placenta (Blaise et al., 2003; Blond et al., 2000; Bonnaud et al., 2004; Mi et al., 2000). Named syncytin-1 and syncytin-2, the fusion mediated by these ERV Env proteins are thought to create a layer of fused cells in the placenta, called syncytiotrophoblasts, which is important for placentogensis and pregnancy. Similar proteins and phenomenon have been described in mice and sheep, and have also been shown to be necessary for a successful pregnancy in sheep (Dunlap et al., 2006; Dupressoir et al., 2005).

An interesting nonhuman gammaretrovirus-related ERV is the koala retrovirus (KoRV). Discovered while studying the high incidence of leukemia and lymphomas in koalas, KoRV seems to be currently undergoing endogenization (Hanger et al., 2000; Tarlinton et al., 2006). Both exogenous and endogenous forms of KoRV can be found, and isolated populations do not have the endogenous form, suggesting that the endogenization has occurred in the past 100 years after the separation of koala populations (Tarlinton et al., 2006). How infection by this virus is related to the leukemia and lymphoma is unclear, but plasma viral RNA levels correlate positively with disease (Tarlinton et al., 2005).

In humans, class II ERVs consist of ten subfamilies which align most closely with each other in phylogenetic analysis than with any other retrovirus known, and are known as HERV-K superfamily due to their tRNA binding site (Bannert and Kurth, 2004; Lander et al., 2001). Betaretroviruses MMTV and Jaagsiekte sheep retrovirus (JSRV), which infect mice and sheep respectively, exist in both endogenous and exogenous forms and align closely with HERV-K superfamily (Gifford and Tristem, 2003). Class III has around 20 families in the human genome, including HERV-L and HERV-S (Cordonnier et al., 1995; Jern et al., 2005; Lander et al., 2001; Yi et al., 2004). Recently, endogenous lentiviruses from rabbits (RELIK) and grey mouse lemurs ($pSIV_{gml}$) have been discovered, and likely form a separate class of ERVs (Gifford et al., 2008; Gilbert et al., 2009; Katzourakis et al., 2007).

In humans, around 8% of the genome is composed of sequences of retroviral origin, independent from retrotransposon sequences (Lander et al., 2001). All known HERVs are defective, replete with mutations, premature stop codons, and truncations. The retroviral insertions are found on all chromosomes, although the number does not correlate with chromosomal size (Kim et al., 2004a; Villesen et al., 2004). Most insertions are found outside of genes, and when in genes, are more often in the opposite transcriptional orientation relative to the gene (van de Lagemaat et al., 2006; Villesen et al., 2004).

As all retroviral insertions found in the human genome are defective, how the insertions proliferate in the genome is worth considering. There are two general mechanisms by which ERVs may proliferate. ERVs can proliferate via bone fide extracellular particle formation and infection of a new target cell, which requires an intact

virus sequence. Interestingly, all HERVs with a high copy number seem to have proliferated via means additional to true replication (Gifford and Tristem, 2003). Defective ERVs may proliferate by methods other than bone fide virus replication, such as via exogenous infection events following complementation in trans, where functional proteins are supplied by other endogenous or exogenous viruses. This may result in retention of only sequences important for transcription and packaging, hence protein ORFs should deteriorate over time. HERV-H and ERV-9 may have proliferated by this method (Belshaw et al., 2005b). Alternatively, envelope-independent retrotransposition in cis may occur, where an element copies itself and inserts into a new genomic locus within the same cell, forgoing the normal extracellular phase of the retroviral life cycle. Via this method, all ORFs except *env* are likely to retain a functional sequence, as Gag and Pol will still be required for replication. HERV-K(HML-3) is a potential example of this scenario, as its *pol* sequence seems to have been maintained (dN/dS = 0.15), while *env* ORF has been changing more rapidly (dN/dS = 0.73) (Belshaw et al., 2005b). Defective proviruses can also proliferate as a result of long interspersed element 1 (LINE-1) aided retrotransposition, where the enzymes of LINE-1 provide necessary function for the viral sequence to be reverse transcribed and integrated, akin to the proliferation of Alu elements. In this scenario, only the sequences required for transcription and packaging will be required, and most protein ORFs will not be maintained. Indeed, two major groups of the HERV-W family can be traced back to either independent replication or LINE-1 dependent replication, based on the proviral and flanking genomic sequence (Belshaw et al., 2005b; Costas, 2002).

Currently, there is no known replication competent HERV, but one family of Class II HERV, HERV-K(HML-2), seems to have been replicating until recently. The HERV-K(HML-2) family includes human specific members and elements that are polymorphic in the human population, suggesting replication since the divergence of humans and chimpanzees 6 million years ago and potentially more recently as well (Barbulescu et al., 1999; Turner et al., 2001).

Although inactive, many HERVs are expressed in various tissues, and as a result many diseases have been connected to HERV expression of RNA, protein, or virus particles, or generation of anti-HERV antibodies [reviewed in (Blomberg et al., 2005; Moyes et al., 2007)]. This includes various cancers (melanoma, breast cancer, germ cell tumors), neurologic diseases (schizophrenia, bipolar disorder, multiple sclerosis), autoimmune diseases (psoriasis, arthritis, systemic lupus erythematosus, type I diabetes), HIV-1, and pregnancy. Notably, healthy individuals also express various HERV products. Overall, none of the correlations with diseases establish clear evidence of HERVs as the cause of the disease. The best correlation between HERV expression and disease is between HERV-K(HML-2) and seminoma, where two independent studies found that a high percentage (60% and 70%) of patients are positive for anti-Gag antibody (Boller et al., 1997; Sauter et al., 1995). This type of correlation may be useful for diagnostic or therapeutic vaccine purposes.

4. History of the discovery of HERV-K superfamily

Much of HERV research has focused on HERV-K(HML-2), a subfamily of the Class II HERV-K superfamily. HERV-K(HML-2) first garnered attention due to its

sequence similarity to MMTV, and more recently due to its young age and replication potential.

The discovery of HERV-K superfamily was spurred by the desire to find a viral etiology for breast cancer. By 1940s, it was already known that MMTV, at the time known as Bittner agent or milk agent, could cause spontaneous mammary tumors in mice regardless of genetic background, and that this agent was filterable and passed to offspring via milk (Bittner, 1936, 1942). In an attempt to find an analogous agent in humans, scientists looked for signs of virus in human milk and breast cancers. In 1969, Moore et al reported finding particles morphologically similar to MMTV in human milk, the occurrence which correlated with breast cancer (Moore et al., 1969). A few years later, they also found that sera from some women with breast cancer could neutralize MMTV infection in mice, and that MMTV-like virus particles were found in milk of women with family history or women from populations with high incidence of breast cancer more frequently than those who are not (Charney and Moore, 1971; Moore et al., 1971). These particles showed RNA-dependent DNA polymerase activity and sensitivity to ribonuclease (Ohno et al., 1977; Schlom et al., 1971), as well as the existence of a high molecular weight RNA in human milk (Schlom et al., 1972), two features thought to be unique to oncogenic RNA viruses at the time. Simultaneously, Axel et al showed that human breast cancer cells express RNA that is homologous to MMTV RNA (Axel et al., 1972), supporting the idea that an MMTV-like virus exists in human breast, and may be associated with cancerous states.

Despite these signs of a retrovirus, no clear candidate virus emerged. As it was already established that the Bittner agent could integrate into infected cells and become

heritable (Moore, 1963), the search for the "human breast cancer virus" eventually turned to genomic DNA. During early 1980s, two groups independently detected an MMTV-like sequence in the human genome using probes designed based on the sequence of MMTV *gagpol* (Callahan et al., 1982; May et al., 1983). This sequence had a recognizable retroviral organization, with similarities in parts to known retroviruses based on both DNA hybridization and sequence (Callahan et al., 1985; Deen and Sweet, 1986; May and Westley, 1986). Studies using probes based on the reverse transcriptase region of the Syrian hamster intracisternal A particle (IAP) also identified similar human endogenous retroviral sequences (Ono, 1986). This study named the group of the identified proviruses HERV-K, based on the lysine tRNA binding site sequence, the presumed primer for reverse transcription. The same group also sequenced the HERV-K10 provirus in its entirety, thus establishing the first prototype HERV-K (Ono et al., 1986). After realizing that the MMTV-based probes hybridized to different clones with varying strength, Franklin et al teased apart their collection into nine subgroups, NMWV-1 to NMWV-9, based on hybridization differences (Franklin et al., 1988). The groups remain largely intact today after analysis based on sequence data, but have been renamed human MMTV like-1 (HML-1) to HML-10, in an altered order (Andersson et al., 1999). Furthermore, studies showed that sequences similar to HERV-K is found in other hominoids and Old World monkeys, but not New World monkeys or prosimians (Mariani-Costantini et al., 1989), suggesting that the first germ cell integration occurred around 35 million years ago after the divergence of New World and Old World monkeys.

In a separate line of inquiry, ultrastructural examination and electron microscopy of cells derived from a testicular cancer revealed retrovirus-like particles budding from

the cell, which were named human teratocarcinoma derived virus (HTDV) (Bronson et al., 1979; Bronson et al., 1978). Transfer of supernatant from virus like particle (VLP) producing cells to fresh cells did not result in detectable virus particles in the new cells, suggesting either that these viruses were not infectious, or that the target cells were nonpermissive (Lower et al., 1984). However, the supernatant contained high molecular weight RNA and detectable reverse transcriptase activity at a sucrose density characteristic of animal retroviruses of 1.16g/ml (Boller et al., 1983; Lower et al., 1987). Furthermore, these viruses did not cross react with antibodies against other animal retroviruses, and were visually and immunologically distinguishable from HTLV-1 and HIV-1, the two human retroviruses known at the time (Lower et al., 1987). However, excitement at discovering a new human retrovirus was soon dashed, as studies using antibody derived against HERV-K Gag for immunoelectron microscopy and western blot analysis of teratocarcinoma cell lines showed that HDTV and HERV-K(HML-2) are identical viruses (Boller et al., 1993b).

5. Activity and replication potential of HERV-K(HML-2)

Since its discovery, research on HERV-K(HML-2) has focused on characterizing the multiple proviruses in the genome, testing the activity of its individual components in an effort to understand viral replication, and searching for a replication competent provirus. Prior to the completion of the human genome sequencing project, screening was manually conducted by Southern blotting analysis, Northern blotting analysis, or PCR from cellular genomic DNA or libraries (Andersson et al., 1999; Barbulescu et al., 1999; Barbulescu et al., 2001; Mayer et al., 1999a; Medstrand and Blomberg, 1993b;

Medstrand and Mager, 1998; Sugimoto et al., 2001; Tonjes et al., 1999; Turner et al., 2001; Zsiros et al., 1998). After genome sequencing, both manual genomic screens via blots and bioinformatic approaches have been used (Belshaw et al., 2005a; Belshaw et al., 2004; Costas, 2001; Hughes and Coffin, 2004, 2005; Romano et al., 2006).

Collectively, these studies show that components of HERV-K are still functional. Structural proteins are capable of assembling into particles and budding from cells (Bieda et al., 2001; Mueller-Lantzsch et al., 1993; Tonjes et al., 1999; Tonjes et al., 1997), and protease, RT, and IN are functional in vitro (Berkhout et al., 1999; Kitamura et al., 1996; Mueller-Lantzsch et al., 1993; Schommer et al., 1996). However, as most of the clones tested were derived from PCR reactions that did not distinguish between the proviruses, the identity of the proviruses to which the functional components belong is unknown. And while not all HERV-K *env* ORFs encode an active protein, one Env that can pseudotype SIV particles and permit infection has been identified (Dewannieux et al., 2005).

More work has been conducted on an accessory protein, Rec/K-Rev. Retroviruses often use overlapping ORFs or multiple splice products to enable expression of multiple protein products. The splice sites, while useful for generating new transcripts, creates another problem, namely nuclear retention of full length or partially spliced viral transcripts. Some retroviruses, such as HIV-1, HTLV-1, and MMTV, have solved this problem by encoding a nuclear-cytoplasm shuttling protein, which specifically binds full length and partially spliced viral transcripts and exports them to the cytoplasm, named Rev, Rex, and Rem, respectively (Chang and Sharp, 1989; Emerman and Malim, 1998; Hidaka et al., 1988; Indik et al., 2005a; Mertz et al., 2005a).

In HERV-K, northern blot analysis of polyA$^+$ RNA from the GH teratocarcinoma cell line using an U5 region specific probe showed the expected transcripts of GagPol and Env, as well as smaller transcripts around 1.8-1.5 kb (Lower et al., 1993). Sequencing these transcripts showed a doubly spliced mRNA encoding a protein whose amino acid sequence revealed an arginine-rich motif reminiscent of HIV-1 Rev and HTLV-1 Rex proteins. This protein, named Rec or K-Rev, has been shown to be expressed in teratocarcinoma cell lines such as GH and Tera-2, and localizes to the nucleolus (Lower et al., 1995). Like its HIV-1 counterpart Rev, Rec/K-Rev contains a leucine-rich nuclear export signal (NES) and an arginine-rich nuclear localization signal (NLS), and can export unspliced or partially spliced viral transcripts via a CRM-1 dependent manner (Boese et al., 2001; Boese et al., 2000; Magin et al., 2000; Magin et al., 1999; Yang et al., 1999). The Rec response element (RcRE), an RNA element that Rec/K-Rev recognizes on the full length and partially spliced transcripts, has been mapped to the U3 region of the 3' LTR, which is likely folds into a complex structure (Magin-Lachmann et al., 2001; Magin et al., 1999; Yang et al., 2000; Yang et al., 1999).

HERV-K also encodes for a dUTPase, a ubiquitously expressed protein that catalyzes the conversion of dUTP to dUMP and PP_i. This protein, encoded by herpesviruses (EBV, VZV, and HSV-1), poxviruses (Orf virus and vaccinia virus) and retroviruses (Baldo and McClure, 1999), is thought to reduce the fatal incorporation of dUTP into the viral genome during replication. The acquisition of dUTPase by viruses is believed to have occurred as a result of horizontal transfer of the dUTPase gene from a host (Baldo and McClure, 1999). In retroviruses, three groups contain the dUTPase gene, namely betaretroviruses, non-primate lentiviruses, and an ancient mammalian ERV-L

family (Cordonnier et al., 1995). These three groups encode dUTPase in different locations in their genomes, suggesting that the acquisition of dUTPase may have occurred independently three times in retroviral evolution.

Like MMTV and MPMV, HERV-K families encode for a dUTPase protein (Harris et al., 1997a; Mayer and Meese, 2003). A consensus dUTPase, derived from an alignment of PCR derived HERV-K dUTPase sequences from human cell lines, is functional in vitro (Harris et al., 1997a; Harris et al., 1999). HERV-K dUTPase is expressed in some human cell lines, but the proviral origin and its functional capability or effect on host is unknown (Harris et al., 2000).

Beyond individual proteins, scientists were also interested in discovering new HERV-K proviruses, with the hopes of finding a replication competent provirus. In screens prior to the human genome sequencing, multiple full length proviruses were identified (Barbulescu et al., 1999; Barbulescu et al., 2001; Mayer et al., 1999a; Sugimoto et al., 2001). The most exciting of these finds are the human specific and the polymorphic proviruses, the youngest and the most likely to be active (Barbulescu et al., 1999; Belshaw et al., 2005a; Herrera et al., 2006; Hughes and Coffin, 2004; Macfarlane and Simmonds, 2004; Turner et al., 2001).

There are multiple signs in the human specific proviruses that hint at their young age. First, the human specificity indicates that the proviruses likely integrated after the divergence of humans and chimpanzees, estimated to have occurred around six million years ago (Bannert and Kurth, 2004; Barbulescu et al., 1999). The polymorphic insertions may be even younger (Barbulescu et al., 1999; Turner et al., 2001), is based on the idea that fixation time for neutral mutations in the human population with an effective

population size of 10,000 and generation time of 20 years will be around 800,000 years (Graur and Li, 2000; Hughes and Coffin, 2004). Hence, any unfixed neutral insertion is likely to have inserted less than 800,000 years ago. Furthermore, insertions that have identical LTR sequences are thought to be less than approximately 200,000 years old. The two LTRs of a provirus, identical at the time of integration due to the nature of reverse transcription, will accumulate mutations independently from each other over time, hence the number of differences between the two LTRs may serve as a molecular clock for the insertion (Johnson and Coffin, 1999). Based on calculations of the divergence between insertions in both humans and chimpanzees and the time since that divergence, one change between the two LTRs is expected to occur per 200,000 to 450,000 years (Johnson and Coffin, 1999; Turner et al., 2001). Thus, identical LTRs in a provirus is an indication that the insertion may have occurred as recently as 200,000 or fewer years ago. One provirus, K113, incited particular excitement, as it exhibited all the characteristics just described, in addition to its complete set of open reading frames that exhibited no obvious mutations or truncations (Turner et al., 2001).

Bioinformatic search is a more thorough method to identify HERV-K(HML-2) elements in the human genome. One study identified 553 HERV-K(HML-2) insertions, of which 113 are human specific (Belshaw et al., 2005a). Most identified elements, including all the human specific ones, likely inserted independently (Belshaw et al., 2004; Romano et al., 2006). Two characteristics of these insertions suggest that HERV-K(HML-2) was replication competent until recent times. First, the independently inserted elements show a purifying selection on all ORFs, suggesting that the protein sequence and likely its function has been preserved (Belshaw et al., 2005c; Belshaw et al., 2004;

Costas, 2001). This is true for Env ORF as well, suggesting that retrotransposition in cis (retrotransposition within the same cell) is unlikely to be a major method of proliferation. Secondly, stop codons do not seem to be inherited from one provirus to another, which is likely to occur if the viruses replicated via retrotransposition in trans.

The low activity of most HERVs is not surprising, as high levels of viral protein expression or continuous replication is likely harmful to the host. One way by which hosts may defend themselves against constant genomic assault is via antiviral host proteins called restriction factors.

6. Restriction factors

Various SIVs and HIVs are highly similar to each other in sequence and behavior, but also exhibit species-specific characteristics that may govern interspecies transmission and adaptation in a new host. One determinant of the transmission or the lack thereof between closely related species may be restriction factors, host proteins that are expressed in many types of cells, including non-immune cells. Three major classes of restriction factors described and studied here are the Fv-1 and TRIM5 group of proteins, the APOBEC family of proteins, and tetherin.

Fv-1, which confers resistance to the Friend strain of MLV in various strains of mice in a dominant and heritable by single locus manner (Lilly, 1967) [reviewed in (Bieniasz, 2004a; Goff, 2004)], is a Gag like protein similar to that of the MuERV-L endogenous retrovirus in mice (Benit et al., 1997; Best et al., 1996; Qi et al., 1998). Two major alleles of the locus were found in different strains of mice which dictated their susceptibility to different strains of MLV: cells from NIH3T3 mice carrying only the

Fv1^n allele are infectable by N-tropic MLV but block B-tropic MLV, while cells with only the *Fv1*^{b} allele, found in BALB/c mice, show the opposite susceptibility phenotype (Gardner et al., 1980; Goff, 2004). Mice with both alleles block infection from both strains of viruses, showing that resistance is dominant (Odaka et al., 1978). The restriction occurs at a step after reverse transcription, but before the nuclear transport of the virus preintegration complex and the target of Fv-1 on the virus maps to amino acid 110 of MLV CA (DesGroseillers and Jolicoeur, 1983; Kozak and Chakraborti, 1996; Rommelaere et al., 1979). Furthermore, Fv-1's restriction activity saturates at high quantities of incoming virus, suggesting that a limiting factor is being used up.

Although intriguing in itself, the restriction phenotype of Fv-1 is also important because a similar block in retrovirus infection occurs in primates (Besnier et al., 2003; Cowan et al., 2002; Towers et al., 2003). The factor in human cells that restrict N-tropic MLV was termed Ref-1 (restriction factor 1), and a similar factor in various nonprimate cells that restrict HIV-1 and other retroviruses was termed Lv-1 (lentivirus susceptibility factor 1).

Lv-1 was identified from rhesus macaque cells as TRIM5α (tripartite motif 5α) (Stremlau et al., 2004). Subsequent studies showed that Ref-1 is the human homolog of Lv-1, and characterized the susceptibility of numerous retroviruses to each protein (Hatziioannou et al., 2004; Keckesova et al., 2004; Perron et al., 2004; Yap et al., 2004). TRIM5 protein is a member of the TRIM family, whose characteristic tripartite motifs include an N terminal RING domain, a B-box domain, and a coiled-coil domain, with an additional C terminal domain in many members [reviewed in (Johnson and Sawyer, 2009b)].

TRIM5α, the longest splice variant of the *trim5* gene, contains a C-terminal B30.2/SPRY domain, which is the most variable portion of the protein between species and determines restriction specificity (Nakayama et al., 2005; Perez-Caballero et al., 2005; Sawyer et al., 2005a; Sebastian and Luban, 2005; Stremlau et al., 2005; Yap et al., 2005). The functions of RING and B box domains in TRIM5α remain unclear but both are important for the protein's restriction activity (Diaz-Griffero et al., 2006; Javanbakht et al., 2005; Perez-Caballero et al., 2005; Stremlau et al., 2004). The coiled-coil domain is important for multimerization, which is necessary for restriction (Mische et al., 2005; Perez-Caballero et al., 2005). The mechanism of TRIM5α action is still unclear, but may involve the disruption of disassembly of the viral core (Anderson et al., 2006; Perron et al., 2007; Sebastian and Luban, 2005; Shi and Aiken, 2006; Stremlau et al., 2006; Wu et al., 2006). Furthermore, TRIM5α has been shown to function as a RING-type finger E3 ubiquitin ligase (Yamauchi et al., 2008), suggesting a potential role independent from retroviral infection related to protein degradation.

In a twist to the story, TRIM-Cyp, a TRIM5 protein with a retrotransposed CypA, was described in owl monkeys that is also capable of restricting HIV-1 (Nisole et al., 2004a; Sayah et al., 2004a). Astoundingly, an independent retrotransposition event resulting in similar TRIM-Cyp protein have been found in macaque species as well (*Macaca nemestrina*, *Macaca fascicularis*, and *Macaca mulatta*) (Brennan et al., 2008; Liao et al., 2007; Newman et al., 2008; Virgen et al., 2008; Wilson et al., 2008).

The second family of restriction factors is the APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like) family of cytidine deaminases. Prior to their identification as an antiretroviral factor, signatures of the activity of APOBEC3 proteins

had already been detected in the HIV-1 grown in ex vivo cultures and peripheral blood mononuclear cells (PBMCs) of HIV-1^+ patients as mutations of Gs to As in the context of either GG or GA dinucleotides (Borman et al., 1995; Janini et al., 2001; Vartanian et al., 1991). These changes were thought to be a result of misincorporation by RT, which is highly error prone (Preston et al., 1988; Roberts et al., 1988).

The actual cause of these changes was identified in a search for the host target of HIV-1 accessory protein Vif. It was already established that Vif was necessary for the production of replication competent HIV-1 particles from cells such as CEM cells, but not in others such as CEM-SS and 293T cells (Gabuzda et al., 1992; von Schwedler et al., 1993). A complementary DNA subtraction strategy was used to selectively amplify cDNA unique to the nonpermissive cell population (CEM) from a related, permissive cell population (CEM-SS) (Sheehy et al., 2002). These cDNAs were used as probes in Northern blotting analysis to identify APOBEC (A3G) as the factor in nonpermissive cells responsible for producing noninfectious particles from Vif-deficient HIV-1 (Sheehy et al., 2002). In the absence of Vif, A3G is packaged into virus particles during assembly in the producer cell and inactivates the virus for the next round of infections (Bogerd and Cullen, 2008; Khan et al., 2007; Khan et al., 2005; Sheehy et al., 2002; Zennou et al., 2004).

The initial studies of APOBEC proteins showed that the cytidine deaminase activity, which resulted in a “hypermutation” of the viral genome, was important for its restriction activity (Bishop et al., 2004b; Harris et al., 2003; Mangeat et al., 2003; Zhang et al., 2003). A3G deaminates Cs to Us on the ssDNA, negative strand intermediate of reverse transcription, which is read by RT as Ts and thus transcribed as As in the positive

strand (Iwatani et al., 2006; Yu et al., 2004c). This results in the G to A change ultimately detected on the positive strand of the viral DNA sequence. Other APOBEC3 family members were also shown to have cytidine deaminase and antiviral activity (Bishop et al., 2004a; Dang et al., 2008; Dang et al., 2006; Harris and Liddament, 2004b; OhAinle et al., 2006; Yu et al., 2004a; Zennou and Bieniasz, 2006). Furthermore, other viruses and repeat elements, including hepatitis B virus (HBV), HPV, LINE-1, Alu elements, HTLV-1, foamy viruses, adeno-associated virus (AAV), MLV, and murine endogenous retroelements, were also shown to be susceptible to antiviral effects of APOBEC family of proteins (Bogerd et al., 2006a; Chen et al., 2006; Chiu and Greene, 2006; Delebecque et al., 2006; Esnault et al., 2006; Muckenfuss et al., 2006; Niewiadomska et al., 2007; Sasada et al., 2005; Stenglein and Harris, 2006; Turelli et al., 2004a; Vartanian et al., 2008b).

Ensuing studies using mutant APOBEC proteins that lack cytidine deaminase activity showed that hypermutation may not be the sole component of restriction (Bishop et al., 2006; Newman et al., 2005; Opi et al., 2006; Shindo et al., 2003). Blocks in reverse transcription and integration have been proposed as additional mechanisms, with the caveat that the studies were conducted using APOBEC overexpression systems (Guo et al., 2006; Holmes et al., 2007; Mbisa et al., 2007; Yang et al., 2007). However, when A3G is titrated to natural levels, the catalytic activity seems to be the main determinant of A3G's antiviral activity (Browne et al., 2009; Miyagi et al., 2007).

Unfortunately, human APOBEC proteins cannot restrict HIV-1. HIV-1 Vif prevents the encapsidation of human A3G (hAG) and A3F (hA3F) into the virus particles by recruiting the cullin-RING ubiquitin ligase complex which leads to the

polyubiquitination and degradation of A3G and A3F (Kao et al., 2003; Liu et al., 2005a; Mehle et al., 2004; Sheehy et al., 2003; Yu et al., 2003; Yu et al., 2004e).

Studies of Vpu, another HIV-1 accessory protein, led to the discovery of tetherin. Vpu was known to enhance budding of a variety of retroviruses, suggesting that it likely acts on a host factor rather than directly on the virus particles (Gottlinger et al., 1993; Klimkait et al., 1990; Strebel et al., 1989; Strebel et al., 1988; Terwilliger et al., 1989). Its effect is cell-type specific: Vpu defective particles assemble and bud normally in simian cells such as African green monkey COS-1 cells but are blocked in others, and accumulate at the plasma membrane as fully assembled and mature particles (Sakai et al., 1995). Heterokaryons of permissive and nonpermissive cells result in a block, suggesting that the factor is present in the nonpermissive cells and is dominant (Varthakavi et al., 2003). Furthermore, IFN treatment enhances the block of Vpu defective HIV-1 release, suggesting that the host factor responsible for the block is upregulated by IFN (Neil et al., 2007).

Comparative microarray assay of permissive and non-permissive cell lines, as well as IFNα treated and untreated permissive cells pinpointed the factor responsible for blocking HIV-1 particle release in the absence of Vpu as tetherin (Neil et al., 2008a; Van Damme et al., 2008b). Tetherin, also known as CD317, Bst-2, and HM1.24, is a membrane protein with unknown function, but is highly expressed in B cells and multiple myeloma-derived cells (Goto et al., 1994; Kupzig et al., 2003). Its extracellular domain contains a coiled coil domain and three cysteines, while the C-terminus is anchored to the membrane via a GPI-anchor and the N-terminus via a transmembrane domain (Kupzig et al., 2003). Prior to its association with HIV-1, tetherin was thought to participate in

vesicular trafficking due to its localization at both the plasma membrane and intracellular compartments and its ability to be internalized (Kupzig et al., 2003). The mechanism of tetherin's activity on virus particles and the converse mechanism of Vpu on tetherin is currently under investigation.

There are a few characteristics of the restriction factors described above that further support their antiviral nature. The most prominent characteristic is positive selection, detected by a high dN/dS ratio (nonsynonymous changes to synonymous changes). This ratio quantifies the nucleotide changes that result in amino acid change relative to nucleotide changes that do not, and reflect the need for and pattern of evolution of the host in the face of evolving or new pathogens, particularly their evasion of the host immune defense. TRIM5α, six of seven APOBEC3 family members, and tetherin all exhibit this sign of positive selection (Conticello et al., 2005; Liu et al., 2005b; McNatt et al., 2009a; Rhodes et al., 2005; Sawyer et al., 2004; Sawyer et al., 2005a). Another sign of adaptation is the expansion of the locus, whereby duplicate copies of the same gene can each adapt in different ways to different pressures, and thereby specialize in function. For example, the APOBEC3 protein family is composed of seven highly related members that are capable of targeting different pathogens, as noted earlier [reviewed in (Harris and Liddament, 2004a)]. In humans, the TRIM5 locus includes three other family members TRIM6, TRIM34, and TRIM22, the latter protein also having evolved under under positive selection; in cows, TRIM5 locus has expanded to five potential coding genes (Sawyer et al., 2007). Furthemore, these proteins are upregulated by IFN, suggesting that they may be a part of a concerted effort by the host

to combat pathogens (Asaoka et al., 2005; Neil et al., 2007; Sakuma et al., 2007; Tanaka et al., 2006).

The constant adaptation of restriction factors is suggestive of an ongoing battle between host and old and new pathogens. The identity of these pathogens is unknown, as current state of restriction factors is likely a result of selective pressure from multiple pathogens. Isolating the influence of a single pathogen is difficult, and also does not help identify the pathogen. Furthermore, many of these battles likely occurred millions of years ago, potentially against ancestors of modern pathogens, or against pathogens that have been successfully controlled by the host and may no longer exist.

The only ancient pathogens whose genomic information is available are those which have integrated their genome into that of the host, such as endogenous retroviruses and other repeat elements. Endogenous retroviruses are fossils of ancient infections, and their sequence provides a snapshop of the pathogen from millions of years ago. Furthermore, as constant genomic companions to the host, any virus activity such as the expression of viral proteins or virus replication may provide continuous interaction between the virus and the host. Thus, ERVs likely exerted a constant selective pressure on the host, and may have played a large role in shaping how hosts defend themselves against current retroviruses.

A major problem of working with ERVs is the deteriorated state of the sequence, which over years in the host genome has accumulated numerous mutations and truncations that shrouds the original sequence at the time of infection. In this body of work, this problem is circumvented by deducing a consensus sequence from the youngest set of HERV proviruses. Named HERV-K_{CON}, the function of its individual components

and the replication potential of this virus is tested, and its life cycle is characterized. Lastly, HERV-K_{CON}'s interaction with current anti-retroviral host proteins is accessed, and evidence of the same interaction occurring in vivo is presented.

II. MATERIALS AND METHODS

1. Cloning of endogenous HERV-K proviruses

BAC plasmid containing HERV-K113 provirus (BAC RP11-398B1) was prepared according to instructions from the manufacturer (Invitrogen). K113 was PCR amplified using the purified BAC DNA as template into two fragments using primers targeting the flanking genomic sequence and SacI restriction site (4415) in *pol* ORF. For clarification, all the restriction sites in HERV-K113 and HERV-K_{CON} (discussed in Results) used for construction of HERV-K plasmids are noted in Figure 6. The primers targeting the flanking genomic sequence were designed based on the NCBI Nucleotide entries AF387847 and AF387849. The 3' half of K113 was first cloned into pXF3 using SfiI and NheI restriction sites in the multiple cloning site (MCS) of pXF3 (K113-3' For SfiI: TCAACCCATGGGGCCGAGGCGGCCTGGG, K113-3' Rev NheI: CATGTTT CCTG CTAGCCCACAAACACATGCAGACG). The 5' half of K113 was amplified (K113-5' For SfiI: TCATTCTAGGCCGAGGCGGCCTTGCATGGGGAGATTCAGAA CC, K113 Rev1: TCTTATCAGATGCTATTGCCAGTCC) and cloned into pXF3 containing the 3' half of K113 into SacI (4415) and SfiI restriction sites (pXF3 MCS) to construct pXF3/K113 (Figure 6). pXF\3 is a low-copy plasmid for expression in mammalian cells.

CMVP was cloned into the 5' LTR U3 region up to the TATA box by overlap PCR to construct pXF3/CMVP K113 using ClaI (pXF3 MCS) and SalI (1037) restriction sites (Figure 6). BAC plasmid DNA containing the K108 provirus was prepared according to instructions from the manufacturer from BAC 33P21 (Invitrogen). K108

sequence from SacII (6488) to the end was amplified from purified DNA (HERVK SacII For: GCACCTCCGCGGAGACGGAGACATCGCAATCG, K108 Flanking 3': GACCA GCCTGACCAAGATGGTGAAACCTGTAGGGGTG), and inserted into SacII (6488) and NheI (pXF3 MCS) restriction sites of either pXF3/K113 or pXF3/CMVP K113 to construct pXF3/K113 K108 and pXF3/CMVP K113 K108 (Figure 6).

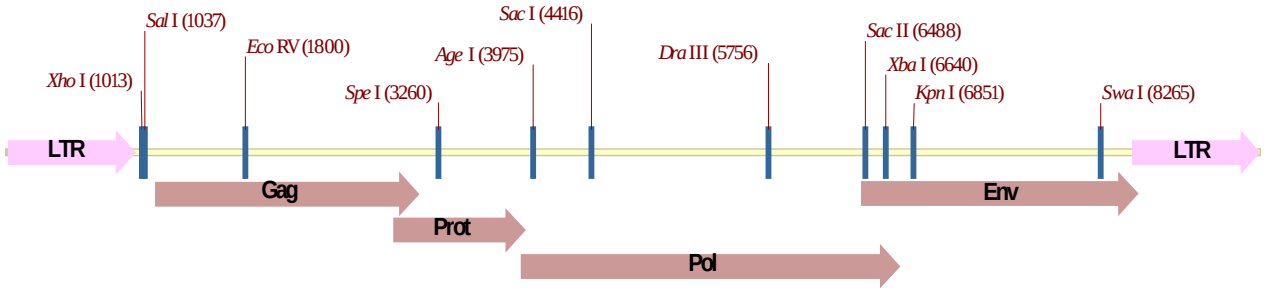


Figure 6 Location of restriction sites in HERV-K(HML-2)

The restriction sites used for constructing HERV-K113 and HERV-K_{CON} plasmids are listed by name and location in the genome. The locations are identical for K113 and HERV-K_{CON}.

2. Derivation and synthesis of HERV-K_{CON}

The complete HERV-K113 proviral sequence was used to search human genome sequence using National Center for Biotechnology Information nucleotide-to-nucleotide BLAST. Multiple entries of the same HERV-K proviruses were identified by inspection of flanking genomic sequence and only the most recently sequenced entries were used for the alignment. The first ten hits were aligned using AlignX program (Vector NTI Advance 10.0.1, Invitrogen, http://www.invitrogen.com) to derive a consensus sequence that was termed HERV-K_{CON}. Namely, the ten hits were K101, K102, K104, K107, K108,

K109, K113, K115, 11q22, and 12q13. Information about these proviruses is listed on Table 2.

The complete HERV-K_{CON} proviral sequence was synthesized using overlapping oligonucleotides of approximately 60 bases spanning the entire genome. Oligonucleotides were assigned to 13 groups corresponding to 13 HERV-K_{CON} fragments of approximately 700 nucleotides and assembled using sequential PCRs. 1µl of each 100µM oligonucleotide was mixed with other oligonucleotides of each group to make an equal molar mix. In the first round PCR, 1µl of the mix was used in the reaction and 15 cycles of synthesis were executed using Pfu DNA polymerase (94 °C for 10 s, 45 °C for 20 s, 72 °C for 30 s), which fills the gaps between the oligonucleotide overlaps. Thereafter, 2µl of the reaction product was subjected to amplification using the 5′ and 3′ oligonucleotides in each fragment group (94 °C for 20 s, 45 °C for 20 s, 72 °C for 3 min; 15 cycles). Fragments from regions of the HERV-K genome lacking convenient restriction sites were assembled into longer fragments of up to 1.5 kb via overlap extension PCR. A derivative of the low-copy-number plasmid vector, pXF3, was cloned by inserting a synthetic oligonucleotide encoding the restriction sites ClaI (pXF3 MCS), XhoI (1013), EcoRV (1800), SpeI (3260), AgeI (3975), SacI (4416), DraIII (5756), XbaI (6640), and NheI (pXF3 MCS) that corresponded to convenient restriction sites in the HERV-K_{CON} genome (Figure 6). These sites were used to sequentially insert the synthesized DNA fragments, thereby generating the final pXF3/HERV-K_{CON} proviral plasmid. The entire HERV-K_{CON} was sequenced.

3. HERV-K_{CON}–derived expression plasmids

Various versions of HERV-K_{CON} were synthesized as packageable viral genome for experiments by replacing the U3 region of 5’ LTR with cytomegalovirus (CMV) promoter/enhancer sequence to enhance transcription and inserting a selectable marker into the Env open reading frame to detect infection (Figure 21). CHKCG was created from the pXF3/HERV-K_{CON} proviral plasmid by first replacing HERV-K U3 sequences 5′ to the TATA box of the 5’ LTR with CMV promoter/enhancer sequences using overlapping PCR and ClaI (pXF3 MCS) and EcoRV (1800) restriction sites to generate pXF3/CMVP HERV-K_{CON} (Figure 6). In parallel, an EGFP cDNA (Clontech) was inserted into pCR3.1 (Invitrogen), followed by PCR amplification and insertion of the CMVP-EGFP into KpnI (6851) site of pXF3/CMVP HERV-K_{CON} to create CHKCG. Similarly, a Puro cDNA was digested from pMSCV Puro (Clontech) with HinDIII and XbaI and inserted into pCR3.1. Thereafter, a CMVP-Puro cassette was PCR amplified and cloned into the KpnI (6851) restriction site of pXF3/CMVP-HERV-K_{CON} to construct CHKCP.

CCGBX, a derivative of CHKCG, was constructed by inserting a CMVP-EGFP cassette into XbaI (6640) and SwaI (8265) restriction sites of CHKCG, resulting in a vector genome that is slightly smaller than CHKCG and gives higher infectious titers. CCGBX-P was derived from CCGBX by inserting a 53-bp HIV-1-derived sequence (GATCTGAGCCTGGGAGCTCTCTGGCTTGTGACTCTGGTAACTAGAGATCCCTC) into the 5' end of the 3' long terminal repeat (LTR) to allow specific amplification of de novo HERV-K_{CON} sequences upon infection of human cells. The insertion was made using overlap extension PCR and SwaI (8265) and NheI (pXF3 MCS) restriction sites in

the HERV-K_{CON} vector sequence. CCBXS is a modified CCGBX with a blasticidin resistance gene in place of EGFP for clone selection. Reverse transcriptase mutant versions of all Pol containing plasmids were constructed by mutating the conserved YIDD (amino acids 195-198) RT domain into AIAA, then inserting the amplified fragment with nucleotide changes into *pol* ORF using SacII (6488) and DraIII (5756) restriction sites.

HERV-K protein expression plasmids pCRVI/Gag, pCRVI/Gag-PR, and pCRVI/Gag-PR-Pol were generated by insertion of the respective ORFs from pXF3/HERV-K_{CON} into the NotI restriction site of pCRVI, an HIV-1 based expression plasmid that also expresses HIV-1 accessory proteins Tat, Rev, and Vpu (Figure 7). Similarly, a PCR-amplified HERV-K_{CON} Env-encoding fragment was inserted using EcoRI and NotI restriction sites, generating pCRVI/Env. A mutant version of pCRVI/Gag-PR, pCRVI/Gag-PR(mut), was generated by substituting the conserved putative active site residues DTG (amino acids 33-35) to AAA. The two Rec exons were PCR amplified from BAC RP11-33P21 (Invitrogen) containing the HERV-K108 sequence, which encodes a Rec protein that is identical to the consensus sequence, joined using overlapping PCR, and inserted into the EcoRI and XhoI sites of pCR3.1 to generate pCR3.1/Rec. pCR3.1 is a commercially available plasmid (Invitrogen) designed to transiently express a cloned sequence at high levels in mammalian cells and includes the human cytomegalovirus virus promoter and the BGH polyA signal. pCRVI/Gag-GFP was cloned by first deleting an NcoI restriction site (1251) near the start of *gag* ORF by site-directed mutagenesis. This Gag sequence was cloned into pCAGGS-HA via NcoI and EcoRI sites, then the EGFP sequence inserted into the EcoRI site, and screened for

correct EGFP orientation. Like pCR3.1, pCAGGS is a plasmid designed to transiently express cloned inserts at high levels in mammalian cells, and includes the enhancer from human cytomegalovirus immediate early promoter, the chicken β-actin/rabbit β-globin hybrid promoter, and the rabbit β-globin polyadenylation signal. Gag-GFP was then cloned into the NotI restriction site in pCRVI.

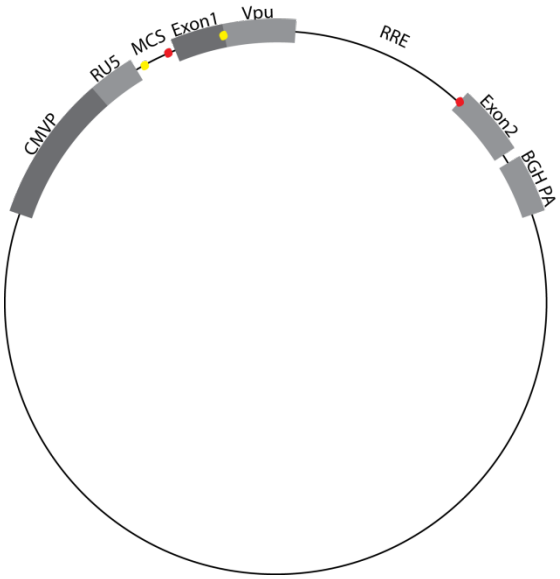


Figure 7 Map of pCRVI

pCRVI is a high copy plasmid for transient expression in mammalian cells derived by combining sequences of HIV-1 and pCR3.1 (Invitrogen). Inserts cloned into the MCS are expressed at high levels by the CMV promoter enhancer sequences and the expression of HIV-1 accessory protein, Tat. Tat activation region, the viral sequence Tat recognizes to promote transcript elongation is located in the RU5 region. The transcripts are exported efficienty by Rec and Rec response element (RRE) by CRM-1 dependent RNA export pathway. Exon1 and exon2 encode for Tat and Rev in different open reading frames. Splice donor and acceptor sites are marked by yellow and red dots, respectively.

pCR3.1/K113 LTR GFP, a plasmid that expresses GFP under the control of HERV-K113 LTR, was constructed by a three-way ligation with K113 LTR from pXF3/K113 described earlier (SpeI and EcoRI sites), EGFP from pEGFP-N1 (EcoRI and NotI sites, Invitrogen), and pCR3.1 (NotI and SpeI sites). pEGFP-N1 is a commercially

available plasmid designed to express the cloned insert in fusion with EGFP transiently at high levels in mammalian cells (Invitrogen). The EGFP contains two amino acid changes from GFP sequence (F64L and S65T) for higher expression and brighter fluorescence (Invitrogen). This cloning strategy eliminates the CMVP from pCR3.1.

4. Other expression plasmids

CSGW, a packageable HIV-1 vector plasmid expressing enhanced green fluorescent protein (EGFP), and HIV-1_{NL4-3} GagPol expression plasmids have been described elsewhere (Bainbridge et al., 2001; Cowan et al., 2002). Plasmids expressing various human APOBEC3 proteins, namely hA3A, hA3B, hA3C, hA3F, and hA3G, have been described previously (Bishop et al., 2004a). Plasmids expressing additional human APOBEC3 proteins (hA3DE and hA3H) were constructed using the same pCMV4-HA vector and HindIII and XbaI restriction sites.

5. Cell lines

293T, MDTF (mouse), NIH3T3 (mouse), Pindak (squirrel monkey), TE671 (human), HeLa, CRFK (cat), and HT1080 (human) cells were maintained in Dulbecco's modified Eagle's medium (DMEM), CHO745 (Chinese hamster) cells and TRIM5-expressing derivatives in Ham's F-12 medium, and CEM (human) cells in RPMI medium, all supplemented with 10% fetal calf serum and gentamicin.

6. Transfection

To generate virus-like particles, 293T cells were seeded on 10-cm plates at 6×10^6 cells per plate or in six-well plates at 1×10^6 cells per well. The following day, the cells were transfected using 4 μg of polyethylenimine (PEI) per μg of DNA.

To generate VSVG pseudotyped HERV-K_{CON} particles, 293T cells in six-well plates were transfected with 1.3 μg of HERV-K_{CON} packageable genome plasmid, 1 μg of pCRVI/Gag-PR-Pol, 0.5 μg of pCR3.1/Rec, and 0.2 μg of VSVG. Alternatively, 293T cells in 10-cm dishes were transfected with 6.5 μg of HERV-K_{CON} packageable genome plasmid, 4 μg of pCRVI/Gag-PR-Pol, 3 μg of pCR3.1/Rec, and 1.5 μg of VSVG. To generate infectious HIV-1 virions, 293T cells were transfected in six-well plates with 0.75 μg of CSGW, 0.75 μg HIV-1 Gag-Pol, and 0.2 μg VSVG. To generate HIV-1 (HERV-K_{CON}) pseudotypes, 293T cells in 10-cm plates were transfected with 6 μg of HIV-1-GagPol, 6 μg of CSGW, and either 3 μg of pCRVI/HERV-K_{CON} Env or empty pCRVI as a control. Additional plasmids or empty control vectors were transfected when necessary and as noted for each experiment. Medium was changed 5 or 12 hours after transfection with fresh medium containing 5μM sodium butyrate, and virus-containing supernatants were collected 48 hours after transfection. No Vif protein was expressed during the generation of virions for APOBEC experiments in Figures 38, 39, and 40.

7. Infection

For infection, cells were seeded on 24-well plates the previous day (293T and TE671 cells at 5×10^4 cells per well, CHO745 cells at 4×10^4 cells per well, HT1080 cells at 3×10^4 cells per well, Pindak, CRFK, NIH3T3, and MDTF cells at 2×10^4 cells

per well). Filtered supernatant from transfected cells (0.2 μm) was layered onto cells with fresh medium supplemented with 5 μg of polybrene/ml. For experiments in Chapter 4, cells were also spinoculated at 2,000 rpm for 2 hours at room temperature. Two days after infection, the infected cells were quantified either by counting foci microscopically, by fluorescence-activated cell sorter analysis for GFP^+ cells, or by selecting cells for drug resistance depending on the packageable genome plasmid used for particle production. For drug selection, cells were expanded 24 hours post infection, and placed under blasticidin (0.25 μg/ml) or puromycin (2.5 μg/ml) for approximately ten to fourteen days.

8. Reverse Transcription Assay

Reverse transcriptase activity in 293T culture supernatants was measured using a commercially available Lenti RT Activity Kit (Cavidi, http://www.cavidi.se) in which BrdUTP is incorporated into a plate-bound oligo(dT)/poly(rA) substrate. Thereafter, solid phase polymerized BrdU is detected using an anti–BrdU–alkaline phosphatase conjugate and a colorimetric substrate. Activity is standardized using a recombinant HIV-1 RT standard. The Lenti RT Activity Kit was used, as HERV-K(HML-2) RT had been shown to preferentially use Mg^{2+} over Mn^{2+} as the divalent cation like HIV-1 (Berkhout et al., 1999).

9. Generation of anti-HERV-K CA antibody

To generate an anti-CA polyclonal antiserum, the N-terminal cleavage site of CA was first determined by Edman sequencing of the putative CA protein isolated from purified HERV-K_{CON} virion particles. 293T cells were transfected with pCRVI/Gag-PR,

and the supernatant was replaced 12 hours post transfection with fresh medium containing 5μM of sodium butyrate. After an additional 48 hours, the virus particles were purified through a 30% sucrose gradient, and the proteins separated in a protein gel by SDS-PAGE. The proteins were stained using amido black stain, and two bands at 30kDa and approximately 15kDa were cut from the gel for Edman sequencing. The bands were sent to University of Texas Medican Branch (UTMB) at Galveston Texas Protein Chemistry Core for Edman sequencing (http://www.utmb.edu/brf/cores/ProteinChemistry/index.html).

The exact location of cleavage is depicted in Figure 9. As CA is estimated to be around 30 kDa, the position of the C-terminal CA cleavage site was estimated based on the determined position of the CA N terminus. The deduced CA-encoding sequence was cloned into pGEX-6P-1 (GE Healthcare Life Sciences) to express a glutathione *S*-transferase (GST)-tagged CA protein that was purified using glutathione-agarose beads. The GST tag was eliminated by PreScission protease cleavage as per the manufacturer's instructions (GE Healthcare Life Sciences). The purified recombinant CA protein was used to generate the antiserum (Covance).

10. HERV-K protein analysis

The 293T cells were transfected with HERV-K plasmids as described above. Supernatant was collected two days post-transfection, filtered (0.2 μm), and ultracentrifuged through a 20% sucrose layer at 100,000*g* for 90 min at 4 °C to pellet virus-like particles (VLPs). Transfected cells were lysed via sonication in SDS-PAGE loading buffer and separated on 10% SDS-PAGE gels (Bio-Rad, http://www.biorad.com).

Proteins were transferred onto nitrocellulose membrane and probed with a commercially purchased anti–HERV-K Gag antibody (Austral Biologicals, http://www.australbiologicals.com), or the generated anti-HERV-K CA polyclonal antiserum. Alternatively, VLPs were separated on 4% to 20% gradient or 10% SDS-PAGE gel (Bio-Rad) and silver stained using a kit, as per the manufacturer's instructions (Sigma-Aldrich, http://www.sigmaaldrich.com).

11. Analysis of de novo integrated HERV-K_{CON} proviral DNA

CHO745 cells were infected with CHKCP-carrying HERV-K_{CON}(VSVG) virus stock and transduced cells selected in 2.5 μg/ml puromycin for approximately 10 days. From the puromycin-resistant population comprising several hundred colonies, four single cell clones were derived by limiting dilution and expanded in culture for approximately 2 weeks. Total DNA was extracted from each clone using the Qiagen DNeasy Blood and Tissue Extraction kit as per manufacturer's instructions (http://www.qiagen.com). The extracted DNA was used as template for PCR analysis using HERV-K_{CON} *gag*-specific primers Gag-S (nucleotides 1236 to 1262) and Gag-AS (nucleotides 1991 to 1946). Additionally, host DNA sequences flanking the integrated CHKCP proviral DNA were cloned using the GenomeWalker kit (Clontech) according to the manufacturer's instructions and PCR primers directed to the HERV-K_{CON} LTRs. Specifically, LTR-AS (GCA AGA GAG ATC AGA TTG TTA CTG TGT CTG) and LTR-S (TAC GAG AAA CAC CCA CAG GTG TGT AGG) oligonucleotides were used to clone sequences flanking the 5′ and 3′ LTRs, respectively. Additional PCR primers, targeting flanking hamster DNA sequences identified via the GenomeWalker approach,

were used to authenticate the presence of preintegration sites in uninfected CHO745 cells and integrated provirus in three CHKCP transduced cell clones. In the example (clone No. 1) shown in Figure 27D, the primers Ham-S (GCT ACC CTG AAG ATT TGA GCC AGT GTG C) and Ham-AS (TCT TGC AAG TTG TCC TGT GGC ATG G) were used. For all PCR reactions, 30 cycles of amplification were completed using 200 ng of cellular DNA, with no DNA, uninfected CHO cell DNA, or human DNA templates analyzed as negative and positive controls, as appropriate.

12. HERV- K sequence analysis

Full-length HERV-K(HML-2) sequences in the human genome were identified using a TBLASTX search (www.ensembl.org/Multi/blastview) of the human genome with the HERV-K_{CON} Gag sequence as the query sequence. Sixteen unique (by chromosomal location) human-specific full-length HERV-K(HML-2) proviruses were identified by cross-referencing with insertions identified by Belshaw et al. and Romano et al. (Belshaw et al., 2005a; Romano et al., 2006). All identified insertions were included in subgroup N, as defined by Romano et al. (Romano et al., 2006). Specifically, the proviruses included K101, K102, K104, K106, K107, K108, K109, K113, K115, 11q22, 12q14, 19q12, 1p31 (K4), 3q27 (K50b), 3q21 (KI), and 21q21 (K60). All GenBank accession numbers for these sequences are found in the reports by Barbulescu et al. and Romano et al. In addition, the K60 sequence was deduced from GenBank entry AL109763 (Barbulescu et al., 1999; Romano et al., 2006). The proviruses were aligned to HERV-K_{CON} using AlignX (Invitrogen) for comparison. Figures 34 and 38, below, depict G-to-A differences between the provirus and HERV-K_{CON} and were derived using the

HYPERMUT program (http://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html). The *p* values that accompany the sequence analyses were calculated by a chi-square test of independence to determine whether the frequency at which each nucleotide occurred at each position flanking each mutation was significantly different relative to its expected frequency based on nucleotide composition.

13. HERV-K hypermutation assay

Infectious HERV-K_{CON} particles were generated as described above using CCGBX-P in place of CCGBX. Prior to infection, supernatant was supplemented with 10 mM $MgCl_2$ and treated with DNase I (0.1 U per μl; Roche) for 1 h at 37°C to eliminate residual transfected DNA. Fresh 293T cells were infected as described above. Ten hours postinfection, total DNA was collected using the DNeasy blood and tissue kit (Qiagen). Partial EGFP (Clontech) and HERV-K sequence of 762 nucleotides were amplified using oligos designed to target EGFP (CGC ACC ATC TTC TTC AAG GAC GAC G) and the inserted HIV-1 sequence (GAG GGA TCT CTA GTT ACC AGA GTC ACA AGC C) using Phusion polymerase (Invitrogen) (98°C for 10 s, 55°C for 10 s, and 72°C for 15 s; 30 cycles). Amplified DNA was purified using a gel extraction kit (Qiagen) and cloned into pCR-Blunt II-TOPO according to the manufacturer's instructions (Invitrogen). To confirm the complete elimination of transfected DNA, amplification of plasmid sequence using primers targeting the plasmid backbone sequence and the untranslated region of HERV-K using similar PCR conditions was attempted. Also, to confirm that the amplified sequences were derived from de novo reverse-transcribed DNA, HERV-K

particles containing a mutationally inactivated reverse transcriptase were subjected to the same infection procedure and PCR analysis. Twelve clones were sequenced for each APOBEC3 protein studied, as well as the empty vector control, and compared to the original CCGBX-P sequence for evidence of hypermutation.

14. Aphidicolin induced cell arrest and infection

HT1080 cells were seeded in 24-well plate at 5 x 10^4 cells per well and treated with 2 µg/ml of aphidicolin 24 hours prior to infection to arrest at G1/S phase. The same concentration of aphidicolin was maintained during infection. 48 hours post-infection, the cells were fixed with 4% PFA, permeabilized in 0.1% Triton/1X PBS, and stained with propidium iodide (1 µg/ml), and checked for cell cycle arrest and infection by FACS.

15. Infection and recovery of integration sites

For infection, 293T and HT1080 cells were seeded at 2.5×10^5 and 1.5×10^5 cells per well, respectively, in six-well plates the previous day. Cells were spinoculated with the DNase-treated HERV-K_{CON} virus at 2000 rpm for 2 h at room temperature. Total DNA was collected 48 h post-infection.

Recovery of integration sites was performed as described (Wang et al., 2007). Two micrograms of genomic DNA were digested overnight with MseI or ApoI, ligated to linkers overnight at 16°C, and digested a second time with PstI and DpnI. Nested PCR was then carried out under stringent conditions using LTR primers complementary to HERV U3 sequences. Oligonucleotides used in this study are listed in Supplemental Table 1. DNA barcodes were included in the second-round PCR primers in order to track

sample origin (Hoffmann et al., 2007). Amplification products were gel-purified and sequenced by massively parallel pyrophosphate sequencing. Only sequences that uniquely aligned to the human genome by BLAT (hg18, version 36.1, >98% match score) and began within 3 bp of the LTR end were used in downstream analyses. Integration sites sequences have been deposited in GenBank under the accession numbers FI497131–FI498695.

Of the 25,102 sequences analyzed, 6873 showed a high-quality match to the HERV-K_{CON} vector using BLAT. Sequences were classified as 2-LTR circle if there was a match to the U5 LTR end in the expected orientation, while allowing indels of 100 bp. One-thousand-fifty-eight were an internal fragment derived from the internal U3 LTR and flanking sequences. A total of 3784 sequences showed the viral DNA end abutting internal HERV-K_{CON} sequences and were classified as autointegration products. Another 14 sequence reads had complex structures and were not included in the above categories.

Analysis of other retroviruses and genomic HERVs integration site data sets published previously (Table 3) were analyzed using the bioinformatics pipeline mentioned above. Discrepancies in data set sizes likely result from differences in quality-control thresholds compared with the original publications. The ERV2 data set was generated using RepeatMasker and the human genome (hg18, version 36.1). For the HML2(85) data set, the HERV-K_{CON} LTR sequence was used as a query to search for sequences 85% or higher in nucleotide similarity and longer than 600 bp using the Ensembl BLASTN (http://www.ensembl.org/Multi/blastview). The remaining sequences were organized by chromosomal location, and LTRs <9000 bp apart were manually determined as either solo-LTRs or LTRs of the same provirus based on the LTR flanking

sequence and identification of target site duplication sequence. Duplicate hits due to genome duplications or belonging to the same provirus were condensed into a single entry.

16. Analysis of integration site distributions

Analyses were carried out as described (Berry et al., 2006; Marshall et al., 2007). Analyses of gene expression used data from 293T cells, with expression measured using the Affymetrix HU133 plus 2.0 gene chip array. Expression values were ranked and divided into eight bins according to rank. Consensus sequence analysis at the point of integration was performed using WebLogo (http://weblogo.Berkeley.edu/logo.cgi), and the primary sequence features found to match those expected for HERV-K integration (data not shown).

$CD4^+$ T cells were used to generate ChIP-Seq data (Barski et al., 2007), differing from the cell types studied here. However, genome-wide surveys of modification densities in different cell types from the ENCODE project show that a substantial fraction of epigenetic marks are common to most cell types analyzed probably because a large fraction of transcription is from “housekeeping genes.” For example, for HIV data sets, there is no stronger correlation with epigenetic marks measured in T cells than for integration site data sets from T cells than from other cell types (C.C. Berry, T.L. Brady, F.D. Bushman, and K. Ronen, unpublished.). Furthermore, differences due to experimental error were generally greater than differences due to cell type (ENCODE Project Consortium, 2004). Thus the data from Barski et al represent a useful approximation to the cell types studied here.

III. RESULTS

Chapter 1. Endogenous HERV-K proviruses

1.1 Introduction to human endogenous retroviruses

Retrovirus-like LTR elements, also known as human endogenous retroviruses (HERV), make up more than 8% of the human genome (Lander et al., 2001). There are three main groups of human endogenous retroviruses: class I, II, and III, which are similar to gammaretroviruses, betaretroviruses, and spumaretroviruses, respectively (Bannert and Kurth, 2004). Regardless of genus affiliation, all HERVs known to date are heavily mutated with insertions, deletions, premature stop codons, and truncations, and are unlikely to be replication competent.

The class II HERVs are also known as the HERV-K superfamily, a designation based on the tRNA-lys binding site on the viral sequence, the interaction of which is believed to prime reverse transcription (Ono, 1986). The superfamily is subdivided into ten families from HML-1 to HML-10 based on LTR and RT sequence (Andersson et al., 1999; Medstrand and Mager, 1998). Each family is composed of insertions that are around 80% or more similar to each other in RT sequence (Andersson et al., 1999; Medstrand and Blomberg, 1993a). Of the ten subfamilies, HERV-K(HML-2) family has been a source of particular interest, as it is believed to be the youngest of the HERV-K superfamily, and most likely to include replication competent members.

1.2 HERV-K(HML-2)

Many insertions of HERV-K(HML-2) family are human-specific, indicating that they inserted after the human-chimpanzee divergence approximately six million years ago; some are polymorphic within human population indicating an even more recent insertion time (Belshaw et al., 2005a; Hughes and Coffin, 2004; Turner et al., 2001). Table 1 list the HERV-K(HML-2) LTRs that were found by Ensembl nucleotide to nucleotide BLAST using HERV-K_{CON} LTR (HERV-K_{CON} discussed later), organized by chromosome location. Only hits that were 600 bps or longer and 85% or higher in sequence similarity to HERV-K_{CON} LTR were included. The 600 bps limit was selected to eliminate SVA elements, a repeat element which include partial Env and 3' LTR sequences from HERV-K(HML-2) of around 490 bps in length and greatly outnumbers HERV-K elements (Wang et al., 2005). The 85% sequence similarity cut off was based on sequence similarity within HML-2 group and sequence divergence between HML-2 and HML-1, the subfamily closest in sequence to HML-2 (Medstrand and Blomberg, 1993a). 402 insertions were found. Including K113 and K103, which were discovered in a BAC screen and is not present in the NCBI human genome sequencing project sample, a total of 404 insertions are listed in Table 1 (Turner et al., 2001).

The number of insertions were compared to chromosome size (Lander et al., 2001). If HERV-K integrates randomly, the insertions should be found evenly across the chromosomes despite the differences in size, hence the number of insertions in a given chromosome should correlate proportionately with its size. Figure 8 displays the ratio of proportion of HERV-K insertions in each chromosome of total insertions to

Table 1 HERV-K(HML-2) LTRs in the human genome (pages 55-61)

HERV-K_{CON} LTR sequence was used as query in the Ensembl BLASTN search tool (http://www.ensembl.org/Multi/blastview) to identify HERV-K(HML-2) LTRs. Hits that were less than 600 bps were eliminated in final list. All LTRs are 85% or more identical to HERV-K_{CON} LTR, and organized by chromosome. Both solo-LTRs and LTRs associated with proviruses are listed. The LTRs in proviral context are only listed once, and the provirus identified. LTRs that were less than the distance of a full length HERV-K(HML-2) of 9472 bps were manually checked for proviral sequence or flanking five to six nucleotide repeats to confirm status.

			Chromosomal location		Statistics		Notes					Chromosomal location		Statistics		Notes
Start	End	Length	Start	End	E-val	% ID			Start	End	Length	Start	End	E-val	% ID	
Chromosome 1									1	965	970	108861590	108862554	0	92.9	
1	968	969	1335050	1336016	0	96.8			14	737	732	111704102	111704821	9E-305	88.9	
1	858	859	1505586	1506443	0	96.3			1	965	974	114402293	114403262	0	92.2	
252	965	717	10409558	10410267	0	92.9			1	725	732	143314771	143315496	0	90.4	
123	965	845	15578104	15578945	0	96.0			254	849	602	144213168	144213765	0	91.0	
1	965	972	25786928	25787892	0	92.1			1	725	731	144779880	144780605	0	91.0	
1	965	966	29537184	29538146	0	96.1			1	639	647	145044648	145045283	1E-246	88.4	
123	965	845	29558546	29559386	0	95.9			1	737	744	145178373	145179105	0	90.1	
1	668	676	33301704	33302369	2E-245	88.3			14	965	956	145672607	145673553	0	92.5	
5	668	670	36727459	36728118	2E-279	90.3			1	725	731	146076983	146077708	0	91.0	
1	965	970	40810285	40811250	0	93.5			1	725	732	146723221	146723946	0	90.7	
4	965	964	45766362	45767324	0	96.4			1	725	732	146832410	146833135	0	90.6	
1	965	975	46558555	46559519	0	92.3			1	965	967	153835955	153836920	0	96.1	
220	965	757	46568022	46568771	0	91.2			1	968	968	153863081	153872260	0	99.3	K102
1	968	970	52244505	52245471	0	97.9			1	968	968	154415638	154416605	0	99.3	
5	965	965	65378071	65379030	0	93.8			1	965	970	154417876	154418833	0	92.9	
123	968	847	66663086	66663931	0	98.1			1	965	965	158004210	158005174	0	95.9	
3	737	740	70698933	70699668	0	91.1			1	729	738	158888553	158889277	5E-303	89.3	
1	965	967	73367572	73368536	0	93.8			1	965	968	158927199	158936427	0	95.0	K18
1	968	968	75615359	75621731	0	99.0	K4		1	737	745	159185475	159186205	1E-300	88.7	
1	851	854	78221263	78222109	0	96.7			1	961	967	159253249	159254211	0	95.0	
123	968	849	89305134	89305980	0	96.4			1	639	649	163328444	163329078	4E-204	86.1	
1	968	968	93514927	93515894	0	99.1			1	965	972	205875083	205876042	0	90.3	

			Chromosomal location		Statistics		Notes					Chromosomal location		Statistics		Notes
Start	End	Length	Start	End	E-val	% ID			Start	End	Length	Start	End	E-val	% ID	
1	968	968	222594156	222595123	0	99.0			1	968	968	47276028	47276995	0	99.1	
1	965	966	226122667	226123632	0	95.5			123	968	848	50532441	50533283	0	97.5	
1	670	673	244312610	244313279	0	94.2			1	968	969	53986581	53987548	0	97.4	
123	965	844	245245711	245246553	0	96.7			1	716	721	75525034	75525743	7E-284	89.9	
Chromosome 2									1	616	621	75777169	75777782	0	88.9	
14	737	730	26826246	26826970	0	91.0			157	965	810	101474031	101474838	0	94.6	
1	968	968	27536350	27537317	0	99.3			5	965	973	101537863	101538821	0	91.9	
1	968	968	30689842	30690809	0	98.9			304	965	663	101867678	101868338	0	95.6	
1	965	972	32354158	32355122	0	93.0			1	965	970	102901581	102902546	0	95.6	olo LTRs
1	968	968	37305965	37306932	0	99.0			1	965	969	102905370	102906335	0	95.9	
1	965	972	39401803	39402769	0	94.1			125	965	850	113765620	113766460	0	90.9	
293	965	676	55360205	55360871	0	91.9			123	968	846	114225814	114234858	0	99.3	K106
1	965	967	86341933	86342898	0	95.2			1	639	649	119741740	119742376	0	87.5	
14	737	729	98112994	98113717	0	91.8			123	786	664	127092106	127101103	0	97.9	KI
1	737	744	100667760	100668495	0	89.7			1	786	786	130842811	130843595	0	97.3	
1	965	972	112436894	112437857	0	93.0			1	732	738	131258822	131259548	9E-274	89.6	
1	965	972	113408302	113409266	0	93.2			300	965	672	135717474	135718133	0	91.1	
123	965	848	130436011	130436854	0	95.6			14	736	730	146909964	146910682	0	89.2	
254	965	717	193114916	193115627	0	91.6			14	737	730	147722750	147723473	3E-302	90.7	
1	965	973	201711970	201712932	0	92.2			1	968	969	177106032	177106999	0	97.1	
123	968	847	207609916	207610762	0	99.1			1	968	968	186763030	186772209	0	98.6	K50B
1	598	602	208772524	208773125	9E-296	95.0			123	965	845	188087360	188088203	0	96.6	
307	965	661	215376160	215376818	0	96.1			1	965	966	188093703	188094666	0	96.8	
115	965	858	223763553	223764402	0	91.1			1	965	967	190451857	190452821	0	94.4	
1	965	967	228936605	228937569	0	96.8			1	968	968	197138793	197139760	0	99.4	
123	968	848	231416598	231417441	0	97.2										
123	968	846	232149853	232150698	0	97.9			Chromosome 4							
292	965	679	232996251	232996916	0	89.8			1	968	968	135521	136488	0	99.1	
4	965	968	234741451	234742413	0	92.2			4	734	740	191826	192551	3E-287	88.5	
									1	732	738	3978297	3979022	8E-283	89.3	
Chromosome 3									63	732	676	9342737	9343403	2E-285	88.9	
1	737	746	12690614	12691346	2E-290	89.5			51	767	718	9738282	9738990	0	91.9	
1	968	968	14107686	14108653	0	99.3			294	965	676	15687679	15688350	0	93.6	
1	668	675	14289349	14290011	7E-278	90.4			1	769	773	47982337	47983105	0	89.9	
1	965	966	23561161	23562120	0	95.6			1	968	969	63489191	63490158	0	97.7	
293	965	680	39434797	39435469	0	91.5			1	965	971	66907073	66908036	0	93.1	

			Chromosomal location		Statistics		Notes					Chromosomal location		Statistics		Notes
Start	End	Length	Start	End	E-val	% ID			Start	End	Length	Start	End	E-val	% ID	
1	968	968	73213805	73214772	0	99.6			1	965	966	180186875	180187835	0	95.6	
123	965	845	118323838	118324681	0	96.5			1	737	744	180626704	180627433	0	91.7	
1	968	968	120483137	120484102	0	99.2										
1	965	967	120536969	120537932	0	96.2			Chromosome 6							
123	968	846	157445364	157446209	0	99.1			1	965	969	2853896	2854860	0	94.4	
1	668	675	161117393	161118053	8E-282	88.9			1	737	740	24508490	24509226	0	91.9	
1	968	968	161799388	161800355	0	99.3			1	965	973	26107438	26108401	0	91.6	
4	965	966	166046119	166047082	0	96.2			1	737	745	26864459	26865189	3E-265	88.9	
1	965	970	166130985	166131949	0	93.0			1	965	968	27850221	27851186	0	96.4	
1	965	973	166136290	166143515	0	92.3	K5		1	965	970	32853826	32854790	0	95.0	
1	725	730	175540755	175541480	2E-285	91.2			123	968	847	33885715	33886560	0	97.8	
1	725	730	175558486	175559210	2E-266	90.1			1	965	968	34795218	34796183	0	96.2	
									1	668	669	44404059	44404724	0	96.0	
Chromosome 5									1	668	674	52734587	52735250	0	88.7	
1	968	971	1649092	1650059	0	97.4			123	968	847	52895551	52896397	0	98.9	
1	965	968	4978041	4979006	0	96.1			1	679	683	74571757	74572435	0	94.3	
1	968	969	8990854	8991820	0	98.0			123	968	846	78483381	78492688	0	99.4	K109
1	965	967	18614320	18615285	0	96.2			1	968	968	79625224	79626191	0	99.1	
123	968	848	30522517	30531624	0	97.8	K104		1	968	968	89148026	89148993	0	99.1	
1	968	971	35212244	35213212	0	97.6			2	968	969	93939805	93940769	0	97.7	
304	965	664	43616583	43617245	0	93.7			1	872	873	99983968	99984839	0	96.5	
1	968	969	44766346	44767313	0	98.1			254	965	720	121294551	121295255	0	88.2	
123	968	846	54902779	54903624	0	98.8			1	965	971	126138689	126139654	0	93.7	
1	737	746	55488581	55489314	1E-275	89.1			1	968	969	135021941	135022908	0	98.5	
1	965	967	58795369	58796334	0	95.9			294	965	673	151818041	151818712	0	97.5	
123	968	846	74937414	74938259	0	99.7			1	737	743	158093470	158094201	0	89.2	
1	613	619	92818429	92819040	2E-294	90.2			1	965	967	160135762	160136717	0	94.4	
14	640	632	105623100	105623722	1E-285	90.7			123	968	851	169488305	169489152	0	96.2	
123	968	847	116184848	116185693	0	98.1										
1	786	788	119558459	119559245	0	98.5			Chromosome 7							
1	636	646	149302586	149303220	4E-217	85.9			1	965	967	2395571	2396533	0	94.4	
14	737	733	149813747	149814464	0	88.7			1	968	968	4588583	4606557	0	99.3	K108
1	968	968	156017295	156026474	0	99.2	K107		1	732	741	6908153	6908879	7E-282	88.7	
123	968	846	169356457	169357302	0	98.1			115	917	814	7007983	7008786	7E-304	89.8	
5	618	620	178157687	178158298	5E-230	90.2			1	968	968	16203872	16204839	0	99.6	
1	968	968	178873570	178874537	0	98.0			1	968	968	23046000	23046967	0	99.2	

			Chromosomal location		Statistics		Notes					Chromosomal location		Statistics		Notes
Start	End	Length	Start	End	E-val	% ID			Start	End	Length	Start	End	E-val	% ID	
1	668	677	23763659	23764323	1E-226	88.3			123	965	845	145003872	145004708	0	95.7	
1	965	968	27747807	27748765	0	94.7			1	965	966	145992083	145993047	0	96.8	
14	737	733	47995810	47996523	3E-302	89.2										
115	818	707	54710547	54711250	0	91.9			Chromosome 9							
123	965	846	64340835	64341677	0	96.2			1	737	748	11884757	11885489	9E-260	88.0	
1	737	745	65721614	65722348	3E-278	88.6			1	968	970	17435359	17436326	0	97.4	
1	965	967	100579858	100580822	0	96.0			366	965	603	26705151	26705751	0	93.0	
1	968	969	104175605	104176571	0	98.4			1	968	968	31622877	31623844	0	98.6	
123	965	846	112768812	112769654	0	96.7			1	668	677	33505170	33505834	3E-264	89.2	
1	643	652	118527860	118528500	2E-262	89.3			123	968	846	67767163	67768008	0	98.7	
1	968	968	123207927	123208894	0	99.1			1	968	968	71567873	71568840	0	98.8	
1	968	968	124648460	124649427	0	99.3			2	965	966	74891946	74892910	0	96.0	
1	968	968	125595450	125596417	0	98.4			123	968	846	110396828	110397673	0	98.6	
1	965	972	138802691	138803656	0	93.4			1	737	744	112628352	112629081	2E-256	88.2	
1	968	968	157722244	157723211	0	99.5			14	737	733	113677413	113678133	3E-272	89.4	
									1	968	968	123231205	123232172	0	99.2	
Chromosome 8									123	968	846	133222041	133222884	0	99.1	
1	968	968	7342807	7352155	0	99.3	K115		123	965	846	135812837	135813681	0	97.0	
1	968	968	18695739	18696706	0	99.3			4	965	964	135947382	135948344	0	97.2	
1	968	968	37170044	37171011	0	99.9			1	668	680	136746896	136747562	0	87.7	
1	737	744	39627707	39628434	8E-274	89.5										
1	965	968	42769997	42770963	0	95.7			Chromosome 10							
1	968	970	43713839	43714806	0	96.9			1	965	969	6906150	6915609	0	93.0	K33
280	968	690	48206821	48207509	0	98.0			1	968	968	27222405	27223372	0	99.2	
1	968	968	55106794	55107761	0	99.6			1	968	968	27182398	27183366	0	99.0	K103
123	968	848	58274716	58275563	0	97.1			1	968	969	41846062	41847027	0	98.3	
1	737	745	59747085	59747817	5E-302	90.7			1	968	971	43152657	43153624	0	97.5	
1	968	968	91765436	91766403	0	98.6			1	965	968	66838997	66839960	0	94.2	
123	968	847	91766405	91767250	0	98.5			1	968	969	69955435	69956399	0	97.6	
1	732	738	113853579	113854311	0	90.9			268	908	645	99166645	99167278	0	92.9	
110	962	860	120998825	120999675	0	90.8			123	754	633	101570798	101571430	0	96.8	
14	733	726	138496009	138496728	2E-296	90.5			1	707	709	104141192	104141899	0	96.9	
1	968	968	140541331	140542298	0	98.6			1	965	972	104204125	104205082	0	91.1	
1	736	746	143907261	143907988	6E-237	88.2			1	965	968	132316211	132317170	0	94.1	
1	965	967	144440838	144441796	0	95.8										
1	968	970	144986147	144987115	0	97.5										

			Chromosomal location		Statistics		Notes					Chromosomal location		Statistics		Notes
Start	End	Length	Start	End	E-val	% ID			Start	End	Length	Start	End	E-val	% ID	
Chromosome 11									1	668	673	57223515	57224172	6E-268	89.8	
1	968	968	10369432	10370399	0	98.6			1	737	742	71689007	71689742	5E-296	89.9	
14	639	635	18878249	18878870	9E-242	87.2			1	670	676	84387145	84387812	2E-285	90.8	
123	968	846	24424330	24425174	0	98.6			1	965	967	91992310	91993273	0	96.3	
1	958	963	54955047	54956001	0	92.2			1	965	968	104396470	104397435	0	96.3	
5	737	742	55581713	55582438	4E-277	89.1			1	968	968	109492226	109493193	0	99.2	
1	968	969	61179061	61180024	0	96.3			123	965	845	117028286	117029128	0	96.9	
1	965	968	61718708	61719673	0	96.3			14	732	726	121801603	121802322	0	91.1	
14	737	732	61850657	61851377	3E-290	89.6			275	965	692	122543380	122544070	0	95.8	
123	965	847	62382569	62383410	0	96.3										
1	965	967	63052022	63052986	0	96.8			Chromosome 13							
1	968	968	63054362	63055329	0	98.6			1	968	969	19072358	19073325	0	97.7	
123	968	846	67127401	67128246	0	99.3			1	737	741	22286024	22286759	1E-297	91.5	
1	737	741	67281936	67282660	7E-280	89.3			1	965	967	49071159	49072124	0	96.6	
123	968	847	67392011	67392856	0	98.4			1	965	973	53738946	53739902	0	89.9	
1	968	968	71553066	71554033	0	99.0			1	965	973	69087014	69087975	0	91.7	
1	737	743	95962343	95963072	3E-282	89.6			1	664	679	94718276	94718939	8E-225	87.6	
1	968	968	101071004	101080469	0	99.4	11q22									
1	737	747	104640788	104641519	1E-265	88.6			Chromosome 14							
1	965	967	118096937	118106093	0	96.3	K37		1	965	972	19622586	19623552	0	92.1	
264	965	706	118415195	118415894	0	91.5			1	965	969	19806922	19807886	0	93.5	
									1	965	967	22263737	22264701	0	95.6	
Chromosome 12									1	965	967	37657055	37658020	0	95.7	
2	965	973	4700661	4701626	0	93.2			1	968	968	64515059	64516026	0	98.8	
1	668	675	5516444	5517111	8E-289	91.0			1	737	745	77171609	77172341	3E-277	88.6	
1	968	968	6865285	6866246	0	97.0			1	965	970	77197497	77198461	0	94.2	
1	965	968	8505795	8506759	0	94.9			123	965	846	77330773	77331616	0	96.2	
1	968	969	9644328	9645295	0	98.5			1	965	971	99847884	99848848	0	91.8	
1	737	745	10615370	10616102	2E-272	89.8										
1	968	970	29976254	29977221	0	98.5			Chromosome 15							
1	968	968	32143712	32144679	0	99.1			1	716	721	56913073	56913787	0	91.1	
5	732	734	36418393	36419120	2E-282	88.8			1	968	970	63305933	63306900	0	97.6	
1	968	968	50134338	50135305	0	99.2			1	965	967	63813045	63814010	0	96.0	
1	968	969	54013482	54014450	0	98.8			3	965	964	73950253	73951216	0	95.6	
1	964	964	55080416	55081376	0	98.6			1	968	968	86884787	86885754	0	99.0	
1	968	968	57007509	57016965	0	99.3	K41		1	968	970	100217703	100218670	0	97.9	

			Chromosomal location		Statistics		Notes					Chromosomal location		Statistics		Notes
Start	End	Length	Start	End	E-val	% ID			Start	End	Length	Start	End	E-val	% ID	
Chromosome 16									Chromosome 19							
1	964	964	5744394	5745357	0	99.6			130	965	838	337801	338634	0	96.1	
1	968	968	8168764	8169731	0	99.0			127	769	648	8285858	8286501	0	90.7	
4	965	967	14637508	14638461	0	92.8			1	680	680	12203926	12204605	0	96.2	
14	618	610	21140961	21141561	1E-252	90.7			14	668	663	20054160	20054811	1E-291	89.0	
123	968	848	23518420	23519265	0	98.1			1	737	742	20440977	20441705	0	89.2	
123	965	844	34090800	34091640	0	97.8			1	961	964	21542584	21543546	0	94.4	
1	965	967	35047430	35048394	0	95.5			1	968	968	21633376	21633376	0	99.0	K113
1	968	968	46455871	46456838	0	99.3			246	965	722	22258423	22259142	0	96.7	
1	957	957	73390803	73391759	0	99.3			83	769	688	22549664	22556398	0	92.6	K51
									76	968	893	32820338	32821230	0	98.8	
Chromosome 17									1	965	970	33313377	33314342	0	93.3	
1	968	969	4918699	4919665	0	98.0			1	965	967	33872062	33873027	0	95.9	
1	965	969	19348866	19349827	0	94.9			1	668	678	40102922	40103586	3E-255	88.9	
2	737	746	25914572	25915305	7E-272	89.0			1	737	741	41416343	41417074	0	91.5	
1	968	968	26051185	26052152	0	99.0			123	786	664	41429889	41430552	0	99.0	
1	965	969	31491551	31492519	0	96.4			1	965	968	42188901	42189859	0	94.9	
123	965	844	38787962	38788804	0	95.1			1	639	645	42298271	42298906	5E-301	88.8	
1	965	970	41716760	41717724	0	93.6			14	885	876	42712628	42713500	0	93.6	
1	968	969	49633934	49634901	0	98.0			1	885	886	42812769	42813653	0	97.6	
1	965	969	60356602	60357560	0	92.7			1	968	968	43048657	43049624	0	99.7	
115	965	859	62753427	62754279	0	91.5			137	965	836	43813622	43814444	0	91.8	
123	968	847	64111728	64112573	0	97.8			110	965	860	45156981	45157833	0	90.7	
1	965	967	76139606	76140571	0	96.3			1	968	969	49789258	49790225	0	98.4	
1	965	968	78295700	78296662	0	95.4			1	965	965	54084705	54085664	0	95.9	
									1	668	674	57100757	57101423	1E-275	91.1	
Chromosome 18									1	965	967	57103260	57104222	0	95.7	
1	968	968	1990815	1991782	0	98.2			1	965	967	57238197	57239162	0	96.9	
1	968	969	4907278	4908246	0	98.0			1	618	624	57534186	57534797	2E-295	87.2	
1	965	967	22035800	22036764	0	95.8			38	965	931	57616024	57616952	0	94.5	
1	668	674	27644462	27645126	4E-253	89.6			5	737	742	57680607	57681333	7E-261	88.1	
1	737	745	37017830	37018561	1E-265	89.0			1	737	744	57753762	57754488	5E-262	88.2	
1	736	744	62895237	62895969	7E-295	90.2			1	962	965	57892811	57893771	0	93.9	
1	737	747	63923695	63924425	3E-268	88.5			1	965	968	59033692	59034655	0	93.7	
14	732	728	64760725	64761441	1E-285	89.8			1	959	964	60153473	60154429	0	92.3	
1	965	967	64861072	64862037	0	96.0			3	731	748	63026807	63027532	2E-180	84.1	

			Chromosomal location		Statistics		Notes					Chromosomal location		Statistics		Notes
Start	End	Length	Start	End	E-val	% ID			Start	End	Length	Start	End	E-val	% ID	
Chromosome 20									Chromosome 22							
1	767	771	7884847	7885613	0	92.2			1	968	968	17306187	17315361	0	98.9	K101
1	668	678	23623299	23623964	1E-224	87.8			1	737	745	21322053	21322789	1E-279	89.5	
123	947	827	23789256	23790081	0	95.8			1	737	745	22208249	22208982	3E-273	89.7	
1	733	742	23911278	23912007	4E-253	89.0			1	965	967	22579271	22580233	0	93.7	
14	965	960	25162477	25163429	0	92.6			123	965	846	22936863	22937703	0	96.0	
1	965	972	25169882	25170848	0	93.2										
1	968	969	33313343	33314310	0	97.4			Chromosome X							
1	968	968	40032950	40033917	0	99.1			83	737	661	57378544	57379199	0	89.9	
									1	668	678	89354986	89355649	5E-243	88.5	
Chromosome 21									282	965	693	122641474	122642160	0	89.8	
1	965	970	14114072	14115037	0	92.8			1	639	647	133992041	133992677	3E-231	88.1	
1	821	825	14575284	14576104	0	90.3										
1	961	967	17988412	17989372	0	92.5			Chromosome Y							
1	965	970	18226910	18227874	0	92.7			1	968	969	6676930	6677896	0	97.0	
1	966	967	18862868	18863833	0	97.8			1	965	969	13084687	13085651	0	92.7	
1	965	965	41714669	41715633	0	95.5			1	968	970	13725131	13726097	0	96.5	
123	968	847	43440309	43441154	0	98.4			1	670	681	16394742	16395407	1E-202	86.9	
115	965	855	44527174	44528018	0	90.5			1	965	967	20320758	20321721	0	93.8	
									1	968	968	23448574	23449541	0	99.7	
									1	968	968	25082323	25083290	0	99.7	
									1	968	968	25697883	25698850	0	99.6	

proportion of each chromosome length of total chromosome length. A ratio of 1 should indicate that the number of insertions reflect the size of the chromosome. Unexpectedly, chromosome 19 has far more insertions than expected by size while chromosomes 13, 15, and X have far less. Other chromosomes have modest or no differences relative to expected according to chromosome size. Aside from chromosomal size, other factors, such as the selection of insertions based on their effect on host fitness and the fixation of insertions at the population level may also determine HERV accumulation after infection. A more detailed study is needed to determine the true integration preference and forces influencing the fixation of HERV-K(HML-2) in primate genome. How chromosomal characteristic affects HERV-K integration preference will be addressed in Chapter 4.

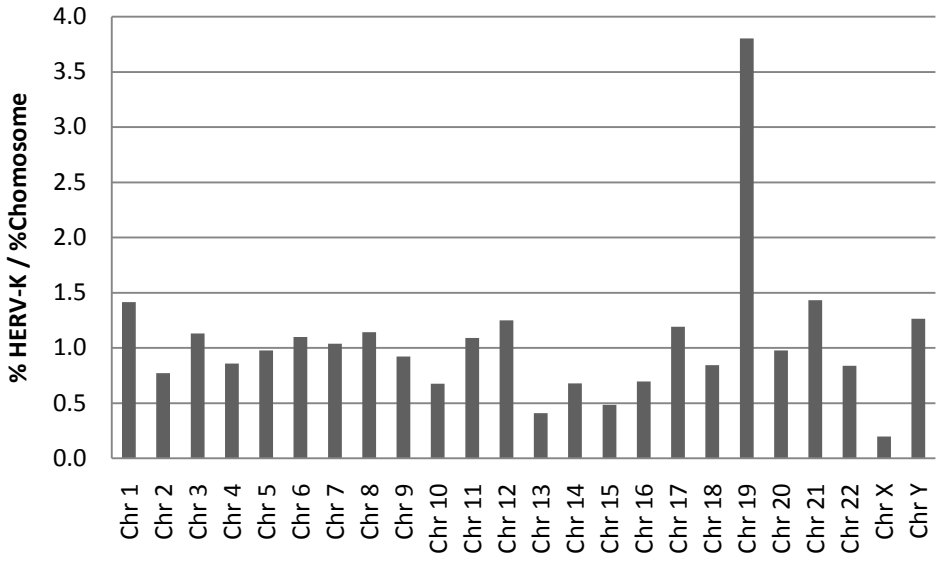


Figure 8 Chromosomal distribution of HERV-K(HML-2) LTRs.

The chromosomal distribution of HERV-K(HML-2) LTRs is represented as the ratio between the proportion of HERV-K insertions per chromosome of total insertions and proportion of each chromosome length of total genome length. A ratio of 1 should indicate that the number of insertions reflect the size of the chromosome. The chromosome size is taken from Lander et al, and the number of insertions from Table 1 (Lander et al., 2001).

Although LTRs are the most abundant forms of HERV, there are also numerous full length and truncated proviruses. Table 2 lists the insertions that include proviral sequence in addition to LTR, including both full-length and truncated insertions. The amino acid sequence of HERV-K_{CON} Gag (discussed later) was used as query to search for HERV-K insertions with coding sequence via TBLASTN search tool. Only hits that were 85% or higher in sequence similarity to HERV-K_{CON} were included based on reasons described above. This search yielded forty one hits, of which twenty three are full length, arbitrarily defined as 90% or greater in total sequence length relative to the full length HERV-K (9472 bp) (Ono et al., 1986). Of these, seventeen proviruses are found in human but not chimpanzee reference genome, although the sequence and location of K103 is unknown (Barbulescu et al., 1999). Three proviruses, K4, K106, and K113, have identical LTRs, suggesting insertion events 200,000 years ago or less (Belshaw et al., 2005a; Johnson and Coffin, 1999).

A few insertions had non-matching flanking nucleotide repeats, suggesting that the insertion contains sequences derived from more than one provirus. Other groups have suggested that such insertions likely reflect homologous recombination or gene conversion of the genomic locus with a solo-LTR, which outnumber full-length elements by ten-fold or more, rather than a recombination of two full length elements at internal ORF sequences (Belshaw et al., 2004; Hughes and Coffin, 2001; Johnson and Coffin, 1999). Inspection of other primate genomes may be able to resolve this issue depending on the length of host genomic sequence included in the recombination or gene conversion.

Table 2 Proviral HERV-K(HML-2) in the human genome (pages 64-65)

Amino acid sequence of HERV-K_{CON} Gag was used as query to search for HERV-K(HML-2) insertions with internal sequence in the Ensemble TBLASTN search tool (http://www.ensembl.org/Multi/blastview). Only sequences 85% or higher in similarity were included. One flanking short repeat sequence is included when the two ends are identical, two ends are included when they do not match or cannot be clearly determined. The numbering of each insertion is according to the HERV-K_{CON} sequence. The proviruses described by Hughes et al have been cross-referenced (Hughes and Coffin, 2001). The names of insertions are from previous publications (Barbulescu et al., 1999; Hughes and Coffin, 2004; Romano et al., 2006; Sugimoto et al., 2001; Turner et al., 2001). FL, full length; HS, human specific; ID LTRs, identical LTRs.

Location	Flanking	Name	Contig	Features	State
1p31.1	ATGGAA	K4	AC093156	HS, ID LTRs	1-3505, 6313-6502, 6795-end
1q22	GGGATG	K102, K50A	AL353807	HS	FL. By ensembl blast, only one provirus found.
1q23.3	TGAGAC	K110, K18	AL121985		FL
1q32.2	GCATTC	None	AL137789		5929-end
2q21.1	AGAACT	None	AC079776		6094-end
3p25.3	CTTGGT/GAAAGT	K11	AC018809		Missing 3684-5940
3q12.3	GAGGT	KII	AC084198		FL
3q13.2	GGCTGG	K106, K68	AC078785	HS, ID LTRs	FL
3q21.2	GGCCC	KI	AC092903	HS	FL
3q24	Unknown	None	AC069410		961-4887
3q27.2	GGTACA	K50B	AC099661	HS	FL
4q32.3	CTTTCT/TTTTAT	K5	AC106872		Missing 3686-5992.
5p13.3	CAGAAC	K104, K50d	AC025757	HS	FL. K104 and 50a are the same element.
5p11	CTCCC	K8	AC126750		FL, with insertion
5q33.3	ACTGC	K107, K10	AC016577	HS	FL
6p22.1	CCTGGG	HERV-K20	AL121932		196-end, has internal solo LTR with flanking seq GATCCC.
6q14.1	ATATGC	K109	AL590785	HS	FL
7p22.1	GGTTTC	K108, C7	AC072054	HS	FL
7q22.2	Unknown	None	AC079796		969-4887
7q34	Unknown	None	AC004979		937-4886
8p23.1	CCTTT	K115	AC134684	HS	FL

Location	Flanking	Name	Contig	Features	State
8q24.3	Unknown	None	AC087354		947-2246
10p12	ATGGGG	K103	AF164611	HS	FL
10p14	TCATTC	K33	AL392086		FL
10q24.2	CAGGTG	None	AL392107		940-end, missing chunk of Env
11q12.3	TGGATT/ATCATT	None	AP003064		Two parts, start-6966, 6955-end.
11q22.1	TTGTG	11q22, K36	AP000776	HS	FL
11q23.3	AGCCT	K37	AP002954		FL
12p11.1	CTGCTC/unclear	K50e	AC144535		FL, 88% identical to HERV-Kcon
12q14.1	TTGGTA	12q14, K41	AC025420	HS	FL
12q24.11	AGTATT/Unknown	None	AC002350		1-1484
16p11.2	CTGAGG	None	AC135776		6795-end
19p13.3	Unknown/CAGGTC	None	AC010641		6922-end
19p12	Unknown/TGTAAT	K51	AC011467		83-end, missing 3685-5940
19p12	CTCTAT	K113	AY037928	HS, ID LTRs	FL
19q12	AGGTAT	None	AC112702	HS	947-end, but flanking sites both there.
19q13.12	Unknown	None	AD000090		3329-3922, 5328-8625
19q13.42	Unknown/GGCTGA	None	AC010467		4788-5713, insertion of 1011bps, 5709-end.
20q11.22	Unknown, unclear	None	AL031668		4604-6161, insertion of 1916bps, 6610-end
21q21.1	GCCAGG/Unknown	K60	AL109763, AF2	HS	Start - 8761
22q11.21	ACCCAG	K101	AC007326	HS	FL

1.3 HERV-K113 LTR

Of the twenty three full length HERV-K(HML-2) proviruses (Table 1), HERV-K113 is believed to be the most likely candidate to be replication competent (Turner et al., 2001). This idea is based on three major points: its polymorphic state, the identical sequence of its LTRs, and its complete open reading frames (Burmeister et al., 2004). As described in the Introduction, an element that is polymorphic in the human population is thought be less than 800,000 years old based on Kimura's neutral theory (Graur and Li, 2000). The identical nature of K113's LTRs decrease its estimated age to around 200,000, based roughly on substitution rate between humans and chimpanzees (Johnson and Coffin, 1999). Lastly, K113 is the only provirus described to date with the complete open reading frames that lack obvious mutations such as premature stop codons, mutations that inactivate essential motifs, or large truncations of sequence. Hence, replication potential of K113 was examined. The LTRs were chosen as the first subject of inquiry.

Transcriptional tropism of a virus depends on the expression level of the necessary host factors the LTR interacts with to induce transcription. HERV-K(HML-2)'s expression tropism in teratocarcinomas was discovered in the late 1970s, when virus particles were detected in cells derived from teratocarcinomas by EM (Boller et al., 1993b; Bronson et al., 1978; Lower et al., 1981). To become endogenized, all ERVs must infect germ cells or progenitors, but the actual target cell of any ERV is unknown. Teratocarcinomas, as malignant cancer of germ cells derived from the testes, are closely related to potential natural target cells for ERV, and transcriptional activity in these cells may reflect the natural situation of HERV-K infection and endogenization. Work by Ruda et al have shown that HERV-K LTR is as active as SV40 promoter in

teratocarcinoma cell line, Tera-1, which is tenfold or more active relative to other cell lines (Ruda et al., 2004). Also, Ruprecht et al showed that numerous HERV-K(HML-2) transcripts are expressed in Tera-1 cell line, especially HERV-K101, and packaged preferentially into particles over other RNA (Ruprecht et al., 2008). Thus, it is well established that the HERV-K LTR can induce transcription and is active in teratocarcinoma-derived cells.

To test whether K113 LTR is functional, 293T cells were transfected with pCR3.1/K113 LTR GFP and HTLV-1 Env. 12 hours post transfection, 293T cells were fused to either fresh 293T or NCCIT cells, which are teratocarcinoma cells known to express HERV-K proteins at high levels which assemble into particles (Bieda et al., 2001). As expected, 293T cells expressed an increased quantity of GFP when fused to NCCIT cells, but not 293T cells (Figure 9). This data indicates that K113 LTR is functional, and transcription from it and the resulting virus replication is cell type dependent. This also suggests that factor or factors responsible for the increased expression from K113 LTR is dominant and expressed in NCCIT cells, but not 293T cells.

1.4 HERV-K113 and YY2

Retrovirus long terminal repeats contain regulatory elements that control the transcription of the viral genome, which in turn controls the replication of the virus. The LTR is divided into three main regions U3 (unique 3'), R (repeat), and U5 (unique 5'), which are defined by the start (U3-R) and end (R-U5) of transcription. R region is especially important for reverse transcription, ensuring that all necessary virus sequence

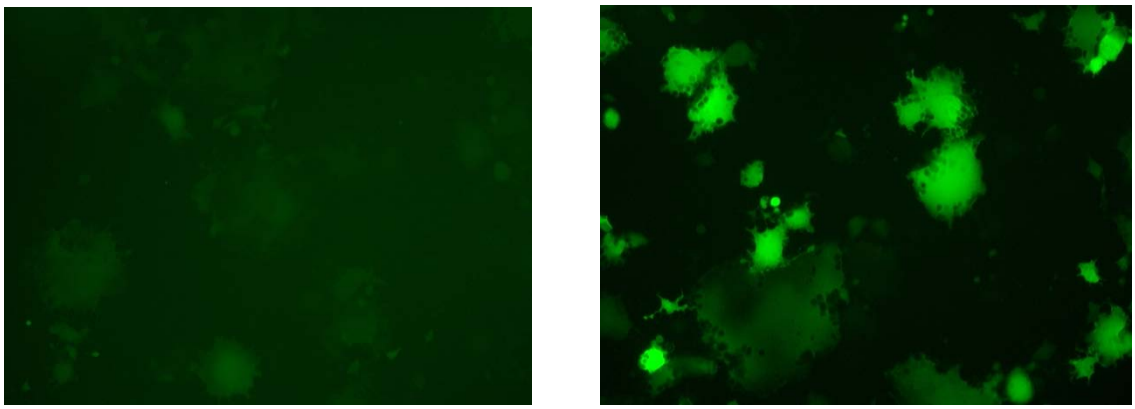


Figure 9 The activity of HERV-K(HML-2) LTR inducing factor or factors in NCCIT cells

293T cells were transfected with pCR3.1/K113 LTR-GFP and HTLV-1 Env, which uses GLUT-1 glucose transporter as a receptor to enable fusion of membranes expressing the two proteins. 12-hours post transfection, the cells were incubated with untransfected 293T cells or NCCIT teratocarcinoma cells, and grown for an additional 24 hours. 293T cells fused to 293T cells do not show an increase in GFP levels detectable by eye, while those fused to NCCIT cells are noticeably brighter. The pictures shown are representative of three experiments.

The TATA box is located towards the end of U3 at the 5' LTR. The 3' LTR also contains the same functional promoter sequences, but only the 5' LTR is used for the proviral transcription. The remaining U3 region likely contains regulatory sequences to aid transcription, such as binding sites for transcription factors. The HERV-K LTR is nearly ten-fold more active in teratocarcinoma derived cell line than non-teratocarcinoma derived human cell lines, but the determinants of this tropism is unknown (Casau et al., 1999; Ruda et al., 2004). A putative glucocorticoid responsive element sequence (nucleotides 75-88) has been found in the HERV-K LTR but remains functionally untested (Ono, 1986; Ono et al., 1986). Another partially overlapping fragment of around 20 nucleotides (nucleotides 62 to 83) was found to be important for transcription enhancement in teratocarcinoma derived cells via several DNA-binding complexes by two independent groups (Akopov et al., 1998; Knossl et al., 1999). One complex was found to include the transcription factor YYI, but this complex was not responsible for

the teratocarcinoma cell line specificity of HERV-K transcription, as it was found also in non-teratocarcinoma cell lines HeLa and HepG2 as well (Knossl et al., 1999). Another complex, which was specific to teratocarcinoma cell lines tested and binds to the same sequence region, remains unidentified.

Since the study of YY1 and its effects on HERV-K LTR, a second family member has been discovered. YY2 is the result of a YY1 mRNA retrotransposition event that occurred around 60 to 100 million years ago, as it is found only in placental mammals (Kim et al., 2007). YY2 shares 56.2% sequence similarity to YY1 in amino acid sequence, and 86.4% similarity in the DNA binding zinc finger region (Nguyen et al., 2004). It has also been shown to have the same DNA binding motif as YY1, and can bind and modulate sequences that YY1 binds and regulates (Kim et al., 2007; Nguyen et al., 2004). Thus, YY2 is a valid alternative regulatory factor to YY1, and a candidate for the teratocarcinoma cell line -specific factor for enhancing HERV-K transcription.

To test this idea, 293T cells were transfected with plasmids expressing GFP under the control of either HIV-1 or HERV-K113 LTR and pCR3.1/Flag-YY2 or control plasmid pCR3.1. As shown in Figure 10, expression of GFP increased modestly in both LTRs when coexpressed with YY2: 1.5-fold in HIV-1 LTR and less than 3-fold in K113 LTR relative to controls not expressing YY2. Similar results were obtained with untagged YY2 and HERV-K LTR, suggesting that the flag tag did not interfere with protein activity. Although GFP expression did increase when YY2 was overexpressed, the three fold increase does not account for the tenfold or higher increase detected in other teratocarcinoma cell lines (Casau et al., 1999; Ruda et al., 2004). This data suggests

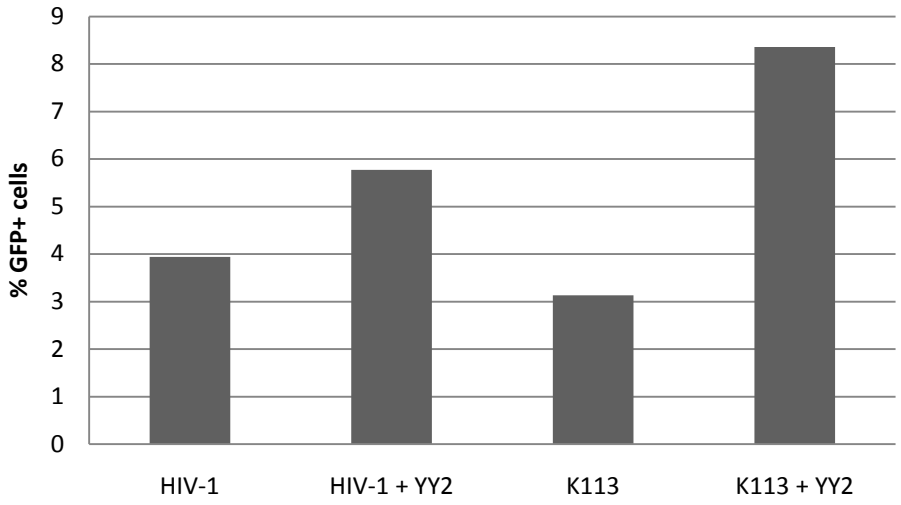


Figure 10 The modest transcription enhancing activity of YY2

1.5 µg of pCR3.1/HIV-1 LTR-GFP or pCR3.1/K113 LTR-GFP plasmids were transfected into 293T cells in a 6-well format with 1.5 µg of FLAG-YY2 or pCR3.1. Two days post transfection, the percent of GFP^+ cells was accessed by FACS.

that YY2 is unlikely to be the only or the main upregulator of HERV-K expression in teratocarcinomas.

1.5 Replication capacity of K113 and K108

Thus far, with its active but tissue specific LTR, K113 still remains as the most likely candidate fo r replication competent HERV. Hence, K113 was cloned to test its replication potential. The constructed proviral plasmids are depicted in Figure 11. As K113 LTR was shown to be inactive in 293T cells, a derivative where the 5' LTR's U3 region was replaced with a cytomegalovirus promoter sequence to enhance transcription was also constructed. However, one study showed that K113 Env could not induce infection of retroviral VLPs despite K108 Env being able to do so, suggesting that K113 Env, while full length, is functionally defective (Dewannieux et al., 2005). Despite its functional Env, HERV-K108 has a premature stop codon in the Gag ORF, a frameshift in

protease, as well as a mutation of a highly conserved YIDD motif in reverse transcriptase into CIDD (Mayer et al., 1999b; Reus et al., 2001). This mutation is believed to render the full length virus replication incompetent, making K108 an unlikely candidate for replication.

Although K113 and K108 may be individually nonfunctional, it was possible that a hybrid between the two proviruses may be replication competent. To test this hypothesis, K113 K108 hybrid construct was cloned by replacing the K113 Env and 3' LTR sequence with that of K108 (Figure 11). A CMVP containing K113 K108 hybrid was constructed as well.

The plasmids were transfected into 293T cells; two days post-transfection the cells were lysed and the VLPs in the supernatant spun through 20% sucrose gradient to purify and concentrate and the particles. Western blotting analysis of the cell lysate and VLPs using the commercially available antibody (Austral Biologicals) showed no expression of HERV-K Gag or VLPs (data not shown), indicating that both K113 and K113K108 hybrid proviruses may not express viral proteins at a detectable level regardless of the promoter used.

The HERV-K genome has an unusual nucleotide composition in that it is relatively A-rich (Zsíros et al., 1999). For example, K113 is composed of 32% A, 26% T, 21% G, and 21%C. This feature, which is characteristic of lentiviruses such as HIV-1, is partly responsible for the nuclear retention of HIV-1 mRNAs and contributes to the requirement for Rev in mediating export of incompletely spliced HIV-1 transcripts.

Indeed, HERV-K encodes a functional ortholog of the Rev protein, termed K-Rev

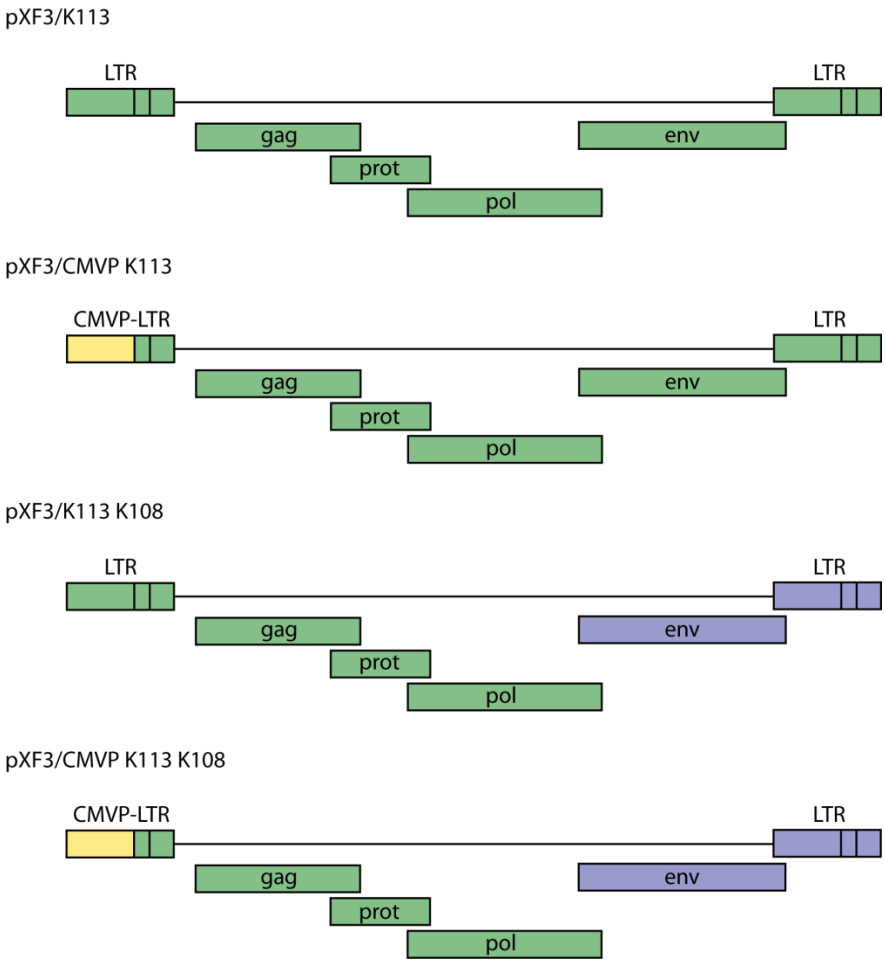


Figure 11 K113 and K108 derived proviral constructs

K113 construct and its derivatives CMVP-K113, K113 K108, and CMVP K113 K108 are depicted. LTRs and ORFs are depicted as boxes. The 5' LTR U3 region is replaced with the CMVP up to the TATA box, represented by the yellow box. The green and blue boxes represent K113 and K108 sequences, respectively. K108 sequences are inserted into the SacII restriction site in Env ORF, which is 37 nucleotides from the ATG of Env and is identical in amino acid sequence between the two proviruses.

or Rec, which mediates nuclear export of HERV-K RNA (Boese et al., 2000; Magin et al., 2000; Magin et al., 1999; Yang et al., 2000; Yang et al., 1999). Therefore, because of the likely requirement for a Rev-like post-transcriptional activator for efficient HERV-K mRNA export, K108 Rec was cloned to test its effect on K113 and K113-K108 replication. K108 Rec differs only in amino acid 52 (Ser to Thr) from a functionally tested Rec (Magin et al., 1999; Yang et al., 1999). Thus, K108 Rec was co-transfected

with viral genomic plasmids to enhance nuclear export of viral transcripts. No VLPs were detected in the supernatant via western blotting analysis (data not shown).

Thus far, infection capability of HERV-K(HML-2) was tested by enhancing transcription by insertion of CMVP, replacing a defective Env with a functional Env, and increasing transcript nuclear export by co-expression of Rec. The expression and function of viral proteins were also tested independently from rest of the virus via cloning the ORFs into a different expression plasmid. Gag, Gag-protease (PR), and Gag-PR-Pol of HERV-K113 were cloned into HIV-1 based expression vector pCRVI, which eliminates potential expression problems of K113 due to expression of HIV-1 accessory proteins that aid in transcription, and nuclear export of transcript (Figure 7). Hence, any sequence cloned into this plasmid should be expressed at high levels. Despite these optimal conditions, the transfection of pCRVI based HERV-K plasmids resulted in poorly expressed proteins that were inefficiently released as VLPs, like the whole genomic plasmids (data not shown).

Chapter 2. Derivation of HERV-K_{CON} and the single-cycle infection system

2.1 Derivation of HERV-K_{CON}

Despite the aid for transcription from CMVP and transcript export from Rec, and the cloning of the ORFs into high-expression plasmids, K113 and K113-K108 failed to produce detectable quantities of VLPs. Reasons may be numerous, but as there are no obvious mutations in K113, it is difficult to locate the source or sources of the problem. To bypass this difficulty, a consensus HERV-K sequence was derived. This idea was based on the assumption that any inherent replication defects encoded within HERV-K(HML-2) proviruses present in contemporary human DNA are either unique to each provirus or shared only by a minority of recently integrated proviruses. If this assumption is correct, then each individual defect should be selected out from a sequence representing the consensus of a collection of proviruses, even if each individual provirus that contributes to the consensus is defective.

To derive the consensus sequence, HERV-K113 was used to search for similar full-length HERV-K proviruses in the human genome via nucleotide to nucleotide BLAST. Only human-specific and non-redundant insertions were selected. The top ten matches were chosen for alignment, which are K101, K102, K104, K107, K108, K109, K113, K115, and proviruses K11q22 and K12q14, named after their genomic location (Table 2). All of these proviruses are human specific, indicating integration into the germ-line within the last 6 million years (Barbulescu et al., 1999; Belshaw et al., 2005a; Hughes and Coffin, 2004; Turner et al., 2001). Moreover, several show insertional polymorphism in humans, with intact preintegration sites present in a fraction of the

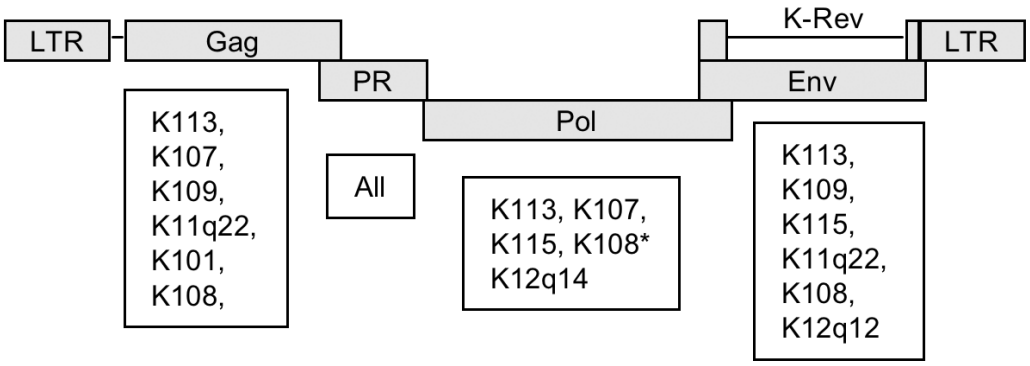


Figure 12 Undisrupted open reading frames in HERV-K proviruses

Diagram of HERV-K(HML-2) provirus. ORFs are depicted as boxes. Proviruses used in the design of HERV-K_{CON} that contain intact versions of Gag, protease, Pol, and Env are listed under each ORF. *K108 encodes a full-length Pol ORF, but a presumed essential YIDD motif is mutated.

human population (K108, K109, K113, K115, K11q22, K12q14), suggesting an even more recent replication for these proviruses. While all insertions except HERV-K113 encoded an obvious defect in at least one ORF, all proviruses also had an undisrupted ORF for at least one of the putative HERV-K proteins (Figure 12).

2.2 HERV-K_{CON} sequence

The nucleotide encoded by the majority of the ten proviruses was deduced for each of the 9,472 nucleotide positions to derive the consensus sequence using the AlignX program. This sequence was named HERV-K_{CON}. Thereafter, using a set of synthetic, approximately 60 base oligonucleotides spanning the entire HERV-K_{CON} sequence and a PCR-based strategy to progressively link them together, a plasmid containing the entire HERV-K_{CON} proviral genome was constructed. As expected, the HERV-K_{CON} sequence

was positioned close to the root of a phylogenetic tree constructed using HERV-K_{CON} itself and each of the ten proviruses used to derive it (Figure 13).

HERV-K_{CON} is 9472 nucleotides long, with intact major protein open reading frames (Figure 14). The long terminal repeats are 968 bps, with a consensus TATA box and poly-A signal. All identifiable motifs important for retroviral protein function can be

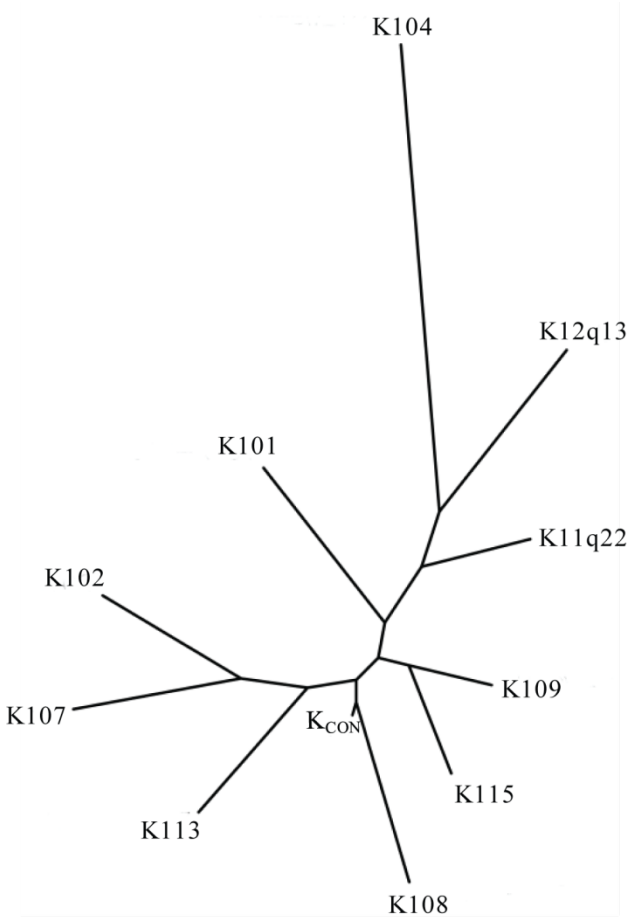


Figure 13 Phylogenetic analysis of HERV-K proviruses and HERV-K_{CON}

HERV-K_{CON} and the ten proviruses used to deduce it were phylogenetically analyzed. The tree was constructed using Kimura 2-parameter algorithm in the Treemaker program after gap-stripping the sequence alignment (http://www.hiv.lanl.gov/content/hivdb/CONTAM/Treemker/TreeMaker.html).

found on HERV-K_{CON} sequence. The MA protein contains the consensus myristoylation signal at the start (MGXXXS/T). A PTAP motif, often found in late assembly domains and important for the final scission of the particle membrane from the host membrane, is found in Gag as well, although its function has not been tested. The CA protein contains the major homology region (QXXXEXXXXAromaticXXR) that is present in all retroviruses, and the NC protein contains two CCHC-type zinc binding motifs, presumably for binding the viral RNA for packaging. The dUTPase is the sole protein with possible functional defect in HERV-K_{CON}: of the five highly conserved motifs, two motifs contain an amino acid change each (motif 3 GVVDSDYKG to SVVDSDYKG and motif 5 KRIGGFGSTD to KRIGGLGSTD) (Harris et al., 1997b). Of the sixteen human-specific proviruses whose sequence is available, three proviruses each encode for a G in motif 3 (19q12, K50B, and K60) and a F in motif 5 (K50B, K104, and 3q21). Thirteen proviruses encode for the potentially inactivating amino acid, suggesting that the correct HERV-K consensus sequence was derived. Whether Con dUTPase retains its enzymatic activity and whether its activity is still necessary for replication is unknown. Protease, RT, and IN each retain the correct, highly conserved DTG, YIDD, and DD(35)E motifs, respectively. Incidentally, HERV-K_{CON} Rec is identical to K108 Rec in amino acid sequence, and contains the arginine rich motif at N terminus (nuclear localization signal) and a leucine rich stretch at the C terminal (nuclear export signal).

Figure 14 Sequence of HERV-K_{CON} (pages 78-88)

The nucleic sequence of HERV-K_{CON} and the amino acid sequence of the open reading frames are shown. The start and end of both LTRs are marked above the nucleic sequence by arrows. The TATA box (TATAAAA) and polyA signal (AATAAA) are underlined. The presumed primer binding site is bolded. Gag, Pro, Pol, Env, and Rec ORFs are shown in red, green, blue, purple, and orange. The C terminus cleavage sites in Gag of p15 and CA (discussed later) are shown in pink. The myristoylation signal of MA, PTAP motif, MHR of CA, Cis-His boxes of NC, DTG motif of Prot, YIDD motif of RT, and DD(35)E motif of IN are highlighted. The DD(35)E motif of IN was estimated based on alignments with other retrovirus IN proteins.

→

1 TGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAGACATAGGAGACTCCATTTTGTTAT 80

81 GTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCAACCCCGTGCTCTCTGAAACGTGTG 160

161 CTGTGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAACAGATGCTTGAAGGCAGCATGCTC 240

241 CTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGGAAGGCCGCAGGGACCTCTGCCTAG 320

321 GAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTCGTGGGAAGGGAAAGACCTGACCGT 400

401 CCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAAGGAATGCCTCTTGCAGTTGAGACA 480

481 AGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAAACCCGATTGTATGCTCCATCTACT 560

561 GAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTGCTTTGTAAAGCACTGAGATGTTTA 640

641 TGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAAAGACCTTTGTTCACGTGTTTGTCT 720

721 GCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAAACACCCACAGATGATCAATAAATA 800

801 CTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGGGTCCCCTTATTTCTTTCTCTATAC 880

881 TTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACACCCACAGGTGTGTAGGGGCAACCCA 960

←

961 CCCCTACATC**TGGTGCCCAACGTGGAGG**CTTTTCTCTAGGGTGAAGGTACGCTCGAGCGTGGTCATTGAGGACAAGTCGA 1040

Figure 14 continued

```
                                                                                                M  G  Q
1041    CGAGAGATCCCGAGTACGTCTACAGTCAGCCTTACGGTAAGCTTGTGCGCTCGGAAGAAGCTAGGGTGATAATGGGGCAA  1120

         T  K  S  K  I  K  S  K  Y  A  S  Y  L  S  F  I  K  I  L  L  K  R  G  G  V  K  V
1121    ACTAAAAGTAAAATTAAAAGTAAATATGCCTCTTATCTCAGCTTTATTAAAATTCTTTTAAAAAGAGGGGGAGTTAAAGT  1200

           S  T  K  N  L  I  K  L  F  Q  I  I  E  Q  F  C  P  W  F  P  E  Q  G  T  L  D
1201    ATCTACAAAAAATCTAATCAAGCTATTTCAAATAATAGAACAATTTTGCCCATGGTTTCCAGAACAAGGAACTTTAGATC  1280

         L  K  D  W  K  R  I  G  K  E  L  K  Q  A  G  R  K  G  N  I  I  P  L  T  V  W  N
1281    TAAAAGATTGGAAAAGAATTGGTAAGGAACTAAAACAAGCAGGTAGGAAGGGTAATATCATTCCACTTACAGTATGGAAT  1360

         D  W  A  I  I  K  A  A  L  E  P  F  Q  T  E  E  D  S  V  S  V  S  D  A  P  G  S
1361    GATTGGGCCATTATTAAAGCAGCTTTAGAACCATTTCAAACAGAAGAAGATAGCGTTTCAGTTTCTGATGCCCCTGGAAG  1440

           C  I  I  D  C  N  E  N  T  R  K  K  S  Q  K  E  T  E  G  L  H  C  E  Y  V  A
1441    CTGTATAATAGATTGTAATGAAAACACAAGGAAAAAATCCCAGAAAGAAACGGAAGGTTTACATTGCGAATATGTAGCAG  1520

         E  P  V  M  A  Q  S  T  Q  N  V  D  Y  N  Q  L  Q  E  V  I  Y  P  E  T  L  K  L
1521    AGCCGGTAATGGCTCAGTCAACGCAAAATGTTGACTATAATCAATTACAGGAGGTGATATATCCTGAAACGTTAAAATTA  1600

         E  G  K  G  P  E  L  V  G  P  S  E  S  K  P  R  G  T  S  P  L  P  A  G  Q  V  P
1601    GAAGGAAAAGGTCCAGAATTAGTGGGGCCATCAGAGTCTAAACCACGAGGCACAAGTCCTCTTCCAGCAGGTCAGGTGCC  1680

           V  T  L  Q  P  Q  K  Q  V  K  E  N  K  T  Q  P  P  V  A  Y  Q  Y  W  P  P  A
1681    CGTAACATTACAACCTCAAAAGCAGGTTAAAGAAAATAAGACCCAACCGCCAGTAGCCTATCAATACTGGCCTCCGGCTG  1760

         E  L  Q  Y  R  P  P  P  E  S  Q  Y  G  Y  P  G  M  P  P  A  P  Q  G  R  A  P  Y
1761    AACTTCAGTATCGGCCACCCCCAGAAAGTCAGTATGGATATCCAGGAATGCCCCCAGCACCACAGGGCAGGGCGCCATAC  1840

         P  Q  P  P  T  R  R  L  N  P  T  A  P  P  S  R  Q  G  S  E  L  H  E  I  I  D  K
1841    CCTCAGCCGCCCACTAGGAGACTTAATCCTACGGCACCACCTAGTAGACAGGGTAGTGAATTACATGAAATTATTGATAA  1920
```

Figure 14 continued

```
          S  R  K  E  G  D  T  E  A  W  Q  F  P  V  T  L  E  P  M  P  P  G  E  G  A  Q
1921  ATCAAGAAAGGAAGGAGATACTGAGGCATGGCAATTCCCAGTAACGTTAGAACCGATGCCACCTGGAGAAGGAGCCCAAG  2000

          E  G  E  P  P  T  V  E  A  R  Y  K  S  F  S  I  K  M  L  K  D  M  K  E  G  V  K
2001  AGGGAGAGCCTCCCACAGTTGAGGCCAGATACAAGTCTTTTTCGATAAAAATGCTAAAAGATATGAAAGAGGGAGTAAAA  2080

          Q  Y  G  P  N  S  P  Y  M  R  T  L  L  D  S  I  A  H  G  H  R  L  I  P  Y  D  W
2081  CAGTATGGACCCAACTCCCCTTATATGAGGACATTATTAGATTCCATTGCTCATGGACATAGACTCATTCCTTATGATTG  2160

          E  I  L  A  K  S  S  L  S  P  S  Q  F  L  Q  F  K  T  W  W  I  D  G  V  Q  E
2161  GGAGATTCTGGCAAAATCGTCTCTCTCACCCTCTCAATTTTTACAATTTAAGACTTGGTGGATTGATGGGGTACAAGAAC  2240

          Q  V  R  R  N  R  A  A  N  P  P  V  N  I  D  A  D  Q  L  L  G  I  G  Q  N  W  S
2241  AGGTCCGAAGAAATAGGGCTGCCAATCCTCCAGTTAACATAGATGCAGATCAACTATTAGGAATAGGTCAAAATTGGAGT  2320

          T  I  S  Q  Q  A  L  M  Q  N  E  A  I  E  Q  V  R  A  I  C  L  R  A  W  E  K  I
2321  ACTATTAGTCAACAAGCATTAATGCAAAATGAGGCCATTGAGCAAGTTAGAGCTATCTGCCTTAGAGCCTGGGAAAAAAT  2400

          Q  D  P  G  S  T  C  P  S  F  N  T  V  R  Q  G  S  K  E  P  Y  P  D  F  V  A
2401  CCAAGACCCAGGAAGTACCTGCCCCTCATTTAATACAGTAAGACAAGGTTCAAAAGAGCCCTATCCTGATTTTGTGGCAA  2480

          R  L  Q  D  V  A  Q  K  S  I  A  D  E  K  A  R  K  V  I  V  E  L  M  A  Y  E  N
2481  GGCTCCAAGATGTTGCTCAAAAGTCAATTGCCGATGAAAAAGCCCGTAAGGTCATAGTGGAGTTGATGGCATATGAAAAC  2560

          A  N  P  E  C  Q  S  A  I  K  P  L  K  G  K  V  P  A  G  S  D  V  I  S  E  Y  V
2561  GCCAATCCTGAGTGTCAATCAGCCATTAAGCCATTAAAAGGAAAGGTTCCTGCAGGATCAGATGTAATCTCAGAATATGT  2640

          K  A  C  D  G  I  G  G  A  M  H  K  A  M  L  M  A  Q  A  I  T  G  V  V  L  G
2641  AAAAGCCTGTGATGGAATCGGAGGAGCTATGCATAAAGCTATGCTTATGGCTCAAGCAATAACAGGAGTTGTTTTAGGAG  2720

          G  Q  V  R  T  F  G  G  K  C  Y  N  C  G  Q  I  G  H  L  K  K  N  C  P  V  L  N
2721  GACAAGTTAGAACATTTGGAGGAAAATGTTATAATTGTGGTCAAATTGGTCACTTAAAAAAGAATTGCCCAGTCTTAAAC  2800
```

Figure 14 continued

```
          K  Q  N  I  T  I  Q  A  T  T  T  G  R  E  P  P  D  L  C  P  R  C  K  K  G  K  H
2801  AAACAGAATATAACTATTCAAGCAACTACAACAGGTAGAGAGCCACCTGACTTATGTCCAAGATGTAAAAAAGGAAAACA  2880

            W  A  S  Q  C  R  S  K  F  D  K  N  G  Q  P  L  S  G  N  E  Q  R  G  Q  P  Q
                                                K  W  A  T  I  V  G  K  R  A  K  G  P  A  S  G
2881  TTGGGCTAGTCAATGTCGTTCTAAATTTGATAAAAATGGGCAACCATTGTCGGGAAACGAGCAAAGGGGCCAGCCTCAGG  2960

          A  P  Q  Q  T  G  A  F  P  I  Q  P  F  V  P  Q  G  F  Q  G  Q  Q  P  P  L  S  Q
             P  T  T  N  W  G  I  P  N  S  A  I  C  S  S  G  F  S  G  T  T  T  P  T  V  P
2961  CCCCACAACAAACTGGGGCATTCCCAATTCAGCCATTTGTTCCTCAGGGTTTTCAGGGACAACAACCCCCACTGTCCCAA  3040

          V  F  Q  G  I  S  Q  L  P  Q  Y  N  N  C  P  P  P  Q  A  A  V  Q  Q  *
          S  V  S  G  N  K  P  V  T  T  I  Q  Q  L  S  P  A  T  S  G  S  A  A  V  D  L  C
3041  GTGTTTCAGGGAATAAGCCAGTTACCACAATACAACAATTGTCCCCCGCCACAAGCGGCAGTGCAGCAGTAGATTTATGT  3120

          T  I  Q  A  V  S  L  L  P  G  E  P  P  Q  K  I  P  T  G  V  Y  G  P  L  P  E  G
3121  ACTATACAAGCAGTCTCTCTGCTTCCAGGGGAGCCCCCACAAAAAATCCCCACAGGGGTATATGGCCCCCTGCCTGAGGG  3200

            T  V  G  L  I  L  G  R  S  S  L  N  L  K  G  V  Q  I  H  T  S  V  V  D  S  D
3201  GACTGTAGGACTAATCTTGGGAAGATCAAGTCTAAATCTAAAAGGAGTTCAAATTCATACTAGTGTGGTTGATTCAGACT  3280

          Y  K  G  E  I  Q  L  V  I  S  S  S  I  P  W  S  A  S  P  G  D  R  I  A  Q  L  L
3281  ATAAAGGCGAAATTCAATTGGTTATTAGCTCTTCAATTCCTTGGAGTGCCAGTCCAGGAGACAGGATTGCTCAATTATTA  3360

          L  L  P  Y  I  K  G  G  N  S  E  I  K  R  I  G  G  L  G  S  T  D  P  T  G  K  A
3361  CTCCTGCCATATATTAAGGGTGGAAATAGTGAAATAAAAAGAATAGGAGGGCTTGGAAGCACTGATCCAACAGGAAAGGC  3440

            A  Y  W  A  S  Q  V  S  E  N  R  P  V  C  K  A  I  I  Q  G  K  Q  F  E  G  L
3441  TGCATATTGGGCAAGTCAGGTCTCAGAGAACAGACCTGTGTGTAAGGCCATTATTCAAGGAAAACAGTTTGAAGGGTTGG  3520

          V  D  T  G  A  D  V  S  I  I  A  L  N  Q  W  P  K  N  W  P  K  Q  K  A  V  T  G
3521  TAGACACTGGAGCAGATGTCTCTATCATTGCTTTAAATCAGTGGCCAAAAAATTGGCCTAAACAAAAGGCTGTTACAGGA  3600
```

Figure 14 continued

```
        L  V  G  I  G  T  A  S  E  V  Y  Q  S  T  E  I  L  H  C  L  G  P  D  N  Q  E  S
3601    CTTGTCGGCATAGGCACAGCCTCAGAAGTGTATCAAAGTACGGAGATTTTACATTGCTTAGGGCCAGATAATCAAGAAAG  3680

          T  V  Q  P  M  I  T  S  I  P  L  N  L  W  G  R  D  L  L  Q  Q  W  G  A  E  I
3681    TACTGTTCAGCCAATGATTACTTCAATTCCTCTTAATCTGTGGGGTCGAGATTTATTACAACAATGGGGTGCGGAAATCA  3760

        T  M  P  A  P  L  Y  S  P  T  S  Q  K  I  M  T  K  M  G  Y  I  P  G  K  G  L  G
3761    CCATGCCCGCTCCATTATATAGCCCCACGAGTCAAAAAATCATGACCAAGATGGGATATATACCAGGAAAGGGACTAGGG  3840
                                                N  K  S  R  K  R  R  N  R  V  S  F  L  G
        K  N  E  D  G  I  K  V  P  V  E  A  K  I  N  Q  E  R  E  G  I  G  Y  P  F  *
3841    AAAAATGAAGATGGCATTAAAGTTCCAGTTGAGGCTAAAATAAATCAAGAAAGAGAAGGAATAGGGTATCCTTTTTAGGG  3920

          A  A  T  V  E  P  P  K  P  I  P  L  T  W  K  T  E  K  P  V  W  V  N  Q  W  P  L
3921    GCGGCCACTGTAGAGCCTCCTAAACCCATACCATTAACTTGGAAAACAGAAAAACCGGTGTGGGTAAATCAGTGGCCGCT  4000

            P  K  Q  K  L  E  A  L  H  L  L  A  N  E  Q  L  E  K  G  H  I  E  P  S  F  S
4001    ACCAAAACAAAAACTGGAGGCTTTACATTTATTAGCAAATGAACAGTTAGAAAAGGGTCATATTGAGCCTTCGTTCTCAC  4080

        P  W  N  S  P  V  F  V  I  Q  K  K  S  G  K  W  R  M  L  T  D  L  R  A  V  N  A
4081    CTTGGAATTCTCCTGTGTTTGTAATTCAGAAGAAATCAGGCAAATGGCGTATGTTAACTGACTTAAGGGCTGTAAACGCC  4160

          V  I  Q  P  M  G  P  L  Q  P  G  L  P  S  P  A  M  I  P  K  D  W  P  L  I  I  I
4161    GTAATTCAACCCATGGGGCCTCTCCAACCCGGGTTGCCCTCTCCGGCCATGATCCCAAAAGATTGGCCTTTAATTATAAT  4240

            D  L  K  D  C  F  F  T  I  P  L  A  E  Q  D  C  E  K  F  A  F  T  I  P  A  I
4241    TGATCTAAAGGATTGCTTTTTTACCATCCCTCTGGCAGAGCAGGATTGCGAAAAATTTGCCTTTACTATACCAGCCATAA  4320

        N  N  K  E  P  A  T  R  F  Q  W  K  V  L  P  Q  G  M  L  N  S  P  T  I  C  Q  T
4321    ATAATAAAGAACCAGCCACCAGGTTTCAGTGGAAAGTGTTACCTCAGGGAATGCTTAATAGTCCAACTATTTGTCAGACT  4400

          F  V  G  R  A  L  Q  P  V  R  E  K  F  S  D  C  Y  I  I  H  Y  I  D  D  I  L  C
4401    TTTGTAGGTCGAGCTCTTCAACCAGTTAGAGAAAAGTTTTCAGACTGTTATATTATTCATTATATTGATGATATTTTATG  4480
```

Figure 14 continued

```
          A  A  E  T  K  D  K  L  I  D  C  Y  T  F  L  Q  A  E  V  A  N  A  G  L  A  I
4481  TGCTGCAGAAACGAAAGATAAATTAATTGACTGTTATACATTTCTGCAAGCAGAGGTTGCCAATGCTGGACTGGCAATAG  4560

          A  S  D  K  I  Q  T  S  T  P  F  H  Y  L  G  M  Q  I  E  N  R  K  I  K  P  Q  K
4561  CATCTGATAAGATCCAAACCTCTACTCCTTTTCATTATTTAGGGATGCAGATAGAAAATAGAAAAATTAAGCCACAAAAA  4640

          I  E  I  R  K  D  T  L  K  T  L  N  D  F  Q  K  L  L  G  D  I  N  W  I  R  P  T
4641  ATAGAAATAAGAAAAGACACATTAAAAACACTAAATGATTTTCAAAAATTACTAGGAGATATTAATTGGATTCGGCCAAC  4720

          L  G  I  P  T  Y  A  M  S  N  L  F  S  I  L  R  G  D  S  D  L  N  S  K  R  M
4721  TCTAGGCATTCCTACTTATGCCATGTCAAATTTGTTCTCTATCTTAAGAGGAGACTCAGACTTAAATAGTAAAAGAATGT  4800

          L  T  P  E  A  T  K  E  I  K  L  V  E  E  K  I  Q  S  A  Q  I  N  R  I  D  P  L
4801  TAACCCCAGAGGCAACAAAAGAAATTAAATTAGTGGAAGAAAAAATTCAGTCAGCGCAAATAAATAGAATAGATCCCTTA  4880

          A  P  L  Q  L  L  I  F  A  T  A  H  S  P  T  G  I  I  I  Q  N  T  D  L  V  E  W
4881  GCCCCACTCCAACTTTTGATTTTTGCCACTGCACATTCTCCAACAGGCATCATTATTCAAAATACTGATCTTGTGGAGTG  4960

          S  F  L  P  H  S  T  V  K  T  F  T  L  Y  L  D  Q  I  A  T  L  I  G  Q  T  R
4961  GTCATTCCTTCCTCACAGTACAGTTAAGACTTTTACATTGTACTTGGATCAAATAGCTACATTAATCGGTCAGACAAGAT  5040

          L  R  I  I  K  L  C  G  N  D  P  D  K  I  V  V  P  L  T  K  E  Q  V  R  Q  A  F
5041  TACGAATAATAAAATTATGTGGAAATGACCCAGACAAAATAGTTGTCCCTTTAACCAAGGAACAAGTTAGACAAGCCTTT  5120

          I  N  S  G  A  W  Q  I  G  L  A  N  F  V  G  I  I  D  N  H  Y  P  K  T  K  I  F
5121  ATCAATTCTGGTGCATGGCAGATTGGTCTTGCTAATTTTGTGGGAATTATTGATAATCATTACCCAAAAACAAAGATCTT  5200

          Q  F  L  K  L  T  T  W  I  L  P  K  I  T  R  R  E  P  L  E  N  A  L  T  V  F
5201  CCAGTTCTTAAAATTGACTACTTGGATTCTACCTAAAATTACCAGACGTGAACCTTTAGAAAATGCTCTAACAGTATTTA  5280

          T  D  G  S  S  N  G  K  A  A  Y  T  G  P  K  E  R  V  I  K  T  P  Y  Q  S  A  Q
5281  CTGATGGTTCCAGCAATGGAAAAGCAGCTTACACAGGGCCGAAAGAACGAGTAATCAAAACTCCATATCAATCGGCTCAA  5360
```

Figure 14 continued

```
          R  A  E  L  V  A  V  I  T  V  L  Q  D  F  D  Q  P  I  N  I  I  S  D  S  A  Y  V
5361  AGAGCAGAGTTGGTTGCAGTCATTACAGTGTTACAAGATTTTGACCAACCTATCAATATTATATCAGATTCTGCATATGT  5440

          V  Q  A  T  R  D  V  E  T  A  L  I  K  Y  S  M  D  D  Q  L  N  Q  L  F  N  L
5441  AGTACAGGCTACAAGGGATGTTGAGACAGCTCTAATTAAATATAGCATGGATGATCAGTTAAACCAGCTATTCAATTTAT  5520

          L  Q  Q  T  V  R  K  R  N  F  P  F  Y  I  T  H  I  R  A  H  T  N  L  P  G  P  L
5521  TACAACAAACTGTAAGAAAAAGAAATTTCCCATTTTATATTACTCATATTCGAGCACACACTAATTTACCAGGGCCTTTG  5600

          T  K  A  N  E  Q  A  D  L  L  V  S  S  A  L  I  K  A  Q  E  L  H  A  L  T  H  V
5601  ACTAAAGCAAATGAACAAGCTGACTTACTGGTATCATCTGCACTCATAAAAGCACAAGAACTTCATGCTTTGACTCATGT  5680

          N  A  A  G  L  K  N  K  F  D  V  T  W  K  Q  A  K  D  I  V  Q  H  C  T  Q  C
5681  AAATGCAGCAGGATTAAAAAACAAATTTGATGTCACATGGAAACAGGCAAAAGATATTGTACAACATTGCACCCAGTGTC  5760

          Q  V  L  H  L  P  T  Q  E  A  G  V  N  P  R  G  L  C  P  N  A  L  W  Q  M  D  V
5761  AAGTCTTACACCTGCCCACTCAAGAGGCAGGAGTTAATCCCAGAGGTCTGTGTCCTAATGCATTATGGCAAATGGATGTC  5840

          T  H  V  P  S  F  G  R  L  S  Y  V  H  V  T  V  D  T  Y  S  H  F  I  W  A  T  C
5841  ACGCATGTACCTTCATTTGGAAGATTATCATATGTTCATGTAACAGTTGATACTTATTCACATTTCATATGGGCAACTTG  5920

          Q  T  G  E  S  T  S  H  V  K  K  H  L  L  S  C  F  A  V  M  G  V  P  E  K  I
5921  CCAAACAGGAGAAAGTACTTCCCATGTTAAAAAACATTTATTGTCTTGTTTTGCTGTAATGGGAGTTCCAGAAAAAATCA  6000

          K  T  D  N  G  P  G  Y  C  S  K  A  F  Q  K  F  L  S  Q  W  K  I  S  H  T  T  G
6001  AAACTGACAATGGACCAGGATATTGTAGTAAAGCTTTCCAAAAATTCTTAAGTCAGTGGAAAATTTCACATACAACAGGA  6080

          I  P  Y  N  S  Q  G  Q  A  I  V  E  R  T  N  R  T  L  K  T  Q  L  V  K  Q  K  E
6081  ATTCCTTATAATTCCCAAGGACAGGCCATAGTTGAAAGAACTAATAGAACACTCAAAACTCAATTAGTTAAACAAAAAGA  6160

          G  G  D  S  K  E  C  T  T  P  Q  M  Q  L  N  L  A  L  Y  T  L  N  F  L  N  I
6161  AGGGGGAGACAGTAAGGAGTGTACCACTCCTCAGATGCAACTTAATCTAGCACTCTATACTTTAAATTTTTTAAACATTT  6240
```

Figure 14 continued

```
        Y  R  N  Q  T  T  T  S  A  E  Q  H  L  T  G  K  K  N  S  P  H  E  G  K  L  I  W
6241    ATAGAAATCAGACTACTACTTCTGCAGAACAACATCTTACTGGTAAAAAGAACAGCCCACATGAAGGAAAACTAATTTGG  6320

         W  K  D  N  K  N  K  T  W  E  I  G  K  V  I  T  W  G  R  G  F  A  C  V  S  P  G
6321    TGGAAAGATAATAAAAATAAGACATGGGAAATAGGGAAGGTGATAACGTGGGGGAGAGGTTTTGCTTGTGTTTCACCAGG  6400

                                                                M  N  P  S  E  M  Q  R  K  A
          E  N  Q  L  P  V  W  I  P  T  R  H  L  K  F  Y  N  E  P  I  G  D  A  K  K  S
                                                                M  N  P  S  E  M  Q  R  K  A
6401    AGAAAATCAGCTTCCTGTTTGGATACCCACTAGACATTTGAAGTTCTACAATGAACCCATCGGAGATGCAAAGAAAAGCA  6480
         P  P  R  R  R  R  H  R  N  R  A  P  L  T  H  K  M  N  K  M  V  T  S  E  E  Q  M
        T  S  A  E  T  E  T  P  Q  S  S  T  V  D  S  Q  D  E  Q  N  G  D  V  R  R  T  D
         P  P  R  R  R  R  H  R  N  R  A  P  L  T  H  K  M  N  K  M  V  T  S  E  E  Q  M
6481    CCTCCGCGGAGACGGAGACACCGCAATCGAGCACCGTTGACTCACAAGATGAACAAAATGGTGACGTCAGAAGAACAGAT  6560

          K  L  P  S  T  K  K  A  E  P  P  T  W  A  Q  L  K  K  L  T  Q  L  A  T  K  Y
         E  V  A  I  H  Q  E  G  R  A  A  D  L  G  T  T  K  E  A  D  A  V  S  Y  K  I  S
          K  L  P  S  T  K  K  A  E  P  P  T  W  A  Q  L  K  K  L  T  Q  L  A  T  K  Y
6561    GAAGTTGCCATCCACCAAGAAGGCAGAGCCGCCGACTTGGGCACAACTAAAGAAGCTGACGCAGTTAGCTACAAAATATC  6640

        L  E  N  T  K  V  T  Q  T  P  E  S  M  L  L  A  A  L  M  I  V  S  M  V
          R  E  H  K  G  D  T  N  P  R  E  Y  A  A  C  S  L  D  D  C  I  N  G  G  K  S
        L  E  N  T  K  V  T  Q  T  P  E  S  M  L  L  A  A  L  M  I  V  S  M  V  V  S  L
6641    TAGAGAACACAAAGGTGACACAAACCCCAGAGAGTATGCTGCTTGCAGCCTTGATGATTGTATCAATGGTGGTAAGTCTC  6720

        P  Y  A  C  R  S  S  C  S  *
         P  M  P  A  G  A  A  A  A  N  Y  T  Y  W  A  Y  V  P  F  P  P  L  I  R  A  V  T
6721    CCTATGCCTGCAGGAGCAGCTGCAGCTAACTATACCTACTGGGCCTATGTGCCTTTCCCGCCCTTAATTCGGGCAGTCAC  6800

          W  M  D  N  P  I  E  V  Y  V  N  D  S  V  W  V  P  G  P  I  D  D  R  C  P  A
6801    ATGGATGGATAATCCTATAGAAGTATATGTTAATGATAGTGTATGGGTACCTGGCCCCATAGATGATCGCTGCCCTGCCA  6880
```

Figure 14 continued

```
       K  P  E  E  E  G  M  M  I  N  I  S  I  G  Y  R  Y  P  P  I  C  L  G  R  A  P  G
6881   AACCTGAGGAAGAAGGGATGATGATAAATATTTCCATTGGGTATCGTTATCCTCCTATTTGCCTAGGGAGAGCACCAGGA  6960

        C  L  M  P  A  V  Q  N  W  L  V  E  V  P  T  V  S  P  I  S  R  F  T  Y  H  M  V
6961   TGTTTAATGCCTGCAGTCCAAAATTGGTTGGTAGAAGTACCTACTGTCAGTCCCATCAGTAGATTCACTTATCACATGGT  7040

         S  G  M  S  L  R  P  R  V  N  Y  L  Q  D  F  S  Y  Q  R  S  L  K  F  R  P  K
7041   AAGCGGGATGTCACTCAGGCCACGGGTAAATTATTTACAAGACTTTTCTTATCAAAGATCATTAAAATTTAGACCTAAAG  7120

       G  K  P  C  P  K  E  I  P  K  E  S  K  N  T  E  V  L  V  W  E  E  C  V  A  N  S
7121   GGAAACCTTGCCCCAAGGAAATTCCCAAAGAATCAAAAAATACAGAAGTTTTAGTTTGGGAAGAATGTGTGGCCAATAGT  7200

        A  V  I  L  Q  N  N  E  F  G  T  I  I  D  W  A  P  R  G  Q  F  Y  H  N  C  S  G
7201   GCGGTGATATTACAAAACAATGAATTTGGAACTATTATAGATTGGGCACCTCGAGGTCAATTCTACCACAATTGCTCAGG  7280

         Q  T  Q  S  C  P  S  A  Q  V  S  P  A  V  D  S  D  L  T  E  S  L  D  K  H  K
7281   ACAAACTCAGTCGTGTCCAAGTGCACAAGTGAGTCCAGCTGTTGATAGCGACTTAACAGAAAGTTTAGACAAACATAAGC  7360

       H  K  K  L  Q  S  F  Y  P  W  E  W  G  E  K  G  I  S  T  P  R  P  K  I  V  S  P
7361   ATAAAAAATTGCAGTCTTTCTACCCTTGGGAATGGGGAGAAAAAGGAATCTCTACCCCAAGACCAAAAATAGTAAGTCCT  7440

        V  S  G  P  E  H  P  E  L  W  R  L  T  V  A  S  H  H  I  R  I  W  S  G  N  Q  T
7441   GTTTCTGGTCCTGAACATCCAGAATTATGGAGGCTTACTGTGGCCTCACACCACATTAGAATTTGGTCTGGAAATCAAAC  7520

         L  E  T  R  D  R  K  P  F  Y  T  V  D  L  N  S  S  L  T  V  P  L  Q  S  C  V
7521   TTTAGAAACAAGAGATCGTAAGCCATTTTATACTGTCGACCTAAATTCCAGTCTAACAGTTCCTTTACAAAGTTGCGTAA  7600

       K  P  P  Y  M  L  V  V  G  N  I  V  I  K  P  D  S  Q  T  I  T  C  E  N  C  R  L
7601   AGCCCCCTTATATGCTAGTTGTAGGAAATATAGTTATTAAACCAGACTCCCAGACTATAACCTGTGAAAATTGTAGATTG  7680

        L  T  C  I  D  S  T  F  N  W  Q  H  R  I  L  L  V  R  A  R  E  G  V  W  I  P  V
7681   CTTACTTGCATTGATTCAACTTTTAATTGGCAACACCGTATTCTGCTGGTGAGAGCAAGAGAGGGCGTGTGGATCCCTGT  7760
```

Figure 14 continued

```
          S  M  D  R  P  W  E  A  S  P  S  V  H  I  L  T  E  V  L  K  G  V  L  N  R  S
7761   GTCCATGGACCGACCGTGGGAGGCCTCACCATCCGTCCATATTTTGACTGAAGTATTAAAAGGTGTTTTAAATAGATCCA  7840

          K  R  F  I  F  T  L  I  A  V  I  M  G  L  I  A  V  T  A  T  A  A  V  A  G  V  A
7841   AAAGATTCATTTTTACTTTAATTGCAGTGATTATGGGATTAATTGCAGTCACAGCTACGGCTGCTGTAGCAGGAGTTGCA  7920

          L  H  S  S  V  Q  S  V  N  F  V  N  D  W  Q  K  N  S  T  R  L  W  N  S  Q  S  S
7921   TTGCACTCTTCTGTTCAGTCAGTAAACTTTGTTAATGATTGGCAAAAAAATTCTACAAGATTGTGGAATTCACAATCTAG  8000

          I  D  Q  K  L  A  N  Q  I  N  D  L  R  Q  T  V  I  W  M  G  D  R  L  M  S  L
8001   TATTGATCAAAAATTGGCAAATCAAATTAATGATCTTAGACAAACTGTCATTTGGATGGGAGACAGACTCATGAGCTTAG  8080

          E  H  R  F  Q  L  Q  C  D  W  N  T  S  D  F  C  I  T  P  Q  I  Y  N  E  S  E  H
8081   AACATCGTTTCCAGTTACAATGTGACTGGAATACGTCAGATTTTTGTATTACACCCCAAATTTATAATGAGTCTGAGCAT  8160

          H  W  D  M  V  R  R  H  L  Q  G  R  E  D  N  L  T  L  D  I  S  K  L  K  E  Q  I
8161   CACTGGGACATGGTTAGACGCCATCTACAGGGAAGAGAAGATAATCTCACTTTAGACATTTCCAAATTAAAAGAACAAAT  8240

          F  E  A  S  K  A  H  L  N  L  V  P  G  T  E  A  I  A  G  V  A  D  G  L  A  N
8241   TTTCGAAGCATCAAAAGCCCATTTAAATTTGGTGCCAGGAACTGAGGCAATTGCAGGAGTTGCTGATGGCCTCGCAAATC  8320

          L  N  P  V  T  W  V  K  T  I  G  S  T  T  I  I  N  L  I  L  I  L  V  C  L  F  C
8321   TTAACCCTGTCACTTGGGTTAAGACCATTGGAAGTACTACGATTATAAATCTCATATTAATCCTTGTGTGCCTGTTTTGT  8400

                   S  A  G  V  P  N  S  S  E  E  T  A  T  I  E  N  G  P  *
          L  L  L  V  C  R  C  T  Q  Q  L  R  R  D  S  D  H  R  E  R  A  M  M  T  M  A  V
8401   CTGTTGTTAGTCTGCAGGTGTACCCAACAGCTCCGAAGAGACAGCGACCATCGAGAACGGGCCATGATGACGATGGCGGT  8480
                                  →
          L  S  K  R  K  G  G  N  V  G  K  S  K  R  D  Q  I  V  T  V  S  V  *
8481   TTTGTCGAAAAGAAAAGGGGGAAATGTGGGGAAAAGCAAGAGAGATCAGATTGTTACTGTGTCTGTGTAGAAAGAAGTAG  8560

8561   ACATAGGAGACTCCATTTTGTTATGTACTAAGAAAAATTCTTCTGCCTTGAGATTCTGTTAATCTATGACCTTACCCCCA  8640
```

Figure 14 continued

8641 ACCCCGTGCTCTCTGAAACGTGTGCTGTGTCAACTCAGAGTTGAATGGATTAAGGGCGGTGCAGGATGTGCTTTGTTAAA 8720

8721 CAGATGCTTGAAGGCAGCATGCTCCTTAAGAGTCATCACCACTCCCTAATCTCAAGTACCCAGGGACACAAAAACTGCGG 8800

8801 AAGGCCGCAGGGACCTCTGCCTAGGAAAGCCAGGTATTGTCCAAGGTTTCTCCCCATGTGATAGTCTGAAATATGGCCTC 8880

8881 GTGGGAAGGGAAAGACCTGACCGTCCCCCAGCCCGACACCCGTAAAGGGTCTGTGCTGAGGAGGATTAGTAAAAGAGGAA 8960

8961 GGAATGCCTCTTGCAGTTGAGACAAGAGGAAGGCATCTGTCTCCTGCCTGTCCCTGGGCAATGGAATGTCTCGGTATAAA 9040

9041 ACCCGATTGTATGCTCCATCTACTGAGATAGGGAAAAACCGCCTTAGGGCTGGAGGTGGGACCTGCGGGCAGCAATACTG 9120

9121 CTTTGTAAAGCACTGAGATGTTTATGTGTATGCATATCTAAAAGCACAGCACTTAATCCTTTACATTGTCTATGATGCAA 9200

9201 AGACCTTTGTTCACGTGTTTGTCTGCTGACCCTCTCCCCACAATTGTCTTGTGACCCTGACACATCCCCCTCTTTGAGAA 9280

9281 ACACCCACAGATGATCAATAAATACTAAGGGAACTCAGAGGCTGGCGGGATCCTCCATATGCTGAACGCTGGTTCCCCGG 9360

9361 GTCCCCTTATTTCTTTCTCTATACTTTGTCTCTGTGTCTTTTTCTTTTCCAAATCTCTCGTCCCACCTTACGAGAAACAC 9440

9393 CCACAGGTGTGTAGGGGCAACCCACCCCTACA 9472

2.3 HERV-K_{CON} proteins

To determine whether HERV-K_{CON} proteins were capable of assembling into retrovirus-like particles, pCRVI based plasmids expressing the consensus Gag, Gag-PR and Gag-PR-Pol ORFs were constructed. Transfection of the plasmids resulted in the expression of a protein of approximately 70 to 80 kDa, detected by Western blotting using a commercially available antibody raised against HERV-K Gag (Covance) (Figure 15A). This approximated to the expected size (74 kDa) of intact HERV-K Gag. A concurrent analysis of proteins pelleted from culture supernatant through 20% sucrose revealed that Gag expression alone could efficiently generate extracellular particles (Figure 15A). In addition to the 74-kDa Gag precursor, a protein of approximately 40 kDa that reacted with the HERV-K Gag antibody was detected in lysates of cells transfected with Gag-PR and Gag-PR-Pol expression plasmids. While the precise identity of the 40-kDa protein is unknown, it likely represents a proteolytically processed form of Gag, which suggests that the HERV-K_{CON} protease is active.

Analysis of extracellular particles by SDS-PAGE and silver staining revealed that HERV-K_{CON} Gag expression alone produced a protein band around 74 kDa as expected for the full length Gag (Figure 15B and C). Particles generated by Gag-PR contained a dominant protein of 30 kDa, which based on previous studies likely represents HERV-K_{CON} capsid (CA) (Bieda et al., 2001; Boller et al., 1993a; Mueller-Lantzsch et al., 1993). A smaller protein or proteins of 20 kDa were also observed in Gag-PR particles, which presumably represent other mature Gag processed products such as MA or NC proteins (Figure 15B). Additionally, a protein of 40 kDa that likely corresponded to the 40-kDa band detected by Western blotting was also observed on

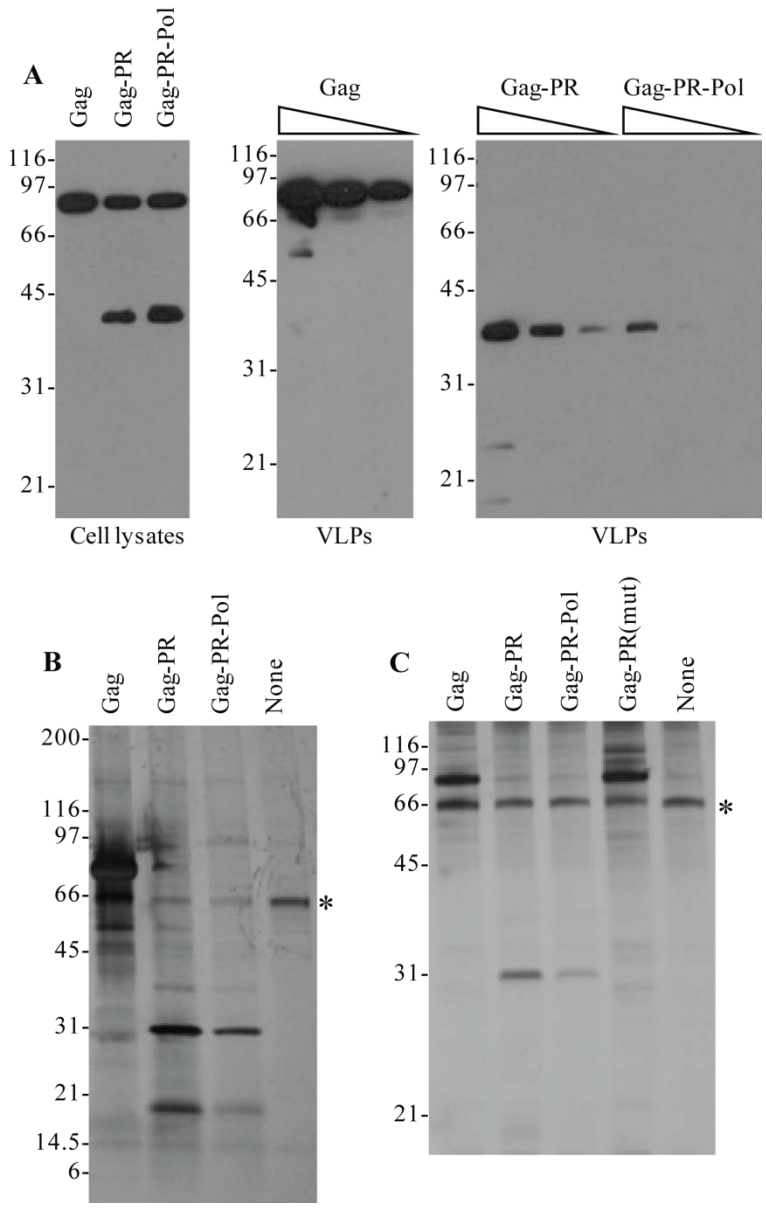


Figure 15 Expression of HERV-K proteins and release of virus-like particles

293T cells were transfected with HERV-K protein expression vectors, pCRVI/Gag, pCRVI/Gag-PR, pCRVI/Gag-PR-Pol, or pCRVI/Gag-PR(mut), and harvested two days after transfection. A. Western blot analysis of HERV-K protein expression in cell lysates (left) or virus-like particle expression in the cell supernatant (center and right) using a commercially available anti-HERV-K Gag antibody. The expected size of full length Gag polyprotein is 74kDa. Decreasing quantities of virion lysate (0.1, 0.05, and 0.025µl for Gag, center panel, or 0.4, 0.2, and 0.1µl for Gag-PR and Gag-PR-Pol, right panel) were loaded to estimate relative levels of VLP production. B. Silver stain analysis of a 4% to 20% gradient SDS-PAGE gel loaded with VLPs harvested from 293T cells transfected with HERV-K protein expression plasmids or empty plasmid control. An asterisk marks a nonspecific 66kDa protein band, most likely BSA. C. Silver stain analysis of VLPs harvested from 293T cells transfected with HERV-K protein expression plasmids including pCRVI/Gag-PR(mut) which has an active site mutation (DTG to AAA) in protease. An asterisk marks a nonspecific 66kDa band.

silver-stained gels. However, the 40-kDa protein was a minor species in Gag-PR particles, and it is therefore likely that this protein represents a partly processed intermediate. HERV-K Gag-PR-Pol expression also yielded particles containing the same apparently processed Gag proteins as those generated by Gag-PR but at slightly lower levels (Figure 15B). The 30-kDa putative CA protein on silver stained gels was abolished when three predicted active site residues (Asp-Thr-Gly) in the HERV-K_{CON} protease ORF were mutated to Ala-Ala-Ala (Figure 15C). Additionally, a higher molecular-weight protein, possibly representing the Gag-PR precursor, was observed in particles harvested from cells expressing the mutant Gag-PR protein (Figure 15C). These data are consistent with proper translation of the proteins, assembly and budding of particles, and cleavage of polyprotein by protease.

Although their low abundance relative to contaminating extraneous cellular proteins and the lack of available antibodies precluded unambiguous identification of Pol

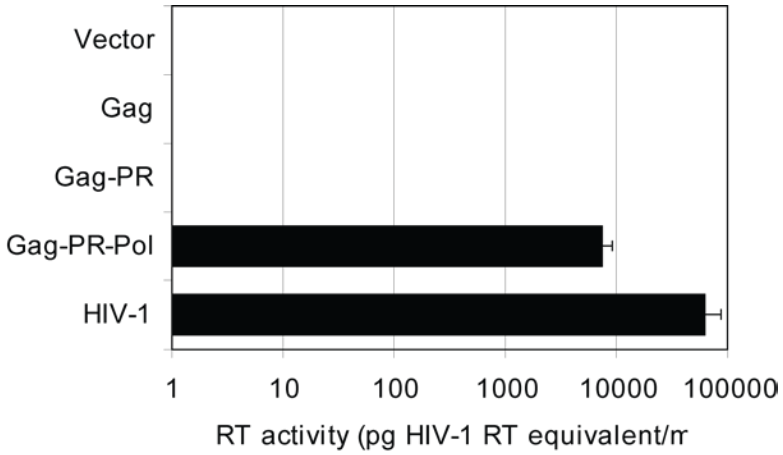


Figure 16 RT activity of HERV-K_{CON} VLPs

Reverse transcriptase activity was assayed in culture supernatants of 293T cells transfected with empty pCRV1 (vector) or vectors expressing HERV-K_{CON} Gag, Gag-PR, or Gag-PR-Pol proteins using commercially available RT assay (Cavidi). Enzymatic activity was determined relative to a recombinant HIV-1 reverse transcriptase standard and is representative of three experiments. Supernatants from 293T cells transfected with an HIV-1–based proviral plasmid are included for comparison.

proteins in SDS-PAGE analyses of HERV-K_{CON} VLPs, supernatants of 293T cell cultures transfected with the HERV-K_{CON} Gag-PR-Pol expression plasmid contained quite high levels of reverse transcriptase activity, as detected by an ELISA-based assay designed for the detection of HIV-1 reverse transcriptase (Figure 16). No reverse transcriptase activity was detected in control cultures transfected with HERV-K_{CON} Gag or Gag-PR expression plasmids.

To determine the location of protease cleavage sites in Gag, the N terminus of two bands (p30 and p15) detected in the silver stain analysis of HERV-K_{CON} VLPs in the presence of protease were identified by Edman sequencing. The result shows that the p30 band, predicted to be the CA protein, begins at the 283rd amino acid of Gag (PVTLE). The smaller band showed a cleavage site at the 149th amino acid of Gag (YNQLQ). This data implies that the MA protein is 148 amino acids in length and estimated to weigh 16.7 kDa, and may correspond to the smallest band detected in silver stain of VLPs between 20 and 15 kDa (Figure 15B). The protein that resides between MA and CA (amino acids 149-282) is estimated to be 15 kDa, and includes a late domain PTAP motif. The location of p15 between MA and CA is to the same as that of p24 in MPMV, which also contains late domains (PPPY and PTAP motifs) that have been shown to be important for MPMV particle release (Gottwein et al., 2003), indicating a similar role for p15 in HERV-K. Based on the molecular weights of MA, p15, and CA, the molecular weight of NC is estimated from the total molecular weight of Gag to be around 12 kDa, but there is no direct evidence of this estimation.

The molecular weight of CA is estimated from previous studies to be around 30 kDa (Bieda et al., 2001; Boller et al., 1993b). Despite not knowing the exact cleavage site

between CA and NC, knowing the N-terminus of CA allows for a rough calculation of the position of the C-terminus based on the molecular weight of CA. Based on this estimation, CA was expressed and purified for the generation of anti-CA polyclonal antiserum in rabbits (Covance). CA (amino acids 289 to 532, the C terminus conservatively estimated) was cloned into bacterial expression vector as a glutathione S-transferase (GST) tagged protein, which was purified using glutathione-agarose beads. The GST tag was cleaved, and the purified recombinant CA protein used to generate the antiserum (Figure 17). The western blotting analysis using HERV-K_{CON} Gag transfected 293T cell lysate show both a full length 74 kDa band and a smaller 30 kDa band (Figure 18), confirming that the p30 protein is CA. The proteins recognized by this polyclonal antibody is different from the commercially available anti-HERV-K Gag antibody, which recognizes the full length 74 kDa band and a band around 40 kDa that is a minor band in silver stain of VLPs (Figure 15A).

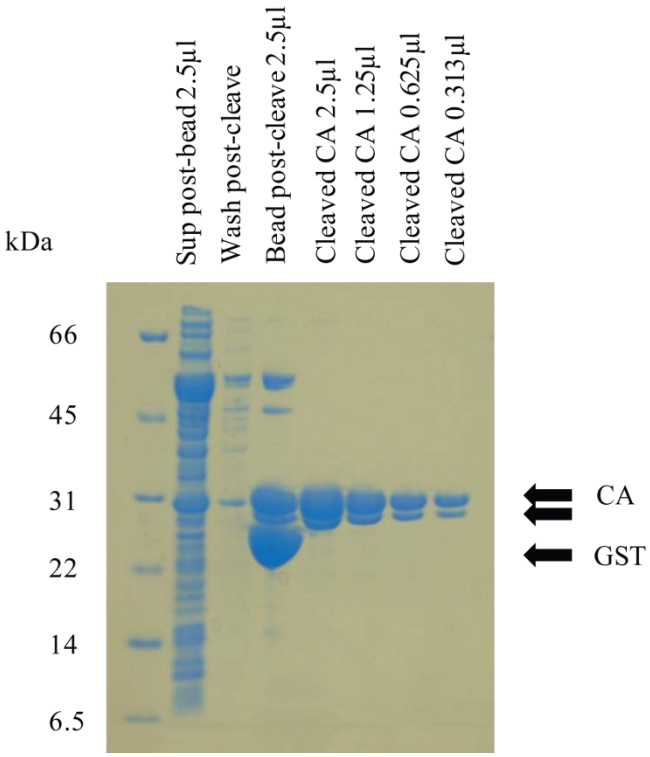


Figure 17 In vitro purification of HERV-K_{CON} CA

The deduced HERV-K_{CON} CA sequence was cloned into pGEX-6P-1 (GE Healthcare Life Sciences) to express a glutathione S-transferase (GST)-tagged CA protein that was purified using glutathione-agarose beads. The GST tag was eliminated by PreScission protease cleavage as per the manufacturer's instructions (GE Healthcare Life Sciences). The purified CA protein was separated in SDS-PAGE gel, and stained with Coomassie dye for visualization. Expected sizes of CA and GST are 30kDa and 26kDa, respectively, and are noted by the arrows. The minor band between CA and GST (middle arrow) is believed to be a minor degradation or cleavage product of CA. The concentration of CA was measured by Bradford assay to be 7.35 µg/µl.

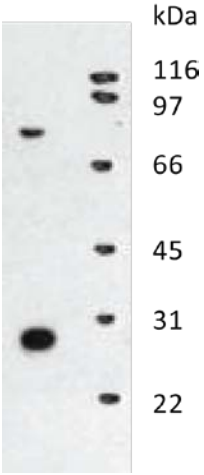


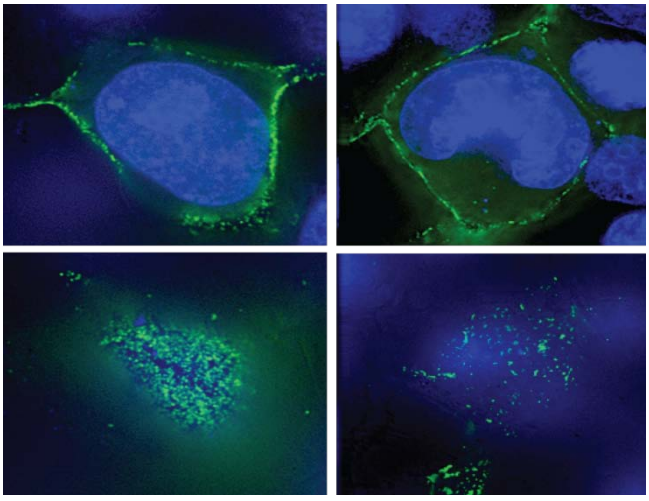
Figure 18 Efficacy of polyclonal HERV-K_{CON} CA antibody

Cell lysates separated by SDS-PAGE on PVDF membrane was probed with polyclonal HERV-K_{CON} CA antibody at 1:5000. The first lane contains the 293T cell lysate transfected with pCRVI/Con Gag-PR-Pol, and the second lane contains untransfected 293T cell lysate. The expected sizes of full length Gag and CA are 74kDa, and 30kDa. Gag-PR and Gag-PR-Pol are not visible.

The localization of Gag for virus particle assembly and budding was also examined. Co-expression of HERV-K_{CON} Gag and Gag–green fluorescent protein (GFP) fusion proteins in 293T cells revealed that HERV-K_{CON} Gag localized predominantly to the plasma membrane, where numerous fluorescent puncta were observed (Figure 19). Moreover, electron microscopic examination of 293T cells expressing HERV-K_{CON} Gag-PR revealed the presence of cell-associated retrovirus-like particles and structures that appeared to represent assembly intermediates (Figure 20). Most particles appeared to be between 100 to 150 nm in diameter, apparently spherical immature virions, with a minority assembled as aberrant particles that appeared as two or more connected, partly assembled, virions. While no unambiguously mature virions associated with the surface of Gag-PR

Figure 19 Plasma membrane localization of HERV-K_{CON} Gag

Two representative 293T cells transfected with HERV-K_{CON} Gag and Gag-GFP expression plasmids. Cells were fixed 18 hours post-transfection, and nuclei were stained with DAPI (blue) prior to visualization by deconvolution microscopy. Top, images acquired at the mid-section of the cell to show localization of Gag-GFP proteins; bottom, focused on the bottom of the cell to show accumulated VLPs at the cell-coverslip interface.



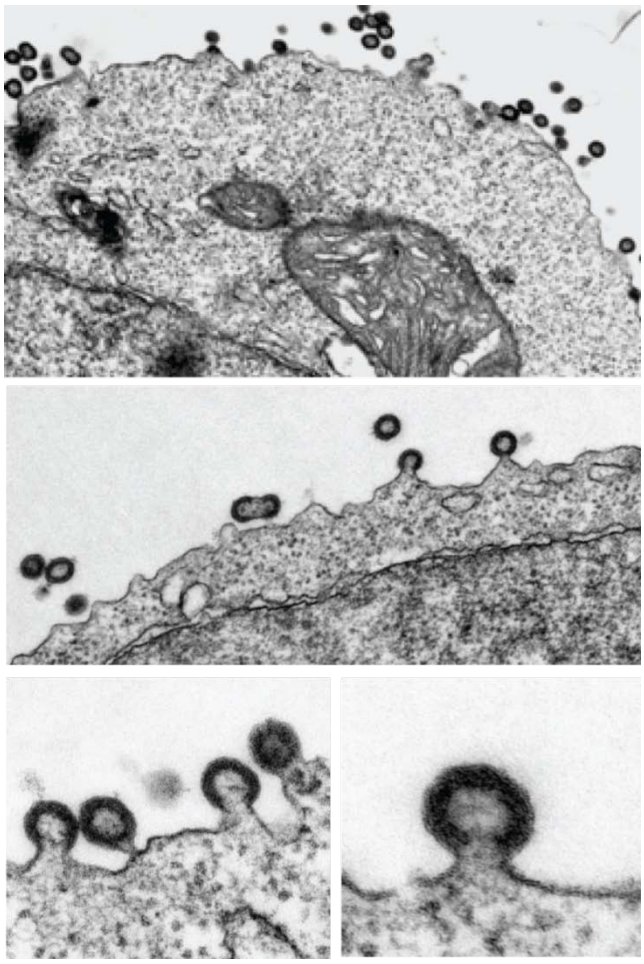


Figure 20 Electron microscopy of HERV-K_{CON} VLPs

Gallery of images of 293T cells transfected with a Gag-PR expressing plasmid. Black scale bars in the upper and middle panels represent 500 nm, while scale bars in the lower two panels represent 100 nm.

expressing cells were observed, it is possible that full maturation, which was clearly indicated by the biochemical analysis of extracellular VLPs (Figure 15 and 18), occurred only after the completion of particle release from cells. Completely or incompletely assembled particles appeared exclusively at the plasma membrane with a morphology resembling partly assembled alpharetroviruses or gammaretroviruses. Even though betaretroviruses represent HERV-Ks closest exogenous retrovirus relatives, no cytoplasmic, nonenveloped particles, typically observed in betaretroviruses, were found.

2.4 Establishment of single-cycle infection system

After confirming that the viral proteins can assemble into VLPs and that protease and RT are functional, a single-cycle infection system was established using these components along with a viral genome that should be packaged into VLPs, reverse transcribed, and integrated to complete the cycle. Frequently, a marker is inserted into the genomic sequence to allow for detection of a successful infection. Therefore various versions of the HERV-K_{CON} genome were constructed with CMVP-GFP or CMVP-drug resistance genes inserted into the Env ORF via KpnI restriction site (Figure 21). This strategy disrupts the Env ORF, but infectious pseudotyped particles could, in principle, be generated by the co-expression of vesicular stomatitis virus glycoprotein (VSVG). VSVG is known to have a wide tropism, and thus helpful in single-cycle infection systems as any cell can be used as target cells to study other steps of virus infection. The HERV-K_{CON} which contains the CMVP in the 5' LTR and the CMVP-marker in the Env ORF is named CMVP-HERV-K_{CON}-CMVP marker (CHKCG for GFP and CHKCP for puromycin-resistance gene).

As the insertion of the marker lengthens the total packaged sequence (additional 1464 bps for CMVP-GFP), the modified genome may not be packaged into the virus particle as effectively. To address this issue, versions of the genome where the sequence has been shorted to reflect the actual length of HERV-K (9472 bps) were constructed. Deletions of the viral sequence around Gag, PR and Pol ORFs (Bgl II to Bgl II, a deletion of 3918 bps which includes most of Gag to the 3' half of Pol; Hpa I to Hpa I, a deletion of 2526 bps which includes the 3' half of Gag to 5' third of Pol; Age I to Dra III, a deletion of 1782 bps which includes the 5' two thirds of Pol) did not result in infection

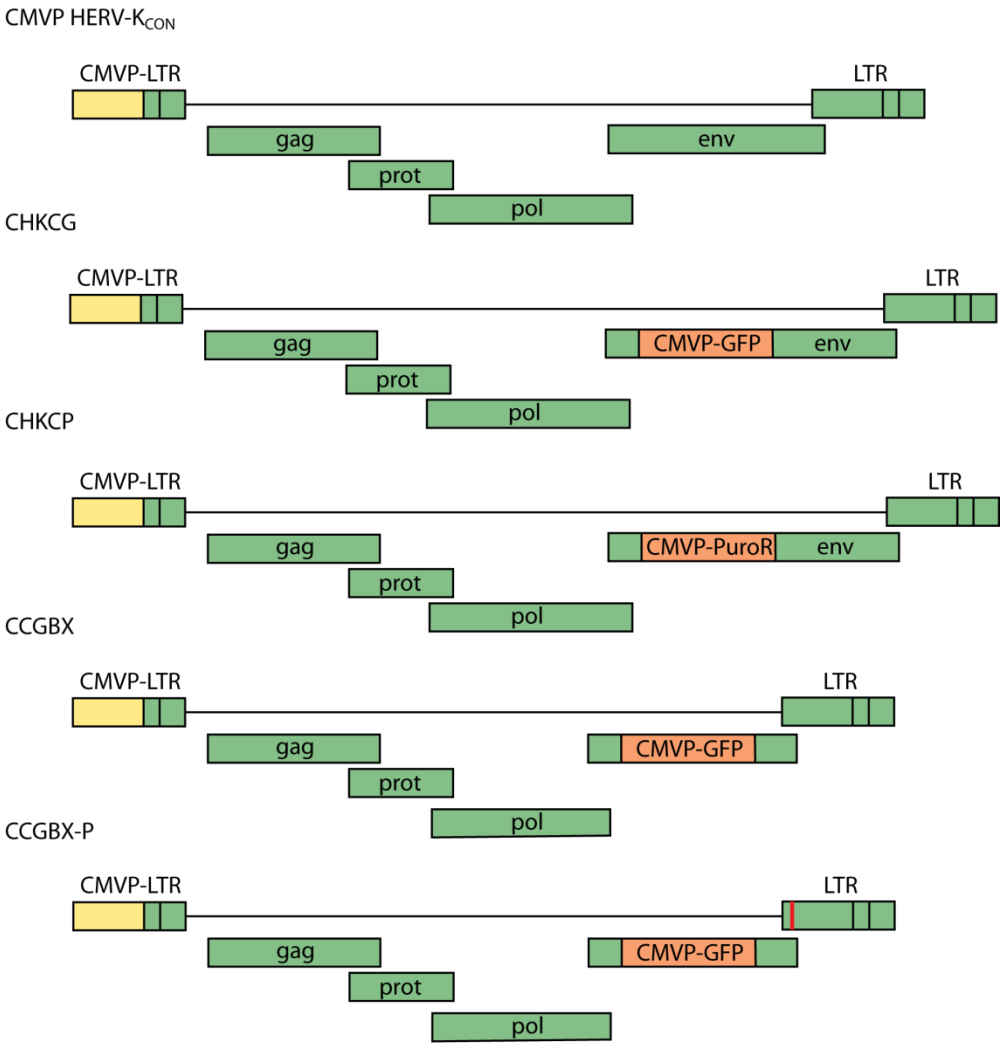


Figure 21 Diagram of HERV-K$_{CON}$ derived packageable genome constructs

The packageable genomes derived from HERV-K$_{CON}$ with the various features are depicted. CMVP-HERV-K$_{CON}$ contains a CMVP in the 5' LTR U3 region, and is shown for length comparisons. CHKCG and CHKCP contain CMVP-GFP and CMVP-puromycin resistance gene inserted into the SwaI restriction site of Env ORF. CCGBX is shortened in the Pol and Env ORFs (6643-7784) to reduce the length of the construct to 9287 bps for efficient packaging. CCGBX-P includes a 53 bp HIV-1 sequence in the 3' LTR, as depicted by the red bar.

(data not shown). These deletions all include the elimination of Age I to Hpa I fragment of 827 bps located in the Pol ORF encoding for RT, which is near the center of the viral genome. In HIV-1, a sequence located around that position called the central polypurine tract (cPPT) is important for maximum efficiency of reverse transcription and is highly

conserved (Goff, 2007). MMTV and MPMV, the closely related betaretroviruses, are not known to contain a cPPT; none was found in HERV-K as well. Hence, why these deletions resulted in the lack of transfer of the marker is unknown.

The final shortened genome that gave the highest virus titer retains the CMVP in the 5' LTR, but the CMVP-marker is cloned into the Kpn I and Swa I restriction sites of Env ORF, eliminating 1414 bps of Env sequence. An additional truncation from the Xba I to KpnI restriction site in the Env ORF eliminates an additional 211 bps. This final construct was named CCGBX, reflecting the GFP as the marker (Figure 21). The length of CCGBX is 9287 bps, 185 bps shorter than HERV-K_{CON}. Other markers are named accordingly; for example, CCBBX encodes for the blasticidin resistance gene. For each experiment, the genome used is noted in the figure legend for clarity.

To determine whether particles containing the HERV-K_{CON} genome, Gag, PR, and Pol proteins were capable of infectious transfer of the HERV-K_{CON} genome to target cells, the plasmids were transfected into 293T cells for VLP production (Figure 22). As expected, transfection of the Env-defective CHKCG construct resulted in GFP expression in transfected 293T cells, but inoculation of target cells with 0.2-μm filtered supernatant harvested from these cells did not result in infectious transfer of the reporter gene. However, when VSVG was expressed in trans, GFP expression was observed in rare foci of target cells inoculated with filtered supernatant from CHKCG-transfected cells. Moreover, when Rec was expressed in trans with CHKCG and VSVG, infectious particle yield was in excess of 10^2 infectious units/ml. Similarly, when the HERV-K_{CON} Gag-PR-Pol expression plasmid was provided in trans for the combined expression of CHKCG, VSV-G, HERV-K Gag-Pol, and Rec, the highest infectious titers were

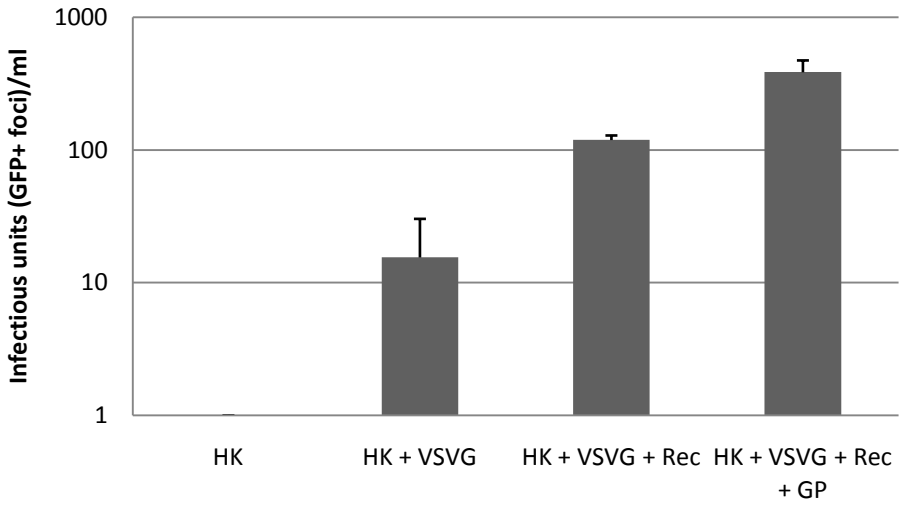


Figure 22 Generation of single cycle infectious HERV-K_{CON} VLPs

Infectious titers of HERV-K_{CON} VSV-G pseudotyped virions generated following transfection with the indicated plasmid mixtures using 293T target cells. GFP-positive foci were enumerated visually and expressed as infectious units per milliliter of virion-containing supernatant. CHKCG was used as the packageable genome.

detected (up to 10^3 IU/ml, Figure 22). Thus, this combination of plasmids (or an improved packageable HERV-K genome as noted) was used to generate infectious HERV-K_{CON} VSV-G pseudotyped particles in subsequent studies, which generated titers approaching 10^4 IU/ml. While this infectious titer is low compared to that generated by many exogenous retroviruses (e.g., murine leukemia virus [MLV] and HIV-1), the yield of infectious HERV-K particles was of the same order as or greater than that obtained with similarly constructed human T-cell lymphotropic virus-1 (HTLV-1) based vector systems ((Derse et al., 2001) and unpublished data).

HERV-K_{CON} particles with mutations in the reverse transcriptase protein were also tested as controls. VLPs were generated using CCBXS RTX (HERV-K_{CON} with a blasticidin resistance gene and RT mutation) and either WT or RTX Con Gag-PR-Pol. RTX plasmids were synthesized by mutating the highly conserved YIDD RT motif to

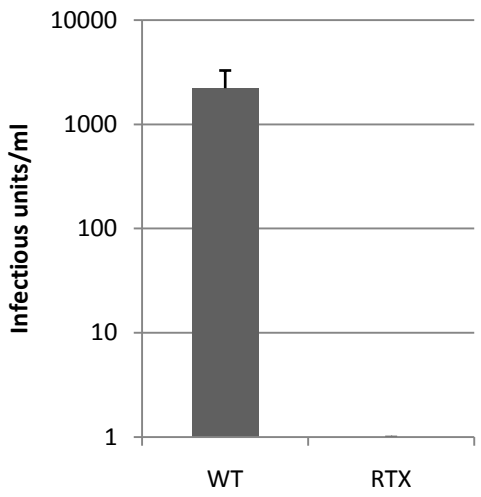


Figure 23 Lack of infection by RT mutant HERV-K_{CON} VLPs

CCBXS was used as the packageable genome for production of VLPs. 293T cells were used as target cells for infection. 12 hours post-infection, infected cells were selected via addition of blasticidin (0.5 µg/ml) in cell culture media for two weeks. The resistant colonies were quantified by eye, and infectious units were calculated based on eye count.

AIAA. In RTX VLPs, no blasticidin resistant colonies were observed after selection of infected cells, while numerous colonies were observed after infection with the WT VLPs (Figure 23).

HERV-K108, a provirus which exists as multiple alleles, contains a YIDD to CIDD mutation in its RT, and therefore assumed to encode an inactive protein (Mayer et al., 1999b; Reus et al., 2001). However, when the CIDD mutation was tested in the context of HERV-K_{CON} for activity in an exogenous RT assay, CIDD RT was in fact consistently slightly more active than YIDD RT (Figure 24). As K108 RT differs from HERV-K_{CON} in three additional amino acids, the activity of CIDD HERV-K RT does not confirm that K108 RT is active.

Additionally, 293T cells were inoculated with VSVG pseudotyped HERV-K_{CON} particles containing the CHKCG genome in the presence of azidothymidine (AZT), a

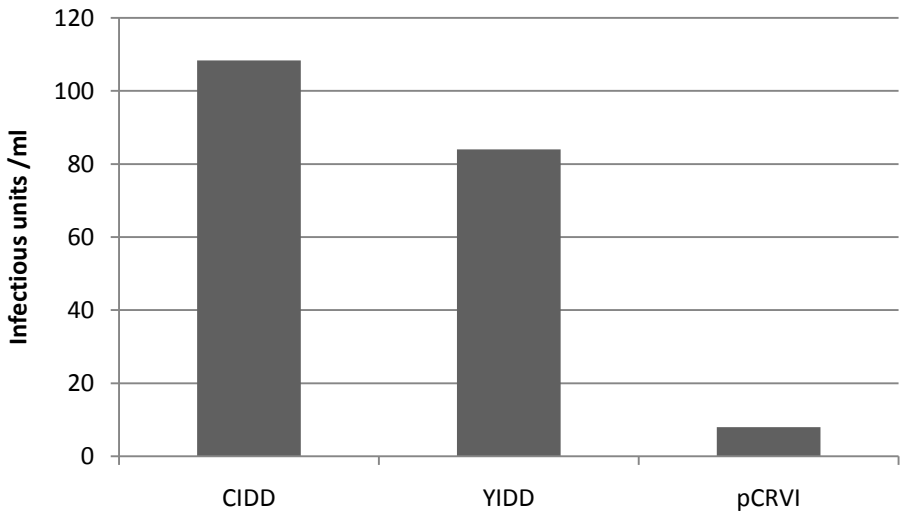


Figure 24 Infection of HERV-K_{CON} VLPs with RT mutation

CHKCG was used as the packageable genome for VLP production with either HERV-K_{CON} Gag-Prot-Pol with WT RT (YIDD motif), or mutant RT (CIDD motif). Two days post transfection, supernatant was used to infect fresh 293T cells. GFP^+ cells were quantified by FACS two days post infection.

reverse transcriptase inhibitor. AZT is a thymidine analog chain terminator and is known to inhibit reverse transcriptases from a wide variety of retroviruses (Rosenblum et al., 2001). As can be seen in Figure 25, application of AZT to target cells inhibited HERV-K–mediated reporter gene transduction by approximately 30-fold, suggesting that the reporter gene transfer by HERV-K_{CON} was dependent on reverse transcription.

In some cases, low levels of reporter gene expression mediated by retroviral gene transfer can be mediated by reverse-transcribed but nonintegrated retroviral DNA, which can exist as linear or circular forms in target cells (Saenz et al., 2004; Wu and Marsh, 2001; Yanez-Munoz et al., 2006). However, these retroviral DNA forms are diluted during cell division and eventually lost. Stable retrovirus-mediated gene transfer that is transferred to both daughter cells requires that retroviral DNA be integrated into the target cell genome. While the formation of clear multicellular foci of GFP positive cells

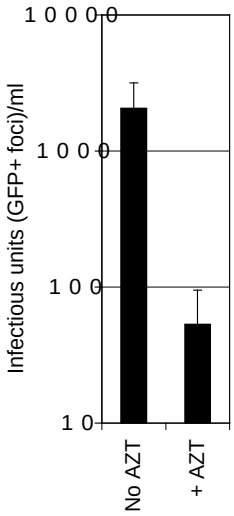


Figure 25 AZT sensitivity of HERV-K_{CON} infection

Infectious titers of CHKCG containing VSVG pseudotyped HERV-K_{CON} using 293T target cells in the presence or absence of 50 µM AZT was accessed.

suggested the reporter gene was maintained in daughter cells, integration events are most effectively assayed by daughter cell colony formation under antibiotic selection using retroviral genomes that carry resistance markers. Therefore, VSVG pseudotyped HERV-K_{CON} particles carrying CHKCP, packageable genome with a puromycin resistance gene, were used to infect new target cells (Figure 26). Puromycin-resistant colonies formed following exposure of 293T target cells to these virions and antibiotic selection for two weeks post infection, suggesting that true integration had occurred. Indeed, the infectious titers of puromycin resistance transducing particles were similar to that of GFP-transducing particles.

To further demonstrate that HERV-K_{CON} genomes were capable of integration, hamster CHO745 cells were infected with HERV-K_{CON} particles carrying the CHKCP

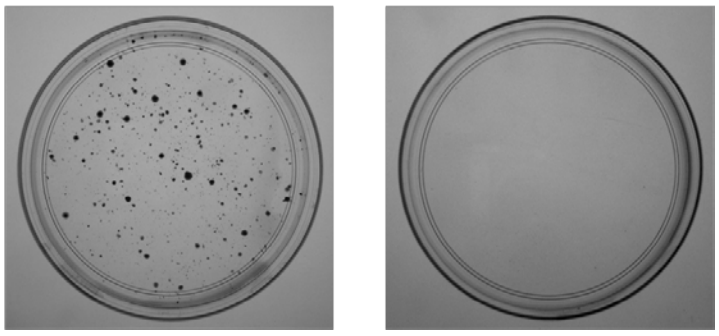


Figure 26 Transduction by HERV-K_{CON} Gag-PR-Pol and genomes

Puromycin-resistant colonies of 293T cells infected with either VSVG pseudotyped (left) or Env defective (right) virions carrying the CHKCP genome. Infected 293T cells were selected in 0.5 µg/ml puromycin for 2 weeks and then fixed and stained to reveal colonies of viable cells. Data are representative of at least three experiments.

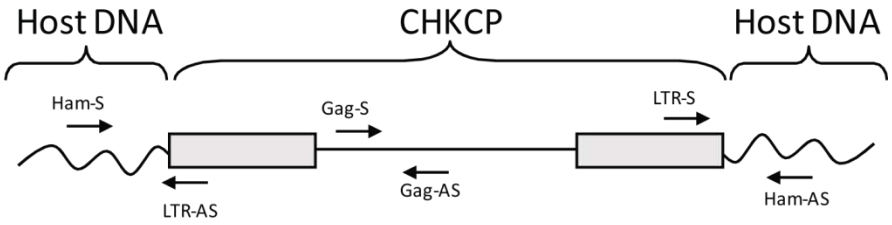
genome and four single cell clones were derived from the resulting puromycin-resistant cell population by limiting dilution. Cellular genomic DNA was extracted following expansion of the clones for two weeks in culture and analyzed for the presence of integrated HERV-K DNA using a PCR-based strategy (Figure 27). Hamster CHO745 cells were used for these experiments because they were found to be as sensitive as human cells to HERV-K_{CON} infection (see below), but unlike human cells, they lack endogenous HERV-K proviruses that would complicate detection and analysis of de novo HERV-K integration events. As can be seen in Figure 27B, PCR analysis using HERV-K *gag* specific PCR primers revealed that each of the CHKCP-transduced clones, but not parental CHO745 cells, carried HERV-K DNA. Next, sequences flanking the integrated proviruses were identified using a PCR-based strategy (GenomeWalker kit; Clontech, http://www.clontech.com) and in each case revealed the presence of a six nucleotide duplicated sequence immediately flanking the provirus (Figure 27C). For three CHKCP-transduced CHO745 cell clones, PCR primers were designed that targeted hamster DNA sequences flanking the integrated HERV-K_{CON} provirus (Figure 27A), and these were used to authenticate the presence of the intact preintegration site in uninfected hamster cells (e.g., Figure 27D). Moreover, PCRs using combinations of the hamster DNA-specific and HERV-K–specific PCR primers were used to authenticate the presence HERV-K provirus/hamster cellular DNA junctions in three of the CHKCG-transduced clones (e.g., Figure 27D). Overall, these experiments demonstrate that HERV-K genomes can be replicated via exogenous infection in a reverse transcriptase–dependent manner, resulting in stable and authentic integration into the target cell genome.

Next, it was determined whether VSVG pseudotyped HERV-K_{CON} particles could

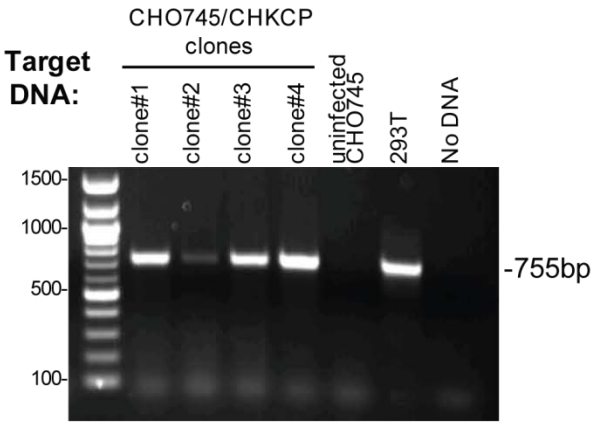
Figure 27 Identification and confirmation of HERV-K_{CON} integration sites

A. Experimental strategy for detection of HERV-K_{CON} proviruses in CHO745 cells using PCR primers targeted to HERV-K_{CON} Gag and LTR sequences, or flanking hamster DNA sequences. B. PCR amplification of HERV-K_{CON} gag DNA using Gag-S and Gag-AS primers in four expanded clones of puromycin-resistant CHO745 cells transduced with CHKCP-containing HERV-K_{CON} particles. C. Nucleotide sequences at the 59 and 39 ends of integrated CHKCP proviral DNA, revealing six nucleotide duplicated sequences at the CHKCP integration sites. D. Verification of the presence and absence of an integrated provirus and the empty preintegration site in CHKCP-transduced and naive CHO745 cells using combinations of HERV-K and hamster DNA targeted PCR primers (see [B] for primer design strategy). DNA templates and PCR primer pairs used are indicated above each lane, and the expected PCR product size is given below each lane. A representative analysis of a single CHKCG-carrying CHO745 cell clone is shown; similar results were obtained with two additional clones. Uninfected CHO745 cells and human 293T cells serve as controls.

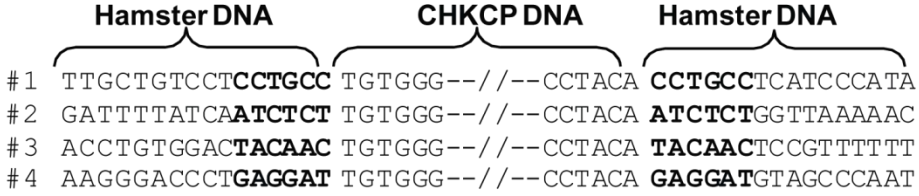
A


Host DNA
CHKCP
Host DNA
Ham-S
Gag-S
LTR-S
LTR-AS
Gag-AS
Ham-AS

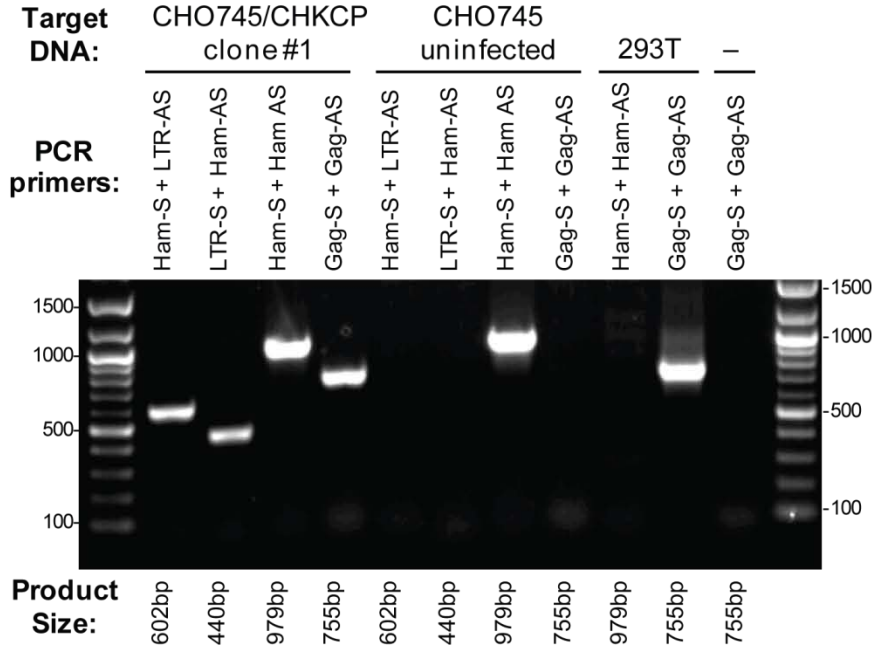
B


CHO745/CHKCP clones
Target DNA:
clone#1
clone#2
clone#3
clone#4
uninfected CHO745
293T
No DNA
1500
1000
500
100
-755bp

C


Hamster DNA
CHKCP DNA
Hamster DNA
#1 TTGCTGTCCTCCTGCC TGTGGG--//--CCTACA CCTGCCTCATCCCATA
#2 GATTTTATCAATCTCT TGTGGG--//--CCTACA ATCTCTGGTTAAAAAC
#3 ACCTGTGGACTACAAC TGTGGG--//--CCTACA TACAACTCCGTTTTTT
#4 AAGGGACCCTGAGGAT TGTGGG--//--CCTACA GAGGATGTAGCCCAAT

D


Target DNA:
CHO745/CHKCP clone #1
CHO745 uninfected
293T
–
PCR primers:
Ham-S + LTR-AS
LTR-S + Ham-AS
Ham-S + Ham AS
Gag-S + Gag-AS
Ham-S + LTR-AS
LTR-S + Ham-AS
Ham-S + Ham AS
Gag-S + Gag-AS
Ham-S + Ham-AS
Gag-S + Gag-AS
Gag-S + Gag-AS
1500
1000
500
100
Product Size:
602bp
440bp
979bp
755bp
602bp
440bp
979bp
755bp
979bp
755bp
755bp

transduce reporter genes into cells other than 293T and CHO745. As can be seen in Figure 28A, several target cells of human, squirrel monkey, feline, and rodent origin could be infected by HERV-K_{CON}. However, it was noticeable that murine NIH3T3 cells and squirrel monkey Pindak cells were somewhat less sensitive to HERV-K_{CON}, compared to the human and feline cells. The human cells were each quite similar in their sensitivity even though 293T cells display little or no TRIM5α-dependent resistance to retroviruses such as EIAV or N-tropic MLV, while TE671 and HT1080 exhibit strong TRIM5α-dependent resistance to N-tropic MLV and EIAV. This finding suggested that HERV-K_{CON} may not be sensitive to human TRIM5α.

Additionally, to test whether the HERV-K_{CON} envelope sequence was functional, it was inserted into the HIV-1–based expression vector pCRV1 and expressed along with HIV-1 Gag-PR-Pol proteins and the packageable GFP-expressing HIV-1 vector CSGW.

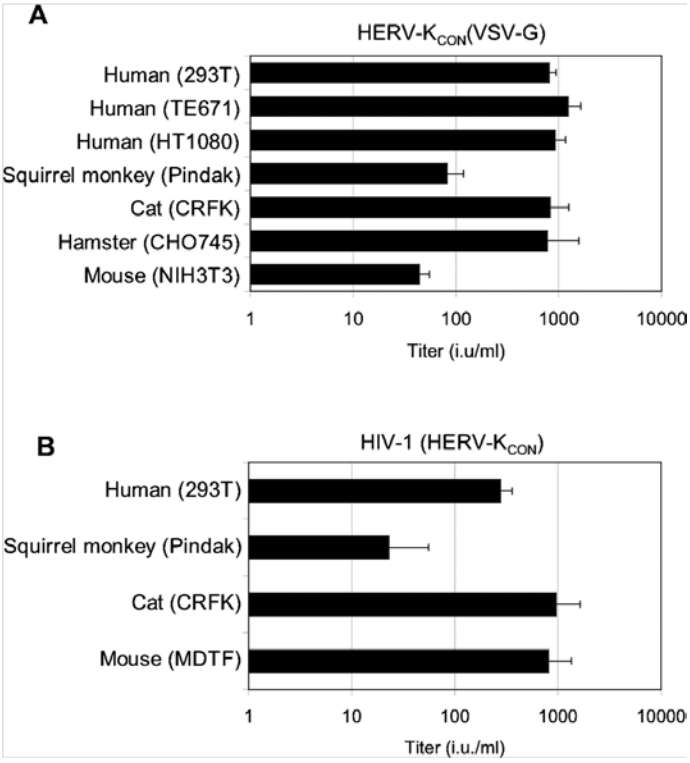


Figure 28 Tropism of HERV-K_{CON}

A. Human, squirrel monkey, feline, hamster, or murine cells were infected with VSVG pseudotyped HERV-K_{CON} particles. Two days postinfection, GFP^+ foci were quantified microscopically, and titers are expressed as number of infectious units (i.u.) per milliliter of virus containing supernatant applied. B. Human, squirrel monkey, feline, or murine cells were infected with HERV-K_{CON} Env pseudotyped HIV-1 particles as in A. Two days postinfection, GFP-positive foci were quantified. All data are representative of at least three experiments.

This transfection mixture should generate HIV-1 particles, putatively pseudotyped with the HERV-K_{CON} envelope protein. Notably, these particles were capable of infecting 293T cells, with titers of around 3 x 10^2 IU/ml (Figure 28B), while particles generated in the absence of HERV-K_{CON} Env were noninfectious. Inoculation of cells from a small panel of mammalian species revealed that several, including those of human, squirrel monkey, murine, and feline origin, could be infected with HERV-K_{CON} Env pseudotyped HIV-1 VLPs (Figure 28B).

While attempts were made to generate infectious particles that contained both HERV-K_{CON} cores and Env proteins, infection events using this combination were undetectable. Nevertheless, these experiments indicate that the HERV-K_{CON} genome contains all functional components required to complete an exogenous retroviral replication cycle.

As mentioned previously, HERV-K_{CON} is a betaretrovirus. Of the seven genera of Retroviridae, lentiviruses are unique in their ability to infect non-diving cells. Most other retroviruses depend on the cell cycle to dissolve the nuclear membrane for the PIC to access the host genomic DNA for integration. How HERV-K accesses the host genome was examined by treating HT1080 cells with aphidicolin, which blocks the cell cycle at early S phase, prior to infection (Figure 29). When the target cells were effectively blocked in cell cycle, HERV-K_{CON} was unable to infect HT1080. This was in contrast to HERV-K_{CON} infection of untreated HT1080s or HIV-1 infection of treated HT1080 cells. This data clearly demonstrates that HERV-K, like most nonlentiviral retroviruses, depends on the cell cycle for successful infection.

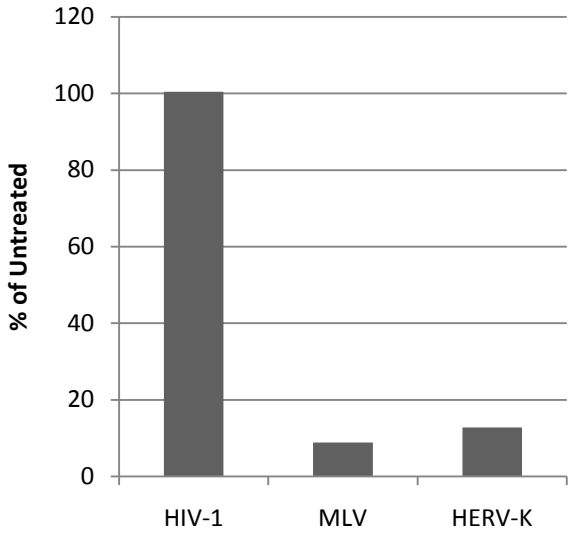


Figure 29 Cell cycle dependency of HERV-K_{CON}

HT1080 cells were treated with 2µg/ml of aphidicolin 24 hours, then infected with HIV-1, MLV, or HERV-K_{CON} VLPs. The cells were treated with aphidicolin at the same concentration during the entire experiment. For HIV-1 VLPs, 293T cells were transfected with CSGW and NL4-3 GPol plasmids. For MLV VLPs, cells were transfected with CNCG and MLV GPol plasmids. For HERV-K_{CON} VLPs, cells were transfected with CCGBX, pCRVI/Con Gag-Prot-Pol, and pCR3.1/K108 Rec. All VLPs are pseudotyped with VSVG. Successful cell cycle arrest and infection was checked via FACS.

Chapter 3. Restriction factors and HERV-K

3.1 Introduction to restriction factors

Restriction factors are host proteins that help defend the host cell against virus infections (Bieniasz, 2004b). Although also up-regulated by IFN, they are constitutively expressed, potentially in any cell, providing an immediate front-line defense against invading viruses (Bieniasz, 2004b). To date, three major groups of restriction factors have been described for retroviruses; all were discovered in relation to HIV-1, but some have been shown since to be active against other retroviruses and non-retroviruses as well (Esnault et al., 2005; Harris et al., 2003; Jouvenet et al., 2009b; Turelli et al., 2004b; Vartanian et al., 2008a; Yu et al., 2004b).

The first group of proteins, including Fv-1, TRIM5α and TRIM5-Cyp, target the incoming viral core, although the exact mechanism of restriction is unclear (Bieniasz, 2004b; Johnson and Sawyer, 2009a). The second group of restriction factors is the APOBEC3 cytidine deaminase family. APOBEC3G (A3G) was first described as an antiretroviral factor capable of inhibiting Vif deficient HIV-1 (Sheehy et al., 2002), but other APOBEC3 proteins have since been shown to have antiviral activity as well (Bishop et al., 2004a; Dang et al., 2008; Dang et al., 2006; OhAinle et al., 2006; Yu et al., 2004a; Zennou and Bieniasz, 2006). The major mechanism of restriction by A3G is likely by extensive mutation of Cs to Us in the minus strand of the ssDNA viral genome during reverse transcription (Bishop et al., 2004a; Harris et al., 2003; Mangeat et al., 2003; Miyagi et al., 2007). The most recently described retroviral restriction factor is tetherin, which tethers fully assembled and budded particles to the infected cell's plasma

membrane and prevents the virus from moving away and infecting a new cell (Neil et al., 2008b; Van Damme et al., 2008a).

TRIM5α and A3G proteins have been under positive selection in primates for at least 35 million years (Sawyer et al., 2004; Sawyer et al., 2005b), as has tetherin for at least 25 million years (McNatt et al., 2009b). Likely, these proteins have been functioning against many retroviruses and perhaps other viruses during that time (Johnson and Sawyer, 2009a). As an ancient retrovirus that infected primate and Old World monkey ancestors, HERV-K_{CON} is an ideal candidate to test the retroviral activity of these restriction factors in an effort to understand the long-standing interaction between the proteins and retroviruses.

3.2 Effects of TRIM proteins on HERV-K_{CON} infection

To test the sensitivity of HERV-K to retrovirus restriction factors that it might encounter in human cells and might be responsible for attenuation or extinction of replication therein, unmodified or human TRIM5α expressing hamster (CHO)-derived cell lines were challenged with VSVG pseudotyped HERV-K_{CON}. The human TRIM5α expressing cell line was greater than 100-fold resistant to N-tropic MLV relative to the control cell line or B-tropic MLV, confirming that the transduced human TRIM5α was functioning as expected in the cell line (Figure 30A). Unlike MLV, HERV-K_{CON} infected unmanipulated and human TRIM5α -expressing cells with nearly identical efficiency (Figure 30B). Additionally, CHO cells expressing rhesus macaque TRIM5α or the unusual owl monkey variant of TRIM5 (TRIM-Cyp) were also similarly sensitive to HERV-K_{CON} infection compared to unmanipulated control cells (Figure 30B). This was

despite the fact that CHO cells expressing rhesus monkey TRIM5α and owl monkey TRIMCyp were about 30-fold and 100-fold, respectively, resistant to HIV-1 infection compared to HIV-1 carrying an SIV_{MAC} CA (Figure 30A). Whether HERV-K(HML-2) has evolved to resist the effects of human TRIM5α and by chance is also resistant to rhesus macaque TRIM5α and owl monkey TRIM-Cyp or is completely unaffected by TRIM5 proteins is unknown.

Human TRIM5 exists as a gene cluster of four TRIM proteins on chromosome 11, TRIM6, TRIM34, TRIM5, and TRIM22, which likely resulted from tandem gene duplication (Sawyer et al., 2007). Like TRIM5, TRIM22 is up-regulated by IFN

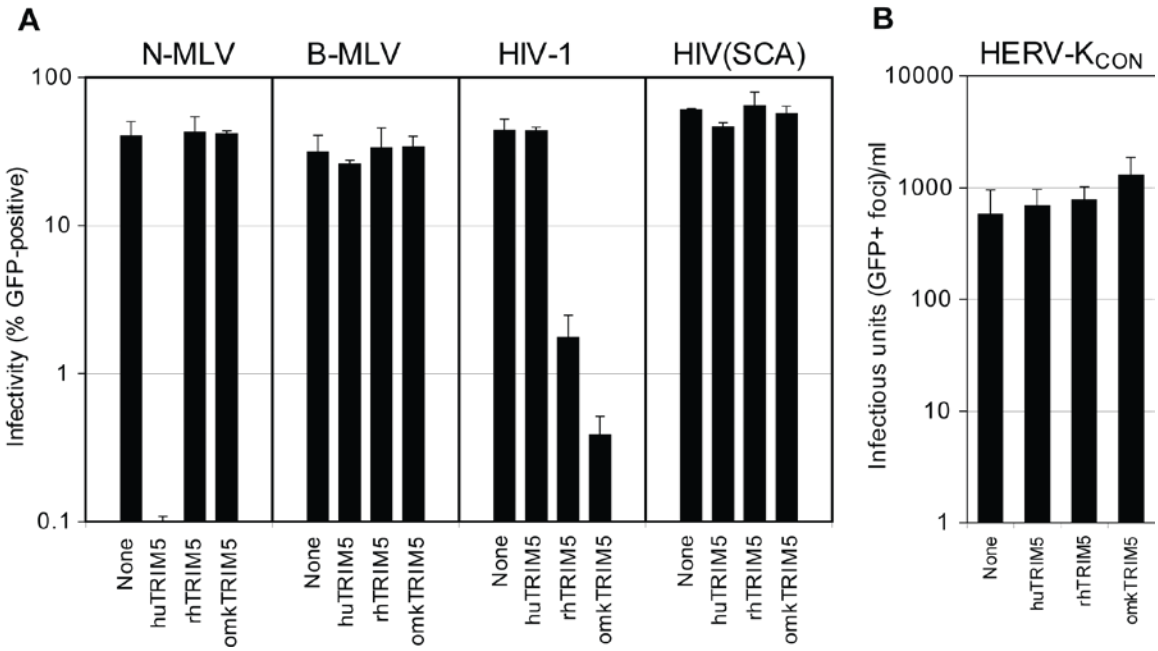


Figure 30 Effect of TRIM5 proteins on HERV-K_{CON} infection

A. Unmanipulated CHO cells or variants stably expressing human TRIM5α, rhesus monkey TRIM5α, or owl monkey TRIM-Cyp were infected with VSVG pseudotyped retroviral vectors that are sensitive to one or more of the TRIM5 proteins (N-MLV or HIV-1) or TRIM5-resistant controls (B-MLV or HIV-1 carrying SIV_{mac} CA HIV(SCA)), as indicated. Two days postinfection, the percentage of GFP^+ cells was determined using FACS. B. The same panel of CHO-derived TRIM5-expressing CHO cell lines were inoculated with VSVG pseudotyped HERV-K_{CON}. Two days postinfection, GFP^+ foci were quantified by FACS.

treatment and, unusually, has been under episodic positive selection in primates over the past 23 million years (Bouazzaoui et al., 2006; Gongora et al., 2000; Sawyer et al., 2007). TRIM22 also has been shown to reduce HIV-1 replication in primary human macrophages and cell lines such as HOS and HeLa cells, but the mechanism by which TRIM22 restricts the virus remains unclear (Barr et al., 2008; Bouazzaoui et al., 2006; Gongora et al., 2000).

To test the restrictive potential of TRIM22 on HERV-K_{CON}, CHO745 cells were stably transduced with TRIM22 proteins from human, gorilla, rhesus macaque, owl monkey, and squirrel monkey. The localization of these proteins, examined via confocal microscopy of GFP tagged versions, shows that human, gorilla, rhesus macaque, and squirrel monkey TRIM22 proteins are punctate and cytoplasmic (Figure 31A). Squirrel monkey TRIM22 also existed as small ring structures within the nucleus. Owl monkey TRIM22 is found as both punctate cytoplasmic form and as a large nuclear mass, resembling the nucleolus. Regardless, when these cells were infected with HERV-K_{CON}, no change in infection was seen (Figure 31B), suggesting that TRIM22 proteins tested do not restrict HERV-K between virus entry and integration.

3.3 Effect of tetherin on HERV-K_{CON} release

Next, the effect of tetherin on HERV-K infection was tested. Tetherin was discovered as the target of an HIV-1 accessory protein Vpu, the interaction which leads to increase HIV-1 particle release (Neil et al., 2008b; Van Damme et al., 2008a). Expression of tetherin seems to block the release of other virus particles as well (Jouvenet et al., 2009b). How Vpu antagonizes the effects of tetherin is unknown. Co-expression of human

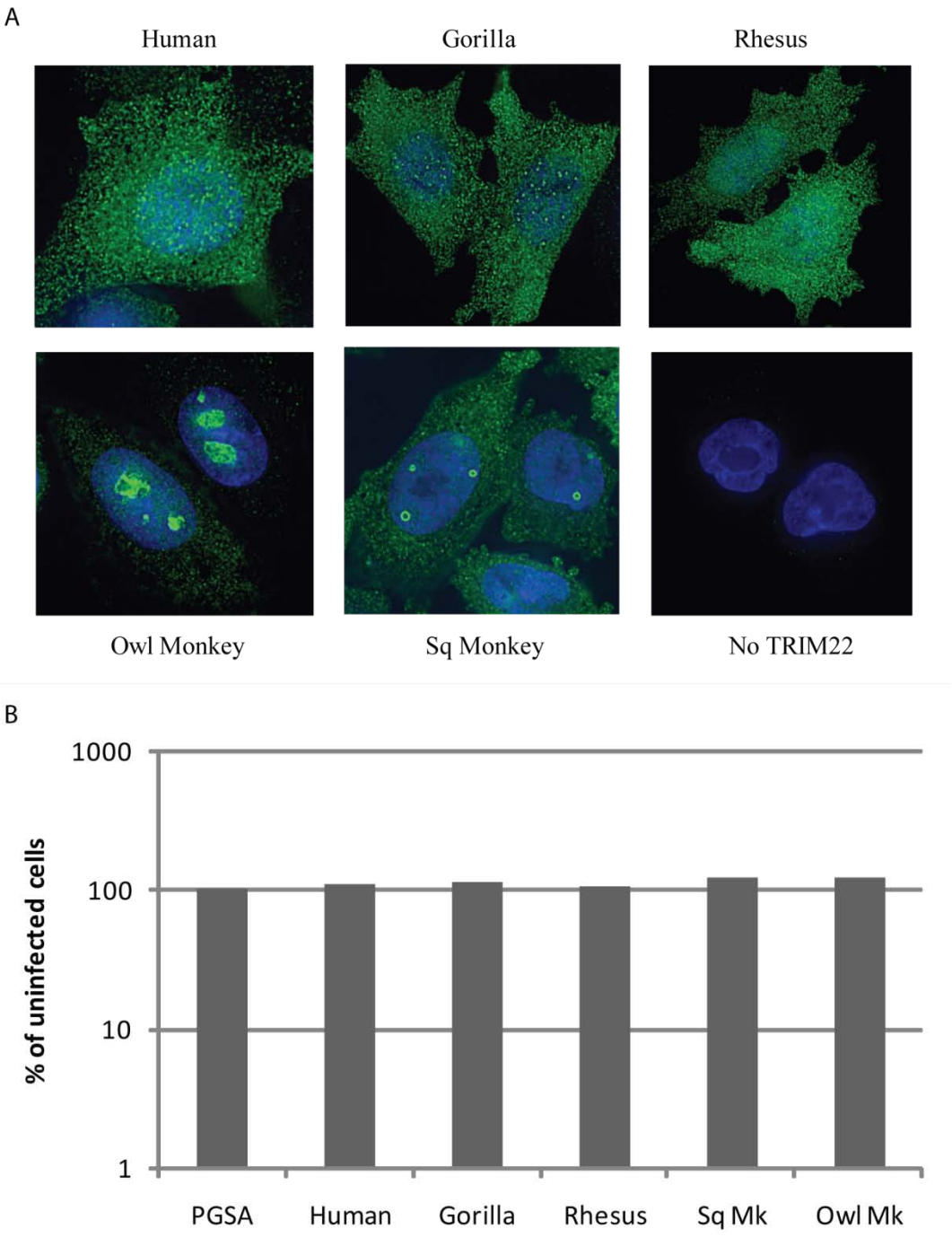


Figure 31 Localization and effect of TRIM22 on HERV-K_{CON} infection

A. Confocal microscopy of PGSA cells stably transduced with human, gorilla, rhesus macaque, squirrel monkey, and owl monkey TRIM22 proteins fused with an HA tag. The cells were permeabilized, fixed with 4% PFA, and stained with anti-HA antibody. B. PGSA cells stably transduced with TRIM22 proteins were infected with HERV-K_{CON} VLPs and fixed two days after for FACS analysis. CCGBX packageable genome was used to make the VLPs.

tetherin with pCRVI/Con Gag-PR-Pol in 293T cells resulted in a dramatic decrease of VLPs in the supernatant, from a ten-fold decrease at the lowest concentration of tetherin tested to almost complete inhibition at the highest concentration (Figure 32) (Jouvenet et al., 2009b). Co-expression of HIV-1 Vpu partially restored the amount of VLP in the supernatant to within half of the level released when no tetherin was co-expressed (Figure 32).

This demonstrates that tetherin can effectively reduce the amount of HERV-K particles produced from an infected cell. As the entire virus was not used in the experiment, it is possible that proteins not expressed in pCRVI/Gag-PR-Pol such as Env may have a Vpu-like activity to counter the antiviral effects of tetherin. Indeed, the Env protein of HIV-2_{ROD10} strain has a Vpu-like ability to enhance retrovirus release (Bour and Strebel, 1996). Whether HERV-K_{CON} Env has similar ability has not been tested. Furthermore, as tetherin targets a very wide array of enveloped viruses, it is unclear whether HERV-K truly interacted with tetherin in the past, or is susceptible because of tetherin's general mechanism of action. Furthermore, as tetherin is only expressed in

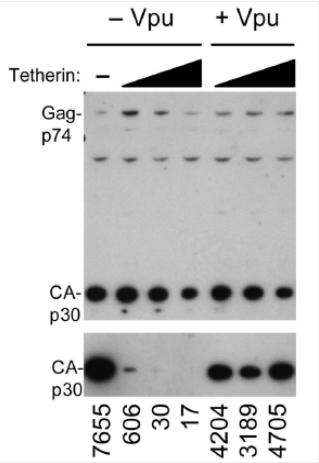


Figure 32 Effect of tetherin on HERV-K_{CON} VLP release

HERV-K particles were generated by transfecting 1.8 µg of CCGBX, 0.5 µg of pCRV1/ConGag-PR-Pol, and 0.5 µg of pCR3.1/Rec, along with 0, 100, 200, or 400 ng of pCR3.1/Tetherin-HA, in the presence or absence of pCR3.1/Vpu. Semi-quantitative analysis of western blot was carried out by scanning using an Alpha Innotech imaging system and rendering the blots as TIFF files. Band intensities (given in arbitrary units) associated with released VLPs, at each amount of transfected tetherin expression plasmid, were quantitated using Image J software (W. S. Rasband, U.S. National Institutes of Health, Bethesda, MD [http://rsb.info.nih.gov/ij/], 1997 to 2008). Histograms of pixel intensity for each lane were generated using the plot lane function in the Gels Analysis toolbox, and the area under the curve of the histogram peak was then calculated.

the presence of type-I IFN in most cells, if HERV-K does not induce an IFN response during infection, then tetherin may not affect its replication in vivo at all (Neil et al., 2007). Hence, it is difficult to conclude that HERV-K_{CON} has influenced the evolution of tetherin.

3.4 Effect of APOBEC proteins on HERV-K_{CON} infection

Like tetherin, A3G was discovered as the target of an HIV-1 accessory protein (Sheehy et al., 2002). While human A3G cannot block infection of HIV-1 due to the actions of Vif, it may still counteract infection from other viruses which have not evolved to avoid it. The effect of human APOBEC3 proteins on HERV-K infection was determined. Although it would be optimal to test infection using physiologically relevant APOBEC3 protein levels, the expression level of most APOBEC3 proteins in various tissues is unknown. Moreover, the tissue tropism of HERV-K is unknown; hence it is difficult to find the APOBEC expression level at which HERV-K naturally infects. Instead, by titrating the amount of expression vector in the assays, APOBEC3 proteins were expressed at various, and relatively low, levels that mimicked the range of protein levels at which hA3G, hA3F, and hA3B inhibit HIV-1 infection in the absence of Vif (Figure 33A). HERV-K_{CON} VLPs were generated in 293T cells in the presence of each of the C-terminally HA-tagged human APOBEC3 proteins. Western blot analysis of cell lysates showed that APOBEC3 protein expression varied (Figure 33A), but the range of expression levels for most of the proteins overlapped as a result of transfecting various levels of the corresponding expression plasmids. However, hA3DE and hA3H were

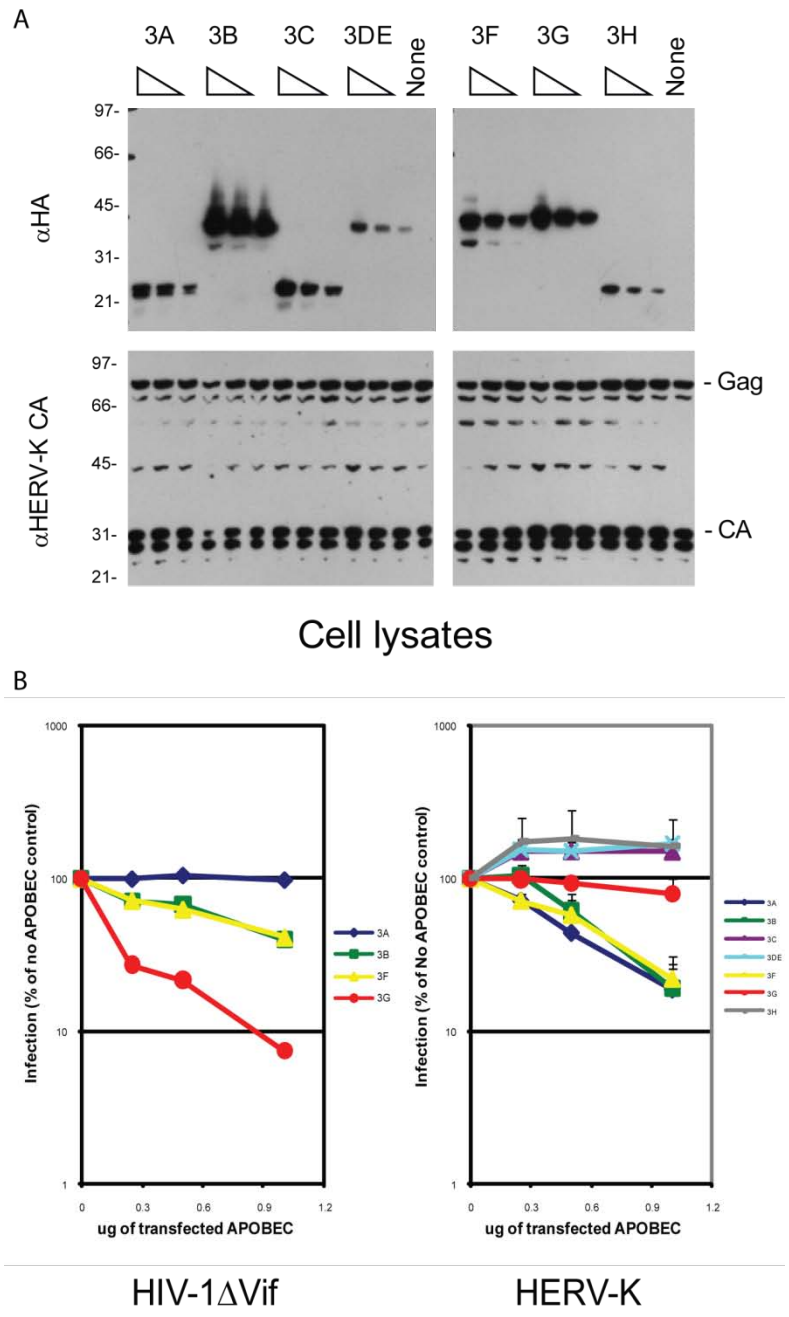


Figure 33 Expression and effect of human APOBEC3 proteins on HERV-K_{CON} infection

A. Anti-HA (top panel) and anti-HERV-K CA (bottom panel) Western blot of 293T cell lysates transfected with HERV-K_{CON}, VSVG, and APOBEC-HA plasmids. B. Infection of CEM cells with HIV-1 (left panel) or 293T cells with HERV-KCON virions, generated in the presence of the indicated APOBEC3-HA proteins. Infectious units were quantified as GFP^+ cells using fluorescence-activated cell sorter analysis 2 days postinfection, and are expressed as a percentage of the number of infected cells (typically 15 to 30%) that were obtained in the absence of an APOBEC3 protein.

comparatively poorly expressed. Importantly, HERV-K Gag was expressed equally in all conditions. The only exception to this was at high levels of hA3B expression, which appeared to slightly reduce the levels of HERV-K Gag expression, presumably due to marginal toxicity. Fresh 293T target cells were infected with HERV-K virions generated in the presence of each of the APOBEC3 proteins. As can be seen in Figure 33B, hA3A, hA3B, and hA3F inhibited HERV-K_{CON} infection by approximately fivefold at the highest concentration tested. Only marginal inhibition of infection was seen with hA3G, while hA3C, hA3DE, and hA3H did not inhibit infection. Clearly, the relative sensitivity of HERV-K to the various APOBEC3 proteins differ greatly from that of HIV-1 (Figure 33B).

3.5 Hypermutated HERV-K proviruses in modern human DNA

While several APOBEC3 proteins appeared capable of inhibiting HERV-K infection in vitro, to determine whether restriction of HERV-K infection might have occurred in vivo, evidence of APOBEC3-induced mutation in HERV-K proviruses that are present in modern human DNA was sought. Specifically, 16 human-specific full-length HERV-K(HML-2) proviruses were examined and especially for biases in the patterns of mutation therein, relative to the pseudo-ancestral HERV-K_{CON} sequence.

Overall, G-to-A and C-to-T substitutions were the most abundant changes in the proviruses as a whole (Figure 34A), as would be expected from genomic sequences that are not under purifying selection. Fourteen HERV-K proviruses showed a comparatively minor increase in the frequency of G-to-A changes and C-to-T changes relative to other

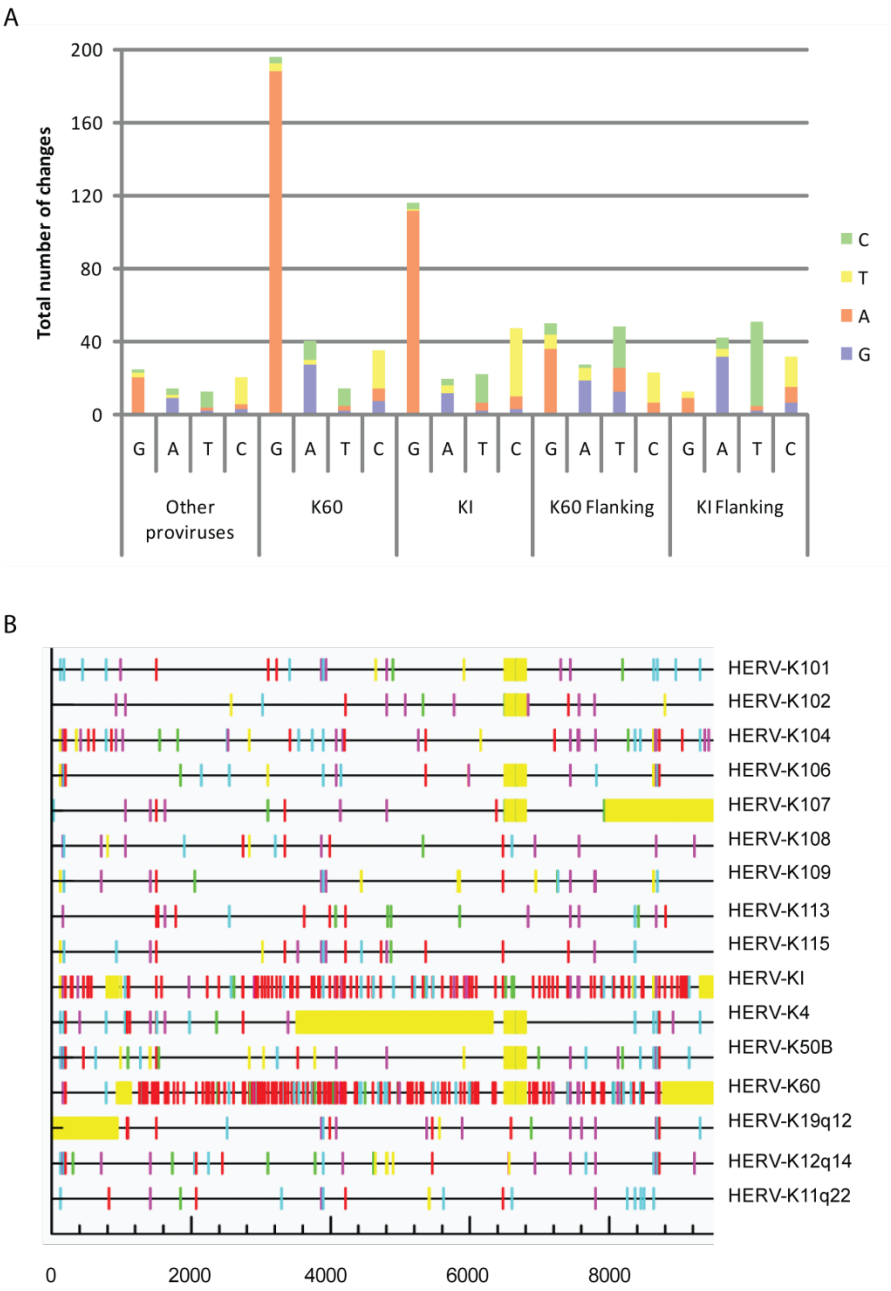


Figure 34 Nucleotide changes in human-specific HERV-K proviruses relative to HERV-K_{CON}

A. The numbers of changes of each type from the HERV-K_{CON} to sequences in endogenous proviruses are plotted. For comparison of sequences flanking HERV-K60 and HERV-KI, 2 kb of genomic sequence immediately proximal to the 5' and 3' ends of the proviruses were compared for changes from chimpanzee to the orthologous human sequence. For each sequence comparison, the numbers of changes were normalized to enable direct comparison with the numbers of changes in the individual HERV-K60 and HERV-KI proviruses. B. Graphical representation of nucleotide changes relative to HERV-K_{CON} in the 16 human specific proviruses. Red, GG to AG; cyan, GA to AA; green, GC to AC; magenta, GT to AT; black, non-G-to-A transitions; yellow, gaps in the sequence.

changes (Figure 34A). However, two proviruses, HERV-K60 and HERV-KI, were exceptional in the total quantity and type of changes. Overall, they exhibited similar frequencies of C-to-T mutation as did the other 14 proviruses, but both HERV-K60 and HERV-KI exhibited a very high frequency of G-to-A changes relative to the HERV-K_{CON} (Figure 34A and B). Indeed, each of these individual proviruses had more G-to-A mutations than the other 14 proviruses combined; two-thirds and one-half of all the changes in K60 and KI, respectively, were G-to-A mutations (Figure 34A).

To ensure that the exceptional properties of the two apparently hypermutated proviruses were not due to their insertion into an unusually hypermutated region of the human genome, 2 kb of flanking genomic sequence at each end of the two proviruses were examined for evidence of hypermutation. Specifically, changes between the equivalent loci in humans compared to chimpanzees were examined. This is a conservative approach, since these HERV-K insertions are absent in chimpanzees and have, therefore, been resident in the human genome for less time than the flanking sequences have been diverging in the two species. As expected, G-to-A, C-to-T, and reciprocal A-to-G and T-to-C changes were found to be most abundant in comparisons of the human and chimpanzee flanking sequences (Figure 34A). However, G-to-A changes did not greatly outnumber other changes, suggesting that that the apparent hypermutation in the inserted proviruses occurred independently of the genomic context sequence, likely prior to their integration.

In addition to the 16 aforementioned proviruses, additional human-specific partial HERV-K sequences that lacked LTRs, as well as HERV-K proviruses (defined as group N by Romano et al.) that were not human specific, and group N CERV-K proviruses

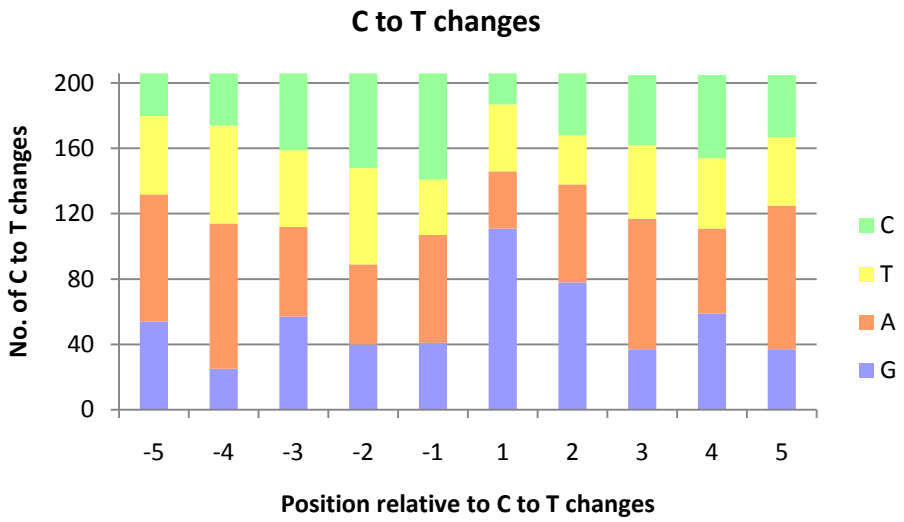
(chimpanzee counterparts of HERV-K), including four chimpanzee-specific insertions, were also examined, as were related group O proviruses. No evidence of hypermutation was evident in these sequences (data not shown).

3.6 Flanking nucleotide characteristics of G-to-A changes in hypermutated HERV-K proviruses.

A common cause of G-to-A and C-to-T substitutions in genomic DNA is spontaneous cytosine deamination. This occurs most often after methylation of cytosines in CG dinucleotides, followed by spontaneous deamination of 5-methylcytosine to a thymine (CG to ^{m}CG to TG; altered nucleotide underlined). This series of events would lead to an overabundance of plus-strand C-to-T changes with G in the +1 position relative to the C-to-T mutation (CG to TG). Conversely, the same deamination event on the minus strand would lead to plus strand G-to-A changes with an overabundance of C at the -1 position relative to the mutated nucleotide (CG to CA). Indeed, in most of the HERV-K proviruses examined (the 14 nonhypermutated proviruses), C-to-T changes were significantly enriched for G in the +1 position (Figure 35A), and G-to-A changes were significantly enriched for C in the -1 position (Figure 35B), suggesting most C to T and G-to-A changes were a result of spontaneous cytosine deamination, as would be expected of DNA elements that are long-term residents of the human genome.

To determine whether the excessive G-to-A changes present in HERV-K60 and HERV-KI were indicative of APOBEC3- induced hypermutation, and if so, to determine the identity of the responsible protein, the nucleotides flanking the G-to-A changes were examined (Figure 36). At least some of the APOBEC3 proteins have signature

A



B

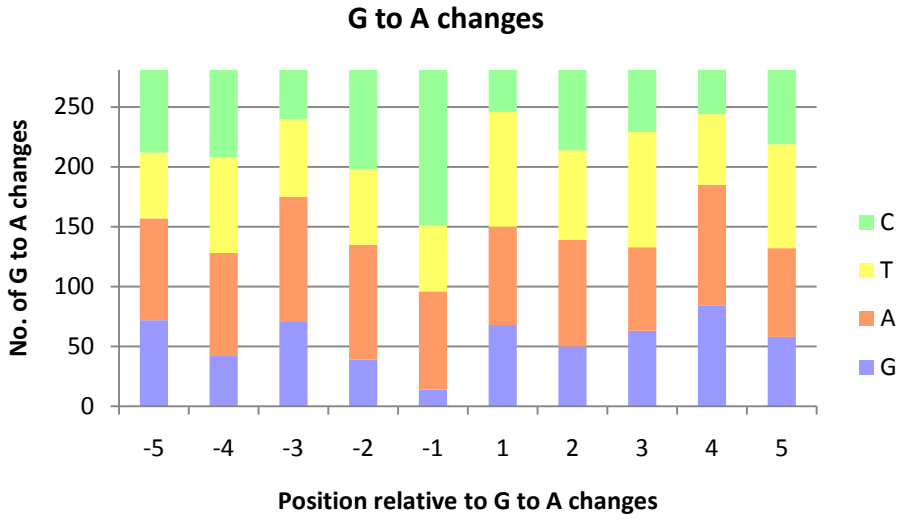


Figure 35 Flanking nucleotides of C-to-T and G-to-A changes in non-hypermutated HERV-K

Nucleotide occurrence at five positions 5' and 3' to (A) C to T and (B) G-to-A changes was catalogued, using the HERV-K_{CON} sequence as a reference in non-hypermutated HERV-K proviruses. The absolute number of times that each nucleotide occurred at each position relative to each change in the nonhypermutated proviruses is plotted.

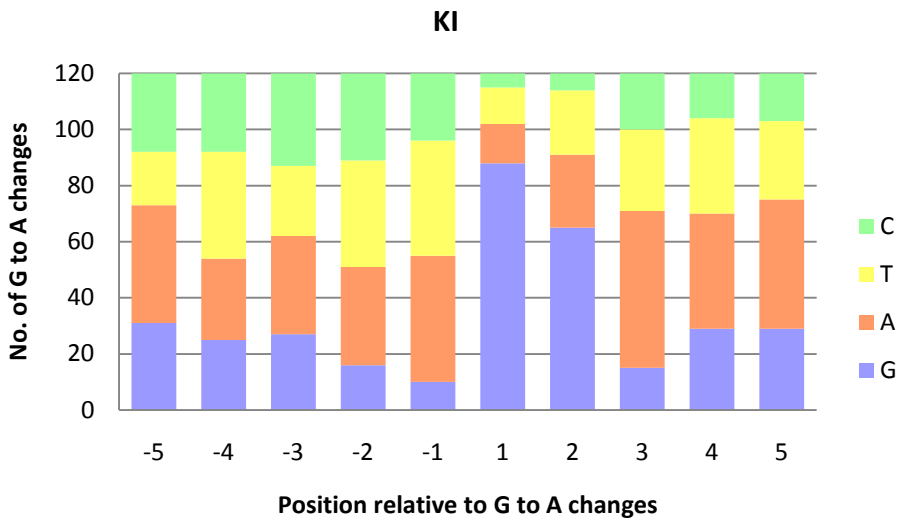
dinucleotide preferences; for example, hA3B and hA3F prefer to deaminate cytosines within TC dinucleotides, resulting in GA-to-AA mutations on the viral plus strand, while hA3G exhibits a bias for deaminating CC dinucleotide, leading to plus-strand GG-to-AG changes. In fact, of all the human APOBEC3 proteins, only hA3G exhibits the bias toward GG-to-AG changes.

In both HERV-K60 and HERV-KI, a strong GG dinucleotide bias was detected at G-to-A changes (Figures 36 and 37A). This preference indicates that hA3G was likely responsible for the excessive G-to-A mutations in HERV-K60 and HERV-KI. In other proviruses, and the genomic DNA flanking HERV-K60 and HERV-KI, no such dinucleotide preference was detected (Figure 35B and data not shown). Furthermore, a strong, statistically significant bias for GGG trinucleotides at G-to-A mutated positions was also evident upon examination of the third nucleotide in all GG-to-AG substitutions (Figure 37B). Notably, the GGG preference has been detected in previous studies of hA3G with a Vif-deficient HIV-1 (Yu et al., 2004d), further supporting the notion that hA3G was responsible for the excessive G-to-A mutations in HERV-KI and HERV-K60.

3.7 Hypermutation of HERV-K by APOBEC3 proteins during in vitro replication.

To test whether the in vivo hypermutation changes could be recapitulated in vitro, HERV-K_{CON} virions were generated in the presence of each human APOBEC3 protein using the HERV-K packaging construct CCGBX-P (Figure 21). This construct contains GFP as the marker and a 53 bp HIV-1 sequence (GATCTGAGCCTGGGAGCTCT CTG GCTTGTGACTCTGGTAACTAGAGATCCCTC) 34 bps after the start of 3’ LTR to

A



B

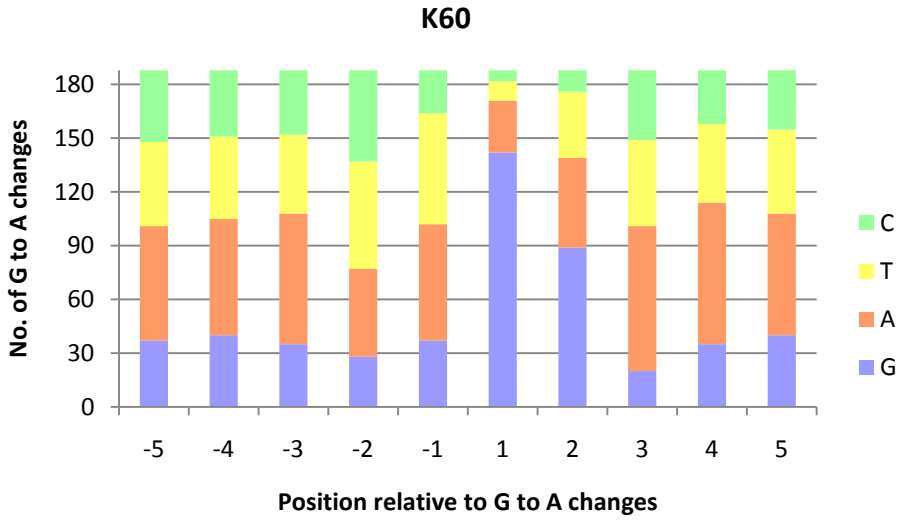


Figure 36 Flanking nucleotides of G-to-A changes in hypermutated HERV-K

Nucleotide occurrence at five positions 5' and 3' to G-to-A changes was catalogued, using the HERV-K_{CON} sequence as a reference. The absolute number of times that each nucleotide occurred at each position relative to each G-to-A mutation in (A) HERV-K60 and (B) HERV-KI is plotted. The *P* value in panel B for deviation from random nucleotides at +1 and +2 positions was > 0.0001, calculated using a chi-square test of independence.

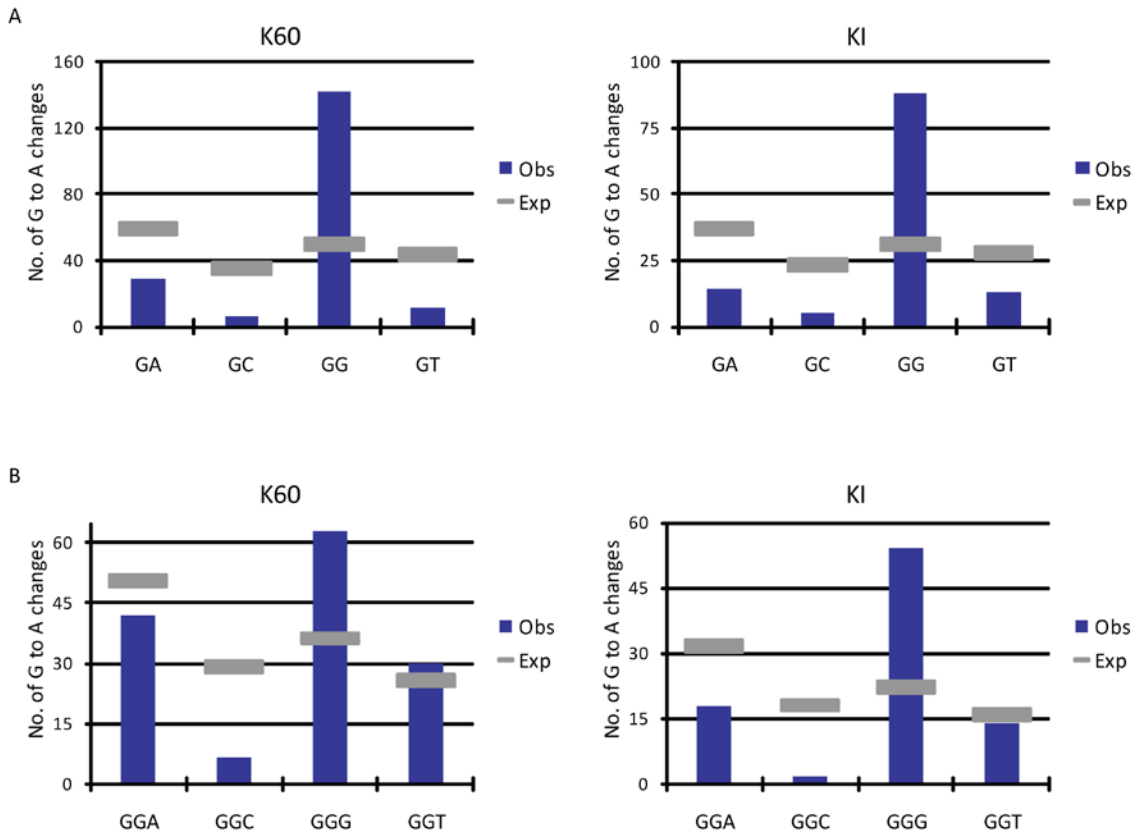


Figure 37 Frequency of di- and trinucleotide for all G-to-A and GG-to-AG changes

Frequencies of each di- (A) and trinucleotide (B) for all G-to-A and GG-to-AG changes, respectively, in HERV-K60 and HERV-KI are plotted as black bars. The expected numbers of G-to-A mutations in each sequence context, based on di- and trinucleotide composition, of HERV-K_{CON} is represented as a horizontal gray line. The *P* value, for deviation from random di- and trinucleotide preference, was > 0.0001, calculated by a chi-square test of independence.

allow selective amplification of newly synthesized HERV-K in a background of existing HERV-K proviruses present in human cells. This extra sequence should be copied into both LTRs during reverse transcription, and provide the means to distinguish de novo integration events from endogenous HERV-K proviruses. As integration is largely determined by the IN protein independently of the virus sequence, this extra sequence should have minimal effect on the integration site preference (Lewinski et al., 2006).

HERV-K and APOBEC plasmids were transfected into 293T cells. Five hours post transfection, the medium was replaced with fresh DMEM. Two days post transfection, the supernatant was DNase treated at 37°C for an hour, then used to infect fresh 293T cells. 11 hours post infection, the cells were harvested for total DNA extraction using QiaAmp Blood and Tissue Kit (QIAGEN). A 762-bp sequence of nascent HERV-K_{CON} DNA was then amplified from the total DNA using primers targeting the EGFP insert in the vector genome and the inserted HIV-1 sequence. Two controls were done to establish the success of this strategy and to show that DNase treatment reduced contaminating plasmid DNA in the virion preparations to subdetectable levels. First, PCR amplifications were done using primers targeting the plasmid backbone. Second, infections were done using virions harboring an inactivating point mutation in the HERV-K reverse transcriptase. In both cases, PCR products were not detected, indicating that the sequences generated following HERV-K infection genuinely represented infection-dependent, de novo-synthesized HERV-K DNA, not a transfer of residual transfected DNA.

hA3A, hA3B, hA3F, and hA3G were found to induce hypermutation of HERV-K_{CON} during in vitro infection. However the patterns and frequency of hypermutation were different (Figure 38). In particular, nearly all HERV-K clones generated in the presence of hA3G had G-to-A mutations, but each had a low to moderate number of changes (median of six G-to-A changes per clone). Conversely, hA3A, hA3B, and hA3F each induced hypermutation in a minority of clones, but hA3A and hA3B hypermutated clones had a very high burden of mutations (37 changes for the sole hA3A hypermutated clone, and a median of 43.5 G-to-A changes for hA3B hypermutated clones). Only a few

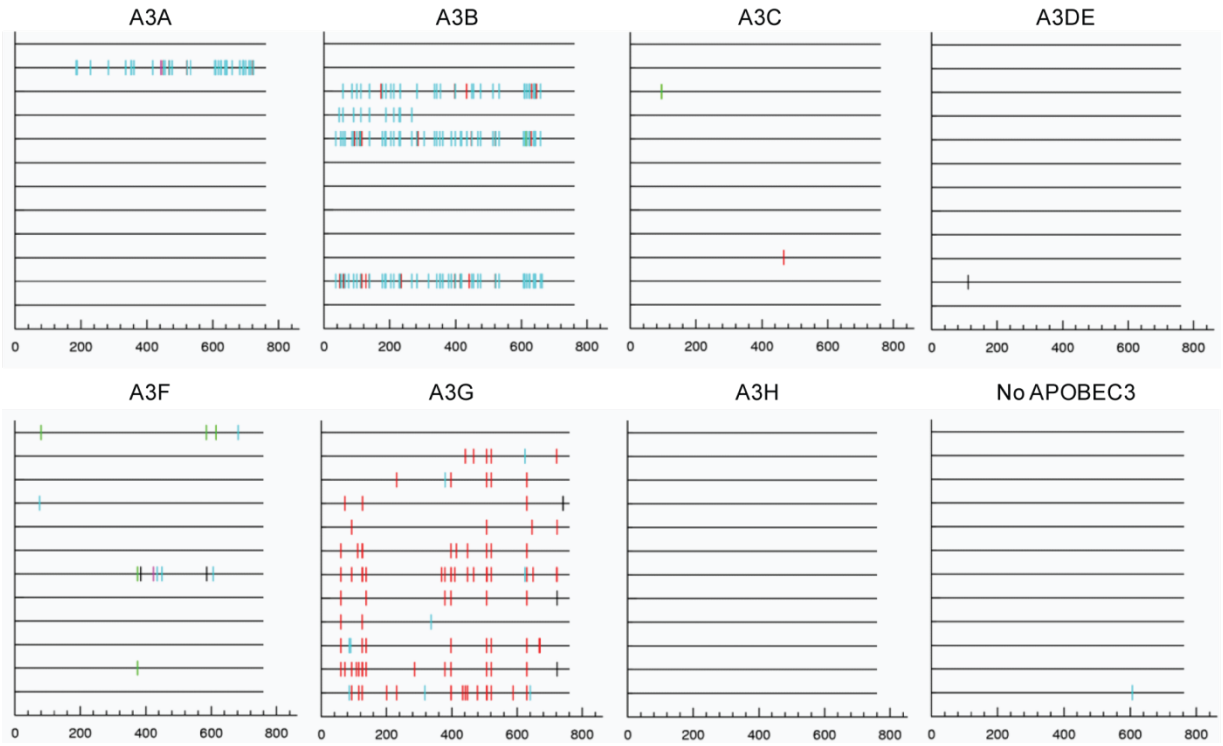


Figure 38 Hypermutation of HERV-K_{CON} during in vitro infection

All changes in HERV-K_{CON} reverse transcripts relative to preinfection sequence were analyzed for +1 site preference using HYPERMUT, and the sequences of 12 HERV-K clones generated during infection in the presence of each APOBEC protein are represented as horizontal lines. Mutations are indicated as vertical lines. Red, GG to AG; cyan, GA to AA; green, GC to AC; magenta, GT to AT; black, non-G-to-A transitions.

changes were seen in HERV-K DNA generated in the presence of hA3F (median of four changes per clone). Of the four human APOBEC3 proteins found to hypermutate HERV-K in vitro, only hA3G exhibited the GG dinucleotide and GGG trinucleotide bias for the generation of G-to-A mutations (Figures 38 and 39), as has previously been reported for hA3G mutation of HIV-1 (Yu et al., 2004c). Moreover, hA3F and hA3B exhibited the expected GA dinucleotide bias at positions where G-to-A mutations were generated (Figures 38 and 39). hA3A also showed the same GA bias, with the caveat that only a single HERV-K_{CON} clone was found to be mutated by hA3A (Figures 38 and 39). Hence,

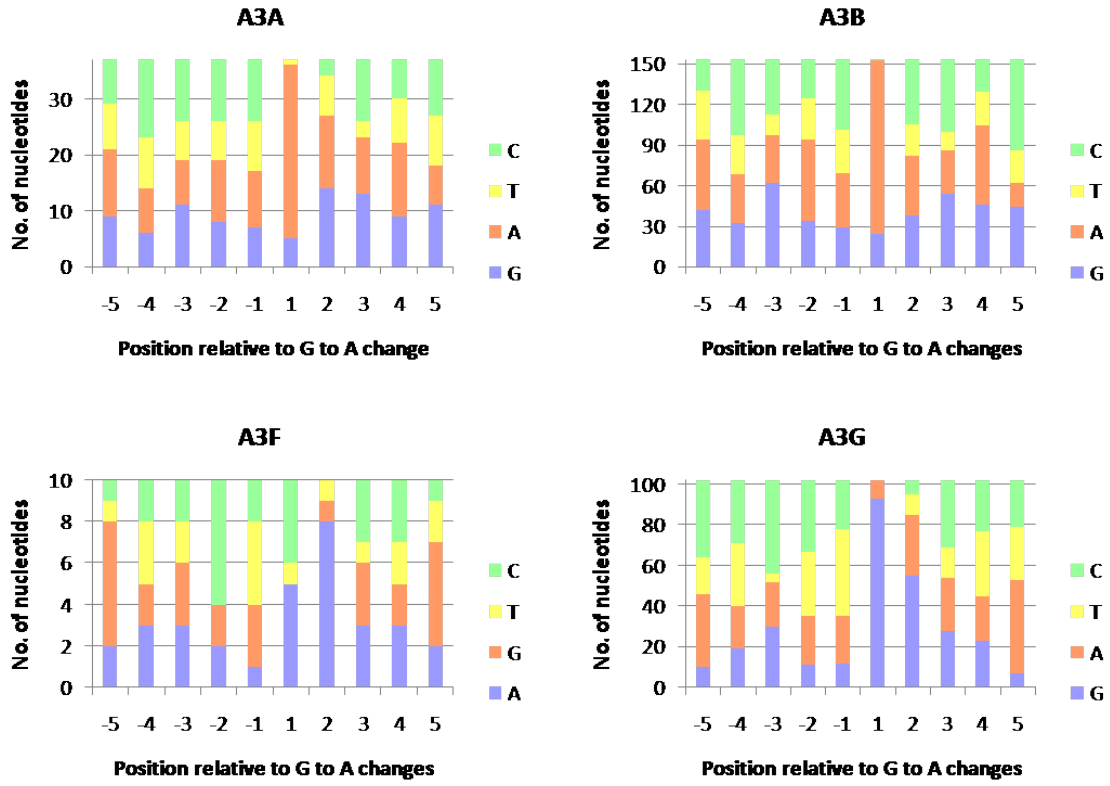


Figure 39 Flanking nucleotides of G to A changes incurred during in vitro infection of HERV-K_{CON}

Flanking nucleotide sequence context surrounding mutations generated by APOBEC3 proteins during HERV-K infection in the presence of hA3A, hA3B, hA3F, and hA3G. The numbers of times that each nucleotide occurred at five positions 5' and 3' to each G-to-A change were plotted.

among all seven of the human APOBEC3 proteins, four appear intrinsically capable of inducing hypermutation in HERV-K, but only hA3G was capable of hypermutating HERV-K during in vitro infection with a characteristic bias that very closely resemble mutations found in the endogenous HERV-K60 and HERV-KI proviruses (compare Figures 36 and 39).

Moreover, when the 273 bp of HERV-K_{CON} sequence that was analyzed in the hA3G mutagenesis assay were compared with the corresponding sequence in HERV-K60 and HERV-KI, 10 out of the possible 75 G-to-A mutations were found in either or both

HERV-K60 and HERV-KI, while 15 out of the possible 75 G-to-A mutations were represented in corresponding sequences that were mutated by hA3G during in vitro infection (Figure 40). Notably, six of these G-to-A mutations were at identical positions, a highly significant correlation (P = 0.003, Fisher's exact test), lending further supporting to the notion that hA3G was responsible for hypermutation of HERV-K60 and HERV-KI.

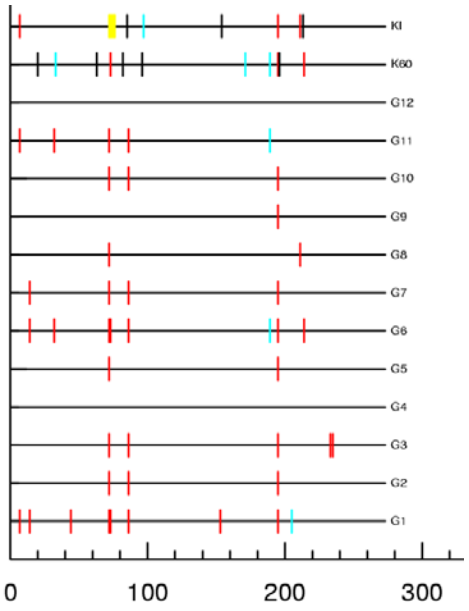


Figure 40 Comparison of in vivo and in vitro mutations

Comparison of mutations relative to HERV-K_{CON} that appear in naturally hypermutated proviruses (HERV-K60 and HERV-KI) and in the 273 nucleotides of mutated HERV-K sequence generated in vitro in the presence of hA3G. Changes relative to HERV-K_{CON} are represented graphically on horizontal lines and are color coded according to the nucleotide appearing at the -1 site for HERV-K60 and HERV-KI and 12 HERV-K clones generated during infection in the presence of hA3G. Mutations are indicated as vertical lines. Red, GG to AG; cyan, GA to AA; green, GC to AC; magenta, GT to AT; black, non-G-to-A transitions; yellow, gaps in sequence.

Chapter 4. Integration of HERV-K_{CON}

4.1 Introduction to retrovirus integration

Integration is one of two hallmark events of retrovirus replication. Where a retrovirus integrates may affect its transcriptional efficiency, as well as the host cell's viability due to insertional mutagenesis. Each retrovirus has its own integration site preference, but the preference usually relates to genomic features, such as CpG islands or transcription units, rather than a specific sequence that the integrase recognizes. For example, studies have shown that HIV-1 prefers to integrate into active transcription units, MLV prefers CpG-rich promoter regions, while ASLV integrates relatively randomly with a minor preference for transcription units (Barr et al., 2005; Mitchell et al., 2004; Narezkina et al., 2004; Schröder et al., 2002b; Wu et al., 2003). The preference based on genomic features means that the number of potential sites for integration is large and predicting the exact insertion location is impossible, presenting an obvious drawback for the use of retroviruses as vectors for gene therapy.

As no replicating HERV-K was available until recently, studies of HERV-K integration have thus far focused on examining the resident proviruses in the human genome sequence. Previous studies, and examination of the 404 insertions in this work identified in Table 1 relative to chromosome size suggests that HERV-K accumulation is not a random event, which would predict the number of insertions per chromosome to positively correlate with chromosome size (Figure 8) (Medstrand et al., 2002; van de Lagemaat et al., 2006; Villesen et al., 2004). Rather, HERV-K is found more frequently outside of transcription units and when in genes, in reverse transcriptional orientation

relative to the gene (Medstrand et al., 2002; Smit, 1999; van de Lagemaat et al., 2006; Villesen et al., 2004). Whether this integration pattern reflects a true preference of HERV-K or is a result of post-integration selection is unknown, but understanding the preference will help assess the mutagenic potential of HERV-K.

The following work was conducted as a collaboration between the Bieniasz laboratory and Bushman laboratory at University of Pennsylvania. The infection of 293T and HT1080 cells, DNA extraction, and HML2(85) data set collection was conducted by myself, the preparation of the DNA samples for the pyrophosphate mass sequencing by Dr. Troy Brady, the MMTV data set was derived by Dr. Keshet Ronen, the analysis of de novo insertions relative to genomic markers by Dr. Charles Berry (UCSD), and the analysis of de novo insertions relative to endogenous HERVs by Dr. Troy Brady and Dr. Nirav Malani.

4.2 HERV-K integration preference relative to genomic markers

The HERV-K_{CON} single cycle infection system is an optimal system to test integration preference. Attempts were made to find a germ cell-like cell line as target cells, as these would most closely mimic the cell populations that would have been infected to generate endogenous proviruses. Teratocarcinoma cells Tera-1, Tera-2, and PA-1 were tested for their susceptibility to retrovirus infection, but they were all relatively resistant to HERV-K infection (less than 10% GFP^+ cells, data not shown). Thus two human cell lines, 293T and HT1080, already known to be susceptible to HERV-K_{CON} infection, were used as target cells. Numerous studies of HIV-1, ASLV, and MLV integration showed that the preference did not change between different cell lines,

including cell lines of a different species, hence it is likely that using 293T and HT1080 cells will not result in a significant discrepancy relative to the native target cells (Barr et al., 2006; Barr et al., 2005; Mitchell et al., 2004). Furthermore, 293T and HT1080 cells are both human cell lines, but are derived from different organs (293T and HT1080 cell lines are derived from fetal kidney and fibrosarcoma, respectively), but this difference did not affect the integration site preference, further supporting the notion that these cell lines are representative of natural HERV-K target cells for integration preference.

To approximate the native conditions of HERV-K infection as closely as possible and enable use of human cells as target cells, CCGBX-P was used as the packaging genome, the same construct used to identify APOBEC hypermutated HERV-K sequence in Chapter 3.7. CCGBX-P and relevant HERV-K expression plasmids were transfected into 293T cells for virus production. Two days post transfection, the supernatant was treated with DNase for 1 hour at 37°C, then used to infect fresh target cells. The infection efficiency determined by FACS was between 10-20%. The genomic DNA of target cells was harvested two days post-infection and digested (MseI for 293T cells, MseI or ApoI for HT1080 cells) and ligated to linker sequences. The integration sites were PCR amplified using primers specific to the inserted HIV-1 sequence and linker sequence, and the amplicons used for template in pyrophosphate mass sequencing. A total of 1565 de novo integration sites (1064 sites from 293T cells, 501 sites from HT1080 cells) were identified and characterized. Despite the different organ and gender origins of 293T and HT1080 cells, comparison of integration relative to genomic markers did not show any discrepancy between the two cell lines, except insertion into Y chromosome in HT1080

cells that is not expected in 293T cell line, as it is derived from a female donor. Hence, data for the two cell lines are combined for analysis below, except when noted.

Controls were derived for each HERV-K_{CON} integration site. A large library of random sites was generated in silico, then the distances to restriction enzyme recognition sites were scored. Each experimental site was matched with three random control sites that were positioned the same number of nucleotides from a restriction site for the enzyme used to isolate the experimental site. That is, if an integration site was isolated after cleavage with ApoI, and the distance from the ApoI site to the edge of the HERV-K_{Con} sequence was 80 bp, then three random control sites were drawn from the pool that were also 80 bp from an ApoI site. Integration sites and matched random controls were annotated for proximity to genomic features, and the distributions were compared. The data is depicted as the ratio between observed HERV-K_{CON} sites to control matches. A ratio of 1 indicates that integration is random for the particular genomic feature, where as a ratio more or less than 1 indicates a preference or an aversion for the genomic feature.

The genomic characteristics examined in retrovirus integration include transcription units, gene density, gene expression, CpG islands, and DNase sites. Active transcription units have been hypothesized to be potential integration sites due to either the accessible nature of the genome during transcription, or the binding of viral proteins with transcription factors associated with the region (Schröder et al., 2002a). Gene density and gene expression are correlated features, as highly expressed genes tend to reside in regions of high gene density. These genomic features correlate with HIV-1 integration (Mitchell et al., 2004; Schröder et al., 2002a; Wu et al., 2003). CpG islands

and DNase sites are indicators of promoter regions, and correlate with MLV integration (Bird et al., 1985; Lander et al., 2001; Mitchell et al., 2004; Wu et al., 2003).

HERV-K_{CON} integration showed a minor but significant preference for transcription units, and correlated positively with gene density and expression (Figure 41A, D, E). More significantly, HERV-K_{CON} displayed a two and a half fold increased preference for proximity to CpG islands and a two-fold preference for proximity to DNase sites over matched controls (Figure 41B, C), which is similar but more subdued than the integration preference of MLV. Integration site information derived from studies of other viruses is analyzed together with the data derived from this study of HERV-K to give a relative perspective. The source of data is listed in Table 3. Notably, the retrovirus to which HERV-K is the most closely related whose integration preference has been analyzed is MMTV. Despite their sequence similarity, the integration preferences of these two retroviruses of the same genera seem to be different. The difference between HERV-K and MMTV is unlike the integration preference of two other related retroviruses, HIV-1 and EIAV, whose preference are highly similar (Hacker et al., 2006).

Insertions were also examined by chromosome (Figure 42). The null hypothesis of de novo insertions equaling the random insertions was statistically rejected ($p < 10^{-16}$), indicating that insertions do not correlate to the size of the chromosome. Chromosomes 1, 17 and 19 were highly favored, and chromosomes 4, 13, 18, X, Y were heavily disfavored. Earlier in this work, the number of endogenous HERV-K as a ratio to the size of chromosome was presented (Figure 8). In that data set, the insertions were heavily favored in chromosome 19 and disfavored in chromosomes 13, 15, and X. The de novo HERV-K integration preference discussed above may explain some of these

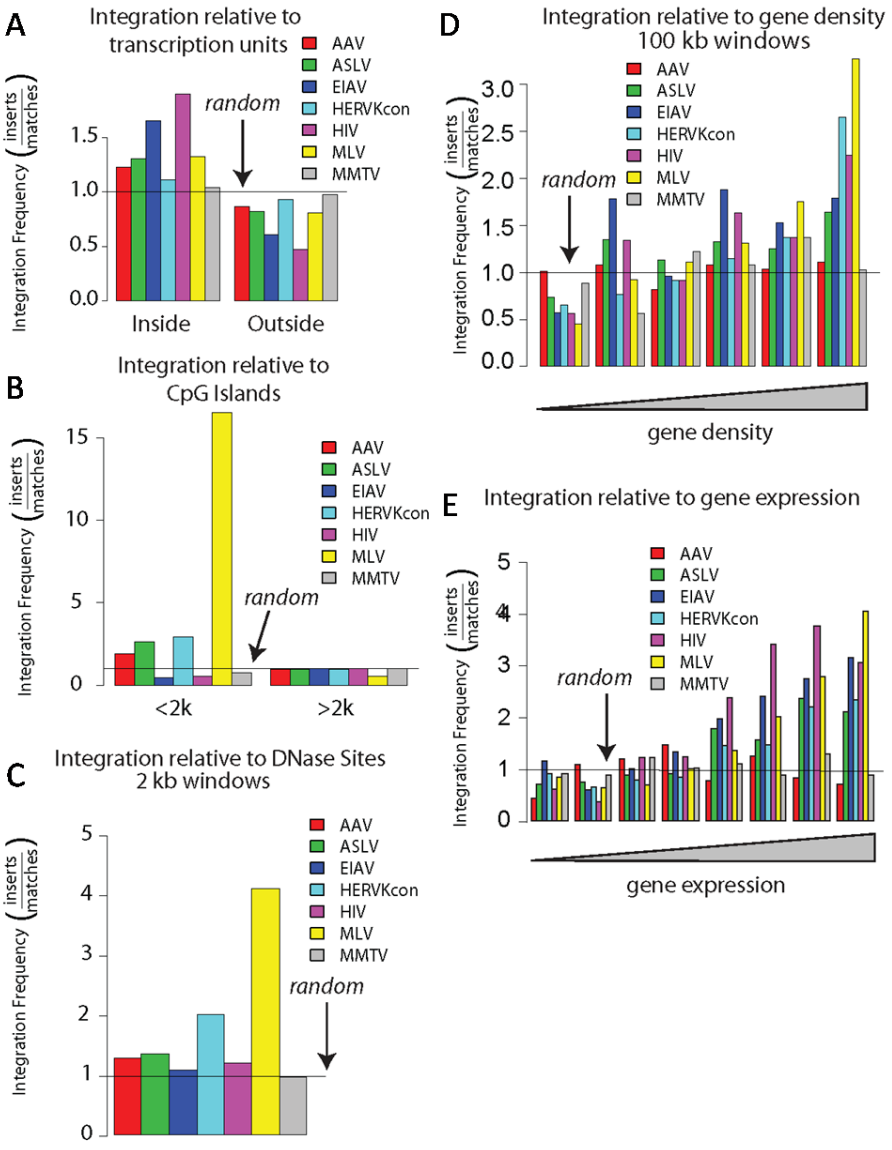


Figure 41 Integration target site selection of HERV-K_{CON} relative to other retroviruses

Values are reported as the proportion of integration events divided by random events. The bar at 1.0 represents the expected random distribution. The statistical significance of differences from the matched random controls is shown by the asterisks next to the legends. (*) 0.05 > P > 0.01; (**) 0.01 > P > 0.001; (***) P < 0.001. A. Integration frequency within RefSeq genes. B. Integration frequency as a function of gene density. The X-axis shows six bins of increasing gene density from lowest (left) to highest (right). C. Integration frequency relative to gene expression. All genes tested in 293T cells using the Affymetrix 133 array were divided into eight equal bins, then the proportions of integration sites in genes at each activity level were quantified and compared with random. The X-axis shows bins of increasing expression rank from lowest (left) to highest (right). D. Integration frequency relative to CpG islands, scored as the proportion of integration sites within 2 kb of an annotated CpG island. E. Integration frequency relative to sites of DNase I cleavage (Crawford et al. 2004), scored as the proportion of integration sites within 2 kb of an annotated cleavage site.

Table 3 Integration data sets used

Set	Size	Cell type	Enzyme	Reference
AAV	436	MHF2	MfeI, AvrII	Miller et al, 2005
ASLV	557	293T	AvrII, SpeI, NheI	Mitchell et al, 2004
EIAV	747	SupT1	MseI	Marshall et al, 2007
HERVKcon	1064	293T	MseI	This study
HERVKcon	501	HT1080	MseI, ApoI	This study
HIV	729	293T	MseI	Ciuffi et al, 2005
MLV	1588	293T	MseI	This study
MMTV	236	578T	MseI	Faschinger et al, 2008
ERV2	10573	NA	NA	Lander et al, 2001
HML2(85)	402	NA	NA	This study

(Ciuffi et al., 2005; Faschinger et al., 2008; Lander et al., 2001; Marshall et al., 2007; Miller et al., 2005; Mitchell et al., 2004)

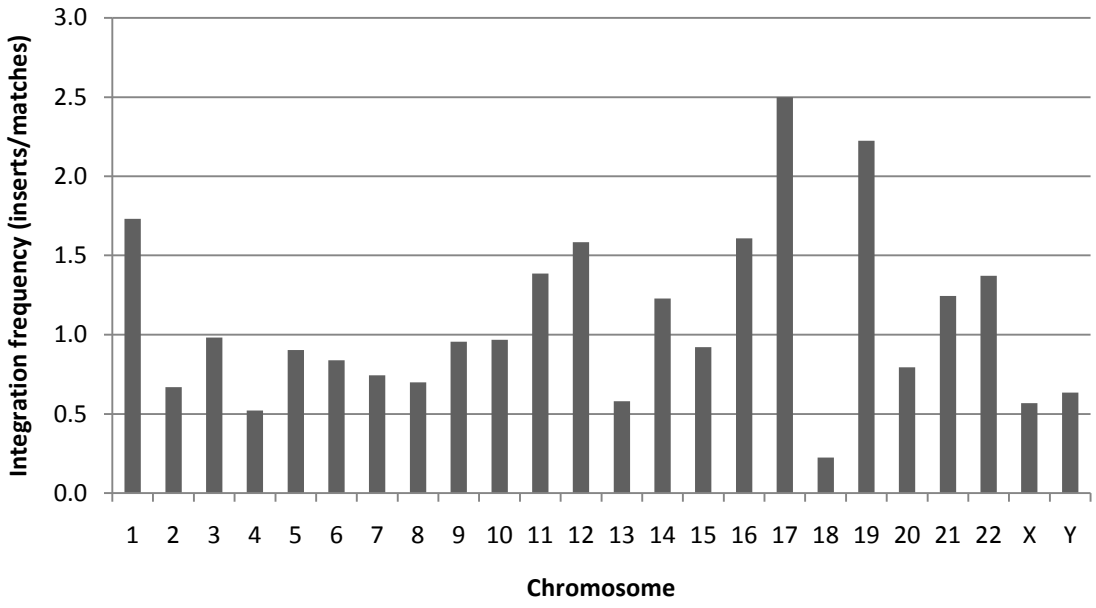


Figure 42 Integration target site selection of HERV-K_{CON} by chromosome

Values are reported as the proportion of integration events divided by random events. A ratio of 1.0 represents the expected random distribution. The statistical significance (p) of differences from the matched random controls is less than 10^{-16}.

characteristics. HERV-K_{CON}'s preference to integrate into transcription units and promoter regions, which should also be affiliated with transcription units, and chromsome 19's characteristic as most gene dense chromosome in the human genome, explains why chromosome 19 is highly favored for HERV-K insertions (Grimwood et al., 2004). In contrast, chromosome 13 contains the fewest endogenous HERV insertions and HERV-K_{CON} insertions likely because it has the lowest gene density of all chromosomes (Dunham et al., 2004). As insertions will be under selection pressure independent from the integration preference later on, it is not surprising that the distribution of de novo HERV-K_{CON} and endogenous HERV-K differ.

Both sex chromosomes are known for their high density of repeat elements. The two sex chromosomes, X and Y, were originally a pair of autosomes that have evolved into sex chromosomes over the past 300 million years (Ross et al., 2005). They now have diverged so far that only 5% of the Y chromosome overlap with sequences in X chromosome (Graves, 2006a; Skaletsky et al., 2003). Hence, unlike autosomal chromosomes, X chromosome only undergoes homologous recombination in females, while Y chromosome does not, except the 5% of its sequence that still resembles X chromosome (Ross et al., 2005). The heavy disfavoring for both de novo integration and endogenous insertions into X chromosome may be due to the low density of genes relative to autosomal chromosomes (estimated 1098 genes, 7.1 genes per Mb) (Ross et al., 2005). The reduced amount of homologous recombination, an opportunity to eliminate unwanted sequences, may explain the high quantity of other HERVs on the X chromosome despite the low level of HERV-K de novo integration. A more detailed analysis of Chromosome Y is presented later.

4.3 HERV-K_{CON} integration preference relative to endogenous HERV-K integration sites

For various biologically interesting reasons, the pattern of HERV-K insertions in the human genome may not match that generated following HERV-K_{CON} infection. To analyze the differences, the sites of de novo HERV-K_{CON} insertions were compared with those of two groups of endogenous proviruses: the HERV-K(HML-2) subfamily, and HERV-K superfamily. For the HERV-K(HML-2) data set, sequences with 85% matches to the HERV-K_{CON} LTR sequences were collected, marking the most recently acquired and evolutionarily youngest HERV-K (HML-2) elements. This dataset was termed termed HML2(85) and contained a total of 402 integration sites. The 85% cutoff was determined by comparison of the percent nucleotide identity among HML-2 elements, which ranges from 99% to 85%, with the percent identity between HML-2 and the next closest HERV-K subfamily, HML-1, which ranges between 70% and 80% (Medstrand and Blomberg, 1993a). For the HERV-K superfamily data set, RepeatMasker (http://www.repeatmasker.org) was used to generate a large set of all Class II HERV related sequences in the human genome (ERV2 data set; 10,573 integration sites). The ERV2 set combines all subfamilies of the HERV-K superfamily (HML1 through HML10), including both old and young ERV2s.

Integration of each data set relative to transcription units, CpG islands, DNase sites, gene expression, gene density, and GC content were examined. As previously noted, HERV-K_{CON} has a minor preference for integration in transcription units (Figure 41A and 43A). HML2(85) and ERV datasets, however, have a clear preference for integration outside transcription units (Figure 43A), a statistically significant difference with HERV-

K_{CON} data set. HERV-K_{CON} showed a two-fold preference over random matched controls for CpG islands, while HML2(85) and ERV2 showed a one and a half fold and just over one fold preference over controls (Figure 43B). For DNase sites, HERV-K_{CON} had a two-fold preference, while HML2(85) and ERV2 were disfavored from DNase I sites (Figure 43C). The preference of HERV-K_{CON} increases gradually proportionally to GC content, while the peak and decline before the highest GC content group, and ERV2 peaks before HML(85) data set and declines as well (Figure 43D). Hence, for all genome feature examined, a general pattern emerged where HERV-K_{CON} data set displayed a preference at one end of a range and the ERV2 data set at another, with HML2(85) having an intermediate phenotype between the two data sets. This suggests a two-step model of HERV accumulation where HERV-K integrates according to its preference, then the proviruses are selected upon over time. Hence, HML-2(85) data set may represent insertions that had the phenotype of HERV-K_{CON} insertions before, and will become like the ERV2 data set later on.

The preference for gene density increased proportionally for all data sets, with the biggest increase for HERV-K_{CON} and smaller increases for the older data sets (Figure 43E). Gene expression also correlated positively with HERV-K_{CON} integration, but were disfavored by HML2(85) and ERV2 data sets, except at the highest gene expression group for HML(85), where integration was preferred at one and a half fold over controls (Figure 43F). Combined with the HERV-K_{CON}'s preference for transcription units, these data suggest that after integration into gene dense regions, HERV-K elements which have inserted into expressed parts of gene dense region are selected out. Hence, in gene dense regions, HERV-K will accumulate outside of genes over time.

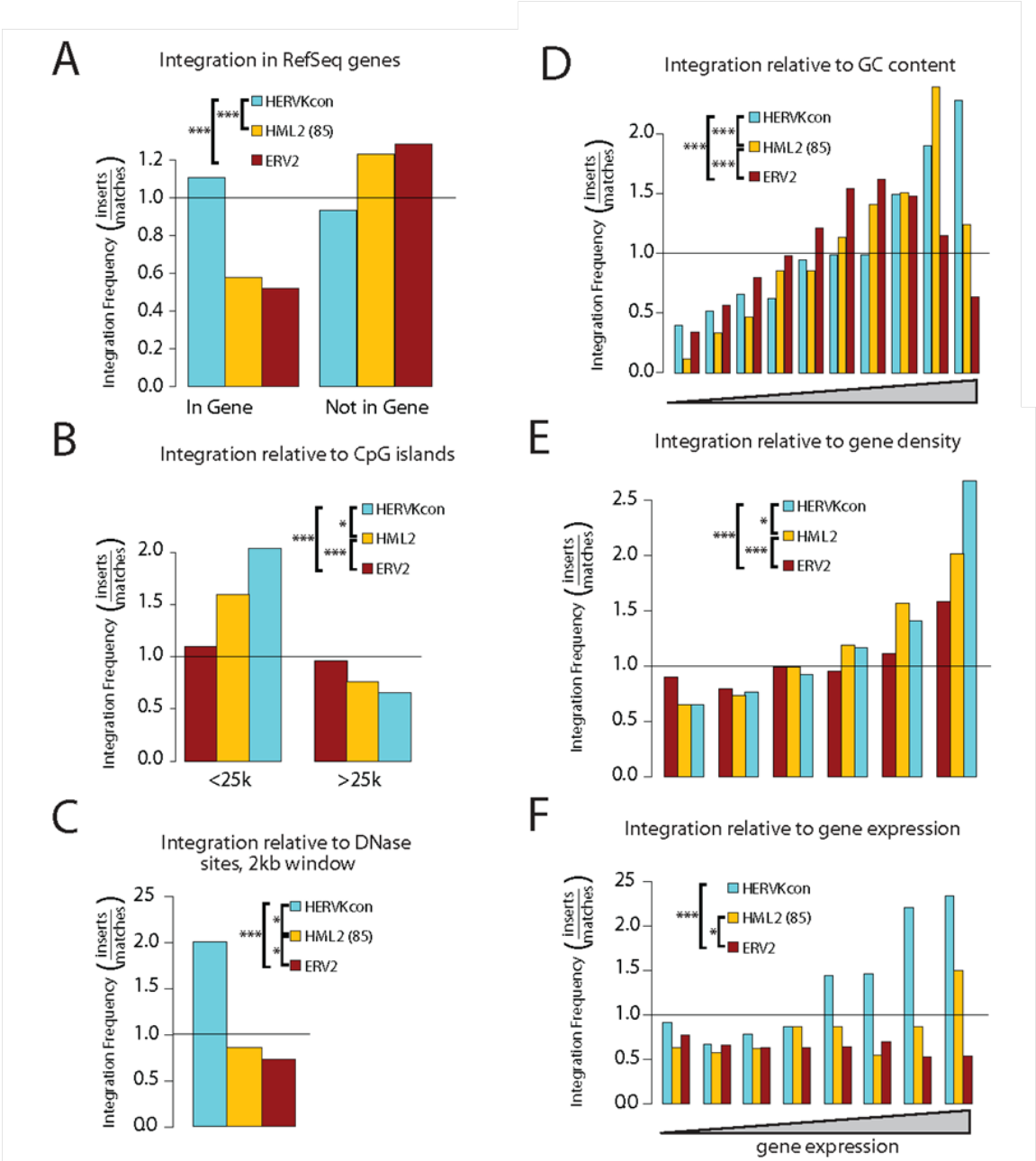


Figure 43 Integration of HERV-K_{CON} versus resident ERV2 elements

Values are reported as the proportion of integration events divided by random events. The bar at 1.0 represents the expected random distribution. The statistical significance of differences between data sets is shown by the asterisks next to the legends: (*) 0.05 > P > 0.01; (**) 0.01 > P > 0.001; (***) P < 0.001. A. Integration frequency relative to transcription units as defined by the RefSeq database. B. Integration frequency relative to CpG islands. C. Integration frequency relative to DNAse I cleavage sites, 2-kb windows. D. Integration frequency relative to G/C content, 5-kb windows. E. Integration frequency relative to gene density. F. Integration frequency relative to gene activity. In this plot, Affymetrix microarray analysis was used to rank the activity of all genes queried, then the ranks were distributed into eight bins. The genes hosting integration events were then distributed into the bins and the frequencies compared with matched random controls.

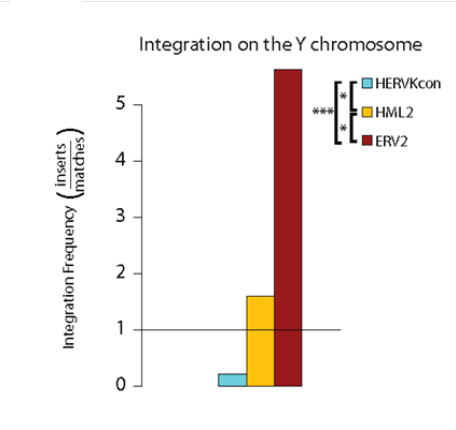


Figure 44 HERV-K integration into the Y chromosome

Values are reported as the proportion of integration events divided by random events. The bar at 1.0 represents the expected random distribution. The statistical significance of differences between data sets is shown by the asterisks next to the legends: (*) 0.05 > P > 0.01; (**) 0.01 > P > 0.001; (***) P < 0.001. Only the HT1080 data set was used in this analysis, as it is derived from a male cell line, while 293T cells are not.

The Y chromosome has been noted previously to contain an extraordinarily high quantity of repeat elements, including an average of 14 HERVs per Mb (Graves, 2006b; Kim et al., 2004b; Kunkel et al., 1976; Skaletsky et al., 2003; Villesen et al., 2004). To test whether the abundance of HERVs in the Y chromosome is caused by a natural integration preference or an accumulation post selection, the de novo insertion of HERV-K_{CON} was examined. Compared to matched random controls, HERV-K_{CON} insertions in Y chromosome were underrepresented but the difference did not reach statistical significance (Figure 44). The HML2(85) data set displayed a one and a half fold preference to insert over random matches, but ERV2 integration sites were enriched on the Y chromosome five fold over random matches, in agreement with studies of other HERVs (Kim et al., 2004b; Villesen et al., 2004). The small number of HERV-K_{CON} insertions in Y chromosome is expected based on its preference for transcription units and CpG islands and the dearth of genes in the Y chromosome. The low number of de novo insertions suggests that the high quantity of resident HERVs in Y chromosome is not due to a natural preference of integration. Y chromosome is a small, gene poor chromosome of 60 Mbs, of which 95% of the sequence does not participate in

homologous recombination, and therefore, unlike other chromosomes, lacks a potential method to eliminate deleterious insertions, which may explain the abundance of HERVs (Graves, 2006b). Furthermore, like the analysis relative to genomic markers, the pattern for the three data sets has a gradient pattern, suggesting that the HML2(85) data set represents a intermediate state between the de novo integrations and the more ancient ERV2 integrations.

Lastly, the orientation of insertions into transcription units was examined. Similar to other HERVs described, ERV2 and HML2(85) were integrated with a heavy bias in the opposite transcriptional orientation relative to the gene (Figure 45).

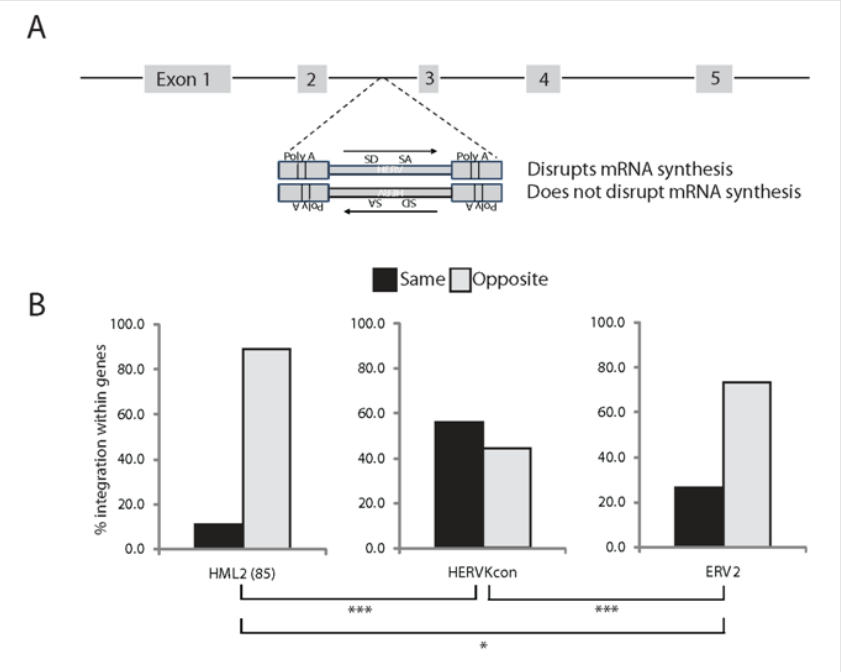


Figure 45 Proviral orientation of de novo HERV-K_{CON} vs resident HERVs

A. Diagram showing proviral orientations and the potential for transcriptional disruption by provirusencoded transcription signals. (SD) Splice donor; (SA) splice acceptor; (PolyA) polyA signal. B. Transcriptional orientation of ERV2, HML2(85), and HERV-K_{CON} sequences found within gene-coding regions as defined by the RefSeq database.

Surprisingly, the bias was stronger for the HML2(85) data set than ERV2 data set. This phenotype may attest to the presence and the stronger influence of splice acceptors and donors and premature polyA signals in the younger group. The ERV2 group contains members with more sequence deterioration, and may include insertions without effective SA and SD sites and polyA signals, hence may be less detrimental to the transcript expression, and thus under less pressure to be negatively selected from the population.

In contrast to HML2(85) and ERV2, HERV-K_{CON} showed no preference in orientation when inserted in a gene. This is perhaps the strongest evidence of a two-step model of HERV-K genomic accumulation. Integration likely occurs according to the actual preference of HERV-K, as determined here using HERV-K_{CON}, then over time individual insertions are negatively selected based on their negative influences on host survival, such as abnormal gene expression due to integration into or near a gene. What remains in the host genome is a result of constant 'cleansing' of detrimental repeat elements at the population level.

IV. DISCUSSION

Endogenous retroviruses are fossils of ancient infections that contain hidden information about host-pathogen interactions, genomic rearrangements, and history of population migration and speciation. The first human endogenous retroviruses were discovered in the early 1980s: a type-C-like endogenous retrovirus similar to MLV and Baboon endogenous retrovirus (BaEV), and a type-D-like endogenous retrovirus similar to MMTV (Callahan et al., 1982; Martin et al., 1981; May et al., 1983). These two HERVs are now known as HERV-E and HERV-K, which belong to the genera gammaretroviruses and betaretroviruses, respectively (Ono, 1986; Steele et al., 1984). Further studies showed that HERV-K is in fact a family of closely related viruses, and is now referred to as the HERV-K superfamily (Franklin et al., 1988). Of the ten subfamilies of HERV-K, the HERV-K(HML-2) subfamily is of particular interest because it is believed to be the youngest of all HERVs known, and thought perhaps to include replication competent members.

This body of work has focused on the replication of HERV-K(HML-2), and its interaction with antiretroviral restriction factors. In Chapter 1, the HERV-K(HML-2) elements in the human genome gathered by BLAST search are presented, with details on their location. Many of these elements have been previously described or characterized (Barbulescu et al., 1999; Belshaw et al., 2005a; Belshaw et al., 2004; Macfarlane and Simmonds, 2004; Romano et al., 2006). Characterization of two proviruses, K113 and K108, is also presented. Chapter 2 describes the derivation of HERV-K_{CON}, a consensus sequence of ten human specific proviruses and the establishment of a single-cycle

infection system using this virus. Known anti-retroviral factors are tested against HERV-K_{CON} in Chapter 3, along with a more extensive analysis of the previous and current interactions between HERV-K_{CON} and APOBEC3G protein. Lastly, where HERV-K_{CON} prefers to integrate is described in Chapter 4, as well as the comparison of de novo integrations with those of endogenous HERV-K elements.

Chapter 1.

In this chapter, a BLAST search of the human genome for endogenous HERV-K insertions using either the LTR sequence or the amino acid sequence of HERV-K_{CON} Gag is presented (Table 1, Table 2). These collections of insertions likely reflect HERV-K infections of hominoid and Old world monkey ancestors after the divergence from New world monkeys. A previous study found 553 HERV-K(HML-2) insertions, but as the identity of the insertions or the criteria for distinguishing HML-2 from other HERV-K subfamilies is not discussed in the previous study, a comparison is not possible (Belshaw et al., 2005a). Similar to other HERVs, 90% of insertions are solo-LTRs and 10% are associated with complete or partial internal sequence (41 of 404 insertions, Table 2). The 404 insertions are integrated across all chromosomes and largely correlate with the size of the chromosome except for chromosome 19, which as more insertions than expected by size, and chromosomes 13, 15, and X, which have less than expected. These exceptions are discussed later.

No naturally occurring replication competent provirus has been identified, and the replication potential of the best candidate, K113, has been disappointing. However, the apparent expression of HERV-K proteins in teratocarcinoma-derived cell lines may yield

some useful information about the nature of the factor or factors that induce transcription from HERV-K LTR and the natural tropism of HERV-K. A recent publication which examined the identity of HERV-K(HML-2) transcripts in Tera-1 teratocarcinoma cell line via sequencing, eight proviruses were identified as major contributors of transcripts encoding either Gag or Env sequences (Ruprecht et al., 2008). Consistent with this study and the finding that HERV-K proteins are expressed in teratocarcinoma cells, K113 LTR was more active in NCCIT teratocarcinoma cell line than 293T cell line (Figure 9). The heterokaryon experiment shows that the factor or factors are dominant and present in the NCCIT cells, but its identity is still unknown. Discovery of this factor or factors may narrow down the list of potential natural target cells, as its expression is likely be a requirement for a productive infection. The low level expression induced in the overexpression of YY2 does not account for the much larger induction in teratocarcinoma cell lines detected by Knossl et al, and hence unlikely to be the sole or the main factor responsible for the upregulation in transcription from HERV-K LTR.

Despite the efforts to induce higher levels of transcription by insertion of a CMVP and a functional Env of K108, CMVP-K113 and CMVP-K113K108 constructs did not produce detectable levels of Gag by Western blotting (data not shown). The expression of Rec in trans enhanced Gag expression, but it was still difficult to detect via Western blotting. This shows that Rec is expressed and that K108 Rec and RcRE of K113 and K108 are functional, but there are likely other defects in the proviruses which results in efficient expression of Gag.

The undetectable level of Gag from a full length genome and the lack of replication of K113 may be due to multiple factors. Aside from the low level of activity

of K113 LTR in 293T cells and the nonfunctional Env, K113 also has a substitution in CA (I516M), which greatly reduces viral particle formation in the context of HERV-K_{CON} Gag (Heslin et al., 2009). Furthermore, K113 RT is inactive, although its activity can be restored by changing six amino acids that differ from RT sequence of HERV-K10 (Beimforde et al., 2008). Hence, despite its complete nucleotide sequence, K113 is defective on multiple levels, and is demonstrably inactive.

In contrast, K108 has become more promising in its replication potential since its discovery. The K108 locus, also known as HERV-K(C7) and HERV-K(HML2-HOM), is composed of two almost identical proviruses that share an internal LTR (Reus et al., 2001). K108 alleles contain three major mutations, a stop codon in Gag, a frameshift in Prot, and a YIDD to CIDD in the highly conserved RT, in various combinations (Mayer et al., 1999a; Reus et al., 2001). Its LTRs differ from each other by six nucleotides, suggesting that it inserted more than a million years ago (Mayer et al., 1999a). For these reasons, K108 was thought to be incapable of replication. However, against expectation, the CIDD mutation does not inactivate RT (Figure 24). The most common allele of K108 lacks the stop codon and frameshift mutations, which are clearly inactivating mutations, and only includes the CIDD RT mutation (Reus et al., 2001). Given that the Env and Rec proteins of K108 are functional (Dewannieux et al., 2005), and Gag-Prot-Pol ORFs are complete in some alleles, K108 may now represent the most likely candidate for a replication competent HERV-K.

Chapter 2. Derivation of HERV-K_{CON} and the single-cycle infection system

Here, a HERV-K provirus whose sequence resembles that of an ancestral human-specific HERV-K(HML-2) was constructed. All viral proteins encoded by this provirus were demonstrated to be capable of functioning in the context of a retroviral replication cycle. While some recent studies have reconstituted "live" viruses from synthetic DNA (Cello et al., 2002; Tumpey et al., 2005), this and a similar study of HERV-K published nearly simultaneously (Dewannieux et al., 2006) are the first examples in which the replication cycle of a virus has been reconstituted using a group of sequences that represent ancient fossils and are demonstrably defective. The methods used here are conceptually similar to those applied to the reconstitution of the transposable element Sleeping Beauty, in which a functional Tc1/mariner-type transposon present only in defective forms in fish DNA was reconstituted (Ivics et al., 1997). Successful reconstitution in that study was achieved using a majority consensus sequence to synthesize an active transposase protein and selecting *cis*-acting sequences from a representative element that closely resembled those of the majority consensus sequence (Ivics et al., 1997).

It was not obvious what the optimal approach to reconstitute functional HERV-K sequences would be, since variation in HERV-K sequence could arise through natural variation via error-prone reverse transcription, mutational degradation after deposition in the primate germ-line, or cytidine deamination before, during, or after during initial germ-line deposition (see below). Moreover, it was possible that the population of proviruses accessible to us in modern DNA represented a highly biased sample of HERV-K genomes where defects might have been positively selected during primate

evolution. Thus, rather than attempt reconstruct the evolutionary history of HERV-K in primates, a conservative approach to reconstitute functional sequences was adopted, selecting ten proviruses that were most similar to K113, reasoning that these were the least likely to have undergone substantial sequence degradation. As described earlier, all of the selected proviruses were unique to human DNA, and some were polymorphic in humans, suggesting comparatively recent replication. While it was possible that all of the selected proviruses would have a common lethal defect, this appeared not to be the case. Indeed, by compiling a simple majority consensus sequence, individual lethal defects represented in the group of proviruses that contribute to the consensus sequence were successfully removed, allowing replication of the consensus genome in a bona fide reverse transcription–dependent manner that resulted in the stable integration of HERV-K_{CON} genomes into target cells.

Assembly of HERV-K virions at the plasma membrane is notable (Figures 19 and 20), given that the exogenous retroviruses that are most closely related to HERV-K include mouse mammary tumor virus and Mason-Pfizer monkey virus, both of which are betaretroviruses that assemble complete capsids within the cytoplasm of infected cells. Nonetheless, previous analyses have suggested that the small number of human cell lines that express HERV-K exhibit plasma membrane localized assembly intermediates (Bieda et al., 2001; Boller et al., 1993b), as was observed here for HERV-K_{CON}. Moreover, previous work has shown that a single amino acid mutation in MPMV Gag protein can change its assembly characteristics from cytoplasmic to plasma membrane associated assembly (Rhee and Hunter, 1990b). Thus, it should not be surprising that HERV-K assembly appears morphologically different to that of its betaretrovirus relatives.

Chapter 3. Restriction factors and HERV-K

In recent years, several gene products with antiretroviral activity have been identified, such as TRIM5, tetherin, and the APOBEC3 family of proteins. Positive selection pressure has been placed on many of these genes during primate evolution (Conticello et al., 2005; McNatt et al., 2009b; Sawyer et al., 2004; Sawyer et al., 2005b; Song et al., 2005). As HERV-K has been repeatedly colonizing the genomes of Old World primates since the divergence of Old and New World monkeys approximately 35 million years ago, it is a potential source of recurrent selective pressure on primate hosts (Bannert and Kurth, 2004).

HERV-K infection was not inhibited by the TRIM5 proteins that were tested (Figure 30). In the case of human TRIM5α, this was not unexpected, because HERV-K_{CON} was derived from human-specific proviruses that must, by definition, have replicated in humans at some point in their evolution and may, therefore, have evolved resistance to human TRIM5α. However, HERV-K_{CON} was also resistant to rhesus monkey TRIM5α and also TRIM-Cyp, a form of TRIM5 that is unique to owl monkeys (Nisole et al., 2004b; Sayah et al., 2004b), a New World monkey species that does not carry HERV-K. At present, therefore, there is no evidence that TRIM5 proteins and HERV-K have exerted reciprocal evolutionary pressure during primate evolution. However, analysis of CA sequences reconstructed from more ancient groups of HERV-K proviruses and inserted into HERV-K_{CON}, as well as inclusion of more TRIM5α variants, may be illuminating. The studies described herein suggest that such approaches to study interactions between ancient retroviruses and their hosts should be feasible.

Unlike TRIM5α, tetherin was able to restrict HERV-K_{CON} virus release (Figure 32). Thus far, tetherin seems capable of restricting release of particles from all viruses tested, including retroviruses from six of seven genera, as well as Marburg and Ebola filoviruses and the arenavirus Lassa virus (Jouvenet et al., 2009a; Neil et al., 2008a; Sakuma et al., 2009). However, most of these assays were conducted using viral structural proteins rather than full length viruses. Hence, any anti-tetherin activity in other parts of these viruses will not be detected. Indeed, some SIVs that do not encode for Vpu seem to use another accessory protein, Nef, to counter the effects of tetherin (Jia et al., 2009; Zhang et al., 2009). It is possible that other enveloped viruses have their own mechanism of counteracting tetherin.

The study of the interaction between APOBEC proteins and HERV-K has been more illuminating. When comparing full-length human-specific HERV-K proviruses to HERV-K_{CON}, an abundance of G-to-A and C-to-T substitutions were found (Figure 34). These substitutions, the most common change found in genomes (Lander et al., 2001), can occur when DNA methyltransferase methylates cytosines in CG dinucleotides to 5-methylcytosine, which spontaneously deaminates to thymine, resulting in a CG-to-TG change. These methylation events, important for development via genomic imprinting and X chromosome inactivation, can also silence retroelements. This effect has been demonstrated in mice, where knocking out DNA methyltransferase Dnmt1 or Dnmt3L leads to transcriptional activation of mouse retroelements intracisternal A particles and LINE-1 (Bourc'his and Bestor, 2004; Walsh et al., 1998). In addition, previous studies of a selection of HERV-K LTRs in the teratocarcinoma cell line Tera-1 showed that methylation and transcription are inversely related (Lavie et al., 2004). Given these two

facts, one would expect to find that HERV-K proviruses would be cytosine methylated by the host, and consequently G-to-A and C-to-T mutations should be abundant. This was indeed the case, and in 14 of 16 HERV-K proviruses examined, the CG dinucleotide methylation/spontaneous deamination pathway appeared to be the major source of G-to-A and C-to-T mutations (Figure 35).

Another common cause of G-to-A, and less frequently C-to-T (Bishop et al., 2004b), changes in viral DNA is APOBEC3-mediated cytidine deamination. Fortuitously, the two events are easily distinguished. DNA methyltransferases methylate cytosines in CG dinucleotides, while APOBEC proteins deaminate cytosines in XC dinucleotides, where X can differ depending on the APOBEC protein that is responsible for deamination. Hence, by examining the nucleotides immediately 5' and 3' to the mutated cytosine, one can largely assign G-to-A and C-to-T changes to either mechanism. One exception is when both DNA methyltransferase and APOBEC preferred nucleotides flank the altered cytosine, such as CCG trinucleotides, where CC represents the dinucleotide preference of hA3G and CG the preference of DNA methyltransferase. However, exclusion of these ambiguous samples in the analyses did not alter the conclusions.

The characteristics of HERV-K hypermutation found both in vivo and in vitro match several of the characteristics previously observed in the context of APOBEC-induced mutations in other retrovirus infections. First, G-to-A mutations constituted a large fraction of the total mutations in HERV-K60 and HERV-KI (Figures 34 and 36), as has previously been found for hypermutated viral sequences in HIV- and HBV-infected patients (Janini et al., 2001; Simon et al., 2005; Suspene et al., 2006; Vartanian et al., 1991). Furthermore, the GGG trinucleotide preference found during in vitro HERV-K

infection in the presence of hA3G has been documented in HIV-1 infection assays by several groups (Figures 36 and 37) (Bishop et al., 2004a; Wiegand et al., 2004; Yu et al., 2003). The combination of these two major characteristics found in HERV-K60 and HERV-KI, plus the failure of any other human APOBEC3 protein to induce a similar pattern of mutation during HERV-K replication in vitro, makes a strong argument for hA3G as the sole source of hypermutation in ancient HERV-K proviruses.

Another reported characteristic of viral hypermutation by hA3G is the gradient of changes along the viral genome. This characteristic is thought to derive from the position-dependent length of time that the nascent viral DNA is in the form of single-stranded DNA, the preferred nucleic substrate of hA3G. However, this was not observed in the HERV-K60 and HERV-KI sequences (data not shown). The reasons for this are unclear at present. Nonetheless, the ability to fairly precisely recreate the hypermutation patterns present in ancient proviruses specifically using hA3G during in vitro HERV-K replication assays suggests that the interaction between this protein and HERV-K occurred and was physiologically and evolutionarily relevant. HERV-K is therefore an eminently reasonable candidate for an infectious agent that has applied selective pressure on A3G during primate evolution. However, it is notable that HERV-K is one of a number of agents that could potentially have imposed selective pressure on antiretroviral defenses. Other abundant endogenous retroelements, such as Alu and LINE-1 elements in humans, have also been shown to be restricted by APOBEC3 proteins (Bogerd et al., 2006b; Chiu et al., 2006; Hulme et al., 2007; Niewiadomska et al., 2007; Stenglein and Harris, 2006), and these elements as well as other exogenous and endogenous retroviruses may also have contributed to the expansion and positive selection that is evident in APOBEC3

genes. Indeed, among the ancient retroviruses, only those that colonized the germ line are accessible to this type of analysis, and it is completely unknown what fraction of ancient retroviruses that replicated in ancestral primates are fossilized in modern DNA. Nonetheless, A3G has been under positive selection for many millions of years (Sawyer et al., 2004), and HERV-K could, potentially, have contributed to this pressure.

Given that HERV-K(HML-2) appears intrinsically mutable by hA3G and a hypermutated provirus is likely to be less harmful than an intact provirus and hence more likely to become fixed in a host genome, it is perhaps surprising that only 2 out of 16 HERV-K human-specific proviruses and none of 4 chimpanzee-specific proviruses were clearly hypermutated. Several factors may have contributed to this, and perhaps the most important influence would be viral tropism. The appearance of a hypermutated provirus in human DNA indicates that HERV-K likely replicated in an A3G-expressing maternal or paternal tissue immediately prior to deposition of the provirus into the germ line. Conversely, the deposition of a nonhypermutated virus suggests that the preceding generations involved replication in APOBEC3G-negative tissues. The simplest explanation for the appearance of hypermutated and nonhypermutated proviruses in the human genome is that HERV-K replicated in both A3G-expressing and nonexpressing somatic cells prior to germ line infection.

Moreover, the frequency of hypermutated proviruses in modern genomes may not reflect the frequency at which hypermutation occurred during ancient infections. Indeed, while hypermutation would generally inactivate a particular provirus, hypermutation itself is unlikely to be always necessary or sufficient to result in fixation of the element, where chance-influenced mechanisms, such as drift or bottlenecking, may play a dominant role

in provirus fixation. Of note, older HERV-K sequences belonging to group O as defined by Romano et al. (Romano et al., 2006) did not exhibit signs of hypermutation compared to HERV-K_{CON}. Moreover, a previous study of endogenous murine leukemia viruses also found that only a minority of proviruses were overtly hypermutated, perhaps for the same aforementioned reasons.

While there was a reasonable qualitative correlation between the appearance of APOBEC3-induced G-to-A mutations and infection inhibition during in vitro HERV-K replication, there was a notable lack of a quantitative correlation between the burden of mutations and the extent to which infection was inhibited. Specifically, hA3A, hA3B, and hA3F caused mutation in a minority of nascent HERV-K reverse transcripts yet inhibited infection to a greater degree than hA3G, which mutated the majority of nascent HERV-K DNA molecules (Figure 38). Since no evidence of hA3A, hA3B, or hA3F hypermutation was found in endogenous proviruses, inhibition of HERV-K infection by these cytidine deaminases may be physiologically irrelevant or might occur primarily via mechanisms that would not leave remnants of the viral encounter with the APOBEC protein, such as inhibition of DNA synthesis or integration (Bishop et al., 2006; Guo et al., 2006; Holmes et al., 2007; Mbisa et al., 2007; Yang et al., 2007).

The lack of strong inhibition of HERV-K infection by hA3G in the single-cycle replication assay should not be overinterpreted as suggesting that hA3G lacks anti-HERV-K activity in vivo. As documented here and elsewhere, A3G appears to have evolved to target GG dinucleotides, especially GGG trinucleotides. This property makes it a particularly efficient mutator of tryptophan codons. G-to-A mutation of tryptophan codons invariably leads to the generation of new stop codons, which would almost always

be lethal to a retrovirus, even if a provirus were successfully established with a relatively low burden of A3G-induced mutation. Importantly, the HERV-K infection assay requires a single cycle of infection by a reporter virus that encodes the commonly used EGFP as the reporter gene. EGFP contains only a single tryptophan, and thus a moderate level of hA3G-induced mutations might not score as strong inhibition during an in vitro single-cycle infection assay but would abolish further rounds of replication in an in vivo spreading infection. Indeed, HERV-K60 and HERV-KI represent clear examples of viral sequences that have been fossilized in the human genome in defective form as a consequence of hA3G-induced hypermutation.

Chapter 4. Integration of HERV-K_{CON}

Here, a study of integration target site selection by HERV-K_{CON} and its relationship to the distribution of fixed HERV sites in the human genome is reported. Sites of HERV-K_{CON} integration were slightly enriched in transcription units, in gene-dense regions, and in a collection of features associated with gene activity (Figure 41). The endogenous HERV-K elements, ERV2 and HML2(85), showed a very different distribution and were enriched outside genes (Figure 43). The HERV-K(HML-2), represented by the HML2(85) data set, showed a pattern intermediate between the older HERV-K superfamily ERV2 sites and the de novo integrated HERV-K_{CON} (Figures 43 and 45). These data support a two-step model for accumulation of fixed HERV elements in the human genome, in which integration targeting preferences dictated the initial placement of integration sites, while subsequent purifying selection eliminated the majority of insertions because they were deleterious to host viability.

Previous studies have reported differing distributions among the distinct HERV element families and investigated the mechanisms that mediate gene disruption upon integration (Mager et al., 1999; Smit, 1999; van de Lagemaat et al., 2006). These studies support the idea that strong splice sites and poly(A) sites within HERV elements can disrupt gene transcription, as has been seen with other genomic parasites (Britten, 1996; Jordan et al., 2003; van de Lagemaat et al., 2006). However, these studies did not identify a distinctive distribution pattern for the most recently integrated HERV sequences. This study used homology searching to form a collection of the most similar, hence youngest, genomic HERV-K(HML-2) elements, and it was by analyzing this collection that the intermediate distribution of HML2(85) sequences between the older ERV2 elements and newly integrated HERV-K_{CON} was detected. In contrast to the relationship to genomic features, the orientation bias within genes is evident and similar for both the HML2(85) and ERV2 data sets. This is consistent with the idea that particularly disruptive proviruses integrated within genes may be removed relatively quickly by purifying selection.

Previous studies of retroviral integration targeting have shown that retroviruses from the same genus tend to share the same targeting patterns, but HERV-K_{CON} appears to be an exception. HERV-K_{CON} is most closely related to exogenous betaretroviruses, and its integrase protein sequence clusters with MMTV, a prototype betaretrovirus rather than integrases from other retroviral genera (data not shown). As discussed above, HERV-K_{CON} integration is more frequent in gene-rich regions and genomic features associated with active transcription, somewhat resembling MLV (Wu et al., 2003). Surprisingly, the reported MMTV distribution is almost perfectly random (Faschinger et al., 2008). The only other data set with such a random distribution is AAV, but AAV is

believed to become integrated at cellular DNA double-strand breaks by the action of cellular DNA repair enzymes (Rutledge and Russell, 1997; Song et al., 2001). MMTV, in contrast, encodes an integrase protein, and MMTV integration events show the usual sequence features associated with retroviral integration. It will be useful to obtain more data on integration site distributions from the betaretrovirus genus to clarify this puzzling observation.

Another previously noted surprising difference between the members of betaretroviridae is the location of assembly. Assembly of related betaretrovirus MPMV takes place at a perinuclear region (Rhee and Hunter, 1987), whereas HERV-K assembly takes place at the plasma membrane. These two phenotypic differences (assembly and integration targeting) within betaretroviridae suggest that the genus may not be monophyletic. Thus, although the integration targeting data for HERV-K_{CON} seems likely to model trends for all of the HERV-K elements, it is uncertain to what extent, if any, the data for HERV-K_{CON} models the other HERV families that most closely resemble exogenous retroviruses of other genera.

The two-step model for HERVs accumulation is likely operating on endogenous retroviruses and other integrating elements of many vertebrates (Bushman, 2001; Han and Boeke, 2005; Kazazian, 2004). In a previous study, Barr et al. compared de novo ASLV integration events in chicken cells to fixed proviruses in the chicken germline that were derived from the same retroviral group (Barr et al., 2005). They found that de novo ASLV integration showed a modest preference for transcription units, while fixed ASLVs in the germline accumulated outside of transcription units. Fixed ASLVs in the germline also showed an orientation bias, so that proviruses within genes tended to accumulate in

opposite transcriptional orientation relative to the host gene, while the de novo integration events showed no such bias. Similar biases in endogenous provirus accumulation have also been observed in mouse and rat (Barr et al., 2005; van de Lagemaat et al., 2006). These findings suggest that purifying selection is operating similarly on the endogenous retroviruses inhabiting the genomes of many vertebrates (Barr et al., 2005; Brookfield, 2005; Cutter et al., 2005; Lowe et al., 2007; Roy-Engel et al., 2005).

Other questions

Is HERV-K still replication competent?

The findings from the HERV-K sequence data, such as the purifying selection on ORFs, supports the idea that HERV-K has been replicating until recent evolutionary time (Belshaw et al., 2005a; Costas, 2001). However, so far all insertions have been identified via manual screening of cell lines, a handful of human samples, or BAC libraries and mining the genomic sequence data, meaning that only a few individuals' genomes have been screened. To identify a replication competent virus, more polymorphic elements need to be identified, as replication competent virus is likely to be among these sequences, if it exists at all. However, currently, the detection method is limited, as mass sequencing of individuals is expensive, and various blotting techniques are less sensitive. Furthermore, even after a wide array of human individuals and populations have been screened, proving de novo infection by HERV-K will be difficult, as it will require identification of an insertion in an offspring that does not exist in either parent. Hence, proof of natural replication of HERV-K in vivo is still currently unlikely.

Which cells did HERV-K infect?

ERVs must infect germ cells or germ cell progenitors to become endogenized, but it is unclear whether these are the natural or rare targets. If cells other than germ cells and progenitors are HERV-K's most common target cells, then the virus may be capable of horizontal transmission to new, unrelated hosts. This idea is based on the concept of purifying selection. If HERV-K infects somatic cells but does not infect new individuals, the infection in the somatic cells will be lost once the host dies, as no effective additional copies of the virus have been generated. In other words, if the somatic infection had not occurred, there would be no difference to the virus population. Without replication, there is no selection to retain the sequence, and the somatic cell tropism will likely be lost over time. However, if upon infection of somatic cells, the virus can transmit to a new host, the propagation and hence selection to retain sequences important for exogenous life cycle may be retained. What these other natural target cells may be is entirely unknown.

Another factor in question in identifying the target cell is the gender of the host. Unlike most somatic cells such as $CD4^+$ T cells of HIV-1, gametogenesis is a gender specific process; thus, HERV infection may be gender specific. It is certain that male germ cells and/or progenitors are infected, as Y chromosome harbors HERV insertions. The insertions found on X chromosome may have occurred in either male or female host, hence it is not possible to rule out female germ cells as potential targets.

The stage of germ cell development at which HERV-K infection occurs is also difficult to identify. The transcriptional activity of HERV-K LTR in teratocarcinoma-derived cell lines, a tumor in the testes which includes pluripotent germ cells, suggests

that HERV-K LTR activity may be linked to earlier parts of germ cell development. Furthermore, study of HERV-K LTR in human and mouse teratocarcinoma cells showed higher LTR activity relative to nonteratocarcinoma cells, which was lost after differentiation of teratocarcinoma cell lines, suggesting that there may be an unknown point in development after which the LTR become inactive (Casau et al., 1999). Also, in HERV-K LTR transgenic mice, the LTR was active in the testes, especially in the undifferentiated spermatocytes (Casau et al., 1999). In human samples, HERV-K(HML-2) elements were expressed at low levels in both the testes and ovary (Seifarth et al., 2005). Collectively, the LTR seems to be more active in the earlier male germ cell development stages than later, and suggests that the virus may be more replication competent during this time.

Can HIV-1 become endogenized?

As a retrovirus, it is possible that HIV-1 can also become endogenized by either infecting developing or mature gametes. In oocytes, in vitro incubation of oocytes and cell free HIV-1 does not result in infection; in vivo infection is unknown (Baccetti et al., 1999). In males, HIV-1 is found in semen soon after infection (Tindall et al., 1992). Furthermore, virus DNA can be detected in sperm at various stages of development (spermatogonia, spermatids, and spermatocytes) from testes of HIV-1 infected men via PCR in situ hybridization and PCR (Muciaccia et al., 1998; Nuovo et al., 1994). In animal models, viral RNA and proteins are found in testes and epipididymis of pig-tail macaques (*Macaca nemestrina*) infected with SIV_{mac251}or $SHIV_{mn229}$, suggesting that these locations of spermatogensis contain productively infected cells (Shehu-Xhilaga et

al., 2007). A caveat to these studies is the potential of presence by cell-free virus or infected lymphocytes. HIV-1 may also infect spermatozoa in the semen, where cell-free virus is found (Krieger et al., 1991). Spermatozoa purified from other cells and semen fluid have been shown to contain HIV-1 DNA, but no direct evidence of a provirus has been shown (Muciaccia et al., 2007) [reviewed in (Cardona-Maya et al., 2006)]. Hence, infection of spermatozoa and progenitors is still inconclusive.

Another problem of endogenization of HIV-1 is the lack of CD4, the HIV-1 Env receptor, in both gametes (Baccetti et al., 1999; Gil et al., 1995). As an alternate, mannose receptor CD206 has been identified from solubilized spermatozoa proteins as binding partners of cell free HIV-1 virus particles or HIV-1 Env protein, and proposed to act as the receptor on spermatozoa in the absence of CD4 (Bandivdekar et al., 2003; Fanibunda et al., 2008). Mannose receptor has already been shown to mediate HIV-1 Env binding in astrocytes, macrophages, and dendritic cells, although in the latter two cell types, this binding is thought to mediate virus transfer to $CD4^+$ T cells, rather than inducing infection in macrophages and DCs (Liu et al., 2004; Nguyen and Hildreth, 2003; Turville et al., 2001). In astrocytes, the interaction between HIV-1 Env and mannose receptor is thought to enable infection (Liu et al., 2004). However, there is no direct evidence of spermatozoa infection via the mannose receptor. Hence, thus far, there is no direct evidence that HIV-1 will become endogenized.

V. REFERENCES

Akopov, S.B., Nikolaev, L.G., Khil, P.P., Lebedev, Y.B., and Sverdlov, E.D. (1998). Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. FEBS letters *421*, 229-233.

Alkhatib, G., Combadiere, C., Broder, C.C., Feng, Y., Kennedy, P.E., Murphy, P.M., and Berger, E.A. (1996). CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science *272*, 1955-1958.

Anderson, J.L., Campbell, E.M., Wu, X., Vandegraaff, N., Engelman, A., and Hope, T.J. (2006). Proteasome inhibition reveals that a functional preintegration complex intermediate can be generated during restriction by diverse TRIM5 proteins. Journal of virology *80*, 9754-9760.

Andersson, M.L., Lindeskog, M., Medstrand, P., Westley, B., May, F., and Blomberg, J. (1999). Diversity of human endogenous retrovirus class II-like sequences. J Gen Virol *80*, 255-260.

Asaoka, K., Ikeda, K., Hishinuma, T., Horie-Inoue, K., Takeda, S., and Inoue, S. (2005). A retrovirus restriction factor TRIM5alpha is transcriptionally regulated by interferons. Biochemical and biophysical research communications *338*, 1950-1956.

Axel, R., Schlom, J., and Spiegelman, S. (1972). Presence in Human Breast Cancer of RNA homologous to Mouse Mammary Tumour Virus RNA. Nature *235*, 32-36.

Baccetti, B., Benedetto, A., Collodel, G., Crisa, N., di Caro, A., Garbuglia, A.R., and Piomboni, P. (1999). Failure of HIV-1 to Infect Human Oocytes Directly. Journal of Acquired Deficiency Syndromes *21*, 355-361.

Bainbridge, J.W., Stephens, C., Parsley, K., Demaison, C., Halfyard, A., Thrasher, A.J., and Ali, R.R. (2001). In vivo gene transfer to the mouse eye using an HIV-based lentiviral vector; efficient long-term transduction of corneal endothelium and retinal pigment epithelium. Gene therapy *8*, 1665-1668.

Baldo, A.M., and McClure, M.A. (1999). Evolution and Horizontal Transfer of dUTPase-Encoding Genes in Viruses and Their Hosts. J Virol *73*, 7710-7721.

Baltimore, D. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. Nature *226*, 1209-1211.

Bandivdekar, A.H., Velhal, S.M., and Raghavan, V.P. (2003). Identification of CD4-independent HIV receptors on spermatozoa. Am J Reprod Immunol *50*, 322-327.

Bannert, N., and Kurth, R. (2004). Retroelements and the human genome: New perspectives on an old relation. Proceedings of the National Academy of Sciences *101*, 14572-14579.

Barbulescu, M., Turner, G., Seaman, M.I., Deinard, A.S., Kidd, K.K., and Lenz, J. (1999). Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. Current Biology *9*, 861.

Barbulescu, M., Turner, G., Su, M., Kim, R., Jensen-Seaman, M.I., Deinard, A.S., Kidd, K.K., and Lenz, J. (2001). A HERV-K provirus in chimpanzees, bonobos and gorillas, but not humans. Current Biology *11*, 779.

Barr, S.D., Ciuffi, A., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F.D. (2006). HIV Integration Site Selection: Targeting in Macrophages and the Effects of Different Routes of Viral Entry. Mol Ther *14*, 218-225.

Barr, S.D., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F.D. (2005). Integration Targeting by Avian Sarcoma-Leukosis Virus and Human Immunodeficiency Virus in the Chicken Genome. J Virol *79*, 12035-12044.

Barr, S.D., Smiley, J.R., and Bushman, F.D. (2008). The Interferon Response Inhibits HIV Particle Production by Induction of TRIM22. PLoS Pathog *4*, e1000007.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., *et al.* (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science *220*, 868-871.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Beimforde, N., Hanke, K., Ammar, I., Kurth, R., and Bannert, N. (2008). Molecular cloning and functional characterization of the human endogenous retrovirus K113. Virology *371*, 216-225.

Belshaw, R., Dawson, A.L.A., Woolven-Allen, J., Redding, J., Burt, A., and Tristem, M. (2005a). Genomewide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K(HML2): Implications for Present-Day Activity. J Virol *79*, 12507-12514.

Belshaw, R., Katzourakis, A., Paces, J., Burt, A., and Tristem, M. (2005b). High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. Molecular biology and evolution *22*, 814-817.

Belshaw, R., Katzourakis, A., Paces, J., Burt, A., and Tristem, M. (2005c). High Copy Number in Human Endogenous Retrovirus Families is Associated with Copying Mechanisms in Addition to Reinfection. Mol Biol Evol *22*, 814-817.

Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A., and Tristem, M. (2004). Long-term reinfection of the human genome by endogenous retroviruses. PNAS *101*, 4894-4899.

Belshaw, R., Watson, J., Katzourakis, A., Howe, A., Woolven-Allen, J., Burt, A., and Tristem, M. (2007). Rate of Recombinational Deletion among Human Endogenous Retroviruses. J Virol *81*, 9437-9442.

Benit, L., De Parseval, N., Casella, J.F., Callebaut, I., Cordonnier, A., and Heidmann, T. (1997). Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. Journal of virology *71*, 5652-5657.

Berkhout, B., Jebbink, M., and Zsiros, J. (1999). Identification of an Active Reverse Transcriptase Enzyme Encoded by a Human Endogenous HERV-K Retrovirus. J Virol *73*, 2365-2375.

Berry, C., Hannenhalli, S., Leipzig, J., and Bushman, F.D. (2006). Selection of target sites for mobile DNA integration in the human genome. PLoS computational biology *2*, e157.

Besnier, C., Ylinen, L., Strange, B., Lister, A., Takeuchi, Y., Goff, S.P., and Towers, G.J. (2003). Characterization of murine leukemia virus restriction in mammals. Journal of virology *77*, 13403-13406.

Best, S., Le Tissier, P., Towers, G., and Stoye, J.P. (1996). Positional cloning of the mouse retrovirus restriction gene Fv1. Nature *382*, 826-829.
Bieda, K., Hoffmann, A., and Boller, K. (2001). Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. J Gen Virol *82*, 591-596.
Bieniasz, P.D. (2004a). Intrinsic immunity: a front-line defense against viral attack. Nature immunology *5*, 1109-1115.
Bieniasz, P.D. (2004b). Intrinsic immunity: a front-line defense against viral attack. Nat Immunol *5*, 1109-1115.
Bieniasz, P.D. (2006). Late budding domains and host proteins in enveloped virus release. Virology *344*, 55-63.
Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell *40*, 91-99.
Bishop, K.N., Holmes, R.K., and Malim, M.H. (2006). Antiviral Potency of APOBEC Proteins Does Not Correlate with Cytidine Deamination. J Virol *80*, 8450-8458.
Bishop, K.N., Holmes, R.K., Sheehy, A.M., Davidson, N.O., Cho, S.-J., and Malim, M.H. (2004a). Cytidine Deamination of Retroviral DNA by Diverse APOBEC Proteins. Current Biology *14*, 1392.
Bishop, K.N., Holmes, R.K., Sheehy, A.M., and Malim, M.H. (2004b). APOBEC-mediated editing of viral RNA. Science *305*, 645.
Bittner, J.J. (1936). Some Possible Effects of Nursing on the Mammary Gland Tumor Incidence in Mice. Science *84*, 162.
Bittner, J.J. (1942). The Milk-Influence of Breast Tumors in Mice. Science *95*, 462-463.
Blaise, S., de Parseval, N., Benit, L., and Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proceedings of the National Academy of Sciences of the United States of America *100*, 13013-13018.
Blomberg, J., Ushameckis, D., and Jern, P. (2005). Evolutionary Aspects of Human Endogenous Retroviral Sequences (HERVs) and Disease. Retroviruses and Primate Genome Evolution, 204 - 238.
Blond, J.L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F., and Cosset, F.L. (2000). An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. Journal of virology *74*, 3321-3329.
Blot, V., Perugi, F., Gay, B., Prevost, M.C., Briant, L., Tangy, F., Abriel, H., Staub, O., Dokhelar, M.C., and Pique, C. (2004). Nedd4.1-mediated ubiquitination and subsequent recruitment of Tsg101 ensure HTLV-1 Gag trafficking towards the multivesicular body pathway prior to virus budding. Journal of cell science *117*, 2357-2367.
Boese, A., Galli, U., Geyer, M., Sauter, M., and Mueller-Lantzsch, N. (2001). The Rev/Rex homolog HERV-K cORF multimerizes via a C-terminal domain. FEBS letters *493*, 117-121.
Boese, A., Sauter, M., and Mueller-Lantzsch, N. (2000). A Rev-like NES mediates cytoplasmic localization of HERV-K cORF. FEBS Letters *468*, 65-67.

Bogerd, H.P., and Cullen, B.R. (2008). Single-stranded RNA facilitates nucleocapsid: APOBEC3G complex formation. RNA (New York, NY *14*, 1228-1236.
Bogerd, H.P., Echarri, A., Ross, T.M., and Cullen, B.R. (1998). Inhibition of human immunodeficiency virus Rev and human T-cell leukemia virus Rex function, but not Mason-Pfizer monkey virus constitutive transport element activity, by a mutant human nucleoporin targeted to Crm1. Journal of virology *72*, 8627-8635.
Bogerd, H.P., Wiegand, H.L., Doehle, B.P., Lueders, K.K., and Cullen, B.R. (2006a). APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic acids research *34*, 89-95.
Bogerd, H.P., Wiegand, H.L., Hulme, A.E., Garcia-Perez, J.L., O'Shea, K.S., Moran, J.V., and Cullen, B.R. (2006b). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. Proceedings of the National Academy of Sciences of the United States of America *103*, 8780-8785.
Boller, K., Frank, H., Lower, J., Lower, R., and Kurth, R. (1983). Structural Organization of Unique Retrovirus-like Particles Budding from Human Teratocarcinoma Cell Lines. J Gen Virol *64*, 2549-2559.
Boller, K., Janssen, O., Schuldes, H., Tonjes, R.R., and Kurth, R. (1997). Characterization of the antibody response specific for the human endogenous retrovirus HTDV/HERV-K. J Virol *71*, 4581-4588.
Boller, K., Konig, H., Sauter, M., Mueller-Lantzsch, N., Lower, R., Lower, J., and Kurth, R. (1993a). Evidence That HERV-K Is the Endogenous Retrovirus Sequence That Codes for the Human Teratocarcinoma-Derived Retrovirus HTDV. Virology *196*, 349.
Boller, K., König, H., Sauter, M., Mueller-Lantzsch, N., Löwer, R., Löwer, J., and Kurth, R. (1993b). Evidence That HERV-K Is the Endogenous Retrovirus Sequence That Codes for the Human Teratocarcinoma-Derived Retrovirus HTDV. Virology *196*, 349-353.
Bonnaud, B., Bouton, O., Oriol, G., Cheynet, V., Duret, L., and Mallet, F. (2004). Evidence of selection on the domesticated ERVWE1 env retroviral element involved in placentation. Molecular biology and evolution *21*, 1895-1901.
Borman, A.M., Quillent, C., Charneau, P., Kean, K.M., and Clavel, F. (1995). A Highly Defective HIV-1 Group O Provirus: Evidence for the Role of Local Sequence Determinants in G --> A Hypermutation during Negative-Strand Viral DNA Synthesis. Virology *208*, 601.
Bouazzaoui, A., Kreutz, M., Eisert, V., Dinauer, N., Heinzelmann, A., Hallenberger, S., Strayle, J., Walker, R., Rübsamen-Waigmann, H., Andreesen, R., *et al.* (2006). Stimulated trans-acting factor of 50 kDa (Staf50) inhibits HIV-1 replication in human monocyte-derived macrophages. Virology *356*, 79-94.
Bour, S., and Strebel, K. (1996). The human immunodeficiency virus (HIV) type 2 envelope protein is a functional complement to HIV type 1 Vpu that enhances particle release of heterologous retroviruses. J Virol *70*, 8285-8300.
Bourc'his, D., and Bestor, T.H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature *431*, 96-99.
Bowerman, B., Brown, P.O., Bishop, J.M., and Varmus, H.E. (1989). A nucleoprotein complex mediates the integration of retroviral DNA. Genes & development *3*, 469-478.

Braun, I.C., Herold, A., Rode, M., Conti, E., and Izaurralde, E. (2001). Overexpression of TAP/p15 heterodimers bypasses nuclear retention and stimulates nuclear mRNA export. The Journal of biological chemistry *276*, 20536-20543.

Bray, M., Prasad, S., Dubay, J.W., Hunter, E., Jeang, K.T., Rekosh, D., and Hammarskjold, M.L. (1994). A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. Proceedings of the National Academy of Sciences of the United States of America *91*, 1256-1260.

Brennan, G., Kozyrev, Y., and Hu, S.L. (2008). TRIMCyp expression in Old World primates Macaca nemestrina and Macaca fascicularis. Proceedings of the National Academy of Sciences of the United States of America *105*, 3569-3574.

Britten, R.J. (1996). DNA sequence insertion and evolutionary variation in gene regulation. Proc Natl Acad Sci U S A *93*, 9374-9377.

Bronson, D.L., Fraley, E.E., Fogh, J., and Kalter, S.S. (1979). Induction of retrovirus particles in human testicular tumor (Tera-1) cell cultures: an electron microscopy study. Journal of National Cancer Institute *63*, 337-339.

Bronson, D.L., Ritzi, D.M., Fraley, E.E., and Dalton, A.J. (1978). Morphological evidence for retrovirus production by epithelial cells derived from a human testicular tumor metastasis. Journal of National Cancer Institute *60*, 1305-1308.

Brookfield, J.F. (2005). The ecology of the genome - mobile DNA elements and their hosts. Nature reviews *6*, 128-136.

Brown, P.O., Bowerman, B., Varmus, H.E., and Bishop, J.M. (1989). Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proceedings of the National Academy of Sciences of the United States of America *86*, 2525-2529.

Browne, E.P., Allers, C., and Landau, N.R. (2009). Restriction of HIV-1 by APOBEC3G is cytidine deaminase-dependent. Virology *387*, 313-321.

Bukrinsky, M.I., Sharova, N., McDonald, T.L., Pushkarskaya, T., Tarpley, W.G., and Stevenson, M. (1993). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. Proceedings of the National Academy of Sciences of the United States of America *90*, 6125-6129.

Burmeister, T., Ebert, A.D., Pritze, W., Loddenkemper, C., Schwartz, S., and Thiel, E. (2004). Insertional polymorphisms of endogenous HERV-K113 and HERV-K115 retroviruses in breast cancer patients and age-matched controls. AIDS research and human retroviruses *20*, 1223-1229.

Bushman, F.D. (2001). Lateral DNA transfer: Mechanisms and consequences (Cold Spring Harbor, NY., Cold Spring Harbor Laboratory Press).

Busschots, K., Vercammen, J., Emiliani, S., Benarous, R., Engelborghs, Y., Christ, F., and Debyser, Z. (2005). The Interaction of LEDGF/p75 with Integrase Is Lentivirus-specific and Promotes DNA Binding. J Biol Chem *280*, 17841-17847.

Callahan, R., Chiu, I.-M., Wong, J.F.H., Tronick, S.R., Roe, B.A., Aaronson, S.A., and Schlom, J. (1985). A New Class of Endogenous Human Retroviral Genomes. Science *228*, 1208-1211.

Callahan, R., Drohan, W., Tronick, S., and Schlom, J. (1982). Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. PNAS *79*, 5503-5507.

Campbell, S., and Vogt, V.M. (1995). Self-assembly in vitro of purified CA-NC proteins from Rous sarcoma virus and human immunodeficiency virus type 1. Journal of virology *69*, 6487-6497.

Cann, A.J., Rosenblatt, J.D., Wachsman, W., Shah, N.P., and Chen, I.S. (1985). Identification of the gene responsible for human T-cell leukaemia virus transcriptional regulation. Nature *318*, 571-574.

Cardona-Maya, W., Lopez-Herrera, A., Nelilla-Hernadez, P., Rugeles, M.T., and Cadavid, A.P. (2006). The Role of Mannose Receptor on HIV-1 Entry into Human Spermatozoa. American Journal of Reproductive Immunology *55*, 241-245.

Carriere, C., Gay, B., Chazal, N., Morin, N., and Boulanger, P. (1995). Sequence requirements for encapsidation of deletion mutants and chimeras of human immunodeficiency virus type 1 Gag precursor into retrovirus-like particles. Journal of virology *69*, 2366-2377.

Casau, A.E., Vaughan, J.E., Lozano, G., and Levine, A.J. (1999). Germ Cell Expression of an Isolated Human Endogenous Retroviral Long Terminal Repeat of the HERV-K/HTDV Family in Transgenic Mice. J Virol *73*, 9976-9983.

CDC (1981). Pneumocystis Pneumonia - Los Angeles (CDC), pp. 1-3.

Cello, J., Paul, A.V., and Wimmer, E. (2002). Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science *297*, 1016-1018.

Chang, D.D., and Sharp, P.A. (1989). Regulation by HIV Rev depends upon recognition of splice sites. Cell *59*, 789-795.

Charney, J., and Moore, D.H. (1971). Biological Sciences: Neutralization of Murine Mammary Tumour Virus by Sera of Women with Breast Cancer. Nature *229*, 627-628.

Chen, H., Lilley, C.E., Yu, Q., Lee, D.V., Chou, J., Narvaiza, I., Landau, N.R., and Weitzman, M.D. (2006). APOBEC3A Is a Potent Inhibitor of Adeno-Associated Virus and Retrotransposons. Current Biology *16*, 480.

Cherepanov, P. (2007). LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro. Nucl Acids Res *35*, 113-124.

Cherepanov, P., Maertens, G., Proost, P., Devreese, B., Van Beeumen, J., Engelborghs, Y., De Clercq, E., and Debyser, Z. (2003). HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells. J Biol Chem *278*, 372-381.

Chiu, Y.L., and Greene, W.C. (2006). Multifaceted antiviral actions of APOBEC3 cytidine deaminases. Trends in immunology *27*, 291-297.

Chiu, Y.L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T., and Greene, W.C. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proceedings of the National Academy of Sciences of the United States of America *103*, 15588-15593.

Choe, H., Farzan, M., Sun, Y., Sullivan, N., Rollins, B., Ponath, P.D., Wu, L., Mackay, C.R., LaRosa, G., Newman, W., *et al.* (1996). The beta-chemokine receptors

CCR3 and CCR5 facilitate infection by primary HIV-1 isolates. Cell *85*, 1135-1148.
Chopra, H.C., and Mason, M.M. (1970). A new virus in a spontaneous mammary tumor of a rhesus monkey. Cancer research *30*, 2081-2086.
Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R., and Bushman, F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. Nat Med *11*, 1287-1289.
Conticello, S.G., Thomas, C.J.F., Petersen-Mahrt, S.K., and Neuberger, M.S. (2005). Evolution of the AID/APOBEC Family of Polynucleotide (Deoxy)cytidine Deaminases. Mol Biol Evol *22*, 367-377.
Cordonnier, A., Casella, J.F., and Heidmann, T. (1995). Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence. J Virol *69*, 5890-5897.
Costas, J. (2001). Evolutionary Dynamics of the Human Endogenous Retrovirus Family HERV-K Inferred from Full-Length Proviral Genomes. Journal of Molecular Evolution *53*, 237-243.
Costas, J. (2002). Characterization of the Intragenomic Spread of the Human Endogenous Retrovirus Family HERV-W. Mol Biol Evol *19*, 526-533.
Cowan, S., Hatziioannou, T., Cunningham, T., Muesing, M.A., Gottlinger, H.G., and Bieniasz, P.D. (2002). Cellular inhibitors with Fv1-like activity restrict human and simian immunodeficiency virus tropism. Proceedings of the National Academy of Sciences of the United States of America *99*, 11914-11919.
Cutter, A.D., Good, J.M., Pappas, C.T., Saunders, M.A., Starrett, D.M., and Wheeler, T.J. (2005). Transposable element orientation bias in the Drosophila melanogaster genome. Journal of molecular evolution *61*, 733-741.
D'Souza, V., and Summers, M.F. (2005). How retroviruses select their genomes. Nature reviews *3*, 643-655.
Dalgleish, A.G., Beverley, P.C., Clapham, P.R., Crawford, D.H., Greaves, M.F., and Weiss, R.A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature *312*, 763-767.
Dang, Y., Siew, L.M., Wang, X., Han, Y., Lampen, R., and Zheng, Y.-H. (2008). Human cytidine deaminase APOBEC3H restricts HIV-1 replication. J Biol Chem, M707586200.
Dang, Y., Wang, X., Esselman, W.J., and Zheng, Y.H. (2006). Identification of APOBEC3DE as another antiretroviral factor from the human APOBEC family. Journal of virology *80*, 10522-10533.
Deen, K.C., and Sweet, R.W. (1986). Murine mammary tumor virus pol-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus pol gene. Journal of Virology *57*, 422-432.
Delebecque, F., Suspene, R., Calattini, S., Casartelli, N., Saib, A., Froment, A., Wain-Hobson, S., Gessain, A., Vartanian, J.P., and Schwartz, O. (2006). Restriction of foamy viruses by APOBEC cytidine deaminases. Journal of virology *80*, 605-614.
Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., Di Marzio, P., Marmon, S., Sutton, R.E., Hill, C.M., *et al.* (1996). Identification of a major co-receptor for primary isolates of HIV-1. Nature *381*, 661-666.

Derse, D., Crise, B., Li, Y., Princler, G., Lum, N., Stewart, C., McGrath, C.F., Hughes, S.H., Munroe, D.J., and Wu, X. (2007). Human T-Cell Leukemia Virus Type 1 Integration Target Sites in the Human Genome: Comparison with Those of Other Retroviruses. J Virol *81*, 6731-6741.

Derse, D., Hill, S.A., Lloyd, P.A., Chung, H.-k., and Morse, B.A. (2001). Examining Human T-Lymphotropic Virus Type 1 Infection and Replication by Cell-Free Infection with Recombinant Virus Vectors. J Virol *75*, 8461-8468.

DesGroseillers, L., and Jolicoeur, P. (1983). Physical mapping of the Fv-1 tropism host range determinant of BALB/c murine leukemia viruses. Journal of virology *48*, 685-696.

Desrosiers, R.C. (2007). Nonhuman lentiviruses. In Field's Virology, D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, and S.E. Straus, eds. (Philadelphia, Lippincott Williams & Wilkins), pp. 2215-2243.

Dewannieux, M., Blaise, S., and Heidmann, T. (2005). Identification of a Functional Envelope Protein from the HERV-K Family of Human Endogenous Retroviruses. J Virol *79*, 15573-15577.

Dewannieux, M., Harper, F., Richaud, A.l., Letzelter, C., Ribet, D., Pierron, G.r., and Heidmann, T. (2006). Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Research *16*, 1548-1556.

Diaz-Griffero, F., Vandegraaff, N., Li, Y., McGee-Estrada, K., Stremlau, M., Welikala, S., Si, Z., Engelman, A., and Sodroski, J. (2006). Requirements for capsid-binding and an effector function in TRIMCyp-mediated restriction of HIV-1. Virology *351*, 404-419.

Doranz, B.J., Rucker, J., Yi, Y., Smyth, R.J., Samson, M., Peiper, S.C., Parmentier, M., Collman, R.G., and Doms, R.W. (1996). A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. Cell *85*, 1149-1158.

Dunham, A., Matthews, L.H., Burton, J., Ashurst, J.L., Howe, K.L., Ashcroft, K.J., Beare, D.M., Burford, D.C., Hunt, S.E., Griffiths-Jones, S., *et al.* (2004). The DNA sequence and analysis of human chromosome 13. Nature *428*, 522-528.

Dunlap, K.A., Palmarini, M., Varela, M., Burghardt, R.C., Hayashi, K., Farmer, J.L., and Spencer, T.E. (2006). Endogenous retroviruses regulate periimplantation placental growth and differentiation. Proceedings of the National Academy of Sciences of the United States of America *103*, 14390-14395.

Dunn, C.A., Medstrand, P., and Mager, D.L. (2003). An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. Proc Natl Acad Sci USA *100*, 12841 - 12846.

Dupressoir, A., Marceau, G., Vernochet, C., Benit, L., Kanellopoulos, C., Sapin, V., and Heidmann, T. (2005). Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. Proceedings of the National Academy of Sciences of the United States of America *102*, 725-730.

Emerman, M., and Malim, M.H. (1998). HIV-1 Regulatory/Accessory Genes: Keys to Unraveling Viral and Host Cell Biology. Science *280*, 1880-1884.

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636-640.
Engelman, A., Mizuuchi, K., and Craigie, R. (1991). HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer. Cell *67*, 1211-1221.
Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A.J., Heidmann, T., and Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. Nature *433*, 430-433.
Esnault, C., Millet, J., Schwartz, O., and Heidmann, T. (2006). Dual inhibitory effects of APOBEC family proteins on retrotransposition of mammalian endogenous retroviruses. Nucleic acids research *34*, 1522-1531.
Fanibunda, S.E., Velhal, S.M., Raghavan, V.P., and Bandivdekar, A.H. (2008). CD4 independent binding of HIV gp120 to mannose receptor on human spermatozoa. Journal of acquired immune deficiency syndromes (1999) *48*, 389-397.
Faschinger, A., Rouault, F., Sollner, J., Lukas, A., Salmons, B., Gunzburg, W.H., and Indik, S. (2008). Mouse Mammary Tumor Virus Integration Site Selection in Human and Mouse Genomes. J Virol *82*, 1360-1367.
Fassati, A., and Goff, S.P. (1999). Characterization of intracellular reverse transcription complexes of Moloney murine leukemia virus. Journal of virology *73*, 8919-8925.
Fassati, A., and Goff, S.P. (2001). Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. Journal of virology *75*, 3626-3635.
Felber, B.K., Hadzopoulou-Cladaras, M., Cladaras, C., Copeland, T., and Pavlakis, G.N. (1989). rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. Proceedings of the National Academy of Sciences of the United States of America *86*, 1495-1499.
Felber, B.K., Paskalis, H., Kleinman-Ewing, C., Wong-Staal, F., and Pavlakis, G.N. (1985). The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. Science (New York, NY *229*, 675-679.
Feng, Y., Broder, C.C., Kennedy, P.E., and Berger, E.A. (1996). HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science *272*, 872-877.
Flint, S.J., Enquist, L.W., Racaniello, V.R., and Skalka, A.M. (2004). Principles of virology, 2nd Edition edn (Washington, DC, ASM Press).
Forshey, B.M., von Schwedler, U., Sundquist, W.I., and Aiken, C. (2002). Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. Journal of virology *76*, 5667-5677.
Franklin, G.C., Chretien, S., Hanson, I.M., Rochefort, H., May, F.E., and Westley, B.R. (1988). Expression of human sequences related to those of mouse mammary tumor virus. Journal of Virology *62*, 1203-1210.
Fujiwara, T., and Mizuuchi, K. (1988). Retroviral DNA integration: structure of an integration intermediate. Cell *54*, 497-504.
Gabuzda, D.H., Lawrence, K., Langhoff, E., Terwilliger, E., Dorfman, T., Haseltine, W.A., and Sodroski, J. (1992). Role of vif in replication of human immunodeficiency virus type 1 in CD4+ T lymphocytes. Journal of virology *66*, 6489-6495.

Gardner, M.B., Rasheed, S., Pal, B.K., Estes, J.D., and O'Brien, S.J. (1980). Akvr-1, a dominant murine leukemia virus restriction gene, is polymorphic in leukemia-prone wild mice. Proceedings of the National Academy of Sciences of the United States of America *77*, 531-535.

Garrus, J.E., von Schwedler, U.K., Pornillos, O.W., Morham, S.G., Zavitz, K.H., Wang, H.E., Wettstein, D.A., Stray, K.M., Cote, M., Rich, R.L., *et al.* (2001). Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell *107*, 55-65.

Gifford, R., and Tristem, M. (2003). The evolution, distribution and diversity of endogenous retroviruses. Virus Genes *26*, 291-315.

Gifford, R.J., Katzourakis, A., Tristem, M., Pybus, O.G., Winters, M., and Shafer, R.W. (2008). A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. Proceedings of the National Academy of Sciences *105*, 20362-20367.

Gil, T., Castilla, J.A., Hortas, M.L., Molina, J., Redondo, M., Samaniego, F., Garrido, F., Vergara, F., and Herruzo, A. (1995). CD4+ cells in human ejaculates. Human reproduction (Oxford, England) *10*, 2923-2927.

Gilbert, C.m., Maxfield, D.G., Goodman, S.M., and Feschotte, C.d. (2009). Parallel Germline Infiltration of a Lentivirus in Two Malagasy Lemurs. PLoS Genet *5*, e1000425.

Gilboa, E., Goff, S., Shields, A., Yoshimura, F., Mitra, S., and Baltimore, D. (1979a). In vitro synthesis of a 9 kbp terminally redundant DNA carrying the infectivity of Moloney murine leukemia virus. Cell *16*, 863-874.

Gilboa, E., Mitra, S.W., Goff, S., and Baltimore, D. (1979b). A detailed model of reverse transcription and tests of crucial aspects. Cell *18*, 93-100.

Goff, S.P. (2004). Retrovirus restriction factors. Molecular cell *16*, 849-859.

Goff, S.P. (2007). Retroviridae: The Retroviruses and Their Replication. In Field's Virology, D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, and S.E. Straus, eds. (Philadelphia, Lippincott Williams & Wilkins), pp. 1999-2069.

Goff, S.P., Gilboa, E., Witte, O.N., and Baltimore, D. (1980). Structure of the Abelson murine leukemia virus genome and the homologous cellular gene: studies with cloned viral DNA. Cell *22*, 777-785.

Gongora, C., Tissot, C., Cerdan, C., and Mechti, N. (2000). The Interferon-Inducible Staf50 Gene Is Downregulated During T Cell Costimulation by CD2 and CD28. Journal of Interferon & Cytokine Research *20*, 955-961.

Goto, T., Kennel, S.J., Abe, M., Takishita, M., Kosaka, M., Solomon, A., and Saito, S. (1994). A novel membrane antigen selectively expressed on terminally differentiated human B cells. Blood *84*, 1922-1930.

Gottlinger, H.G., Dorfman, T., Cohen, E.A., and Haseltine, W.A. (1993). Vpu protein of human immunodeficiency virus type 1 enhances the release of capsids produced by gag gene constructs of widely divergent retroviruses. Proceedings of the National Academy of Sciences of the United States of America *90*, 7381-7385.

Gottlinger, H.G., Dorfman, T., Sodroski, J.G., and Haseltine, W.A. (1991). Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle

release. Proceedings of the National Academy of Sciences of the United States of America *88*, 3195-3199.
Gottwein, E., Bodem, J., Muller, B., Schmechel, A., Zentgraf, H., and Krausslich, H.-G. (2003). The Mason-Pfizer Monkey Virus PPPY and PSAP Motifs Both Contribute to Virus Release. J Virol *77*, 9474-9485.
Graur, D., and Li, W.-H. (2000). Fundamentals of Molecular Evolution, 2nd Edition edn (Sunderland, MA, Sinauer Associates, Inc.).
Graves, J.A.M. (2006a). Sex Chromosome Specialization and Degeneration in Mammals. *124*, 901-914.
Graves, J.A.M. (2006b). Sex Chromosome Specialization and Degeneration in Mammals. Cell *124*, 901-914.
Grimwood, J., Gordon, L.A., Olsen, A., Terry, A., Schmutz, J., Lamerdin, J., Hellsten, U., Goodstein, D., Couronne, O., Tran-Gyamfi, M., *et al.* (2004). The DNA sequence and biology of human chromosome 19. Nature *428*, 529-535.
Gross, L. (1951). "Spontaneous" leukemia developing in C3H mice following inoculation in infancy, with AK-leukemic extracts, or AK-embrvos. Proceedings of the Society for Experimental Biology and Medicine *76*, 27-32.
Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K., and Izaurralde, E. (1998). TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. Molecular cell *1*, 649-659.
Guo, F., Cen, S., Niu, M., Saadatmand, J., and Kleiman, L. (2006). Inhibition of Formula-Primed Reverse Transcription by Human APOBEC3G during Human Immunodeficiency Virus Type 1 Replication. J Virol *80*, 11710-11722.
Guzik, B.W., Levesque, L., Prasad, S., Bor, Y.C., Black, B.E., Paschal, B.M., Rekosh, D., and Hammarskjold, M.L. (2001). NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of intron-containing RNA in mammalian cells. Molecular and cellular biology *21*, 2545-2554.
Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., *et al.* (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. The Journal of clinical investigation *118*, 3132-3142.
Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., *et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science (New York, NY *302*, 415-419.
Hacker, C.V., Vink, C.A., Wardell, T.W., Lee, S., Treasure, P., Kingsman, S.M., Mitrophanous, K.A., and Miskin, J.E. (2006). The Integration Profile of EIAV-Based Vectors. Mol Ther *14*, 536-545.
Hammarskjold, M.L., Heimer, J., Hammarskjold, B., Sangwan, I., Albert, L., and Rekosh, D. (1989). Regulation of human immunodeficiency virus env expression by the rev gene product. Journal of virology *63*, 1959-1966.
Han, J.S., and Boeke, J.D. (2005). LINE-1 retrotransposons: modulators of quantity and quality of mammalian gene expression? Bioessays *27*, 775-784.

Hanafusa, H., Halpern, C.C., Buchhagen, D.L., and Kawai, S. (1977). Recovery of avian sarcoma virus from tumors induced by transformation-defective mutants. The Journal of experimental medicine *146*, 1735-1747.

Hanger, J.J., Bromham, L.D., McKee, J.J., O'Brien, T.M., and Robinson, W.F. (2000). The nucleotide sequence of koala (Phascolarctos cinereus) retrovirus: a novel type C endogenous virus related to Gibbon ape leukemia virus. Journal of virology *74*, 4264-4272.

Hansen, M.S., and Barklis, E. (1995). Structural interactions between retroviral Gag proteins examined by cysteine cross-linking. Journal of virology *69*, 1150-1159.

Harris, J.M., Haynes, R.H., and McIntosh, E.M. (1997a). A consensus sequence for a functional human endogenous retrovirus K (HERV-K) dUTPase. Biochemistry and cell biology *75*, 143-151.

Harris, J.M., Haynes, R.H., and McIntosh, E.M. (1997b). A consensus sequence for a functional human endogenous retrovirus K (HERV-K) dUTPase. Biochemistry and Cell Biology *75*, 143-151.

Harris, J.M., McIntosh, E.M., and Muscat, G.E. (2000). Expression and cytoplasmic localisation of deoxyuridine triphosphate pyrophosphatase encoded by a human endogenous retrovirus. Archives of Virology *145*, 353-363.

Harris, J.M., McIntosh, E.M., and Muscat, G.E.O. (1999). Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target. Journal of Molecular Biology *288*, 275-287.

Harris, R.S., Bishop, K.N., Sheehy, A.M., Craig, H.M., Petersen-Mahrt, S.K., Watt, I.N., Neuberger, M.S., and Malim, M.H. (2003). DNA Deamination Mediates Innate Immunity to Retroviral Infection. Cell *113*, 803.

Harris, R.S., and Liddament, M.T. (2004a). Retroviral restriction by APOBEC proteins. Nat Rev Immunol *4*, 868-877.

Harris, R.S., and Liddament, M.T. (2004b). Retroviral restriction by APOBEC proteins. Nat Rev Immunol *4*, 868-877.

Hatziioannou, T., and Goff, S.P. (2001). Infection of nondividing cells by Rous sarcoma virus. Journal of virology *75*, 9526-9531.

Hatziioannou, T., Perez-Caballero, D., Yang, A., Cowan, S., and Bieniasz, P.D. (2004). Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5alpha. Proceedings of the National Academy of Sciences of the United States of America *101*, 10774-10779.

Heinzinger, N.K., Bukinsky, M.I., Haggerty, S.A., Ragland, A.M., Kewalramani, V., Lee, M.A., Gendelman, H.E., Ratner, L., Stevenson, M., and Emerman, M. (1994). The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. Proceedings of the National Academy of Sciences of the United States of America *91*, 7311-7315.

Hematti, P., Hong, B.-K., Ferguson, C., Adler, R., Hanawa, H., Sellers, S., Holt, I.E., Eckfeldt, C.E., Sharma, Y., Schmidt, M., *et al.* (2004). Distinct Genomic Integration of MLV and SIV Vectors in Primate Hematopoietic Stem and Progenitor Cells. PLoS Biology *2*, e423.

Henderson, L.E., Krutzsch, H.C., and Oroszlan, S. (1983). Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational proteins

modification. Proceedings of the National Academy of Sciences of the United States of America *80*, 339-343.
Herniou, E., Martin, J., Miller, K., Cook, J., Wilkinson, M., and Tristem, M. (1998). Retroviral diversity and distribution in vertebrates. Journal of virology *72*, 5955-5966.
Herrera, R.J., Lowery, R.K., Alfonso, A., McDonald, J.F., and Luis, J.R. (2006). Ancient retroviral insertions among human populations. J Hum Genet *51*, 353-362.
Heslin, D.J., Murcia, P., Arnaud, F., Van Doorslaer, K., Palmarini, M., and Lenz, J. (2009). A single amino acid substitution in a segment of the CA protein within Gag that has similarity to human immunodeficiency virus type 1 blocks infectivity of a human endogenous retrovirus K provirus in the human genome. J Virol *83*, 1105-1114.
Hidaka, M., Inoue, J., Yoshida, M., and Seiki, M. (1988). Post-transcriptional regulator (rex) of HTLV-1 initiates expression of viral structural proteins but suppresses expression of regulatory proteins. EMBO Journal *7*, 519-523.
Hoffmann, C., Minkah, N., Leipzig, J., Wang, G., Arens, M.Q., Tebas, P., and Bushman, F.D. (2007). DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic acids research *35*, e91.
Holmes, R.K., Koning, F.A., Bishop, K.N., and Malim, M.H. (2007). APOBEC3F Can Inhibit the Accumulation of HIV-1 Reverse Transcription Products in the Absence of Hypermutation: COMPARISONS WITH APOBEC3G. J Biol Chem *282*, 2587-2595.
Howe, S.J., Mansour, M.R., Schwarzwaelder, K., Bartholomae, C., Hubank, M., Kempski, H., Brugman, M.H., Pike-Overzet, K., Chatters, S.J., de Ridder, D., *et al.* (2008). Insertional mutagenesis combined with acquired somatic mutations causes leukemogenesis following gene therapy of SCID-X1 patients. The Journal of clinical investigation *118*, 3143-3150.
Hughes, J.F., and Coffin, J.M. (2001). Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat Genet *29*, 487-489.
Hughes, J.F., and Coffin, J.M. (2004). Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution. PNAS *101*, 1668-1672.
Hughes, J.F., and Coffin, J.M. (2005). Human endogenous retroviral elements as indicators of ectopic recombination events in the primate genome. Genetics *171*, 1183-1194.
Hulme, A.E., Bogerd, H.P., Cullen, B.R., and Moran, J.V. (2007). Selective inhibition of Alu retrotransposition by APOBEC3G. Gene *390*, 199-205.
Indik, S., Günzburg, W.H., Salmons, B., and Rouault, F. (2005a). A novel, mouse mammary tumor virus encoded protein with Rev-like properties. Virology *337*, 1-6.
Indik, S., Gunzburg, W.H., Salmons, B., and Rouault, F. (2005b). A novel, mouse mammary tumor virus encoded protein with Rev-like properties. Virology *337*, 1-6.
Inoue, J., Yoshida, M., and Seiki, M. (1987). Transcriptional (p40x) and post-transcriptional (p27x-III) regulators are required for the expression and replication

of human T-cell leukemia virus type I genes. Proc Natl Acad Sci U S A *84*, 3653-3657.
Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvak, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell *91*, 501-510.
Iwatani, Y., Takeuchi, H., Strebel, K., and Levin, J.G. (2006). Biochemical activities of highly purified, catalytically active human APOBEC3G: correlation with antiviral effect. Journal of virology *80*, 5992-6002.
Jacks, T., and Varmus, H.E. (1985). Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science (New York, NY *230*, 1237-1242.
Janini, M., Rogers, M., Birx, D.R., and McCutchan, F.E. (2001). Human Immunodeficiency Virus Type 1 DNA Sequences Genetically Damaged by Hypermutation Are Often Abundant in Patient Peripheral Blood Mononuclear Cells and May Be Generated during Near-Simultaneous Infection and Activation of CD4+ T Cells. J Virol *75*, 7973-7986.
Javanbakht, H., Diaz-Griffero, F., Stremlau, M., Si, Z., and Sodroski, J. (2005). The contribution of RING and B-box 2 domains to retroviral restriction mediated by monkey TRIM5alpha. The Journal of biological chemistry *280*, 26933-26940.
Jensen, E.M., Zelljadt, I., Chopra, H.C., and Mason, M.M. (1970). Isolation and propagation of a virus from a spontaneous mammary carcinoma of a rhesus monkey. Cancer research *30*, 2388-2393.
Jern, P., Sperber, G., and Blomberg, J. (2005). Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology *2*, 50.
Jia, B., Serra-Moreno, R., Neidermyer, W., Rahmberg, A., Mackey, J., Fofana, I.B., Johnson, W.E., Westmoreland, S., and Evans, D.T. (2009). Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. PLoS pathogens *5*, e1000429.
Johnson, W., and Sawyer, S. (2009a). Molecular evolution of the antiretroviral TRIM5 gene. Immunogenetics.
Johnson, W.E., and Coffin, J.M. (1999). Constructing primate phylogenies from ancient retrovirus sequences. Proceedings of the National Academy of Sciences of the United States of America *96*, 10254-10260.
Johnson, W.E., and Sawyer, S.L. (2009b). Molecular evolution of the antiretroviral TRIM5 gene. Immunogenetics *61*, 163-176.
Jordan, I.K., Rogozin, I.B., Glazko, G.V., and Koonin, E.V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet *19*, 68-72.
Jouvenet, N., Neil, S.J., Zhadina, M., Zang, T., Kratovac, Z., Lee, Y., McNatt, M., Hatziioannou, T., and Bieniasz, P.D. (2009a). Broad-spectrum inhibition of retroviral and filoviral particle release by tetherin. J Virol *83*, 1837-1844.
Jouvenet, N., Neil, S.J.D., Zhadina, M., Zang, T., Kratovac, Z., Lee, Y., McNatt, M., Hatziioannou, T., and Bieniasz, P.D. (2009b). Broad-Spectrum Inhibition of Retroviral and Filoviral Particle Release by Tetherin. J Virol *83*, 1837-1844.
Kao, S., Khan, M.A., Miyagi, E., Plishka, R., Buckler-White, A., and Strebel, K. (2003). The human immunodeficiency virus type 1 Vif protein reduces intracellular

expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. Journal of virology *77*, 11398-11407.
Kao, S.Y., Calman, A.F., Luciw, P.A., and Peterlin, B.M. (1987). Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product. Nature *330*, 489-493.
Kapitonov, V.V., and 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 - 251.
Katahira, J., Strasser, K., Podtelejnikov, A., Mann, M., Jung, J.U., and Hurt, E. (1999). The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. The EMBO journal *18*, 2593-2609.
Katzourakis, A., Tristem, M., Pybus, O.G., and Gifford, R.J. (2007). Discovery and analysis of the first endogenous lentivirus. Proceedings of the National Academy of Sciences of the United States of America *104*, 6261-6265.
Kazazian, H.H., Jr. (2004). Mobile Elements: Drivers of Genome Evolution. Science *303*, 1626-1632.
Keckesova, Z., Ylinen, L.M., and Towers, G.J. (2004). The human and African green monkey TRIM5alpha genes encode Ref1 and Lv1 retroviral restriction factor activities. Proceedings of the National Academy of Sciences of the United States of America *101*, 10780-10785.
Khan, M.A., Goila-Gaur, R., Opi, S., Miyagi, E., Takeuchi, H., Kao, S., and Strebel, K. (2007). Analysis of the contribution of cellular and viral RNA to the packaging of APOBEC3G into HIV-1 virions. Retrovirology *4*, 48.
Khan, M.A., Kao, S., Miyagi, E., Takeuchi, H., Goila-Gaur, R., Opi, S., Gipson, C.L., Parslow, T.G., Ly, H., and Strebel, K. (2005). Viral RNA is required for the association of APOBEC3G with human immunodeficiency virus type 1 nucleoprotein complexes. Journal of virology *79*, 5870-5874.
Kikonyogo, A., Bouamr, F., Vana, M.L., Xiang, Y., Aiyar, A., Carter, C., and Leis, J. (2001). Proteins related to the Nedd4 family of ubiquitin protein ligases interact with the L domain of Rous sarcoma virus and are required for gag budding from cells. Proceedings of the National Academy of Sciences of the United States of America *98*, 11199-11204.
Kim, J.D., Faulk, C., and Kim, J. (2007). Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. Nucl Acids Res *35*, 3442-3452.
Kim, T.H., Jeon, Y.J., Yi, J.M., Kim, D.S., Huh, J.W., Hur, C.G., and Kim, H.S. (2004a). The distribution and expression of HERV families in the human genome. Molecules and cells *18*, 87-93.
Kim, T.H., Jeon, Y.J., Yi, J.M., Kim, D.S., Huh, J.W., Hur, C.G., and Kim, H.S. (2004b). The distribution and expression of HERV families in the human genome. Molecules and Cells *18*, 87-93.
Kitamura, Y., Ayukawa, T., Ishikawa, T., Kanda, T., and Yoshiike, K. (1996). Human endogenous retrovirus K10 encodes a functional integrase. J Virol *70*, 3302-3306.
Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.C., and Montagnier, L. (1984). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature *312*, 767-768.

Klimkait, T., Strebel, K., Hoggan, M.D., Martin, M.A., and Orenstein, J.M. (1990). The human immunodeficiency virus type 1-specific protein vpu is required for efficient virus maturation and release. Journal of virology *64*, 621-629.

Knossl, M., Lower, R., and Lower, J. (1999). Expression of the Human Endogenous Retrovirus HTDV/HERV-K Is Enhanced by Cellular Transcription Factor YY1. J Virol *73*, 1254-1261.

Kozak, C.A., and Chakraborti, A. (1996). Single amino acid changes in the murine leukemia virus capsid protein gene define the target of Fv1 resistance. Virology *225*, 300-305.

Krieger, J.N., Coombs, R.W., Collier, A.C., Ross, S.O., Chaloupka, K., Cummings, D.K., Murphy, V.L., and Corey, L. (1991). Recovery of human immunodeficiency virus type 1 from semen: minimal impact of stage of infection and current antiviral chemotherapy. Journal of Infectious Diseases *163*, 386-388.

Kunkel, L.M., Smith, K.D., and Boyer, S.H. (1976). Human Y-chromosome-specific reiterated DNA. Science *191*, 1189-1190.

Kupzig, S., Korolchuk, V., Rollason, R., Sugden, A., Wilde, A., and Banting, G. (2003). Bst-2/HM1.24 is a raft-associated apical membrane protein with an unusual topology. Traffic (Copenhagen, Denmark) *4*, 694-709.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., and Baldwin, J. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Landry, J.R., Rouhi, A., Medstrand, P., and Mager, D.L. (2002). The opitz syndrome gene mid1 is transcribed from a human endogenous retroviral promoter. Mol Biol Evol *19*, 1934 - 1942.

Lavie, L., Medstrand, P., Schempp, W., Meese, E., and Mayer, J. (2004). Human Endogenous Retrovirus Family HERV-K(HML-5): Status, Evolution, and Reconstruction of an Ancient Betaretrovirus in the Human Genome. J Virol *78*, 8788-8798.

Lewinski, M.K., Yamashita, M., Emerman, M., Ciuffi, A., Marshall, H., Crawford, G., Collins, F., Shinn, P., Leipzig, J., Hannenhalli, S., *et al.* (2006). Retroviral DNA Integration: Viral and Cellular Determinants of Target-Site Selection. PLoS Pathog *2*, e60.

Lewis, P., Hensel, M., and Emerman, M. (1992). Human immunodeficiency virus infection of cells arrested in the cell cycle. The EMBO journal *11*, 3053-3058.

Lewis, P.F., and Emerman, M. (1994). Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. Journal of virology *68*, 510-516.

Liao, C.H., Kuang, Y.Q., Liu, H.L., Zheng, Y.T., and Su, B. (2007). A novel fusion gene, TRIM5-Cyclophilin A in the pig-tailed macaque determines its susceptibility to HIV-1 infection. AIDS (London, England) *21 Suppl 8*, S19-26.

Lilly, F. (1967). Susceptibility to two strains of Friend leukemia virus in mice. Science (New York, NY *155*, 461-462.

Linial, M.L. (1999). Foamy Viruses Are Unconventional Retroviruses. J Virol *73*, 1747-1755.

Liu, B., Sarkis, P.T.N., Luo, K., Yu, Y., and Yu, X.-F. (2005a). Regulation of Apobec3F and Human Immunodeficiency Virus Type 1 Vif by Vif-Cul5-ElonB/C E3 Ubiquitin Ligase. J Virol *79*, 9579-9587.

Liu, H.L., Wang, Y.Q., Liao, C.H., Kuang, Y.Q., Zheng, Y.T., and Su, B. (2005b). Adaptive evolution of primate TRIM5alpha, a gene restricting HIV-1 infection. Gene *362*, 109-116.

Liu, Y., Liu, H., Kim, B.O., Gattone, V.H., Li, J., Nath, A., Blum, J., and He, J.J. (2004). CD4-independent infection of astrocytes by human immunodeficiency virus type 1: requirement for the human mannose receptor. J Virol *78*, 4120-4133.

Llano, M., Delgado, S., Vanegas, M., and Poeschla, E.M. (2004a). Lens Epithelium-derived Growth Factor/p75 Prevents Proteasomal Degradation of HIV-1 Integrase. J Biol Chem *279*, 55570-55577.

Llano, M., Vanegas, M., Fregoso, O., Saenz, D., Chung, S., Peretz, M., and Poeschla, E.M. (2004b). LEDGF/p75 Determines Cellular Trafficking of Diverse Lentiviral but Not Murine Oncoretroviral Integrase Proteins and Is a Component of Functional Lentiviral Preintegration Complexes. J Virol *78*, 9524-9537.

Lowe, C.B., Bejerano, G., and Haussler, D. (2007). Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc Natl Acad Sci U S A *104*, 8005-8010.

Lower, J., Lower, R., Stegmann, J., Frank, H., and Kurth, R. (1981). Retrovirus particle production in three of four human teratocarcinoma cell lines. Haematology and blood transfusion *26*, 541-544.

Lower, J., Wondrak, E.M., and Kurth, R. (1987). Genome Analysis and Reverse Transcriptase Activity of Human Teratocarcinoma-derived Retroviruses. J Gen Virol *68*, 2807-2815.

Lower, R., Boller, K., Hasenmaier, B., Korbmacher, C., Muller-Lantzsch, N., Lower, J., and Kurth, R. (1993). Identification of Human Endogenous Retroviruses with Complex mRNA Expression and Particle Formation. PNAS *90*, 4480-4484.

Lower, R., Lower, J., Frank, H., Harzmann, R., and Kurth, R. (1984). Human Teratocarcinomas Cultured in vitro Produce Unique Retrovirus-like Viruses. J Gen Virol *65*, 887-898.

Lower, R., Lower, J., and Kurth, R. (1996). The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. PNAS *93*, 5177-5184.

Lower, R., Tonjes, R.R., Korbmacher, C., Kurth, R., and Lower, J. (1995). Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. J Virol *69*, 141-149.

Lyons, M.J., and Moore, D.H. (1962). Purification of the Mouse Mammary Tumour Virus. Nature *194*, 1141-1142.

Macfarlane, C., and Simmonds, P. (2004). Allelic Variation of HERV-K(HML-2) Endogenous Retroviral Elements in Human Populations. Journal of Molecular Evolution *59*, 642-656.

Maddon, P.J., Dalgleish, A.G., McDougal, J.S., Clapham, P.R., Weiss, R.A., and Axel, R. (1986). The T4 gene encodes the AIDS virus receptor and is expressed in the immune system and the brain. Cell *47*, 333-348.

Maertens, G., Cherepanov, P., Pluymers, W., Busschots, K., De Clercq, E., Debyser, Z., and Engelborghs, Y. (2003). LEDGF/p75 Is Essential for Nuclear and Chromosomal Targeting of HIV-1 Integrase in Human Cells. J Biol Chem *278*, 33528-33539.

Mager, D.L., Hunter, D.G., Schertzer, M., and Freeman, J.D. (1999). Endogenous retroviruses provide the primary polyadenylation signal for two new human genes (HHLA2 and HHLA3). Genomics *59*, 255 - 263.

Magin-Lachmann, C., Hahn, S., Strobel, H., Held, U., Lower, J., and Lower, R. (2001). Rec (Formerly Corf) Function Requires Interaction with a Complex, Folded RNA Structure within Its Responsive Element rather than Binding to a Discrete Specific Binding Site. J Virol *75*, 10359-10371.

Magin, C., Hesse, J., Lower, J., and Lower, R. (2000). Corf, the Rev/Rex Homologue of HTDV/HERV-K, Encodes an Arginine-Rich Nuclear Localization Signal That Exerts a trans-Dominant Phenotype When Mutated. Virology *274*, 11.

Magin, C., Lower, R., and Lower, J. (1999). cORF and RcRE, the Rev/Rex and RRE/RxRE Homologues of the Human Endogenous Retrovirus Family HTDV/HERV-K. J Virol *73*, 9496-9507.

Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature *424*, 99-103.

Mariani-Costantini, R., Horn, T.M., and Callahan, R. (1989). Ancestry of a human endogenous retrovirus family, pp. 4982-4985.

Marsh, M., and Helenius, A. (1989). Virus entry into animal cells. Advances in virus research *36*, 107-151.

Marshall, H.M., Ronen, K., Berry, C., Llano, M., Sutherland, H., Saenz, D., Bickmore, W., Poeschla, E., and Bushman, F.D. (2007). Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. PloS one *2*, e1340.

Martin-Serrano, J., Eastman, S.W., Chung, W., and Bieniasz, P.D. (2005). HECT ubiquitin ligases link viral and cellular PPXY motifs to the vacuolar protein-sorting pathway. The Journal of cell biology *168*, 89-101.

Martin-Serrano, J., Yarovoy, A., Perez-Caballero, D., and Bieniasz, P.D. (2003). Divergent retroviral late-budding domains recruit vacuolar protein sorting factors by using alternative adaptor proteins. Proceedings of the National Academy of Sciences of the United States of America *100*, 12414-12419.

Martin-Serrano, J., Zang, T., and Bieniasz, P.D. (2001). HIV-1 and Ebola virus encode small peptide motifs that recruit Tsg101 to sites of particle assembly to facilitate egress. Nature medicine *7*, 1313-1319.

Martin, J., Herniou, E., Cook, J., Waugh O'Neill, R., and Tristem, M. (1997). Human endogenous retrovirus type I-related viruses have an apparently widespread distribution within vertebrates. Journal of virology *71*, 437-443.

Martin, M.A., Bryan, T., Rasheed, S., and Khan, A.S. (1981). Identification and cloning of endogenous retroviral sequences present in human DNA. Proc Natl Acad Sci U S A *78*, 4892-4896.

Martineau, D., Bowser, P.R., Renshaw, R.R., and Casey, J.W. (1992). Molecular characterization of a unique retrovirus associated with a fish tumor. Journal of virology *66*, 596-599.

Martineau, D., Bowser, P.R., Wooster, G.A., and Armstrong, L.D. (1990). Experimental transmission of a dermal sarcoma in fingerling walleyes (Stizostedion vitreum vitreum). Veterinary pathology *27*, 230-234.

May, F.E., and Westley, B.R. (1986). Structure of a human retroviral sequence related to mouse mammary tumor virus. Journal of Virology *60*, 743-749.
May, F.E.B., Westley, B.R., Rochefort, H., Buetti, E., and Diggelmann, H. (1983). Mouse mammary tumour virus related sequences are present in human DNA. Nucleic Acid Research *11*, 4127-4139.
Mayer, J., and Meese, E. (2003). Presence of dUTPase in the Various Human Endogenous Retrovirus K (HERV-K) Families. Journal of Molecular Evolution *57*, 642-649.
Mayer, J., Sauter, M., Racz, A., Scherer, D., Mueller-Lantzsch, N., and Meese, E. (1999a). An almost-intact human endogenous retrovirus K on human chromosome 7. Nat Genet *21*, 257.
Mayer, J., Sauter, M., Racz, A., Scherer, D., Mueller-Lantzsch, N., and Meese, E. (1999b). An almost-intact human endogenous retrovirus K on human chromosome 7. Nat Genet *21*, 257-258.
Mbisa, J.L., Barr, R., Thomas, J.A., Vandegraaff, N., Dorweiler, I.J., Svarovskaia, E.S., Brown, W.L., Mansky, L.M., Gorelick, R.J., Harris, R.S., *et al.* (2007). Human Immunodeficiency Virus Type 1 cDNAs Produced in the Presence of APOBEC3G Exhibit Defects in Plus-Strand DNA Transfer and Integration. J Virol *81*, 7099-7110.
McClure, M.O., Sommerfelt, M.A., Marsh, M., and Weiss, R.A. (1990). The pH independence of mammalian retrovirus infection. The Journal of general virology *71 (Pt 4)*, 767-773.
McNatt, M.W., Zang, T., Hatziioannou, T., Bartlett, M., Fofana, I.B., Johnson, W.E., Neil, S.J., and Bieniasz, P.D. (2009a). Species-specific activity of HIV-1 Vpu and positive selection of tetherin transmembrane domain variants. PLoS pathogens *5*, e1000300.
McNatt, M.W., Zang, T., Hatziioannou, T., Bartlett, M., Fofana, I.B., Johnson, W.E., Neil, S.J.D., and Bieniasz, P.D. (2009b). Species-Specific Activity of HIV-1 Vpu and Positive Selection of Tetherin Transmembrane Domain Variants. PLoS Pathog *5*, e1000300.
Medstrand, P., and Blomberg, J. (1993a). Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. J Virol *67*, 6778-6787.
Medstrand, P., and Blomberg, J. (1993b). Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. J Virol *67*, 6778-6787.
Medstrand, P., Landry, J.R., and Mager, D.L. (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 - 1903.
Medstrand, P., and Mager, D.L. (1998). Human-Specific Integrations of the HERV-K Endogenous Retrovirus Family. J Virol *72*, 9782-9787.
Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement Distributions in the Human Genome: Variations Associated With Age and Proximity to Genes. Genome Research *12*, 1483-1495.

Mehle, A., Goncalves, J., Santa-Marta, M., McPike, M., and Gabuzda, D. (2004). Phosphorylation of a novel SOCS-box regulates assembly of the HIV-1 Vif-Cul5 complex that promotes APOBEC3G degradation. Genes & Development *18*, 2861-2866.

Mertz, J.A., Simper, M.S., Lozano, M.M., Payne, S.M., and Dudley, J.P. (2005a). Mouse Mammary Tumor Virus Encodes a Self-Regulatory RNA Export Protein and Is a Complex Retrovirus. J Virol *79*, 14737-14747.

Mertz, J.A., Simper, M.S., Lozano, M.M., Payne, S.M., and Dudley, J.P. (2005b). Mouse mammary tumor virus encodes a self-regulatory RNA export protein and is a complex retrovirus. Journal of virology *79*, 14737-14747.

Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., *et al.* (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature *403*, 785-789.

Miller, D.G., Trobridge, G.D., Petek, L.M., Jacobs, M.A., Kaul, R., and Russell, D.W. (2005). Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. Journal of virology *79*, 11434-11442.

Miller, M.D., Farnet, C.M., and Bushman, F.D. (1997). Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. Journal of virology *71*, 5382-5390.

Mische, C.C., Javanbakht, H., Song, B., Diaz-Griffero, F., Stremlau, M., Strack, B., Si, Z., and Sodroski, J. (2005). Retroviral restriction factor TRIM5alpha is a trimer. Journal of virology *79*, 14446-14450.

Mitchell, R.S., Beitzel, B.F., Schroder, A.R.W., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. PLoS Biology *2*, e234.

Miyagi, E., Opi, S., Takeuchi, H., Khan, M., Goila-Gaur, R., Kao, S., and Strebel, K. (2007). Enzymatically Active APOBEC3G Is Required for Efficient Inhibition of Human Immunodeficiency Virus Type 1. J Virol *81*, 13346-13353.

Miyauchi, K., Kim, Y., Latinovic, O., Morozov, V., and Melikyan, G.B. (2009). HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. Cell *137*, 433-444.

Moebes, A., Enssle, J., Bieniasz, P.D., Heinkelein, M., Lindemann, D., Bock, M., McClure, M.O., and Rethwilm, A. (1997). Human foamy virus reverse transcription that occurs late in the viral replication cycle. Journal of virology *71*, 7305-7311.

Moore, D.H. (1963). Mouse Mammary Tumour Agent and Mouse Mammary Tumours. Nature *198*, 429-433.

Moore, D.H., Charney, J., Kramarsky, B., Lasfargues, E.Y., Sarkar, N.H., Brennan, M.J., Burrows, J.H., Sirsat, S.M., Paymaster, J.C., and Vaidya, A.B. (1971). Search for a Human Breast Cancer Virus. Nature *229*, 611-615.

Moore, D.H., Sarker, N.H., E., K.C., Pillsbury, N., and Charney, J. (1969). Type B particles in human milk. Texas reports on biology and medicine *27*, 1027-1039.

Mothes, W., Boerger, A.L., Narayan, S., Cunningham, J.M., and Young, J.A. (2000). Retroviral entry mediated by receptor priming and low pH triggering of an envelope glycoprotein. Cell *103*, 679-689.

Moyes, D., Griffiths, D.J., and Venables, P.J. (2007). Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. Trends Genet *23*, 326-333.

Muciaccia, B., Corallini, S., Vicini, E., Padula, F., Gandini, L., Liuzzi, G., Lenzi, A., and Stefanini, M. (2007). HIV-1 viral DNA is present in ejaculated abnormal spermatozoa of seropositive subjects. Hum Reprod *22*, 2868-2878.

Muciaccia, B., Uccini, S., Filippini, A., Ziparo, E., Paraire, F., Baroni, C.D., and Stefanini, M. (1998). Presence and cellular distribution of HIV in the testes of seropositive subjects: an evaluation by in situ PCR hybridization. FASEB J *12*, 151-163.

Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., Flory, E., Schumann, G.G., and Munk, C. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. The Journal of biological chemistry *281*, 22161-22172.

Mueller-Lantzsch, N., Sauter, M., Weiskircher, A., Kramer, K., Best, B., Buck, M., and Grasser, F. (1993). Human Endogenous Retroviral Element K10 (HERV-K10) Encodes a Full-Length Gag Homologous 73-kDa Protein and a Functional Protease. AIDS Research and Human Retroviruses *9*, 343-349.

Nakayama, E.E., Miyoshi, H., Nagai, Y., and Shioda, T. (2005). A specific region of 37 amino acid residues in the SPRY (B30.2) domain of African green monkey TRIM5alpha determines species-specific restriction of simian immunodeficiency virus SIVmac infection. Journal of virology *79*, 8870-8877.

Narezkina, A., Taganov, K.D., Litwin, S., Stoyanova, R., Hayashi, J., Seeger, C., Skalka, A.M., and Katz, R.A. (2004). Genome-Wide Analyses of Avian Sarcoma Virus Integration Sites. J Virol *78*, 11656-11663.

Neil, S.J., Sandrin, V., Sundquist, W.I., and Bieniasz, P.D. (2007). An interferon-alpha-induced tethering mechanism inhibits HIV-1 and Ebola virus particle release but is counteracted by the HIV-1 Vpu protein. Cell host & microbe *2*, 193-203.

Neil, S.J., Zang, T., and Bieniasz, P.D. (2008a). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature *451*, 425-430.

Neil, S.J.D., Zang, T., and Bieniasz, P.D. (2008b). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature *451*, 425-430.

Neville, M., Stutz, F., Lee, L., Davis, L.I., and Rosbash, M. (1997). The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. Curr Biol *7*, 767-775.

Newman, E.N., Holmes, R.K., Craig, H.M., Klein, K.C., Lingappa, J.R., Malim, M.H., and Sheehy, A.M. (2005). Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. Curr Biol *15*, 166-170.

Newman, R.M., Hall, L., Kirmaier, A., Pozzi, L.A., Pery, E., Farzan, M., O'Neil, S.P., and Johnson, W. (2008). Evolution of a TRIM5-CypA splice isoform in old world monkeys. PLoS pathogens *4*, e1000003.

Nguyen, D.G., and Hildreth, J.E. (2003). Involvement of macrophage mannose receptor in the binding and transmission of HIV by macrophages. European journal of immunology *33*, 483-493.

Nguyen, N., Zhang, X., Olashaw, N., and Seto, E. (2004). Molecular Cloning and Functional Characterization of the Transcription Factor YY2. J Biol Chem *279*, 25927-25934.

Niewiadomska, A.M., Tian, C., Tan, L., Wang, T., Sarkis, P.T.N., and Yu, X.-F. (2007). Differential Inhibition of Long Interspersed Element 1 by APOBEC3 Does Not Correlate with High-Molecular-Mass-Complex Formation or P-Body Association. J Virol *81*, 9577-9583.

Nisole, S., Lynch, C., Stoye, J.P., and Yap, M.W. (2004a). A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. Proceedings of the National Academy of Sciences of the United States of America *101*, 13324-13328.

Nisole, S.b., Lynch, C., Stoye, J.P., and Yap, M.W. (2004b). A Trim5-cyclophilin A fusion protein found in owl monkey kidney cells can restrict HIV-1. Proceedings of the National Academy of Sciences of the United States of America *101*, 13324-13328.

Nuovo, G.J., Becker, J., Simsir, A., Margiotta, M., khalife, G., and Shevchuk, M. (1994). HIV-1 nucleic acids localize to the spermatogonia and their progeny. A study by polymerase chain reaction in situ hybridization. American Journal of Pathology *144*, 1142-1148.

Odaka, T., Ikeda, H., Moriwaki, K., Matsuzawa, A., Mizuno, M., and Kondo, K. (1978). Genetic resistance in Japanese wild mice (Mus musculus molossinus) to an NB-tropic Friend murine leukemia virus. Journal of the National Cancer Institute *61*, 1301-1306.

OhAinle, M., Kerns, J.A., Malik, H.S., and Emerman, M. (2006). Adaptive Evolution and Antiviral Activity of the Conserved Mammalian Cytidine Deaminase APOBEC3H. J Virol *80*, 3853-3862.

Ohno, T., Sweet, R.W., Dejak, D., and Spiegelman, S. (1977). Purification and characterization of the DNA polymerase of human breast cancer particles. Proceedings of the National Academy of Sciences of the United States of America *74*, 764-768.

Ono, M. (1986). Molecular Cloning and Long Terminal Repeat Sequences of Human Endogenous Retrovirus Genes Related to Types A and B Retrovirus Genes. Journal of Virology *58*, 937-944.

Ono, M., Yasunaga, T., Miyata, T., and Ushikubo, H. (1986). Nucleotide Sequence of Human Endogenous Retrovirus Genome Related to the Mouse Mammary Tumor Virus Genome. Journal of Virology *60*, 589-598.

Opi, S., Takeuchi, H., Kao, S., Khan, M.A., Miyagi, E., Goila-Gaur, R., Iwatani, Y., Levin, J.G., and Strebel, K. (2006). Monomeric APOBEC3G is catalytically active and has antiviral activity. Journal of virology *80*, 4673-4682.

Perez-Caballero, D., Hatziioannou, T., Yang, A., Cowan, S., and Bieniasz, P.D. (2005). Human tripartite motif 5alpha domains responsible for retrovirus restriction activity and specificity. Journal of virology *79*, 8969-8978.

Perron, M.J., Stremlau, M., Lee, M., Javanbakht, H., Song, B., and Sodroski, J. (2007). The human TRIM5alpha restriction factor mediates accelerated uncoating of the N-tropic murine leukemia virus capsid. Journal of virology *81*, 2138-2148.

Perron, M.J., Stremlau, M., Song, B., Ulm, W., Mulligan, R.C., and Sodroski, J. (2004). TRIM5alpha mediates the postentry block to N-tropic murine leukemia viruses in human cells. Proceedings of the National Academy of Sciences of the United States of America *101*, 11827-11832.

Poiesz, B.J., Ruscetti, F.W., Gazdar, A.F., Bunn, P.A., Minna, J.D., and Gallo, R.C. (1980). Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America *77*, 7415-7419.

Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. (1984). Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science *224*, 497-500.

Preston, B.D., Poiesz, B.J., and Loeb, L.A. (1988). Fidelity of HIV-1 reverse transcriptase. Science *242*, 1168-1171.

Qi, C.F., Bonhomme, F., Buckler-White, A., Buckler, C., Orth, A., Lander, M.R., Chattopadhyay, S.K., and Morse, H.C., 3rd (1998). Molecular phylogeny of Fv1. Mamm Genome *9*, 1049-1055.

Reus, K., Mayer, J., Sauter, M., Scherer, D., Müller-Lantzsch, N., and Meese, E. (2001). Genomic Organization of the Human Endogenous Retrovirus HERV-K(HML-2.HOM) (ERVK6) on Chromosome 7. Genomics *72*, 314-320.

Rhee, S.S., Hui, H.X., and Hunter, E. (1990). Preassembled capsids of type D retroviruses contain a signal sufficient for targeting specifically to the plasma membrane. Journal of virology *64*, 3844-3852.

Rhee, S.S., and Hunter, E. (1987). Myristylation is required for intracellular transport but not for assembly of D-type retrovirus capsids. J Virol *61*, 1045-1053.

Rhee, S.S., and Hunter, E. (1990a). A single amino acid substitution within the matrix protein of a type D retrovirus converts its morphogenesis to that of a type C retrovirus. Cell *63*, 77-86.

Rhee, S.S., and Hunter, E. (1990b). A single amino acid substitution within the matrix protein of a type D retrovirus converts its morphogenesis to that of a type C retrovirus. Cell *63*, 77.

Rhodes, D.A., de Bono, B., and Trowsdale, J. (2005). Relationship between SPRY and B30.2 protein domains. Evolution of a component of immune defence? Immunology *116*, 411-417.

Roberts, J.D., Bebenek, K., and Kunkel, T.A. (1988). The accuracy of reverse transcriptase from HIV-1. Science *242*, 1171-1173.

Roe, T., Chow, S.A., and Brown, P.O. (1997). 3'-end processing and kinetics of 5'-end joining during retroviral integration in vivo. Journal of virology *71*, 1334-1340.

Romano, C.M., Ramalho, R.F., and Zanotto, P.M.d.A. (2006). Tempo and mode of ERV-K evolution in human and chimpanzee genomes. Archives of Virology *151*, 2215.

Rommelaere, J., Donis-Keller, H., and Hopkins, N. (1979). RNA sequencing provides evidence for allelism of determinants of the N-, B- or NB-tropism of murine leukemia viruses. Cell *16*, 43-50.

Rosenblum, L., Patton, G., Grigg, A., Frater, A., Cain, D., Erlwein, O., Hill, C., Clarke, J., and McClure, M. (2001). Differential susceptibility of retroviruses to nucleoside analogues. Antiviral Chemistry and Chemotherapy *12*, 91-97.

Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., *et al.* (2005). The DNA sequence of the human X chromosome. Nature *434*, 325-337.

Ross, S.R., Schofield, J.J., Farr, C.J., and Bucan, M. (2002). Mouse transferrin receptor 1 is the cell entry receptor for mouse mammary tumor virus. Proceedings of the National Academy of Sciences of the United States of America *99*, 12386-12390.

Roth, M.J., Schwartzberg, P.L., and Goff, S.P. (1989). Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. Cell *58*, 47-54.

Rous, P. (1910). A TRANSMISSIBLE AVIAN NEOPLASM. (SARCOMA OF THE COMMON FOWL.). J Exp Med *12*, 696-705.

Rous, P. (1911). A SARCOMA OF THE FOWL TRANSMISSIBLE BY AN AGENT SEPARABLE FROM THE TUMOR CELLS. J Exp Med *13*, 397-411.

Roy-Engel, A.M., El-Sawy, M., Farooq, L., Odom, G.L., Perepelitsa-Belancio, V., Bruch, H., Oyeniran, O.O., and Deininger, P.L. (2005). Human retroelements may introduce intragenic polyadenylation signals. Cytogenetic and genome research *110*, 365-371.

Ruda, V.M., Akopov, S.B., Trubetskoy, D.O., Manuylov, N.L., Vetchinova, A.S., Zavalova, L.L., Nikolaev, L.G., and Sverdlov, E.D. (2004). Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. Virus Research *104*, 11-16.

Ruprecht, K., Ferreira, H., Flockerzi, A., Wahl, S., Sauter, M., Mayer, J., and Mueller-Lantzsch, N. (2008). Human Endogenous Retrovirus Family HERV-K(HML-2) RNA Transcripts Are Selectively Packaged into Retroviral Particles Produced by the Human Germ Cell Tumor Line Tera-1 and Originate Mainly from a Provirus on Chromosome 22q11.21. J Virol *82*, 10008-10016.

Rutledge, E.A., and Russell, D.W. (1997). Adeno-associated virus vector integration junctions. Journal of virology *71*, 8429-8436.

Saenz, D.T., Loewen, N., Peretz, M., Whitwam, T., Barraza, R., Howell, K.G., Holmes, J.M., Good, M., and Poeschla, E.M. (2004). Unintegrated Lentivirus DNA Persistence and Accessibility to Expression in Nondividing Cells: Analysis with Class I Integrase Mutants. J Virol *78*, 2906-2920.

Sakai, H., Tokunaga, K., Kawamura, M., and Adachi, A. (1995). Function of human immunodeficiency virus type 1 Vpu protein in various cell types. The Journal of general virology *76 (Pt 11)*, 2717-2722.

Sakuma, R., Mael, A.A., and Ikeda, Y. (2007). Alpha interferon enhances TRIM5alpha-mediated antiviral activities in human and rhesus monkey cells. Journal of virology *81*, 10201-10206.

Sakuma, T., Noda, T., Urata, S., Kawaoka, Y., and Yasuda, J. (2009). Inhibition of Lassa and Marburg virus production by tetherin. J Virol *83*, 2382-2385.

Sakurai, A., Yasuda, J., Takano, H., Tanaka, Y., Hatakeyama, M., and Shida, H. (2004). Regulation of human T-cell leukemia virus type 1 (HTLV-1) budding by ubiquitin ligase Nedd4. Microbes and infection / Institut Pasteur *6*, 150-156.

Sasada, A., Takaori-Kondo, A., Shirakawa, K., Kobayashi, M., Abudu, A., Hishizawa, M., Imada, K., Tanaka, Y., and Uchiyama, T. (2005). APOBEC3G targets human T-cell leukemia virus type 1. Retrovirology *2*, 32.

Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Dolken, G., Lemm, I., Buck, M., Best, B., Neumann-Haefelin, D., and Mueller-Lantzsch, N. (1995). Human

endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas. J Virol *69*, 414 - 421.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2004). Ancient Adaptive Evolution of the Primate Antiviral DNA-Editing Enzyme APOBEC3G. PLoS Biology *2*, e275.

Sawyer, S.L., Emerman, M., and Malik, H.S. (2007). Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. PLoS pathogens *3*, e197.

Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005a). Positive selection of primate TRIM5alpha identifies a critical species-specific retroviral restriction domain. Proceedings of the National Academy of Sciences of the United States of America *102*, 2832-2837.

Sawyer, S.L., Wu, L.I., Emerman, M., and Malik, H.S. (2005b). Positive selection of primate TRIM5Î± identifies a critical species-specific retroviral restriction domain. Proceedings of the National Academy of Sciences of the United States of America *102*, 2832-2837.

Sayah, D.M., Sokolskaja, E., Berthoux, L., and Luban, J. (2004a). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature *430*, 569-573.

Sayah, D.M., Sokolskaja, E., Berthoux, L., and Luban, J. (2004b). Cyclophilin A retrotransposition into TRIM5 explains owl monkey resistance to HIV-1. Nature *430*, 569-573.

Schlom, J., Spiegelman, S., and Moore, D. (1971). RNA-dependent DNA Polymerase Activity in Virus-like Particles isolated from Human Milk. Nature *231*, 97-100.

Schlom, J., Spiegelman, S., and Moore, D.H. (1972). Detection of High-Molecular-Weight RNA in Particles from Human Milk. Science *175*, 542-544.

Schommer, S., Sauter, M., Krausslich, H.-G., Best, B., and Mueller-Lantzsch, N. (1996). Characterization of the human endogenous retrovirus K proteinase. J Gen Virol *77*, 375-379.

Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002a). HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. Cell *110*, 521-529.

Schröder, A.R.W., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002b). HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *110*, 521-529.

Sebastian, S., and Luban, J. (2005). TRIM5alpha selectively binds a restriction-sensitive retroviral capsid. Retrovirology *2*, 40.

Seifarth, W., Frank, O., Zeilfelder, U., Spiess, B., Greenwood, A.D., Hehlmann, R., and Leib-Mosch, C. (2005). Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. J Virol *79*, 341 - 352.

Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature *418*, 646.

Sheehy, A.M., Gaddis, N.C., and Malim, M.H. (2003). The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nature medicine *9*, 1404-1407.

Shehu-Xhilaga, M., Kent, S., Batten, J., Ellis, S., Van der Meulen, J., O'Bryan, M., Cameron, P., Lewin, S., and Hedger, M. (2007). The testis and epididymis are productively infected by SIV and SHIV in juvenile macaques during the post-acute stage of infection. Retrovirology *4*, 7.

Shi, J., and Aiken, C. (2006). Saturation of TRIM5 alpha-mediated restriction of HIV-1 infection depends on the stability of the incoming viral capsid. Virology *350*, 493-500.

Shih, T.Y., Weeks, M.O., Young, H.A., and Scholnick, E.M. (1979). Identification of a sarcoma virus-coded phosphoprotein in nonproducer cells transformed by Kirsten or Harvey murine sarcoma virus. Virology *96*, 64-79.

Shindo, K., Takaori-Kondo, A., Kobayashi, M., Abudu, A., Fukunaga, K., and Uchiyama, T. (2003). The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not a sole determinant of its antiviral activity. The Journal of biological chemistry *278*, 44412-44416.

Simon, V., and Ho, D.D. (2003). HIV-1 dynamics in vivo: implications for therapy. Nature reviews *1*, 181-190.

Simon, V., Zennou, V., Murray, D., Huang, Y., Ho, D.D., and Bieniasz, P.D. (2005). Natural variation in Vif: differential impact on APOBEC3G/3F and a potential role in HIV-1 diversification. PLoS pathogens *1*, e6.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., *et al.* (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825-837.

Smit, A.F.A. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. Current Opinion in Genetics & Development *9*, 657-663.

Song, B., Gold, B., O'Huigin, C., Javanbakht, H., Li, X., Stremlau, M., Winkler, C., Dean, M., and Sodroski, J. (2005). The B30.2(SPRY) domain of the retroviral restriction factor TRIM5alpha exhibits lineage-specific length and sequence variation in primates. J Virol *79*, 6111-6121.

Song, S., Laipis, P.J., Berns, K.I., and Flotte, T.R. (2001). Effect of DNA-dependent protein kinase on the molecular fate of the rAAV2 genome in skeletal muscle. Proc Natl Acad Sci U S A *98*, 4084-4088.

Steele, P.E., Rabson, A.B., Bryan, T., and Martin, M.A. (1984). Distinctive termini characterize two families of human endogenous retroviral sequences. Science *225*, 943-947.

Stehelin, D., Varmus, H.E., Bishop, J.M., and Vogt, P.K. (1976). DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. Nature *260*, 170-173.

Stein, B.S., Gowda, S.D., Lifson, J.D., Penhallow, R.C., Bensch, K.G., and Engleman, E.G. (1987). pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. Cell *49*, 659-668.

Stenglein, M.D., and Harris, R.S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. The Journal of biological chemistry *281*, 16837-16841.

Stevenson, M. (2003). HIV-1 pathogenesis. Nature medicine *9*, 853-860.

Strack, B., Calistri, A., Craig, S., Popova, E., and Gottlinger, H.G. (2003). AIP1/ALIX is a binding partner for HIV-1 p6 and EIAV p9 functioning in virus budding. Cell *114*, 689-699.

Strebel, K., Klimkait, T., Maldarelli, F., and Martin, M.A. (1989). Molecular and biochemical analyses of human immunodeficiency virus type 1 vpu protein. Journal of virology *63*, 3784-3791.

Strebel, K., Klimkait, T., and Martin, M.A. (1988). A novel gene of HIV-1, vpu, and its 16-kilodalton product. Science (New York, NY *241*, 1221-1223.

Stremlau, M., Owens, C.M., Perron, M.J., Kiessling, M., Autissier, P., and Sodroski, J. (2004). The cytoplasmic body component TRIM5alpha restricts HIV-1 infection in Old World monkeys. Nature *427*, 848-853.

Stremlau, M., Perron, M., Lee, M., Li, Y., Song, B., Javanbakht, H., Diaz-Griffero, F., Anderson, D.J., Sundquist, W.I., and Sodroski, J. (2006). Specific recognition and accelerated uncoating of retroviral capsids by the TRIM5alpha restriction factor. Proceedings of the National Academy of Sciences of the United States of America *103*, 5514-5519.

Stremlau, M., Perron, M., Welikala, S., and Sodroski, J. (2005). Species-specific variation in the B30.2(SPRY) domain of TRIM5alpha determines the potency of human immunodeficiency virus restriction. Journal of virology *79*, 3139-3145.

Sugimoto, J., Matsuura, N., Kinjo, Y., Takasu, N., Oda, T., and Jinno, Y. (2001). Transcriptionally Active HERV-K Genes: Identification, Isolation, and Chromosomal Mapping. Genomics *72*, 137-144.

Suspene, R., Rusniok, C., Vartanian, J.-P., and Wain-Hobson, S. (2006). Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. Nucl Acids Res *34*, 4677-4684.

Tanaka, Y., Marusawa, H., Seno, H., Matsumoto, Y., Ueda, Y., Kodama, Y., Endo, Y., Yamauchi, J., Matsumoto, T., Takaori-Kondo, A., *et al.* (2006). Anti-viral protein APOBEC3G is induced by interferon-alpha stimulation in human hepatocytes. Biochemical and biophysical research communications *341*, 314-319.

Tarlinton, R., Meers, J., Hanger, J., and Young, P. (2005). Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. The Journal of general virology *86*, 783-787.

Tarlinton, R.E., Meers, J., and Young, P.R. (2006). Retroviral invasion of the koala genome. Nature *442*, 79-81.

Telesnitsky, A., and Goff, S.P. (1997). Reverse Transcriptase and the Generation of Retroviral DNA. In Retroviruses, J.M. Coffin, S.H. Hughes, and H.E. Varmus, eds. (Cold Spring Harbor, Cold Spring Harbor Laboratory Press), pp. 121-160.

Temin, H.M., and Mizutani, S. (1970). Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. Nature *226*, 1211-1213.

Terwilliger, E.F., Cohen, E.A., Lu, Y.C., Sodroski, J.G., and Haseltine, W.A. (1989). Functional role of human immunodeficiency virus type 1 vpu. Proceedings of the National Academy of Sciences of the United States of America *86*, 5163-5167.

Tindall, B., Evans, L., Cunningham, P., McQueen, P., Hurren, L., Vasak, E., Mooney, J., and Cooper, D.A. (1992). Identification of HIV-1 in semen following primary HIV-1 infection. AIDS (London, England) *6*, 949-952.

Tonjes, R.R., Czauderna, F., and Kurth, R. (1999). Genome-Wide Screening, Cloning, Chromosomal Assignment, and Expression of Full-Length Human Endogenous Retrovirus Type K. J Virol *73*, 9187-9195.

Tonjes, R.R., Limbach, C., Lower, R., and Kurth, R. (1997). Expression of human endogenous retrovirus type K envelope glycoprotein in insect and mammalian cells. J Virol *71*, 2747-2756.

Towers, G.J., Hatziioannou, T., Cowan, S., Goff, S.P., Luban, J., and Bieniasz, P.D. (2003). Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. Nature medicine *9*, 1138-1143.

Tristem, M. (2000). Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. Journal of virology *74*, 3715-3730.

Tumpey, T.M., Basler, C.F., Aguilar, P.V., Zeng, H., Solorzano, A., Swayne, D.E., Cox, N.J., Katz, J.M., Taubenberger, J.K., Palese, P., *et al.* (2005). Characterization of the reconstructed 1918 Spanish influenza pandemic virus. Science *310*, 77-80.

Turelli, P., Mangeat, B., Jost, S., Vianin, S., and Trono, D. (2004a). Inhibition of hepatitis B virus replication by APOBEC3G. Science (New York, NY *303*, 1829.

Turelli, P., Mangeat, B., Jost, S., Vianin, S., and Trono, D. (2004b). Inhibition of Hepatitis B Virus Replication by APOBEC3G. Science *303*, 1829-.

Turner, G., Barbulescu, M., Su, M., Jensen-Seaman, M.I., Kidd, K.K., and Lenz, J. (2001). Insertional polymorphisms of full-length endogenous retroviruses in humans. Current Biology *11*, 1531.

Turville, S.G., Arthos, J., Donald, K.M., Lynch, G., Naif, H., Clark, G., Hart, D., and Cunningham, A.L. (2001). HIV gp120 receptors on human dendritic cells. Blood *98*, 2482-2488.

UNAIDS (2008). 2008 Report on the global AIDS epidemic.

Van Beveren, C., Galleshaw, J.A., Jonas, V., Berns, A.J., Doolittle, R.F., Donoghue, D.J., and Verma, I.M. (1981). Nucleotide sequence and formation of the transforming gene of a mouse sarcoma virus. Nature *289*, 258-262.

Van Damme, N., Goff, D., Katsura, C., Jorgenson, R.L., Mitchell, R., Johnson, M.C., Stephens, E.B., and Guatelli, J. (2008a). The Interferon-Induced Protein BST-2 Restricts HIV-1 Release and Is Downregulated from the Cell Surface by the Viral Vpu Protein. Cell Host and Microbe *3*, 245-252.

Van Damme, N., Goff, D., Katsura, C., Jorgenson, R.L., Mitchell, R., Johnson, M.C., Stephens, E.B., and Guatelli, J. (2008b). The interferon-induced protein BST-2 restricts HIV-1 release and is downregulated from the cell surface by the viral Vpu protein. Cell host & microbe *3*, 245-252.

van de Lagemaat, L., Medstrand, P., and Mager, D. (2006). Multiple effects govern endogenous retrovirus survival patterns in human gene introns. Genome Biology *7*, R86.

Vartanian, J.-P., Guetard, D., Henry, M., and Wain-Hobson, S. (2008a). Evidence for Editing of Human Papillomavirus DNA by APOBEC3 in Benign and Precancerous Lesions. Science *320*, 230-233.

Vartanian, J.P., Guetard, D., Henry, M., and Wain-Hobson, S. (2008b). Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. Science (New York, NY *320*, 230-233.

Vartanian, J.P., Meyerhans, A., Asjo, B., and Wain-Hobson, S. (1991). Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. J Virol *65*, 1779-1788.

Varthakavi, V., Smith, R.M., Bour, S.P., Strebel, K., and Spearman, P. (2003). Viral protein U counteracts a human host cell restriction that inhibits HIV-1 particle production. Proceedings of the National Academy of Sciences of the United States of America *100*, 15154-15159.

Verderame, M.F., Nelle, T.D., and Wills, J.W. (1996). The membrane-binding domain of the Rous sarcoma virus Gag protein. Journal of virology *70*, 2664-2668.

VerPlank, L., Bouamr, F., LaGrassa, T.J., Agresta, B., Kikonyogo, A., Leis, J., and Carter, C.A. (2001). Tsg101, a homologue of ubiquitin-conjugating (E2) enzymes, binds the L domain in HIV type 1 Pr55(Gag). Proceedings of the National Academy of Sciences of the United States of America *98*, 7724-7729.

Villesen, P., Aagaard, L., Wiuf, C., and Pedersen, F. (2004). Identification of endogenous retroviral reading frames in the human genome. Retrovirology *1*, 32.

Virgen, C.A., Kratovac, Z., Bieniasz, P.D., and Hatziioannou, T. (2008). Independent genesis of chimeric TRIM5-cyclophilin proteins in two primate species. Proceedings of the National Academy of Sciences of the United States of America *105*, 3563-3568.

Vodicka, M.A., Koepp, D.M., Silver, P.A., and Emerman, M. (1998). HIV-1 Vpr interacts with the nuclear transport pathway to promote macrophage infection. Genes & development *12*, 175-185.

von Schwedler, U., Song, J., Aiken, C., and Trono, D. (1993). Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. Journal of virology *67*, 4945-4955.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet *20*, 116-117.

Wang, G.P., Ciuffi, A., Leipzig, J., Berry, C.C., and Bushman, F.D. (2007). HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res *17*, 1186-1194.

Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A., and Batzer, M.A. (2005). SVA Elements: A Hominid-specific Retroposon Family. Journal of Molecular Biology *354*, 994-1007.

Weinberg, J.B., Matthews, T.J., Cullen, B.R., and Malim, M.H. (1991). Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. The Journal of experimental medicine *174*, 1477-1482.

Weiss, R.A. (2006). The discovery of endogenous retroviruses. Retrovirology *3*, 67.

Wiegand, H.L., Doehle, B.P., Bogerd, H.P., and Cullen, B.R. (2004). A second human antiretroviral factor, APOBEC3F, is suppressed by the HIV-1 and HIV-2 Vif proteins. EMBO Journal *23*, 2451-2458.

Wills, J.W., Cameron, C.E., Wilson, C.B., Xiang, Y., Bennett, R.P., and Leis, J. (1994). An assembly domain of the Rous sarcoma virus Gag protein required late in budding. Journal of virology *68*, 6605-6618.

Wilson, S.J., Webb, B.L., Ylinen, L.M., Verschoor, E., Heeney, J.L., and Towers, G.J. (2008). Independent evolution of an antiviral TRIMCyp in rhesus macaques. Proceedings of the National Academy of Sciences of the United States of America *105*, 3557-3562.

Wu, X., Anderson, J.L., Campbell, E.M., Joseph, A.M., and Hope, T.J. (2006). Proteasome inhibitors uncouple rhesus TRIM5alpha restriction of HIV-1 reverse transcription and infection. Proceedings of the National Academy of Sciences of the United States of America *103*, 7465-7470.

Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration. Science *300*, 1749-1751.

Wu, Y., and Marsh, J.W. (2001). Selective Transcription and Modulation of Resting T Cell Activity by Preintegrated HIV DNA. Science *293*, 1503-1506.

Xiang, Y., Cameron, C.E., Wills, J.W., and Leis, J. (1996). Fine mapping and characterization of the Rous sarcoma virus Pr76gag late assembly domain. Journal of virology *70*, 5695-5700.

Yamauchi, K., Wada, K., Tanji, K., Tanaka, M., and Kamitani, T. (2008). Ubiquitination of E3 ubiquitin ligase TRIM5 alpha and its potential role. The FEBS journal *275*, 1540-1555.

Yanez-Munoz, R.J., Balaggan, K.S., MacNeil, A., Howe, S.J., Schmidt, M., Smith, A.J., Buch, P., MacLaren, R.E., Anderson, P.N., Barker, S.E., *et al.* (2006). Effective gene therapy with nonintegrating lentiviral vectors. Nat Med *12*, 348-353.

Yang, J., Bogerd, H., Le, S.Y., and Cullen, B.R. (2000). The human endogenous retrovirus K Rev response element coincides with a predicted RNA folding region. RNA *6*, 1551-1564.

Yang, J., Bogerd, H.P., Peng, S., Wiegand, H., Truant, R., and Cullen, B.R. (1999). An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. PNAS *96*, 13404-13408.

Yang, Y., Guo, F., Cen, S., and Kleiman, L. (2007). Inhibition of initiation of reverse transcription in HIV-1 by human APOBEC3F. Virology *365*, 92.

Yap, M.W., Nisole, S., Lynch, C., and Stoye, J.P. (2004). Trim5alpha protein restricts both HIV-1 and murine leukemia virus. Proceedings of the National Academy of Sciences of the United States of America *101*, 10786-10791.

Yap, M.W., Nisole, S., and Stoye, J.P. (2005). A single amino acid change in the SPRY domain of human Trim5alpha leads to HIV-1 restriction. Curr Biol *15*, 73-78.

Yasuda, J., Hunter, E., Nakao, M., and Shida, H. (2002). Functional involvement of a novel Nedd4-like ubiquitin ligase on retrovirus budding. EMBO reports *3*, 636-640.

Yi, J.M., Kim, T.H., Huh, J.W., Park, K.S., Jang, S.B., Kim, H.M., and Kim, H.S. (2004). Human endogenous retroviral elements belonging to the HERV-S family from human tissues, cancer cells, and primates: expression, structure, phylogeny and evolution. Gene *342*, 283-292.

Yoshida, M., Miyoshi, I., and Hinuma, Y. (1982). Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proceedings of the National Academy of Sciences of the United States of America *79*, 2031-2035.

Yoshinaka, Y., Katoh, I., Copeland, T.D., and Oroszlan, S. (1985). Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proceedings of the National Academy of Sciences of the United States of America *82*, 1618-1622.

Yu, Q., Chen, D., Konig, R., Mariani, R., Unutmaz, D., and Landau, N.R. (2004a). APOBEC3B and APOBEC3C are potent inhibitors of simian immunodeficiency virus replication. The Journal of biological chemistry *279*, 53379-53386.

Yu, Q., Chen, D., Konig, R., Mariani, R., Unutmaz, D., and Landau, N.R. (2004b). APOBEC3B and APOBEC3C Are Potent Inhibitors of Simian Immunodeficiency Virus Replication. J Biol Chem *279*, 53379-53386.

Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., and Landau, N.R. (2004c). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol *11*, 435.

Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J.M., and Landau, N.R. (2004d). Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. Nat Struct Mol Biol *11*, 435-442.

Yu, S.F., Eastman, S.W., and Linial, M.L. (2006). Foamy virus capsid assembly occurs at a pericentriolar region through a cytoplasmic targeting/retention signal in Gag. Traffic (Copenhagen, Denmark) *7*, 966-977.

Yu, S.F., Sullivan, M.D., and Linial, M.L. (1999). Evidence that the human foamy virus genome is DNA. Journal of virology *73*, 1565-1572.

Yu, X., Yu, Y., Liu, B., Luo, K., Kong, W., Mao, P., and Yu, X.-F. (2003). Induction of APOBEC3G Ubiquitination and Degradation by an HIV-1 Vif-Cul5-SCF Complex. Science *302*, 1056-1060.

Yu, Y., Xiao, Z., Ehrlich, E.S., Yu, X., and Yu, X.-F. (2004e). Selective assembly of HIV-1 Vif-Cul5-ElonginB-ElonginC E3 ubiquitin ligase complex through a novel SOCS box and upstream cysteines. Genes & Development *18*, 2867-2872.

Zennou, V., and Bieniasz, P.D. (2006). Comparative analysis of the antiretroviral activity of APOBEC3G and APOBEC3F from primates. Virology *349*, 31.

Zennou, V., Perez-Caballero, D., Gottlinger, H., and Bieniasz, P.D. (2004). APOBEC3G incorporation into human immunodeficiency virus type 1 particles. Journal of virology *78*, 12058-12061.

Zhang, F., Wilson, S.J., Landford, W.C., Virgen, B., Gregory, D., Johnson, M.C., Munch, J., Kirchhoff, F., Bieniasz, P.D., and Hatziioannou, T. (2009). Nef Proteins from Simian Immunodeficiency Viruses Are Tetherin Antagonists. Cell host & microbe.

Zhang, H., Yang, B., Pomerantz, R.J., Zhang, C., Arunachalam, S.C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature *424*, 94-98.

Zhou, W., Parent, L.J., Wills, J.W., and Resh, M.D. (1994). Identification of a membrane-binding domain within the amino-terminal region of human immunodeficiency virus type 1 Gag protein which interacts with acidic phospholipids. Journal of virology *68*, 2556-2569.

Zsiros, J., Jebbink, M.F., Lukashov, V.V., Voute, P.A., and Berkhout, B. (1998). Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses. J Gen Virol *79*, 61-70.

Zsíros, J., Jebbink, M.F., Lukashov, V.V., Voûte, P.A., and Berkhout, B. (1999). Biased Nucleotide Composition of the Genome of HERV-K Related Endogenous Retroviruses and Its Evolutionary Implications. Journal of Molecular Evolution *48*, 102-111.