The Expression of Human Endogenous Retroviruses is modulated by the Tat protein of HIV-1

by

Marta Jeannette Gonzalez-Hernandez

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Immunology)
in The University of Michigan
2012

Doctoral Committee

Professor David M. Markovitz, Chair
Professor Gary Huffnagle
Professor Michael J. Imperiale
Associate Professor David J. Miller
Assistant Professor Akira Ono
Assistant Professor Christiane E. Wobus

For my family and friends, the most fantastic teachers I have ever had.

Acknowledgements

First, and foremost, I would like to thank David Markovitz for his patience and his scientific and mentoring endeavor. My time in the laboratory has been an honor and a pleasure.

Special thanks are also due to all the members of the Markovitz laboratory, past and present. It has been a privilege, and a lot of fun, to work near such excellent scientists and friends. You all have a special place in my heart.

I would like to thank all the members of my thesis committee for all the valuable advice, help and jokes whenever needed.

Our collaborators from the Bioinformatics Core, particularly James Cavalcoli, Fan Meng, Manhong Dai, Maureen Sartor and Gil Omenn gave generous support, technical expertise and scientific insight to a very important part of this project. Thank you.

Thanks also go to Mariana Kaplan's and Akira Ono's laboratory for help with experimental designs and for being especially generous with time and reagents.

Table of Contents

List of Figures

List of Tables

CHAPTER I

INTRODUCTION

HUMAN ENDOGENOUS RETROVIRUS TYPE K (HML-2) AND HIV-1 INFECTION

Part I. Retroviruses and their general properties

Vertebrate Retroviruses are a type of infectious agent belonging to the *Retroviridae* family of viruses, and comprising a diverse and growing range of microorganisms. They are a specialized type of virus, with its genetic material being comprised of Ribonucleic Acid (RNA), instead of Deoxyribonucleic Acid (DNA), and characterized by two particularly interesting and biologically important features:

1) The ability to convert information from genomic viral RNA into DNA;

2) The ability to integrate a DNA copy of the viral RNA into the host genome [1, 2].

Indeed, retroviruses get their name from the Latin *retro* (meaning backwards or reverse), due to their ability to reverse what is thought of as the central dogma of molecular biology, in that their genetic information goes from RNA to DNA to RNA to protein, instead of from DNA to RNA to protein. Additionally, through the use of their ability for genomic integration, these viruses can replicate as cellular genes, taking permanent residence in the genetic

material of the infected host. This particular feature can cause detrimental, as well as beneficial, consequences for the host.

Although it was not until 1974 that retroviruses gained their name, the diseases that they cause had been observed much earlier [3, 4]. From as early as the 19th century, Bovine leukosis and Jaagsiekte in sheep were diseases afflicting animals later found to be caused by retroviruses [4, 5]. Later on, from the early 1900s up to 1914, different groups showed that filtrates could transmit infection of equine anemia (Vallée and Carré in 1904), or transfer sarcoma in chickens (Peyton Rous in 1911; Fujinami and Inamoto in 1914). Even prior to the Rous finding, in 1908, Bang and Ellerman showed that a filtrate transmitted chicken erythroleukemia, though the data were largely ignored until the 1930s when leukemia became regarded as cancerous [6-8]. At the time, these illnesses were thought to be transmitted by extracts, fluids or aerosols with no thought of viruses as a cause. The question of whether the agent was a "living fluid" or a particle was still open. These data were the first pieces of proof that what we now know as *retroviruses* were indeed transmissible agents that had the capacity to affect cellular host replication, and could in fact cause cancer. The first tumor virus discovered was later called the Rous Sarcoma Virus (RSV, found by Peyton Rous, 1911), shown in 1961 to contain RNA in its particles [8] and later demonstrated to be a retrovirus. Thus retroviruses were first called oncogenic viruses (Oncoviruses) or RNA tumor viruses. Several other cancer-causing retroviruses have since been described, and throughout the years many of their interesting characteristics have been ascertained.

One of the interesting and characteristic properties of retroviruses was discovered through the observation that cells that were transformed by RSV maintained a similar and stable phenotype throughout many cycles of division and replication. This phenotype induced by the virus showed a certain heritability that remained even in the absence of viral replication [9]. Thus, this observation led Howard Temin in the 1960s to postulate that in the infected cell the virus must be making a stable copy of itself and integrating it into the host chromosomal DNA [10, 11]. Temin called his concept the DNA provirus hypothesis [10, 11], as an analogy with the integrated prophage of the temperate bacteriophage. The integration of viral DNA into cellular DNA had already been suggested by André Lwoff (who discovered prophages and lysogeny) [12], and the concept was later confirmed experimentally in 1968 [13]. However, how an RNA virus transforms its RNA genome into DNA prior to integration into the host cell was not understood until, in 1965, Howard Temin described the process: ***Reverse Transcription***. A virus whose RNA genome was *reverse transcribed* into complementary DNA (cDNA), then integrated into the host's genome and expressed from that template. The viral enzyme Reverse Transcriptase was then independently described in 1970 by Howard Temin and David Baltimore [14, 15]. These findings led to the understanding that some diseases with previously unknown etiologies were caused by these sorts of viruses, and slowly led to the discovery in 1980 of the first human retrovirus: Human T-cell Leukemia Virus (HTLV) [16], which causes adult T-cell leukemia. In 1983-4, another human.

retrovirus, Human Immunodefiency Virus (HIV), was discovered independently by two researchers [17, 18].

We now know that mammalian retroviruses are a large and complex group of viruses that not only have the potential to cause tumors but are also associated with immunodeficiency, autoimmunity and neurological disorders [19-22]. These characteristics have thus prompted the term Oncovirus to be replaced by Retrovirus, since it more broadly describes their main distinguishing factor, and does not associate their pathogenesis only with cancer. This term is also more accurate as, of course, other viruses that are not retroviruses have been shown to be oncogenic.

A. Structure of Retroviruses

a. General Properties

The structure of the retroviral particle, or virion, reflects the requirements needed to establish an efficient infection that can conclude in successful propagation of the virus in the infected host (Figure 1.1). These requirements are summarized as follows: successful incorporation of the viral genetic material into stable particles, the

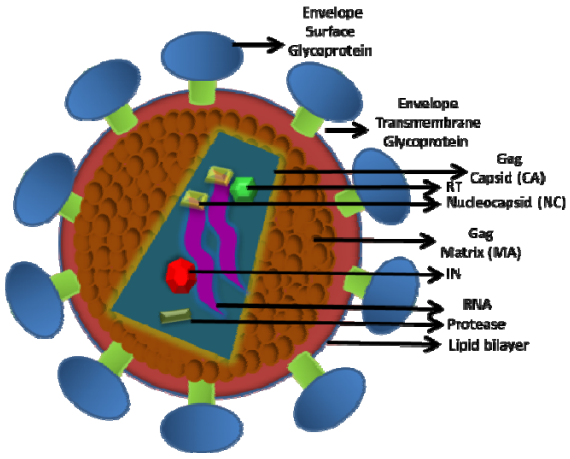


Figure 1.1 Basic structure of the retroviral virion. The retroviral particle is surrounded by a lipid bilayer derived from the host cell, which contains the Envelope glycoproteins. Inside this bilayer are two protein shells: the viral Gag Matrix protein that surrounds the Gag Capsid protein. Encapsidated in the protein core is the viral genome (as 2 single stranded RNA molecules), the viral Reverse Transcriptase enzyme (RT), the viral Protease, the Nucleocapsid (NC) and the viral Integrase enzyme (IN).

recognition of, and entry into, the appropriate host cells, successful replication and integration of the viral genome, the translation of viral messenger RNA to yield new viral proteins, and assembly and packaging of viral proteins. Structural information for retroviruses has shown them to be roughly spherical, enveloped, and containing a complex group of proteins. These proteins carry out several functions for the virion: 1) the condensation of the genome into an RNA-protein complex; 2) the encapsidation of this complex in a protein shell; 3) the enclosure of the shell in a lipid membrane, or envelope; 4) the addition of surface proteins to the envelope for recognition of cellular receptors; and 5) the copying of the RNA in the newly infected cell. Two or more proteins can share each function, or one protein can carry out multiple activities [23].

Retroviral virions are 80-150 nanometers in diameter, and they contain a diploid positive-sense RNA genome. Their life cycle is mediated by a DNA intermediate, which is a DNA molecule complementary to the genomic RNA (i.e. a cDNA or complementary DNA). This occurs as an intermediate step attributed to the presence of an RNA-dependent DNA polymerase called “Reverse Transcriptase” (RT). This enzyme uses a single stranded RNA template to produce a double stranded DNA copy, and this DNA copy is later integrated into the genome by the virally-encoded Integrase (IN) enzyme [23, 24]. Retroviruses are composed of an outer envelope made of a lipid bilayer, which is derived from the membrane of infected cells as the virus buds off when it finalizes its infection cycle [23, 24]. Virally-encoded glycoproteins are embedded in this lipid bilayer envelope, and their numbers vary greatly amongst different retroviruses and viral

isolates. These proteins play a significant role in the initial stage of infection since they are involved in interacting with cellular receptors, therefore being important in the determination of host range and of target cell and tissue tropism within an infected host.

All retroviruses can, additionally, be categorized based upon their genome structure as *simple retroviruses* or as *complex retroviruses*. Simple retroviruses (e.g. Avian Leukemia Virus, ALV) carry just four core genes (*gag, pro, pol* and *env*) while complex retroviruses (e.g. HIV-1) contain the same core genes in addition to a number of accessory genes responsible for aiding and facilitating replication [23, 24].

b. Genomic Composition

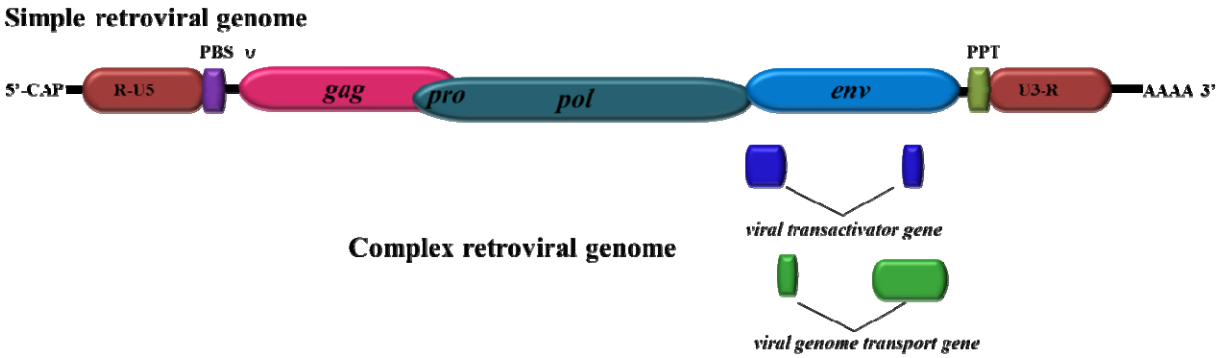


Figure 1.2 Organization of the retroviral genome. Simple retroviral genomes contain two partial Long Terminal Repeats (composed of R-U5 at the 5'-end, U3-R at the 3'-end) surrounding four core genes (*gag*, *pro*, *pol* and *env).* Complex retroviral genomes have additional accessory genes. A Primer Binding Site (PBS) and a Polypurine Tract (PPT) are important sequences for reverse transcription. Ψ is the packaging signal for incorporation of the viral genomic RNA into viral particles.

Genome Organization

Retroviral RNA genomes are approximately 7-13 kilobases in length (including the 5'-cap and 3'-Poly-A tail), and the individual RNA strands are joined at the 5`-end by a self-complementary region called a dimer linkage structure [19, 23, 25]. The significance of having two copies is not clearly understood but it is thought to act as a biological buffer that compensates for the high rate of error of reverse transcription [25]. Retroviral genomes are organized as two partial Long Terminal Repeats (LTR), which are composed of U3, R and U5 regions and separated by the core genes (and accessory genes for complex retroviruses), followed by a poly-A tail added by cellular enzymes (Figure 1.2).

A description of each gene and gene segment is shown in Table 1.1 (modified from [25]). The provirus (the integrated form of the retroviral genome) will have all these same genes but with the full LTR, formed during reverse transcription.

Table 1.1. Retroviral Genomic Composition

Genes and functional elements	*Description*
R	20-300(+) bases in sequence. Repeated at both ends.
U5	75-200 bases. Unique sequence at the 5`-end, with transcription initiator signal. Involved in viral integration.
PBS	Primer binding site. Site for binding of cellular tRNA for reverse transcription.
Gag	Encodes for structural proteins Matrix, Capsid and Nucleocapsid.
Pol	Encodes the enzymes Reverse Transcriptase and Integrase.
Pro	Encodes the Protease, which cleaves the Gag polyprotein precursor. It can be encoded as part of *Gag* or as part of *gag-pol* (depending on the retrovirus).
Env	Encodes the surface and transmembrane components of the envelope protein.
PPT	Polypurine tract; a short stretch of A and G nucleotides. Resistant to degradation and used as a primer during cDNA synthesis, at the end of the RNA.
U3	Unique sequence at the 3`-end. Involved in gene expression.

Long terminal repeats (LTRs)

The viral protein coding genes are bracketed by two identical stretches of sequence termed the long terminal repeats (LTR), which can each be further separated into three separate domains: U3, R and U5 (Figure 1.3). The LTRs are exact copies of one another at opposing ends of the retroviral genome, playing significant roles in the initiation and transcription stages of the retroviral lifecycle [26]. U3 is derived from the

sequence unique to the 3'-end of the RNA, R is derived from a sequence repeated at both ends of the RNA, and U5 is derived from the sequence unique to the 5'-end of the RNA. The synthesis of the full LTR region occurs through the process of reverse transcription, when the enzyme "jumps" from one end of the template to the other[25]. The sizes of these three elements vary considerably among different retroviruses. The U3 region is similar to the U5 region in that it is a unique non-coding region, yet forms the 5'-end of the LTR and can be several hundreds to thousands of nucleotides in length. U3 contains most of the transcriptional control elements of the virus, which include the promoter proper, and multiple enhancer sequences responsive to cellular, and in some cases viral, transcriptional activator proteins [26]. This sequence also plays an important role in initiation and facilitation of replication via a number of binding sites for cellular transcription factors along its length, which are able to interact with a large number of host cell factors, cytokines and cellular transcription signals [26-28]. Consequently, the U3 region has a significant influence over the cell-tropism and tissue-specificity of the virus, and the dynamics between cellular factors and viral sequences are important in regulation of viral expression [27]. The U5 region is a unique non-coding region of 75-250 nucleotides in length that forms the 3' end of the LTR. The R, or repeated, region of the LTR contains the transcription initiation site in most retroviruses, and is crucial for formation of the full LTR sequence during reverse transcription [25]. The Primer Binding Site (PBS), which is located at the end of the U5 region, spans approximately 18 nucleotides and is complementary to the 3'-end of a specific cellular tRNA. This complementarity causes the tRNA to bind to this region, where it is used to prime DNA

synthesis by the viral reverse transcriptase. Each different group of viruses shares a specific tRNA primer for this function. As different types of retroviruses use different tRNAs as primers, a taxonomy system for classifying them based upon their use of tRNA primers is often used [26].

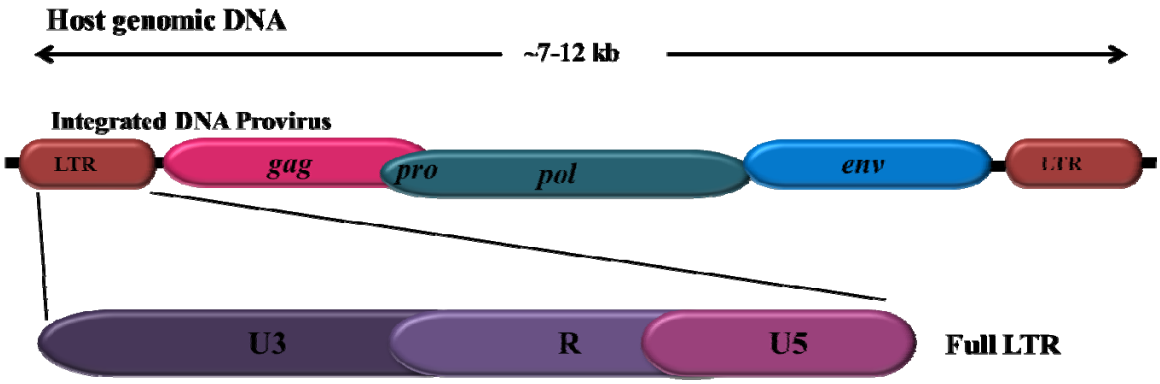


Figure 1.3 Proviral Genome and LTR. Organization of the integrated viral genome with LTR subdivisions is shown.

c. Core proteins

Gag Protein

The *gag* gene in retroviral genomes encodes the Gag protein, which is the precursor protein to the internal structure of all retroviruses. Its name stands for *Group-associated/specific Antigen* and is thus called because antibodies produced against it cross-react with similar proteins of related retroviruses. It is initially synthesized as an immature polyprotein. This Gag polyprotein self-assembles at the cell membrane to

form immature virions. During maturation it is proteolytically cleaved by the viral protease into the Matrix, Capsid, Nucleocapsid and several low mass cleavage products. Matrix, Capsid and Nucleocapsid all have the same basic amino acid organization from their amino to their carboxyl termini [23, 24]. For most retroviruses except HIV-1, the Matrix protein is the largest of the three domains and the most antigenic, and separates the viral envelope from the viral Capsid [20, 23]. The core contains the viral genetic material stabilized through binding with the Nucleocapsid (a small and basic protein) as a ribonucleoprotein complex, as well as the viral enzymes RT, protease and integrase [23] (Figure 1.1). The Nucleocapsid displays a high affinity for RNA and may play a role in helping to form and stabilize the RNA dimer of the diploid retroviral genome [29].

Pol Protein- Polymerase

The *pol* gene in retroviral particles encodes two important viral enzymes, is the most conserved of all retroviral genes, and is essential for viral replication [30]. It contains two coding domains, since it is a precursor for both the Reverse Transcriptase (RT) and Integrase (IN) proteins. Being an RNA-dependent DNA polymerase with RNaseH properties, the RT enzyme allows the degradation of DNA-RNA hybrid templates and is essential in ensuring the synthesis of DNA complementary to genomic viral RNA template (cDNA) [14]. Reverse transcription mostly occurs at a stage shortly after viral entry into a target cell (prior to entry into the nucleus), but can start in the viral particle [31-33], or even in the cell of origin (e.g. Spumaviruses) [34, 35]. After reverse transcription is finished, the IN enzyme participates in the insertion of the cDNA into the

host cell's chromosomal DNA [23]. After integration, the viral cDNA is referred to as proviral DNA, or "provirus".

Pro Protein- Protease

The Protease (PR or Pro) is an enzyme encoded between the *gag* and *pol* genes in the retroviral genome. It acts late in the viral life cycle, aiding in the assembly of Gag proteins through proteolytically cleaving the precursor polyprotein. The Protease is generally translated as a Gag-Pol fusion product produced by a ribosomal frameshift [23, 36, 37]. The enzyme is released from the polyprotein by an autocatalytic mechanism [38], and is active inside the viral particle.

Env Protein- Envelope

The last of the core genes in retroviruses is the *env* gene, which encodes a polyprotein that is cleaved to form two components of the viral envelope - the Surface (SU) protein and the Transmembrane (TM) protein [20, 23]. While the SU protein is primarily responsible for receptor binding and anchoring particles to the surface of cells, the TM protein is linked to cellular fusion and entry. Consequently, both play central roles for cell and receptor tropism.

Figure 1.1 shows the distribution of these core proteins in a typical exogenous retrovirus particle.

Retroviral accessory proteins

Products of retroviral accessory genes occur in the complex retroviruses, namely the lentiviruses [20, 23]. These products are usually not incorporated into virions (with exceptions occurring in primate lentiviruses), and are important for controlling transcription of the virus, immune evasion, and protein production. Examples of accessory genes include those coding for the Tat and Tax transactivator proteins from HIV-1 and HTLV-1, respectively, the HIV-1 Rev protein (involved in exporting viral RNAs), and the HIV-1 Vif, Vpu and Nef proteins.

B. Replication of retroviruses

As summarized in [20, 25], retroviral replication starts with the recognition of receptors on an appropriate target cell, followed by receptor-mediated membrane fusion. Within the cytoplasm, partial uncoating of the virus occurs and reverse transcription of the viral genomic RNA starts within the partially uncoated virus. Replication of the virus genome involves a step of reverse transcription (to synthesize negative-sense complementary DNA), followed by the synthesis of positive-sense complementary DNA (while RNA is degraded) and the formation of the LTRs. The viral cDNA migrates into the nucleus by varied mechanisms depending on retrovirus type, with most requiring the cell to be undergoing active cellular division with nuclear membrane breakdown (with the exception of lentiviral retroviruses). Within the nucleus, the cDNA integrates into the cellular genome through the use of the Integrase enzyme, forming a provirus. Full length viral RNA, and viral mRNA for protein

production, can be produced from this provirus by the cellular machinery. Newly produced viral RNA can be either spliced to generate mRNA encoding various viral proteins, or it migrates to the cytoplasm unspliced where it is encapsidated to generate a new virus particle (the only exception to this rule being Gag protein expression, as Gag is translated from unspliced viral RNA). Viral envelope proteins are incorporated in the cell membrane, and it is from this same cellular membrane that immature viral particles are subsequently released. This process leads to a virus with an outer envelope of cellular lipid bilayer, carrying with it additional viral polyproteins that were synthesized in the cytoplasm from viral mRNA. After budding, maturation of the virion occurs by cleavage of said precursor proteins [39]. Figure 1.4 shows a basic retroviral replication cycle.

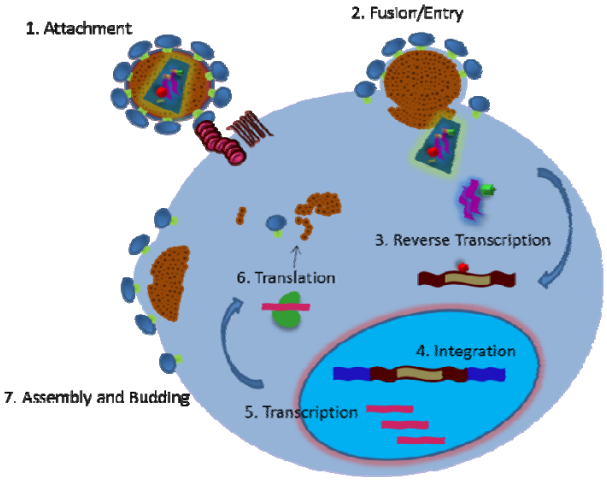


Figure 1.4 Basic retroviral lifecycle. The steps necessary for a successful retroviral infection are highlighted. Infection begins with viral attachment to cellular receptors, followed by fusion with the plasma membrane, uncoating of the capsid and reverse transcription, integration into the host genome, viral gene expression, particle assembly and subsequent viral release.

C. Retroviral classification

Retroviruses are classified into two subfamilies by the International Committee on Taxonomy of Viruses (ICTV): *Orthoretrovirinae* and *Spumaretrovirinae*. Classification is

mainly based upon both phylogenetic and sequence analyses of the RT gene (Table 1.2, see ICTV database http://ictvonline.org/index.asp, [40]). *Spumaretrovirinae* only has one genus (*Spumavirus),* whereas *Orthoretrovirinae* includes six genera: *Alpharetrovirinae*, *Betaretrovirinae*, and *Gammaretrovirinae,* considered simple retroviruses, and *Deltaretrovirinae*, *Epsilonretrovirinae* and *Lentivirinae*, described as complex (see ICTV database http://ictvonline.org/index.asp). The majority of these groups are found in piscine (*Epsilonretrovirinae*), avian (*Alpharetrovirinae*) and other nonprimate vertebrates. Amongst those groups previously reported in humans are the *Spumavirinae, Betaretrovirinae, Deltaretrovirinae* and the *Lentivirinae.*

In addition to the phylogenetic classification of retroviruses, further classification involves phenotypic characteristics (such as morphology), host organism, and the type of disease caused. Phenotypic characteristics, like viral core morphology, are used to differentiate retroviral particles. Morphology of retroviral cores divides particles into four main types: Types A, B, C, and D [39]. Type A particles represent the immature form of the virus, and are mainly seen in mutant viruses with defective proteolysis of viral protein precursors and in infectious viruses early in infection (prior to budding and maturation). They have a doughnut-like appearance by electron microscopy due to their electroluscent center, surrounded by one or two concentric rings. Mature viruses, however, display one of three morphological forms: Type B particles that have an acentrically placed, rounded and electron-dense core, Type C particles that have a centrally located, rounded and slightly angular core, and Type D particles that have a characteristic cylindrical or bar-shaped core [39]. All genera display Type C morphology, except three: *Betaretrovirinae,* which have B or D type morphologies; *Lentivirinae,* which have cylindrical or conical cores (Type D-like), and

Spumavirinae, which display immature morphological forms with characteristic central uncondensed cores.

Table 1.2. Classification of *Retroviridae* according to ICTV*

Subfamily	Genus	Host	Example
Orthoretrovirinae	*Alpharetrovirus*	Avian	Avian Leukemia Virus
	Betaretrovirus	Sheep, primates, mice	Mouse Mammary Tumor Virus
	Gammaretrovirus	Birds, mammals, reptiles	Murine Leukemia Virus
	Deltaretrovirus	Mammals	Human T Lymphotropic Virus
	Epsilonretrovirus	Fish and reptiles	Walleye Dermal Sarcoma Virus
	Lentivirus	Mammals	Human Immunodeficiency Virus Type 1
Spumaretrovirinae	*Spumavirus*	Mammals	Human Foamy Virus
Unclassified Retroviruses			Human Endogenous Retrovirus
			Murine Endogenous Retrovirus
			Avian Endogenous Retrovirus

Most of these viruses are widespread in mammals and cause different types of cytopathic effects. Some, like the Lentiviruses, cause infections with characteristic long incubation periods and potentially fatal diseases, and some have oncogenic potential.

It must be stressed, however, that classification of *Retroviridae* is still incomplete, with many retroviruses still unclassified. Some retroviruses, for example, that are still not fully classified include the Human Endogenous Retroviruses (HERVs), which will be the focus of this thesis. These are very interesting and peculiar retroviruses, that have managed to remain fixed in the human genome and become vertically transmitted from parent to progeny.

Part II. Human Endogenous Retroviruses (HERVs)

A. General properties of HERVs

Thanks to the efforts of the Human Genome Project we now know that just 1.5% of our DNA sequence codes for protein, while the remaining 98.5% is made up of so-called "junk" DNA [41]. "Junk" DNA usually refers to noncoding DNA, much of which has no known biological function. This type of DNA includes repetitive elements and sequences, such as *Alu* repeats, LINE and SINE elements (Long and Short Interspersed Elements, respectively), and Human Endogenous Retroviruses (HERVs), the latter of which makes up about 8-10% of our genome [4, 41-43]. Though some believe that endogenous retroviruses (ERVs) could have evolved from pre-existing genomic elements such as LTR-retrotransposons [44], ERVs rather appear to be the result of the interesting, and rare, successful infection events of germline cells that occurred throughout millions of years [2, 43, 45-47].

Generally, retroviruses are only able to infect somatic cells. When the cell is proliferating, the insertion is present at the same chromosomal location in all progeny cells and vanishes with the last cell of the clone, which might not occur until the host dies. Occasionally however, a retrovirus is able to infect a cell belonging to the germline. Any offspring developing from an infected germline cell carries the provirus in every single nucleated cell of the organism (at the same chromosomal position) and this proviral element is then passed to the descendants according to Mendelian inheritance laws. They are thus termed *endogenous* retroviruses in order to distinguish them from their exogenous counterparts [3]. Their transmission mode (from one host generation to the next) has been termed “vertical” to distinguish it from the “horizontal” spread (from somatic cell to somatic cell) of exogenous retroviruses.

Endogenous retroviruses were discovered in the late 1960s after experiments with RSV yielded a possible “endogenous” envelope for the transmission of the virus [4]. They closely resemble infectious retroviruses, and integrated into the genome of our ancestors at least 25 million years ago [42, 48], with just one single exception: HERV-K. While different HERV-Ks have integrated into the genome over millions of years, some have entered the human genome as recently as 200,00 years ago [43, 48]. ERVs are also detected in plants and in all vertebrates studied to date [42, 45], ranging from mice to humans [49].

Unless proviral insertions are disadvantageous to the host, ERVs are transmitted vertically to host progeny [2], increasing their frequency within the host population through mechanisms such as mutational drift and hitch-hiking [50]. Distribution of HERVs across the genome has been considered non-random, with chromosomes 4, 19, X and Y having the highest number of HERVs, with maximal concentrations identified on the Y

chromosome [43, 48]. HERVs were also reported to preferentially migrate across the genome to regions of chromosomes with increased levels of heterochromatin, thereby increasing their chances of being retained within the host genome [48, 51].

Retrotransposition or reinfection of the germline can generate further proviral insertions, augmenting the number of a particular lineage in the genome [2, 52, 53]. A rapid period of amplification follows before activity decreases as mutations start to accumulate within the open reading frames of various viral genes [2]. This decline is attributable to host selection mechanisms and the lack of a purifying selection within the retroviral coding regions [2]. The final point at which mutation accumulations lead to inability to express proteins is termed “extinction” resulting in intergenic DNA [54]. Although extinction is an inevitable occurrence for most lineages, some retain varied levels of expression for millions of years (e.g. HERV-K) [54]. This retention of coding ability may be due to mechanisms such as complementation *in trans* [55, 56], or if the initial colonization and subsequent amplifications conferred advantages upon the host [57-59]. In humans, ERVs comprise approximately 200,000 entities, including more than 200 full length proviruses (with greater than 8,300 elements containing *pol*-related sequences) and more than 3,000 of these with full or partial open reading frames [41, 43, 60]. While many HERVs are defective, a limited number retain the potential to produce viral products and, in the case of HERV-K, form viral particles [46, 61]. Some have been reported to be active, at least at a transcriptional level, within several tissues [46, 62, 63].

In contrast to humans, several animals contain ERVs highly related to exogenous retroviruses [64, 65] or endogenized forms of current exogenous retroviruses: mouse mammary tumour virus (MMTV), Jaagsiekte sheep retrovirus (JSRV), and Koala retrovirus (KoRV) [66-69]. The most striking example is represented by the koala, in which a retrovirus is currently undergoing the process of endogenization since it still displays many of the features of an exogenous virus (e.g. the ability to produce infectious viral particles and variability in proviral copy number and sequence) [68].

It is still unclear whether these endogenous retroviruses have been retained because they performed, or perform, a useful biological function. Indeed, some HERVs have been shown to encode physiologically important proteins (e.g. syncitin) [57]. However, even though their exact function in the human genome is not very well understood, they are still thought to play an important role in genomic modeling and plasticity through gene transposition, and in cellular protection from other retroviral infections [46].

B. Classification of HERVs

HERV families show broad diversity concerning their relation to exogenous retroviruses, copy number and expression. Having originated from their exogenous counterparts, HERVs share the same genetic structure and organization [47], so their classification follows that of exogenous retroviruses, with around 26 to 30 distinct families or lineages [1, 2, 48, 69, 70]. Less well-defined sequence comparisons,

interestingly, suggest that there may be well over 100 different HERV groups [71]. Most, though, are grouped into three classes (Table 1.3) [2, 47, 48, 72]. Of note, at the genetic level, identifiable common structures such as the 5' LTR, PBS, Gag, Pro, Pol, Env, PPT and 3' LTR may or may not be present in a HERV locus due to the accumulation of mutations and deletions throughout human evolution [47]. Therefore, this classification is based only on the sequence homology of their *pol* regions with that of the *pol* genes of exogenous gammaretroviruses, betaretroviruses and spumaviruses [2]. The earliest classification was based on the tRNA specificity of the PBS by adding the one-letter code for the specific amino acid to "HERV" [73, 74]. For example, one of the most studied HERVs, HERV-K, uses a lysine (K) tRNA as its specific primer. However, this classification system can cause confusion when two viruses use the same tRNA primer, so it is complemented by other ways to designate HERVs. To date, HERVs have not been included in the retrovirus taxonomy since most sequences are fragmented or incomplete.

Class I HERVs

HERVs belonging to the first class of human endogenous retroviruses, Class I, share sequence similarities with gammaretroviruses such as Murine Leukemia Virus (MLV). This class is the largest of the group, with approximately twenty members [75]. The majority of its members exhibit genomes organized as above. The most notable exceptions would be HERV-H (which lacks an *env* gene) and HERV-FRD (which possesses an additional *gag*) [75]. The viruses comprising this class also exhibit varying copy

numbers, ranging from one (ERV3) to several hundred, although this is dependent upon the family and selection criteria used [2]. The vast majority, if not almost all, of the members of Class I elements are highly defective; only some, like HERV-W, HERV-H and HERV-F, show evidence of intact, or nearly intact, open reading frames [76-80]. Three families within this class show homology with MLV and Baboon endogenous virus (BaEV) in the highly conserved *pol*, *gag* and *env* regions.

Class II HERVs

Class II elements show the greatest sequence homology to alpha- and betaretroviruses (particularly to MMTV) [47, 81], and are often called the "HERV-K superfamily" since most, if not all, of its members are composed of different HERV-Ks (though not all members possess a tRNA-Lysine primer binding site) [75]. All members of this class of retroviruses (both endogenous and exogenous) produce a Gag-Pol polyprotein via one or two ribosomal frameshifting sites [82], and many of these viruses share a short glycine-rich region related to the G-patch domain found in many RNA-binding proteins [83]. Viruses belonging to this class were initially identified through hybridization studies using the *pol* gene of MMTV and Syrian hamster Type A particles [84, 85]. One of the viruses belonging to this class has become the prototypic sequence for the HERV-K family, yet it lacks a 292 base pair fragment from its *env* gene [43, 84]. Subsequent studies using cross-hybridization have identified 9 subgroups within the HERV-K family, at the time termed HERV-K (NMWV 1 through 9) [86, 87]. This was followed up by a further study using degenerate primers to HERV-K10 *pol*, which identified 6 subgroups,

termed HERV-K (HML-1-6), with HML standing for Human Murine-Like, or Human MMTV-like [88]. HERV-K (HML-2) contains the prototypic sequence for the group: HERV-K108 [89]. The HERV-K (HML-2) subfamily has attracted recent attention due to its increased number of insertional polymorphisms [43, 90], and will be the focus of this thesis.

Class III HERVs

Originally described on the basis of their sequence similarity to spumaretroviruses [91], Class III HERVs are composed of only four families [75, 92] with the HERV-L family being the largest (copy number of 200-500), followed closely by HERV-S [2].

C. HERV expression within humans

HERV transcriptional activity occurs in all human cell and tissue types at different basal levels for those viruses that have intact genes, as well as for those with mutated cistrons [93-97]. It appears that HERV expression is heterogeneous, varying depending upon individual host factors like ethnicity [96-99], cell populations [93, 100, 101], and tissue types [60, 102, 103]. Endogenous retroviral elements, such as LTRs, can be involved in physiological processes such as those regulating transcription of genes like INSL4, β1,3-GT, Endothelin B Receptor, and Tissue-Specific Salivary Amylase [46, 104-108]. Additionally, expression of certain HERV proteins has important physiological

Table 1.3. Classification of HERVs*

HERV Family	tRNA Primer
Class I	
HERV-ADP	$tRNA^{Thr}$
HERV-E	$tRNA^{Glu}$
HERV-F	$tRNA^{Phe}$
HERV-F type b	$tRNA^{Phe}$
HERV-F type c	$tRNA^{Phe}$
HERV-FRD	$tRNA^{His}$
HERV-H	$tRNA^{His}$
HERV-H49C23	No LTR
HERV-I	$tRNA^{Ile}$
RRHERV-I	$tRNA^{Ile}$
HERV-P	$tRNA^{Pro}$
HERV-R (really ERV-3)	$tRNA^{Arg}$
HERV-R type b	$tRNA^{Arg}$
HERV-T	$tRNA^{Thr}$
HERV-W	$tRNA^{Trp}$
HERV-XA	$tRNA^{Phe}$
HERV-Z69907	Not Determined
ERV-9	$tRNA^{Arg}$
Class II	
HERV-K (HML-1-4)	$tRNA^{Lys}$
HERV-K (HML-5)	$tRNA^{Ile}$
HERV-K (HML-6)	$tRNA^{Lys}$
HERV-K (HML-7-9)	Not Determined
Class III	
HERV-L	$tRNA^{Leu}$
HERV-S	$tRNA^{Ser}$
HERV-U2	Not Determined
HERV-U3	Not Determined

*Modified from [2]

functions, such as in placental development [58, 109-111], and may also provide mechanisms for protecting against exogenous virus infection [46, 112]. However, in general, how or why HERV genes are expressed, and the mechanisms responsible for expression, is not clearly understood. It is known that exogenous viral infections, viral transactivators, processes such as inflammation, chemical agents, cytokines, hormones, and stress conditions can contribute to the activation and transcription of transposable genetic elements, with HERV-K (HML-2) being an example of an activated HERV [113-131]. Interestingly, evidence has been building that specific families of HERVs are differentially regulated in different tissue and cell types [132-135].

Class I endogenous retroviruses appear more transcriptionally active in the skin, uterus and cervix than are Class II HERVs, which exhibit higher levels of transcription in Peripheral Blood Mononuclear cells (PBMCs), brain and mammary glands [60]. This differential expression may be due to the presence of different cellular transcription factors. Indeed, binding sites for such cellular factors are found along the length of the U3 LTR region, and these sites have been suggested to play central roles in the initiation, efficiency and regulation of retroviral replication [136-140]. Class II ERVs additionally differ from other groups as they are preferentially integrated into regions of increased gene density, specifically about 5-20 kilobases (kb) upstream of gene coding sequences [141]. Integrations into transcriptionally active areas of the genome such as these increases the likelihood of host gene disruption, supporting the idea that insertional inactivation of genes has strongly been selected against in the lineage leading to modern humans. However, some level of control must still exist so as not to damage

normal cellular gene expression. This can be seen in the fact that approximately 80% of HERVs in intragenic regions are in anti-sense orientation to the gene sequence [142]. One example where this is observed within the human genome is with regards to the human C4 gene, which is involved in the complement cascade and is important for the innate immune response [143].

Reports of HERVs in areas of gene transcription have lent credence to a new role for HERVs within the human genome, i.e. one of a regulatory nature contributing to tissue specific expression by providing perhaps promoter/enhancer sites for tissue-specific transcription factors. Documented examples supporting this hypothesis include Salivary amylase and HERV-E [107], HERV-K and *ISNL4* placental expression [104, 109], a HERV LTR driving Beta1,3-galactosyltransferase beta3Gal-T5 expression [105, 108] and an ERV9 LTR driving Alcohol Dehydrogenase 1C (ADH1C) expression [144]. Additionally, it is noteworthy that the majority of documented HERV insertions into gene coding regions, resulting in HERV driven tissue-specific expression, have been in genes concerned with energy metabolism. HERV transcriptional activity appears elevated in those tissues or cell types with higher rates of proliferation or metabolic activity, as opposed to highly specialized non-proliferative cells (e.g. muscle cells) [60].

Class III HERVs seem to have lower transcriptional activity levels [60], and this has been suggested to be the result of their elimination from gene-rich regions via purifying selection through evolution [145]. In fact, activity might be correlated to relative time of entrance into the human genome since the most transcriptionally active member of the HERVs, the HERV-K family, is also the most recent entrant [43].

Just like HERV transcript expression, HERV protein expression might also occur at a basal level and in a tissue-specific manner [96]. To date, this has been seen mainly in the *env* gene, which contains a full-length Open Reading Frame (ORF) in a limited number of HERVs [146]. Only HERV-K, a Class II member, has been found to contain ORFs for all of its proteins, though these are not necessarily expressed from the same provirus (there are at least 91 HERV-K copies scattered around the genome) [43]. Detectable HERV gene expression has been found to be involved in both human physiology and pathology [46].

HERVs can confer a selective advantage onto its host, resulting in a purifying selection keeping the intact provirus and its reading frames. An example of HERV coding regions that have been maintained by natural selection in the human genome are the HERV-W *env* gene (Syncytin-1), which is vital to the formation of the syncytiotrophoblast layer of the placenta [57, 77, 147]. Dysregulation or disruption of the syncytin protein causes defects in the cell fusion process that can lead to development of hypertensive disorders (e.g. preeclampsia and HELLP syndrome) [148]. The HERV *env* encoded protein may also play a role in immunosuppressing the maternal immune system (so it does not attack the fetus) [148, 149], and in the origins of the salivary amylase gene [150]. Other HERV proteins have shown immunosuppressive or immunomodulatory domains, including peptide domains in HERV-H *env* [151] and the HERV-E CKS–17 peptide [152]. Such immunomodulatory domains, both immunosuppressive and immunostimulatory (i.e. superantigens), have been suggested to play central roles in the development of aspects of the immune system, such as in peripheral tolerance in the thymus and the

evolution of the diverse specificity of the immune response [46, 153-155]. Expression of HERV sequences as mRNA transcripts or translated protein may also confer a selective advantage upon the host through protection against exogenous retroviruses [46, 156], as cellular resistance against infection of spleen necrosis virus has been demonstrated by transfection of cultured cells with an expression plasmid containing a HERV-W *env* insert [157].

As HERV sequences have the potential to be integrated in any location of the genome, they consequently may alter the structure and function of other genes. Therefore, they can be involved in genetic disorders through disruption of genes at their integration site and generation of truncated proteins or other isoforms through alternative splicing. For example, a chimeric cDNA clone of a fusion transcript containing the 5'-LTR sequences, and leader region, of a HERV-H element fused to the human c*albindin* gene (coding for a cytosolic calcium-binding protein) was identified in a cell line derived from a prostate metastasis [158]. The HERV-H element splices to the second exon of the *calbindin* gene, introducing a different leader peptide sequence in its place, and thus altering the calcium binding motif structure of the protein [158]. HERVs may even abnormally regulate the expression of particular transcripts by means of the newly inserted LTRs, or by undergoing fusion with cellular transcripts [159]. Therefore, epigenetic control has to notably be involved in silencing most of these genetic elements.

D. Pathological role of HERVs

In humans, exogenous retroviruses are known to cause a number of diseases (e.g. immunodeficiency, neurological disease and cancer). For HERVs, however, despite

the fact that in the past years a number of studies associating them with clinical conditions (Table 1.4) have emerged, none, so far, have shown any conclusive evidence that implicate HERVs specifically as causative agents of disease. Many associations with disease do exist, however, in the form of circulating anti-retroviral antibodies and increased levels of HERV gene expression in disease samples versus controls. Therefore, while endogenous retroviruses are firmly established pathogens in other species, HERVs could potentially be considered as emerging pathogens.

HERV families have been implicated in the pathogenesis of human diseases including melanoma [160], Type 1 diabetes [161], and autoimmune diseases [162] like Multiple Sclerosis [163-167], Rheumatoid Arthritis [168-173], Psoriasis [174-176] and Systemic Lupus Erythematosus [177, 178]. The exact mechanisms behind these ocurrances are unknown but have been hypothesized to involve violation of tolerance to HERV proteins or expression of superantigens [179-181]. In addition, there is an established correlation between the HLA-DQ locus with HERV-K (HML-6) LTRs (DQ-LTR3) as a haplotype in rheumatoid arthritis [182].

Recently, an increasing number of reports have suggested the association of HERVs, in particular HERV-H/F, ERV9 and HERV-W, with schizophrenia, bipolar disorder/depression and motor neuron disease [166, 167, 183-187]. HIV-1 infection has also been associated with increased HERV-K expression, where HERV-K RNA, protein and viral particles have been detected in the plasma of HIV-1-infected individuals and in HIV-1-associated lymphoma patients, but not in hepatitis C virus (HCV) infected patients or in seronegative controls [116, 131, 188-192]. The level of HERV-K titers was found to

correlate significantly with that of HIV-1, and to be high in HIV-1 patients with non-suppressive HAART (Highly Active Antiretroviral Therapy) and undetectable or low in HIV-1 patients with effective, suppressive HAART [116, 188, 189, 192]. The increase in HERV-K expression in HIV-1 patients might have a role in the disease pathogenesis and might also be related to HIV-associated cancers [116, 131, 191].

HERVs have also been associated with a variety of other diseases [63] (Table1.4). HERV-E mRNA was detected by reverse transcription PCR in 38.8% of prostatic cancer but not in normal specimens [193]. HERV-W mRNA encoding *env* gene product syncytin 1 was found to be over expressed in endometrial cancer specimens and is thought to play an important role in cell-cell fusion in endometrial carcinoma [194] and in cell-endothelial fusion in breast cancer [195]. In addition, HERV-K, HERV-E, and ERV-3 *env* mRNAs were detected in ovarian cancer and not in controls, while anti-HERV antibodies were also detected in a high percentage of patient sera (55%, 40%, and 30%, respectively) [196, 197]. HERV-K transcripts were detected at high levels in the plasma of patients with lymphoma and breast carcinoma, which decreased with treatment [131, 190]. Furthermore, HERV-K *rec*, which originates from *env* transcripts, has been detected in germ cell tumors and it was found that the induction of *rec* expression in transgenic mice induced developmental disturbances and carcinoma *in situ* [198], giving the first evidence that the induction of HERV-K *rec* expression might play an essential role in tumorigenesis. However, the mere expression of retroviral transcripts and proteins in these pathologies does not implicate a causal association of these viruses with disease, and their exact roles in these processes need to be further evaluated.

E. Human Endogenous Retrovirus Type K (HML-2)

Human Endogenous Retrovirus Type K, besides being the most recent entrant into the human genome [43, 98, 199], is also the most transcriptionally active of the endogenous retroviruses [43, 96, 200-203]. A series of HERV-K proviral sequences was first discovered using both MMTV *pol* [85, 204] and Syrian hamster intracisternal A-particle (IAP) *pol* probes [205]. The existence of HERV-K proviruses is restricted to the lineage of Old World monkeys and hominoids [205-207], with extensive conservation of *pol* and *env* gene sequences between monkeys, apes, and humans [43, 207]. The members of the HML-2 subgroup have been shown to be human specific [98], and genome-wide screening has revealed them to contain high levels of insertional polymorphisms [90, 208] demonstrated by the presence of polymorphic HERV-K proviruses in a portion of the human population. Only members of the HERV-K (HML-2) provirus family have open reading frames (ORFs) for most, if not all, viral genes, and therefore have received attention in relation to diseases and will be the main focus of this thesis [43, 96, 209, 210].

HERV-K (HML-2) proviruses exist as one of two forms (Type 1 and Type 2) differing only in a 292-bp deletion in the envelope gene in Type 1 viruses [43]. As of yet it has not been demonstrated that a single full-length HERV-K proviral clone (HML-2 or otherwise) has all ORFs for their viral genes intact. Mostly, viruses have one or two intact ORFs, and the rest of the genes are riddled with mutations. This, however, most interestingly in the case of the HML-2 family, does not diminish their mRNA expression:

normal tissues such as placenta and peripheral blood lymphocytes have low, steady-state levels of full-length mRNA expression [96]. Sensitive Reverse Transcriptase Polymerase Chain Reaction (RT-PCR)-based assays have also revealed spliced HERV-K (HML-2) transcripts in a variety of normal and tumor tissues [211]. In addition, HERV-K (HML-2) and related sequences appear to be expressed at different levels in several individuals who have been studied [97]. While it is still unclear whether a single full-length HERV-K (HML-2) genome with ORFs for all proteins exists, in some cell types (particularly malignant ones) HERV-K (HML-2) has retained the capacity to form retroviral-like particles. This is demonstrated in human teratocarcinoma, germ-cell tumours, melanoma, breast cancers and in megakaryocytes from patients with essential thrombocythemia [101, 201, 212-216]; in the plasma of lymphoma patients [131, 190], and possibly in the human placenta [217, 218]. However, to date, the infectivity of these retroviral-like particles, and if they originate from a single provirus, has yet to be determined. It might be that complementation *in trans* is occurring in these situations, leading to viral particle production.

Table 1.4. HERVs associated with human diseases.

Disease Type	HERV	Reference Number
Autoimmune		
Psoriasis	HERV-E, K, W, ERV-9	[175-177]
Opitz Syndrome	HERV-K, E	[145]
Rheumatoid Arthritis	HERV-K, L, ERV-3	[169-174, 183]
Essential thrombocytopenia	HERV-K10	[214]
Type 1 Diabetes	HERV-K18	[162]
Systemic Lupus Erythematosus	HERV-E, HERV-K	[178, 179, 215]
Multiple Sclerosis	HERV-W, HERV-F, HERV-H, ERV-3	[164-168], [219]
Congenital Heart Block	ERV-3	[220, 221]
Sjögren's Syndrome	HERV-K 113	[99, 222]
Mixed Connective Tissue Disease	HERV-E (4-1)	[223]
Disease Type	**HERV**	**Reference Number**
Neurological		
Bipolar disorder/Depression	HERV-K (HML-2), HERV-W	[187]
Schizophrenia	HERV-K (HML-2), ERV-9, HERV-W	[167, 168, 184-187]
Motor Neuron Disease	HERV-W	[188]
Disease Type	**HERV**	**Reference Number**
Cancer		
Ovarian	HERV-W, HERV-E	[197, 198]
Germ cell-Seminoma	HERV-K	[221]
Melanoma	HERV-K	[161, 213, 215, 226]
Prostate	HERV-E, HERV-K	[158, 159, 194] [193]
Lymphoma	HERV-K, ERV-9	[16, 191] [224]
Leukemias (CLL/CML/ALL)	HERV-K	[225]
Childhood Lymphoblastic Leukemia	Miscellaneous HERVs	[226]
Breast	HERV-K	[97, 123, 191, 194, 196,203]
Gastrointestinal Cancer	HERV-H	[227]
Stem cell myeloproliferative disorder	HERV-K	[228]
Disease Type	**HERV**	**Reference Number**
Miscellaneous		
Interstitial Lung Disease	HERV-E (4-1)	[229]

AIDS	HERV-K, Miscellaneous HERVs	[188, 189, 192, 230, 231]

Protein expression for these viruses has been detected in multiple circumstances, and corresponds to what would be expected with a betaretrovirus [96]. Gag proteins are formed as precursors of 76 kilodaltons (kDa), can be myristoylated, and are cleaved into major capsid, matrix, and nucleocapsid components [96, 232-236]. A HERV-K (HML-2) Protease can also be made, as is indicated by the presence of processed Gag proteins in teratocarcinoma cell lines [96, 237]. Envelope proteins can be made in some cell types (mostly melanomas, breast cancer and teratocarcinomas), and are usually synthesized as the full-length 80- to 90-kDa glycosylated precursor [96, 202, 203, 238, 239]. However, the precursor protein is not necessarily cleaved to the outer surface unit (SU) and transmembrane (TM) glycoproteins, although the consensus SU/TM cleavage site is present [96], and in any other cell types besides the ones mentioned above Env protein is rarely detected at significant levels. Insufficient production of Env could be another reason for the lack of apparent HERV-K infectivity. HERV-K (HML-2) also encodes the accessory oncogenes *np9* and *rec* (*nuclear protein of 9 kDa* and *regulator of expression encoded by cORF*) [240-243], which are expressed by either HERV-K (HML-2) Type 1 or Type 2, respectively. The HERV-K (HML-2) accessory proteins Rec and Np9 provide a potential link between HERV-K (HML-2) and oncogenesis [62, 126, 239, 240, 244-247]. The major function of the Rec protein (formerly called cORF, or also referred to as K-Rev) is in shuttling HERV-K RNA from the nucleus to the cytoplasm of a cell; it is an HIV-1-Rev-homologue protein that binds to a complex secondary structure in HERV-K (HML-2) unspliced or incompletely-spliced RNAs [242,

243, 248-252]. Both Rec and Np9 have been shown to stimulate c-Myc expression by binding and inhibiting the c-myc gene repressor Promyelocytic Leukemia Zinc-Finger Protein (PLZF; [246]), and Rec has also recently been shown to interact with the Testicular Zinc-Finger Protein, another transcriptional repressor [245]. Additionally, Rec overexpression leads to testicular carcinoma *in situ* in transgenic mice [198, 234, 253]. Np9 transcripts are detected with high frequency in tumor samples and, though no direct evidence exists that links it to oncogenesis, Np9 has been shown to interact with a member of the cancer-associated Notch signaling pathway [247]. Thus, increased expression of the HERV-K (HML-2) proteins Rec and Np9 has the potential to contribute to oncogenesis.

Antibodies against HERV-K (HML-2) have also been found in the blood of patients with a number of different clinical conditions [130, 210, 211, 239, 254, 255]. The humoral immune response against expressed HERV-K proteins has long been investigated through epitope mapping of Gag and Env proteins[211]. Antibodies were detected at very low titers in normal blood donors, in accordance with the finding that HERV-K mRNA is expressed at low levels in normal tissue samples (see above). When groups of patient sera were screened for antibodies, the survey revealed that an anti-HERV-K (HML-2) humoral response is high in leukemias, after pregnancies and, particularly, in patients suffering from testicular tumors [211, 234, 253, 256]. Yet, although the antibody titers are elevated compared to those of normal blood donors, they hardly ever reach the titers observed after retroviral infections such as human immunodeficiency virus-1 (HIV-1) [112, 116, 130, 131, 188-192, 257, 258]. One of the

highest percentages of antibodies against these retroviruses is seen in HIV-1 infected patients, where approximately 70% show a response against HERV-K (HML-2) antigens [121, 211, 239, 255]. Additionally, it has been shown that HERV-K antigens promote a T cell response against HIV-1 [112, 259], although Gag- and Env-specific T cell responses are infrequent [258].

Part III. Human Endogenous Retrovirus Type K and HIV-1 infection

A. Expression of HERV-K (HML-2) during HIV-1 infection

Infection of CD4 T-cell lines or stimulated peripheral blood mononuclear cells (PBMCs) with HIV-1 *in vitro* has been shown to upregulate the production of HERV-K (HML-2) RNA and proteins [116]. Additionally, plasma samples from HIV-1- infected individuals show significant levels of HERV-K (HML-2) RNA expression [112, 191] with titers as high as 10^6-10^{10}copies/mL [131, 188, 189]. HERV-K *pol* RNA was amplified in more than 95% of HIV-1-infected samples, as compared to only 5-8% of controls that included Hepatitis C virus-infected patients and healthy individuals [191]. HERV-K (HML-2) transcripts derived from multiple genomic loci were seen in the blood of HIV-1 infected patients, with transcripts from Type 2 proviruses located at chromosome 4q35 predominating, and with a minority of transcripts corresponding to HERV-K (HML-3) [131]. Additionally, HERV-K (HML-2) recombinant genomes were detected in plasma from some HIV-1 infected patients but not in breast cancer patients, and genetic

variation was observed in *env* (primarily synonymous substitutions indicative of purifying selection), with preservation of glycosylation sites together suggestive of replication through reverse transcription [131]. Specifically, HERV-K 102, a member of the HML-2 family, is quite significantly upregulated during HIV-1 infection [130], with approximately 76% of blood plasma samples of HIV-1 infected patients testing positive for particle associated HERV-K 102 *pol* transcripts, versus about 3% of healthy controls [130].

B. Mechanism of activation of HERV-K (HML-2) during HIV-1 infection

There are several potential ways that a viral infection could be causing the activation of HERV-K (HML-2) and its expression during disease, yet it is not clear what is the precise mechanism of HERV-K (HML-2) induction by HIV-1 infection. It is likely that a mix of both direct (e.g. HIV-1 proteins acting on HERV-K (HML-2) elements) and indirect (e.g. immune response to HIV-1) mechanisms contribute to the overall endogenous retrovirus activation. As an indirect mechanism of activation, cytokine responses (such as those triggered by IL-12, IL-1α, TNF-α and IFN-γ in response to inflammation, or as part of the innate response to exogenous viruses) could trigger signals that activate cellular responses and transcription factors that (randomly) increase HERV expression [122, 137, 140, 260-265]. HERV LTRs can exhibit distinct promoter/enhancer activities in different cell lineages, since they contain many potential transcription factor binding

sites [137, 140, 266-268], and cellular activation due to an inflammatory response or an antiviral response (to an exogenous virus) could cause changes in the pool of available transcription factors that could modulate HERV expression [137, 140, 261-265]. Transcription factor mobilization could thus result in binding to a HERV LTR and an increase in transcription from said HERV LTR. However, proviruses (e.g. HERVs, including the HERV-K (HML-2) family, as well as HIV-1 proviruses) are normally transcriptionally silenced by CpG methylation [269-271]. Yet it is possible that HERV expression in HIV-1-infected cells (or cells infected with any other exogenous virus) could be enhanced if methylation levels of proviral DNA are reduced by the exogenous infection. This way, potential transcription factors activated through cellular responses to infection would be able to bind the HERV LTR and activate transcription of endogenous retroviral genes. However, as of yet it has not been investigated whether HIV actively influences proviral methylation levels.

Another indirect mechanism of HERV-K (HML-2) upregulation by HIV-1 infection could be as the result of additional opportunistic viral infections. Since it is known that HIV-1 infection deregulates (and eventually destroys) the immune system, loss of immune control would facilitate the replication of any additional pathogen that might infect the host. Some of these pathogens, as is the case for Herpesviruses, are virtually ubiquitous in the human population and have the ability to remain latent, once they infect, during the lifetime of the host. Interestingly, Herpesviruses are reactivated from

latency during chronic HIV infection [272]. Epstein-Barr virus (EBV), which has a prevalence rate of around 75% in humans, induces the transcription of the *env* gene of HERV-K18 (which encodes a superantigen) , as do Human Herpesvirus (HHV) 6A and 6B [113, 181, 273, 274], and most likely other Human Herpesviruses are also able to activate HERV transcription. Some Herpesviruses, like HHV-6 and HHV-7, infect CD4 T-cells (as does HIV-1), and could thus (partly) be responsible for the HERV-K increases that are seen in PBMCs after HIV-1 infection. EBV commonly infects B-cells and could likewise contribute to the observed rise in PBMC HERV-K expression.

A more direct mechanism for activation of HERV-K (HML-2) expression during HIV-1 infection could result from the interaction of HIV-1 proteins with HERV-K (HML-2) elements. Expression of the HIV-1 accessory protein Vif is known to counteract the viral restriction factor APOBEC3G, which is known to exert control over retroviral elements (both endogenous and exogenous), and could result in abrogation of cellular HERV control and *de novo* replication of HERV elements by reverse transcription [275]. In fact, HERV-K (HML-2) elements have indeed been shown to be restricted during reverse transcription by APOBEC3G, both during *in vitro* replication as well as in the mutation pattern of ancient integrations [276]. Another essential protein for HIV-1 that might be involved in HERV-K (HML-2) activation is Rev, which assists in transporting unspliced and incompletely spliced viral mRNAs from the nucleus to the cytoplasm (a homolog protein in HTLV-I is Rex, in MMTV is Rem and in HERV-K (HML-2) it is Rec). The trans-acting Rev

protein binds to a viral RNA structure in Env-encoding sequences that is denoted the Rev responsive element (RRE). HIV-1 Rev can bind the HERV-K RRE (though the reverse is not true as Rec does not interact with the HIV-1 RRE) [248], and thus it is possible that HIV-1 Rev is allowing for increase in the transport and the accumulation of HERV-K (HML-2) RNAs in the cytoplasm of the cell that expresses both Rev and HERV-K (HML-2) elements.

Most interestingly, though, it is possible that during HIV-1 infection specific upregulation of endogenous proviruses could occur by stimulation of HERV-K transcription by an HIV-1 protein. For example, LTR-directed transcription of HERV-K (HML-2) can be induced by Herpes Simplex Virus-1 (HSV-1) infection [118]. The effect is mediated by the action of an HSV-1 immediate early protein ICP0, and requires the AP-1 binding site present on the HERV-K LTR [118]. The HTLV-I-encoded transactivator protein Tax is able to activate HERV LTRs, like HERV-K (HML-2), though its effect has mainly been shown on HERV-W and HERV-H [129]. These instances demonstrate that exogenous viral and retroviral transactivators can modulate endogenous retroviral expression. HIV-1 also contains a viral transactivator that regulates HIV-1 viral RNA expression. It is called Tat (for Transcriptional Transactivator), and this regulatory protein is a potent transactivator of the HIV promoter and is essential for viral replication [277]. Additionally, HIV-1 Tat has been shown to act not only on the HIV-1 and HIV-2 promoters, but also on other viral and cellular promoters [120, 278-281]. Most interestingly, the HIV-1 Tat protein has been shown to upregulate the

transcription of *Alu* repeat sequences (repetitive elements in the genome, as HERVs are), an effect mediated by the interaction of the transcription factor TFIIIC with the *Alu* promoter [120]. Therefore, Tat appears to be an interesting candidate HIV-1 protein that might activate HERV-K expression.

THESIS FOCUS AND HYPOTHESIS

This thesis will focus on HIV-1 Tat's role in the activation of HERV-K (HML-2) expression. In view of the above, we hypothesized that the HIV-1 Tat protein provides a functional link between HIV-1 infection and the induction of HERV-K (HML-2) gene expression by causing activation of HERV-K LTR-directed transcription. In the studies presented in this thesis, we show that Tat stimulates HERV-K expression in cell lines and primary lymphocytes. We further demonstrate that the HERV-K (HML-2) transcriptional promoter is responsive to Tat, and that this effect is mediated by NF-κB and NF-AT. We further go on to show, through Next Generation RNA Sequencing experiments, what the Tat-induced HERV-K transcriptome looks like, as well as provide additional evidence for other previously undescribed human genes activated by Tat. These data begin to provide explanations, at least in part, as to why HIV-1 infection is associated with such high levels of HERV-K (HML-2) expression in patients and begin to shine light on to what the possible consequences of this activation could be.

CHAPTER II

EXPRESSION OF HUMAN ENDOGENOUS RETROVIRUS TYPE K (HML-2) IS ACTIVATED BY THE TAT PROTEIN OF HIV-1

The work presented in this chapter was recently published in the Journal of Virology, Volume 86, Issue 15, pages 7790 to 7805; 2012

Abstract

Human Endogenous Retroviruses (HERVs) make up 8% of the human genome. Expression of HERV-K (HML-2), the family of HERVs that most recently entered the genome, is tightly regulated but becomes markedly increased following infection with HIV-1. To better understand the mechanisms involved in this activation, we explored the role of the HIV-1 Tat protein in inducing expression of these endogenous retroviral genes. Administration of recombinant HIV-1 Tat protein caused a 13-fold increase in HERV-K (HML-2) *gag* RNA transcripts in Jurkat T cells and a 10-fold increase in primary lymphocytes, and expression of the HERV-K (HML-2) *rec* and *np9* oncogenes was also markedly increased. This activation was seen especially in lymphocytes and monocytic cells, the natural hosts for HIV-1 infection. Luciferase reporter gene assays demonstrated that the effect of Tat on HERV-K (HML-2) expression occurred at the level

of the transcriptional promoter. The transcription factors NF-κB and NF-AT contribute to the Tat-induced activation of the promoter, as shown by chromatin immunoprecipitation assays, mutational analysis of the HERV-K (HML-2) LTR, and treatments with agents that inhibit NF-κB or NF-AT activation. These studies demonstrate that HIV-1 Tat plays an important role in activating expression of HERV-K (HML-2) in the setting of HIV-1 infection.

Introduction

Human Endogenous Retroviruses (HERVs) are transposable elements that make up 8% of the total human cellular DNA [43, 46, 282, 283]. After a series of germline infections over millions of years [43, 46], HERVs now exist in the genome in proviral forms [284] consisting of the basic retroviral genes (*gag*, *pro, pol* and *env*) flanked by two Long Terminal Repeats (LTRs) [210] formed during reverse transcription, with the 5' LTR serving as the viral transcriptional promoter. Most of these proviral sequences have been rendered non-functional through the passage of time due to the acquisition of multiple inactivating mutations and deletions [46]. However, many individual HERV genes remain intact, leading to their expression in human cells [46, 62]. The endogenous retrovirus HERV-K (HML-2), for example, has been shown to be transcriptionally active [43, 46, 200, 201, 285-287]. It is the most recent entrant into the human genome [98, 199], having last integrated into the human genome between 200,000 to 5 million years ago [210], and it is the only subfamily of endogenous retroviruses with conserved, and

therefore potentially functional, open reading frames (ORFs) for all viral proteins [209, 210]. There are approximately 91 full-length copies of HERV-K (HML-2) [43] per haploid genome and thousands of solitary LTRs. Copies of HERV-K (HML-2) are found on multiple different chromosomes [210, 241]. HERV-K (HML-2) also encodes the accessory oncogenes *np9* and *rec* (*nuclear protein of 9 kDa* and *regulator of expression encoded by cORF*) [240-243], which are expressed respectively by the two types of HERV-K (HML-2), Type 1 and Type 2. Type 1 and Type 2 differ only by a 292-bp deletion at the beginning of the envelope gene in Type 1 [43, 210].

Endogenous retroviral elements can be involved in physiological processes, such as those regulating transcription of genes like INSL4, β1,3-GT, Endothelin B Receptor, and Tissue-Specific Salivary Amylase [46, 104-107]. Additionally, expression of certain HERV proteins has important physiological functions, such as in placental development [58, 109-111], and may also provide mechanisms for protecting against exogenous virus infection [46, 112]. However, in general, how or why HERV genes are expressed, and the mechanisms responsible for expression, is not clearly understood. It is known that exogenous viral infections, viral transactivators, processes such as inflammation, chemical agents, cytokines, hormones, and stress conditions can contribute to the activation and transcription of transposable genetic elements, HERV-K (HML-2) being an example [113-129]. A possible role for HERV-K (HML-2) in pathogenesis has been considered in disorders such as Systemic Lupus Erythematosus (SLE), rheumatoid arthritis, and neuroinflammation [165, 168, 170, 255, 283, 288-290]. Certain

malignancies, most commonly germ-cell tumors, melanoma, breast tumors, and prostate cancer, also show high levels of HERV-K (HML-2) antigen expression [46, 203, 239, 291-293], sometimes accompanied by the production of viral particles [101, 215], yet the actual contribution of HERVs to disease remains to be characterized.

The HERV-K (HML-2) proteins Rec and Np9 provide a potential link between HERV-K (HML-2) and oncogenesis [62, 126, 239, 240, 244-247]. Both proteins have been shown to stimulate c-Myc expression by binding and inhibiting the *c-myc* gene repressor Promyelocytic Leukemia Zinc-Finger Protein (PLZF; [246]), and Rec has also recently been shown to interact with the Testicular Zinc-Finger Protein, another transcriptional repressor [245]. Additionally, Rec overexpression leads to testicular carcinoma *in situ* in transgenic mice [198, 234, 253]. Np9 transcripts are detected with high frequency in tumor samples and, though no direct evidence exists that links it to oncogenesis, Np9 has been shown to interact with a member of the cancer-associated Notch signaling pathway [247]. Thus, increased expression of the HERV-K (HML-2) proteins Rec and Np9 has the potential to contribute to oncogenesis.

Antibodies against HERV-K have been found in the blood of patients with a number of different clinical conditions [130, 210, 211, 239, 254, 255]. One of the highest percentages of antibodies against these retroviruses is seen in HIV-1 infected patients, where approximately 70% show a response against HERV-K (HML-2) antigens [121, 211, 239, 255]. We and others have demonstrated that HERV-K (HML-2) RNA levels are significantly increased in the plasma of HIV-1 infected patients (~10^7-10^8 copies/mL) as

compared to healthy HIV-1-negative controls (0-10^2 copies/mL) [112, 189-192, 257], and we have detected HERV-K (HML-2) proteins and viral particles in the blood of human patients with HIV-1-associated lymphoma [190, 257]. HIV-1 infection of peripheral blood mononuclear cells (PBMCs) isolated from healthy donor blood leads to increased expression of both HERV-K (HML-2) RNA and protein [116]. Additionally, it has been observed that HERV-K elements are over-expressed in brain tissue of AIDS patients who develop neurological complications due to increased immune activation [285]. How HERV-K might be activated by HIV has remained an open question. It is possible that the increased expression of HERV-K in HIV-1 infection is due to immunosuppression, but it could also be a consequence of direct interaction with infectious HIV-1 particles or viral proteins [192]. Interestingly, work from the Cullen and Löwer laboratories has provided evidence that HIV-1 Rev recognizes the cis-acting Rec-response region in the HERV-K (HML-2) RNA, which is similar to the Rev-Response Element of HIV-1, and can actively export HERV-K (HML-2) RNA from the nucleus to the cytoplasm [248, 249].

The HIV-1 regulatory protein Tat is a potent transactivator of the HIV promoter and is essential for viral replication [277]. Tat is produced in the early phase of HIV infection as a 72- or 101- amino acid protein; depending on the viral isolate, a truncated 86- amino acid form can also be produced [277, 294-297]. In addition to activating HIV transcription in the cell where it is made, Tat is actively secreted by infected cells into the extracellular surroundings, where it can be taken up by neighboring cells and exert effects on gene expression [120, 277, 297]. Additionally, HIV-1 Tat has been known to act not only on the HIV-1 and HIV-2 promoters, but also on other viral and cellular

promoters [120, 278-281]. Interestingly, the HIV-1 Tat protein has been shown to upregulate the transcription of *Alu* repeat sequences, an effect mediated by interactions of the transcription factor TFIIIC with the *Alu* promoter [120].

In view of the above, we hypothesized that the HIV-1 Tat protein provides a functional link between HIV-1 infection and the induction of HERV-K (HML-2) gene expression by causing activation of HERV-K LTR-directed transcription. In the studies presented below, we show that Tat stimulates HERV-K expression in cell lines and primary lymphocytes. We further demonstrate that the HERV-K (HML-2) transcriptional promoter is responsive to Tat, and that this effect is mediated by NF-κB and NF-AT. This may explain, at least in part, why HIV-1 infection is associated with such high levels of HERV-K (HML-2) expression in patients.

Results

HERV-K (HML-2) RNA expression is increased in HIV-1-infected cell lines

Recent studies by our group and others have shown that HIV-1 infection increases HERV-K (HML-2) gene expression, both in cell culture and in patients [112, 116, 188-191, 257]. However, the underlying mechanism for this increased expression has remained unknown. To begin to address this issue, we first ascertained what the levels of HERV-K (HML-2) RNA expression were after HIV-1 infection of 2 different T cell lines. We decided to measure expression of HERV-K (HML-2) at the RNA level, and used primers to the *gag* gene due to this gene's similarities among all 91 HERV-K (HML-2)

proviruses. Quantification of HERV-K (HML-2) *gag* RNA by qRT-PCR in cells not infected with HIV-1 showed RNA levels around 10^3 copies per 500 nanograms (ng) of total RNA (Appendix 1.1). These levels increased up to approximately 20-fold with HIV-1 infection (Fig. 2.1). Higher expression of HERV-K (HML-2) *gag* RNA was consistently seen in all HIV-1-infected cells as compared to their uninfected counterparts ($p<0.001$, Fig. 2.1). Interestingly, uninfected Jurkat T cells that stably express the Tat protein from HIV-1 (Jurkat-Tat) showed 8-fold more HERV-K (HML-2) *gag* RNA expression than did Jurkat T cells lacking Tat (Fig. 2.1). As Tat has been shown to activate both viral and cellular genes (15, 52, 86, 90, 110), and as we observed that Jurkat-Tat cells have higher HERV-K (HML-2) *gag* RNA expression as compared to the parental Jurkat counterpart, we hypothesized that the HIV-1 Tat protein might play a role in activating HERV-K (HML-2) expression during HIV-1 infection.

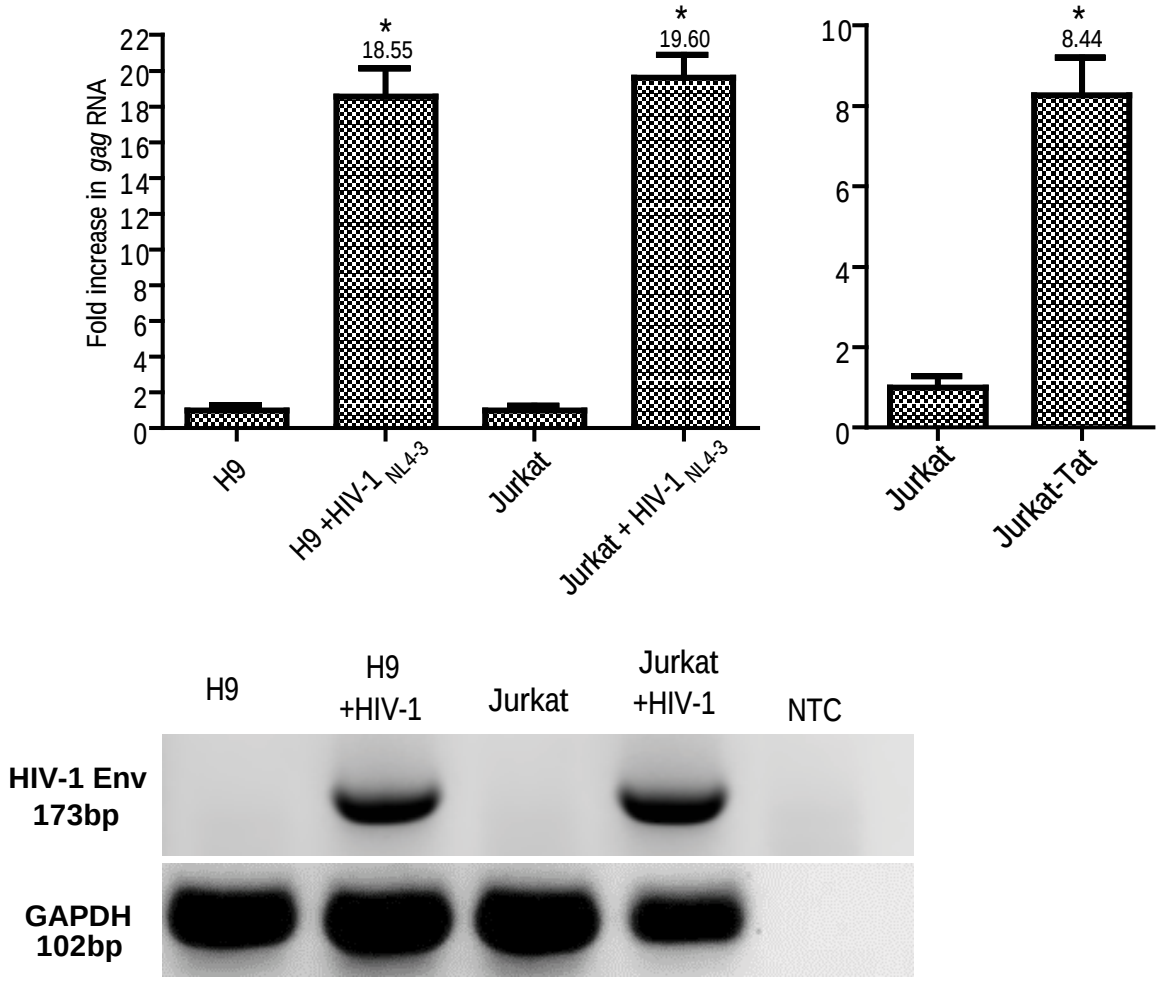


Figure 2.1 HIV-1 infection and Tat stimulate HERV-K (HML-2) gene expression**.** Total cellular RNA was isolated from cells that were infected with HIV-1_{NL4-3} (for one week), or left uninfected. RNA was amplified using primers specific for HERV-K (HML-2) *gag* through One Step qRT-PCR, and quantified using a standard curve generated by amplification of HERV-K (HML-2) *gag* RNA standards. Data are expressed as fold increase over uninfected cells, with uninfected cells shown as normalized to 1 for simplicity of comparison. Rightmost panel shows Jurkat-Tat HERV-K (HML-2) *gag* RNA levels as compared to levels in the parental Jurkat T cell line, with Jurkat T cell *gag* RNA levels normalized to 1. HIV-1 *env* and *GAPDH* One Step RT-PCR amplifications were also performed on the RNA to verify infection status and integrity of the material (NTC=Non–template control). Error bars indicate SD for results of three independent experiments. Significance was calculated by comparing infected samples with their uninfected counterparts using a Student's T-test and significant results are indicated (*p= <0.005).

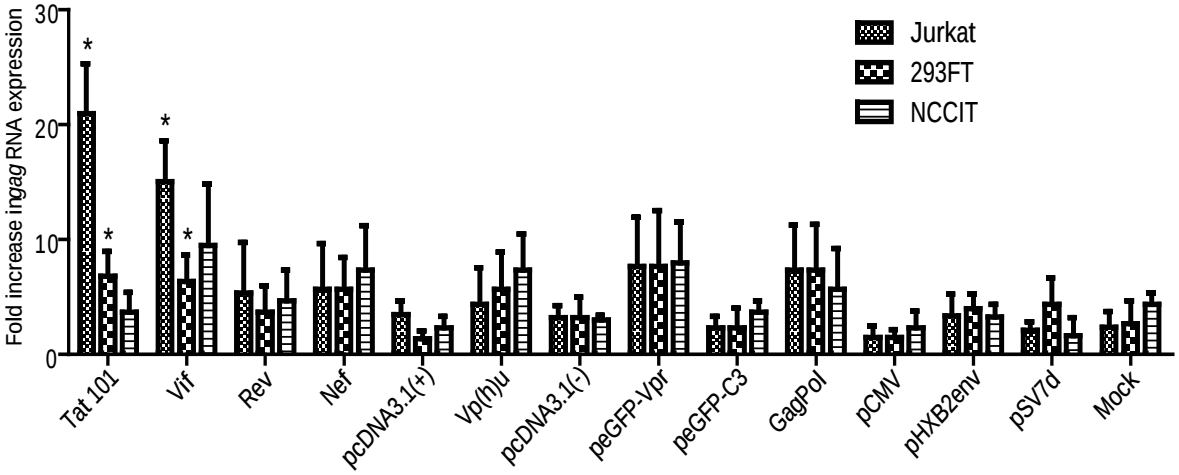
HIV-1 Tat and Vif independently cause an increase in HERV-K (HML-2) gag RNA expression

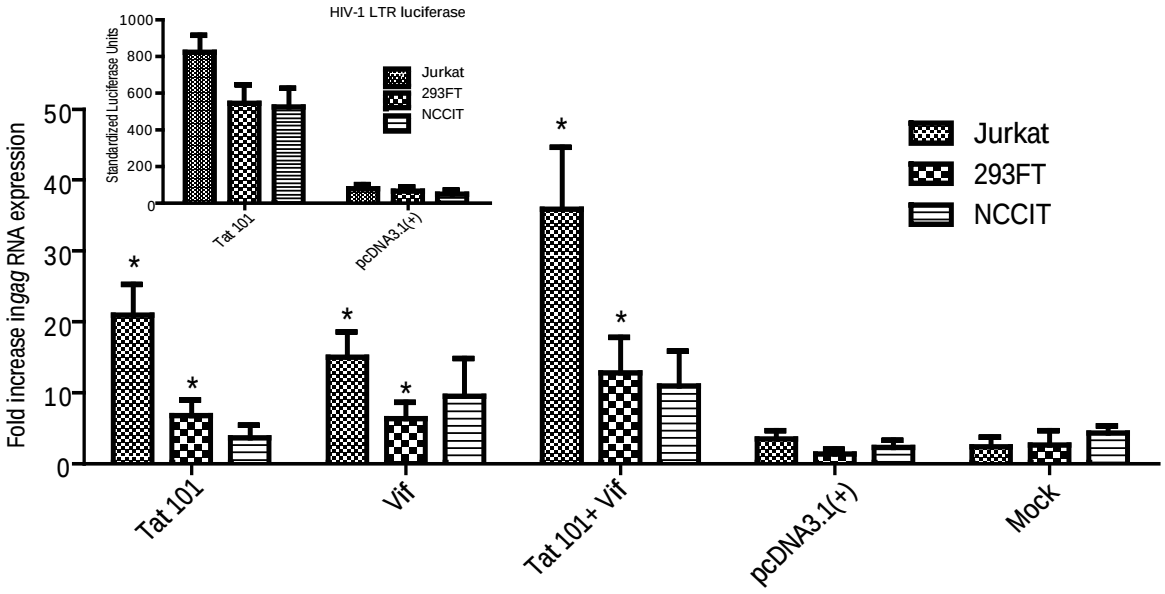
To ascertain whether our hypothesis that HIV-1 Tat activates the synthesis of HERV-K (HML-2) *gag* RNA is correct, we first transfected different cell types with plasmids encoding each of the regulatory, accessory, and structural proteins from HIV-1 and measured the levels of cellular HERV-K (HML-2) *gag* RNA after 24 and 48 hours. We used a T cell line permissive for HIV-1 infection (Jurkat T cells), an easily transfectable line (293FT), and a cell line known for its high expression of HERV-K (HML-2) transcripts and proteins (NCCIT teratocarcinoma cells). Expression of Tat (from a plasmid encoding the full-length, 101 amino acid form) or Vif yielded a significant increase in HERV-K (HML-2) RNA expression (Fig. 2.2A). The presence of Tat or Vif increased HERV-K (HML-2) *gag* RNA by about 21- or 15-fold over untreated Jurkat cells, respectively ($p<0.01$). Similar responses to Tat and Vif were seen in 293FT cells, whereas the teratocarcinoma NCCIT showed a response only to Vif (Fig. 2.2A). Similar increases in HERV-K (HML-2) *gag* RNA expression were observed 48 hours after transfection (Appendix 1.2). The expression of any of the other HIV-1 proteins resulted in no significant increase in HERV-K (HML-2) RNA.

To determine whether Tat and Vif synergistically increase HERV-K (HML-2) transcript levels, we co-transfected plasmids encoding these proteins and observed the effect on transcription after 24 hours. For both Jurkat T cells and 293FT, Tat and Vif had an additive, not synergistic, effect with regards to HERV-K (HML-2) transcription,

whereas in NCCIT no significant difference was seen when compared to Vif-induced expression alone (Fig. 2.2B). Tat functionality was verified in NCCIT cells in parallel co-transfections using an HIV-1 LTR-luciferase reporter assay (Fig. 2.2B, inset). These data show that both Tat and Vif can activate HERV-K (HML-2) transcription, and suggest that Tat alone is sufficient for activation in HIV-1-relevant targets of infection (i.e. Jurkat T cells).

To further test the importance of Tat in HIV-1-mediated activation of HERV-K (HML-2), we transfected an HIV-1 infectious molecular clone (pHXB2) or a mutant version of it lacking the Tat protein (pMtat(-)) into Jurkat T cells and 293FT cells, and measured HERV-K (HML-2) *gag* RNA 48 hours later. As can be seen in Figure 2.2C, Tat expression in Jurkat T cells is important for higher HERV-K (HML-2) RNA expression, as its absence greatly diminishes levels of transcripts (by more than half, $p<0.05$). Of course, this experiment must be interpreted with caution as the lack of Tat also leads to a decrease in the expression of other HIV-1 proteins. However, these data are consistent with the observation that HIV-1 Tat expression is sufficient to increase HERV-K (HML-2) transcript expression.


Fold increase in gag RNA expression
30
20
10
0
Jurkat
293FT
NCCIT
Tat 101
Vif
Rev
Nef
pcDNA3.1(+)
Vp(h)u
pcDNA3.1(-)
peGFP-Vpr
peGFP-C3
GagPol
pCMV
pHXB2env
pSV7d
Mock


HIV-1 LTR luciferase
Standardized Luciferase Units
1000
800
600
400
200
0
Jurkat
293FT
NCCIT
Tat 101
pcDNA3.1(+)
Fold increase in gag RNA expression
50
40
30
20
10
0
Jurkat
293FT
NCCIT
Tat 101
Vif
Tat 101+ Vif
pcDNA3.1(+)
Mock

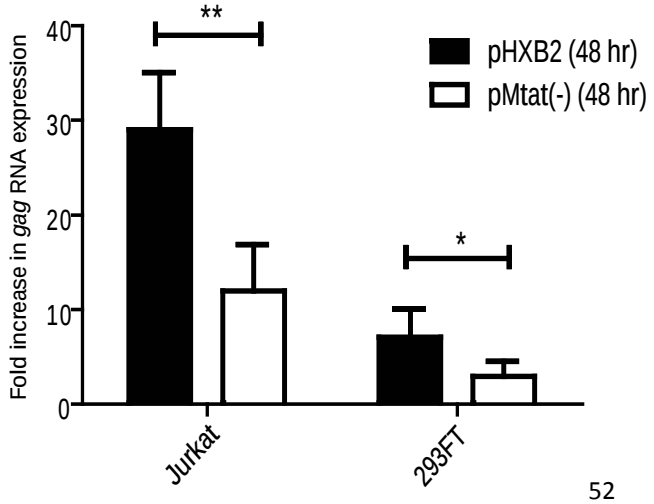

Fold increase in gag RNA expression
40
30
20
10
0
pHXB2 (48 hr)
pMtat(-) (48 hr)
Jurkat
293FT

Figure 2.2 HIV-1 Tat and Vif proteins activate HERV-K (HML-2) RNA expression. (A) Total cellular RNA was isolated from cells 24 hours after transfection with plasmids encoding the individual regulatory and accessory proteins from HIV-1, and subjected to One Step Sybr Green qRT-PCR with primers specific for HERV-K (HML-2) *gag*. HIV-1 full length Tat (Tat 101), Rev and Nef were cloned into the pcDNA3.1(+) expression vector. HIV-1 full length Vif and Vpu were cloned into the pcDNA3.1(-) expression construct. Full length Vpr was cloned into the peGFP-C3 expression vector, and is fused to GFP. Full length HIV-1 Gag and Pol were expressed from the HIV-1 *gagpol* RNA sequence cloned into pCMV and expressed with co-transfection of HIV-1 Rev. Full length Env from HIV-1_{HXB2} was cloned into pSV7D and termed pHXB2Env for simplicity. Respective empty vectors are shown, as well as mock transfections, with data expressed as fold increase in RNA over untreated cells. (B) Cells were transfected with full length Tat and/or Vif encoding plasmids, and total RNA was extracted 24 hours later and subjected to One-Step Sybr Green qRT-PCR using primers specific for HERV-K *gag.* Inset: Tat activity was further verified by co-transfection with an HIV-1 LTR-luciferase reporter vector. Relative luciferase units were normalized to an internal transfection control vector encoding *Renilla* luciferase and expressed as standardized luciferase units. (C) HERV-K (HML-2) *gag* RNA expression from cell lines transfected with an infectious HIV-1 molecular clone (pHXB2) or a mutant version lacking the Tat protein (pMtat(-)). Forty eight hours after transfection, total RNA was isolated and subjected to amplification by One Step Sybr Green qRT-PCR. Data are expressed as fold increase in RNA over untreated cells. All qRT-PCR results were normalized to the GAPDH reference gene after analysis using the $2^{-\Delta\Delta Ct}$ method, and relative expression is plotted. Error bars indicate SD for results of three independent experiments. Significance was calculated by comparing protein treatments to the empty vector controls (i.e. pcDNA3.1, peGFP-C3, pCMV, pSV7D) or by comparing the full length molecular clone to the Tat mutant using a Student's T-test. Significant results are indicated (*p= <0.05, **p=<0.005).

Recombinant HIV-1 Tat increases HERV-K (HML-2) expression

We next analyzed whether addition of physiologically-relevant levels of recombinant HIV-1 Tat to cells could activate HERV-K (HML-2) transcription. We took advantage of the fact that, unlike most transcription factors, HIV-1 Tat is able to cross intact cellular membranes when it is present in the extracellular milieu [277, 297-299]. Time-course experiments in Jurkat T cells to which we added a recombinant 86-amino acid form of Tat and measured HERV-K (HML-2) *gag* RNA expression showed that Tat addition caused a 10.8-fold increase in HERV-K (HML-2) *gag* RNA production after 6 hours (Fig. 2.3A, left panel). The effect peaked at 8 hours (13-fold increase), and then gradually declined (Fig. 2.3A, left panel). Recombinant Tat activity was further verified in parallel experiments by measuring reporter activity from Jurkat T cells transfected with a vector containing the HIV-1 LTR fused to the luciferase reporter gene, which showed the protein to be fully active (Fig. 2.3A, right panel). To verify that endotoxin or other extraneous material in the Tat protein samples were not responsible for HERV-K (HML-2) *gag* activation, we heat-denatured Tat and added it to cells (endotoxin is highly heat-stable, whereas Tat is not). No significant increase in HERV-K (HML-2) *gag* RNA was detected under these conditions, suggesting that activation of HERV-K (HML-2) *gag* gene expression was due to the Tat protein and not to endotoxin (Fig. 2.3A).

Activation of HERV-K (HML-2) RNA synthesis by Tat was not limited to *gag* transcripts in Jurkat T cells, nor was it limited to that cell type alone, as we also detected significantly increased HERV-K (HML-2) transcripts for *gag, rec* and *np9* in HUT-78

lymphoblasts, U-937 monocytes, and 293FT fibroblasts after Tat was added for 8 hours (Fig. 2.3B). This effect was not seen with all cell types tested, as the teratocarcinoma cell line NCCIT did not show any significant increase in HERV-K (HML-2) gene expression with Tat treatment, consistent with the results of the transfection experiments (Fig. 2.2A). Further evidence that this lack of effect was not due to the cells being non-permissive or non-responsive to Tat comes from parallel experiments involving transfection of the HIV-1 LTR-luciferase reporter vector, and subsequent Tat exposure, which showed that Tat could be internalized and activates the HIV-1 LTR (Fig. 2.3B, inset).

As the experiments described above were all performed in cell lines, we sought to verify that the increase in HERV-K (HML-2) transcripts in response to Tat also occurs in primary cells, especially those most relevant to HIV infection. Peripheral blood lymphocytes (PBLs) from healthy individuals were either exposed to stimulatory conditions (PHA and IL-2) or not, and then treated with recombinant Tat protein. RNA expression analyses from unstimulated PBLs showed that Tat treatment leads to expression of HERV-K (HML-2) *gag* RNA as early as 6 hours, which continues to rise and peaks by twelve hours, with an approximately 10-fold increase over cells not exposed to Tat (Fig. 2.3C). We observed that stimulation of PBLs with PHA and IL-2 alone for three days activates transcription of HERV-K (HML-2) (Appendix 1.3), but this effect was sustained for only about a total of seventy-four hours in culture (three days of pre-stimulation, and then experimental treatments), with RNA expression returning to basal levels by ninety-six hours (Fig. 2.3C, Buffer (24hr)), unless Tat was present.

As our data show that Tat activated HERV-K (HML-2) transcription, we next assessed whether the increases in HERV-K (HML-2) transcripts would result in increased protein expression. Using untreated Jurkat T cell lysate with a mix of commercially available antibodies against the full-length HERV-K (HML-2) Gag and its Capsid form, little protein expression was detected in untreated cells, as has previously been reported [212]. When Tat was transfected into Jurkat T cells, full-length HERV-K (HML-2) Gag protein was still minimally expressed but the cleaved Capsid form of Gag (approximately 30 kDa) showed an approximately 18.8-fold increase in expression over the empty vector control (Fig. 2.3D and E). A similar increase in Capsid expression was seen when recombinant Tat was added to the cells, although treatment with the vehicle buffer (which contains the reducing agent DTT) in this case also increases the expression to some degree (Fig. 2.3D). Protein increases with Tat treatment in this setting were approximately 3.92-fold over the vehicle control (Fig. 2.3E). Although not as impressive as the RNA increases seen with Tat, these results show that Tat can substantially increase the expression of HERV-K (HML-2) Gag, and this protein is detected in its Capsid form after it is further processed by a protease. Indeed, it has previously been established that cell lysates can contain both the precursor form of HERV-K (HML-2) Gag as well as the cleaved Capsid [101, 212].

Taken together, these observations demonstrate that HIV-1 Tat stimulates HERV-K (HML-2) gene expression at the RNA and the protein level in what appears to be a cell type-specific manner, with cells that are relevant to HIV biology being the most affected,

and with significant transcriptional activation seen in primary lymphocytes, a major target for HIV infection.

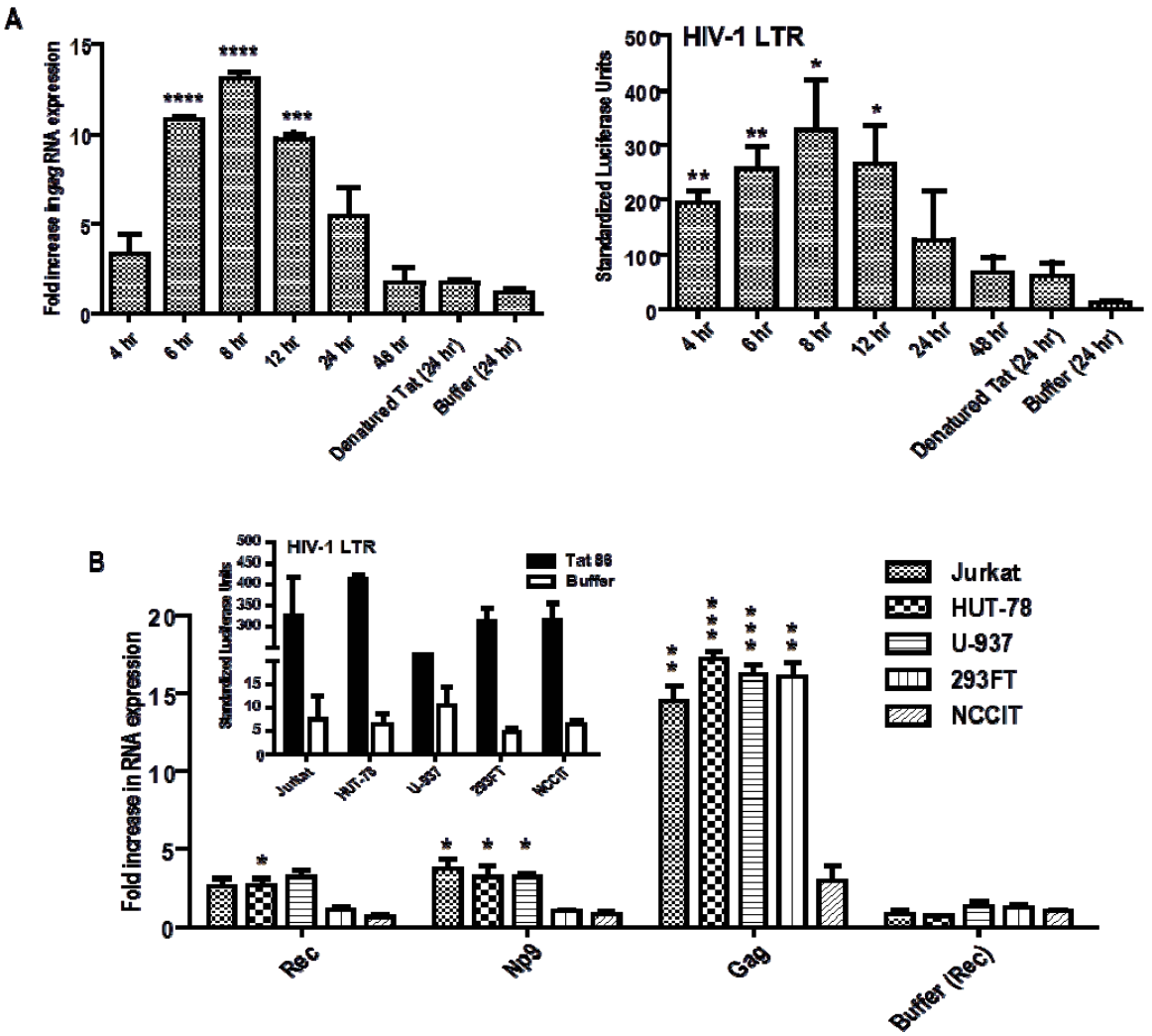


Figure 2.3 Recombinant HIV-1 Tat activates HERV-K (HML-2) gene expression. Total cellular RNA was isolated from cells that were subjected to different Tat treatments and amplified by One-Step Sybr Green Real Time qRT-PCR using primers specific for HERV-K *gag, rec, np9, envelope type 1* and *envelope type 2. (*A) Left panel: HERV-K (HML-2) g*ag* RNA expression from Jurkat T cells treated with 100 ng/mL of purified Tat protein for the specified periods of time. Right panel: Jurkat T cells transfected with an HIV-1 LTR-luciferase reporter construct, which 24hr later were treated with 100 ng/mL purified Tat protein for the specified periods of time. Relative luciferase units were normalized to an internal transfection control vector encoding *Renilla* luciferase, and expressed as standardized luciferase units. Denatured Tat protein and the Tat vehicle buffer were used as negative controls, and only the 24 hour time point is shown for those as representative results. (B) RNA expression of the HERV-K (HML-2) genes *rec, np9, gag, env type 1,* and *env type 2* in different cell lines after 8 hour treatment with 100 ng/mL Tat protein. Inset: Tat activity was verified by transfection with an HIV-1 LTR-luciferase reporter vector. Relative luciferase units were normalized to an internal transfection control vector encoding *Renilla* luciferase, and expressed as standardized luciferase units.

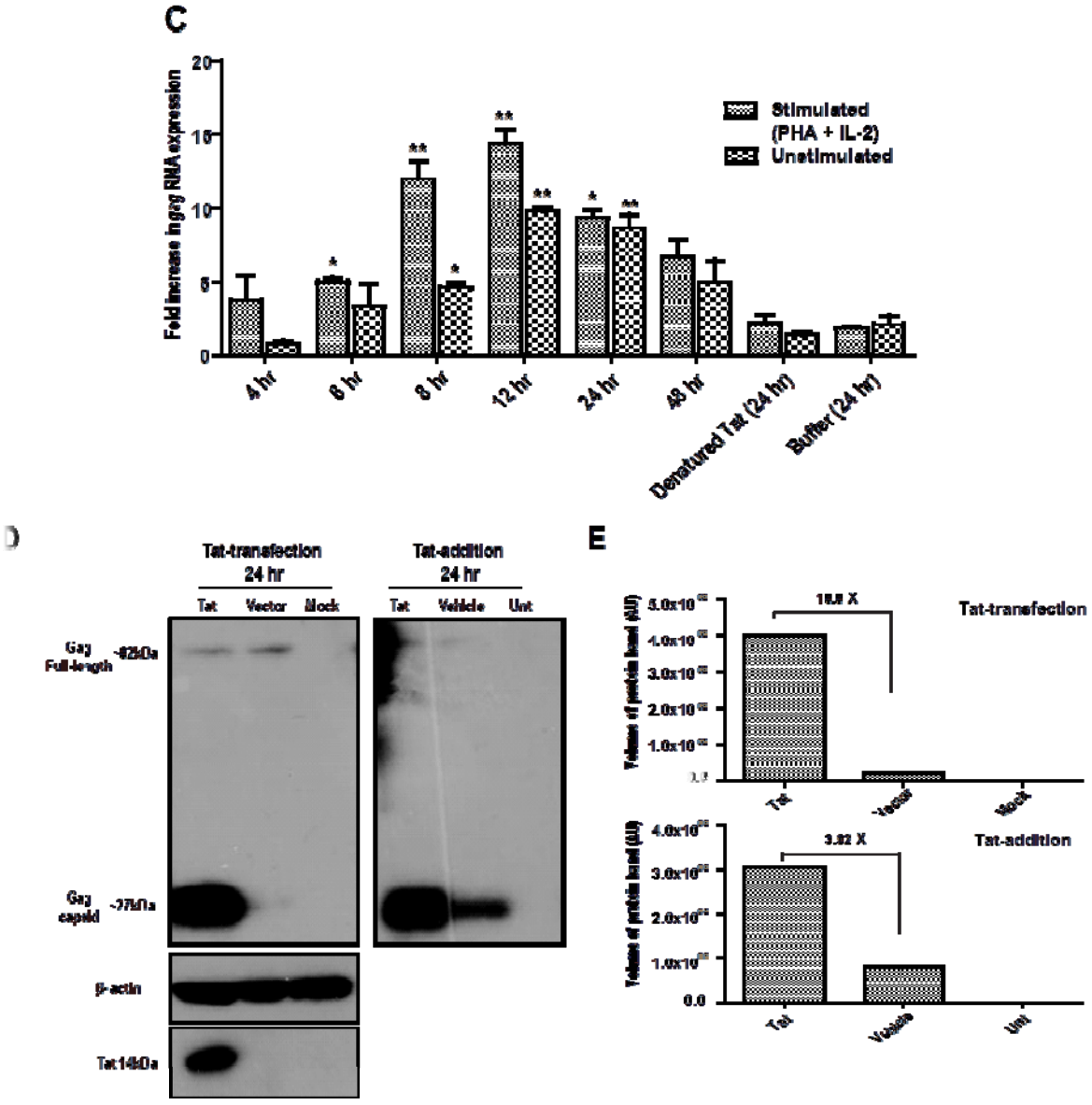

C
Fold increase in gag RNA expression
Stimulated (PHA + IL-2)
Unstimulated
4 hr
8 hr
8 hr
12 hr
24 hr
48 hr
Denatured Tat (24 hr)
Buffer (24 hr)
D
Tat-transfection 24 hr
Tat-addition 24 hr
Tat
Vector
Mock
Vehicle
Unt
Gag Full-length ~62kDa
Gag capsid ~27kDa
β-actin
Tat 14kDa
E
Tat-transfection
18.8 X
Tat-addition
3.82 X
Tat
Vector
Mock
Vehicle
Unt

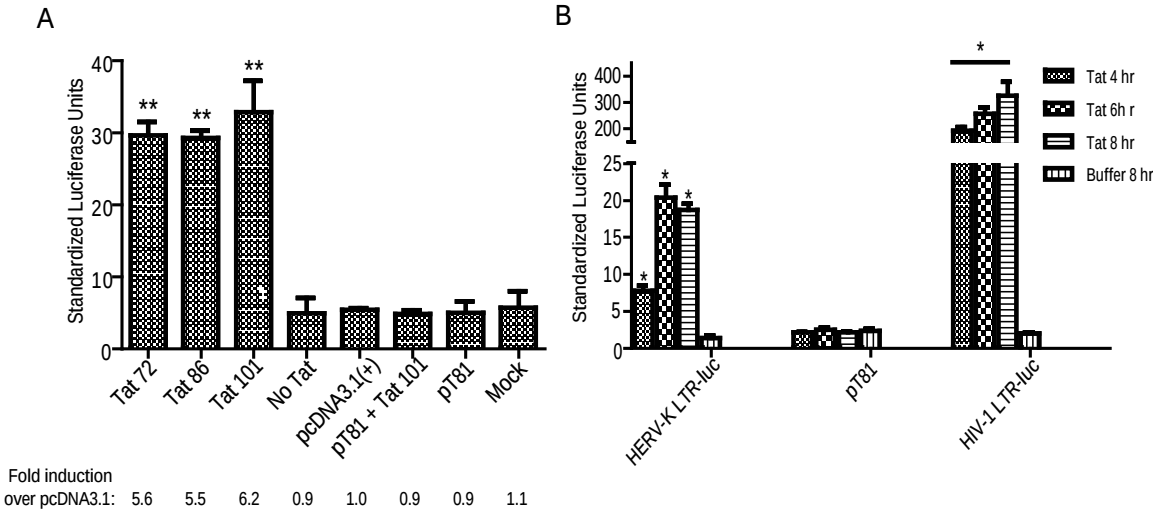
Figure 2.3 (continuation) (C) HERV-K (HML-2) g*ag* RNA expression from peripheral blood lymphocytes (PBLs). PBLs were split in half and either treated with 100 ng/mL Tat protein for 8hr, or pre-stimulated with PHA and IL-2 for 3 days followed by an 8hr Tat treatment. All qRT-PCR results were normalized to the GAPDH reference gene after analysis using the $2^{-\Delta\Delta Ct}$ method, and relative expression is plotted as fold over untreated cells. Error bars indicate SD for results of three independent experiments. Significance was calculated by comparing Tat treatments to buffer controls, at the same time points, using a Student's T-test, and significant results are indicated (****p= <0.001, ***p=<0.005, **p=<0.01, *p=<0.05). (D) Western blot showing HERV-K (HML-2) Gag protein expression in Jurkat T cells after treatment with HIV-1 Tat. For detection of HERV-K (HML-2) Gag protein, two monoclonal antibodies were used simultaneously: anti-HERV-K Gag and anti-HERV-K capsid. Left panel: HERV-K (HML-2) Gag protein expression in cell lysates 24 hours after transfection with Tat 101 ("Tat") or a control vector ("Vector", which is pcDNA3.1) or after mock transfection. Right panel: HERV-K (HML-2) Gag protein expression in cell lysates 48 hours after addition of recombinant Tat protein. Control lanes include the vehicle buffer ("Vehicle", PBS with BSA and DTT), and untreated cell lysate ("Unt"). Respective Gag proteins and sizes are shown. Tat protein expression in transfected cells is shown, and β-actin protein expression is shown as a loading control. (E) Densitometry analysis of the HERV-K (HML-2) Capsid protein bands seen in the Western Blots in D.

Activation of the HERV-K (HML-2) promoter by HIV-1 Tat

As HIV-1 Tat is known to act upon both viral and cellular promoters to increase or decrease gene expression [120, 278-280, 294, 300-303], we tested whether the effect of Tat on HERV-K (HML-2) occurred at the level of the transcriptional promoter. Using a construct containing a HERV-K (HML-2) LTR promoter driving the expression of the luciferase reporter gene, Jurkat T cells were co-transfected with plasmids encoding one of three isoforms of Tat (72, 86 or 101-amino acids). This was done as it is known that all isoforms do not necessarily behave the same with regard to cellular gene activation (e.g. full-length Tat 101, but not one-exon Tat 72, represses MHC-I expression) [294, 304, 305]. The expression of Tat RNA and protein was confirmed by RT-PCR and immunoblot (Appendix 1.4), and its functional integrity confirmed by showing transactivation of the HIV-1 LTR in luciferase reporter gene assays as described above. Consistent with the findings shown in Figures 2.2 and 2.3 above, we found that all HIV-1 Tat isoforms transactivated the HERV-K (HML-2) promoter with almost equal efficiency, approximately 5 to 6-fold over the empty vector control (Fig. 2.4A). This effect was corroborated by addition of purified, recombinant Tat protein (Tat 86 isoform) to Jurkat T cells that were transfected with the HERV-K (HML-2) LTR reporter construct, which lead to an increase in luciferase activity of approximately 4 to 8-fold over the signal generated by the buffer control alone (Fig. 2.4B). Tat showed little effect on the HERV-K (HML-2) promoter after transfection into 293FT or NCCIT cells (Appendix 1.5), which suggests that the Tat-mediated stimulation of HERV-K (HML-2) is cell-type specific.

Interestingly, the degree of activation seen from the HERV-K (HML-2) promoter used in the luciferase reporter assays was consistent with the observed Tat-induced increases in *gag* RNA. These promoter activation data thus support our observations that HIV-1 Tat activates HERV-K (HML-2) at the transcriptional level.

In order to understand whether HERV-K (HML-2) promoter activation occurs by a mechanism similar to that which Tat uses to activate transcription from the HIV-1 LTR, we co-transfected the HERV-K (HML-2) LTR-luciferase constructs with two different Tat mutants: Tat C22G and the naturally-occurring Tat T23N. Tat C22G contains a mutation in a cysteine at position 22 (to a glycine) in the transactivation domain, which renders it unable to interact with cyclin T1 and thus it is HIV-1 LTR activation-deficient [306, 307]. Tat T23N, on the other hand, contains a mutation in the threonine at residue 23 (to asparagine) that increases Tat's ability to activate the HIV-1 LTR by increasing binding of Tat to the cellular kinase positive transcription elongation factor b (P-TEFb, [307]). Neither of these mutations in Tat affected its ability to drive transcription from the HERV-K (HML-2) promoter, whereas they had the predicted effect on the HIV-1 LTR (Fig. 2.4C). These data show that activation of HERV-K (HML-2) transcription by Tat occurs in a different manner than that of HIV-1, and does not appear to involve Tat's interaction with Cyclin T1 or P-TEFb. This is consistent with the data described below.


A
Standardized Luciferase Units
Tat 72
Tat 86
Tat 101
No Tat
pcDNA3.1(+)
pT81 + Tat 101
pT81
Mock
Fold induction
over pcDNA3.1: 5.6 5.5 6.2 0.9 1.0 0.9 0.9 1.1
B
Standardized Luciferase Units
Tat 4 hr
Tat 6h r
Tat 8 hr
Buffer 8 hr
HERV-K LTR-luc
pT81
HIV-1 LTR-luc

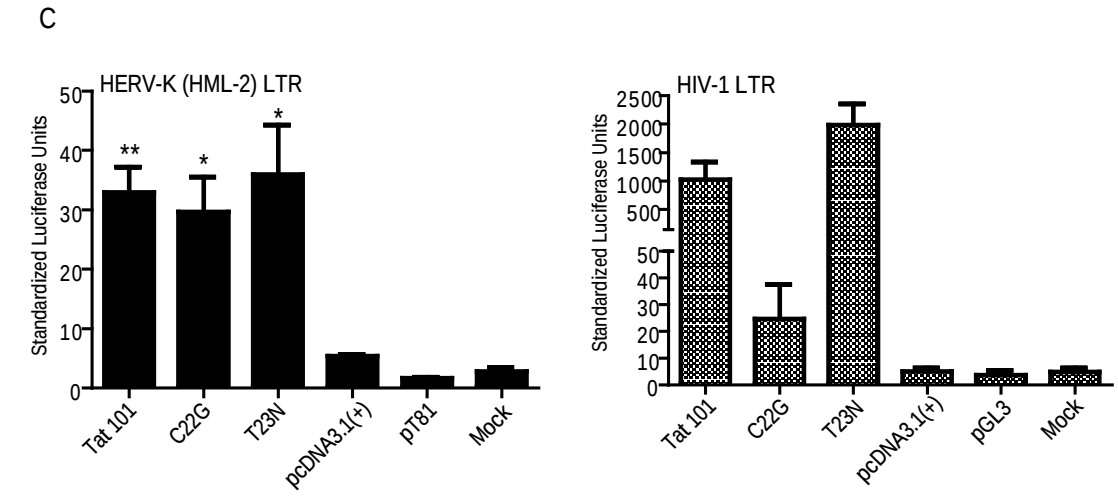

C
HERV-K (HML-2) LTR
Standardized Luciferase Units
Tat 101
C22G
T23N
pcDNA3.1(+)
pT81
Mock
HIV-1 LTR
Standardized Luciferase Units
Tat 101
C22G
T23N
pcDNA3.1(+)
pGL3
Mock

Figure 2.4 HIV-1 Tat activates the HERV-K (HML-2) promoter by a different mechanism than it uses to activate HIV-1. (A) Jurkat T cells were transfected with a HERV-K (HML-2) LTR-luciferase construct (except for the luciferase backbone construct pT81 and mock samples) and co-transfected with the construct shown on the x-axis (Tat 72, Tat 86, Tat 101 or pcDNA3.1). "Mock" corresponds to treatment with transfection reagent alone, whereas "No Tat" corresponds to untreated cells. Activation of the luciferase construct was measured 24 hours after transfection, normalized to *Renilla* luciferase signal, and shown as standardized luciferase units. Fold induction was calculated over the empty vector, pcDNA3.1. Control experiments included a mock transfection and transfection of the backbone luciferase vector (pT81) alone or with a Tat-expressing vector. (B) Jurkat T cells were transfected with the HERV-K (HML-2) LTR-luciferase construct, or a similar HIV-1 LTR luciferase construct as a positive control, and 24 hours after transfection were treated with 500 ng/mL purified Tat protein for the specified periods of time. Buffer control is 8 hr treatment. (C) Jurkat T cells were transfected with the HERV-K (HML-2) LTR- or HIV-1 LTR-luciferase reporter construct and co-transfected with the Tat mutants C22G or T23N, and cells were harvested at 24 hours. These mutants were expressed from the pcDNA3.1 expression vector, and are either unable to activate transcription from the HIV-1 LTR (Tat C22G), or cause an increase in HIV-1 LTR transcriptional activity (Tat T23N). Data are shown as standardized luciferase units. Error bars indicate SD from three independent experiments. Significance was calculated using a Student's T-test comparing Tat-transfection/treatment to pcDNA3.1 (A, C) or buffer control (B), and significant results are indicated (*p=<0.05, ** p=<0.01).

Activation of the HERV-K (HML-2) promoter by Tat is mediated by NF-κB and NF-AT

In addition to stimulating HIV transcriptional elongation through its interactions with Cyclin T1 and P-TEFb, Tat is known to activate cellular genes through regulation and/or interaction with upstream cellular transcription factors. Furthermore, it has been shown that Tat can, in the absence of a functional TAR, remain an important factor for HIV-1 transcription via Sp1 sequence elements in the U3 promoter region [308]. Tat can additionally interact directly with NF-κB, with this interaction not only demonstrating TAR-independent transactivation in HIV-1, but also pointing towards a mechanism of Tat-mediated modulation of cellular genes [309]. To understand the mechanism by which Tat activates HERV-K (HML-2) LTR-directed transcription, we analyzed the sequence of the promoter for potential transcription factor binding sites. Utilizing the ALGGEN-PROMO software [310, 311] and the sequence prediction algorithm TRANSFAC database software (BioBase Co., Beverley, MA), we analyzed the promoter sequence and found that a number of transcription factors might potentially interact with the HERV-K (HML-2) promoter, including AP-1, CREB, CEBP (C/EBP$_\alpha$), c-Rel, NF-AT, CEBP$_\beta$, NF-κB(p50:p52), Rel-A, p53, YY1, c-Myc, Sp1, Sp3, and the STATs. Of the potential sites present, we decided to focus on ones previously shown to be particularly associated with HIV-1 Tat activation: Sp1, NF-κB and NF-AT [312-319]. Two potential NF-κB binding sites, with NF-AT sites embedded in them (here referred to as κB1/N1 and κB3/N3), are found in the most upstream part of the U3 region of the promoter, along with single NF-AT (N2), NF-κB (κB2), and Sp1 sites (Fig. 2.5A). The R region contains a lone NF-κB/NF-AT

site (κB4/N4), and the U5 region has two potential NF-AT binding sites present (N5 and N6, Fig. 2.5A).

In view of the multiple NF-κB sites found in the HERV-K (HML-2) promoter and the known contribution of NF-κB to Tat-mediated activation of the HIV-1 promoter [317, 320], we tested whether NF-κB mediates the Tat effect on the HERV-K (HML-2) LTR. To do so, we first examined whether a dominant-negative inhibitor of NF-κB nuclear translocation would block Tat stimulation of the HERV-K (HML-2) promoter. This dominant negative construct (referred to in the figure as IκBα DN) codes for the Inhibitor of NF-κB alpha (IκBα) and sequesters NF-κB in the cytosol, preventing its translocation into the nucleus upon activation. Forty eight hours after co-transfection, the activation of the HERV-K (HML-2) promoter construct in response to Tat was significantly decreased, by approximately 65%, in the presence of the NF-κB inhibitor (Fig. 2.5B, left panel $p<0.05$). This dependence on NF-κB was corroborated in experiments in which Jurkat T cells were pre-treated for 1 hour with a specific, irreversible inhibitor of IKKα and β kinase activity, Wedelolactone (Wed, [321]) before the addition of recombinant Tat protein. After 6 hours, RNA was isolated and quantitated by qRT-PCR. Figure 2.5B (right panel) shows that inhibition of NF-κB diminishes the increase in HERV-K (HML-2) *gag* RNA seen in response to Tat by half ($p<0.01$), but does not completely abolish it. As a control, we also observed that PMA and Ionomycin- induced transcription is similarly diminished, but not abolished, in the

presence of Wedelolactone. Thus, NF-κB mediates, in part, activation of HERV-K (HML-2) by Tat.

As the HERV-K (HML-2) LTR also contains potential NF-AT binding sites, NF-AT activation could additionally contribute to HERV-K (HML-2) Tat-driven expression, and might compensate for the absence of NF-κB activity. We therefore examined whether inhibiting NF-AT activation would also result in diminished HERV-K (HML-2) responsiveness to Tat. Treatment of Jurkat T cells that were co-transfected with the HERV-K (HML-2) reporter construct and Tat with the immunosuppressive drug Cyclosporin A (CsA) (a calcineurin inhibitor that prevents dephosphorylation of NF-AT and therefore its activation) showed a dose-dependent inhibition of Tat-mediated HERV-K (HML-2) promoter activation (Fig. 2.5C, left panel). This suggests that NF-AT is also involved in HERV-K (HML-2) activation. An MTT assay was performed on cells treated with the described doses of CsA, and showed no significant cell death at any of the concentrations tested (Appendix 1.6). However, knowing that CsA is not an NF-AT-specific drug, we also pre-treated Jurkat T cells with 11R-VIVIT, a cell-permeable peptide inhibitor specific for NF-AT [322] for 1 hour, added recombinant Tat protein for 6 hours, and then isolated RNA. We observed a reduction in *gag* RNA to about half of the starting levels in cells treated with the NF-AT inhibitor in the presence of Tat (Fig. 2.5C, right panel). Treating with both Wedelolactone and 11R-VIVIT simultaneously prior to Tat addition had the strongest effect on diminishing *gag* RNA levels, as this treatment decreased expression by about 75% (Fig. 2.5D). Overall, these data show that both NF-

κB and NF-AT activation in response to HIV-1 Tat drive transcription from the HERV-K (HML-2) promoter.

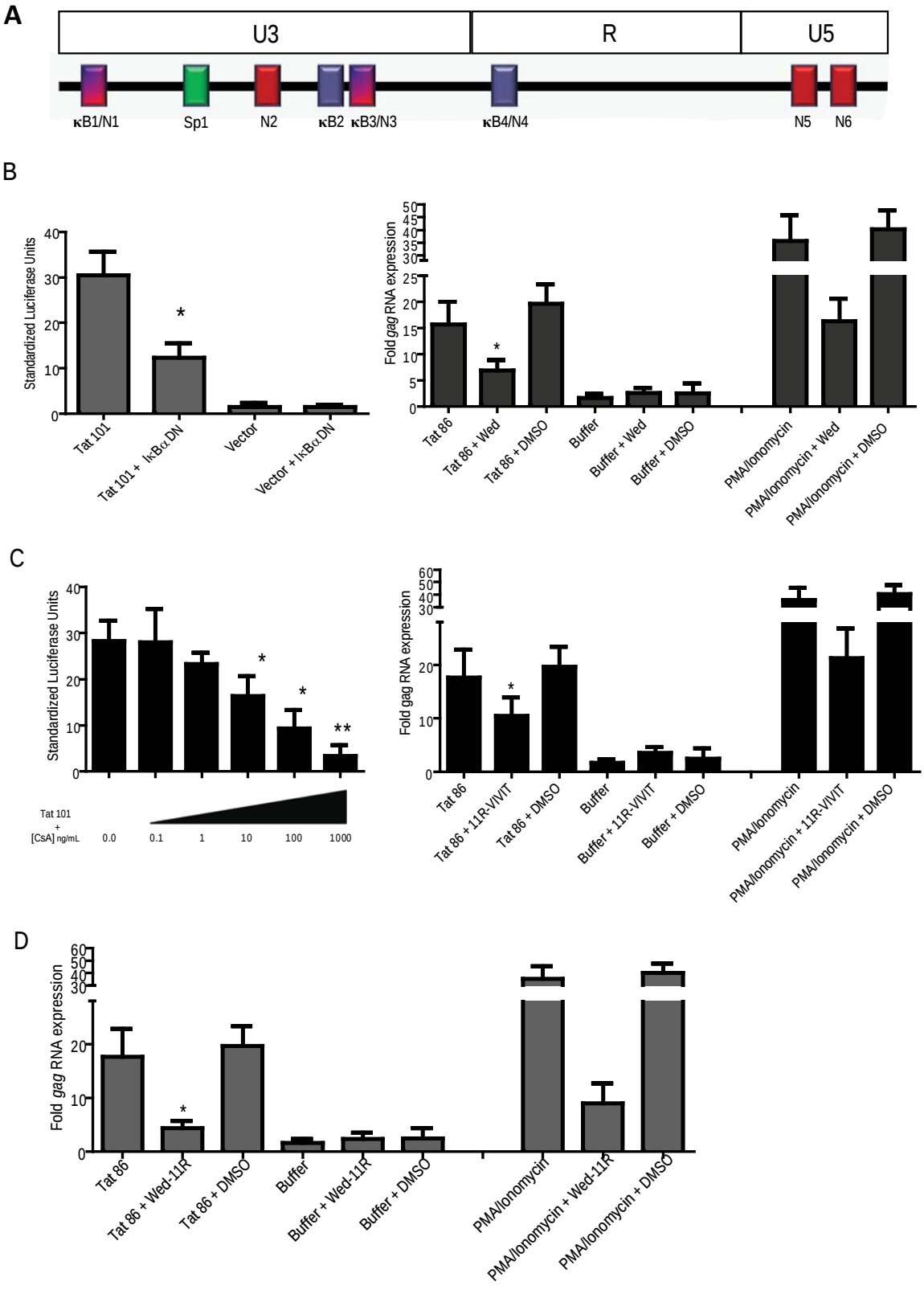

A
U3
R
U5
κB1/N1
Sp1
N2
κB2
κB3/N3
κB4/N4
N5
N6
B
Standardized Luciferase Units
Tat 101
Tat 101 + IκBα.DN
Vector
Vector + IκBα.DN
Fold gag RNA expression
Tat 86
Tat 86 + Wed
Tat 86 + DMSO
Buffer
Buffer + Wed
Buffer + DMSO
PMA/Ionomycin
PMA/Ionomycin + Wed
PMA/Ionomycin + DMSO
C
Standardized Luciferase Units
Tat 101 + [CsA] ng/mL
0.0
0.1
1
10
100
1000
Fold gag RNA expression
Tat 86
Tat 86 + 11R-VIVIT
Tat 86 + DMSO
Buffer
Buffer + 11R-VIVIT
Buffer + DMSO
PMA/Ionomycin
PMA/Ionomycin + 11R-VIVIT
PMA/Ionomycin + DMSO
D
Fold gag RNA expression
Tat 86
Tat 86 + Wed-11R
Tat 86 + DMSO
Buffer
Buffer + Wed-11R
Buffer + DMSO
PMA/Ionomycin
PMA/Ionomycin + Wed-11R
PMA/Ionomycin + DMSO

Figure 2.5 Inhibition of the NF-κB or NF-AT transcription factors suppresses Tat-mediated activation from the HERV-K (HML-2) LTR. *(*A) The HERV-K (HML-2) LTR promoter with U3, R, and U5 regions as well as potential transcription factor binding sites indicated: κB1/N1=first potential NF-κB/NF-AT binding site (one site embedded in the other), Sp1=potential Sp1 binding site, N2= second potential NF-AT binding site, κB2= second potential NF-κB binding site, κB3/N3= third potential NF-κB/NF-AT binding site, κB4/N4= fourth potential NF-κB/NF-AT binding site, N5 and N6= fifth and sixth potential NF-AT binding sites. (B) Left panel: A dominant negative construct coding for the Inhibitor of NF-κB alpha (IκB-α DN) was co-transfected into Jurkat T cells with the full-length HERV-K (HML-2) LTR-luciferase reporter and a Tat expression vector. Luciferase activity was measured 24 hours after transfection, normalized to *Renilla* luciferase, and expressed as standardized luciferase units. Right panel: Jurkat T cells were pre-treated for 1 hour with 10 μM Wedelolactone (a specific inhibitor of NF-κB) and then treated with recombinant HIV-1 Tat for 8 hours, followed by total RNA isolation and HERV-K (HML-2) *gag* amplification by qRT-PCR. (C) Left panel: Jurkat cells co-transfected with the full length HERV-K (HML-2) LTR-luciferase construct and Tat were subsequently treated with varying concentrations of cyclosporin A (CsA), and luciferase activity was measured after 24 hours, normalized and expressed as standardized units. Right panel: Jurkat T cells were pre-treated for 1 hour with 4 μM 11R-VIVIT (a specific inhibitor of NF-AT), and then treated with recombinant HIV-1 Tat for 8 hours, followed by total RNA isolation and HERV-K (HML-2) *gag* amplification by qRT-PCR. (D) Jurkat T cells were pre-treated for 1 hour with both 10 μM Wedelolactone and 4 μM 11R-VIVIT prior to an 8 hour recombinant Tat treatment and subsequent RNA amplification. PMA and Ionomycin treatments served as positive controls for transcription factor activation. Error bars indicate SD from three independent experiments. Significance was calculated using a Student's T-test comparing Tat-transfection/treatment in the absence of inhibitors to Tat-transfection/treatment in the presence of inhibitors. Significant results are indicated (*p=<0.05, **p=<0.01). MTT assays performed to assess the toxicity of the drugs used in the experiments showed no significant cell death.

To verify the physical interaction of NF-κB and NF-AT with the HERV-K (HML-2) promoter, we performed chromatin immunoprecipitation (ChIP) assays. After treating Jurkat T cells with HIV-1 Tat or PMA and Ionomycin for 1 hour, the cells were treated with formaldehyde for chromatin-protein cross-linking, lysed, and the cellular DNA was then purified. DNA-protein complexes were precipitated using anti- NF-κB, anti-NF-AT, or their respective IgG isotype controls, cross-links were reversed, and the DNA was precipitated and purified. We then performed qPCR on the regions of the HERV-K (HML-2) promoter where these transcription factors should bind, using 5 different sets of primers that span the sites of interest (Fig. 2.6A) and provide fragments of optimal size for DNA amplification for ChIP. By measuring the fold enrichment in amplification of the specific antibody immunoprecipitation (IP) over the IgG control IP in a specific region of the promoter of interest we can detect if the transcription factors under study bind to that particular region, and can compare the binding to other regions of the promoter. We detected that, in response to HIV-1 Tat, the order of enrichment (from highest to lowest) for NF-κB was as follows: the κB2-3/N2-3 position, followed by the κB1/N1 and then the κB4/N4 position (Fig. 2.6B). Thus, NF-κB appears to preferentially bind to the κB2-3/N2-3 position in the HERV-K (HML-2) promoter in response to Tat. Similar results were obtained for PMA and Ionomycin-induced NF-κB activation (Appendix 1.7). For NF-AT, enrichment is also highest at the κB2-3/N2-3 position, followed by the κB4/N4 position (Fig. 2.6B). Enrichment levels were of equal proportions when comparing positions κB1/N1 and N5 (Fig. 2.6B). No signal was detected when trying to amplify a product around the N6 position, suggesting that this is not an actual binding site for NF-

AT. Specificity of the IP was further assessed by amplification of the region containing the potential Sp1 transcription factor binding site. qPCR amplification of the Sp1 area never yielded a product or gave a strong signal above background, thus it served as a control for immunoprecipitation (Fig. 2.6B). Comparing the NF-AT IP to the NF-κB IP shows more enrichment for NF-AT in all sites except at the first position (which seems equal to NF-κB enrichment). These data show that there is activation of NF-κB and NF-AT in response to HIV-1 Tat, and that these two transcription factors interact directly with the HERV-K (HML-2) LTR promoter.

We next introduced site-directed mutations into the NF-κB, NF-AT, and/or Sp1 sites in the HERV-K (HML-2) promoter construct. We then transfected the mutated HERV-K (HML-2) promoter constructs with or without HIV-1 Tat into Jurkat T cells, and measured effects on luciferase reporter activity. It should be noted that some mutations of NF-κB sites also destroy the potential binding sites for NF-AT proteins, as they were embedded in their DNA sequence. Interestingly, site-specific mutation of most of these sites led to a clear decrease in Tat responsiveness when compared to the wild-type promoter (Fig. 2.6C). Although others have reported a role for Sp1 in regulation of the basal expression of HERV-K (HML-2) [140], mutation of the Sp1 site did not significantly decrease promoter activity in response to Tat, consistent with our ChIP data (Fig. 2.6B). Whereas mutation of the fourth NF-κB/NF-AT (κB4/N4) binding site did not affect the response to Tat, mutation of the first NF-κB/NF-AT (κB1/N1) and the second NF-κB (κB2) sites led to a decrease in activity (approximately 2 to 2.5-fold). The most

pronounced decrease occurred when the first NF-AT site of the U5 region (N5) was mutated alone (approximately 3.8-fold), further suggesting that NF-AT is a crucial modulator of the response of HERV-K (HML-2) to HIV-1 Tat. Mutation of both potential NF-AT sites in the U5 region simultaneously did not affect the response to Tat any more than did mutation of the single N5 site (data not shown), suggesting that the N6 site is not a functional NF-AT binding site, which is consistent with the ChIP data. Taken together, these data show that both NF-κB and NF-AT are important for HIV-1 Tat-mediated activation of the HERV-K (HML-2) promoter, that one transcription factor may compensate for loss of the activity of the other, and that multiple NF-AT and NF-κB binding sites mediate much of the response to Tat.

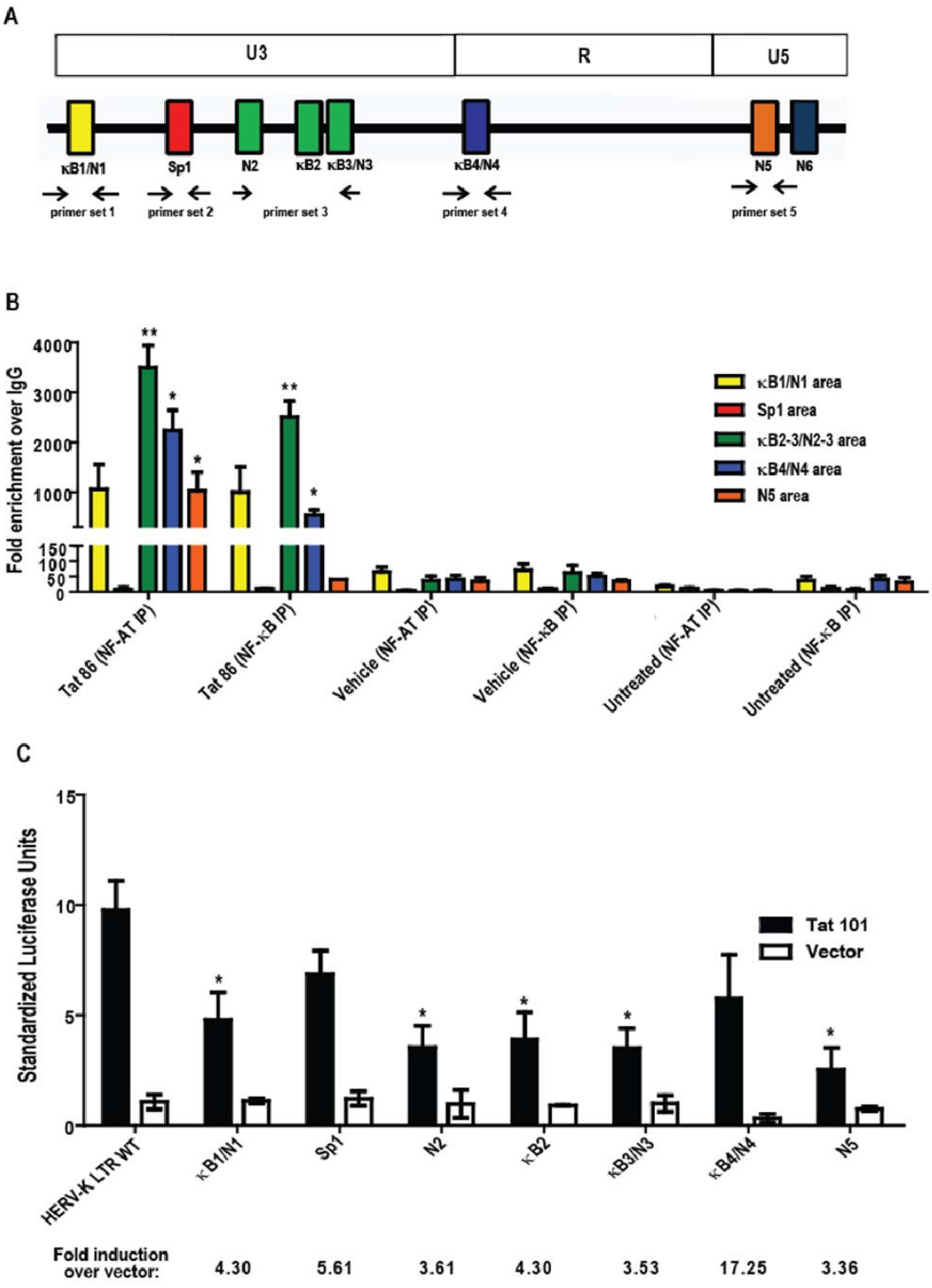

A
U3
R
U5
κB1/N1
Sp1
N2
κB2
κB3/N3
κB4/N4
N5
N6
primer set 1
primer set 2
primer set 3
primer set 4
primer set 5
B
Fold enrichment over IgG
4000
3000
2000
1000
150
100
50
0
κB1/N1 area
Sp1 area
κB2-3/N2-3 area
κB4/N4 area
N5 area
Tat 86 (NF-AT IP)
Tat 86 (NF-κB IP)
Vehicle (NF-AT IP)
Vehicle (NF-κB IP)
Untreated (NF-AT IP)
Untreated (NF-κB IP)
C
Standardized Luciferase Units
15
10
5
0
Tat 101
Vector
HERV-K LTR WT
κB1/N1
Sp1
N2
κB2
κB3/N3
κB4/N4
N5
Fold induction over vector:
4.30
5.61
3.61
4.30
3.53
17.25
3.36

Figure 2.6 Tat responsive elements in the HERV-K (HML-2) promoter. Chromatin immunoprecipitation was performed on the cellular HERV-K (HML-2) LTR promoter in Jurkat T cells using antibodies specific for NF-κB or NF-AT (or non-specific IgG isotype controls) 1 hour after treatment with either recombinant HIV-1 Tat protein or PMA and Ionomycin. qPCR was performed using primer sets designed to target areas of the promoter where the potential transcription factor binding sites are present. (A) Schematic of primer binding areas in the HERV-K (HML-2) promoter. (B) ChIP analysis of potential NF-κB and NF-AT binding sites with data expressed as fold enrichment over IgG. Bars are color-coded to match primer binding areas. Error bars indicate SD from three independent immunoprecipitation experiments. (C) Potential transcription factor binding site mutations in the HERV-K (HML-2) LTR and their response to HIV-1 Tat in Jurkat T cells. WT=Wild Type, κB=NF-κB, N=NF-AT. Luciferase activation was measured 48 hours after transfection. Control experiments included a mock transfection and transfection of an empty vector (pcDNA3.1). Error bars indicate SEM from three independent transfections. Significance was calculated using a Student's T-test comparing Tat treatments to vehicle controls, or full-length LTR activity to mutant-LTR activity. Significant results are indicated (*p=<0.05, **p=<0.005).

Discussion

Eight percent of the human genome consists of fixed retroviral elements, termed HERVs, acquired throughout thousands of years of human evolution. Although most HERV genes seem to be non-functional, there are some endogenous retroviral genes that are expressed, even becoming important contributors to human physiological processes, as can be seen in trophoblast development [58, 109-111]. However, HERV gene expression appears to be tightly controlled under normal circumstances, and only when cellular fitness or integrity is compromised (e.g. with synthetic chemical agents, radiation, stress, cytokines/chemokines, or biological agents acting upon human cells) is there pronounced or increased expression of endogenous retroviral elements [46, 113-116, 285, 323]. Under conditions where an exogenous virus, like HIV-1, establishes a productive infection, the repression of HERV expression can be lifted [192]. HIV-1-infected individuals have abnormally high levels of the endogenous retrovirus HERV-K (HML-2) expressed in cells, and vastly increased RNA titers are present in their plasma [112, 190-192, 257]. The consequences of this activation in the setting of HIV-1 infection are still a matter of speculation, but since HERV-K (HML-2) encodes two putative oncogenes and has been associated with several autoimmune, inflammatory and neurological diseases, it could potentially participate in the development of HIV-1-associated disease [165, 168, 170, 255, 283, 288, 289]. Understanding the initial steps in the activation of HERV-K (HML-2) expression after HIV-1 infection is the first step

towards deciphering how HERV-K (HML-2) elements might play a role in HIV-1 pathogenesis.

Previous research has shown how several proteins from HIV-1 interact with HERV-K products. The HIV-1 protease, for example, can cleave HERV-K gag [324, 325]. Additionally, HIV-1 Rev can actively transport HERV-K RNA from the nucleus to the cytoplasm of cells that express it [248]. However, these interactions likely only partially explain the massive increase in HERV-K (HML-2) RNA expression seen in HIV-1-infected patients [190, 191, 257]. Therefore, we reasoned that another protein from HIV-1 might be involved in stimulating HERV-K (HML-2) expression. As the HIV-1 Tat protein has been shown to be a potent activator of viral [281] and cellular gene expression [278, 300], we thought it to be a likely candidate to stimulate HERV-K (HML-2) transcription.

Here we confirm that HIV-1 infection leads to increased expression of HERV-K (HML-2) and show that the HIV-1 Tat and Vif proteins activate HERV-K (HML-2) expression at the RNA level, with Tat also causing an increase in the expression of at least one HERV-K (HML-2) protein, Capsid. Additionally, treating primary lymphocytes (which are more biologically relevant than continuously passaged cell lines) with HIV-1 Tat also results in increased HERV-K (HML-2) RNA expression. This effect of Tat on HERV-K (HML-2) activation was pronounced, but dependent on the cell-type. For cells that constitutively express high levels of HERV-K (HML-2) transcripts and proteins (like teratocarcinoma cell lines), little response to Tat was detected at the RNA level. However, in cell lines normally expressing low levels of HERV-K (HML-2), or in primary

cells with undetectable or extremely low levels of transcripts, Tat treatment upregulated transcription of these proviral sequences. Importantly this was seen in primary lymphocytes, a major target of HIV-1 infection. This activation occurred whether Tat was added exogenously to the cells or if the cells were transfected with constructs encoding it. The effect occurred at the level of the transcriptional promoter, with cooperative involvement between Tat and the cellular transcription factors NF-κB and NF-AT. As the promoter also contains a number of other potential transcription factor binding sites, we cannot discount the possibility of other factors additionally contributing to activation in the presence of Tat. We must note that other biological agents, chemicals, exogenous infections or stress conditions can activate HERV expression in cells [122-127] and that, while Tat is one important mechanism for HERV-K activation during HIV-1 infection, it is but one contributor among the whole spectrum of factors that cause an increase in HERV expression in a biological system. Interestingly, a recent study has found that the HTLV-I Tax protein has been shown to potently activate the LTRs of different HERVs, including HERV-K [129], demonstrating that other exogenous retroviral transactivators can modulate endogenous retroviral expression. Lastly, we should point out that in the future it will be necessary to verify which HERV-K (HML-2) proteins are expressed at a higher level in the presence of Tat, but the current status of the available antibodies precludes quantitatively accurate experiments.

In summary, our research strongly suggests a role for the HIV-1 Tat protein in the activation of HERV-K (HML-2) transcription, which helps to explain the basis for the

increased expression of HERV-K (HML-2) RNA seen during HIV-1 infection. Interestingly, though our data show marked increases in RNA expression in response to Tat treatment, the degree of stimulation still does not reflect the very high increase in the number of HERV-K (HML-2) RNA copies seen in the plasma of HIV-1-infected patients as compared to uninfected individuals [191, 257]. Thus, other HIV-1 proteins could also increase production of HERV-K (HML-2). For example, as we have demonstrated above, Vif could contribute to activation of HERV-K (HML-2) in patients. Further, a potential cooperativity between HERV-K (HML-2) Rec and HIV-1 Rev (both of which can transport HERV-K (HML-2) RNA out of the nucleus) could lead to increased HERV-K (HML-2) protein expression. Thus, additional interactions between HIV-1 and HERV-K likely contribute to the marked increase in HERV-K (HML-2) expression seen in patients with HIV-1 infection.

Materials and Methods

Plasmid constructs

The HIV-1 molecular clone pNL4-3 has been previously described (NIH AIDS Research and Reference Reagent Program. Dr. Malcolm Martin, [326]). The HIV-1 Tat coding plasmids pcDNA3.1-Tat72 and pcDNA3.1-Tat86 were made from the parent vector pcDNA3.1+/Tat101-flag (PEV280) (NIH AIDS Research and Reference Reagent Program, Dr. Eric Verdin, [327]) by PCR amplification. The Tat mutants Tat C22G and Tat T23N were made through site-directed mutagenesis of the pcDNA3.1+/Tat101-flag parent vector. The pcDNA3.1(+) empty vector control was made by releasing the Tat insert with *Bam*HI followed by religation of the backbone. The HIV-1 Rev coding plasmid was made by releasing the *rev* insert from the pRev-1 plasmid construct (NIH AIDS Research and Reference Reagent Program, Drs. Marie-Louise Hammarskjöld and Dr. David Rekosh, [328]) with *Bam*HI, and ligation into pcDNA3.1(+). The HIV-1 Nef protein is also expressed from pcDNA3.1(+), containing the coding region of HIV-1_{SF2} Nef, and has been previously described (NIH AIDS Research and Reference Reagent Program, Drs. J. Victor Garcia and John Foster, [329]). The HIV-1 Vif and Vpu proteins were expressed from pcDNA3.1(-), with both being codon optimized for expression in human cells (called HVif and Vp(h)u) and are further described elsewhere (NIH AIDS Research and Reference Reagent Program, Drs. Stephan Bour and Klaus Strebel, [330]). Full length Vpr was cloned into the peGFP-C3 expression vector (Promega, Madison, WI), and is fused to GFP (NIH AIDS Research and Reference Reagent Program, Dr. Warner Greene, [331]). A

vector that expresses HIV-1 Gag derived from NL4-3 in an HIV-1 Rev-dependent manner was a kind gift from Dr. Akira Ono at the University of Michigan, and has been previously described (pCMVNLGagPolRRE, [332]). pCMV (pCMV-PL, Dr. Bryan Cullen) was obtained from Addgene (Cambridge, MA). Full-length Env from HIV-1_{HXB2} was cloned into pSV7D, termed pHXB2Env for simplicity, and has been described previously (NIH AIDS Research and Reference Reagent Program, Dr. Kathleen Page and Dr. Dan Littman, [333]). The molecular clone pHXB2 (pHXB2gpt) was a kind gift from Dr. F. Wong-Staal [334, 335]. The molecular clone pMtat(-) contains a termination codon (TGA) in place of the ATG (methionine) initiator codon in the Tat coding region, resulting in a mutant unable to synthesize Tat (NIH AIDS Research and Reference Reagent Program, Dr. Reza Sadaie, [336]). The HERV-K (HML-2) LTR reporter construct (HERV-K LTR-luc) contains a partial version of the HERV-K (HML-2) LTR (GenBank accession no. AF394944) cloned in front of the HSV-1 tk minimal promoter in the pT81 luciferase vector [118], and was kindly provided by Dr. Kyung Lib Jang (Pusan National University, South Korea). The *Renilla* luciferase plasmid, pRL-CMV, was obtained from Promega (Madison, WI). The luciferase construct "HERV-K LTR", used in mutational analyses, was made by addition through PCR amplification of the consensus sequence TGTGGGGAAAAGCAAGAGA to the 5' end of the partial promoter sequence of the HERV-K (HML-2) LTR from pT81, and subcloning it into the *Xho*I and *Hind*III sites of pGL4.10[*luc2*] (Promega, Madison, WI), which does not contain a tk minimal promoter. Site-directed mutagenesis was performed on the LTR region using the QuikChange® Lightning Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA). NF-κB binding site

mutations were introduced into the general consensus sequence GGGRNYYYCC (where R=purine, Y=pyrimidine and N= any nucleotide) as G***CTCTA***YYCC. NF-AT binding site mutants were obtained by mutating the GGAAA region of the consensus sequence (A/T)GGAAA(A/N)(A/T/C)N to ***CTCTA***. In the cases were NF-AT sites were embedded in NF-κB sequences, mutation to ***CTCTA*** was used to eliminate both sites simultaneously. For analyses involving the potential Sp1 binding site, two binding site mutants were generated and tested, taking the consensus sequence GGGCGG(G/A)(G/A)(C/T) and changing it to either G***A***G***ATC***TGC or ***TT***G***A***GGTGC. PCR primers for introduction of mutations were designed using the QuikChange® Primer Design Program available at http://www.stratagene.com/sdmdesigner/default.aspx. Sequences of the plasmids were confirmed by DNA sequencing.

Cell culture and transient transfections

Except for 293FT (a fast growing cell line derived from HEK-293 that contains the SV40 large T antigen, Invitrogen, Carlsbad, CA), H9 (NIH AIDS Research and Reference Reagent Program, Dr. Robert Gallo, [337]), and Jurkat-Tat T cells (stably expressing HIV-1 Tat, NIH AIDS Research and Reference Reagent Program, Drs. Antonella Caputo, William Haseltine, and Joseph Sodroski, [338]), all cell lines used were obtained from the American Type Culture Collection (ATCC, Manassas, VA). All media were obtained from Gibco (Invitrogen, Carlsbad, CA). 293FT were maintained in DMEM, supplemented with 10% fetal bovine serum (FBS, Gibco/Invitrogen, Carslbad, CA) and 100 U/mL penicillin

and streptomycin (Gibco/Invitrogen, Carlsbad, CA), at 37 °C in a 5% CO_2 incubator. Jurkat T cells, U-937, HUT-78, H9 (a derivative of HUT-78) and NCCIT cells were maintained in complete RPMI-1640 media, supplemented as stated above. Jurkat-Tat T cells were maintained in complete RPMI-1640 supplemented with 10% FBS, 800 µg/mL G418 (Invitrogen, Carlsbad, CA), and 100 U/mL penicillin and streptomycin. When needed, cells were cultured under stimulatory conditions with PMA (Phorbol 12-Myristate 13-Acetate) and Ionomycin, with DMSO (Dimethyl Sulfoxide) as their vehicle buffer (all from Sigma-Aldrich, St. Louis, MO). Cells with greater than 90% viability were transfected with endotoxin-free plasmids (2 to 5 µg HIV-1-Tat, Vif, Nef, Rev, Vp(h)u, Vpr, GagPol plasmids, or 2 to 5 µg empty/control vectors, or 5 µg of pHXB2 or pMtat(-) proviral molecular clones) using the transfection reagents Lipofectamine 2000 (Invitrogen, Carlsbad, CA), SuperFect (Qiagen, Valencia, CA) or FuGENE HD (Roche, Indianapolis, IN) as per the manufacturer's protocol. Control experiments included mock transfections.

HIV-1 Production and Infection

Infectious HIV-1 was produced by calcium phosphate-mediated transfection of 293FT cells [339] using pNL4-3 [326]. Tissue culture media was harvested at 24 h, 36 h, and 48 h post-transfection, pooled, and filtered (0.2 µm) to remove cells and large cell debris. Media containing HIV-1_{NL4-3} was frozen and strored at -80 °C until used for infections. The relative concentration of virus in the stocks was determined by RT activity [340] and

the ability to induce luciferase activity upon infection of TZM-Bl cells [340]. HIV infection of cells (2 x 10^6) was conducted using 2 mL of virus in a standard spin infection technique (1048.6 *x g*) for two hours at room temperature. Infected cell cultures were diluted in T25 flasks and maintained between 0.5 to 1 x 10^6 cells per ml for 7 days upon which fresh uninfected cells were added and cells were harvested after an additional week. Cellular RNAs were extracted using TriZol (Invitrogen, Carlsbad, CA). HIV-1 *env* amplification was done to confirm infection status with the following primer sequences, in a One-Step RT-PCR reaction:

HIV-1 Env F 1493-1516 5'-AGGCAAAGAGAAGAGTGGTGCAGA-3'

HIV-1 Env R 1643-1666 5'-CCCTCAGCAAATTGTTCTGCTGCT-3'

Luciferase Assays

All transfections included *Renilla* luciferase as an internal control (100 ng of pRL- CMV) to assess for variation in transfection efficiency. Transfected cells were assayed for luciferase activity using the Dual-Luciferase Assay Kit (Promega, Madison, WI), 4, 6, 8, 12, 24 and 48 hours after transfection in a Tecan GENios luminometer plate reader (Phenix Research Products, Candler, NC). Data were normalized to *Renilla* luciferase signal, and expressed as standardized luciferase units. The amounts of plasmids used for transfections were as follows: 2.5 µg luciferase reporter plasmid (HERV-K (HML-2) LTR or HIV-1 LTR), with 2 to 5 µg of Tat-expression plasmid or 2 to 5 µg of empty/control

vector. Control experiments included mock transfections (Lipofectamine 2000 reagent alone) and no transfection (for background subtraction).

Isolation and Culture of Primary Cells

Peripheral blood mononuclear Cells (PBMCs) were obtained by venipuncture from healthy donors and monocytes were separated from peripheral blood lymphocytes (PBLs) by differential adhesion to plates as previously described [341]. PBLs were washed three times with PBS and stimulated with 5 µg/mL phytohemagglutinin (PHA-P, Sigma-Aldrich, St. Louis, MO) for 3 days in RPMI-1640 complete media containing 10% heat-inactivated FBS and 10 U/mL of Interleukin-2 (IL-2, Sigma-Aldrich, St. Louis, MO).

Addition of exogenous Tat protein

The purified recombinant 86 amino acid form of the HIV-1 Tat protein was obtained from the NIH AIDS Research and Reference Reagent Program (the late Dr. John Brady and DAIDS, NIAID, [342]), or from ProSpec Protein Specialists (Cat. No. HIV-129, ProSpec-Tany TechnoGene Ltd., East Brunswick, NJ). The protein was resuspended in sterile phosphate-buffered saline (PBS, Gibco/Invitrogen, Carlsbad, CA) containing 1 mg/mL bovine serum albumin (BSA) and 0.1 mM dithiothreitol (DTT) (both from Sigma-Aldrich, St. Louis, MO), de-aerated, and protected from light. Tat protein was added in a range of concentrations to cells, for the specified time points as indicated in the text and figure legends.

RNA extraction and Real Time RT-PCR

Total cellular RNA was isolated from cells using the RNeasy Mini Kit (Qiagen, Valencia, CA), and subjected to RNase-free DNase treatment (Qiagen, Valencia, CA) for 15 minutes at room temperature. RNA concentration and purity were measured using a spectrophotometer, calculating the 260/280 ratio. RNA integrity (as well as the absence of DNA contamination) was confirmed by One-Step RT-PCR using GAPDH amplification with primers that can bind both genomic and cDNA, employing the PCR conditions described below, as well as "No RT" controls. If DNA contamination was detected, another round of DNase treatment was performed using the DNA-free DNase removal kit (Ambion, Austin, TX) following the manufacturer's protocol. To assess the difference in HERV-K (HML-2) RNA transcript expression levels, we performed quantitative Real Time RT-PCR using the QuantiTect SYBR Green RT-PCR kit (Qiagen, Valencia, CA), or the Bio-Rad iScript One-Step RT-PCR kit with SYBR Green (Bio-Rad, Hercules, CA). Briefly, 100-500 ng of total cellular RNA and 0.5 μM each of HERV-K (HML-2) *gag* primers (forward primer 5'-AGCAGGTCAGGTGCCTGTAACATT-3' and reverse primer 5'-TGGTGCCGTAGGATTAAGTCTCCT-3'), HERV-K (HML-2) *rec* primers (forward primer 5'-ATCGAGCACCGTTGACTCACAAGA-3' and reverse primer 5'-GGTACACCTGCAGACACCATTGAT-3'), or HERV-K (HML-2) *np9* primers (forward primer 5'-AGATGTCTGCAGGTGTACCCA-3' and reverse primer 5'- CTCTTGCTTTTCCCCACATTTC-3') were used in a 20 μL reaction. RNA was reverse transcribed for 30 min at 50 °C. The PCR reaction consisted of an initial denaturing step of 15 min at 95 °C, followed by 35 to

40 cycles with optimal conditions as follows: 94 °C for 15 sec, 60 °C for 30 sec, 72 °C for 10 sec, and an optimized data collection step at 81 °C for 10 sec (for Gag and Rec amplification) or 78 °C for 10 sec (for Np9 amplification). Fluorescence captured at 78 °C or 81 °C was determined to be absent of signal generated by primer dimers by a melting curve analysis. Data were collected and recorded by iCycler iQ software (Bio-Rad, Hercules, CA). GAPDH amplification was used to normalize samples to an endogenous reference gene, as stated in the figure legends.

Western Blot and Protein Band Analysis

The antibodies used in this study, and their respective dilutions, are as follows: mouse anti-HERV-K Gag (HERM-1841-5, 1:1,000 dilution) and mouse anti-HERV-K Capsid (HERM-1831-5, 1:200 dilution) both from Austral Biologicals; anti-β-actin-HRP (conjugated to horseradish peroxidase, 1:25,000 dilution, Abcam, Cambridge, MA); and mouse anti-HIV-1 Tat (1D9, NIH AIDS Research and Reference Reagent Program, Bethesda, MD). After transfection, or recombinant protein addition, Jurkat T cells were washed twice with PBS and then lysed with hot 2% SDS buffer (Sodium Dodecyl Sulfate, Fisher Scientific, Pittsburgh, PA). Cell lysates were boiled and centrifuged to eliminate DNA-associated viscosity, and protein concentration was measured using the Pierce BCA Protein Assay Reagent kit (Pierce/Thermo Scientific, Rockford, IL). Equal protein concentrations were loaded and separated on 15% SDS-polyacrylamide gels, and blotted onto polyvinylidene difluoride (PVDF) membranes. The membranes were blocked with

5% non-fat dry milk in PBS with 0.1% Tween-20 (PBST, Fischer Scientific, Pittsburgh, PA) for 2 hours at room temperature. All antibodies used were incubated in blocking solution with the blotted membranes overnight at 4° C, with the exception of anti-β-actin-HRP (2 hours at room temperature). The membranes were washed 3 times in PBST and blocked with 5% goat serum for 30 minutes at room temperature. When necessary, the bound primary antibody was incubated with a horseradish peroxidase-conjugated goat anti-mouse secondary antibody for 1 hour. Signal was detected using the Super Signal West Pico system (Pierce/Thermo Scientific, Rockford, IL). Semi-quantification of protein levels was performed by digitization of X-ray films using the Typhoon FLA 7000 scanner (GE Healthcare Life Sciences, Pittsburgh, PA) and subsequent analysis of the gray values of the bands in the resulting images. The ImageQuant TL software was used for analysis of the digitized Western blot images. This software allows the measurement of band volume (average optical density of the band times its area) of the band of interest. Background subtraction was performed using the "Image Rectangle" method as a user-defined area within the image (size as well as position), and once defined it was applied to all lanes of interest. Bands were fitted as tightly as possible, and band volumes are plotted as arbitrary units.

Identification of potential transcription factor binding sites in the LTR of HERV-K (HML-2)

Analyses of the HERV-K (HML-2) LTR for potential transcription factor binding sites were performed using the online prediction software tools ALGGEN PROMO and version 8.3 of TRANSFAC (BioBase Co., Beverley, MA, [310, 311]).

NF-κB Inhibition Assays

The expression vector pIκBαM (kindly provided by Dr. Paul J. Chiao (MD Anderson Cancer Center, Houston, Texas)), and here referred to as IκBα DN (dominant-negative), was co-transfected into Jurkat T cells with the HERV-K (HML-2) LTR-luciferase construct with or without Tat (using *Renilla* luciferase as an internal transfection control). Luciferase activity was measured 48 hours after transfection. IκBα DN encodes a phosphorylation-site and a degradation-site mutant IκBα chain, inhibits translocation of NF-κB from the cytosol to the nucleus upon activation, and has a dominant negative effect on NF-κB function [343-345].

NF-κB inhibition was also accomplished through a one hour pre-treatment of experimental samples with 10 μM Wedelolactone in DMSO (7-Methoxy-5,11,12-trihydroxycoumestan, Sigma-Aldrich, St. Louis, MO), followed by treatment with recombinant Tat protein or its buffer control. Wedelolactone specifically inhibits NF-κB-mediated gene transcription in cells by blocking the phosphorylation and degradation of IκBα [321].

NF-AT Inhibition Assays

Jurkat T cells co-transfected with the HERV-K (HML-2) LTR reporter construct and the Tat expression vector were treated with varying doses of the immunosuppressant drug Cyclosporin A (Sigma-Aldrich, St. Louis, MO) 24 hours after transfection, and luciferase activity was measured 48 hours after transfection, as noted above and in the figure legends.

Specific NF-AT inhibition was accomplished through a one hour pre-treatment of experimental samples with 1 μM 11R-VIVIT in DMSO (Calbiochem, La Jolla, CA), followed by treatment with recombinant Tat protein or its buffer control. 11R-VIVIT is a cell-permeable version of the specific NF-AT Inhibitor (VIVIT) that is modified at the C-terminus with an 11-arginine transduction domain and a 3-glycine linker sequence [322].

Chromatin Immunoprecipitation Assays

To examine interactions of NF-κB and NF-AT with the HERV-K (HML-2) LTR, we performed Chromatin Immunoprecipitation (ChIP) assays using the ChIP-IT Express Enzymatic kit (Active Motif, Carlsbad, CA) according to the manufacturer's protocol. Briefly, experimental samples were lysed and subjected to enzymatic shearing of the DNA (random cleaving). Digestion was performed for four hours at 37 °C, with intermittent vortexing. Two micrograms of specific antibody for NF-κB (H-119X, Santa

Cruz Biotechnology, Santa Cruz, CA), NF-AT (7A6, Santa Cruz Biotechnology, Santa Cruz, CA), Vimentin (V9, Santa Cruz Biotechnology, Santa Cruz, CA), or the IgG isotype control antibodies (Rabbit IgG or Mouse IgG, Sigma-Aldrich, St. Louis, MO) was used for each immunoprecipitation, with overnight incubation. Protein G beads were added to the overnight incubation as well. Beads were washed three times in ChIP Buffer, after which elution of the digested chromatin, reversion of cross-linking, and proteinase K treatment was performed. Immunoprecipitated DNA was detected by PCR, using 5 µL of eluate as a template, to verify success of the precipitation. For amplification of the HERV-K (HML-2) LTR transcription factor binding site areas of interest, the following primers were used (their binding areas are depicted in Fig.6):

KLTR ChiP primer set 1 fwd 5′-TGTGGGGAAAAGCAAGAGA-3′

KLTR ChiP primer set 1 rev 5′-GGTCACAGAATCTCAAGGCAG-3′

KLTR ChiP primer set 2 fwd 5′-GTGACCTTACCCCCAACCCCG-3′

KLTR ChiP primer set 2 rev 5′-TGTTTAACAAAGCACATCCTGC-3′

KLTR ChiP primer set 3 fwd 5′-CTGCCTAGGAAAGCCAGGTA-3′

KLTR ChiP primer set 3 rev 5′-CGGGTATCGGGCTGGGGGACG-3′

KLTR ChiP primer set 4 fwd 5′-CCCTGGGCAATGGAATGTCTCG-3′

KLTR ChiP primer set 4 rev 5′-GCTGCCCGCAGGTCCCACCTC-3′

KLTR ChiP primer set 5 fwd 5′-TGGTTCCCCGGGTCCCCTTAT-3′

KLTR ChiP primer set 5 rev 5′-CCTACACACCTGTGGGTGTTT-3′

Real-Time PCR was performed with the Bio-Rad iQ SYBR Green Supermix (Bio-Rad, Hercules, CA) with 0.3 µM final primer concentration in 20 µL final reaction volume, and the following cycle conditions: 1 cycle of 95 °C for 3 min; 40 cycles of 95 °C for 15 s, 57 °C for 30 s, and 72 °C for 30 s; and 1 cycle of 72 °C for 10 min. A melt curve analysis was also performed to verify specific product amplification. Primers targeting the HERV-K *env* gene and the IL-6 promoter were used as controls, with PCR conditions as stated above.

Statistical Analysis

The mean number of HERV-K (HML-2) mRNA and standardized luciferase units between Tat treatments and controls were compared using an independent student's *T* test for samples exhibiting normally distributed values. Two-tailed *p* values were considered significant at $p<0.05$.

CHAPTER III

HIV-1 TAT CAUSES THE ACTIVATION OF A NOVEL CENTROMERIC HUMAN ENDOGENOUS RETROVIRUS TYPE K (HML-2) DISCOVERED THROUGH HIV INFECTION

Abstract

Human Endogenous Retroviruses (HERVs) entered hominid species millions of years ago and have been transmitted in a Mendelian fashion throughout evolution. The HERV-K (HML-2) family is the most recent of these viruses to enter the hominid genome, and today accounts for approximately 3000 proviral fragments. Activation of HML-2 proviruses occurs in response to the HIV-1 Tat protein, has been detected in the blood of patients with HIV-1 infection and with certain types of cancers, and leads to the production of viral RNA, proteins, and viral-like particles (VLPs). Analysis of these expressed viruses has shown that most sequences are the result of recombination events and, most interestingly, have revealed the presence of a newly discovered subset of these HERV-K (HML-2) viruses, here termed K111, which are not annotated in the most recent human genome assembly. These K111 proviruses were revealed to be inserted into the centromeric or pericentromeric areas of several different human chromosomes, in specific Centromeric Repeat sequences (CER-

elements). These centromeric areas are characterized by dense chromatin structures, which are proven to restrict gene expression, yet infection with HIV-1 or stimulation with the HIV-1 Tat protein leads to the activation of K111 proviruses inserted in these areas. As we show by Chromatin Immunoprecipitation assays, HIV-1 Tat induces loss of heterochromatin in pericentromeric regions, thus opening up the chromatin structure and allowing for active expression of K111. As a result, by studying HERV-K (HML-2) sequences in the blood of patients we have discovered at least 100 new K111 endogenous retroviruses scattered across the centromeres of fifteen human chromosomes. K111 expression appears to be activated solely by HIV-1 infection, and the Tat protein plays a key role in the stimulation of K111 expression.

Introduction

HERV-K (HML-2), has replicated during the evolution of hominids by re-infection [53], and today accounts for approximately 3000 proviral fragments [346, 347], out of which at least 91 are full-length [43]. More than 2500 HML-2 elements also exist as solitary LTRs (soloLTR), which originated by recombination between the 5' and 3' LTRs of full-length proviruses, removing the internal viral genes [282]. Recombination events have certainly allowed HERV-K (HML-2) to become quite diverse, both in number and in sequence, and through the analysis and study of these recombined viruses expressed during HIV-1 infection we have found even more HML-2 proviruses and soloLTRs than previously believed to be in the genome. Of most interest is the fact that these newly discovered viruses are present in centromeric areas of several different chromosomes and, in our studies, seem to be activated only during HIV-1 infection or in the presence of HIV-1 Tat.

As stated in previous chapters of this thesis, activation of HERV-K (HML-2) proviruses is detected in the blood of patients with HIV-1 infection and with certain types of cancers (such as lymphoma and breast cancer) [131, 189, 190]. This activation is strongly influenced by both the Tat and Vif proteins from HIV-1, but more so by Tat, as shown previously in Chapter 2 of this thesis. Isolation of VLPs from the blood of HIV-1-infected individuals, and sequencing of their genomic material, has revealed the presence of at least 32 different HERV-K (HML-2) proviruses [131]. However, further analysis of the HML-2 sequences found in the blood of HIV-1 patients revealed a novel group of HERV-K (HML-2) viruses we termed K111, which are not found in the latest version of the human genome sequence assembly. This interesting fact led us to the assignment of the newly discovered K111 proviruses, through DNA sequencing, at 15 different chromosomal locations. We also observed that multiple copies of K111 proviruses are found in the centromeres of these 15 chromosomes in modern humans. Finally, we further explored the mechanism for the activation of K111 during HIV-1 infection, and found that the HIV-1 Tat protein is responsible, at least in part, for the expression of K111 by inducing loss of heterochromatin in pericentromeric regions. This mechanism of activation is different than the one previously seen involving the transcription factors NF-κB and NF-AT, as the chromosomal location of these newly discovered proviruses is in an inaccessible and condensed area of chromatin. By studying endogenous retroviruses expressed in the blood of living patients, we have uncovered a family of new centromeric viruses hidden in the human genome that can be activated by the HIV-1 Tat protein.

Results

A new HERV-K (HML-2) virus, K111, is discovered in the plasma of HIV-1-infected individuals and is only activated during HIV-1 infection

In experiments performed by colleagues in my laboratory, that go beyond the scope of this thesis but will be summarized here, a phylogenetically distinct HERV-K (HML-2) virus branch was found through sequence analysis of nucleic acids isolated from the plasma of HIV-1-infected individuals (Fig. 3.1A and [131]). In order to better identify these viruses we looked at the current version of the annotated human genome assembly and to our surprise found that these distinct sequences were not present in it. Therefore we referred to the newly discovered virus as K111 and examined whether similar sequences were found in other species closely related to humans. Using this approach, a full-length K111 provirus was indeed found in the chimpanzee genome and will be referred to in here as CERV-K111. Using the CERV-K111 sequence as a starting point, primers were designed to amplify human K111. Phylogenetic reconstruction of HERV-K sequences found in the plasma of HIV-1-individuals revealed the existence of a unique branch, K111, with several K111-related viral RNAs (Fig. 3.1A). Further sequence analysis and quantitation of K111 *env* RNA titers by qRT-PCR in the plasma of patients with HIV-1 and other diseases revealed that these viruses are

found exclusively in the blood of HIV-1 infected patients and not in breast cancer or lymphoma patients (Fig. 3.1B and [131]). However, while K111 RNA was found only in the blood of HIV-1-infected patients, K111 exists at the genomic level as both full-length proviruses and soloLTRs in all human samples tested (schematic can be seen in Fig. 3.1C). BLAST analysis revealed that K111 integrated into the centromeric repeat CER:D22Z3, which

has been assigned to the centromere of chromosome 22 of humans [348, 349], and created a characteristic GAATTC target site duplication after proviral integration. Analyzing the sequences of the flanking CER and LTR of K111 by BLAST to the human genome database revealed four K111-related insertions. Two of them are soloLTR, and were assigned to the centromeres of chromosomes 9 and 22, but have not been annotated in the human genome draft (GRCh37/hg19; Feb. 2009). Two other K111-related full-length proviruses, one termed K112 by us and the other traditionally termed K105, were assigned to the pericentromere and the centromere of chromosome 21, respectively. A detailed BLAST analysis of other human centromeres revealed sequence similarity with K111 and its flanking CERs; sequence similarity was also found between K111 and the repeated region of a telomere.

CER:D22Z3-like elements are present in the centromeres (and one telomere) of approximately 10 chromosomes, as found in public sequence databases. Thus, to ascertain whether K111-related insertions are present in any other human chromosomes, DNA from human/rodent cell hybrids (cell lines harboring one human chromosome) was analyzed by colleagues in the laboratory. Interestingly, K111 insertions were detected in chromosomes 1, 4, 7, 9, 12, 13, 14, 15, 17, 18, 20, 21, 22, X and Y, but not in the other human chromosomes (verbal communication from Dr. Rafael A. Contreras-Galindo). K111 Solo LTRs were detected only in chromosomes 15 and 22. Further phylogenetic reconstruction and the nucleotide mutation sequences of K111 insertions in each chromosome demonstrated the existence of at least greater than 100 K111 proviruses (verbal communication from Dr. Rafael A. Contreras-Galindo). All K111 insertions detected in human chromosomes are flanked by CER:D22Z3 repeats and the same GAATTC target site duplication preceding the proviral LTRs.

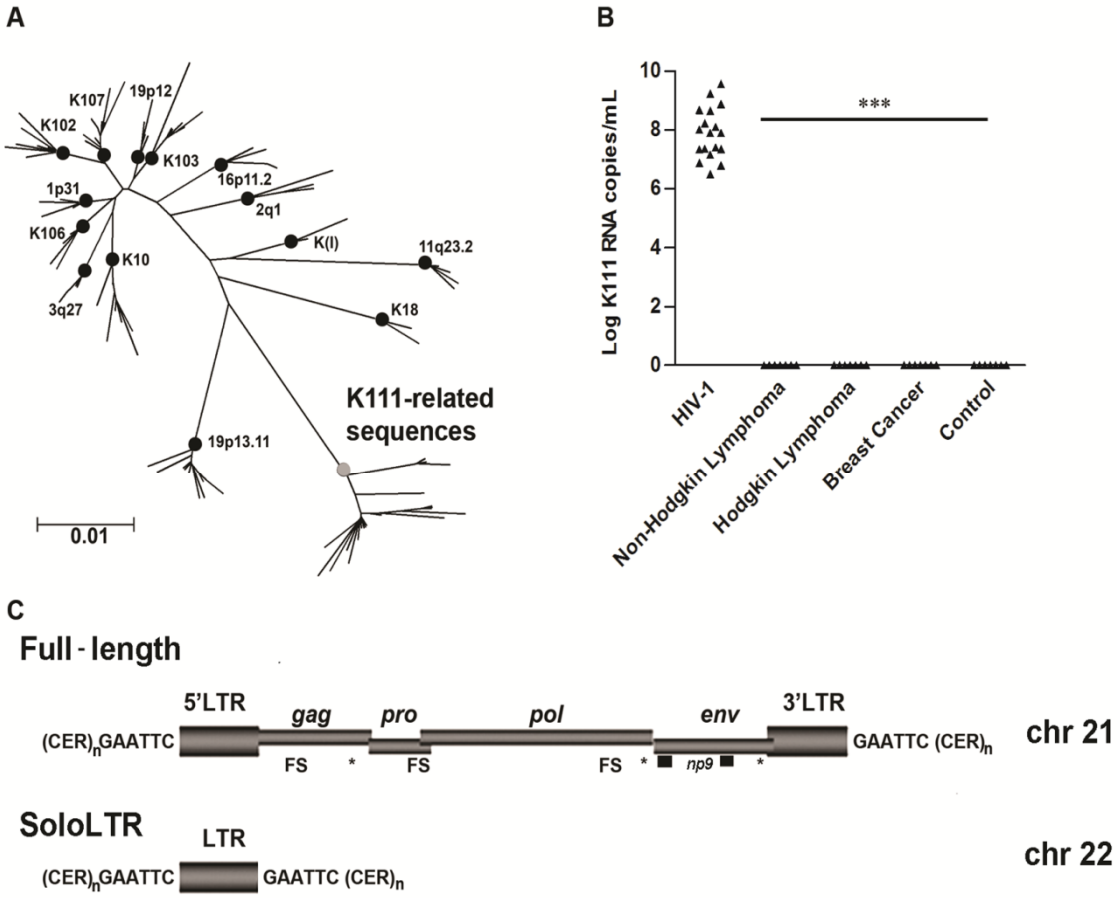


Figure 3.1 Identification and genomic organization of K111 viruses. (A) Phylogenetic reconstruction of HERV-K sequences found in the plasma of HIV-1-individuals reveals the existence of a unique branch of K111-related viral RNAs. (B) Quantitation of K111 *env* titers by qRT-PCR in the plasma of patients with HIV-1 and other diseases shows the expression of these viruses only in HIV infection. (C) Genomic organization of K111 full-length and soloLTR, target site duplication "GAATTC", and centromeric flanking sequences (CER:D22Z3). Frame shift (FS) and stop codon (asterisks) mutations are indicated. Chromosomes 21 and 22 are illustrated as an example. Data derived primarily by Dr. Rafael A. Contreras-Galindo and Dr. Mark H. Kaplan, in collaboration with Marta J. Gonzalez-Hernandez.

In vitro infection of cell lines with HIV-1 causes expression of K111

With the information that HIV-1-infected individuals had expression of a newly discovered HERV-K (HML-2) virus in their plasma, we set out to determine whether the activation of K111 expression in HIV-1 patients was the direct result of HIV-1 infection, or the result of an indirect pathology associated with the infection. This initial approach was chosen because, although we already knew that HIV-1 Tat was responsible for the transcriptional activation of HERV-K (HML-2) proviruses, K111 has the particular characteristic of being inserted in a heterochromatic area. Therefore, activation of NF-κB or NF-AT would most likely not be particularly successful in reaching their target binding sites in the K111 LTR. We had to be sure that HIV-1 infection *per* se, in the absence of any additional pathology or potential drug treatment interactions, was responsible for K111 activation. Using a qRT-PCR approach with a fluorescent probe specific for K111 it was indeed observed that infecting several human cell lines with HIV-1 induced the transcription of K111, otherwise silenced (or below limit of qPCR and regular PCR detection) in most uninfected cells (Fig. 3.2A). Similar results were obtained when quantifying expressed RNAs from primary cells isolated from healthy volunteer: PBLs that were infected with different HIV-1 strains (Fig. 3.2B). These results demonstrated that K111 RNA expression could be recapitulated *in vitro* though infection with HIV-1.

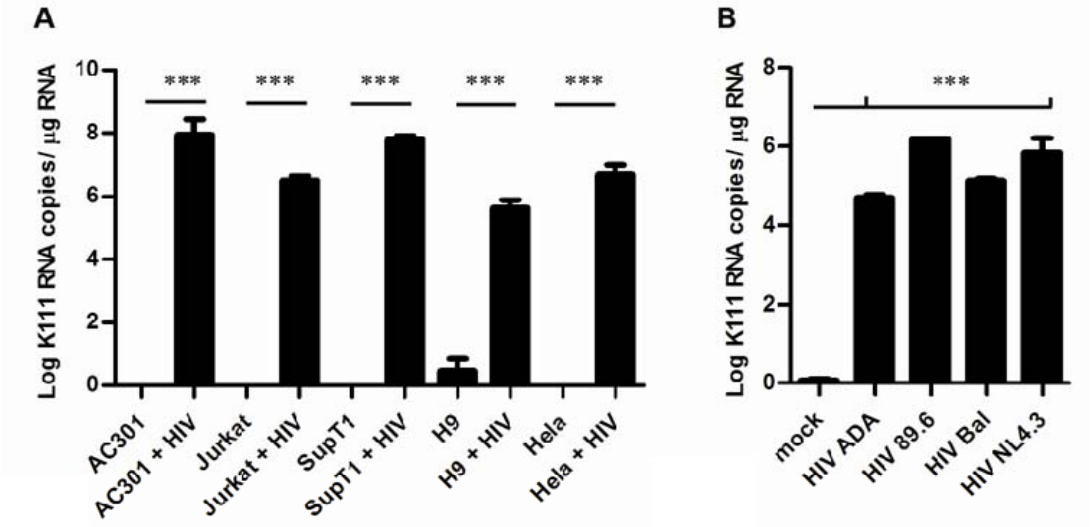


Figure 3.2 HIV-1 infection activates HERV-K (HML-2) K111. Total cellular RNA was isolated 7 days after infection, from cell lines that were infected with HIV-1_{NL4-3} (A), from PBLs that were infected with HIV-1_{ADA}, HIV-$1_{89.6}$, HIV-1_{BAL}, or HIV-$1_{NL4.3}$ (B), or from cells that were left uninfected. RNA was amplified using primers specific for K111 *env* through One Step qRT-PCR, and quantified using a standard curve generated by amplification of K111 RNA standards. Data are expressed as log viral K111 RNA copies per µg of input RNA. The mean number of K111 RNA between HIV-1 infection and controls was compared using an independent student's T test for samples exhibiting normally distributed values. Two-tailed *p* values were considered significant at $p < 0.05$ and are shown as *** = $p < 0.0001$. Data were obtained in collaboration with Dr. Rafael A. Contreras-Galindo.

HIV-1 Tat induces the expression of K111 from centromeric regions through chromatin remodeling

The centromere is the part of a chromosome that links sister chromatids together, and is important for correct segregation of chromosomes during cellular division [350, 351]. Regional centromeres contain large amounts of DNA and are often packaged into condensed areas of heterochromatin [352-354]. In most eukaryotes, the centromere has no defined DNA sequence, and it typically consists of large arrays of repetitive DNA (e.g. satellite DNA) where the sequence within individual repeat elements is similar but not identical [352-354]. In humans, the primary centromeric repeat unit is called α-satellite (or alphoid), although a number of other sequence types are found in this region [352-355], such as the CER elements where K111 was found to be inserted.

Since we had already observed that HIV-1 infection of cell lines and primary cells recapitulated the expression of K111 seen in infected individuals, we next planned to assess whether activation was also occurring as a result of cellular interactions with the HIV-1 Tat protein. We tested whether stimulation of human cell lines and primary cells with Tat was sufficient to induce K111 expression. Indeed, addition of recombinant HIV-1 Tat protein to Jurkat T cells and to PBLs, or overexpression of Tat in cell lines (Hltat, HelaTatIII), led to a marked increase in K111 transcripts when compared to untreated controls (Fig. 3.3). This was very interesting, and a little surprising, seeing as these K111 proviruses are inserted in a very condensed genomic area, where transcription factor binding would be impeded. However, it has been shown that the HIV-1 Tat protein can induce relief of chromatin repression prior to the onset of transcriptional initiation by its interaction with several

chromatin modifiers [356-358]. As expression of K111 proviruses is likely to be repressed at baseline due to the condensed structure of chromatin at the centromere, and as Tat is known to modulate chromatin dynamics, we tested whether expression of Tat leads to a more open chromatin structure over K111 proviruses. Chromatin immunoprecipitation (ChIP) assays using antibodies against the heterochromatin markers H3K9Me3 and H4K20Me3 showed a marked loss of heterochromatin over K111 in Tat-expressing cell lines as compared to their parental cell line (Fig. 3.4A). In contrast, Tat did not induce loss of the same heterochromatin markers over the RPL-30 and MYOD genes, which are regulated epigenetically under other circumstances (Fig. 3.4B). If anything, for these genes, the heterochromatin markers seem to increase in response to Tat. Therefore, Tat appears to activate K111 proviruses by inducing loss of heterochromatin in pericentromeric regions, opening up the chromatin structure and allowing for active transcription.

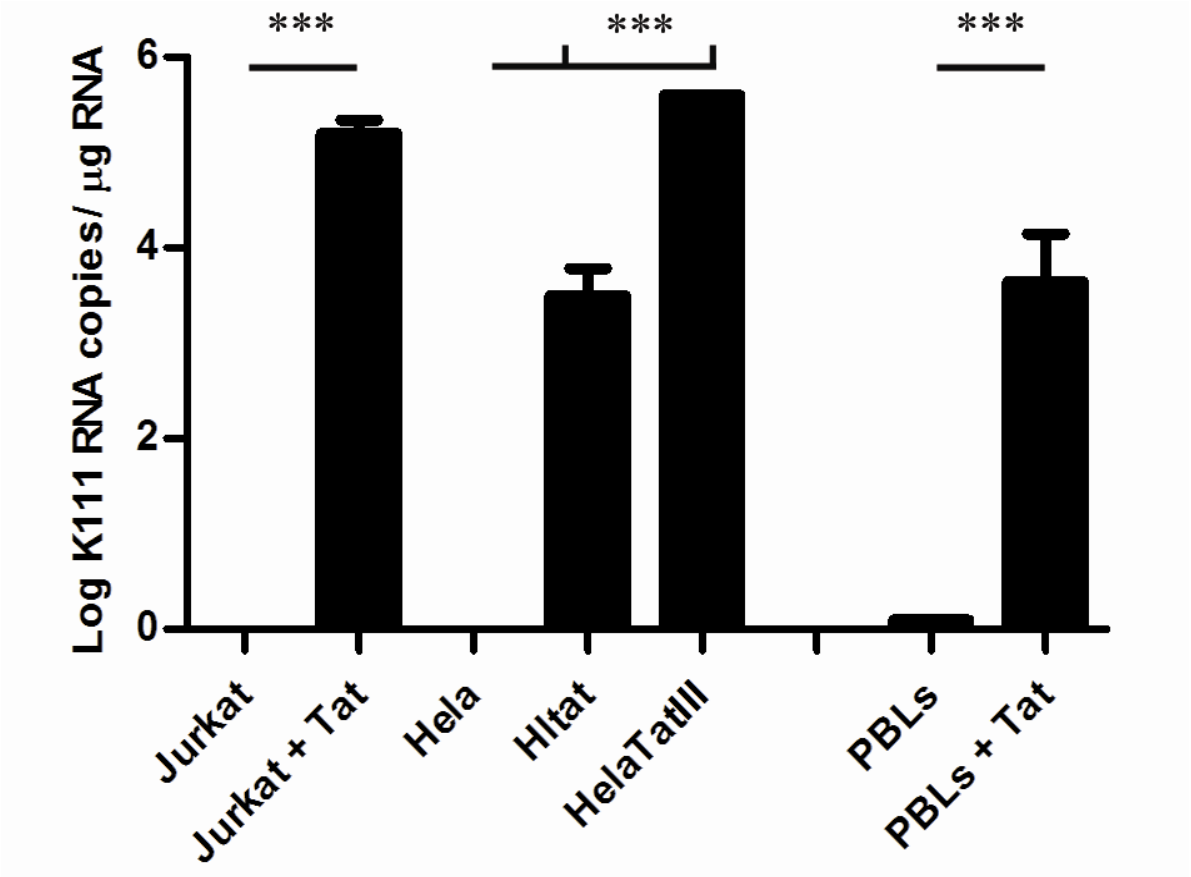


Figure 3.3. Treatment with HIV-1 Tat protein induces expression of centromeric K111. Cell lines or primary cells (PBLs) that were treated with recombinant Tat protein for 24 hours (A), or had stably-expressed Tat (B) had total RNA extracted, DNAse-treated, and amplified by One-Step qRT-PCR using a fluorescent probe specific for the *env* region of K111. Quantification was performed using a standard curve generated by amplification of K111 RNA standards. Data are expressed as log viral K111 RNA copies per µg of input RNA. The mean number of K111 RNA between tat-treatment and controls was compared using an independent student's T test for samples exhibiting normally distributed values. Two-tailed *p* values were considered significant at $p < 0.05$ and are shown as *** = $p < 0.0001$. Tat expressing cell lines: HItat, and HelaTatIII.

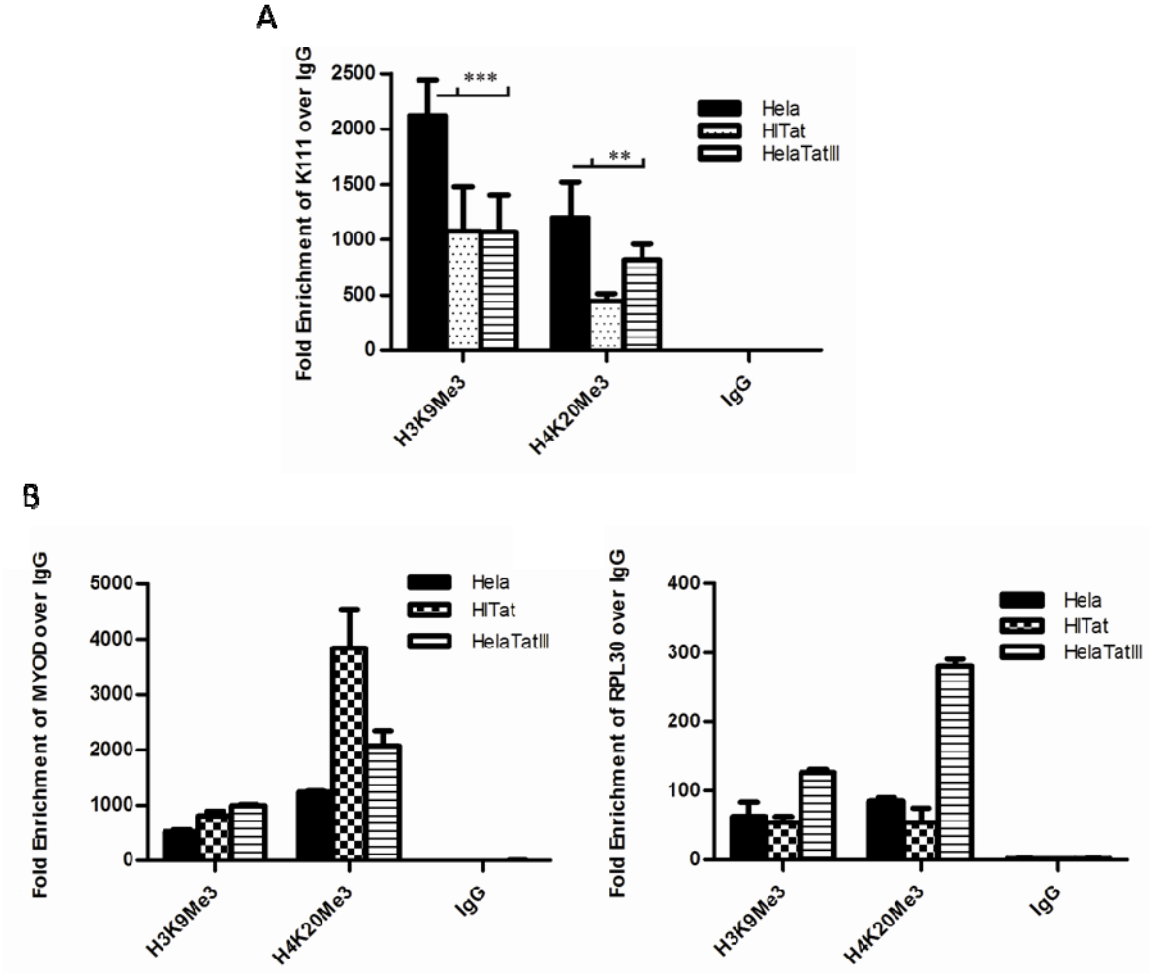


Figure 3.4 HIV-1 Tat protein activates HERV-K111 by inducing loss of heterochromatin. HeLa cell lines stably expressing Tat, as well as their parental cell line, were assayed in a chromatin immunoprecipitation (ChIP) against heterochromatin markers H3K9Me3 and H4K20Me3. K111-specific qPCR was performed on the precipitated chromatin and ChIP assay results show that HIV-1 Tat reduces the heterochromatic markers H3K9Me3 and H4K20Me3 at K111 proviral loci (A). Heterochromatin marks (H3K9Me3 or H4K20Me3) over other (non-HERV-K) genes using a promoter-specific PCR for genes with either constitutively silenced epigenetic marks (MYOD) or generally not under epigenetic control (RPL30) showed no significant decreases in heterochromatic markers even in the presence of Tat, but rather show an increase in them. Data obtained by Dr. Ferdinand Kappes in collaboration with Dr. Rafael A. Contreras-Galindo and Marta J. Gonzalez-Hernandez.

Discussion

This is the first sequence-based study to discover and show the spread of a single human endogenous retrovirus and its progeny into multiple human chromosomes, particularly in their centromeric regions, and suggests that centromeres from different chromosomes exchanged genetic material by homologous recombination during hominid evolution. Additionally, these newly discovered HML-2 proviruses are activated only during HIV-1 infection, and not in any other of the diseases tested (although we realize the limited scope of our analyses). Furthermore, activation of K111 proviruses during HIV-1 infection occurs, at least in part, through the interaction of the HIV-1 Tat protein with chromatin modifier proteins in the centromeric regions of human chromosomes. Relief of chromatin repression by Tat has previously been shown for transactivation at the HIV-1 LTR, and can first occur prior to the onset of transcriptional initiation [356-358]. For example, Tat has been extensively shown to interact with several histone acetyltransferases (HATs) in HIV-1 infected cells, which are involved in transcriptional activation: CBP/p300, p/CAF, GCN5, Tip60, and $TAF_{II}250$ [356]. Tat forms a complex with CBP/p300 and p/CAF proteins in HIV-1-infected cells, and this results in both the stimulation of Tat to activate transcription and in targeting the HATs to the viral promoter and the subsequent nucleosomes [359]. CBP and p300 are close homologs of each other and act as overall transcriptional coactivators; they function by interacting with promoter binding transcription factors such as CREB, as well as HATs such as p/CAF [356]. The presence of Tat induces the acetylation of histones H3 and H4 by CBP/p300 and p/CAF and activates transcription [356]. The presence of Tat also promotes a conformational change in CBP/p300 and significantly increases the HAT activity of p300 on histone H4 [360, 361]. All these interactions between Tat and chromatin remodelers assist in the formation

of a permissive environment that allows the binding of the basal transcription machinery, as well as transcription factors, in an area that is usually heavily silenced. In the absence of Tat, the LTR-bound nucleosomes are hypoacetylated (repressive), yet recruitment of Tat-associated HATs results in their hyperacetylation (activation) and subsequent LTR-directed gene transcription. This mechanism of chromatin remodeling could allow for subsequent demethylation of Histone 3 and Histone 4 in nucleosomes, which could also further allow transcriptional activation through translocation of activated NF-κB and NF-AT and their binding into the HERV-K (HML-2) K111 LTR, as well as any other HERV LTR that is similarly constricted or under strong epigenetic control. As HIV infection and Tat modulate the expression of the K111 viruses, the role they may play in the pathogenesis of HIV must be further explored.

Materials and Methods

Study Subjects

Plasma samples were obtained following protocols approved at the University of Michigan and the North Shore University Hospital. Plasma samples were collected from eighteen HIV-1-infected patients. The HIV-1 viral loads in these plasma samples ranged from < 50 to > 10^6 HIV-1 RNA copies/ml.

This study also included seven HIV-1 negative breast cancer patients, seven HIV-1 negative non-Hodgkin lymphoma patients, seven HIV-1 negative Hodgkin lymphoma patients, and seven control subject donors. Unlike healthy individuals, who typically have very little HERV-K (HML-2) RNA/DNA in their blood, these breast cancer, lymphoma, and HIV-1 patients had HERV-K (HML-2) viral loads within a range of 10^6 to 10^7 copies/mL (personal communication from Dr. Rafael Contreras-Galindo; [131, 190]).

PCR amplification of K111-related proviruses

A full-length proviral genome of K111 and partial genomes of K105 and K112 were amplified using the Expand Long Range dNTPack PCR kit (Roche Applied Science, Indianapolis, IN). PCR reactions contained 100 ng genomic DNA, 2.5 mM $MgCL_2$, 500 µM dNTPs, 300 nM of each primer, and 3.5 units of Expand Long Range Enzyme Mix. SoloLTRs were amplified using the forward P1 or P7 primers and the reverse P2 primer. The full-length forms of K111 were amplified using either P1 or P7 forward primers, the P2 reverse

primer and overlapping primers for the HERV-K *gag*, *pol,* and *env* genes. K112 was selectively amplified by using a primer (P5) that binds to the K112 LTR region which is otherwise mutated in K111 by a 14 bp insertion. K105 was amplified using a primer (P6) that binds a CER element flanking the K105 3'LTR. *In silico* distribution of CER elements flanking the K111 insertions were screened using Repeat Masking from the Genetic Information Research Institute. Only partial sequences of K105 (6361 bp) and K112 (2042 bp) proviruses were sequenced for the purpose of these studies. K111-related proviral insertions and soloLTRs were detected in the DNA of all twenty-eight HIV-1 patients, sixty-five cancer patients and ninety-six Caucasian control subjects. Sequences of proviruses K111, K112, K105 and soloLTR are deposited in the NCBI database with Accession Numbers (JQ790968 - JQ790992; GU476554 - GU476555).

Viral RNA and cellular RNA extraction, and RT-PCR of the HERV-K env gene

Plasma samples were treated with 200 U RNAse-free DNAse for 1 h at 37 °C. Viral RNA was extracted using the QIAmp viral RNA Mini Kit following the manufacturer's instructions (Qiagen, Valencia, CA). Cellular RNA was extracted with the QIAmp RNA blood mini-kit following the manufacturer's instructions (Qiagen, Valencia, CA) and treated with RNAse-free DNAse as recommend by the manufacturer. PCR was performed to verify the absence of genomic HERV-K DNA. The full-length HERV-K *env* gene was amplified using the One-Step RT-PCR kit (Qiagen, Valencia, CA) as described [131]. HERV-K (HML-2) *env* sequences found in the plasma of HIV-1, breast cancer, and lymphoma patients are

deposited in the NCBI database with the accession numbers DQ360503-DQ360809 and EU308642-EU308718.

qRT-PCR specific for K111

Titers of K111 cellular DNA, plasma RNA, supernatant RNA, and cellular RNA were measured by qPCR or qRT-PCR using a probe that specifically discriminates the K111 *env* gene from other HERV-K *env* sequences due to a 6 bp mutation (underlined in the probe sequence K111P shown below). The qPCR was performed as described [131, 190], using the primers K111F and K111R and the FAM-labeled probe K111P. A reverse transcription step of 30 min at 50 °C was included in RT-PCR reactions. The PCR was carried out in 35 cycles consisting of 15 sec of denaturation at 95 °C and 1 min of annealing/hybridization at 60 °C. The K111 titers were estimated using serial dilutions of K111 *env* DNA cloned in the topo TA vector (Invitrogen, Carlsbad, CA). The specificity of the probe was assessed using viral RNA from plasma samples of HIV-1 patients in whom Type-1 HERV-K and K111 *env* were amplified and sequenced and breast cancer patients in whom similar Type-1 HERV-K genotypes, except K111, were detected. In spite of the fact that the primers used in the reaction amplify all Type-1 HERV-K present in HIV-1 and cancer patients, as seen when the PCR products are electrophoresed in agarose gels and stained by ethidium bromide, the K111 FAM-labeled probe fluoresces only in Real-Time reactions carried out using either K111 *env* standards or viral RNA from HIV-1-infected individuals, but not from cancer patients or control subjects.

HIV-1 infection/Tat stimulation of Peripheral Blood Lymphocytes (PBLs) and cell lines

Whole blood was drawn from healthy volunteers and the PBLs were isolated by Ficoll-Hypaque gradient, separated from monocytes by differential plate adhesion, and grown in RPMI medium supplemented with 10% FBS. Ten million PBLs were infected for one day with 1 mL of ~10^3 $TCID_{50}$/ml of the HIV-1 strains ADA, 89.6, BAL, or NL4.3, or, alternatively, stimulated with Tat protein containing the first 86 amino acids (Tat 86: 50 ng/mL) for 24 h. AC301, Jurkat, SupT1, and H9 T-cell lines were infected with HIV-1 NL4.3. Stably HIV-1-infected HeLaLAV cells and non-HIV-1 infected HeLa cells were also used in the study. After infection, cells were washed twice and the viruses were propagated for 7 days at 37 °C. Cellular RNA was extracted from unstimulated or Tat-stimulated PBLs, carcinoma cell lines (PA-1, NCCIT, 293T, U937, Jurkat and HeLa), Tat-expressing cell lines (Jurkat Tat, Hltat, and HeLaTatIII), and HIV-1 infected and uninfected cells. DNA was removed from the RNA samples by DNAse treatment and the absence of DNA was confirmed by PCR. Levels of K111 expression were measured by Real-Time RT-PCR as described above. HIV isolates, Tat 86 protein, and Tat-expressing cell lines were obtained from the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH (Germantown, MD).

Chromatin Immunoprecipitation (ChIP)

ChIP was performed according to standard protocols [362]. The antibodies used here were specific to the heterochromatic marks H3K9Me3 and H4K20Me3 (Abcam, Cambridge, MA). Briefly, formaldehyde was added to the cells to a final concentration of

1% for 10 min in a shaking platform. The cross-linking was stopped by 1/20V of 2.5 M Glycine. The cells were washed with PBS, harvested in the presence of proteinase inhibitors, and resuspended in lysis buffer for 10 min. The samples were pelleted, re-suspended in nuclear lysis buffer, and the chromatin sonicated 12 times at power setting 3.5 for 12 sec with a 2 min cool-down between each sonication to produce average fragments between 200 and 500 bp. The chromatin was pre-cleared using a Salmon sperm DNA/Protein A Agarose-50% slurry and incubated with either H3K9Me3, H4K20Me3, or non-specific IgG antibodies overnight. Antibody-bound chromatin was purified using protein A/agarose, washed extensively, and reverse-cross-linked. Immunoprecipitated DNA and input DNA (5%) were purified by treatment with RNAse A, and proteinase K, and cleaned using the QiAquick PCR purification kit (Qiagen, Valencia, CA).

Heterochromatin occupancy on target K111 was measured by qPCR analysis in HeLa Tat-expressing cells and the parental HeLa cell lines. The fold enrichment (ChIP/input) was determined based on the cycle differences (ΔCt) of the 2 samples. RPL30, an active gene and MYOD, a silenced gene, serve as specificity controls. Primer sequences used for ChIP are listed below.

In silico sequence analysis

The K111-related LTR sequences obtained in the DNA of human cells, DNA from human/rodent chromosomal cell hybrids, as well as the K111-related *env* sequences obtained in the plasma of HIV-1 patients, were BLASTed to the NCBI database. The

sequences were aligned in BioEdit and exported to the MEGA 5 matrix. Phylogenetic trees of the *env* were constructed and corroborated by different methods (neighbor-joining, maximum parsimony, and maximum likelihood), using the statistical bootstrap test (10000 replicates) of inferred phylogeny and the Kimura-2 parameter model [363, 364].

Statistical Analysis

The mean number of K111 RNA or DNA between HIV-1 or Tat treatments and controls were compared using an independent student's T test for samples exhibiting normally distributed values. Two-tailed p values were considered significant at $p < 0.05$.

Data access

All the sequences reported in this manuscript have been deposited in the NCBI database (http://www.ncbi.nlm.nih.gov) under accession numbers DQ360503-DQ360809, EU308642-EU308718, GU476554-GU476555, and JQ790968-JQ790992.

List of Primers

P1: 5'-ACA TCC AGA CCA TGG TAG CCG TGT -3'

P2: 5'-ACA GTG CTG TGT GGG TCT GAA TGA -3'

P4: 5'-GTA CCT TCA CCC TAG AGA AAA GCC T -3'

P5: 5'- TCC AAG GTT TCT GGG AAG GGA AAG A -3'

P6: 5'-ATG CGA TCC AAG AAG ACT GCT GCT-3'

P7: 5’-TCG TTG CAA TGC TCT GGA ATT C-3’

K111F: 5’-AAG AGC ACC AGG ATG CTT AAT GCC-3’

K111R: 5’-AGT GAC ATC CGG CTT ACC ATG TGA-3’

K111P: 5’-FAM-TGC CGG TCC TAA CAG TAG ACT CAC-BHQ1-3’

RPL30 pF1: 5’-CAA GGC AAA GCG AAA TTG GT-3’

RPL30 pR1: 5’-GCC CGT TCA GTC TCT TCG ATT-3’

MYODpF1:5’-CCG CCT GAG CAA AGT AAA TGA-3’

MYODpR1: 5’-GGC AAC CGC TGG TTT GG-3’

CHAPTER IV

THE HERV-K (HML-2) TRANSCRIPTOME INDUCED BY THE HIV-1 TAT PROTEIN

Abstract

As the HIV-1 Tat protein activates expression of HERV-K (HML-2), we set out to determine which of the 91 specific HERV-K (HML-2) proviruses change expression in response to Tat. Next generation sequencing technology applied to total RNA (RNA-Seq) isolated from Tat and vehicle-treated peripheral blood lymphocytes from a healthy donor showed that, out of the 91 annotated HERV-K (HML-2) proviruses, Tat activated the expression of 24 unique viruses, while silencing 28. qRT-PCR validation of the RNA-Seq data performed on Tat-treated PBLs from three additional donors, using viral specific primers, corroborated the results. Most activated and silenced proviruses are Type 2 and do not code for protein, but those few activated Type 1 proviruses do potentially code for the Np9 putative oncoprotein. These data corroborate the dual effect Tat has been shown to have on some cellular genes, for which it behaves as both an activator and a suppressor. Additionally, these findings further demonstrate the complexity of the genome-wide regulation of HERV-K (HML-2) expression.

Introduction

Advancements in modern technology have now allowed us to determine essentially all the genes expressed in a given cell at a given time and/or under specific conditions. The collection of all these expressed genes is often called the cellular transcriptome, and it consists of the set of all RNA molecules (including mRNA (messenger RNA), rRNA (ribosomal RNA), tRNA (transfer RNA), and ncRNAs (non-coding RNAs)) that are produced in a cell or a population of cells. It represents the very small percentage of the genome that is transcribed (less than 5 percent in humans), and which can vary with different external environmental conditions. However, although only a small part of the genome is transcribed, the transcriptome is substantially more complex than the genome that encodes it, since a gene may produce many different types of mRNA molecules [365]. This complexity can now be studied using high-throughput techniques [365] and is called *expression profiling* or RNA-sequencing (RNA-Seq) [365, 366]. Expression profiling, thus, has several different aims that are summarized as follows: cataloguing all species of a transcript (coding, non-coding and small); determining the transcriptional structure of genes, splicing patterns and other post-transcriptional modifications; and quantifying the changing expression levels of each transcript during development and under different conditions.

In order to fully understand the transcriptome, expression profiling uses next-generation sequencing (NGS) technology to analyze information at the nucleotide level with high sensitivity and accuracy [365-369]. NGS technology moves away from the Sanger-DNA sequencing methods to implement various strategies relying on a combination of template preparation, sequencing and imaging, and genome alignment and assembly. These things

make obtaining large amounts of data easier, faster, lower-cost and compatible with high-throughput analysis [365-368]. A typical RNA-Seq experiment has enormous sequencing depth, in the vicinity of 100-1,000 reads per base pair of a transcript, and this allows for an almost complete snapshot of the transcriptome, including rare and low-abundance transcripts [365-368, 370]. Interestingly, RNA-Seq can allow for base-pair level resolution in contrast to alternative high-throughput technologies like microarrays, which rely on prior information that cannot detect novel splice variants, novel genes, and novel transcripts [365-370]. RNA-Seq also has low background noise and high sensitivity, requires less RNA, and is becoming more cost-effective with rapid advances in the technology.

For an RNA-Seq run, total RNA or mRNA is isolated from cells and cDNA is synthesized for a gene library preparation. When analyzing gene expression, the 3' poly-adenylated (poly-A) tail of mRNA is targeted as a purification step or as a cDNA synthesis step, in order to ensure that coding RNA is separated from noncoding RNA (ncRNA). In studies focusing on the portions of the transcriptome that include ncRNAs, several procedures for isolation and amplification can be performed. For example, high-throughput cross-linking immunoprecipitation (HITS-CLIP) is a recently developed method that enables identification of direct target sequences through the sequencing of RNAs from immuno-precipitated cross-linked Argonaute–miRNA–mRNA complexes [371]. Random hexamer primers can also be used for cDNA synthesis from total RNA. Alternatively, RNAs isolated from the flow-through RNA (non-poly-A RNA, from mRNA purification) and random-primed can be used for cDNA synthesis and analysis. Once the cDNA is synthesized, it can then be used for library preparation by being further fragmented to reach the desired length needed for the different NGS platforms currently used for RNA-Seq experiments (e.g. Illumina,

SOLiD, 454). Libraries are made using adaptors, tags or barcodes to further optimize efficiencies and cost savings in transcriptome sequencing. It is important to note that the starting RNA material must be of very high quality in order to obtain similarly high quality reads. This will then make the sequence data analysis and alignment procedures more efficient and more likely to generate meaningful information. With high-quality reads obtained, the task of analyzing the data is then undertaken. RNA-Seq faces several informatics challenges, particularly because of the incredibly large amount of data that can be obtained (on the order of billions of reads), so methods for storing, retrieving and processing such large amounts of data are always being improved in order to minimize errors in image analysis, base-calling, and removal of low-quality reads [365-368, 370]. Additionally, despite the advantages of NGS sequencing technologies, the sequence reads obtained from the common NGS platforms are often very short (35–500 bp) [367, 369]. As a result, it is necessary to reconstruct the full length transcripts by transcriptome assembly, except of course in the case of small classes of RNA, such as microRNAs, PIWI-interacting RNAs (piRNAs), small nucleolar (snRNAs), and small interfering (siRNAs) [366]. This is called annotation. There are two general methods of creating transcriptome annotations: 1. Mapping sequence reads onto a reference genome, and 2. *De novo* transcriptome assembly, which utilizes algorithms built into assembly software to generate transcripts from short sequence reads [366].

In the case of most human gene expression, annotation is straightforward, as most genes can be mapped back to the human genome reference database. However, for

repetitive genetic elements such as HERV sequences, this is not a trivial and straightforward process. Most repetitive sequences from the human genome are not fully annotated, since they are catalogued as “junk DNA”. Therefore, manual labeling and annotation from any sequences that have been archived from the human genome assembly or from BAC clone databases have to be performed prior to any analyses, and is a slow and laborious process. The process is also difficult, as it is necessary to determine differences in each sequence, for sequences that are highly similar, with a relative degree of certainty. Methods based on expressed sequence tags (ESTs) have provided a more comprehensive view of HERV transcriptomes, interestingly, but generally run into trouble when identifying the unique genomic source of expression [63, 372]. Analyzing repetitive element and HERV transcriptome expression, therefore, has been slow and difficult.

Profiling the expression of endogenous retroviral genes in both healthy and diseased tissues has been a widespread effort over the last 10 years, but has proven quite hard as most of the attempts to measure HERV expression use RT-PCR techniques [373]. Reliable PCR systems are difficult to obtain for discriminating between individual HERV elements as these type of genomic sequences have marked phylogenetic similarities [373, 374]. Therefore, most of the profiling attempts have been done either focusing on a specific locus [102, 146, 193, 213, 373-375], or on evaluation of general trends within HERV families or genera [60, 373, 374, 376, 377]. Phylogenetic similarities and the profusion of integration sites, two inherent characteristics of transposable elements and HERVs, make it difficult to study individual locus expression in a large-scale approach [373]. However, new advances in sequencing technology are allowing for better differentiation of retroelement and HERV

expression, down to mapping specific activated proviruses. For example, a NGS technique comprising hemi-specific PCR coupled to Illumina platform sequencing has achieved high coverage of germline polymorphic human-specific L1 (L1Hs) retrotransposition events [378]. A microarray dedicated to a collection of 5,573 HERVs was recently developed using 25-mer probes that minimize cross-hybridization [373], and has shown that almost one third of the HERV repertoire in tissues is indeed transcribed. The data from these studies have shown that the HERV transcriptome follows tropism rules, is sensitive to the state of differentiation, is sensitive to treatment and, unexpectedly, seems not to correlate with the age of the HERV families [373]. These new advances demonstrate that, while still difficult, HERV transcriptome analyses can be performed with a mix of NGS technologies and PCR validations.

Knowing that HIV-1 Tat activates expression of HERV-K (HML-2) transcripts, we were interested in learning exactly which of the 91 HERV-K (HML-2) proviruses are activated by Tat. Taking advantage of newly acquired NGS technologies at the University of Michigan, and the collaboration with Bioinformatics experts in the field of RNA-Seq, we set out to determine the HERV-K (HML-2) transcriptome in response to HIV-1 Tat. Additionally, we analyzed which of the two types of HERV-K (HML-2) were most affected by Tat, if any proviruses had ORFs, and if the same viruses targeted by Tat were also ones that are

expressed in patients during HIV-1 infection. This Tat/HERV-K (HML-2) transcriptome gave us insight into the biology of the activation of HERV-K (HML-2) by HIV-1.

Results

HIV-1 Tat both activates and silences specific HERV-K (HML-2) proviruses in normal PBLs

In order to observe the HERV-K (HML-2) transcriptome in response to HIV-1 Tat, normal PBLs isolated from the blood of a healthy volunteer were kept in culture for three days under PHA and IL-2 stimulation, and then treated with recombinant Tat protein or vehicle buffer for 12 hours, as previously shown. Analysis of total RNA from each sample run as 100 bp x 100 bp paired-end tagged sequences generated around 104 million reads (passing filter) with over 80% above Q30 (quality score of 30, measuring the probability that 1 base in 1000 will be assigned incorrectly, with a corresponding base call accuracy of 99.9%). The data were then Tophat mapped to a reference HERV-K (HML-2) sequence database compiled by us and collaborators from the Bioinformatics Core and the Department of Computational Medicine and Bioinformatics at the University of Michigan. The locations of the currently known 91 HERV-K (HML-2) genomes were organized in a BED formatted file and BEDTools [379] was used to intersect the HERV-K (HML-2) locations in the sequenced samples that had acceptable hits in order to identify all the sequence reads that aligned to the 91 known HERV-K (HML-2). All hits were deduplicated to remove PCR artifacts, subjected to Cufflinks2 analysis, and reads are shown summarized in Table 4.1. This table denotes the specific HERV-K (HML-2) proviruses, with any alternate names next to their chromosomal locations, detected in each treatment, after mapping to our

reference HERV-K (HML-2) genome. Cufflinks2 analysis assembled transcripts, estimated their abundance, and verified differential regulation and/or expression in all RNA-Seq

samples tested. After alignment of RNA-Seq reads is performed, Cufflinks2 accepts the aligned reads and assembles them into a set of transcripts, estimating the relative abundance of each based on how many reads support each one. Each "Total Reads" column shows, for each specific treatment, the number of reads that mapped to a particular HERV-K (HML-2) provirus at a particular chromosomal location, out of all the read hits that that treatment obtained (e.g. total RNA hits for a specific HERV-K (HML-2) provirus expressed out of all the other expressed cellular RNAs that the treatment induced). Each "Unique Reads" column (blue) represents the reads that, out of the "Total Reads" obtained by that treatment, are most likely uniquely mapped to that particular provirus for that particular treatment, and are not shared with any other provirus. This is important to note for two reasons: First, our RNA-Seq analysis was performed using the Paired-End tag (PET) method. Paired-End tags were synthesized during library preparation in order to better map the sequences obtained to the reference genome we designed. The process involves the coming together of two short stretches of sequence: one at the 5' and one at the 3' end of the cDNA fragment of interest. These short paired-ends allow us to map both the beginning part of the sequence of interest, and the end, to the reference genome since they contain enough information and represent a whole DNA fragment of interest. Lastly, since all HERV-K (HML-2) proviruses are highly similar and, with our read length being so small in size (approximately 200 bp long), it is quite possible that more than one virus shares the same exact sequence read at either the same position or different

positions in their genome with another family member. Therefore, filtering out the unique provirus reads from the total reads is important in order to make sure that the reads obtained correspond to just the provirus of interest at the chromosomal location of interest.

As can be seen in Table 4.1, most of these proviruses are already expressed at a basal level (vehicle treatment columns), with modifications in expression after HIV-1 Tat is added to the cells. However, whether or not Tat was added to the sample, numbers were never higher than 1700 total read counts per lane. When compared to approximately 104 million reads obtained for each treatment, these data seem remarkably low. Therefore, though these proviruses are indeed expressed and are affected by Tat treatment, their abundance in comparison to other cellular RNAs in PBLs is low. Interestingly, according to this one biological replicate's RNA-Seq data set, 73 of the 91 HERV-K (HML-2) proviruses had detectable read hits after Tat treatment, but only 42 of those had more than 3 unique read hits each (anything lower than 4 read hits probably means it is not truly a specifically expressed provirus). Most of the reads mapped to different parts of chromosomes 1, 3, 4, 7, 9, 11 and 19. Interestingly, 79 of the 91 known proviruses had detectable read hits in the vehicle control, with 49 having more than 3 unique reads. While increased expression of 42 unique read-mapping proviruses was seen in the Tat treatment group, expression of an additional 7 viruses was seen in the group treated with vehicle alone as compared to the Tat treatment group: 1q32.2, 2q21.1, 4q32.3 (K5/ERV-12), 8q11.1 (K70/K43), 12p11.1 (K50e), 12q14.1 (K(C12)/K119) and 21q21.1 (K60/ERV-K23). These results suggest that Tat treatment silences, as well as activates, HERV-K (HML-2) expression. This was quite

surprising, as we had previously assumed that all HERV-K (HML-2) proviruses would be activated by Tat, with differences only in the degree of the activation (Chapters II and III). Upregulation and downregulation of specific proviruses is detailed in Table 4.2, which shows the relative fold changes seen when comparing the expression of all the viruses that had unique reads in the Tat-treated cells versus those that had unique reads in the vehicle-treated cells. In this table, the resulting fold expression is written as the result of the mathematical formula "# reads in Tat/# reads in vehicle"; the fold decrease is also written in parentheses after calculating the inverse of the "# reads in Tat/# reads in vehicle" result. Compared to the vehicle treatment, 28 viruses had higher expression in the presence of Tat (bold green-colored numbers), whereas 24 viruses were most highly expressed in the control (red-colored numbers). These 24 viruses include the seven viruses that had no hits after Tat-treatment, and 3 additional viruses that had 3 hits or less after Tat-treatment (meaning that they are probably not truly expressed). In view of the surprising findings concerning downregulation of some HERV-K (HML-2) proviruses by Tat, and since the RNA-Seq data were obtained from the PBLs of only one donor, I proceeded to validate the results via qRT-PCR.

Table 4.1 Specific HERV-K (HML-2) proviruses expressed in PBLs with Tat or vehicle treatment

Chrom	Description	Tat-Total Reads	Tat-Unique Reads	Vehicle-Total Reads	Vehicle-Unique Reads
1	1p36.21a	2	0	0	0
1	1p36.21b K(OLDAL023753),K6,K76	11	0	35	0
1	1p36.21c K6,K76	10	0	35	0
1	1p34.3	0	0	0	0
1	1p31.1 K4, K116, ERVK-1	608	91	307	65
1	1q21.3	554	547	182	181
1	1q22 K102, K(C1b),K50a, ERVK-7	1710	429	874	180
1	1q23.3 K110, K18,K(C1a), ERVK-18	327	243	176	136
1	1q24.1 K12	3	2	13	0
1	1q32.2	17	0	17	15
1	1q43	0	0	0	0
2	2q21.1	5	0	20	11
3	3p25.3 K11, ERVK-2	5	5	15	13
3	3p12.3	66	66	1	1
3	3q12.3 K(II), ERVK-5	1234	988	685	449
3	3q13.2 K106, K(C3),K68, ERVK-3	325	3	156	3
3	3q21.2 K(I), ERVK-4	137	45	44	36
3	3q24 ERVK-13	166	0	88	0
3	3q27.2 K50b, K117,ERVK-11	342	23	159	21
4	4p16.3a	3	3	12	12
4	4p16.3b K77	34	28	47	44
4	4p16.1a K17b	39	36	56	56
4	4p16.1b K50c	319	280	69	62
4	4q13.2	0	0	0	0
4	4q32.1	50	25	12	0
4	4q32.3 K5, ERVK-12	2	0	10	6
4	4q35.2	33	31	31	31
5	5p13.3 K104, K50d	157	2	54	1
5	5p12	1	1	29	29
5	5q33.2 K18b	4	4	17	17
5	5q33.3 K107/K10,K(C5), ERVK-10	938	33	480	20
6	6p22.1 K(OLDAL121932),K69,K20	0	0	25	0
6	6p21.1 K(OLDAL035587),KOLD35587	22	22	21	21
6	6p11.2 K23	1	1	31	31
6	6q14.1 K109, K(C6),ERVK-9	642	0	353	2
6	6q25.1	3	3	17	17
7	7p22.1a K108L, K(HML.2-HOM),K(C7), ERVK-6	737	0	371	2
7	7p22.1b K108R, ERVK-6	762	29	413	14
7	7q11.21	0	0	0	0

Table 4.1 Continuation

Chrom	Description	Tat-Total Reads	Tat-Unique Reads	Vehicle-Total Reads	Vehicle-Unique Reads
7	7q22.2 ERVK-14	105	36	26	0
7	7q34 K(OLDAC004979),ERVK-15	510	509	136	136
8	8p23.1a K115, ERVK-8	588	58	192	25
8	8p23.1b K27	0	0	3	0
8	8p23.1c	273	33	112	4
8	8p23.1d KOLD130352	238	37	98	2
8	8p22	9	9	21	21
8	8q11.1 K70, K43	0	0	30	30
8	8q24.3a	68	0	58	0
8	8q24.3b K29	73	73	9	9
9	9q34.11 K31	166	166	85	85
9	9q34.3 K30	3	3	14	14
10	10p14 K(C11a), K33,ERVK-16	10	4	49	44
10	10p12.1 K103, K(C10)	1	1	0	0
10	10q24.2 ERVK-17, c10_B	100	3	30	2
11	11p15.4 K7	239	198	64	55
11	11q12.1	0	0	2	2
11	11q12.3 K(OLDAC004127)	143	130	38	28
11	11q22.1 K(C11c), K36,K118, ERVK-25	552	1	300	5
11	11q23.3 K(C11b), K37,ERVK-20	66	5	54	26
12	12p11.1 K50e	0	0	11	11
12	12q13.2	3	0	2	0
12	12q14.1 K(C12), K41,K119, ERVK-21	282	0	130	8
12	12q24.11	15	1	9	0
12	12q24.33 K42	177	12	103	2
14	14q11.2 K(OLDAL136419),K71	122	122	50	50
14	14q32.33	0	0	0	0
15	**15q25.2**	0	0	0	0
16	16p13.3 K(OLDAC004034	28	28	3	3
16	16p11.2	40	4	18	0
17	17p13.1	4	4	29	29
19	19p13.3 ERVK-22	15	0	9	1
19	19p12a K52	2	2	19	19
19	19p12b K113	0	0	0	0
19	19p12c K51	8	3	3	0
19	19q11 K(C19), ERVK-19	444	33	113	16
19	19q13.12a	32	30	2	0
19	19q13.12b K(OLDAC012309),KOLD12309	533	533	355	349
19	19q13.41	32	32	18	18
19	19q13.42 LTR13	5	5	20	20
20	20q11.22 K(OLDAL136419),K59	23	22	17	17
21	21q21.1 K60, ERVK-23	61	0	18	16
22	22q11.21 K101, K(C22),ERVK-24	569	28	369	2
22	22q11.23 K(OLDAP000345),KOLD345	4	2	54	31

Table 4.1 Continuation

Chrom	Description	Tat-Total Reads	Tat-Unique Reads	Vehicle-Total Reads	Vehicle-Unique Reads
X	Xq11.1	2	2	18	18
X	Xq12	1	0	5	0
X	Xq28a K63	0	0	24	0
X	Xq28b K63	0	0	25	1
Y	Yp11.2	0	0	0	0
Y	Yq11.23a	0	0	0	0
Y	Yq11.23b	0	0	0	0

*Tophat and Cufflinks2 analyses performed by Dr. James Cavalcoli, Bioinformatics.

*The different names after each chromosomal location are all alternative ways to refer to the same virus.

Table 4.2 Changes in fold expression of specific HERV-K (HML-2) proviruses from unique reads in Tat vs vehicle treatments

Chrom	Description	Tat-Unique Reads	Vehicle-Unique Reads	Fold (Tat over Veh)
1	1p31.1 K4, K116, ERVK-1	91	65	1.40
1	1q21.3	547	181	3.02
1	1q22 K102, K(C1b),K50a, ERVK-7	429	180	2.38
1	1q23.3 K110, K18,K(C1a), ERVK-18	243	136	1.79
1	1q32.2	0	15	0.00 (15 fold down)
2	2q21.1	0	11	0.00 (11 fold down)
3	3p25.3 K11, ERVK-2	5	13	0.38 (2.6 fold down)
3	3p12.3	66	1	66.00
3	3q12.3 K(II), ERVK-5	988	449	2.20
3	3q21.2 K(I), ERVK-4	45	36	1.25
3	3q27.2 K50b, K117,ERVK-11	23	21	1.10
4	4p16.3a	3	12	0.25 (4 fold down)
4	4p16.3b K77	28	44	0.64 (1.6 fold down)
4	4p16.1a K17b	36	56	0.64 (1.6 fold down)
4	4p16.1b K50c	280	62	4.52
4	4q32.1	25	0	25.00
4	4q32.3 K5, ERVK-12	0	6	0.00 (6 fold down)
4	4q35.2	31	31	1.00
5	5p12	1	29	0.03 (29 fold down)
5	5q33.2 K18b	4	17	0.24 (4.25 fold down)
5	5q33.3 K107/K10,K(C5), ERVK-10	33	20	1.65
6	6p21.1 K(OLDAL035587),KOLD35587	22	21	1.05
6	6p11.2 K23	1	31	0.031 (31 fold down)
6	6q25.1	3	17	0.18 (5.66 fold down)
7	7p22.1b K108R, ERVK-6	29	14	2.07
7	7q22.2 ERVK-14	36	0	36.00
7	7q34 K(OLDAC004979),ERVK-15	509	136	3.74
8	8p23.1a K115, ERVK-8	58	25	2.32
8	8p23.1c	33	4	8.25
8	8p23.1d KOLD130352	37	2	18.50
8	8p22	9	21	0.43 (2.33 fold down)
8	8q11.1 K70, K43	0	30	0.00 (30 fold down)
8	8q24.3b K29	73	9	8.11
9	9q34.11 K31	166	85	1.95
9	9q34.3 K30	3	14	0.21 (4.66 fold down)
10	10p14 K(C11a), K33,ERVK-16	4	44	0.09 (11 fold down)
11	11p15.4 K7	198	55	3.60
11	11q12.3 K(OLDAC004127)	130	28	4.64
11	11q22.1 K(C11c), K36,K118, ERVK-25	1	5	0.20 (5 fold down)
11	11q23.3 K(C11b), K37,ERVK-20	5	26	0.19 (5.2 fold down)
12	12p11.1 K50e	0	11	0.00 (11 fold down)
12	12q14.1 K(C12), K41,K119, ERVK-21	0	8	0.00 (8 fold down)
12	12q24.33 K42	12	2	6.00

Table 4.2 Continuation

Chrom	Description	Tat-Unique Reads	Vehicle-Unique Reads	Fold increase/decrease
14	14q11.2 K(OLDAL136419),K71	122	50	**2.44**
16	16p13.3 K(OLDAC004034	28	3	**9.33**
16	16p11.2	4	0	**4**
19	19p12a K52	2	19	.105 (9.5 fold down)
19	19p12c K51	3	0	**3**
19	19q13.12a	30	0	**30**
19	19q13.42 LTR13	5	20	0.25 (4 fold down)
20	20q11.22 K(OLDAL136419),K59	22	17	**1.29**
21	21q21.1 K60, ERVK-23	0	16	0.00 (16 fold down)
22	22q11.21 K101, K(C22),ERVK-24	28	2	**14**
22	22q11.23 K(OLDAP000345),KOLD345	2	31	0.065 (15.5 fold down)
X	Xq11.1	2	18	0.111 (9 fold down)

*The different names after each chromosomal location are all alternative ways to refer to the same virus.

qPCR validation of specific HERV-K (HML-2) provirus activation and silencing by HIV-1 Tat corroborates RNA-Seq data analysis

For validation of the RNA-Seq data, we next aligned all HERV-K (HML-2) proviruses that had shown differences between Tat and vehicle treatments (whether it was up or downregulation), and looked for regions that had enough sequence diversity between them so as to be able to be distinguished specifically through qPCR. Twenty-three proviruses fit this description, but only 9 (including K111, whose validation was shown in Chapter III) had specific sequences that could be used for qPCR primer design and could assure their specific detection through qPCR amplification: 3p12.3, 4q32.1, 5p12, 6p11.2 (K23), 7q34 (KOLD/ERVK-15), 8q11.1 (K70/K43), 11q12.3 (KOLD), 16p13.3 (KOLD) and K111.

Among these proviruses, some were upregulated by Tat (3p12.3, 4q32.1, 7q34, 11q12.3 and 16p13.3), while others were downregulated or appeared to be silenced (5p12, 6p11.2, 8q11.1).

Amplification products of the expressed proviruses (Figure 4.1) from the RNA used for RNA-Seq, as well as from the PBLs of three additional healthy volunteers, showed similar fold increases when compared to the data obtained from RNA-Seq (fold values in parentheses correspond to fold values obtained from RNA-Seq) . The same can be said for those viruses that were downregulated by Tat (Figure 4.2). PCR products were sent for sequence verification, and were reported back as the single provirus product the primers were targeting. It is interesting to note that the RNA sample from the PBLs used for the RNA-Seq experiment showed almost the exact fold activation by qPCR as that calculated from RNA-Seq, and the viruses 5p12, 6p11.2 and 8q11.1 were undetectable. However, the additional healthy volunteer PBL RNA samples showed that, although for some viruses the RNA-Seq data reported one or no sequence reads when treated with Tat (5p12, 6p11.2, 8q11.1), these same viruses were indeed detected by qPCR, albeit they were clearly downregulated (Figure 4.2). Therefore, donor-responsiveness or sensitivity to Tat might also play a role in which viruses are activated or silenced after treatment. In conclusion, the qPCR analyses corroborate the RNA-Seq results and show that Tat indeed activates or silences HERV-K (HML-2), depending on the specific provirus.

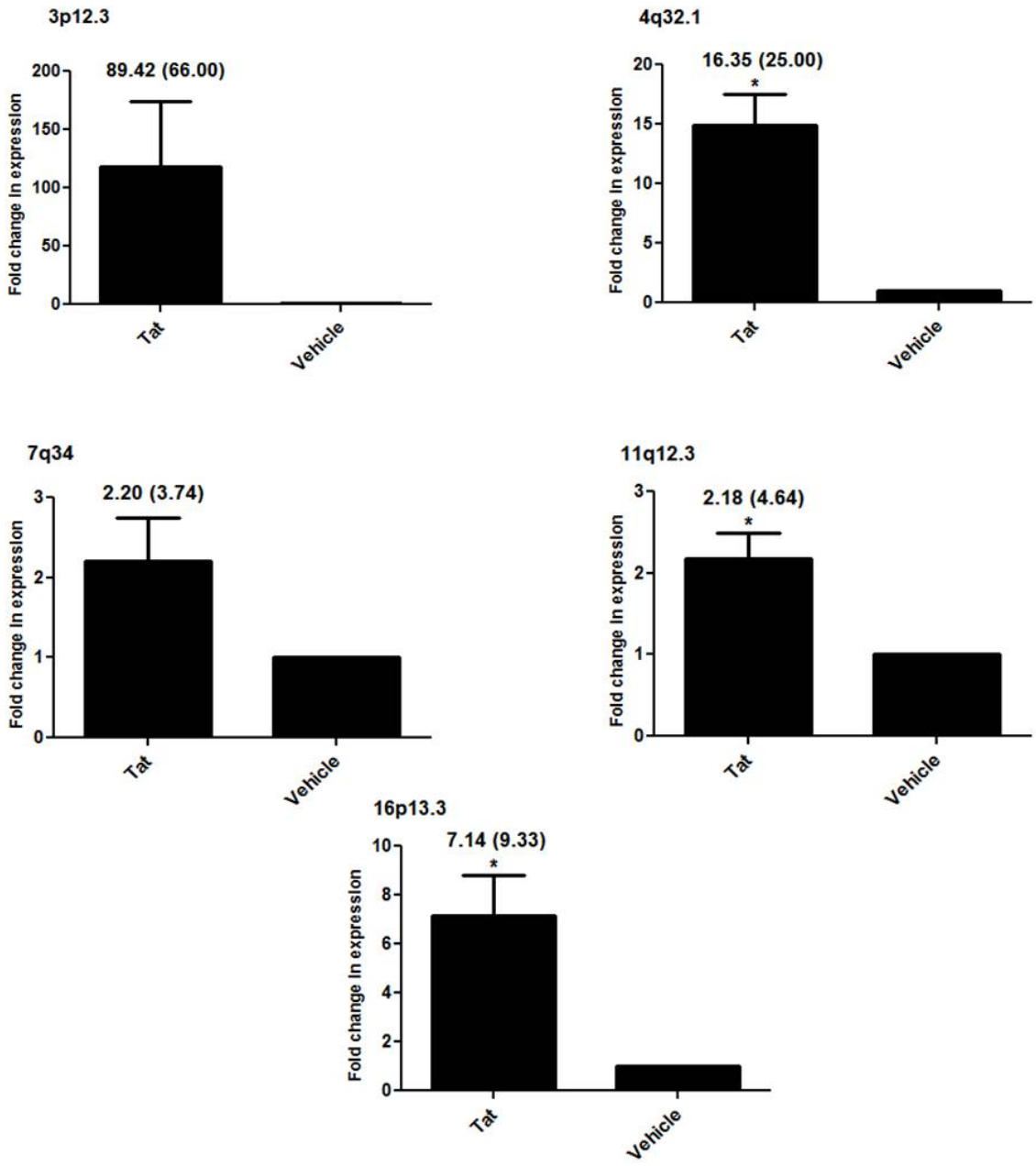


Figure 4.1 HERV-K (HML-2)-specific provirus qPCR validation of RNA-Seq data: upregulation. Total cellular RNA was isolated from primary cells (PBLs) that were treated with recombinant Tat protein or vehicle alone for 12 hours. RNA was DNAse-treated, and cDNA was synthesized and amplified by Sybr Green qPCR. Data are expressed as fold increase in RNA (Tat treatment over vehicle alone), with the fold written on top of the bar (parentheses include the fold obtained by RNA-Seq, for ease in comparison). All qPCR results were normalized to the GAPDH reference gene after analysis using the $2^{-\Delta\Delta Ct}$ method, and relative expression is plotted. Error bars indicate SEM for results of four independent experiments. Significance was calculated by comparing Tat treatment to vehicle, using a Student's T-test. Significant results are indicated (*p= <0.05).

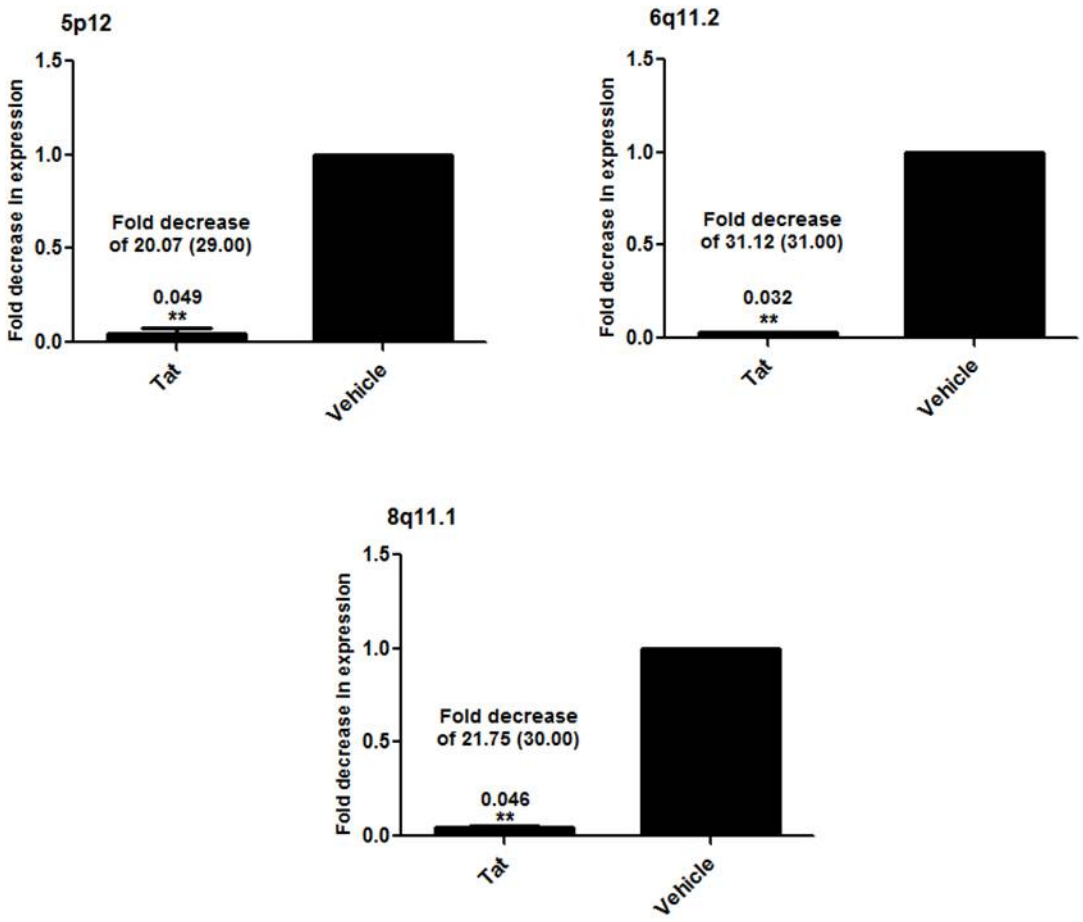


Figure 4.2 HERV-K (HML-2)-specific provirus qPCR validation of RNA-Seq data: downegulation. Total cellular RNA was isolated from primary cells (PBLs) that were treated with recombinant Tat protein or vehicle alone for 12 hours. RNA was DNAse-treated, and cDNA was synthesized and amplified by Sybr Green qPCR. Data are expressed as fold increase in RNA (Tat treatment over vehicle alone), with the fold written on top of the bar (parentheses include the fold obtained by RNA-Seq, for ease in comparison). All qPCR results were normalized to the GAPDH reference gene after analysis using the $2^{-\Delta\Delta Ct}$ method, and relative expression is plotted. Error bars indicate SEM for results of four independent experiments. Significance was calculated by comparing Tat treatment to vehicle, using a Student's T-test. Significant results are indicated (**p=<0.0005).

Nature of the HERV-K (HML-2) Tat-activated transcripts

In order to understand what could be causing the silencing of some HERV-K (HML-2) viruses in response to Tat, I analyzed the sequences of all the detected viruses (induced and silenced by Tat), to verify that they contained intact LTRs. As expected, the upregulated proviruses had complete LTRs with mostly all of the important NF-κB and NF-AT binding sites intact. Unexpectedly, however, viruses 1q21.3, 3p12.3, 14q11.2 and 16p11.2 have no, or almost no, 5'-LTR. Instead, they have an intact 3'-LTR. We had previously observed a similar phenomenon when analyzing expressed HERV-K (HML-2) sequences in the plasma of HIV-1-infected individuals [131], and thus this corroborates those findings and points to the possibility that when the integrity of the 5'-LTR is disrupted, the 3'- LTR acts as the viral promoter [380]. No cellular promoters flanking the 5'-area of these viruses were found that could explain how transcriptional initiation of these viruses takes place. Interesting, and quite puzzling, is the presence of 2 upregulated proviruses that actually have no LTRs (7q34, 16p13.3). However, as these viruses were very close to other highly expressed proviruses that contained either a functional 5' or 3'-LTR (7q22.2 and 16p13.3), it is possible that their expression is regulated by other HERV-K (HML-2) promoters (or solo LTRs) operating *in cis.*

For the downregulated viruses, like in their upregulated counterparts, the LTRs were mostly complete, with minor mutations or deletions. All but one of the downregulated viruses had intact sites for NF-κB and NF-AT, the exception being 8q11.1 (30-fold downregulated) which had no NF-κB binding sites. What was indeed consistent between the downregulated viruses was the absence of either the first or third NF-κB binding sites. From our previous analyses we know that those two binding sites are very important for Tat-

mediated activation, and their absence clearly affects HERV-K (HML-2) expression in response to Tat. Additionally, 1q32.2, 2q21.1, 6q25.1 and 19q13.42 also had only 3'-LTR sequences present, and their potential transcription factor binding sites appeared to be quite mutated.

Most of the Tat-induced HERV-K (HML-2) proviruses are Type 2, yet do not contain ORFs

Knowing that HERV-K (HML-2) encodes two putative oncoproteins, one of which might be involved in the LNX/Numb/Notch pathway, it is important to know whether any of the Tat-induced HERV-K (HML-2) proviruses encode functional ORFs. This information can be helpful in understanding the HERV-K (HML-2) protein expression patterns in a setting like that of HIV-1-associated lymphoma, where high levels of HERV-K (HML-2) Rec and Np9 proteins have been detected (unpublished observations). Table 4.3 summarizes the type of provirus (Type 1 versus Type 2), and ORF status for those viruses that were upregulated by Tat, and Table 4.4 shows the downregulated viruses. Interestingly, Type 2 HERV-K (HML-2) viruses were the most prominent type that was both activated and silenced by Tat. Also, the majority of the expressed Type 2 viruses do not have intact or full length ORFs that code for functional proteins. This finding is not completely surprising, as in our hands the response to Tat has always been much more robust at the RNA level than at the protein level. Those Type 1 proviruses that were activated, though fewer and not as highly expressed (with the exception of one), appear to be able to code for Np9.

Table 4.3 Type of HERV-K (HML-2) viruses expressed in response to Tat and their potential for protein expression

Chromosome and HERV names	Fold Activation	Type1/2	ORF
1p31.1 K4, K116, ERVK-1	1.40	1	Gag, Np9
1q21.3	3.02	ND	No ORF
1q22 K102, K(C1b),K50a, ERVK-7	2.38	1	Np9
1q23.3 K110, K18,K(C1a), ERVK-18	1.79	1	Np9
3p12.3	66.00	1	Np9
3q12.3 K(II), ERVK-5	2.20	1	Np9
3q13.2 K106, K(C3),K68, ERVK-3	1.00	2	No ORF
3q21.2 K(I), ERVK-4	1.25	1	Gag, Pol
4p16.1b K50c	4.52	2	No ORF
4q32.1	25.00	ND	No ORF
5q33.3 K107/K10,K(C5), ERVK-10	1.65	1	Gag, Pol, Np9
7p22.1b K108R, ERVK-6	2.07	2	Pol, Env (Rec)
7q22.2 ERVK-14	36.00	ND	Gag
7q34 K(OLDAC004979),ERVK-15	3.74	ND	Gag
8p23.1a K115, ERVK-8	2.32	2	Gag, Pol, Env
8p23.1c	8.25	2	No ORF
8p23.1d KOLD130352	18.50	2	No ORF
8q24.3b K29	8.11	2	No ORF
9q34.11 K31	1.95	2	No ORF
11p15.4 K7	3.60	2	No ORF
11q12.3 K(OLDAC004127)	4.64	2	Env
12q24.33 K42	6.00	ND	No ORF
14q11.2 K(OLDAL136419),K71	2.44	ND	Unknown
16p13.3 K(OLDAC004034	9.33	2	Unknown
16p11.2	4.00	1	No ORF
19p12c K51	3.00	1	Np9
19q13.12a	30.00	ND	Unknown
20q11.22 K(OLDAL136419),K59	1.29	2	No ORF
22q11.21 K101, K(C22),ERVK-24	14.00	1	Np9

*The different names after each chromosomal location are all alternative ways to refer to the same virus.

Table 4.4 Type of HERV-K (HML-2) viruses downregulated in response to Tat and their potential for protein expression

Chromosome and HERV names	Fold Decrease	Type1/2	ORF
1q32.2	0.00 (15 fold down)	1	No ORF
2q21.1	0.00 (11 fold down)	1	No ORF
3p25.3 K11, ERVK-2	0.38 (2.6 fold down)	1	No ORF
4p16.3a	0.25 (4 fold down)	ND	No ORF
4p16.3b K77	0.64 (1.6 fold down)	2	Rec?
4p16.1a K17b	0.64 (1.6 fold down)	2	Rec?
4q32.3 K5, ERVK-12	0.00 (6 fold down)	1	No ORF
5p12	0.03 (29 fold down)	1	No ORF
5q33.2 K18b	0.24 (4.25 fold down)	2	Gag, Pol, Rec
6p11.2 K23	0.031 (31 fold down)	2	Rec
6q25.1	0.18 (5.66 fold down)	2	Env, Rec
8p22	0.43 (2.33 fold down)	ND	Unknown
8q11.1 K70, K43	0.00 (30 fold down)	2	No ORF
9q34.3 K30	0.21 (4.66 fold down)	2	No ORF
10p14 K(C11a), K33,ERVK-16	0.09 (11 fold down)	2	No ORF
11q22.1 K(C11c), K36,K118, ERVK-25	0.20 (5 fold down)	2	Pol
11q23.3 K(C11b), K37,ERVK-20	0.19 (5.2 fold down)	1	Gag
12p11.1 K50e	0.00 (11 fold down)	2	No ORF
12q14.1 K(C12), K41,K119, ERVK-21	0.00 (8 fold down)	2	Gag, Pol, Env
19p12a K52	.105 (9.5 fold down)	2	No ORF
19q13.42 LTR13	0.25 (4 fold down)	2	No ORF
21q21.1 K60, ERVK-23	0.00 (16 fold down)	1	Np9
22q11.23 K(OLDAP000345),KOLD345	0.065 (15.5 fold down)	2	No ORF
Xq11.1	0.111 (9 fold down)	2	Unknown

*The different names after each chromosomal location are all alternative ways to refer to the same virus.

Correlation of HERV-K (HML-2) proviruses activated by Tat to those detected in the plasma of HIV-1-infected patients

We next analyzed whether the HERV-K (HML-2) proviruses that become expressed after HIV-1 Tat treatment correlate with those viruses whose transcripts we have found to be the most abundant in the plasma of HIV-1-infected individuals [131]. Table 4.5 shows the transcripts of HERV-K (HML-2) viruses most highly detected in HIV-1-infected plasma, as well as which of those transcripts is affected or not by Tat-treatment. Out of 34 HERV-K (HML-2) proviruses, 10 are induced by Tat treatment, 5 are silenced, and 19 are not affected by Tat. These data show that, while Tat is important for activating HERV-K (HML-2) expression, it is clearly not the only viral factor involved in the activation, and other HIV-1 proteins (e.g. Vif), as well as perhaps other cellular proteins, might play a role.

Table 4.5 Comparison between the most abundant HERV-K (HML-2) transcripts found in plasma of HIV-1 infected individuals and those seen after Tat treatment

HERV-K (HML-2) in HIV-1 plasma	Detected after Tat treatment	Direction of expression
K115	yes	up
K108	yes	up
12q14.1	no	n/a
K104	no	n/a
11q22.1	yes	down
K109	no	n/a
KC19	no	n/a
K(2)	no	n/a
10p14	yes	down
K50F	no	n/a
6p22.1	no	n/a
K106	yes	no change
3q27.2	no	n/a
11q23.2	no	n/a
K18/K110	yes	up
K107	yes	up
K101	yes	up
KOLD	yes	up
K52	yes	down
K8	yes	down
K102	yes	up
19q13.13	no	n/a
1p31.1	yes	up
19p13.11	no	n/a
K(I)	no	n/a
K76	no	n/a
1q32.2	yes	down
16p11.2	yes	up
2q1	no	n/a
K103	no	n/a
K113	no	n/a
4q35	no	n/a
19p12	no	n/a
K111	yes	up

*n/a = no change

Discussion

Next generation sequencing technology has allowed us to analyze the transcriptome of HERV-K (HML-2) proviruses in response to HIV-1 Tat in peripheral blood lymphocytes. Out of the 91 annotated HERV-K (HML-2) proviruses, Tat activates 28 and silences 24. Most affected proviruses are Type 2, and do not code for protein, while the expressed Type 1 proviruses are able to code for Np9. Most importantly, the RNA expression levels were validated by amplification through qPCR of RNA isolated from 3 additional healthy volunteers, as well as from the same sample used for RNA-Seq analysis, so we are confident that the results are correct.

Our analysis is consistent with the duality of HIV-1 Tat's effect on cellular genes, as it can be both an activator and a suppressor protein. As it does for HERV-K (HML-2), in terms of activation and silencing, Tat has been shown to behave similarly for several other human genes (e.g. actin, IFN-γ [381-384]). The effects of Tat-mediated regulation of cellular genes are still under study, but some believe that this is one way by which Tat and HIV-1 can induce cell death (e.g. upregulating pro-apoptotic genes while downregulating anti-apoptotic ones) [385, 386].

An additional interesting observation is that some viruses that were activated by Tat actually did not contain any viable 5'-LTR promoter regions. We did not observe cellular promoters close to sequences upstream of these proviruses, so activation mediated by

intact 3'-LTRs and some activation due to nearby HERV-K (HML-2) proviruses must be considered. 3'-LTR-mediated antisense transcription may explain the finding of RNA

transcripts having arisen from these proviruses, as previous investigators have also found expression of other retroviruses when the proviral sequences are devoid of the 5'- LTR [380]. We have previously detected similar transcripts in the plasma of HIV-1-infected patients and have suggested this transcriptional mechanism for HERV-K (HML-2) activation [131]. Transcription driven from the promoter in a 3'-LTR has also been described for gammaretroviruses, HTLV-1, and even for HIV-1 [387-390], so it is quite possible that HERV-K (HML-2) also utilizes this mechanism. Our data, originating from analyses of the HERV-K (HML-2) Tat-induced transcriptome, make it clear that, though NF-κB and NF-AT activation are important for HERV-K (HML-2) transcription, there is a broader and more complex mechanism at play in the regulation of HERV-K (HML-2) expression by HIV. The ramifications of our transcriptome findings are discussed in further detail in the next chapter.

Materials and Methods

Isolation and Culture of Primary Cells

Peripheral blood mononuclear cells (PBMCs) were obtained by venipuncture from healthy donors and monocytes were separated from peripheral blood lymphocytes (PBLs) by differential adhesion to plates as previously described [341]. PBLs were washed three times with PBS and stimulated with 5 μg/mL phytohemagglutinin (PHA-P, Sigma-Aldrich, St. Louis, MO) for 3 days in RPMI-1640 complete media containing 10% heat-inactivated FBS and 10 U/mL of Interleukin-2 (IL-2, Sigma-Aldrich, St. Louis, MO).

Addition of exogenous Tat protein

The purified recombinant 86 amino acid form of the HIV-1 Tat protein was obtained from the NIH AIDS Research and Reference Reagent Program (the late Dr. John Brady and DAIDS, NIAID, [342]), or from ProSpec Protein Specialists (Cat. No. HIV-129, ProSpec-Tany TechnoGene Ltd., East Brunswick, NJ). The protein was resuspended in sterile phosphate-buffered saline (PBS, Gibco/Invitrogen, Carlsbad, CA) containing 1 mg/mL bovine serum albumin (BSA) and 0.1 mM dithiothreitol (DTT) (both from Sigma-Aldrich, St. Louis, MO), de-aerated, and protected from light. Tat protein was added at a concentration of 50 ng/mL for 12 hours.

RNA extraction

Total cellular RNA was isolated from cells using the RNeasy Mini Kit (Qiagen, Valencia, CA), and subjected to RNase-free DNase treatment (Qiagen, Valencia, CA) for 15 minutes at room temperature. RNA concentration and purity were measured using a spectrophotometer, calculating the 260/280 ratio. RNA integrity (as well as the absence of DNA contamination) was confirmed by One-Step RT-PCR using GAPDH amplification with primers that can bind both genomic and cDNA, employing the PCR conditions described below, as well as "No RT" controls. If DNA contamination was detected, another round of DNase treatment was performed using the DNA-free DNase removal kit (Ambion, Austin, TX) following the manufacturer's protocol.

cDNA Library preparation and RNA-Seq platform run

Total RNA from PBLs of a healthy volunteer was isolated, purified and DNAse-treated after 12 hour Tat and vehicle treatments, as specified above. RNA integrity was verified using the Agilent RNA 6000 Nano Kit (Agilent, Santa Clara, CA) RNA nano chip for quality, and the NanoDrop for quantity (Thermo, Rockford, IL). cDNA was prepared for the RNA-Seq library using the NuGen Encore complete library prep kit (Invitrogen, Carlsbad, CA), and quality controlled using the Bioanalyzer DNA 1000 (Agilent, Santa Clara, CA). qPCR for quantification of the library was performed with the KAPA kit (Kapa Biosystems, Woburn, MA). Samples were sequenced on the HiSeq2000 (Illumina, Madison, WI), in a Paired-end 100 cycle run, using TruSeq SBS v3 reagents (Illumina, Madison, WI).

In silico sequence analysis: RNA- Seq

The locations of the currently known 91 HERV-K (HML-2) genomes were organized in a BED formatted file and BEDTools [379] was used to intersect the HERV-K (HML-2) locations in the sequenced samples that had acceptable hits. All hits were deduplicated to remove PCR artifacts, Tophat v2.0.4 [391, 392] mapped and subjected to Cufflinks v2.0.2 [392] analysis. After RNA-Seq reads were obtained (140 million reads per lane), a customized Perl script was written to count the reads with the following criteria:

1. A read was discarded if it spanned the entire HERV-K (HML-2) genome – these were actually reads that were aligned to exons on either side of the genome and did not actually align to HERV-K (HML-2).
2. Uniquely mapping reads were identified by an MQ score of at least 255. Reads that could map to 2 locations (MQ=3) were also reported on a separate column.
3. For this purpose, reads were counted individually regardless of whether the read mapped as a pair or singly.

qRT-PCR

To validate the RNA-Seq data, we performed quantitative PCR on cDNA synthesized from RNA isolated from Tat-treated PBLs from healthy volunteers using the BioRad Supermix iScript SYBR Green kit (Bio-Rad, Hercules, CA). Briefly, 500 ng of total cellular RNA was reverse transcribed for 30 min at 50 °C using the iScript Reverse Transcription Supermix for RT-qPCR (Bio-Rad, Hercules, CA) and diluted to 100 µL. A 20 µL qPCR reaction was made

using 5 µL of cDNA and 0.6 µM specific primers designed and verified by BLAST analysis to amplify 9 specific HERV-K (HML-2) proviruses. The PCR reaction consisted of an initial denaturing step of 15 min at 95 °C, followed by 40 cycles with optimal conditions as follows: 94 °C for 15 sec, 57 °C for 30 sec, 72 °C for 10 sec. This was followed by a melt curve, with fluorescence captured at 78 °C or 81 °C, which determined the product amplified to be absent of signal generated by primer dimers. Data were collected and recorded by ABI StepOne Plus (Applied Biosystems). GAPDH amplification was used to normalize samples to an endogenous reference gene as stated in the figure legends.

List of primers used

The amplification product was designed to either contain one primer or both primers binding to each virus's unique regions. Sequences were verified through BLAST analysis so that any other possible viruses or cellular genes that could be amplified with the primer pairs had 75% or less sequence identity.

7q34 KOLD/ERVK-15 F
CCAACCCTGTGCTCGTAGAAACAA
7q34 KOLD/ERVK-15 R
GCACATCCTACATAGCCCTAAATCC

8q11.1 K70/K43 F
GTGAAGAAGAGGCAGGAAGAGAG
8q11.1 K70/K43 R
TTGCTTTGACTGAGCCACTACGGA

5p12 F
GGTCGAGCTCTTCAACCAGTGAGTTT
5p12 R
CAGATGCTATTGCCAGTCCTGCAT

6p11.2 K23 F
GGTATGGTATGGAATGATTGGGCCA
6p11.2 K23 R
GGGCATCAGACACTGAAACACT

3p12.3 F
CTGTTATAACTGTGGTCAAATCGGTC
3p12.3 R
AGGCCAGGTGCCTCCTTT

4q32.1 F
GCTCGGAAGAAGCTAGGGTGATAA
4q32.1 R
TGGTTCTTCTGTTTGAAATGGCTTG

11q12.3 (KOLD) F
ACAGATGATCGTTGCCCTGCCAAA
11q12.3 (KOLD) R
AACATCCTGGCGCTAAACATCCTG

16p13.3 (KOLD) F
CTCTAAAGAGCCCTACCCTGACTT
16p13.3 (KOLD) R
TGGCCAATTGACATTCCG

Analyses of the LTRs of HERV-K (HML-2) proviruses

Analyses of the HERV-K (HML-2) LTRs for mutations and deletions was performed using published sequences [43, 131]. Potential transcription factor binding sites were analyzed using the online prediction software tools ALGGEN PROMO and version 8.3 of TRANSFAC (BioBase Co., Beverley, MA, [310, 311]).

Statistical Analysis

The mean number of HERV-K (HML-2) cDNA relative fold expression units between Tat treatments and controls were compared using a student's *T* test for samples exhibiting normally distributed values. Two-tailed *p* values were considered significant at $p<0.05$.

CHAPTER V

CONCLUSIONS

The human genome is made of coding genes, regulatory RNAs, repetitive non-coding DNA elements, and retroviral sequences contained in multiple copies, the latter known as human endogenous retroviruses (HERVs). These HERVs entered the germline at various points during human evolution, and are now an integral part of human biology. Most of these viruses have accumulated lethal mutations and deletions that have rendered them inactive or have made them just a shadow of their former selves. However, some of these viruses have managed to maintain partial protein coding capacity and a number of their viral proteins can be expressed under various conditions. Interestingly, one of the viruses that fits this last category is HERV-K (HML-2), the youngest of the HERVs found in our genome, and in which our laboratory has had a long-time interest, as it is found expressed at high levels in patients with HIV-1 infection and/or lymphoma. However, currently it has been difficult to ascertain whether HERVs like HERV-K (HML-2) have any sort of role in normal human physiology or disease, or if they are just genetic passengers that randomly become activated. Yet, as HERVs make up 8% of the genome, it appears quite likely that they will play various biological roles. Multiple studies have analyzed how HERV-K (HML-2) is affected in both health and disease, focusing on their possible biological/physiological roles,

as well as trying to link them to disease pathogenesis. However, no actual conclusion has been reached on the topic. What is known is that a correlation exists between disease states, inflammation, and the expression of these endogenous retroviral sequences. Therefore, although a pathological role for HERV-K (HML-2) has not been established thus far, their increased expression in certain disease states cannot be ignored. Whether or not their activation is the result of the disease state, or the cause of it, is under investigation in a number of laboratories, including our own.

Infection of humans with HIV-1, has profound effects upon the resident endogenous retroviruses, inducing very high expression of transcripts and proteins from diverse HERV-K (HML-2) viruses. Levels of HERV-K (HML-2) become extremely high and remain high if no successful treatment occurs. Interestingly, I and others in our laboratory have found that, in HIV-1-associated lymphoma tissue and in the plasma of patients, very high levels of the Rec and Np9 putative oncoproteins from HERV-K (HML-2) are detected (both at the RNA and at the protein level) ([131, 190], and unpublished observations). These RNA levels drop dramatically with effective treatment of the lymphoma. Further, the presence of HERV-K virus-like particles in the plasma of lymphoma patients was shown using immunoelectron microscopy [131, 190].

Lymphomas are the most common malignancy found in HIV-infected individuals worldwide and contribute substantially to the morbidity and mortality of AIDS patients. At the onset of the AIDS epidemic, the majority of lymphomas found in these patients arose in the central nervous system (CNS) [393, 394], and were almost all Epstein- Barr Virus (EBV)-associated [393, 395]. Since the advent of highly active antiretroviral therapy (HAART), CNS lymphomas have become rare in the developed world [393, 394], but peripheral lymphomas

have remained a quite common and difficult-to-treat malignancy in these often debilitated patients [393, 396-399]. Peripheral lymphomas in AIDS patients are associated only rarely with Human Herpes Virus Type 8 (HHV-8) infection, and with EBV infection less than 50% of the time [394, 396, 397, 399, 400]. The etiology of EBV and HHV-8-negative peripheral lymphomas in AIDS patients has remained unexplained.

As AIDS-related lymphomas are a heterogeneous group, there are likely a number of pathologic factors that influence lymphomagenesis in this group of patients. Clearly, HIV-induced immunosuppression plays an important role, but the induction of cytokines seen in HIV-infected patients has also been hypothesized to stimulate B cell growth and hence favor the development of lymphomas in this group of individuals [395, 401, 402]. Dendritic cell impairment has also been thought to promote lymphomagenesis [394], but viral oncogenesis has been hypothesized to play the dominant role in most HIV-related lymphomas. Ultimately, however, 50% of HIV-associated lymphomas arise in the absence of EBV or HHV-8 and the cause of these tumors remains elusive. Therefore, it is quite possible that other viral agents participate in an opportunistic manner in causing or potentiating the development of lymphomas in HIV-infected individuals. As I have found expression of the HERV-K (HML-2) oncoproteins Np9 and Rec in HIV-1-associated lymphomas using tissue microarrays (MJ Gonzalez-Hernandez, unpublished data), we believe HERV-K (HML-2) to be one such potential opportunistic agent.

Both our published data and those shown above strongly suggest that HIV infection activates HERV-K gene expression, and that this activation may play into pathogenesis, particularly by promoting lymphomagenesis. Therefore, we began to investigate how HIV could activate HERV-K gene expression. There is elegant published work available showing

that the Rev protein of HIV-1 can act upon HERV-K RNA [248, 252]. However, as HIV infection would primarily take place in T cells and monocytic cells, it would be difficult to understand how Rev could impact upon B cell lymphomagenesis. However, the Tat transactivating protein of HIV-1 is known to act upon not only the HIV-1 and HIV-2 promoters, but also on cellular promoters. In addition, Tat has the property of being both secreted and taken up by other cells, thus allowing Tat made in one cell, such as a T cell, to act upon another cell, such as a B cell. For this reason, I examined whether Tat might activate the HERV-K promoter and thus promote the expression of endogenous HERV-K (HML-2) and its oncogenes.

This thesis work has focused on the presence of endogenous retroviral elements in our genome, and how these elements become affected by external stimuli. The main purpose of this project was to elucidate the mechanism by which HIV-1 infection causes increased expression of HERV-K (HML-2) genes. The research also involved the development and optimization of several novel molecular tools and reagents, as well as the use of next generation sequencing technology, to allow the detection and better understanding of HERV-K (HML-2) gene products and their activation.

HERV-K (HML-2) and HIV-1 Tat

Here we have shown that the HIV-1 Tat protein plays a modulatory role in the expression of HERV-K (HML-2): it acts as both an activator and a suppressor protein. We first demonstrated that both transfected and exogenously administered Tat protein causes the activation of the HERV-K (HML-2) promoter. This results in increased HERV-K (HML-2) expression at both the RNA and the protein level. Though the activation is more robust at

the RNA level, there are still significant increases seen when looking at HERV-K (HML-2) Gag protein. Additionally, the mechanism of activation occurs in a manner different from that of how Tat is known to activate HIV-1's promoter: it does not involve interactions with RNA-secondary structures (as there are no known potential TAR-like structures in HERV-K), nor does it involve interactions with Cyclin T1 and P-TEFb. Instead, activation occurs through several mechanisms. First, for those HERVs that are not in particularly dense areas of chromatin, or that are perhaps not under stringent epigenetic control for other reasons, the interaction of the transcription factors NF-κB and NF-AT with upstream promoter elements in the HERV-K (HML-2) LTR in response to Tat causes expression of viral genes. If the target HERVs are in heterochromatic areas, as is the case with the centromeric K111, activation of viral gene transcription occurs through the modification of the chromatin region surrounding the virus. This modification involves decreasing the heterochromatic markers H3K9Me3 and H4K20Me3, allowing for transcription factor binding to the viral LTR.

We then moved from traditional molecular biology systems into newer technologies in order to understand which specific proviruses, out of the known 91 HERV-K (HML-2), are activated by Tat. Using NGS technology platforms (RNA-Seq), we analyzed a Tat-induced HERV-K (HML-2) transcriptome of total RNA isolated from PBLs of a healthy volunteer. We found that Tat activates 24 unique viruses, among them the newly discovered K111, while it silences 28. These results were corroborated through qRT-PCR analyses using primers designed for specific proviruses. The fact that Tat caused silencing of some viruses was somewhat surprising to us, as we believed that it would either activate a given HERV-K (HML-2) or have no effect. However, it should be noted that although the expression of quite a number of proviruses was turned down, the magnitude of the effect was not as

marked as that seen with activated proviruses, in keeping with our earlier observations that Tat increases HERV-K (HML-2) expression on the global level. We also have to keep in context the fact that Tat has been shown to act both as a suppressor and an activator of human genes, so it is logical that the same would occur for human endogenous retroviral genes (as they are by definition also human genes). Interestingly, and surprisingly, RNA-Seq data showed that some of the activated viruses actually either had deletions or mutations in their 5'-LTRs, or lacked 5'-LTRs altogether. This meant that their activation must be occurring through another mechanism: solo LTRs or other viruses with intact LTRs in the vicinity might be contributing to activation, cellular promoters could direct RNA expression, or 3'-LTRs could be driving transcript formation. This new discovery has given us an insight into the fact that the transcription process for these endogenous retroviruses is clearly more complex than previously thought.

Future explorations: testing the consequences of increased HERV-K (HML-2) expression

The results obtained from the experiments discussed in this thesis point towards future research involving understanding the role of HERV-K (HML-2) expression in three very important aspects of the human response to HIV-1. First, fully delineating the mechanism for transcriptional expression of HERV-K (HML-2), the most active LTRs, as well as the biological role of the transcripts, is important for understanding their potential effects in the host (e.g. cancer, impaired immunity) in response to HIV-1 infection. Second, understanding the role of HERV-K (HML-2) in the development, or maintenance, of HIV-1-associated disease is important for finding the best possible means of treatment. Lastly, understanding

the consequences that increased expression of HERV-K (HML-2) has specifically for HIV-1 replication is important for delineating the mechanisms that HIV-1 might be utilizing to better avoid destruction, or that the host might be employing to better control HIV-1 pathogenesis.

From our transcriptome data, we have observed that HERV-K (HML-2) transcriptional expression and regulation in response to HIV-1 are more complex than previously thought. As we have seen that transcription of HERV-K (HML-2) proviruses might not only involve normally promoted transcripts but might include anti-sense transcription (from 3'-LTRs), the prospect of HERV-K (HML-2) being a potential culprit in HIV-1-associated cancer development, or promotion, becomes stronger. 3'-LTR-promoted transcripts are a form of anti-sense transcripts, and might even be c*is*-natural anti-sense transcripts (cis-NATs, or transcripts derived from the same genomic locus as their target but from the opposite DNA strand) [403, 404]. Although natural anti-sense transcripts have been recognized for a long time, their importance has been overlooked due to their heterogeneity, low expression level, and unknown function [405, 406]. Anti-sense transcription has been proposed to contribute to disease through chromosomal changes that result in the altered production of sense transcripts, gene silencing or altered ratios of sense/anti-sense transcripts, which may contribute to metabolic or other alterations associated with cellular transformation [407, 408]. For example, it has been shown that *cis*-NATs of the *CDKN2B* gene are involved in the control of its epigenetically regulated expression through heterochromatin formation in leukemia cells [409]. Also, the HTLV-1 basic leucine zipper factor (HBZ), which is only encoded as a *cis*-NAT generated from the viral 3'-LTR, is believed to contribute to development and maintenance of Acute T-cell

Leukemia [404, 410]. Additionally, it is thought that malignant cancer cells with activated transposable or repetitive elements have a large amount of transcriptional noise that could result in aberrant anti-sense RNA transcripts, which may cause stochastic methylation/hypomethylation of CpG islands associated with promoters of oncogenes and tumor suppressors [411-415]. From this information, it might be hypothesized that anti-sense transcripts derived from HERV-K (HML-2) 3'-LTR promoters could thus affect other cellular, non-viral genes. Indeed, there is evidence that 3'-LTR-generated HERV-K (HML-2) transcripts can regulate human C4 gene expression [143], and that at least two other HERV-K (HML-2) solo LTRs participate in the specific anti-sense regulation of human genes [416]. These 2 solo LTRs are situated in the introns of the genes SLC4A8 (sodium bicarbonate cotransporter) and IFT172 (intraflagellar transport protein 172), in the oppositive DNA strand for transcription, and generate RNAs complementary to the exons of the encoding genes [416]. The anti-sense transcripts formed were shown to decrease the mRNA level of their corresponding genes. Thus, it is possible that the increased expression of 3'-LTR-generated anti-sense transcripts in response to Tat and HIV-1 infection might behave as modulators of host gene expression. Of note is the fact that, potentially, not only can genes involved in oncogenesis be modulated by these antisense transcripts, but expression of genes important in cellular metabolism or immune function may also be disrupted. Therefore, future goals for this project include fully understanding the mechanism for HERV-K (HML-2) transcriptional activation by HIV-1 and HIV-1 Tat, trying to delineate 3'-LTR-driven transcript generation, and determining the potential targets of these anti-sense transcripts, if any.

Our transcriptome data also show that, though most of the transcripts from activated HERV-K (HML-2) viruses do not code for protein, those that do mostly code for Np9 and Rec. Interestingly, due to the types of proteins that these transcripts encode they could potentially have a role during HIV-1-associated pathogenesis. Np9 is a putative oncoprotein associated with the LNX/Numb/Notch pathway that may affect c-myc expression [246, 247, 417]. Transcripts for Np9 are rarely detected in healthy tissue, yet are highly expressed in transformed cells (e.g. EBV-transformed B cells) [417, 418], and we see very high expression levels of Np9 during HIV-1 infection and HIV-1-associated lymphoma. All these facts potentially link Np9 to cancer development, maintenance, or progression. Interestingly, the K111 virus, which is activated by Tat and only detected during HIV-1 infection, has only a single ORF: Np9. HERV-K (HML-2) additionally encodes Rec, a tetrameric RNA export adapter protein that is also a putative oncoprotein. It is functionally homologous to HIV-1 Rev and has been shown to induce carcinoma *in situ* in transgenic mice, to increase the activity of c-myc, and to interact with the androgen receptor (AR) [241, 245, 419, 420]. As the main function of Rec involves transporting unspliced viral RNAs from the nucleus to the cytoplasm, it is possible that this protein could also augment the transport of unspliced or incompletely spliced host cellular RNAs, thus altering the global processing of cellular transcripts. This could lead to malignancy in several ways: if tumor suppressor RNAs, cell cycle regulator RNAs, or pro-apoptotic protein RNA transcripts are transported to the cytosol before they are correctly processed they could either be incorrectly translated, or not translated at all. If the correct conditions are present, this could lead to oncogenesis or tumor maintenance. As we have also detected increased expression of Rec in HIV-1-associated lymphoma, this protein might also play a role in AIDS-

related malignancy or other HIV-1-related pathogenesis. Therefore, these patterns of increased HERV-K (HML-2) oncoprotein expression in response to HIV-1 Tat, and HIV-1 infection, merit further investigation into understanding their role, if any, in the development of malignancy in patients infected with HIV-1.

Another potential question arising from this project focuses not on what effect the expression of HERV-K (HML-2) has on the host, but the effect it has on HIV-1. Interestingly, the potential for HERV-K (HML-2) to produce anti-sense transcripts that may silence expression of host genes during HIV-1 infection could prove beneficial for HIV-1 replication if, for example, silencing of viral innate immune detection molecules (e.g. RIG-I-like receptors, NOD-like receptors, Toll-like Receptors) occurs. This would provide an advantage for the virus in avoiding immune detection. Anti-sense transcription might also interfere with the production of proteins that restrict HIV-1 replication. Therefore, the potential cellular targets of HERV-K (HML-2) anti-sense transcripts that could contribute to increased efficiency of HIV-1 replication or immune evasion should be addressed.

Our observation that expression of the centromeric HERV-K (HML-2) K111 only occurs during HIV-1 infection or after HIV-1 Tat treatment might point to a potential role for K111 in HIV-1 infection. Unpublished observations from our laboratory indicate that a region of K111's sequence is quite similar to that of microRNA-1299 (miR-1299). Such a microRNA could affect HIV-1 replication and disease by modulating expression of cellular cofactors [421, 422]. A regulatory RNA might also affect HIV-1-related disease by promoting tumorigenesis. Interestingly, the potential miRNA to which K111 is similar, miR-1299, is highly expressed in ovarian carcinomas [423], yet its expression in other cancers has not been fully determined. Therefore, silencing expression of K111 during HIV-1 infection, and

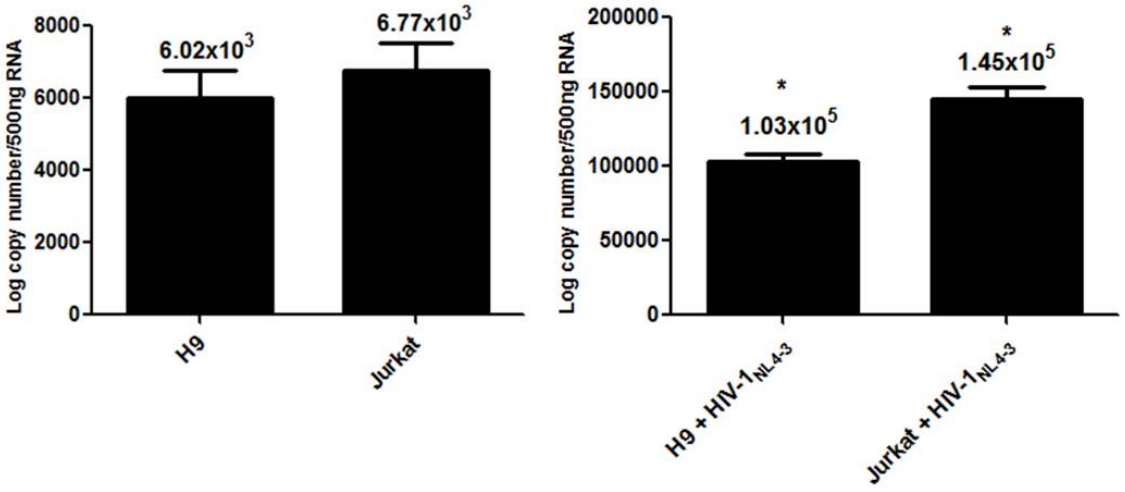
observing the potential effects it might have on HIV-1 replication and on HIV-related cancers like lymphoma, is thus a future goal of this work.

In a paper from our collaborators in the Ono laboratory at the University of Michigan, which has recently been accepted for publication, it was shown that HERV-K (HML-2) Gag is co-packaged in HIV-1 particles and that this co-packaging interferes with HIV-1 release efficiency (Monde K. *et al* 2012, accepted in Journal of Virology). Additionally, HERV-K (HML-2) Gag overexpression reduced not only HIV-1 release efficiency but also its infectivity in a myristylation- and NC-dependent manner. This interesting observation shows a potentially beneficial role of increased HERV-K (HML-2) expression for the host, as increased HERV-K (HML-2) Gag expression can modulate the late phase of HIV-1 replication. Additionally, it has recently been shown that, although Gag- and Env-specific T cell responses are infrequent [258], HERV-K antigens promote a T cell response against HIV-1 [112, 259]. Increased expression of HERV-K (HML-2) following HIV-1 infection thus might actually be helpful in controlling the replication of HIV. Therefore, there are data to suggest both beneficial and detrimental effects from the increased HERV-K (HML-2) expression seen in the setting of HIV-1 infection, and so further studies addressing this issue are clearly indicated.

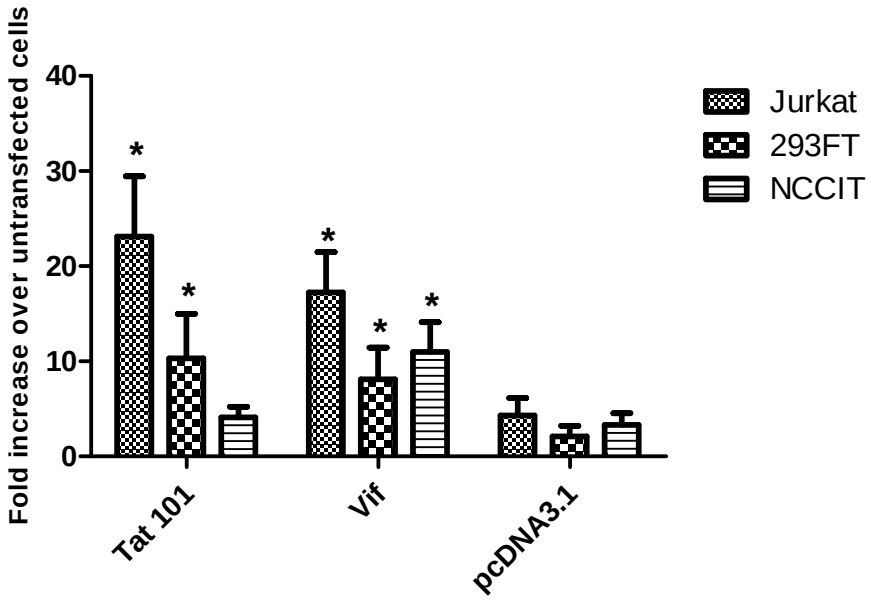
Taken together, the research presented here has added important information to the literature showing how infection of humans with HIV-1 has profound effects upon HERV-K (HML-2) expression. As this endogenous retrovirus is the most likely candidate in our genome to code for a fully functional virus, and as it encodes putative oncoproteins and produces anti-sense transcripts that might modulate host gene expression, the consequences of HERV-K (HML-2) expression could be profound for both the host and for

HIV-1. Future studies in our laboratory will further address the biological and clinical implications of the interactions between these key endogenous and exogenous retroviruses.

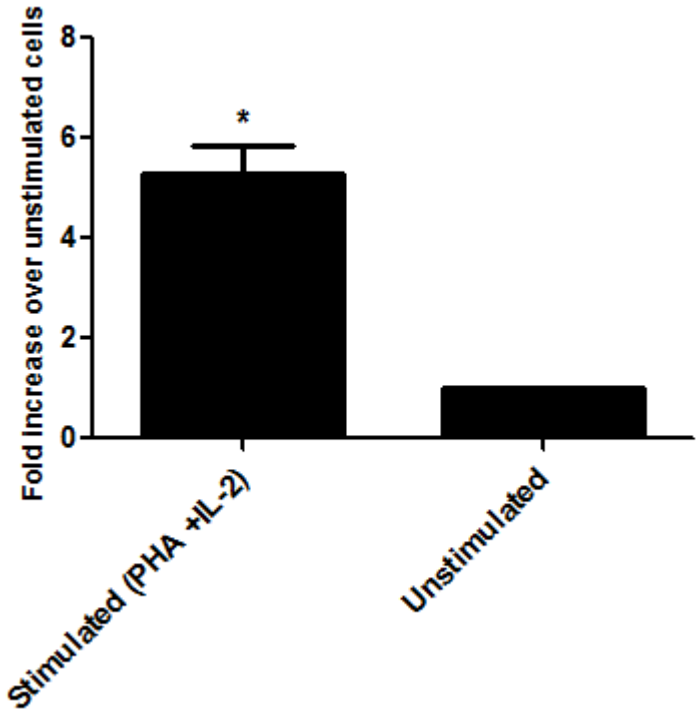
Appendix



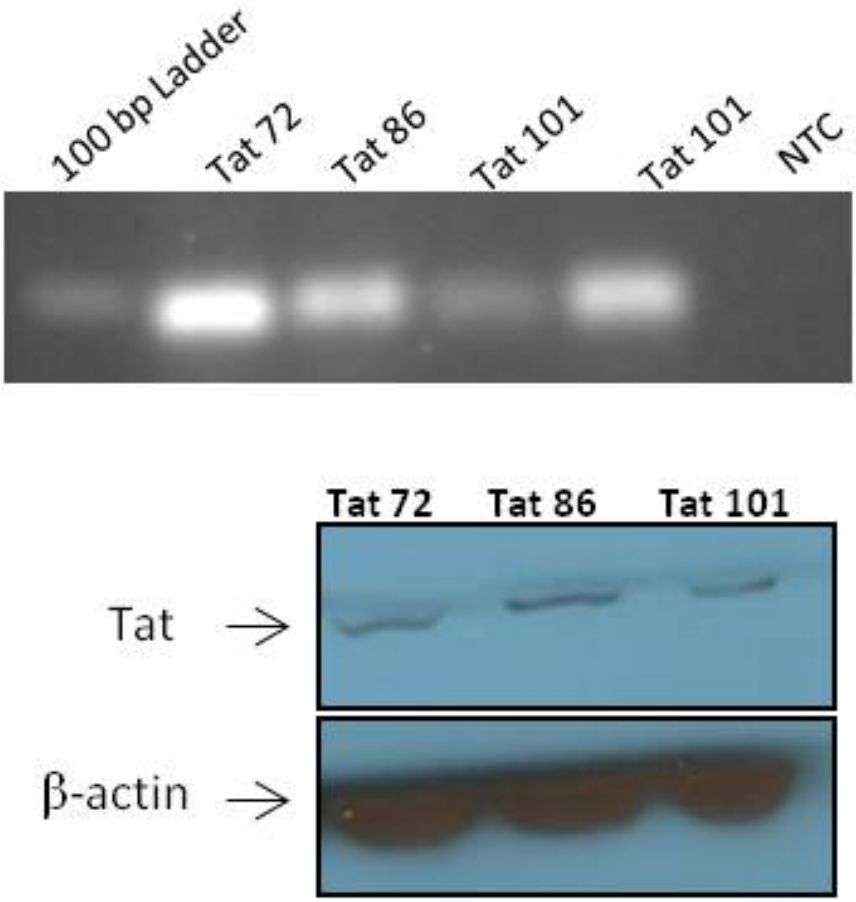
Appendix 1.1 HERV-K (HML-2) *gag* RNA levels are substantially increased in two T cell lines after HIV-1 infection. Total cellular RNA was isolated from cells that were infected with HIV-1_{NL4-3} (for one week), or left uninfected. RNA was amplified using primers specific for HERV-K (HML-2) *gag* through One Step qRT-PCR, and quantified using a standard curve generated by amplification of HERV-K (HML-2) gag RNA standards. Data are expressed as log RNA copy number per 500 ng of input RNA. Error bars indicate SEM for results of three independent experiments. Significance was calculated by comparing infected samples with their uninfected counterparts using a Student's T-test; significant results are indicated (*p= <0.005).



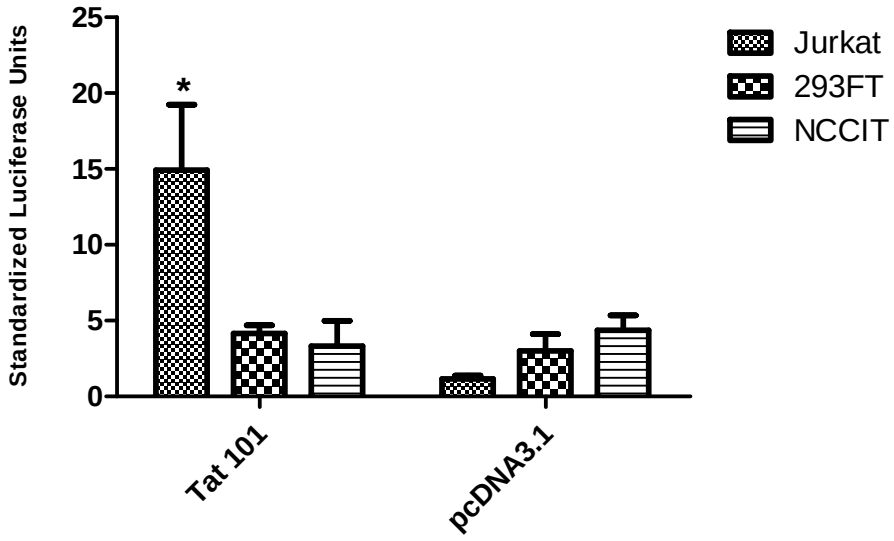
Appendix 1.2 Transfection of HIV-1 Tat or Vif causes an increase in HERV-K (HML-2) *gag* RNA after 48 hours. Cells were transfected with full length Tat and/or Vif encoding plasmids, and total RNA was extracted 48 hours later and subjected to One-Step Sybr Green qRT-PCR using primers specific for HERV-K *gag.* Data are presented as fold increase over untransfected cells. Error bars indicate SEM for results of three independent experiments. Significance was calculated by comparing Tat or Vif transfections to the empty vector transfection (pcDNA3.1) using a Student's T-test, and significant results are indicated (*p= <0.05).



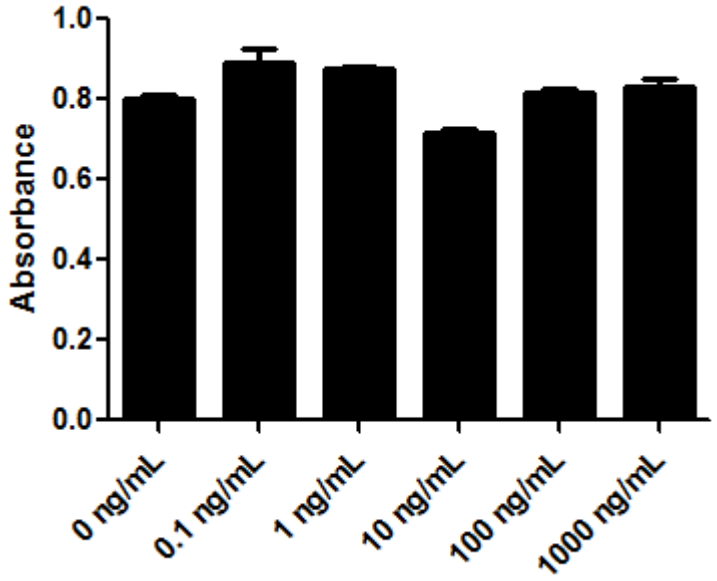
Appendix 1.3 Stimulation with PHA and IL-2 activates HERV-K (HML-2) g*ag* RNA expression in peripheral blood lymphocytes (PBLs) after 72 hours. Isolated PBLs were split in half and either stimulated for 3 days with PHA and IL-2, or left untreated. Total RNA was then isolated and HERV-K (HML-2) *gag* was amplified by One Step Sybr Green qRT-PCR. Results were normalized to the GAPDH reference gene after analysis using the $2^{-\Delta\Delta Ct}$ method, and relative expression is plotted as fold over unstimulated cells. Error bars indicate SEM for results of three independent experiments. Significance was calculated by comparing stimulated samples to untreated controls, using a Student's T-test, and significant results are indicated (*p=<0.05).



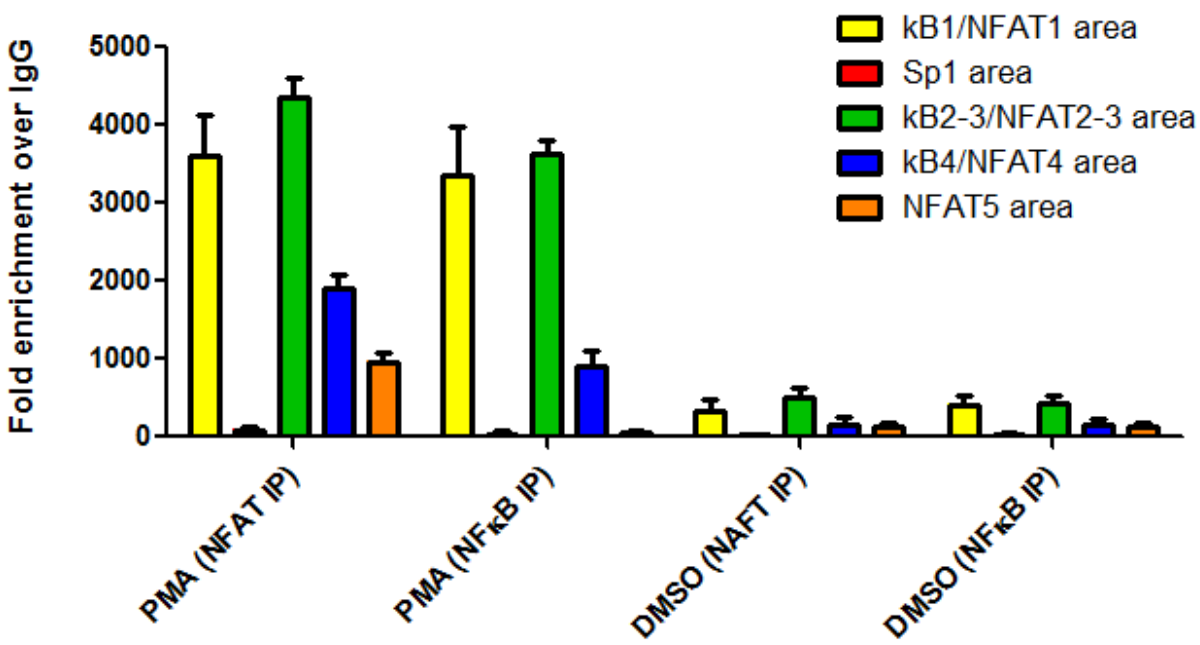
Appendix 1.4 Expression of the different HIV-1 Tat isoforms in Jurkat T cells. Upper panel: Agarose gel electrophoresis of an RT-PCR reaction analyzing the amplification of the different Tat isoforms from total RNA 24 hours after transfection, in Jurkat T cells, with plasmids encoding each of the different HIV-1 Tat proteins (NTC= non-template control). Lower panel: HIV-1 Tat protein isoform expression analysis by Western Blot from Jurkat T cell lysates 48 hours after transfection, using the HIV-1 Tat-specific antibody 1D9.



Appendix 1.5 Expression from the HERV-K (HML-2) LTR in response to Tat is different according to the cell type tested. Jurkat T cells, 293FT cells and NCCIT cells were transfected with a HERV-K (HML-2) LTR-luciferase reporter plasmid and co-transfected with a plasmid encoding the full length Tat protein of HIV-1. Activation of the luciferase construct was measured 24 hours after transfection, normalized to *Renilla* luciferase signal, and shown as standardized luciferase units. Fold induction was calculated over the empty vector, pcDNA3.1. Significance was calculated using a Student's T-test comparing Tat-transfection to pcDNA3.1 and significant results are indicated (*p=<0.05).



Appendix 1.6 MTT assay to determine Cyclosporin A (CsA) cellular toxicity. Jurkat T cells were plated and treated overnight with increasing concentrations of CsA. After each time point, a cellular toxicity assay (MTT assay) was performed and revealed no significant effect on cellular viability at any of the concentrations tested.



Appendix 1.7 NF-κB and NF-AT activation and binding to the HERV-K (HML-2) promoter in response to PMA and Ionomycin. Chromatin immunoprecipitation was performed on the cellular HERV-K (HML-2) LTR promoter in Jurkat T cells using antibodies specific for NF-κB or NF-AT (or non-specific IgG isotype controls) 1 hour after treatment with PMA and Ionomycin, as a positive control. qPCR was performed using primer sets designed to target areas of the promoter where the potential transcription factor binding sites are present. Data are expressed as fold enrichment over IgG. Bars are color-coded to match primer binding areas, as shown in Figure 2.6. Error bars indicate SEM from three independent immunoprecipitation experiments.

Bibliography

1. Gifford, R., et al., *Evolution and distribution of class II-related endogenous retroviruses.* J Virol, 2005. **79**(10): p. 6478-86.
2. Gifford, R. and M. Tristem, *The evolution, distribution and diversity of endogenous retroviruses.* Virus Genes, 2003. **26**(3): p. 291-315.
3. Baltimore, D., *Tumor viruses: 1974.* Cold Spring Harb Symp Quant Biol, 1975. **39 Pt 2**: p. 1187-200.
4. Weiss, R.A., *The discovery of endogenous retroviruses.* Retrovirology, 2006. **3**: p. 67.
5. Vogt, P.K., *Historical Introduction to the General Properties of Retroviruses.* 1997.
6. *Peyton Rous: father of the tumor virus.* J Exp Med, 2005. **201**(3): p. 320.
7. Gross, L., *Oncogenic Viruses*. Third ed. 1983: Pergamon. 1203.
8. Crawford, L.V. and E.M. Crawford, *The properties of Rous sarcoma virus purified by density gradient centrifugation.* Virology, 1961. **13**: p. 227-32.
9. Temin, H.M., *Separation of morphological conversion and virus production in Rous sarcoma virus infection.* Cold Spring Harb Symp Quant Biol, 1962. **27**: p. 407-14.
10. Temin, H.M., *Nature of the provirus of Rous sarcoma.* Natl Cancer Inst Monogr, 1964. **17**: p. 557-570.
11. Temin, H.M., *The DNA provirus hypothesis.* Science, 1976. **192**(4244): p. 1075-80.
12. Lwoff, A., *Tumor viruses and the cancer problem: a summation of the conference.* Cancer Res, 1960. **20**: p. 820-9.
13. Sambrook, J., et al., *The integrated state of viral DNA in SV40-transformed cells.* Proc Natl Acad Sci U S A, 1968. **60**(4): p. 1288-95.
14. Temin, H.M. and S. Mizutani, *RNA-dependent DNA polymerase in virions of Rous sarcoma virus.* Nature, 1970. **226**(5252): p. 1211-3.
15. Baltimore, D., *RNA-dependent DNA polymerase in virions of RNA tumour viruses.* Nature, 1970. **226**(5252): p. 1209-11.
16. Poiesz, B.J., et al., *Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma.* Proc Natl Acad Sci U S A, 1980. **77**(12): p. 7415-9.
17. Gallo, R.C., et al., *Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 865-7.
18. Barre-Sinoussi, F., et al., *Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS).* Science, 1983. **220**(4599): p. 868-71.
19. Coffin, J.M., *Structure and classification of retroviruses. In The Retroviridae.* Retroviruses

ed. J.A. Levy. 1992: Plenum Press NY.
20. Coffin, J.M. and Cold Spring Harbor Laboratory. Press., *Retroviruses*. 2002, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.
21. Temin, H.M., *Origin and general nature of retroviruses. In The Retroviridae.* Retroviruses, ed. J.A. Levy. 1992: Plenum Press NY.

22. Symer D.E., B., J.D., *Retroviruses : molecular biology, genomics, and pathogenesis*, ed. R. Kurth and N. Bannert. 2010, Norfolk, UK: Caister Academic Press. xviii, 454 p.
23. Bannert, N., U. Fiebig, and O. Hohn, *Retroviruses: molecular biology, genomics and pathogenesis*, ed. R. Kurth and N. Bannert, Norfolk, UK: Caister Academic Press. xviii, 454 p.
24. Coffin, J.M., S.H. Hughes, and H. Varmus, *Retroviruses*. 1997, Plainview, N.Y.: Cold Spring Harbor Laboratory Press. xv, 843 p.
25. Wagner, E.K., et al., *Basic Virology*. 3rd ed. 2008, Malden, MA: Blackwell Pub. xxx, 550 p.
26. Urnovitz, H.B. and W.H. Murphy, *Human endogenous retroviruses: nature, occurrence, and clinical implications in human disease.* Clin Microbiol Rev, 1996. **9**(1): p. 72-99.
27. Maniatis, T., S. Goodbourn, and J.A. Fischer, *Regulation of inducible and tissue-specific gene expression.* Science, 1987. **236**(4806): p. 1237-45.
28. Bandziulis, R.J., M.S. Swanson, and G. Dreyfuss, *RNA-binding proteins as developmental regulators.* Genes Dev, 1989. **3**(4): p. 431-7.
29. Fu, W., R.J. Gorelick, and A. Rein, *Characterization of human immunodeficiency virus type 1 dimeric RNA from wild-type and protease-defective virions.* J Virol, 1994. **68**(8): p. 5013-8.
30. Xiong, Y. and T.H. Eickbush, *Origin and evolution of retroelements based upon their reverse transcriptase sequences.* EMBO J, 1990. **9**(10): p. 3353-62.
31. Trono, D., *Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses.* J Virol, 1992. **66**(8): p. 4893-900.
32. Dornadula, G., et al., *Natural endogenous reverse transcription of simian immunodeficiency virus.* Virology, 1997. **227**(1): p. 260-7.
33. Zhang, H., et al., *Intravirion reverse transcripts in the peripheral blood plasma on human immunodeficiency virus type 1-infected individuals.* J Virol, 1994. **68**(11): p. 7591-7.
34. Yu, S.F., M.D. Sullivan, and M.L. Linial, *Evidence that the human foamy virus genome is DNA.* J Virol, 1999. **73**(2): p. 1565-72.
35. Delelis, O., A. Saib, and P. Sonigo, *Biphasic DNA synthesis in spumaviruses.* J Virol, 2003. **77**(14): p. 8141-6.
36. Jacks, T., et al., *Characterization of ribosomal frameshifting in HIV-1 gag-pol expression.* Nature, 1988. **331**(6153): p. 280-3.
37. Dulude, D., M. Baril, and L. Brakier-Gingras, *Characterization of the frameshift stimulatory signal controlling a programmed -1 ribosomal frameshift in the human immunodeficiency virus type 1.* Nucleic Acids Res, 2002. **30**(23): p. 5094-102.
38. Farmerie, W.G., et al., *Expression and processing of the AIDS virus reverse transcriptase in Escherichia coli.* Science, 1987. **236**(4799): p. 305-8.
39. S.P., G., *Virology*. Fifth Edition ed. Retroviridae: the retroviruses and their replication, ed. B.N. Fields, D.M. Knipe, and P.M. Howley. 2006: Lippincott Williams & Wilkins.
40. Zsiros, J., et al., *Biased nucleotide composition of the genome of HERV-K related endogenous retroviruses and its evolutionary implications.* J Mol Evol, 1999. **48**(1): p. 102-11.
41. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
42. Bannert, N. and R. Kurth, *The evolutionary dynamics of human endogenous retroviral families.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 149-73.

43. Subramanian, R.P., et al., *Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses.* Retrovirology, 2011. **8**: p. 90.
44. Temin, H.M., *Origin of retroviruses from cellular moveable genetic elements.* Cell, 1980. **21**(3): p. 599-600.
45. Feschotte, C. and C. Gilbert, *Endogenous viruses: insights into viral evolution and impact on host biology.* Nat Rev Genet, 2012. **13**(4): p. 283-96.
46. Jern, P. and J.M. Coffin, *Effects of retroviruses on host genome function.* Annu Rev Genet, 2008. **42**: p. 709-32.
47. Martins, H. and P. Villesen, *Improved integration time estimation of endogenous retroviruses with phylogenetic data.* PLoS One, 2011. **6**(3): p. e14745.
48. Villesen, P., et al., *Identification of endogenous retroviral reading frames in the human genome.* Retrovirology, 2004. **1**: p. 32.
49. Herniou, E., et al., *Retroviral diversity and distribution in vertebrates.* J Virol, 1998. **72**(7): p. 5955-66.
50. Smith, J.M. and J. Haigh, *The hitch-hiking effect of a favourable gene.* Genet Res, 2007. **89**(5-6): p. 391-403.
51. Kjellman, C., H.O. Sjogren, and B. Widegren, *The Y chromosome: a graveyard for endogenous retroviruses.* Gene, 1995. **161**(2): p. 163-70.
52. Cantrell, M.A., et al., *MysTR: an endogenous retrovirus family in mammals that is undergoing recent amplifications to unprecedented copy numbers.* J Virol, 2005. **79**(23): p. 14698-707.
53. Belshaw, R., et al., *Long-term reinfection of the human genome by endogenous retroviruses.* Proc Natl Acad Sci U S A, 2004. **101**(14): p. 4894-9.
54. Reus, K., et al., *HERV-K(OLD): ancestor sequences of the human endogenous retrovirus family HERV-K(HML-2).* J Virol, 2001. **75**(19): p. 8917-26.
55. Browning, M.T., et al., *Primate and feline lentivirus vector RNA packaging and propagation by heterologous lentivirus virions.* J Virol, 2001. **75**(11): p. 5129-40.
56. Beasley, B.E. and W.S. Hu, *cis-Acting elements important for retroviral RNA packaging specificity.* J Virol, 2002. **76**(10): p. 4950-60.
57. Mi, S., et al., *Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.* Nature, 2000. **403**(6771): p. 785-9.
58. Andersson, A.C., et al., *Developmental expression of HERV-R (ERV3) and HERV-K in human tissue.* Virology, 2002. **297**(2): p. 220-5.
59. Stoye, J.P., *An intracellular block to primate lentivirus replication.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11549-51.
60. Seifarth, W., et al., *Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray.* J Virol, 2005. **79**(1): p. 341-52.
61. Ogasawara, H., et al., *Possible role of DNA hypomethylation in the induction of SLE: relationship to the transcription of human endogenous retroviruses.* Clin Exp Rheumatol, 2003. **21**(6): p. 733-8.
62. Flockerzi, A., et al., *Expression patterns of transcribed human endogenous retrovirus HERV-K(HML-2) loci in human tissues and the need for a HERV Transcriptome Project.* BMC Genomics, 2008. **9**: p. 354.
63. Stauffer, Y., et al., *Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues.* Cancer Immun, 2004. **4**: p. 2.

64. Boeke, M. and J.P. Stoye, *Retrotrasposons, endogenous retroviruses, and the evolution of retroelements. In: Retroviruses.*, ed. H.S.H. Coffin J.M., Varmus H.E. 1997: Cold Spring Harbor Laboratory Press, New York.
65. Arnaud, F., et al., *Coevolution of endogenous betaretroviruses of sheep and their host.* Cell Mol Life Sci, 2008. **65**(21): p. 3422-32.
66. Salmons, B. and W.H. Gunzburg, *Current perspectives in the biology of mouse mammary tumour virus.* Virus Res, 1987. **8**(2): p. 81-102.
67. Palmarini, M., M. Mura, and T.E. Spencer, *Endogenous betaretroviruses of sheep: teaching new lessons in retroviral interference and adaptation.* J Gen Virol, 2004. **85**(Pt 1): p. 1-13.
68. Tarlinton, R.E., J. Meers, and P.R. Young, *Retroviral invasion of the koala genome.* Nature, 2006. **442**(7098): p. 79-81.
69. Gifford, R.J., *Evolution at the host-retrovirus interface.* Bioessays, 2006. **28**(12): p. 1153-6.
70. Katzourakis, A. and M. Tristem, *Phylogeny of human endogenous and exogenous retroviruses.* In: Retroviruses and Primate Genome Evolution, ed. E. Sverdlov. 2005: Landes Bioscience.
71. Jurka, J., *Repbase update: a database and an electronic journal of repetitive elements.* Trends Genet, 2000. **16**(9): p. 418-20.
72. Doolittle, R.F., et al., *Origins and evolutionary relationships of retroviruses.* Q Rev Biol, 1989. **64**(1): p. 1-30.
73. Cohen, M. and E. Larsson, *Human endogenous retroviruses.* Bioessays, 1988. **9**(6): p. 191-6.
74. Larsson, E., N. Kato, and M. Cohen, *Human endogenous proviruses.* Curr Top Microbiol Immunol, 1989. **148**: p. 115-32.
75. Tristem, M., *Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database.* J Virol, 2000. **74**(8): p. 3715-30.
76. Blond, J.L., et al., *Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family.* J Virol, 1999. **73**(2): p. 1175-85.
77. Blond, J.L., et al., *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.* J Virol, 2000. **74**(7): p. 3321-9.
78. Lindeskog, M., D.L. Mager, and J. Blomberg, *Isolation of a human endogenous retroviral HERV-H element with an open env reading frame.* Virology, 1999. **258**(2): p. 441-50.
79. Lindeskog, M., et al., *Coamplification and dispersion of adjacent human endogenous retroviral HERV-H and HERV-E elements; presence of spliced hybrid transcripts in normal leukocytes.* Virology, 1998. **244**(1): p. 219-29.
80. Kjellman, C., H.O. Sjogren, and B. Widegren, *HERV-F, a new group of human endogenous retrovirus sequences.* J Gen Virol, 1999. **80 (Pt 9)**: p. 2383-92.
81. May, F.E. and B.R. Westley, *Structure of a human retroviral sequence related to mouse mammary tumor virus.* J Virol, 1986. **60**(2): p. 743-9.
82. Petropoulus, C., *Retroviral taxonomy, protein structures, sequences and genetic maps.* Retroviruses, ed. H.S.H. Coffin J.M., Varmus H.E. 1997, New York: Cold Spring Harbour Laboratory Press.

83. Aravind, L. and E.V. Koonin, *G-patch: a new conserved domain in eukaryotic RNA-processing proteins and type D retroviral polyproteins.* Trends Biochem Sci, 1999. **24**(9): p. 342-4.
84. Ono, M., et al., *Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome.* J Virol, 1986. **60**(2): p. 589-98.
85. Deen, K.C. and R.W. Sweet, *Murine mammary tumor virus pol-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus pol gene.* J Virol, 1986. **57**(2): p. 422-32.
86. Andersson, M.L., et al., *Diversity of human endogenous retrovirus class II-like sequences.* J Gen Virol, 1999. **80 (Pt 1)**: p. 255-60.
87. Franklin, G.C., et al., *Expression of human sequences related to those of mouse mammary tumor virus.* J Virol, 1988. **62**(4): p. 1203-10.
88. Mayer, J. and E.U. Meese, *Presence of dUTPase in the various human endogenous retrovirus K (HERV-K) families.* J Mol Evol, 2003. **57**(6): p. 642-9.
89. Medstrand, P. and J. Blomberg, *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, 1993. **67**(11): p. 6778-87.
90. Belshaw, R., et al., *Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity.* J Virol, 2005. **79**(19): p. 12507-14.
91. Cordonnier, A., J.F. Casella, and T. Heidmann, *Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence.* J Virol, 1995. **69**(9): p. 5890-7.
92. Benit, L., P. Dessen, and T. Heidmann, *Identification, phylogeny, and evolution of retroviral elements based on their envelope genes.* J Virol, 2001. **75**(23): p. 11709-19.
93. Krieg, A.M., et al., *Heterogeneous expression and coordinate regulation of endogenous retroviral sequences in human peripheral blood mononuclear cells.* AIDS Res Hum Retroviruses, 1992. **8**(12): p. 1991-8.
94. Medstrand, P., M. Lindeskog, and J. Blomberg, *Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals.* J Gen Virol, 1992. **73 (Pt 9)**: p. 2463-6.
95. Brodsky, I., et al., *Expression of HERV-K proviruses in human leukocytes.* Blood, 1993. **81**(9): p. 2369-74.
96. Tonjes, R.R., et al., *HERV-K: the biologically most active human endogenous retrovirus family.* J Acquir Immune Defic Syndr Hum Retrovirol, 1996. **13 Suppl 1**: p. S261-7.
97. Blomberg J, M.P., Yin H, Andersson ML, Lindeskog M, Borg A, Olsson H, *Expression of human endogenous retroviral sequences: differences between individuals and cell types. Increased expression in a human breast cancer.* J Cancer Res Clin Oncol 1995. **121**((Suppl 1)): p. 3.
98. Turner, G., et al., *Insertional polymorphisms of full-length endogenous retroviruses in humans.* Curr Biol, 2001. **11**(19): p. 1531-5.
99. Moyes, D.L., et al., *The distribution of the endogenous retroviruses HERV-K113 and HERV-K115 in health and disease.* Genomics, 2005. **86**(3): p. 337-41.
100. Gourley, M.F., et al., *Role of endogenous retroviruses in autoimmune diseases.* Tohoku J Exp Med, 1994. **173**(1): p. 105-14.

101. Bieda, K., A. Hoffmann, and K. Boller, *Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines.* J Gen Virol, 2001. **82**(Pt 3): p. 591-6.
102. Okahara, G., et al., *Expression analyses of human endogenous retroviruses (HERVs): tissue-specific and developmental stage-dependent expression of HERVs.* Genomics, 2004. **84**(6): p. 982-90.
103. La Mantia, G., et al., *Identification and characterization of novel human endogenous retroviral sequences prefentially expressed in undifferentiated embryonal carcinoma cells.* Nucleic Acids Res, 1991. **19**(7): p. 1513-20.
104. Bieche, I., et al., *Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element.* Biol Reprod, 2003. **68**(4): p. 1422-9.
105. Dunn, C.A., P. Medstrand, and D.L. Mager, *An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12841-6.
106. Medstrand, P., J.R. Landry, and D.L. Mager, *Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans.* J Biol Chem, 2001. **276**(3): p. 1896-903.
107. Ting, C.N., et al., *Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene.* Genes Dev, 1992. **6**(8): p. 1457-65.
108. Mare, L. and M. Trinchera, *Comparative analysis of retroviral and native promoters driving expression of beta1,3-galactosyltransferase beta3Gal-T5 in human and mouse tissues.* J Biol Chem, 2007. **282**(1): p. 49-57.
109. Dunlap, K.A., et al., *Endogenous retroviruses regulate periimplantation placental growth and differentiation.* Proc Natl Acad Sci U S A, 2006. **103**(39): p. 14390-5.
110. Mallet, F., et al., *The endogenous retroviral locus ERVWE1 is a bona fide gene involved in hominoid placental physiology.* Proc Natl Acad Sci U S A, 2004. **101**(6): p. 1731-6.
111. Frendo, J.L., et al., *Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation.* Mol Cell Biol, 2003. **23**(10): p. 3566-74.
112. Garrison, K.E., et al., *T cell responses to human endogenous retroviruses in HIV-1 infection.* PLoS Pathog, 2007. **3**(11): p. e165.
113. Hsiao, F.C., et al., *Cutting edge: Epstein-Barr virus transactivates the HERV-K18 superantigen by docking to the human complement receptor 2 (CD21) on primary B cells.* J Immunol, 2006. **177**(4): p. 2056-60.
114. Tai, A.K., et al., *HHV-6A infection induces expression of HERV-K18-encoded superantigen.* J Clin Virol, 2009. **46**(1): p. 47-8.
115. Turcanova, V.L., B. Bundgaard, and P. Hollsberg, *Human herpesvirus-6B induces expression of the human endogenous retrovirus K18-encoded superantigen.* J Clin Virol, 2009. **46**(1): p. 15-9.
116. Contreras-Galindo, R., et al., *HIV-1 infection increases the expression of human endogenous retroviruses type K (HERV-K) in vitro.* AIDS Res Hum Retroviruses, 2007. **23**(1): p. 116-22.
117. Serafino, A., et al., *The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation.* Exp Cell Res, 2009. **315**(5): p. 849-62.
118. Kwun, H.J., et al., *Transactivation of the human endogenous retrovirus K long terminal repeat by herpes simplex virus type 1 immediate early protein 0.* Virus Res, 2002. **86**(1-2): p. 93-100.

119. Chu, W.M., et al., *Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR.* Mol Cell Biol, 1998. **18**(1): p. 58-68.
120. Jang, K.L., M.K. Collins, and D.S. Latchman, *The human immunodeficiency virus tat protein increases the transcription of human Alu repeated sequences by increasing the activity of the cellular transcription factor TFIIIC.* J Acquir Immune Defic Syndr, 1992. **5**(11): p. 1142-7.
121. Stevens, R.W., et al., *Antibody to human endogenous retrovirus peptide in urine of human immunodeficiency virus type 1-positive patients.* Clin Diagn Lab Immunol, 1999. **6**(6): p. 783-6.
122. Katsumata, K., et al., *Cytokine regulation of env gene expression of human endogenous retrovirus-R in human vascular endothelial cells.* Clin Immunol, 1999. **93**(1): p. 75-80.
123. Ono, M., M. Kawakami, and H. Ushikubo, *Stimulation of expression of the human endogenous retrovirus genome by female steroid hormones in human breast cancer cell line T47D.* J Virol, 1987. **61**(6): p. 2059-62.
124. Hohenadl, C., et al., *Transcriptional activation of endogenous retroviral sequences in human epidermal keratinocytes by UVB irradiation.* J Invest Dermatol, 1999. **113**(4): p. 587-94.
125. Khan, A.S., J. Muller, and J.F. Sears, *Early detection of endogenous retroviruses in chemically induced mouse cells.* Virus Res, 2001. **79**(1-2): p. 39-45.
126. Reiche, J., G. Pauli, and H. Ellerbrok, *Differential expression of human endogenous retrovirus K transcripts in primary human melanocytes and melanoma cell lines after UV irradiation.* Melanoma Res, 2010. **20**(5): p. 435-40.
127. Evans, L.H., et al., *Mobilization of endogenous retroviruses in mice after infection with an exogenous retrovirus.* J Virol, 2009. **83**(6): p. 2429-35.
128. Takeda, S., et al., *Transcriptional activation of the tobacco retrotransposon Tto1 by wounding and methyl jasmonate.* Plant Mol Biol, 1998. **36**(3): p. 365-76.
129. Toufaily, C., et al., *Activation of LTRs from different human endogenous retrovirus (HERV) families by the HTLV-1 tax protein and T-cell activators.* Viruses, 2011. **3**(11): p. 2146-59.
130. Laderoute, M.P., et al., *The replicative activity of human endogenous retrovirus K102 (HERV-K102) with HIV viremia.* AIDS, 2007. **21**(18): p. 2417-24.
131. Contreras-Galindo, R., et al., *Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals.* J Virol, 2012. **86**(1): p. 262-76.
132. Sjottem, E., S. Anderssen, and T. Johansen, *The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3' to the TATA box.* J Virol, 1996. **70**(1): p. 188-98.
133. Andersson, M.L., et al., *Differential expression of human endogenous retroviral sequences similar to mouse mammary tumor virus in normal peripheral blood mononuclear cells.* AIDS Res Hum Retroviruses, 1996. **12**(9): p. 833-40.
134. Baust, C., et al., *Functional activity of HERV-K-T47D-related long terminal repeats.* Virology, 2001. **283**(2): p. 262-72.
135. Schon, U., et al., *Cell type-specific expression and promoter activity of human endogenous retroviral long terminal repeats.* Virology, 2001. **279**(1): p. 280-91.
136. La Mantia, G., et al., *Identification of regulatory elements within the minimal promoter region of the human endogenous ERV9 proviruses: accurate transcription initiation is controlled by an Inr-like element.* Nucleic Acids Res, 1992. **20**(16): p. 4129-36.

137. Knossl, M., R. Lower, and J. Lower, *Expression of the human endogenous retrovirus HTDV/HERV-K is enhanced by cellular transcription factor YY1.* J Virol, 1999. **73**(2): p. 1254-61.
138. de Parseval, N., H. Alkabbani, and T. Heidmann, *The long terminal repeats of the HERV-H human endogenous retrovirus contain binding sites for transcriptional regulation by the Myb protein.* J Gen Virol, 1999. **80 (Pt 4)**: p. 841-5.
139. Lee, W.J., H.J. Kwun, and K.L. Jang, *Analysis of transcriptional regulatory sequences in the human endogenous retrovirus W long terminal repeat.* J Gen Virol, 2003. **84**(Pt 8): p. 2229-35.
140. Fuchs, N.V., et al., *Expression of the human endogenous retrovirus (HERV) group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by transcription factors Sp1 and Sp3.* J Virol, 2011. **85**(7): p. 3436-48.
141. Medstrand, P., L.N. van de Lagemaat, and D.L. Mager, *Retroelement distributions in the human genome: variations associated with age and proximity to genes.* Genome Res, 2002. **12**(10): p. 1483-95.
142. Bushman, F.D., *Targeting survival: integration site selection by retroviruses and LTR-retrotransposons.* Cell, 2003. **115**(2): p. 135-8.
143. Mack, M., K. Bender, and P.M. Schneider, *Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC class III region.* Immunogenetics, 2004. **56**(5): p. 321-32.
144. Chen, H.J., et al., *A retroviral repetitive element confers tissue-specificity to the human alcohol dehydrogenase 1C (ADH1C) gene.* DNA Cell Biol, 2002. **21**(11): p. 793-801.
145. Landry, J.R., et al., *The Opitz syndrome gene Mid1 is transcribed from a human endogenous retroviral promoter.* Mol Biol Evol, 2002. **19**(11): p. 1934-42.
146. de Parseval, N., et al., *Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins.* J Virol, 2003. **77**(19): p. 10414-22.
147. Muir, A., A. Lever, and A. Moffett, *Expression and functions of human endogenous retroviruses in the placenta: an update.* Placenta, 2004. **25 Suppl A**: p. S16-25.
148. Villarreal, L.P., *On viruses, sex, and motherhood.* J Virol, 1997. **71**(2): p. 859-65.
149. Haraguchi, S., R.A. Good, and N.K. Day, *Immunosuppressive retroviral peptides: cAMP and cytokine patterns.* Immunol Today, 1995. **16**(12): p. 595-603.
150. Meisler, M.H. and C.N. Ting, *The remarkable evolutionary history of the human amylase genes.* Crit Rev Oral Biol Med, 1993. **4**(3-4): p. 503-9.
151. Mangeney, M., et al., *The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties.* J Gen Virol, 2001. **82**(Pt 10): p. 2515-8.
152. Cianciolo, G.J., et al., *Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope proteins.* Science, 1985. **230**(4724): p. 453-5.
153. Lewis, S.M. and G.E. Wu, *The origins of V(D)J recombination.* Cell, 1997. **88**(2): p. 159-62.
154. Agrawal, A., Q.M. Eastman, and D.G. Schatz, *Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system.* Nature, 1998. **394**(6695): p. 744-51.
155. Meylan, F., et al., *Negative thymocyte selection to HERV-K18 superantigens in humans.* Blood, 2005. **105**(11): p. 4377-82.
156. Larsson, E. and G. Andersson, *Beneficial role of human endogenous retroviruses: facts and hypotheses.* Scand J Immunol, 1998. **48**(4): p. 329-38.

157. Ponferrada, V.G., B.S. Mauck, and D.P. Wooley, *The envelope glycoprotein of human endogenous retrovirus HERV-W induces cellular resistance to spleen necrosis virus.* Arch Virol, 2003. **148**(4): p. 659-75.
158. Liu, A.Y. and B.A. Abraham, *Subtractive cloning of a hybrid human endogenous retrovirus and calbindin gene in the prostate cell line PC3.* Cancer Res, 1991. **51**(15): p. 4107-10.
159. Tomlins, S.A., et al., *Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer.* Nature, 2007. **448**(7153): p. 595-9.
160. Katoh, I., et al., *Activation of the long terminal repeat of human endogenous retrovirus K by melanoma-specific transcription factor MITF-M.* Neoplasia, 2011. **13**(11): p. 1081-92.
161. Marguerat, S., et al., *Association of human endogenous retrovirus K-18 polymorphisms with type 1 diabetes.* Diabetes, 2004. **53**(3): p. 852-4.
162. Brodziak, A., et al., *The role of human endogenous retroviruses in the pathogenesis of autoimmune diseases.* Med Sci Monit, 2012. **18**(6): p. RA80-88.
163. Martinelli-Boneschi, F., et al., *A genome-wide association study in progressive multiple sclerosis.* Mult Scler, 2012.
164. Perron, H., et al., *Human endogenous retrovirus type W envelope expression in blood and brain cells provides new insights into multiple sclerosis disease.* Mult Scler, 2012.
165. Antony, J.M., et al., *Human endogenous retroviruses and multiple sclerosis: Innocent bystanders or disease determinants?* Biochim Biophys Acta, 2010.
166. Christensen, T., et al., *Expression of sequence variants of endogenous retrovirus RGH in particle form in multiple sclerosis.* Lancet, 1998. **352**(9133): p. 1033.
167. Christensen, T., et al., *Molecular characterization of HERV-H variants associated with multiple sclerosis.* Acta Neurol Scand, 2000. **101**(4): p. 229-38.
168. Herrmann, M., et al., *Retrovirus-associated rheumatic syndromes.* Curr Opin Rheumatol, 1998. **10**(4): p. 347-54.
169. Reynier, F., et al., *Increase in human endogenous retrovirus HERV-K (HML-2) viral load in active rheumatoid arthritis.* Scand J Immunol, 2009. **70**(3): p. 295-9.
170. Freimanis, G., et al., *A role for human endogenous retrovirus-K (HML-2) in rheumatoid arthritis: investigating mechanisms of pathogenesis.* Clin Exp Immunol, 2010. **160**(3): p. 340-7.
171. Ejtehadi, H.D., et al., *The potential role of human endogenous retrovirus K10 in the pathogenesis of rheumatoid arthritis: a preliminary study.* Ann Rheum Dis, 2006. **65**(5): p. 612-6.
172. Nakagawa, K., et al., *Direct evidence for the expression of multiple endogenous retroviruses in the synovial compartment in rheumatoid arthritis.* Arthritis Rheum, 1997. **40**(4): p. 627-38.
173. Takeuchi, K., et al., *Expression of endogenous retroviruses, ERV3 and lambda 4-1, in synovial tissues from patients with rheumatoid arthritis.* Clin Exp Immunol, 1995. **99**(3): p. 338-44.
174. Moles, J.P., A. Tesniere, and J.J. Guilhou, *A new endogenous retroviral sequence is expressed in skin of patients with psoriasis.* Br J Dermatol, 2005. **153**(1): p. 83-9.
175. Bessis, D., et al., *Differential expression of a human endogenous retrovirus E transmembrane envelope glycoprotein in normal, psoriatic and atopic dermatitis human skin.* Br J Dermatol, 2004. **151**(4): p. 737-45.
176. Lai, O.Y., et al., *Protective Effect of Human Endogenous Retrovirus K dUTPase Variants on Psoriasis Susceptibility.* J Invest Dermatol, 2012.

177. Adelman, M.K. and J.J. Marchalonis, *Endogenous retroviruses in systemic lupus erythematosus: candidate lupus viruses.* Clin Immunol, 2002. **102**(2): p. 107-16.
178. Pullmann, R., Jr., et al., *Haplotypes of the HRES-1 endogenous retrovirus are associated with development and disease manifestations of systemic lupus erythematosus.* Arthritis Rheum, 2008. **58**(2): p. 532-40.
179. Posnett, D.N. and A.A. Yarilina, *Sleeping with the enemy--endogenous superantigens in humans.* Immunity, 2001. **15**(4): p. 503-6.
180. Stauffer, Y., et al., *Interferon-alpha-induced endogenous superantigen. a model linking environment and autoimmunity.* Immunity, 2001. **15**(4): p. 591-601.
181. Sutkowski, N., et al., *Epstein-Barr virus transactivates the human endogenous retrovirus HERV-K18 that encodes a superantigen.* Immunity, 2001. **15**(4): p. 579-89.
182. Seidl, C., et al., *An endogenous retroviral long terminal repeat at the HLA-DQB1 gene locus confers susceptibility to rheumatoid arthritis.* Hum Immunol, 1999. **60**(1): p. 63-8.
183. Weis, S., et al., *Reduced expression of human endogenous retrovirus (HERV)-W GAG protein in the cingulate gyrus and hippocampus in schizophrenia, bipolar disorder, and depression.* J Neural Transm, 2007. **114**(5): p. 645-55.
184. Yolken, R.H., et al., *Endogenous retroviruses and schizophrenia.* Brain Res Brain Res Rev, 2000. **31**(2-3): p. 193-9.
185. Huang, W.J., et al., *Human endogenous retroviral pol RNA and protein detected and identified in the blood of individuals with schizophrenia.* Schizophr Res, 2006. **83**(2-3): p. 193-9.
186. Frank, O., et al., *Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders.* J Virol, 2005. **79**(17): p. 10890-901.
187. Oluwole, S.O., et al., *Elevated levels of transcripts encoding a human retroviral envelope protein (syncytin) in muscles from patients with motor neuron disease.* Amyotroph Lateral Scler, 2007. **8**(2): p. 67-72.
188. Contreras-Galindo, R., et al., *Comparative longitudinal studies of HERV-K and HIV-1 RNA titers in HIV-1-infected patients receiving successful versus unsuccessful highly active antiretroviral therapy.* AIDS Res Hum Retroviruses, 2007. **23**(9): p. 1083-6.
189. Contreras-Galindo, R., et al., *A new Real-Time-RT-PCR for quantitation of human endogenous retroviruses type K (HERV-K) RNA load in plasma samples: increased HERV-K RNA titers in HIV-1 patients with HAART non-suppressive regimens.* J Virol Methods, 2006. **136**(1-2): p. 51-7.
190. Contreras-Galindo, R., et al., *Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer.* J Virol, 2008. **82**(19): p. 9329-36.
191. Contreras-Galindo, R., et al., *Detection of HERV-K(HML-2) viral RNA in plasma of HIV type 1-infected individuals.* AIDS Res Hum Retroviruses, 2006. **22**(10): p. 979-84.
192. van der Kuyl, A.C., *HIV infection and HERV expression: a review.* Retrovirology, 2012. **9**(1): p. 6.
193. Wang-Johanning, F., et al., *Detecting the expression of human endogenous retrovirus E envelope transcripts in human prostate adenocarcinoma.* Cancer, 2003. **98**(1): p. 187-97.
194. Strick, R., et al., *Proliferation and cell-cell fusion of endometrial carcinoma are induced by the human endogenous retroviral Syncytin-1 and regulated by TGF-beta.* J Mol Med (Berl), 2007. **85**(1): p. 23-38.
195. Bjerregaard, B., et al., *Syncytin is involved in breast cancer-endothelial cell fusions.* Cell Mol Life Sci, 2006. **63**(16): p. 1906-11.

196. Wang-Johanning, F., et al., *Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer.* Int J Cancer, 2007. **120**(1): p. 81-90.
197. Hu, L., et al., *Expression of human endogenous gammaretroviral sequences in endometriosis and ovarian cancer.* AIDS Res Hum Retroviruses, 2006. **22**(6): p. 551-7.
198. Galli, U.M., et al., *Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors.* Oncogene, 2005. **24**(19): p. 3223-8.
199. Barbulescu, M., et al., *Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans.* Curr Biol, 1999. **9**(16): p. 861-68.
200. Ruda, V.M., et al., *Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR.* Virus Res, 2004. **104**(1): p. 11-6.
201. Seifarth, W., et al., *Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles.* J Virol, 1998. **72**(10): p. 8384-91.
202. Wang-Johanning, F., et al., *Quantitation of HERV-K env gene expression and splicing in human breast cancer.* Oncogene, 2003. **22**(10): p. 1528-35.
203. Wang-Johanning, F., et al., *Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors.* J Natl Cancer Inst, 2012. **104**(3): p. 189-210.
204. Callahan, R., et al., *A new class of endogenous human retroviral genomes.* Science, 1985. **228**(4704): p. 1208-11.
205. Ono, M., *Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes.* J Virol, 1986. **58**(3): p. 937-44.
206. Steinhuber, S., et al., *Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates.* Hum Genet, 1995. **96**(2): p. 188-92.
207. Tönjes RR, K.R., *Comparison of primate cellular sequences homologous to HERV-K reverse transcriptase.* J Cell Biochem, 1994: p. Suppl 18B:41.
208. Moyes, D., D.J. Griffiths, and P.J. Venables, *Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease.* Trends Genet, 2007. **23**(7): p. 326-33.
209. Mayer, J., et al., *An almost-intact human endogenous retrovirus K on human chromosome 7.* Nat Genet, 1999. **21**(3): p. 257-8.
210. Bannert, N. and R. Kurth, *Retroelements and the human genome: new perspectives on an old relation.* Proc Natl Acad Sci U S A, 2004. **101 Suppl 2**: p. 14572-9.
211. Denner, J., Phelps, R. C., Lower, J., Lower, R. & Kurth, R., *ANTIBODY RESPONSE OF PREGNANT WOMEN, TUMOR AND AIDS PATIENTS AGAINST THE HUMAN ENDOGENOUS RETROVIRUS HERV-K.* J. Cancer Res. Clin. Oncol., 1995. **121**(S1): p. 5.
212. Boller, K., et al., *Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles.* J Gen Virol, 2008. **89**(Pt 2): p. 567-72.
213. Buscher, K., et al., *Expression of human endogenous retrovirus K in melanomas and melanoma cell lines.* Cancer Res, 2005. **65**(10): p. 4172-80.
214. Lower, R., et al., *Identification of human endogenous retroviruses with complex mRNA expression and particle formation.* Proc Natl Acad Sci U S A, 1993. **90**(10): p. 4480-4.
215. Muster, T., et al., *An endogenous retrovirus derived from human melanoma cells.* Cancer Res, 2003. **63**(24): p. 8735-41.
216. Morgan, D. and I. Brodsky, *Human endogenous retrovirus (HERV-K) particles in megakaryocytes cultured from essential thrombocythemia peripheral blood stem cells.* Exp Hematol, 2004. **32**(6): p. 520-5.

217. Kalter, S.S., et al., *Brief communication: C-type particles in normal human placentas.* J Natl Cancer Inst, 1973. **50**(4): p. 1081-4.
218. Dirksen, E.R. and J.A. Levy, *Virus-like particles in placentas from normal individuals and patients with systemic lupus erythematosus.* J Natl Cancer Inst, 1977. **59**(4): p. 1187-92.
219. Rasmussen, H.B., et al., *Three allelic forms of the human endogenous retrovirus, ERV3, and their frequencies in multiple sclerosis patients and healthy individuals.* Autoimmunity, 1996. **23**(2): p. 111-7.
220. Li, J.M., et al., *The expression of human endogenous retrovirus-3 in fetal cardiac tissue and antibodies in congenital heart block.* Clin Exp Immunol, 1996. **104**(3): p. 388-93.
221. de Parseval, N., et al., *ERV-3 envelope expression and congenital heart block: what does a physiological knockout teach us.* Autoimmunity, 1999. **30**(2): p. 81-3.
222. Dantec, C.L., et al., *Epigenetics and Sjogren's syndrome.* Curr Pharm Biotechnol, 2012.
223. Hishikawa, T., et al., *Detection of antibodies to a recombinant gag protein derived from human endogenous retrovirus clone 4-1 in autoimmune diseases.* Viral Immunol, 1997. **10**(3): p. 137-47.
224. Strazzullo, M., et al., *Characterization and genomic mapping of chimeric ERV9 endogenous retroviruses-host gene transcripts.* Gene, 1998. **206**(1): p. 77-83.
225. Depil, S., et al., *Expression of a human endogenous retrovirus, HERV-K, in the blood cells of leukemia patients.* Leukemia, 2002. **16**(2): p. 254-9.
226. Iwabuchi, H., et al., *A gene homologous to human endogenous retrovirus overexpressed in childhood acute lymphoblastic leukemia.* Leuk Lymphoma, 2004. **45**(11): p. 2303-6.
227. Wentzensen, N., et al., *Expression of an endogenous retroviral sequence from the HERV-H group in gastrointestinal cancers.* Int J Cancer, 2007. **121**(7): p. 1417-23.
228. Guasch, G., et al., *Endogenous retroviral sequence is fused to FGFR1 kinase in the 8p12 stem-cell myeloproliferative disorder with t(8;19)(p12;q13.3).* Blood, 2003. **101**(1): p. 286-8.
229. Tamura, N., et al., *Alveolar macrophages produce the Env protein of a human endogenous retrovirus, HERV-E 4-1, in a subgroup of interstitial lung diseases.* Am J Respir Cell Mol Biol, 1997. **16**(4): p. 429-37.
230. Li, S.K., et al., *Detection and identification of plasma bacterial and viral elements in HIV/AIDS patients in comparison to healthy adults.* Clin Microbiol Infect, 2011.
231. McIntosh, E.M. and R.H. Haynes, *HIV and human endogenous retroviruses: an hypothesis with therapeutic implications.* Acta Biochim Pol, 1996. **43**(4): p. 583-92.
232. Boller, K., et al., *Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV.* Virology, 1993. **196**(1): p. 349-53.
233. Mueller-Lantzsch, N., et al., *Human endogenous retroviral element K10 (HERV-K10) encodes a full-length gag homologous 73-kDa protein and a functional protease.* AIDS Res Hum Retroviruses, 1993. **9**(4): p. 343-50.
234. Sauter, M., et al., *Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas.* J Virol, 1995. **69**(1): p. 414-21.
235. Kraus, B., et al., *Characterization of the human endogenous retrovirus K Gag protein: identification of protease cleavage sites.* Retrovirology, 2011. **8**: p. 21.
236. George, M., et al., *Identification of the protease cleavage sites in a reconstituted Gag polyprotein of an HERV-K(HML-2) element.* Retrovirology, 2011. **8**: p. 30.
237. Schommer, S., et al., *Characterization of the human endogenous retrovirus K proteinase.* J Gen Virol, 1996. **77 (Pt 2)**: p. 375-9.

238. Dewannieux, M., S. Blaise, and T. Heidmann, *Identification of a functional envelope protein from the HERV-K family of human endogenous retroviruses.* J Virol, 2005. **79**(24): p. 15573-7.
239. Buscher, K., et al., *Expression of the human endogenous retrovirus-K transmembrane envelope, Rec and Np9 proteins in melanomas and melanoma cell lines.* Melanoma Res, 2006. **16**(3): p. 223-34.
240. Armbruester, V., et al., *A novel gene from the human endogenous retrovirus K expressed in transformed cells.* Clin Cancer Res, 2002. **8**(6): p. 1800-7.
241. Boese, A., et al., *Human endogenous retrovirus protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc finger protein.* Oncogene, 2000. **19**(38): p. 4328-36.
242. Boese, A., M. Sauter, and N. Mueller-Lantzsch, *A rev-like NES mediates cytoplasmic localization of HERV-K cORF.* FEBS Lett, 2000. **468**(1): p. 65-7.
243. Lower, R., et al., *Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K.* J Virol, 1995. **69**(1): p. 141-9.
244. Gross, H., et al., *The NP9 protein encoded by the human endogenous retrovirus HERV-K(HML-2) negatively regulates gene activation of the Epstein-Barr virus nuclear antigen 2 (EBNA2).* Int J Cancer, 2010.
245. Kaufmann, S., et al., *Human endogenous retrovirus protein Rec interacts with the testicular zinc-finger protein and androgen receptor.* J Gen Virol, 2010. **91**(Pt 6): p. 1494-502.
246. Denne, M., et al., *Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein.* J Virol, 2007. **81**(11): p. 5607-16.
247. Armbruester, V., et al., *Np9 protein of human endogenous retrovirus K interacts with ligand of numb protein X.* J Virol, 2004. **78**(19): p. 10310-9.
248. Yang, J., et al., *An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein.* Proc Natl Acad Sci U S A, 1999. **96**(23): p. 13404-8.
249. Magin, C., R. Lower, and J. Lower, *cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K.* J Virol, 1999. **73**(11): p. 9496-507.
250. Magin, C., et al., *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, 2000. **274**(1): p. 11-6.
251. Magin-Lachmann, C., et al., *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, 2001. **75**(21): p. 10359-71.
252. Yang, J., et al., *The human endogenous retrovirus K Rev response element coincides with a predicted RNA folding region.* RNA, 2000. **6**(11): p. 1551-64.
253. Sauter, M., et al., *Specificity of antibodies directed against Env protein of human endogenous retroviruses in patients with germ cell tumors.* Cancer Res, 1996. **56**(19): p. 4362-5.
254. Herve, C.A., et al., *Autoantibodies to human endogenous retrovirus-K are frequently detected in health and disease and react with multiple epitopes.* Clin Exp Immunol, 2002. **128**(1): p. 75-82.

255. Löwer, R., J. Löwer, and R. Kurth, *The viruses in all of us:characteristics and biological significance of human endogenous retrovirus sequences.* Proc Natl Acad Sci U S A, 1996. **93**: p. 5177-84.
256. Vogetseder, W., et al., *Antibodies in human sera recognizing a recombinant outer membrane protein encoded by the envelope gene of the human endogenous retrovirus K.* AIDS Res Hum Retroviruses, 1993. **9**(7): p. 687-94.
257. Contreras-Galindo, R.A., et al., *Characterization of Human Endogenous Retroviral Elements in the Blood of HIV-1-Infected Individuals.* J Virol, 2011.
258. Jones, R.B., et al., *Human endogenous retrovirus K(HML-2) Gag- and Env-specific T-cell responses are infrequently detected in HIV-1-infected subjects using standard peptide matrix-based screening.* Clin Vaccine Immunol, 2012. **19**(2): p. 288-92.
259. Tandon, R., et al., *Identification of human endogenous retrovirus-specific T cell responses in vertically HIV-1-infected subjects.* J Virol, 2011. **85**(21): p. 11526-31.
260. Sibata, M., et al., *Human endogenous retroviruses: expression in various organs in vivo and its regulation in vitro.* Leukemia, 1997. **11 Suppl 3**: p. 145-6.
261. Bex, F. and R.B. Gaynor, *Regulation of gene expression by HTLV-I Tax protein.* Methods, 1998. **16**(1): p. 83-94.
262. Fiume, G., et al., *Human immunodeficiency virus-1 Tat activates NF-kappaB via physical interaction with IkappaB-alpha and p65.* Nucleic Acids Res, 2011.
263. Ning, S., J.S. Pagano, and G.N. Barber, *IRF7: activation, regulation, modification and function.* Genes Immun, 2011. **12**(6): p. 399-414.
264. Schwarzer, R. and F. Jundt, *Notch and NF-kappaB signaling pathways in the biology of classical Hodgkin lymphoma.* Curr Mol Med, 2011. **11**(3): p. 236-45.
265. Cullen, B.R., *Regulation of HIV-1 gene expression.* FASEB J, 1991. **5**(10): p. 2361-8.
266. Liu, M. and M.V. Eiden, *Role of human endogenous retroviral long terminal repeats (LTRs) in maintaining the integrity of the human germ line.* Viruses, 2011. **3**(6): p. 901-5.
267. Cohen, C.J., W.M. Lock, and D.L. Mager, *Endogenous retroviral LTRs as promoters for human genes: a critical assessment.* Gene, 2009. **448**(2): p. 105-14.
268. Khodosevich, K., Y. Lebedev, and E. Sverdlov, *Endogenous retroviruses and human evolution.* Comp Funct Genomics, 2002. **3**(6): p. 494-8.
269. Rowe, H.M. and D. Trono, *Dynamic control of endogenous retroviruses during development.* Virology, 2011. **411**(2): p. 273-87.
270. Lavie, L., et al., *CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2).* J Virol, 2005. **79**(2): p. 876-83.
271. Blazkova, J., et al., *CpG methylation controls reactivation of HIV from latency.* PLoS Pathog, 2009. **5**(8): p. e1000554.
272. Griffiths, P.D., *Herpesviruses and AIDS.* J Antimicrob Chemother, 1996. **37 Suppl B**: p. 87-95.
273. Hsiao, F.C., et al., *EBV LMP-2A employs a novel mechanism to transactivate the HERV-K18 superantigen through its ITAM.* Virology, 2009. **385**(1): p. 261-6.
274. Sutkowski, N., et al., *Epstein-Barr virus latent membrane protein LMP-2A is sufficient for transactivation of the human endogenous retrovirus HERV-K18 superantigen.* J Virol, 2004. **78**(14): p. 7852-60.
275. SenGupta, D., et al., *Strong human endogenous retrovirus-specific T cell responses are associated with control of HIV-1 in chronic infection.* J Virol, 2011. **85**(14): p. 6977-85.

276. Armitage, A.E., et al., *Conserved footprints of APOBEC3G on Hypermutated human immunodeficiency virus type 1 and human endogenous retrovirus HERV-K(HML2) sequences.* J Virol, 2008. **82**(17): p. 8743-61.
277. Frankel, A.D. and J.A. Young, *HIV-1: fifteen proteins and an RNA.* Annu Rev Biochem, 1998. **67**: p. 1-25.
278. Scala, G., et al., *The expression of the interleukin 6 gene is induced by the human immunodeficiency virus 1 TAT protein.* J Exp Med, 1994. **179**(3): p. 961-71.
279. Yedavalli, V.S., M. Benkirane, and K.T. Jeang, *Tat and trans-activation-responsive (TAR) RNA-independent induction of HIV-1 long terminal repeat by human and murine cyclin T1 requires Sp1.* J Biol Chem, 2003. **278**(8): p. 6404-10.
280. Romani, B., S. Engelbrecht, and R.H. Glashoff, *Functions of Tat: the versatile protein of human immunodeficiency virus type 1.* J Gen Virol, 2010. **91**(Pt 1): p. 1-12.
281. Tada, H., et al., *Trans-activation of the JC virus late promoter by the tat protein of type 1 human immunodeficiency virus in glial cells.* Proc Natl Acad Sci U S A, 1990. **87**(9): p. 3479-83.
282. Hughes, J.F. and J.M. Coffin, *Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution.* Proc Natl Acad Sci U S A, 2004. **101**(6): p. 1668-72.
283. Nelson, P.N., et al., *Demystified. Human endogenous retroviruses.* Mol Pathol, 2003. **56**(1): p. 11-8.
284. Villarreal, L., *Viruses and host evolution: virus-mediated self identity.* Adv Exp Med Biol, 2012. **738**: p. 185-217.
285. Johnston, J.B., et al., *Monocyte activation and differentiation augment human endogenous retrovirus expression: implications for inflammatory brain diseases.* Ann Neurol, 2001. **50**(4): p. 434-42.
286. Berkhout, B., M. Jebbink, and J. Zsiros, *Identification of an active reverse transcriptase enzyme encoded by a human endogenous HERV-K retrovirus.* J Virol, 1999. **73**(3): p. 2365-75.
287. Zsiros, J., et al., *Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses.* J Gen Virol, 1998. **79 (Pt 1)**: p. 61-70.
288. Ehlhardt, S., et al., *Human endogenous retrovirus HERV-K(HML-2) Rec expression and transcriptional activities in normal and rheumatoid arthritis synovia.* J Rheumatol, 2006. **33**(1): p. 16-23.
289. Sekigawa, I., et al., *Retroviruses and autoimmunity.* Intern Med, 2001. **40**(2): p. 80-6.
290. Sekigawa, I., et al., *A new hypothesis of the possible mechanisms of gender differences in systemic lupus erythematosus.* Clin Exp Rheumatol, 2010. **28**(3): p. 419-23.
291. Schanab, O., et al., *Expression of human endogenous retrovirus K is stimulated by ultraviolet radiation in melanoma.* Pigment Cell Melanoma Res, 2011. **24**(4): p. 656-65.
292. Ishida, T., et al., *Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity.* Cancer Immun, 2008. **8**: p. 15.
293. Ruprecht, K., et al., *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, 2008. **82**(20): p. 10008-16.
294. Caldwell, R.L., K.B. Lane, and V.L. Shepherd, *HIV-1 Tat interaction with cyclin T1 represses mannose receptor and the bone morphogenetic protein receptor-2 transcription.* Arch Biochem Biophys, 2006. **449**(1-2): p. 27-33.

295. Jeang, K.T., H. Xiao, and E.A. Rich, *Multifaceted activities of the HIV-1 transactivator of transcription, Tat.* J Biol Chem, 1999. **274**(41): p. 28837-40.
296. Jones, K.A. and B.M. Peterlin, *Control of RNA initiation and elongation at the HIV-1 promoter.* Annu Rev Biochem, 1994. **63**: p. 717-43.
297. Ensoli, B., et al., *Release, uptake, and effects of extracellular human immunodeficiency virus type 1 Tat protein on cell growth and viral transactivation.* J Virol, 1993. **67**(1): p. 277-87.
298. Ensoli, B., et al., *Tat protein of HIV-1 stimulates growth of cells derived from Kaposi's sarcoma lesions of AIDS patients.* Nature, 1990. **345**(6270): p. 84-6.
299. Yezid, H., et al., *Mechanism for HIV-1 Tat insertion into the endosome membrane.* J Biol Chem, 2009. **284**(34): p. 22736-46.
300. Buonaguro, L., et al., *The human immunodeficiency virus type 1 Tat protein transactivates tumor necrosis factor beta gene expression through a TAR-like structure.* J Virol, 1994. **68**(4): p. 2677-82.
301. Shi, J., et al., *Human immunodeficiency virus type 1 Tat induces B7-H1 expression via ERK/MAPK signaling pathway.* Cell Immunol, 2011. **271**(2): p. 280-5.
302. Leghmari, K., Y. Bennasser, and E. Bahraoui, *HIV-1 Tat protein induces IL-10 production in monocytes by classical and alternative NF-kappaB pathways.* Eur J Cell Biol, 2008. **87**(12): p. 947-62.
303. Leghmari, K., et al., *HIV-1 Tat protein induces TNF-alpha and IL-10 production by human macrophages: differential implication of PKC-betaII and -delta isozymes and MAP kinases ERK1/2 and p38.* Cell Immunol, 2008. **254**(1): p. 46-55.
304. Weissman, J.D., et al., *HIV-1 tat binds TAFII250 and represses TAFII250-dependent transcription of major histocompatibility class I genes.* Proc Natl Acad Sci U S A, 1998. **95**(20): p. 11601-6.
305. Howcroft, T.K., et al., *Repression of MHC class I gene promoter activity by two-exon Tat of HIV.* Science, 1993. **260**(5112): p. 1320-2.
306. Molle, D., et al., *A real-time view of the TAR:Tat:P-TEFb complex at HIV-1 transcription sites.* Retrovirology, 2007. **4**: p. 36.
307. Garber, M.E., et al., *The interaction between HIV-1 Tat and human cyclin T1 requires zinc and a critical cysteine residue that is not conserved in the murine CycT1 protein.* Genes Dev, 1998. **12**(22): p. 3512-27.
308. Das, A.T., A. Harwig, and B. Berkhout, *The HIV-1 Tat protein has a versatile role in activating viral transcription.* J Virol, 2011. **85**(18): p. 9506-16.
309. Dandekar, D.H., K.N. Ganesh, and D. Mitra, *HIV-1 Tat directly binds to NFkappaB enhancer sequence: role in viral and cellular gene expression.* Nucleic Acids Res, 2004. **32**(4): p. 1270-8.
310. Farre, D., et al., *Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN.* Nucleic Acids Res, 2003. **31**(13): p. 3651-3.
311. Messeguer, X., et al., *PROMO: detection of known transcription regulatory elements using species-tailored searches.* Bioinformatics, 2002. **18**(2): p. 333-4.
312. Mingyan, Y., L. Xinyong, and E. De Clercq, *NF-kappaB: the inducible factors of HIV-1 transcription and their inhibitors.* Mini Rev Med Chem, 2009. **9**(1): p. 60-9.
313. Minghetti, L., et al., *Multiple actions of the human immunodeficiency virus type-1 Tat protein on microglial cell functions.* Neurochem Res, 2004. **29**(5): p. 965-78.
314. Kinoshita, S., et al., *The T cell activation factor NF-ATc positively regulates HIV-1 replication and gene expression in T cells.* Immunity, 1997. **6**(3): p. 235-44.

315. Williams, R.D., et al., *Activation domains of transcription factors mediate replication dependent transcription from a minimal HIV-1 promoter.* Nucleic Acids Res, 1996. **24**(4): p. 549-57.
316. Gatignol, A., et al., *Sequential steps in Tat trans-activation of HIV-1 mediated through cellular DNA, RNA, and protein binding factors.* Gene Expr, 1996. **5**(4-5): p. 217-28.
317. Demarchi, F., et al., *Activation of transcription factor NF-kappaB by the Tat protein of human immunodeficiency virus type 1.* J Virol, 1996. **70**(7): p. 4427-37.
318. Sune, C. and M.A. Garcia-Blanco, *Sp1 transcription factor is required for in vitro basal and Tat-activated transcription from the human immunodeficiency virus type 1 long terminal repeat.* J Virol, 1995. **69**(10): p. 6572-6.
319. Vacca, A., et al., *Human immunodeficiency virus type-1 tat enhances interleukin-2 promoter activity through synergism with phorbol ester and calcium-mediated activation of the NF-AT cis-regulatory motif.* Biochem Biophys Res Commun, 1994. **205**(1): p. 467-74.
320. Liu, J., et al., *Specific NF-kappa B subunits act in concert with Tat to stimulate human immunodeficiency virus type 1 transcription.* J Virol, 1992. **66**(6): p. 3883-7.
321. Kobori, M., et al., *Wedelolactone suppresses LPS-induced caspase-11 expression by directly inhibiting the IKK complex.* Cell Death Differ, 2004. **11**(1): p. 123-30.
322. Noguchi, H., et al., *A new cell-permeable peptide allows successful allogeneic islet transplantation in mice.* Nat Med, 2004. **10**(3): p. 305-9.
323. Mameli, G., et al., *Regulation of the syncytin-1 promoter in human astrocytes by multiple sclerosis-related cytokines.* Virology, 2007. **362**(1): p. 120-30.
324. Padow, M., et al., *Analysis of human immunodeficiency virus type 1 containing HERV-K protease.* AIDS Res Hum Retroviruses, 2000. **16**(18): p. 1973-80.
325. Towler, E.M., et al., *Functional characterization of the protease of human endogenous retrovirus, K10: can it complement HIV-1 protease?* Biochemistry, 1998. **37**(49): p. 17137-44.
326. Adachi, A., et al., *Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone.* J Virol, 1986. **59**(2): p. 284-91.
327. Ott, M., et al., *Immune hyperactivation of HIV-1-infected T cells mediated by Tat and the CD28 pathway.* Science, 1997. **275**(5305): p. 1481-5.
328. Lewis, N., et al., *Identification of a cis-acting element in human immunodeficiency virus type 2 (HIV-2) that is responsive to the HIV-1 rev and human T-cell leukemia virus types I and II rex proteins.* J Virol, 1990. **64**(4): p. 1690-7.
329. Raney, A., et al., *Reconstitution and molecular analysis of an active human immunodeficiency virus type 1 Nef/p21-activated kinase 2 complex.* J Virol, 2005. **79**(20): p. 12732-41.
330. Nguyen, K.L., et al., *Codon optimization of the HIV-1 vpu and vif genes stabilizes their mRNA and allows for highly efficient Rev-independent expression.* Virology, 2004. **319**(2): p. 163-75.
331. Schaeffer, E., R. Geleziunas, and W.C. Greene, *Human immunodeficiency virus type 1 Nef functions at the level of virus entry by enhancing cytoplasmic delivery of virions.* J Virol, 2001. **75**(6): p. 2993-3000.
332. Ono, A. and E.O. Freed, *Plasma membrane rafts play a critical role in HIV-1 assembly and release.* Proc Natl Acad Sci U S A, 2001. **98**(24): p. 13925-30.

333. Page, K.A., N.R. Landau, and D.R. Littman, *Construction and use of a human immunodeficiency virus vector for analysis of virus infectivity.* J Virol, 1990. **64**(11): p. 5270-6.
334. Ratner, L., et al., *Complete nucleotide sequences of functional clones of the AIDS virus.* AIDS Res Hum Retroviruses, 1987. **3**(1): p. 57-69.
335. Ratner, L., et al., *Complete nucleotide sequences of functional clones of the virus associated with the acquired immunodeficiency syndrome, HTLV-III/LAV.* Haematol Blood Transfus, 1987. **31**: p. 404-6.
336. Sadaie, M.R., T. Benter, and F. Wong-Staal, *Site-directed mutagenesis of two trans-regulatory genes (tat-III,trs) of HIV-1.* Science, 1988. **239**(4842): p. 910-3.
337. Mann, D.L., et al., *Origin of the HIV-susceptible human CD4+ cell line H9.* AIDS Res Hum Retroviruses, 1989. **5**(3): p. 253-5.
338. Caputo, A., J.G. Sodroski, and W.A. Haseltine, *Constitutive expression of HIV-1 tat protein in human Jurkat T cells using a BK virus vector.* J Acquir Immune Defic Syndr, 1990. **3**(4): p. 372-9.
339. Yang, S., et al., *Generation of retroviral vector for clinical studies using transient transfection.* Hum Gene Ther, 1999. **10**(1): p. 123-32.
340. Arnold, B.A., R.W. Hepler, and P.M. Keller, *One-step fluorescent probe product-enhanced reverse transcriptase assay.* Biotechniques, 1998. **25**(1): p. 98-106.
341. Swanson, M.D., et al., *A lectin isolated from bananas is a potent inhibitor of HIV replication.* J Biol Chem, 2010. **285**(12): p. 8646-55.
342. Bohan, C.A., et al., *Analysis of Tat transactivation of human immunodeficiency virus transcription in vitro.* Gene Expr, 1992. **2**(4): p. 391-407.
343. Dong, Q.G., et al., *The function of multiple IkappaB : NF-kappaB complexes in the resistance of cancer cells to Taxol-induced apoptosis.* Oncogene, 2002. **21**(42): p. 6510-9.
344. Fujioka, S., et al., *Inhibition of constitutive NF-kappa B activity by I kappa B alpha M suppresses tumorigenesis.* Oncogene, 2003. **22**(9): p. 1365-70.
345. Fujioka, S., et al., *Stabilization of p53 is a novel mechanism for proapoptotic function of NF-kappaB.* J Biol Chem, 2004. **279**(26): p. 27549-59.
346. Paces, J., et al., *HERVd: the Human Endogenous RetroViruses Database: update.* Nucleic Acids Res, 2004. **32**(Database issue): p. D50.
347. Sperber, G.O., et al., *Automated recognition of retroviral sequences in genomic data--RetroTector.* Nucleic Acids Res, 2007. **35**(15): p. 4964-76.
348. Dunham, I., et al., *The DNA sequence of human chromosome 22.* Nature, 1999. **402**(6761): p. 489-95.
349. Metzdorf, R., E. Gottert, and N. Blin, *A novel centromeric repetitive DNA from human chromosome 22.* Chromosoma, 1988. **97**(2): p. 154-8.
350. Pollard, T., ed. *Cell Biology*. ed. Saunders. 2007: Philadelphia. pp. 227–230.
351. Pollard, T.D., ed. *Cell Biology.*, ed. Saunders. 2007: Philadelphia. pp. 200–203.
352. Pluta, A.F., et al., *The centromere: hub of chromosomal activities.* Science, 1995. **270**(5242): p. 1591-4.
353. Dalal, Y., *Epigenetic specification of centromeres.* Biochem Cell Biol, 2009. **87**(1): p. 273-82.
354. Chueh, A.C., et al., *Variable and hierarchical size distribution of L1-retroelement-enriched CENP-A clusters within a functional human neocentromere.* Hum Mol Genet, 2005. **14**(1): p. 85-93.

355. Bernad, R., P. Sanchez, and A. Losada, *Epigenetic specification of centromeres by CENP-A.* Exp Cell Res, 2009. **315**(19): p. 3233-41.
356. Easley, R., et al., *Chromatin dynamics associated with HIV-1 Tat-activated transcription.* Biochim Biophys Acta, 2010. **1799**(3-4): p. 275-85.
357. Benkirane, M., et al., *Activation of integrated provirus requires histone acetyltransferase. p300 and P/CAF are coactivators for HIV-1 Tat.* J Biol Chem, 1998. **273**(38): p. 24898-905.
358. Kiernan, R.E., et al., *HIV-1 tat transcriptional activity is regulated by acetylation.* EMBO J, 1999. **18**(21): p. 6106-18.
359. Hottiger, M.O. and G.J. Nabel, *Interaction of human immunodeficiency virus type 1 Tat with the transcriptional coactivators p300 and CREB binding protein.* J Virol, 1998. **72**(10): p. 8252-6.
360. Deng, L., et al., *Acetylation of HIV-1 Tat by CBP/P300 increases transcription of integrated HIV-1 genome and enhances binding to core histones.* Virology, 2000. **277**(2): p. 278-95.
361. Deng, L., et al., *Enhancement of the p300 HAT activity by HIV-1 Tat on chromatin DNA.* Virology, 2001. **289**(2): p. 312-26.
362. Yu, J., et al., *A polycomb repression signature in metastatic prostate cancer predicts cancer outcome.* Cancer Res, 2007. **67**(22): p. 10657-63.
363. J, F., *Confidence limits on phylogenies: an approach using the bootstrap.* Evolution, 1985. **39**: p. 783-791.
364. Hills, D.B., J., *An empirical test of bootstrapping as a method for assesing confidence in phylogenetic analysis.* Syst Biol, 1993. **42**: p. 182-192.
365. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.
366. Martin, J.A. and Z. Wang, *Next-generation transcriptome assembly.* Nat Rev Genet, 2011. **12**(10): p. 671-82.
367. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities.* Nat Rev Genet, 2011. **12**(2): p. 87-98.
368. Marguerat, S. and J. Bahler, *RNA-seq: from technology to biology.* Cell Mol Life Sci, 2010. **67**(4): p. 569-79.
369. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.
370. Chen, G., C. Wang, and T. Shi, *Overview of available methods for diverse RNA-Seq data analyses.* Sci China Life Sci, 2011. **54**(12): p. 1121-8.
371. Chi, S.W., et al., *Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps.* Nature, 2009. **460**(7254): p. 479-86.
372. Oja, M., et al., *Methods for estimating human endogenous retrovirus activities from EST databases.* BMC Bioinformatics, 2007. **8 Suppl 2**: p. S11.
373. Perot, P., et al., *Microarray-Based Sketches of the HERV Transcriptome Landscape.* PLoS One, 2012. **7**(6): p. e40194.
374. Pichon, J.P., et al., *Multiplex degenerate PCR coupled with an oligo sorbent array for human endogenous retrovirus expression profiling.* Nucleic Acids Res, 2006. **34**(6): p. e46.
375. Smallwood, A., et al., *Temporal regulation of the expression of syncytin (HERV-W), maternally imprinted PEG10, and SGCE in human placenta.* Biol Reprod, 2003. **69**(1): p. 286-93.

376. Forsman, A., et al., *Development of broadly targeted human endogenous gammaretroviral pol-based real time PCRs Quantitation of RNA expression in human tissues.* J Virol Methods, 2005. **129**(1): p. 16-30.
377. Muradrasoli, S., et al., *Development of real-time PCRs for detection and quantitation of human MMTV-like (HML) sequences HML expression in human tissues.* J Virol Methods, 2006. **136**(1-2): p. 83-92.
378. Ewing, A.D. and H.H. Kazazian, Jr., *High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes.* Genome Res, 2010. **20**(9): p. 1262-70.
379. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
380. Cullen, B.R., P.T. Lomedico, and G. Ju, *Transcriptional interference in avian retroviruses--implications for the promoter insertion model of leukaemogenesis.* Nature, 1984. **307**(5948): p. 241-5.
381. Coiras, M., et al., *Modifications in the human T cell proteome induced by intracellular HIV-1 Tat protein expression.* Proteomics, 2006. **6 Suppl 1**: p. S63-73.
382. Pocernich, C.B., et al., *Proteomic analysis of oxidatively modified proteins induced by the mitochondrial toxin 3-nitropropionic acid in human astrocytes expressing the HIV protein tat.* Brain Res Mol Brain Res, 2005. **133**(2): p. 299-306.
383. Poggi, A., et al., *NK cell activation by dendritic cells is dependent on LFA-1-mediated induction of calcium-calmodulin kinase II: inhibition by HIV-1 Tat C-terminal domain.* J Immunol, 2002. **168**(1): p. 95-101.
384. Rautonen, N., et al., *HIV-1 Tat induces cytokine synthesis by uninfected mononuclear cells.* AIDS, 1994. **8**(10): p. 1504-6.
385. Roshal, M., Y. Zhu, and V. Planelles, *Apoptosis in AIDS.* Apoptosis, 2001. **6**(1-2): p. 103-16.
386. Li, C.J., et al., *Induction of apoptosis in uninfected lymphocytes by HIV-1 Tat protein.* Science, 1995. **268**(5209): p. 429-31.
387. Cavanagh, M.H., et al., *HTLV-I antisense transcripts initiating in the 3'LTR are alternatively spliced and polyadenylated.* Retrovirology, 2006. **3**: p. 15.
388. Gaudray, G., et al., *The complementary strand of the human T-cell leukemia virus type 1 RNA genome encodes a bZIP transcription factor that down-regulates viral transcription.* J Virol, 2002. **76**(24): p. 12813-22.
389. Lefebvre, G., et al., *Analysis of HIV-1 expression level and sense of transcription by high-throughput sequencing of the infected cell.* J Virol, 2011. **85**(13): p. 6205-11.
390. Rasmussen, M.H., et al., *Antisense transcription in gammaretroviruses as a mechanism of insertional activation of host genes.* J Virol, 2010. **84**(8): p. 3780-8.
391. Trapnell, C., L. Pachter, and S.L. Salzberg, *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics, 2009. **25**(9): p. 1105-11.
392. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.
393. Bashir, R., et al., *Expression of Epstein-Barr virus proteins in primary CNS lymphoma in AIDS patients.* Neurology, 1993. **43**(11): p. 2358-62.
394. Aboulafia, D.M. and A.L. Puswella, *Highly active antiretroviral therapy as the sole treatment for AIDS-related primary central nervous system lymphoma: a case report with implications for treatment.* AIDS Patient Care STDS, 2007. **21**(12): p. 900-7.

395. Knowles, D.M., *Etiology and pathogenesis of AIDS-related non-Hodgkin's lymphoma.* Hematol Oncol Clin North Am, 2003. **17**(3): p. 785-820.
396. Besson, C., et al., *Changes in AIDS-related lymphoma since the era of highly active antiretroviral therapy.* Blood, 2001. **98**(8): p. 2339-44.
397. Biggar, R.J., *AIDS-related cancers in the era of highly active antiretroviral therapy.* Oncology (Williston Park), 2001. **15**(4): p. 439-48; discussion 448-9.
398. Carbone, A., et al., *HIV-associated lymphomas and gamma-herpesviruses.* Blood, 2009. **113**(6): p. 1213-24.
399. Carbone, A., et al., *HIV-associated Hodgkin lymphoma.* Curr Opin HIV AIDS, 2009. **4**(1): p. 3-10.
400. Kaplan, L.D., *HIV-associated lymphoma.* Best Pract Res Clin Haematol, 2012. **25**(1): p. 101-17.
401. Carbone, A., *AIDS-related non-Hodgkin's lymphomas: from pathology and molecular pathogenesis to treatment.* Hum Pathol, 2002. **33**(4): p. 392-404.
402. Gates, A.E. and L.D. Kaplan, *Biology and management of AIDS-associated non-Hodgkin's lymphoma.* Hematol Oncol Clin North Am, 2003. **17**(3): p. 821-41.
403. Wang, X.J., T. Gaasterland, and N.H. Chua, *Genome-wide prediction and identification of cis-natural antisense transcripts in Arabidopsis thaliana.* Genome Biol, 2005. **6**(4): p. R30.
404. Gazon, H., et al., *Human T-cell leukemia virus type 1 (HTLV-1) bZIP factor requires cellular transcription factor JunD to upregulate HTLV-1 antisense transcription from the 3' LTR.* J Virol, 2012.
405. Nishizawa, M., et al., *Regulation of inducible gene expression by natural antisense transcripts.* Front Biosci, 2012. **17**: p. 938-58.
406. Li, Y.Y., et al., *In silico discovery of human natural antisense transcripts.* BMC Bioinformatics, 2006. **7**: p. 18.
407. Zhang, Y., et al., *Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species.* Nucleic Acids Res, 2006. **34**(12): p. 3465-75.
408. Maruyama, R., et al., *Altered antisense-to-sense transcript ratios in breast cancer.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2820-4.
409. Yu, W., et al., *Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA.* Nature, 2008. **451**(7175): p. 202-6.
410. Zhao, T. and M. Matsuoka, *HBZ and its roles in HTLV-1 oncogenesis.* Front Microbiol, 2012. **3**: p. 247.
411. Morrissy, A.S., et al., *Next-generation tag sequencing for cancer gene expression profiling.* Genome Res, 2009. **19**(10): p. 1825-35.
412. Numata, K. and H. Kiyosawa, *Genome-wide impact of endogenous antisense transcripts in eukaryotes.* Front Biosci, 2012. **17**: p. 300-15.
413. Rhee, I., et al., *DNMT1 and DNMT3b cooperate to silence genes in human cancer cells.* Nature, 2002. **416**(6880): p. 552-6.
414. Thrash-Bingham, C.A. and K.D. Tartof, *aHIF: a natural antisense transcript overexpressed in human renal cancer and during hypoxia.* J Natl Cancer Inst, 1999. **91**(2): p. 143-51.
415. Morris, K.V., et al., *Bidirectional transcription directs both transcriptional gene activation and suppression in human cells.* PLoS Genet, 2008. **4**(11): p. e1000258.
416. Gogvadze, E., et al., *Human-specific modulation of transcriptional activity provided by endogenous retroviral insertions.* J Virol, 2009. **83**(12): p. 6098-105.

417. Armbruester, V., et al., *A novel gene from the human endogenous retrovirus K expressed in transformed cells.* Clin Cancer Res, 2002. **8**(6): p. 1800-07.
418. Gross, H., et al., *The NP9 protein encoded by the human endogenous retrovirus HERV-K(HML-2) negatively regulates gene activation of the Epstein-Barr virus nuclear antigen 2 (EBNA2).* Int J Cancer, 2011. **129**(5): p. 1105-15.
419. Hanke, K., et al., *The Rec protein of HERV-K(HML-2) upregulates androgen receptor activity by binding to the human small glutamine-rich tetratricopeptide repeat protein (hSGT).* Int J Cancer, 2012.
420. Langner, J.S., et al., *Biochemical analysis of the complex between the tetrameric export adapter protein Rec of HERV-K/HML-2 and the responsive RNA element RcRE pck30.* J Virol, 2012.
421. Han, Y. and R.F. Siliciano, *Keeping quiet: microRNAs in HIV-1 latency.* Nat Med, 2007. **13**(10): p. 1138-40.
422. Huang, J., et al., *Cellular microRNAs contribute to HIV-1 latency in resting primary CD4+ T lymphocytes.* Nat Med, 2007. **13**(10): p. 1241-7.
423. Kumar, S., et al., *MicroRNA signature of cis-platin resistant vs. cis-platin sensitive ovarian cancer cell lines.* J Ovarian Res, 2011. **4**(1): p. 17.