The Endogenous Retroviral Transcriptome in Melanoma

A thesis submitted by

Farrah Roy

in partial fulfillment of the requirements for the degree of

PhD

in

Genetics

Tufts University

May 2018

Advisor: John M. Coffin, PhD

Abstract

Human endogenous retroviruses (HERVs) are retroviral sequences passed from parent to offspring in a Mendelian fashion. Due to relatively recent endogenization and correlation to human illness, interest in the HERV-K (HML-2) subclade has increased. Expression of HML-2 - specific transcripts and proteins in melanoma cell lines was first discovered about a decade ago. Since there are at least 96 known HML-2 proviruses within the human genome, others have studied the HML-2 transcriptome in melanoma to identify the proviral sequences responsible for such expression. More recently, cDNA cloning frequency was used to detect HML-2 expression in melanoma cell lines and patient samples. Though they were able to detect the expression of several proviruses, cDNA cloning frequency is not as sensitive as RNA-sequencing and therefore they could have missed the expression of poorly expressed proviruses or the expression of rare proviruses. Furthermore, determining transcriptional mechanisms responsible for HML-2 expression in melanoma using cDNA cloning frequency data would be difficult while one could easily do this using a stranded library. Due to the potential role of expressed rare HML-2 proviruses in cancer etiology, and to understand transcriptional mechanisms responsible for HML-2 expression in melanoma, I submitted cellular RNA from five melanoma cell lines and three primary melanocyte populations for RNA-seq. I detected five proviruses that were uniquely expressed in melanoma compared to primary melanocytes. 7q22.2 was the highest expressing provirus followed by 3q12.3, which based on comparison to breast cancer and Tera-1s was only detected in cancerous cell lines. Most proviruses that were expressed in melanoma were young, human specific, and were mostly located in extragenic regions and driven by sense transcription. Two proviruses - 7p22.1a and 7p22.1b - contain intact ORFs for Env, yet I was unable to detect the presence of Env in whole cell lysate. Furthermore, three proviruses were

capable of LTR-driven transcription in some of my melanoma cell lines which appears to be partially driven by transcription factor binding to LTRs. In conclusion, this work expands upon the known HML-2 transcriptome in melanoma while offering a larger view on HML-2 expression in cancer. It also shows one potential cause of HML-2 expression in melanoma and identifies the potential proviruses responsible for previous HML-2 protein production. Future work should include analyzing the effect epigenetic regulation has on HML-2 expression in cancer. Furthermore, future work should focus on analyzing patient samples to identify unique proviruses expressed in cancer so HML-2 expression can be evaluated for potential therapeutic and diagnosis purposes.

Acknowledgements

There are so many that deserve to be acknowledged, yet I'm afraid that I'd end up writing another dissertation on how grateful I am to my chosen family and committee. Blame a religious upbringing on that: you can leave the church, but you can never shake the feeling that you would be nowhere without the grace of others.

I doubt that anyone except the acknowledged would want to read that, so I won't.

When I came to Tufts, I did not expect to travel the path I did to end up here – so it goes. I did not expect John Coffin to be my thesis advisor, yet I'm grateful for it. He's someone who would allow me to learn at my own pace and is available when I needed him. Someone who wants me to succeed yet allows a valley of room for me to learn from failures (and believe me, there were lots of those). Someone who allows my project to go in the path that I believe it should, and someone who gave me a pretty great lab family. John, I never properly thanked you for taking me in. So, thank you.

Speaking of my lab family – Neeru Bhardwaj, Zachary Williams, Meagan Montesion, John Yoon, Joseph Holloway, and – yeah, I guess you too – Michael Freeman -, I'm so happy that I got the chance to work with you all. You are a wonderful group of talented and smart scientists who – thankfully – love food as much as you do your work. I'm going to miss our 2-hour coffee breaks.

Thank you to my Walpole chosen family, who have supported me throughout this process and have been good friends to my better half. I spent a lot of time away from him throughout grad school, so thank you for being good friends to him. Thank you to the members of the Monro lab - especially Natasha Durham, Uri Bulow, Benjamin Brigham, and Ramesh Govindan – for being an extension of my lab family. For all the shenanigans we've gotten into, and all the shenanigans we'll get into on future camping trips. Finally, thank you to my Pathfinder family – Missi Nespolo, Lindsey McColl, Jenna Ellis, Zachary Jacobson, and David Roy – for providing hours of entertainment and

laughter while allowing me as your Dungeon Master to torment you with another difficult monster with three attack turns that each stack poison damage. I wish I could say I'm sorry about that time I almost killed your characters, but I don't like lying.

Special acknowledgements must be given to a few people who have mutually tolerated my nonsense and have been with me the longest. To Missi Nespolo, my oldest and dearest friend who welcomed me into her family when I needed one. For introducing me to nerdy things (anime, manga, high fantasy, video games, and comic books) and for being my tattoo buddy. I can confidently say that without you I would not be the person I am today. To Kristen Kotewitz, for being a confident and supportive lab neighbor who was always willing to help me troubleshoot, talk about Game of Thrones, and offer me whiskey when I wanted to celebrate or needed someone to commiserate. To Anne Weeks, one of the most supportive and loving people I've ever met who is not so secretly the Anne Perkins to my Leslie Knope. For her dad jokes, endless hours of tea and coffee, and being a top-notch cheerleader who kept assuring me that – yes – one day I would defend. To Meagan Montesion, for teaching me all the bioinformatics I know and for occasionally forcing me to put down the pipette to go have fun with her. Your constantly positive attitude kept me motivated on the hard days, and your encouragement helped me get through the final months of grad school. To Zachary Williams, one of the three people who gets acknowledged twice in this section. For being a supportive baymate, for always being ready to answer the questions I have no matter how crazy, and for sharing my love of all things weird and wonderful in this world. I couldn't have asked for a better lab brother.

Special thanks should also go to Gertrude Lambert, John McCollough, and Jeanne Nespolo. You all provided a home when I needed one, a family when I wanted one, and role models when I went looking for one. I wish you all could have been here to see this day.

She'll never be able to read this, but special thanks should also go to my dog, Minnie. It may seem silly to some to acknowledge a pet, but without her I'm not sure I'd have had the emotional fortitude to make it through graduate school. She is there for me through depressed periods, moments of extreme anxiety, and 4 am writing sessions asking nothing more than some of my attention and peanut butter. Minnie, I can confirm that you're the good girl.

To my husband, David Roy. You're the last in this section, but the most important. My better half and strongest supporter. You've picked me up in the middle of snow storms and supported me during my most stressed and worst times. Whenever I'd have a small victory, you'd be adamant that we should celebrate. While I've worked long hours, you've cared for Minnie, kept the house in order, and have worked three jobs (three!) without once complaining. You've kept me calm throughout graduate school, and I couldn't have done this without you. You deserve one billion medals.

Table of Contents

List of Tables

List of Figures

List of Copyrighted Material

Bannert, N., and Kurth, R. 2004. Retroelements and the human genome: New perspectives on an old relation. PNAS. 101 (suppl 2): 14572–14579. DOI:10.1073/pnas.0404838101. Copyright 2004 National Academy of Sciences.

Bannert N., and Kurth R. 2006. The Evolutionary Dynamics of Human Endogenous Retroviral Families. Annual Review of Genomics and Human Geneics. 7:149-173. DOI: 10.1146/annurev.genom.7.080505.115700

Dimitrov, D.S. 2004. Virus entry: molecular mechanisms and biomedical applications. Nature Reviews Microbiology. **2**:109–122. DOI: 10.1038/nrmicro817

Eckwahl, M.J., Telesnitsky, A., and Wolin, S.L. 2016. Host RNA Packaging by Retroviruses: A Newly Synthesized Story. mBio. 7(1), e02025–15. DOI: 10.1128/mBio.02025-15

Handel, A.E., Ebers, G.C., and Ramagopalan, S.V. 2010. Epigenetics: molecular mechanisms and implications for disease. Trends in Molecular Medicine. 16(1): 7-16. DOI: https://doi.org/10.1016/j.molmed.2009.11.003

Jern, P., Sperber, G.O., and Blomberg, J. 2005. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology. 2:50. DOI:https://doi.org/10.1186/1742-4690-2-50

Jern, P. and Coffin, J.M. 2008. Effects of Retroviruses on Host Genome Function. Annual Review of Genetics. 42(2008): 709-732. DOI: 10.1146/annurev.genet.42.110807.091501

Matouk, C.C., and Marsden, P.A. Epigenetic Regulation of Vascular Endothelial Gene Expression. Circulation Research. 2008;102:873-887. DOI: https://doi-org.ezproxy.library.tufts.edu/10.1161/CIRCRESAHA.107.171025

Montesion, M., Bhardwaj, N., Williams, Z., Kuperwasser, C., and Coffin, J.M. 2018. Mechanisms of HERV-K (HML-2) transcription during human mammary epithelial cell transformation. Journal of Virology. 92(1); e01258-17. DOI: 10.1128/JVI.01258-17

Rabson, A.B., and Graves, B.J., 1997. Synthesis and Processing of Viral RNA. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

Rosenberg, N., and Jolicoeur, P. 1997. Retroviral Pathogenesis. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press

Telesnitsky, A., and Goff, S.P. 1997. Reverse Transcriptase and the Generation of Retroviral DNA. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press

Tollervey, J.R., and Lunyak V. 2012. Epigenetics: judge, jury, and executioner of stem cell fate. Epigenetics. 7(8): 823-840. DOI: 10.4161/epi.21141

Vaissiere, T., Sawan, C., and Herceg, Z. 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research. 659(1-2): 40-48. DOI: 10.1016/j.mrrev.2008.02.004

Vogt, P.K., 1997. Historical Introduction to the General Properties of Retroviruses. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

Vogt, V.M., 1997. Retroviral Virions and Genomes. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

Weiss, R.A. 2006. The Discovery of Endogenous Retroviruses. Retrovirology. 3:67. DOI: https://doi.org/10.1186/1742-4690-3-67

List of Abbreviations

ALV	Avian Leukosis Virus
ASLV	Avian Sarcoma/Leukosis Viruses
ATLV	Adult T-cell Leukemia Virus
CA	Capsid
CSD	Chronic Sun Damage
CPSF	Cleavage and Polyadenylation Specific Factor
DNMT	DNA Methyltransferase
enJSRV	endogenous JSRV
ERV	Endogenous retrovirus
FeLV	Feline Leukemia Virus
FPKM	Fragments per Kilobases per Million Mapped Reads
HDAC	Histone Deacetylases
HERV	Human ERV
HIV	Human Immunodeficiency Virus
HME	Human Mammary Epithelial Cells
HML	Human MMTV-Like
HTLV	Human T-Cell Lymphotropic Virus
HuMEC	Human Mammary Epithelial Cells
IAP	Intracisternal Type A Particles
IGV	Integrative Genomics Viewer
IN	Integrase
JSRV	Jaaksetie Squamous Carcinoma Retrovirus
KoRV	Koala Retrovirus
lncRNA	long non-coding RNA
LTR	Long Terminal Repeat
MA	Matrix
MBP	Methylation Binding Protein
miRNA	microRNA
MLV	Murine Leukemia Virus
MMTV	Mouse Mammary Tumor Virus
MYA	Million Years Ago
NC	Nucleocapsid
ORF	Open Reading Frame
PBS	Primer Binding Site
piRNA	PIWI-interacting RNA
PPT	Polypurine Tract
PR	Protease
RAV-0	Rouse-associated Virus
RLU	Relative Light Units
RSV	Rous Sarcoma Virus
RT	Reverse Transcriptase
SA	Splice Acceptor
SD	Splice Donor
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SU	Surface
TF	Transcription Factor
TFBS	Transcription Factor Binding Site

TFIIB	Transcription Factor II B
TM	Transmembrane
vlncRNA	Very lncRNA
VLP	Viral-Like Particle
VSV	Vesicular Stomatitis Virus

Chapter 1: Introduction

1.1 A Brief Look at the History of Retroviruses

Two years after graduating from medical school, Peyton Rous joined the Rockefeller Institute where he was appointed to lead a cancer research lab (6). Shortly after his assignment began, his lab received a Plymouth Barred Rock hen from an inbred flock that presented with a large sarcoma in its right breast (5, 6). A piece of the sarcoma was removed from this hen and re-implanted into its peritoneal cavity and left breast along with two other birds of its flock (6). The original hen died 25 days later due to growth from the transplant in its peritoneal cavity, and one of the two inoculated hens developed tumors that were similar in morphology to the original tumor (5, 6). Transplantation of tumor fragments into birds from the same original flock was weakly successful with a transplantation success rate of 25%, yet the continued passage of these tumors in this inbred flock resulted in highly malignant masses that grew quickly (6). Removal of live cells through grinding and filtration resulted in a supernatant, when injected into relatives of the original hen, resulted in slow-growing tumors at the injection site (6). By 1911, Rous published this work on transplantable tumors caused by a filterable agent, which we now know as Rous Sarcoma Virus (RSV) (1, 5, 6, 19).

Rous was not the only scientist at the time to isolate an infectious virus that caused cancer in chickens. In 1914, Fujinami and Inamoto detected another avian sarcoma virus which they named after Fujinami (1, 19). In 1908, Danish veterinarians Vilhelm Ellermann and Oluf Bang published their work on identifying a viral cause behind chicken leukosis, beating Rous to the punch in identifying a viral origin of chicken cancer (1, 5). This virus - Avian Leukosis Virus (ALV) - is grouped with RSV and other avian sarcoma viruses into the avian virus genus known as alpharetroviruses (1, 5).

Though exciting in hindsight, at the time this work was dismissed: The origin of leukemia as a bone marrow - derived cancer was debated. Cancer was “known” not to be contagious in humans, and chickens were thought to be too unrelated to humans to be reliable models (1, 5). In time, RSV proved to be more useful than anyone at the time anticipated. This virus helped revolutionize oncology through detection of proto-oncogenes and elucidation of their mechanisms, added new tools to the molecular biologist’s toolkit, and changed the way scientist thought of the Central Dogma. Due to this, this virus would indirectly help with the identification of human cancer - and immunodeficiency - causing retroviruses (1, 7).

These were not the first reported instances of cancer transmissibility (5, 19). In 1842, nuns in Verona were reported to have a lower incidence of cervical cancer compared to married women (5). Similarly, bovine leukosis and Jaagsiekte lung carcinoma were also known to be transmissible in the 19th century yet it would take until the 1980s to understand the retroviral origins of these illnesses (5, 19). Post - Rous, additional cancer - causing retroviruses were discovered. It was initially thought that breast cancer in mice was genetic since lines with high or low incidence produced pups of high or low incidence, respectively (19). However, in 1936 John Bittner establish that mammary carcinoma in mice was transmissible by a filterable agent found in milk by fostering pups from low incidence lines with mothers of high incidence background (1, 19). This virus, now known as Mouse Mammary Tumor Virus (MMTV), was finally identified in 1949 by Dmochowski using an electron microscope, which was the common - although often insensitive and cumbersome - technique at the time (19). This was followed by Ludwik Gross who identified a strain of murine leukemia virus (MLV) in 1951 in the AKR mice developed by Jacob Furth in 1933 (1, 19). This began a cascade of cancer - causing retroviral discovery in the following decades in cats, mice, birds, and primates (1, 5).

When the genome of RSV was shown to be RNA in 1961, oncogenic retroviruses came to be known as RNA tumor viruses (19). Interestingly, cells transformed with this virus maintained their transformed state for many rounds of replication even in the absence of replicating virus (19). This led Howard Temin to develop a widely panned hypothesis that retroviruses can produce a DNA copy of themselves within infected cells that is subsequently inserted into the host cell genome, thereby rewriting the Central Dogma (1, 7, 19). He called this the provirus hypothesis, inspired by integrated prophages from bacteriophages (19). Around the same time, David Baltimore was looking for nucleic acid polymerases in RNA viruses, stemming from his experience with vesicular stomatitis virus (VSV) and the discovery of VSV transcriptase (1, 7). Using the RNA tumor viruses MLV and RSV, he initially looked for the presence of RNA polymerase only to come up empty handed. However, when he switched ribonucleoside triphosphates to deoxyribonucleoside triphosphates he was able to detect synthesis of MLV DNA, supporting Temin's hypothesis. The following three years led Baltimore to look at replication in MLV and Temin to do the same in RSV, leading them to simultaneously discover RNA dependent DNA polymerases. Both Temin and Baltimore - technically - presented this finding at the June 1970 Transcription meeting at Cold Spring Harbor, where Baltimore presented his lab's work while Temin sent a publication (7). The discovery of this enzyme, renamed as "reverse transcriptase", would become important to the identification of other retroviruses.

The identification of reverse transcriptase in virions provided a simple and sensitive way to detect and quantify viruses, especially compared to electron microscopy (9). The mechanism by which retroviruses integrate into the host genome also provided a possible mechanism that would explain how RNA viruses could transform cells into cancer cells (9). This mechanism encouraged investigators to look for human retroviruses causing cancer - "human tumor viruses" - as a part of the Special Virus

Cancer Program, an initiative established by the National Cancer Institute to look for retroviruses that cause cancer in humans (9). Reports of retrovirus detection in human tumors started to appear in the early 1970s (8). These were later shown to be due to laboratory contamination from other animal retroviruses - either from passage through mice or from cross - contamination - or from mitochondria (8). After repeated instances of contamination from one source or another, many investigators started to give up the search, calling investigations to identify a human cancer - causing retrovirus as a quest for “human rumor viruses” (9).

There was one group that did not give up. In 1972, Robert Gallo’s group detected a ribonuclease sensitive DNA polymerase in stimulated human lymphocytes (8). To identify what this was, Gallo developed a series of sensitive experiments to isolate any potential viruses found in leukemia and lymphoma cells, part of which was improving culture conditions for T-cells to allow for their continued passage (8, 9). His group finally detected reverse transcriptase in 1979 from an established T-cell line and published on the isolation of the virus now known as Human T-Cell Lymphotropic Virus (HTLV) (7, 8). Interestingly enough, when Gallo first attempted to publish his work on the isolation of HTLV his paper was rejected on the grounds that they were continuing to support a hypothesis that to that point was considered untrue (7, 8, 9, 11). Gallo’s work was followed by two additional papers from Yorio Hinuma on the isolation of a similar virus from primary cultured patient cells and the presence of antibodies against HTLV-1 in Japanese cancer patients, a virus that he named Adult T-cell Leukemia Virus (ATLV) due to the cellular origins that he found it in (8,9). A few years after their papers came out, it was realized that the viruses they individually isolated were isolates of the same virus and so they agreed on the name HTLV.

The methods used to culture T cells and to isolate HTLV was important for isolating another virus. By 1981, the first reported cases of a new infections disease

appeared in California and New York City (15-17). French researchers in the Montagnier team received samples from a young man who presented with lymphadenopathy and were able to isolate a virus (7, 8, 10, 11). Electronic microscopy, DNA inhibitors, and antibodies specific to HTLV suggested that this virus - isolated in 1983 - was a new virus (8). Gallo would follow within the next year with four studies that provided more conclusive evidence that this virus - then known as LAV by Montagnier and HTLV-III by Gallo - was different from HTLV (7, 8, 10, 11). Subsequent years saw the development of new detection methods by Gallo for the identification of patients who carried the virus, and a lawsuit between the French and United States government over the identification of this new virus (8). By 1986, it was proposed that this virus be renamed to Human Immunodeficiency Virus (HIV) (18).

1.2 The Genomic and Virion Structure of Retroviruses

Prior to DNA cloning and sequencing, other biochemical methods had to be used to infer the genetic structure of the retrovirus (2). The size was initially determined through RNA fingerprinting in the 1970s, while restriction endonucleases and Southern blotting were used to make genetic maps of viral genomes (2). By the late 1970s, the genomes of MLV and RSV were mapped thanks to the development of sequencing technology (2).

Retroviruses are a diverse group of enveloped ssRNA viruses where genomic composition, virion structure, and replication requirements vary across each of the seven known genera: alpharetroviruses, betaretroviruses, gammaretroviruses, lentiviruses, deltaretroviruses, epsilonretroviruses, and spumaviruses (1, 2, Figure 1.1, Figure 1.2). Retroviral genomes can range in size from seven kb to 12 kb and are classified as linear, non-segmented, and of positive polarity (1, 2). Their genetic composition is largely conserved where they contain four key ORFs for their structure and function (*gag, pro,*

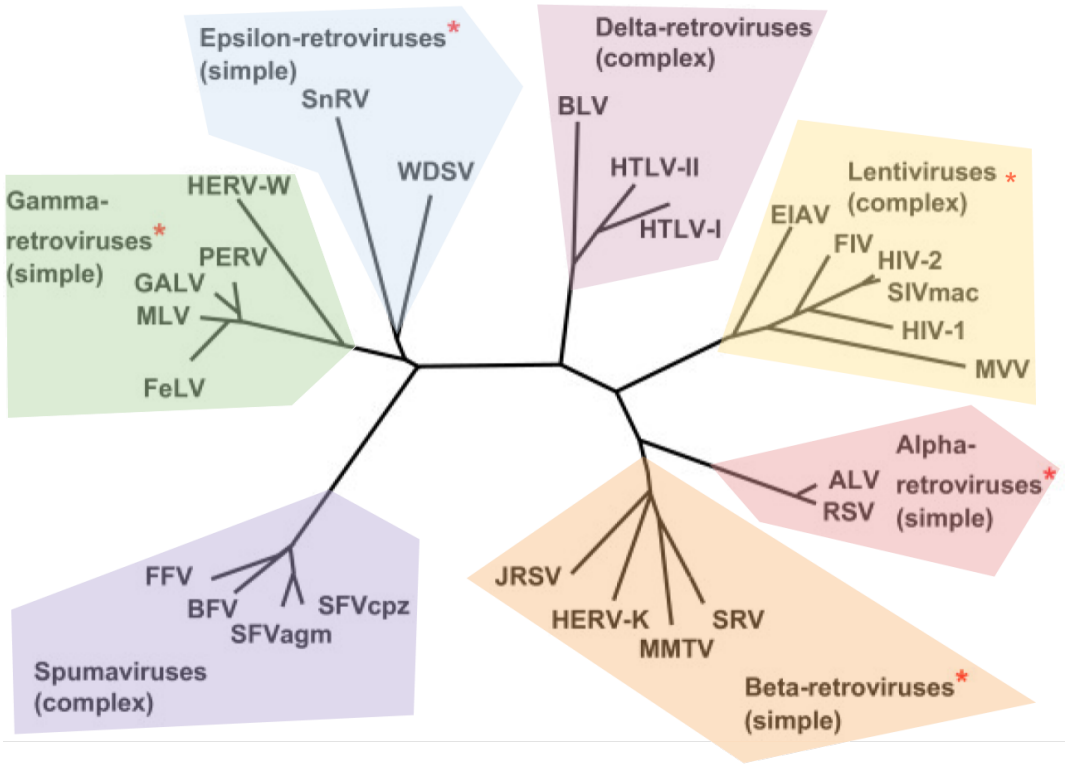


Figure 1.1: Phylogenetic tree of retroviruses.

The seven known genera of retroviruses have been highlighted in various colors to group them and their representative examples within the body of the figure. Groups with a red asterisk contain both endogenous and exogenous viruses. Adapted from Retrovirology with modifications to add color to visually group viruses of the same clade together and to add an asterisk to Lentiviruses to indicate the presence of endogenous and exogenous viruses[1] (19). Journal is open access, so permission was not required.

[1] Weiss, R.A. 2006. The Discovery of Endogenous Retroviruses. Retrovirology. 3:67. DOI: https://doi.org/10.1186/1742-4690-3-67

Figure 1.2: Basic genetic composition of retroviruses.

The genetic structure and organization of retroviruses. All retroviruses contain ORFs for *gag* (blue), *pro* (red), *pol* (yellow), and *env* (green) in that order. Long terminal repeats (LTRs, grey) flank these ORFs and are essential for replication and retroviral transcriptional regulation. Several retroviruses contain additional ORFs for accessory proteins (not shown above) which are typically found around the *env* ORF.

pol, and *env)* with several containing additional reading frames essential to their replication (2, 19, Figure 1.2). For example, HTLV – which contains at least five "accessory genes" - encodes *tax* for transactivation and *rex* for cytoplasmic importation (1, 2). These internal sequences are flanked by two long terminal repeats (LTRs), which are important for viral replication, and transcription/translation regulation (1, 2, Figure 1.2).

LTRs contain a U3, R, and U5 region (2, Figure 1.3). The 5' LTR acts as the promoter and regulator of retroviral expression post insertion through its U3 element and – alongside the 5' LTR's 3' counterpart - is critical in viral genome replication. Prior to integration and after, the ends of the viral cDNA genome consist of R - U5 on the 5' end of the genome and U3 - R on the 3' end of the genome (Figure 1.3).

The *gag* gene encodes virion structural proteins (1, 2, Figure 1.3, Figure 1.4). These proteins include matrix (MA), capsid (CA), and nucleocapsid (NC) (1, 2). MA - sometimes referred to as membrane-associated protein - is closely associated with the lipid bilayer that is picked up from the host cell, therefore forming part of the outer shell of a virion (1, 2). CA forms a protective shell around the viral RNA genome to keep it intact and to transport the genome and necessary replication equipment (enzymes, tRNA primer) into the cytoplasm so it can replicate effectively (1, 2). Collectively, CA and

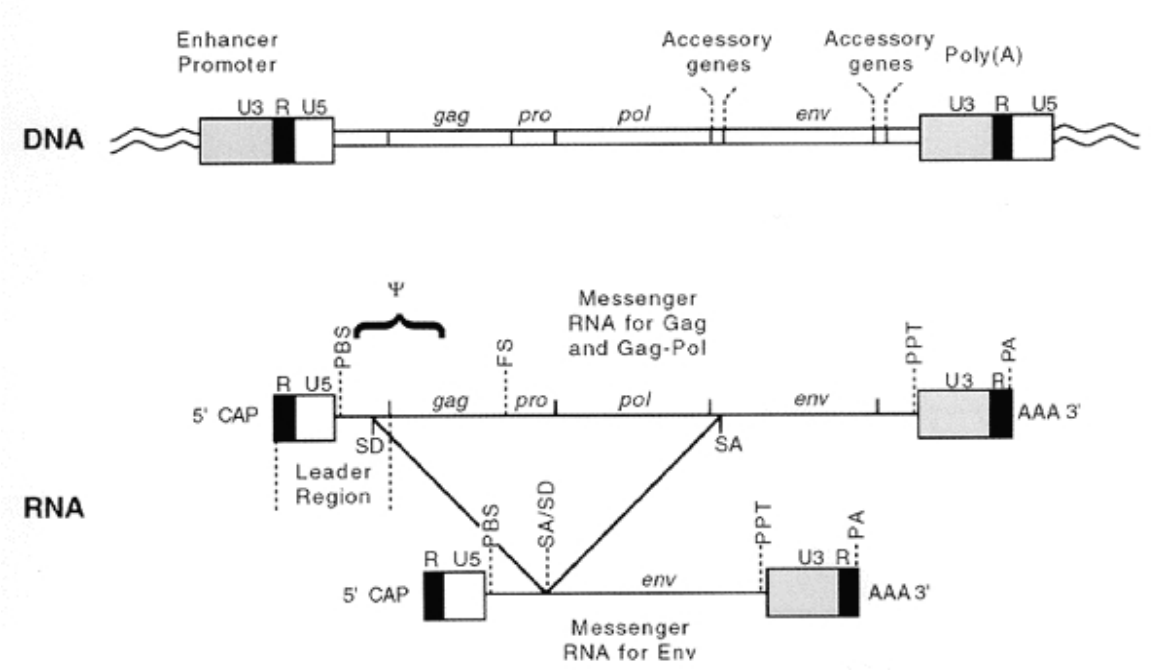


Figure 1.3: Genetic organization of retroviral DNA and RNA.

Structural differences between retroviral DNA and RNA are presented above. Once inserted into the host genome, the provirus (top) is treated as any other gene. ORFs for *gag, pro, pol,* and, *env* are indicated in the provirus sequence with marks representing the approximate placement for accessory genes. The different segments within the 5' and 3' LTRs (U3, R, and U5) are marked by grey, black, and white boxes, respectively. The placement of enhancer and promoter binding motifs are located in U3 of the 5' LTR, while there is a poly(A) sequence in the 3' R sequence. Squiggly lines flanking the provirus represent the host genome. Once expressed, viral mRNA exists in unspliced (middle) and spliced (bottom) forms for translation. Unspliced transcripts are used for the production of Gag, Pro, and Pol, while spliced transcripts are used for the production of Env. The splice sites (splice acceptor (SA) and splice donor (SD)) are marked in both the spliced and unspliced mRNA. Unlike the provirus sequence, mRNA transcripts lack the 5' U3 region and the 3' U5 region. The 5' end of mRNA transcripts contain a 5' cap and the 3' end contains a poly(A) tail. The sites of primer binding (PBS) to initiate first strand synthesis in viral replication and the polypurine tract (PPT) used to initiate second strand synthesis are indicated on both the spliced and unspliced transcript. Reprinted with permission from Cold Spring Harbor Laboratory Press[2] (2).

[2] Vogt, V.M., 1997. Retroviral Virions and Genomes. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

the internal components it is housing are referred to as the viral core. Finally, NC binds to the single-stranded RNA genome to compress the genome so it can fit into the core and to stabilize the structure until reverse transcription can occur (1, 2).

Pol encodes the major enzymatic workhorses of the virion: reverse transcriptase (RT), and integrase (IN) (1, 2, Figure 1.3, Figure 1.4). Post-entry, RT uses its combined DNA polymerase and RNase H activities to reverse transcribe the viral RNA genome into double-stranded cDNA while simultaneously degrading the ssRNA genome. Once the viral genome enters the host cell's nucleus, it then becomes the job of IN to insert viral cDNA into the host through cleavage of the host genome at target sites followed by integration of viral cDNA. *Pro*, the third major enzyme encoded by retroviruses, encodes protease which is responsible for retroviral particle maturation post-budding from the infected host cell.

Env encodes the envelope glycoprotein responsible for entry into the host cell, divided into the surface (SU) and transmembrane (TM) domains (1, 2, Figure 1.3, Figure 1.4). On the surface of virions, Env exists as a dimeric trimer where SU is responsible for identifying cognate receptors that the virus can use to enter a host cell, while TM is responsible for fusion and entry once bound to a target cell. In terms of entry, viruses - especially retroviruses - have two different methods of entry which I will discuss later in this chapter.

There are a few other structural features of the retroviral genome that are critical to viral replication (1, 2, Figure 1.3). The space from the R region of the 5' LTR to *gag* - a region known as the *gag* leader sequence - contains a portion of the packaging (Ψ) sequence, which marks retroviral RNA for packaging into newly synthesized virions, and the primer binding site (PBS), where cellular tRNA binds to viral RNA to initiate first strand synthesis. The leader sequence also contains the splice donor (SD) sequence, which splices with the splice acceptor (SA) site adjacent to *env* to form the spliced

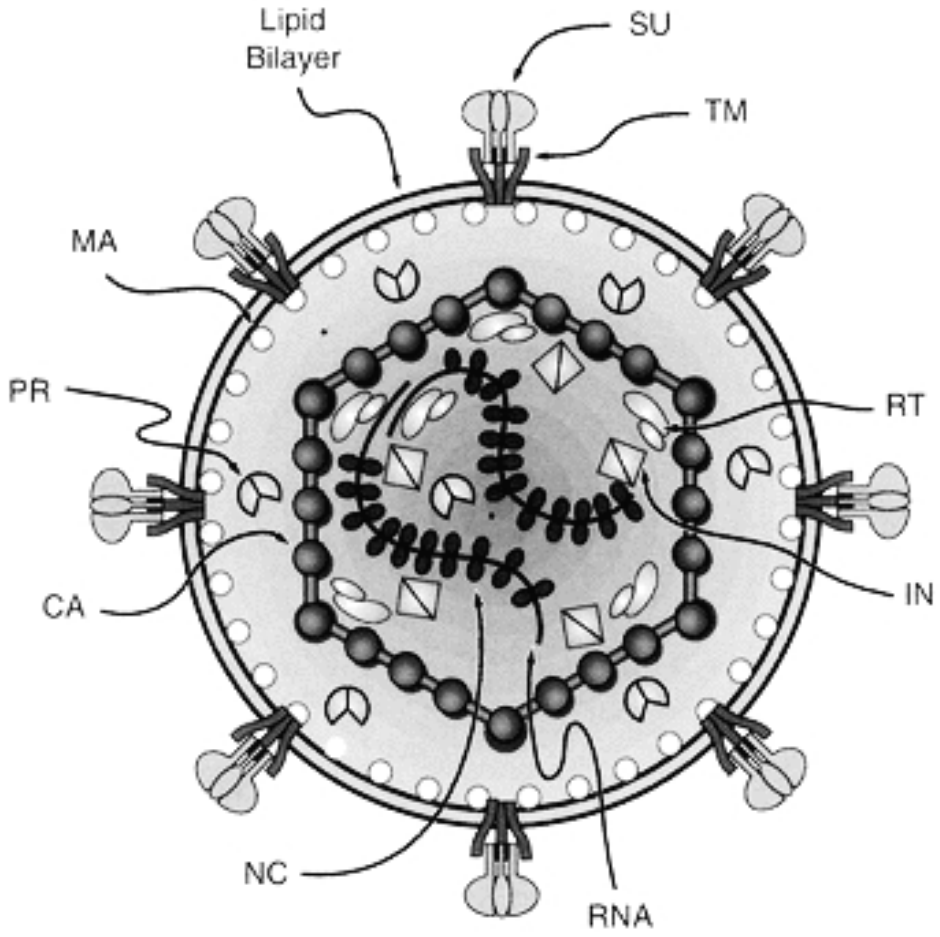


Figure 1.4: Structure of a retroviral virion.

A schematic of the retroviral virion anatomy. The dimeric ssRNA genome is bound to the nucleocapsid (NC) proteins and is contained within the core (hexagon structure). The core also contains reverse transcriptase (RT), integrase (IN), dNTPs (not shown) and tRNAs (not shown). The core is contained by the lipid bilayer - collected from the host cell the virion budded from - which also contains protease (PR). Closely associated with the lipid bilayer is the matrix (MA) protein. Within the lipid bilayer is the envelope protein, consisting of the subunits surface (SU) and transmembrane (TM). Reprinted with permission from Cold Spring Harbor Laboratory Press[3] (1).

[3] Vogt, P.K., 1997. Historical Introduction to the General Properties of Retroviruses. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

transcript used for Env protein production. Finally, preceding the U3 element in the 3' LTR is the polypurine tract (PPT), which serves as an RNase H - resistant tract of RNA used as a primer for second strand synthesis.

1.3 Retrovirus Replication Cycle

When a virion encounters a host cell, it will move along the surface until the SU registers a cell surface receptor that it can bind to strongly (1, Figure 1.5). Once attached, there is a structural change within the envelope protein that results in viral entry into the host cell. There are typically two mechanisms by which retroviruses enter host cells: pH-dependent and pH-independent (Figure 1.6). In the case of pH-dependent entry, the virion is taken into the cell via an endosome. The low pH environment of the endosome results in a conformational change within the viral envelope protein resulting in viral fusion and capsid entry into the intracellular space (Figure 1.6 A). As for pH-independent entry, the virus fuses to the plasma membrane of the host cell after contact with its cognate receptor and the capsid is released into the intracellular space (Figure 1.6 B). A common example of pH-dependent entry for retroviruses would be ALV, while a common example of pH-independent entry would be HIV.

At some point between capsid release and entry into the nucleus, the single-stranded viral RNA genome is reverse transcribed into double-stranded cDNA (1, Figure 1.7). Within each viral capsid are the viral genome, reverse transcriptase (RT), integrase, nucleotides, and a tRNA primer taken from the previous host cell. This primer binds to the PBS where RT transcribes from the PBS to R, where it encounters the end of the genome. As this strand is being synthesized, RNase H has started to degrade the already-transcribed RNA. Once R has been transcribed, the newly synthesized DNA strand is transferred for the first time to the 3' end of the viral RNA genome and binds to

the other R sequence. Here synthesis of the first DNA strand continues, with RNase H destroying the remainder of the viral RNA genome except for the PPT. As mentioned above, the PPT serves as a primer for second strand synthesis so at this step the second strand begins synthesis from PPT to the tRNA bound to the first strand. Second strand synthesis is halted once transcription reaches the tRNA, reverse transcribing a

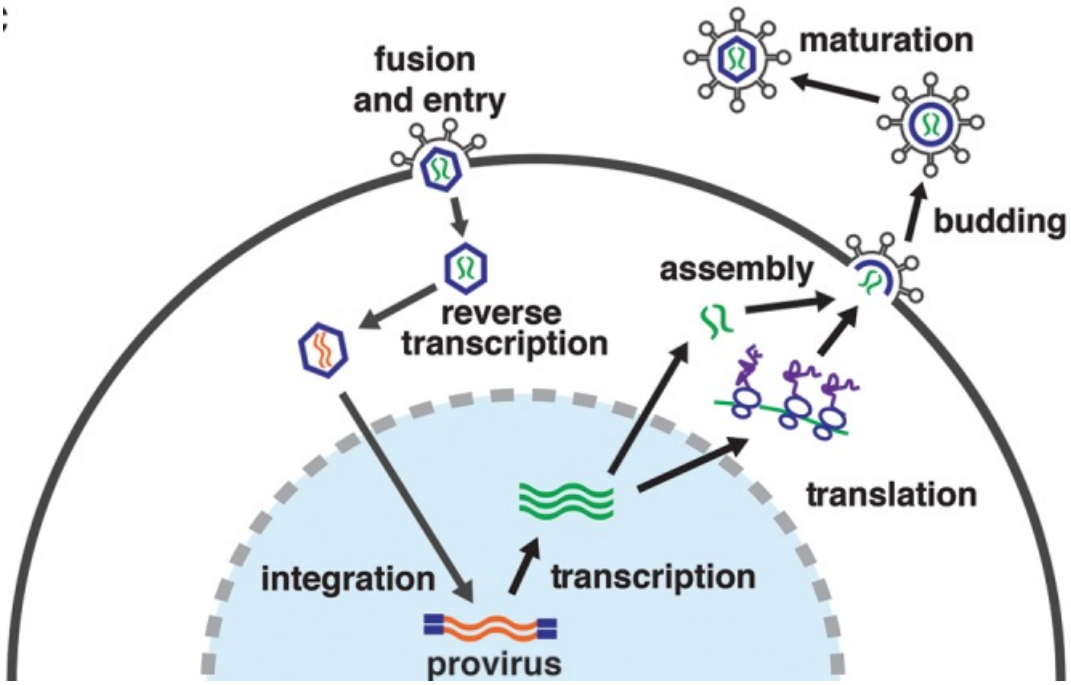


Figure 1.5 Schematic of retroviral replication cycle.

An infectious virus encounters its cognate receptor on the surface of a host cell. Upon binding, the virus enters the host cell and the core (blue hexagon) is released. While in the cytoplasm, reverse transcription of the ssRNA viral genome (green) to dsDNA (orange) occurs and is imported into the nucleus. Once within the nucleus, viral integrase inserts the dsDNA into the host genome (blue), creating a provirus. Upon cellular replication, the provirus is transcribed and translated (purple) to produce new virions. Adapted from mBio and modified to reproduce only panel C from original text [4] (158). Since this comes from an open-access article, permission to use was not needed.

[4] Eckwahl, M.J., Telesnitsky, A., and Wolin, S.L. 2016. Host RNA Packaging by Retroviruses: A Newly Synthesized Story. mBio. 7(1), e02025–15. DOI: 10.1128/mBio.02025-15

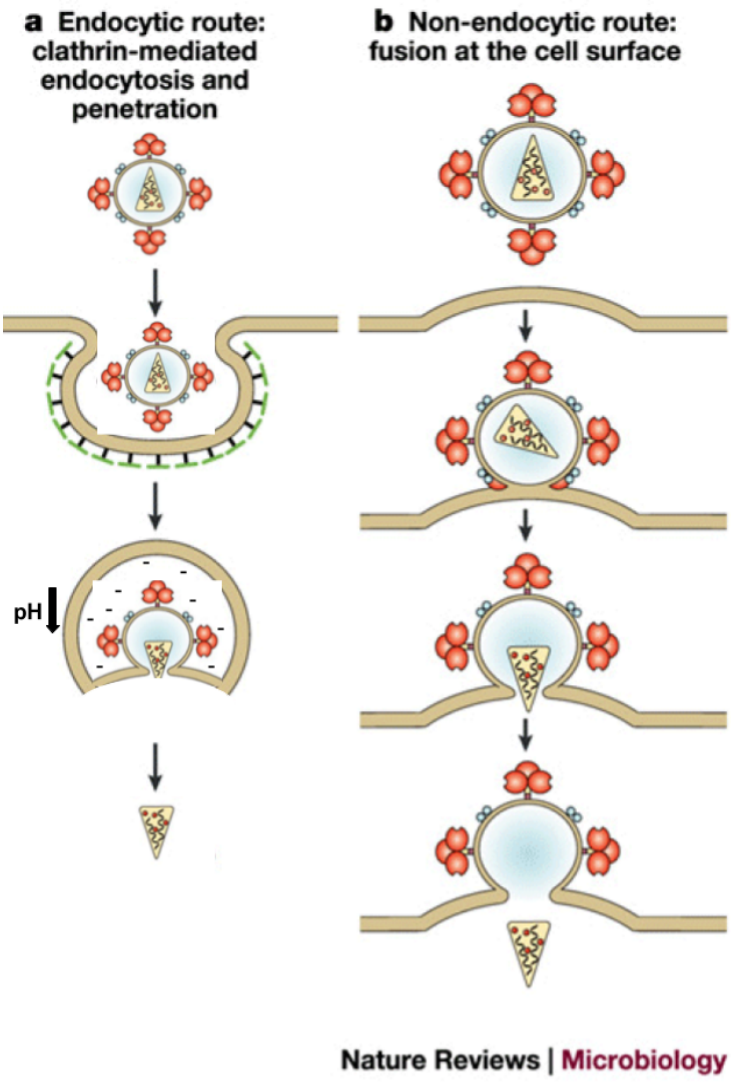


Figure 1.6 Schematic of pH dependent and independent entry.

Two models of retroviral entry into a host cell are represented above: pH dependent (A) and independent (B) entry. (A) A virus encounters its cognate receptor and is pulled into an endosome. When pH decreases, the envelope glycoprotein on the surface of the virion undergoes a conformational change, resulting in the release of the viral core (yellow triangle) into the cytoplasm. (B) A virus encounters its cognate receptor, triggering a conformational change in the viral envelope glycoprotein and the subsequent release of the viral core into the cytoplasm. Reprinted with permission from Nature Reviews Microbiology[5] (153). Adapted from original to exchange non-retroviral virion for retroviral virion and to add indication of a pH drop for fusion and core release to take place.

[5] Dimitrov, D.S. 2004. Virus entry: molecular mechanisms and biomedical applications. Nature Reviews Microbiology. **2**:109–122. DOI: 10.1038/nrmicro817

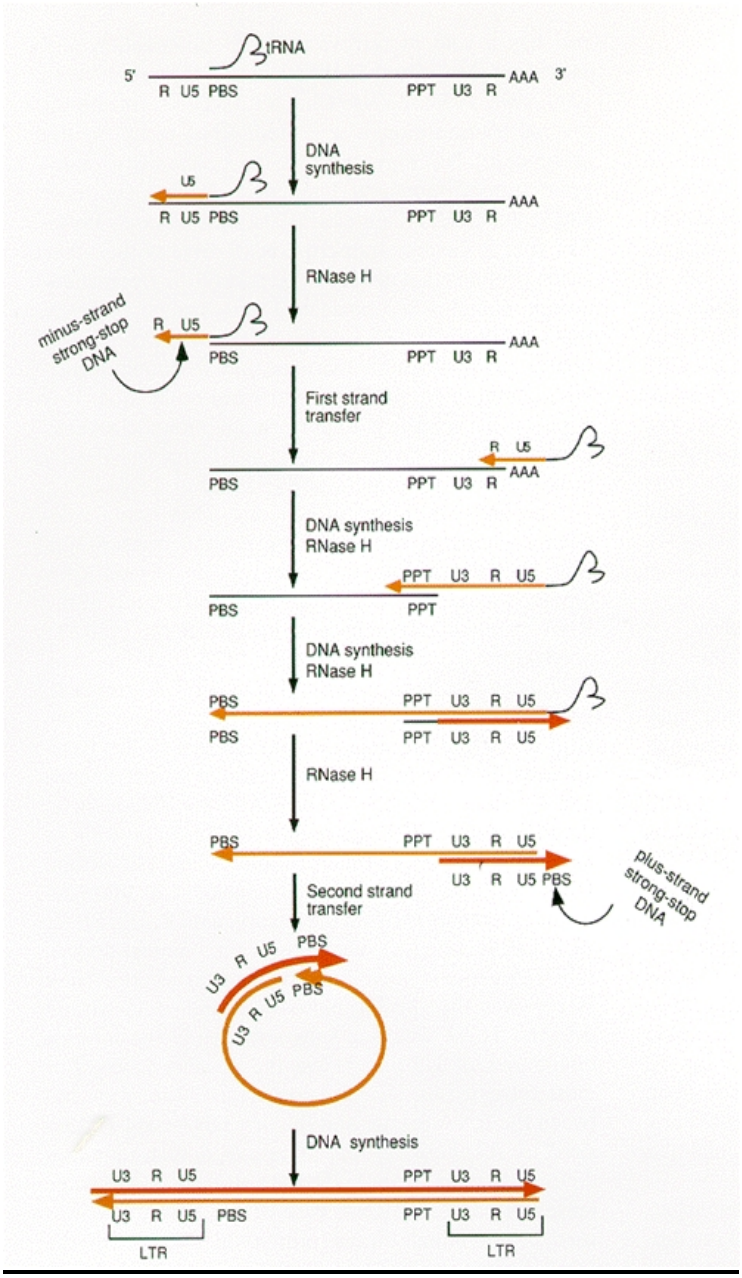


Figure 1.7: Schematic of first and second strand reverse transcription of retroviral RNA.

RNA is represented by a black line, while cDNA as orange. See text for a description of the process. Reprinted with permission from Cold Spring Harbor Laboratory Press [6] (155).

[6] Telesnitsky, A., and Goff, S.P. 1997. Reverse Transcriptase and the Generation of Retroviral DNA. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press

portion of it. RNase-H then removes the primer tRNA, which exposes a portion of the sequence that overlaps with the first strand (the PBS). Due to this overlap, the cDNA loops on itself to complete synthesis of the second strand, resulting in a double-stranded cDNA sequence containing two LTRs.

Once reverse transcribed and after entry into the nucleus, the linear double stranded viral DNA is integrated into the host via integrase. Integration happens in two catalytic steps involving nucleophilic reactions (162). The first step involves the removal of a dinucleotide from the terminal ends of the 3' strand at both the 5' and 3' LTR, generating exposed 3' ends. In the second step, the LTRs are positioned adjacent to the target integration site which triggers the second nucleophilic reaction via the exposed 3' strand. This second reaction results in the insertion of the retroviral DNA genome, yet this sequence contains gaps and two overhangs on the 5' strands. To finish the integration process, the 5' overhangs are removed, and the gaps are repaired. Upon integration into the host genome, proviruses are mostly forever: these sequences are susceptible to mutations that can occur based upon a host's natural genetic mutation rate, which varies by species. These mutations can range from insertions or deletions that can corrupt ORFs, or removal of an entire provirus through homologous recombination leaving behind a single solo LTR. These mutations are not restricted to the internal sequence of a provirus. Mutations can also happen within the 5' and 3' LTRs, which results in the two sequences - identical at insertion - to diverge.

Post-insertion, the retrovirus relies almost entirely on host cellular machinery to complete its replication cycle (157). LTRs are essential for this step, promoting and regulating both transcription and translation in the host cell. They contain important cis-acting elements responsible for regulating expression of retroviral genes. For example, the U3 region of the 5' LTR contains a variety of transcription factor binding sites, capable of recruiting both core transcription factors such as Transcription Factor II B

(TFIIB) and other regulatory transcription factors (ex. SP1, SOX 10, etc) (157, Figure 1.8, Figure 1.9). These transcription factors, similar to other core regulatory elements in the pre-initiation complex, bind to the 5' LTR in a sequence specific manner. Binding of core transcription factors (ex. TFIIB, TFIID, etc) to these elements within the 5' LTR results in the formation of the pre-initiation complex and the recruitment of RNA polymerase II (Figure 1.9).

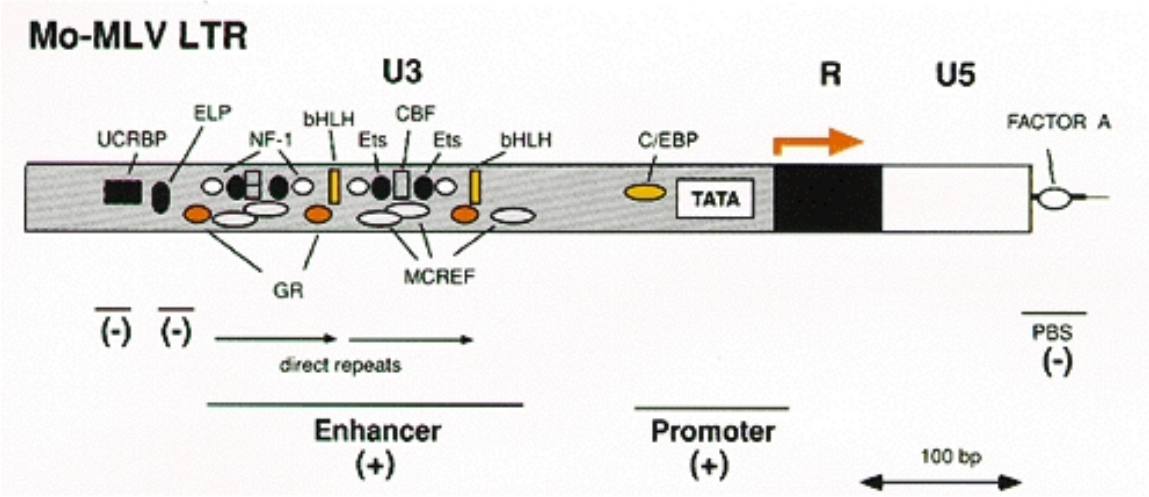


Figure 1.8: Example of LTR structure

LTR structure of Moloney MLV's 5' LTR. Cis-acting elements such as transcription factor binding motifs and the TATA box are found within the U3 (grey box) region of the LTR. The TATA box and neighboring transcription factor binding sites (such as C/EBP in the above example) are a part of the promoter sequence. Other transcription factor binding sites upstream of the promoter (such as ELP, NF-1, etc) are considered enhancer elements. Transcription initiates at R (black box) and is indicated by orange arrow. Reprinted with permission from Cold Spring Harbor Laboratory Press [7] (157).

[7] Rabson, A.B., and Graves, B.J., 1997. Synthesis and Processing of Viral RNA. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

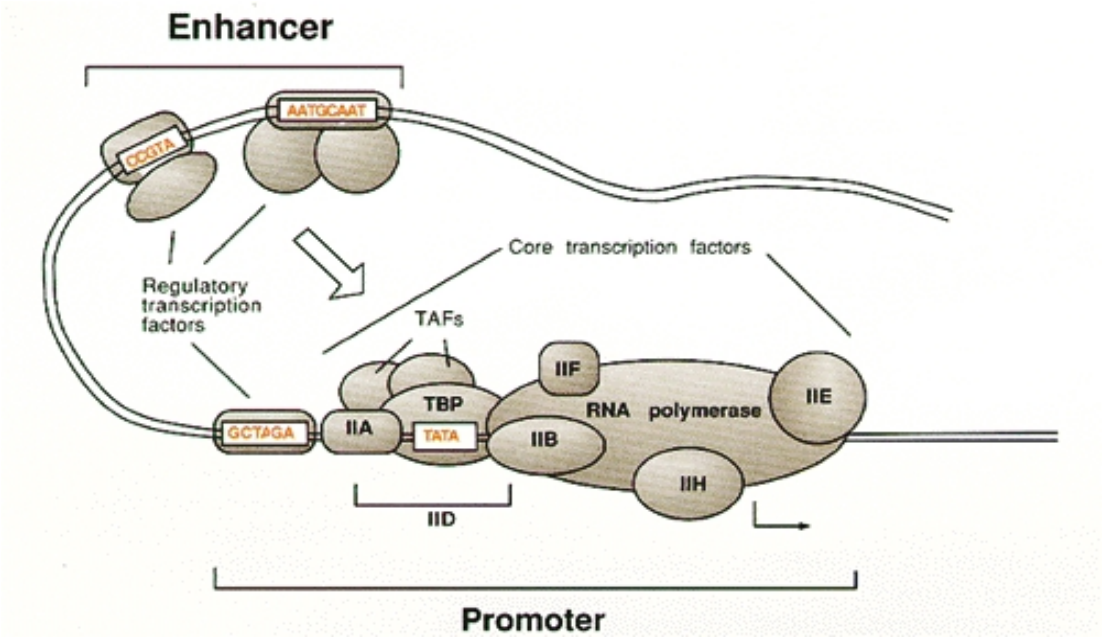


Figure 1.9: Structure of the pre-initiation complex and enhancer elements within eukaryotes

Core transcription factors (TFIIB, TFIIF, TFIIH, TFIIE, TFIID) recruit RNA polymerase to the promoter element. Additional binding of other regulatory transcription factors to enhancer motifs (orange) help further enhance transcription of genes. Reprinted with permission from Cold Spring Harbor Laboratory Press [7] (157).

[7] Rabson, A.B., and Graves, B.J., 1997. Synthesis and Processing of Viral RNA. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press.

Due to transcription factors' role in recruiting RNA polymerase II, the importance of transcription factors in gene regulation is undisputable. Other regulatory cis-acting factors can bind upstream of the TATA box - around the R-U5 region - which are either immediately upstream of the pre-initiation complex, therefore making them a part of the promoter element, or further upstream of the promoter motif (157, Figure 1.8, Figure 1.9). These regulatory elements - such as additional transcription factor binding motifs - can be found within the U3 of the 5' LTR or further upstream (upwards of 100 kb) (157, Figure 1.8, Figure 1.9). The range on these elements is thought to come from the flexibility of chromatin, where a bend in the genome can come within range of the promoter element (157, Figure 1.9). Expression of genes encoding transcription factors varies in terms of intensity and diversity depending on cell type and health, especially in cancer where numerous transcription factors have been reported as either upregulated or downregulated depending on the cancer.

Once the pre-initiation complex is assembled, RNA polymerase II transcribes the provirus in a similar fashion to what is done with other host genes. These elements will transcribe the rest of the provirus, and these transcripts will be targeted for nuclear export. Prior to nuclear export, retroviral transcripts will be modified through the addition of a 5' cap and a poly(A) tail (157). In the case of poly(A) tail addition, readthrough of RNA polymerase II through the 3' LTR can occur and result in a longer transcript. The addition of a 3' poly(A) tail occurs through the binding of a cleavage and polyadenylation specific factor (CPSF) in addition to the presence of a conserved polyadenylation signal (AAUAAA) located in the R or - in the case of viruses such as HIV - U3 region (157). Cleavage and addition of the poly(A) tail occurs at the 3' LTR at the R-U5 border due to these two cis-acting elements in addition to the presence of GU rich areas downstream of the polyadenylation signal (157).

The addition of these elements is for a combination of stability and for ribosomal binding to the mRNA transcript for translation. Once posttranscriptionally modified, retroviral mRNA transcripts will be transported to the cytoplasm to either produce Gag and Gag-Pro-Pol for retroviral particle formation or packaging into budding infectious retroviral particles. Some retroviral mRNA will be processed further prior to nuclear export to produce a subgenomic length transcript used for Env production. In this case, members of the spliceosome will bind to the transcript and direct cleavage of the internal sequence to produce the truncated product. Both full length and spliced retroviral transcripts are exported from the nucleus into the cytoplasm using trans-acting elements such as Rev in HIV or Rec for HERV-K(HML-2). Once imported into the cytoplasm, mRNA transcripts will be bound by host cell ribosomes to produce Gag, Pro, Pol, and Env.

These contents - except for spliced retroviral RNA - are trafficked to the surface of a cell where a new immature virion buds. While spliced retroviral RNA can be packaged into a budding virion, full length retroviral RNA is favored due to the presence of the packaging signal. After budding, retroviral virions undergo a final step to become infectious virions where protease cleaves CA and MA to form a proper structure before proceeding to infect new cells.

1.4 A Brief Overview of Endogenous Retroviruses (ERVs) and an Introduction to HERV-K (HML-2)

Transposable elements encompass 45% of the human genome (20, 21, 29, 38, 39, Figure 1.10). The most abundant transposable elements are retroelements, which constitute 42% of the human genome (38, 39, Figure 1.10). Retroelements are further split into LTR and non-LTR elements, where non-LTR elements such as LINEs and SINES account for 34% of the total human genome and LTR elements such as

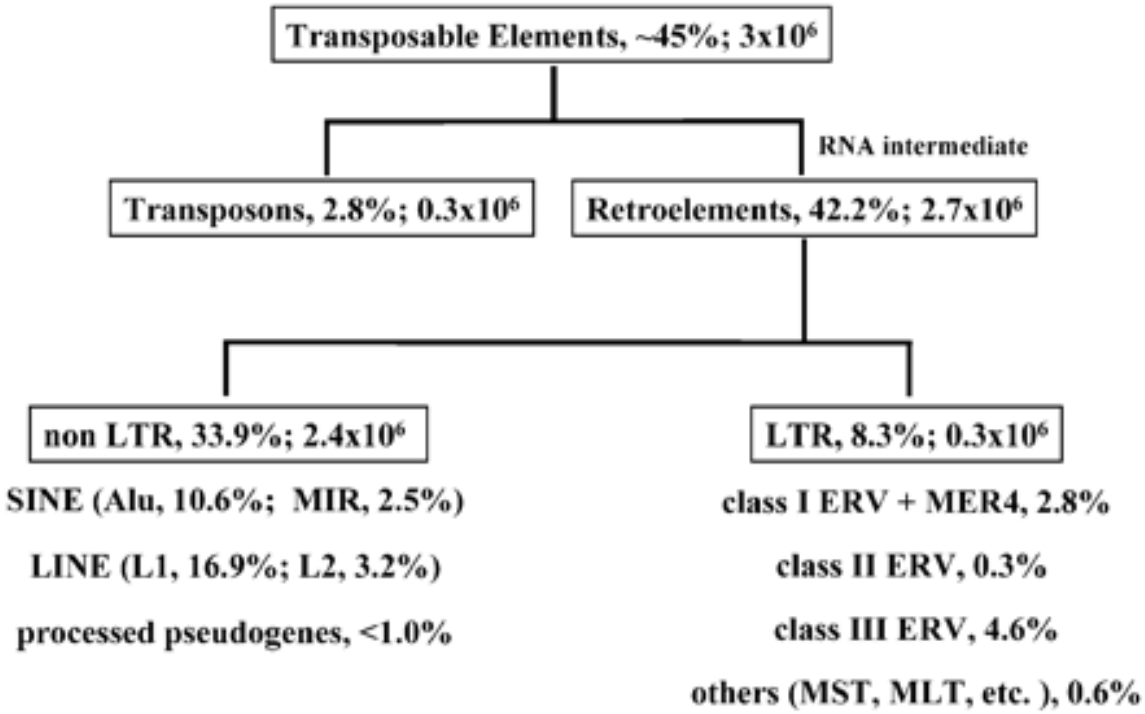


Figure 1.10 Identity of transposable elements in human genome.

A dissection of the different types of transposable elements (TEs) within the human genome, in addition to their abundance within the genome (%). Percentage of each element within the human genome is specified next to the group name. Reprinted from the National Academy of Sciences [8] (38). Since this figure is being used for educational purposes, formal permission is not required.

[8] Bannert, N., and Kurth, R. 2004. Retroelements and the human genome: New perspectives on an old relation. PNAS. 101 (suppl 2): 14572–14579. DOI:10.1073/pnas.0404838101. Copyright 2004 National Academy of Sciences.

retrotransposons and endogenous retroviruses (ERVs) account for 8% (19, 22, 29, 30, 38, 39, 41, Figure 1.10).

As mentioned above, the creation of a provirus by insertion of the retroviral genome into the host cell is a critical step in retroviral replication (1, 2, 21, 30, 31, 38, 40, Figure 1.5). Once a provirus is created, it is a permanent fixture within the host cell except for occasional deletion of the internal ORFs and an LTR through homologous recombination, leaving a solo LTR as a reminder of the provirus that once existed (21). While retroviruses typically infect somatic cells, they can occasionally infect gametes (19 - 22, 30, 31, 38, 40, Figure 1.11). If an infected gamete results in viable offspring, the provirus - now an ERV - will be present in every cell at the same insertion site of the offspring. This ERV will be treated as any other gene in terms of activation and Mendelian inheritance.

ERVs have been found in all vertebrate species analyzed, such as humans, mice, cats, sheep, chickens, and - more recently - koalas (19, 20, 23 - 28, 30 - 33, 39, 41). Depending on the species, there could be anywhere from a few to a few thousand ERVs and retroviral elements present due to reinfection by a replication-competent ERV (42, 43), retrotransposition (44, 45), or infection of a host by an exogenous retrovirus to produce a new ERV (46, 47) (20, 22, Figure 1.12). Their number provides an opportunity for studying their evolution, the evolution of the host, and evolution of the host-virus relationship (40). Due to these mechanisms and their abundance in some species, ERVs have colonized vertebrate genomes over a prolonged period, ranging from decades to millions of years (20, 36, 48-52).

ERVs can be roughly grouped based on their integration time into “ancient” and “modern” proviruses (20, 30, 31, 40). Ancient proviruses are regarded as retroviruses that integrated into the germline prior to speciation, while modern proviruses are

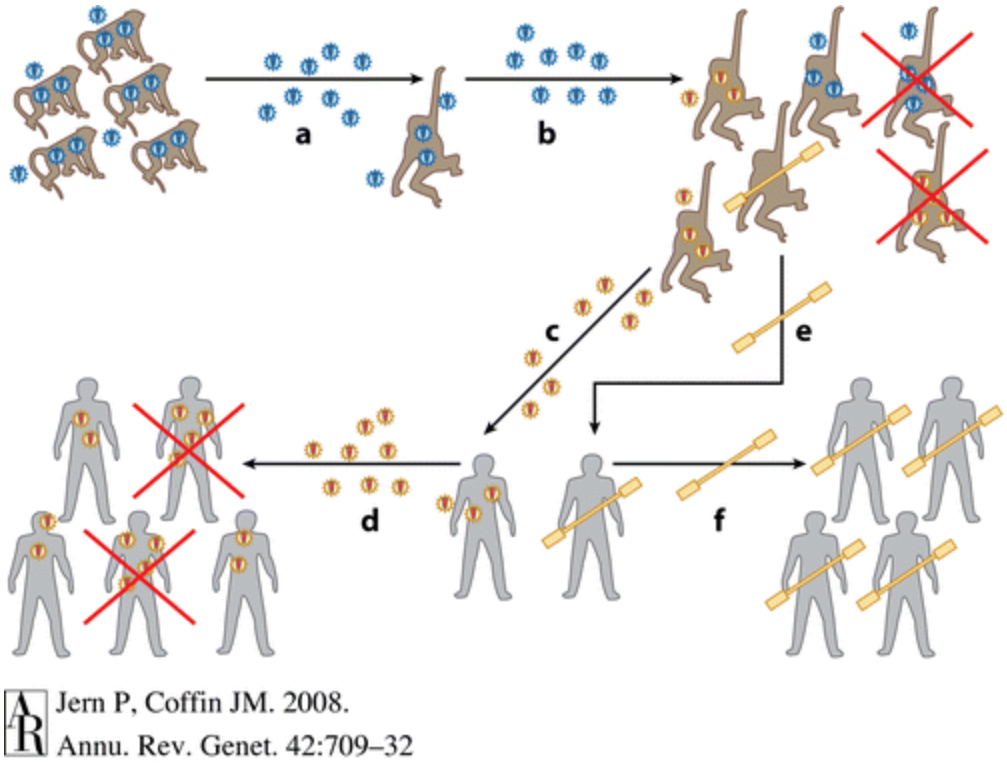


Figure 1.11: Modes of transmission for endogenous and exogenous retroviruses.

Transmission of a virus from host to host (A and B, blue virions) can result in lethality (red X), continued vertical transmission of a virus (A and B, D), zoonotic transmission (C, yellow virions), or endogenization (E, yellow bar) upon infection of a gamete which can be passed on in an Mendelian fashion (F). Reprinted from the Annual Review of Genetics [9] (41). Permission was not needed for a dissertation.

[9] Jern, P. and Coffin, J.M. 2008. Effects of Retroviruses on Host Genome Function. Annual Review of Genetics. 42(2008): 709-732. DOI: 10.1146/annurev.genet.42.110807.091501

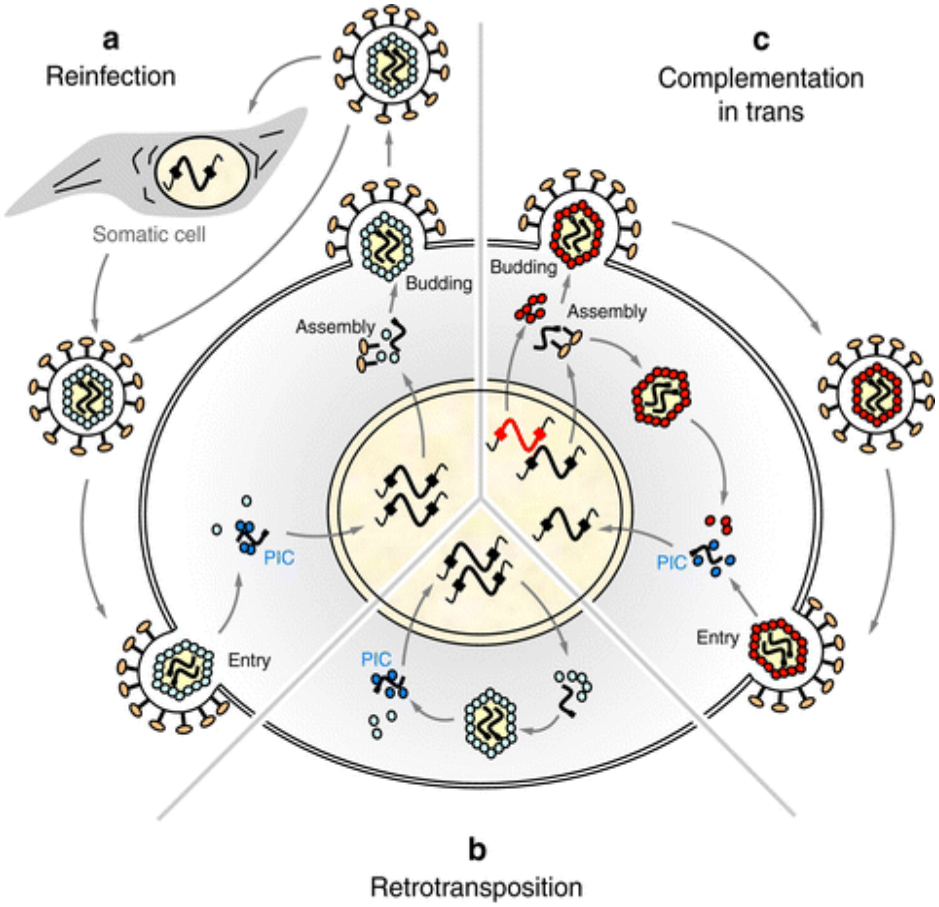


Figure 1.12: Potential methods of ERV multiplication.

ERVs can re-infect their host through reinfection (A), retrotransposition (B), and complementation in trans (C). (A) An endogenous retrovirus (black bars) can produce an infectious virion that could reinfect the host through re-infection of other gametes or through reinfection via a somatic cell. (B) An endogenous provirus could re-integrate into the host (C) An endogenous provirus could use another infectious exogenous retrovirus (red bar) to complement defects within its genome and produce a new infectious virus. Reprinted from Annual Review of Genomics and Human Genetics [10] (22). Permission was not needed for a dissertation.

[10] Bannert N., and Kurth R. 2006. The Evolutionary Dynamics of Human Endogenous Retroviral Families. Annual Review of Genomics and Human Geneics. 7:149-173. DOI: 10.1146/annurev.genom.7.080505.115700

regarded as retroviruses that integrated into the germline after speciation (20, 30, 31, 40). This can be determined by the presence of proviruses within a species and related species at the same location while comparing the approximate integration time of the provirus to relevant speciation dates through LTR divergence based on the host mutation rate and presence in related species (20, 30, 31). Once a provirus is formed, it is a permanent fixture in the host and its offspring: it will be present in the same location in every cell of an individual. Some proviruses that do not cause harm to the host - or can be beneficial to the host - can be passed down from parent to offspring in a Mendelian fashion. At the time of integration, the 5' and 3' LTRs are identical yet over evolutionary time will diverge due to the host mutation rate (3). Due to their advanced age, ancient proviruses are degraded through deletions, frameshifts, and premature stop codons that prevent the production of infectious virus (19, 20, 30, 31). Likewise, modern proviruses are typically more intact with the ability to produce protein products and virions, which are potentially infectious (20, 30, 40). Because of their youth, these proviruses are typically polymorphic in a species (20, 30, 40).

There are several examples of ERVs that are capable of expression in their host species (20, 22). The most well studied avian ERVs are found in domestic chickens and are closely related to ASLV (54). One such ERV, dubbed Rous-associated virus (RAV-0) which is capable of releasing virions, is exclusively found in domestic chickens and red jungle fowl indicating a relatively recent integration (20, 53). Endogenous Jaaksetie Squamous Carcinoma Retrovirus (enJSRV) is also closely related to exogenous JSRV, indicating that enJSRV recently integrated (20, 21, 55). Unlike its exogenous counterpart, enJSRV does not correlate with ovine pulmonary carcinoma and is currently thought to protect sheep from pulmonary carcinoma by blocking cellular entry of exogenous JSRV (19, 20, 30, 31, 56, 57). Mice - which have several ERVs - contain

endogenous proviruses originating from ecotropic (mouse only) and xenotropic (mouse and non-mouse) MLVs which can release infectious virions (20). Finally, a more recent example of endogenization has been occurring within the koala population, where the spread of exogenous and endogenous Koala retroviruses (KoRVs) allows virologists to track the process of endogenization in real time (21 - 24, 26, 27, 48).

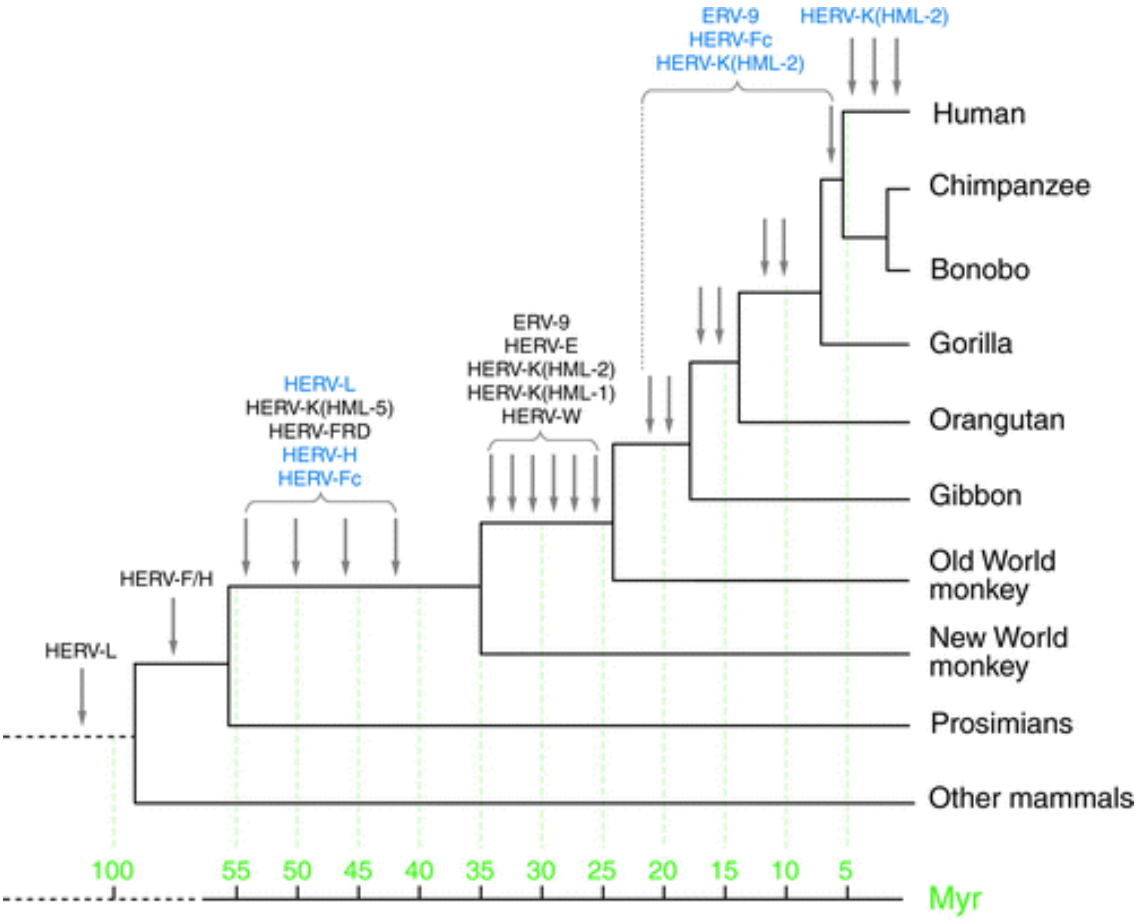
In some instances, ERVs are responsible for cancer onset. Bittner observed that female mouse pups from a high incidence background would still develop mammary adenocarcinoma when fostered on mice of a low incidence background. This suggested the presence of an inherited and infectious virus - later identified as MMTV - capable of causing murine breast cancer which Peter Bentvelzen would later confirm (1, 19, 20, 58 - 60). While feline ERVs such as the endogenous feline leukemia virus (FeLV) - which is also very similar to its exogenous counterpart - are not infectious, recombination between an endogenous FeLV and the exogenous FeLV subgroup A (FeLV-A) results in a *de novo* synthesized FeLV-B which causes neoplastic disease in infected animals (19, 20, 25, 31, 33, 61, 62).

The story of human ERVs (HERVs) detection begins with the identification of cancer-causing ERVS in vertebrates, especially in mice (20, 38, 63, 64). The identification of replication competent cancer-associated ERVs drove research to look for retroviruses that were also associated with human cancer (22, 38). The first HERVs cloned were identified by using hybridization probes for Southern blot analysis designed from conserved *pol* regions specific for MMTV (63, 64) and the hamster endogenous retroelement intracisternal type A particles (IAP) (65) under relaxed conditions (20, 22, 38). Since then, more sensitive methods - PCR and deep sequencing, for example – have been used to detect HERVs in non-human primates and humans. These methods have shown the presence of numerous distinct HERV clades across the primate lineage

with many HERV sequences integrating approximately tens of millions of years ago (22, 34 - 38, 41, Figure 1.13, Figure 1.14).

Several HERV subclades are very old, heavily mutated, and therefore have a low likelihood of producing infectious virus (22, Figure 1.13, Figure 1.14). In this respect, the subclade known as HERV-K (Human MMTV-Like (HML)) is unique. HERV-K (HML) received its name due to the binding of a lysine tRNA to the PBS during reverse transcription and due to the detection of these sequences with the above-mentioned MMTV *pol* probes (20, 22, 38, 41, 63-65). HERV-K (HML) is broken down into 11 subclades numbered 1-11 which integrated into the primate genome at different times and have had many different integration events throughout primate evolution (20, 22, 34, 36, Figure 1.13). For example, the subclade HERV-K (HML-2) has had ~ 1000 fixed integration events starting approximately 35 million years ago (MYA) up to as recent as 100,000 years ago (21, 22, 34, 36, 41).

Of these different HERV-K (HML) subclades, I am most interested in HERV-K (HML-2), which I will refer to as HML-2 (Figure 1.13). As stated above, HML-2 began colonizing the primate genome approximately 35 MYA and as recently as 100,000 years ago, making them the subclade containing the youngest HERVs (21, 22, 34, 36, 38, Figure 1.13). As a result, the most recently integrated and human-specific primate proviruses are within this subclade. Due to their youth, some HML-2 proviruses have not been devastated by time and possess LTRs that are nearly identical and possess ORFs potentially capable of encoding Gag, Pro, Pol, and Env proteins (20, 34, 36, 38). That some HML-2 proviruses contain intact ORFs indicates that some of these proviruses are potentially capable of producing HML-2 - specific proteins that could participate in pathology, protect the host from infection by similar viruses, or produce infectious virus that would facilitate their further spread in the population. This last line of thinking was



Bannert N, Kurth R. 2006.
Annu. Rev. Genomics Hum. Genet. 7:149–73

Figure 1.13: Primate phylogenetic tree with HERV integration times.

Phylogenetic tree of primate divergence over millions of years (MYR). Integration times for various HERVs are marked with grey arrows. The length of time (X axis) ranges from modern day (right) to 100s of millions of years ago (left). Primate species represented in this tree are present on the right with their divergence shown within the figure. Reprinted with permission from Annual Review of Genomics and Human Genetics [10] (22). Permission was not needed for a dissertation.

[10] Bannert N., and Kurth R. 2006. The Evolutionary Dynamics of Human Endogenous Retroviral Families. Annual Review of Genomics and Human Geneics. 7:149-173. DOI: 10.1146/annurev.genom.7.080505.115700

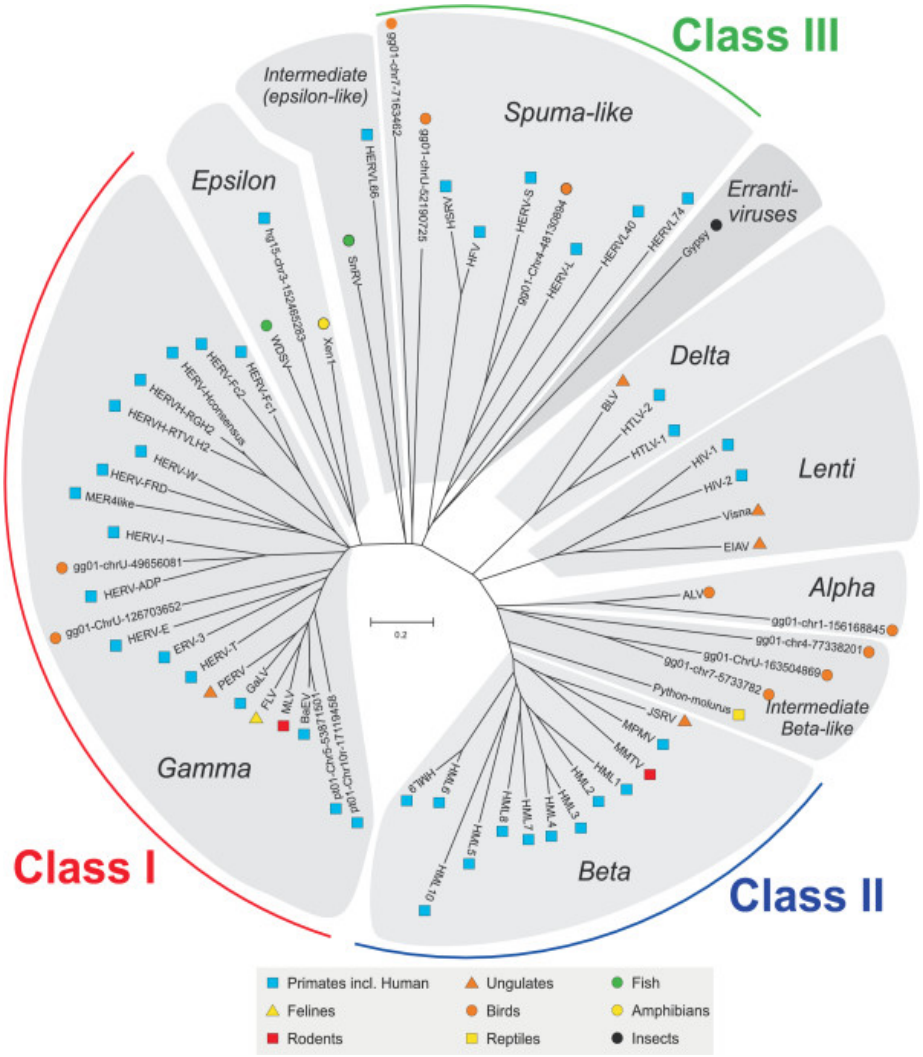


Figure 1.14: Representative dendrogram of the seven known retroviral groups.

The dendrogram is an unrooted neighbor joining tree of Pol showing both endogenous and exogenous members of each group. The classes of endogenous retroviruses (i.e. class I, II, and III) are represented on the periphery of the dendrogram. Symbols indicate the species each virus is found in, while viral sequences that were new at the time of this figure's home paper's publication are indicated by their chromosomal location. Reprinted from Retrovirology [11] (159). Since the original article is open-access, no additional permissions are required.

[11] Jern, P., Sperber, G.O., and Blomberg, J. 2005. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology. 2:50. DOI:https://doi.org/10.1186/1742-4690-2-50

further encouraged due to their polymorphic nature in humans and the presence of human-specific proviruses (22, 35, 36, 42).

Recently, the Coffin lab and others have looked for the presence of infectious HML-2s (34-37). There are at least 96 known HML-2 proviruses and at least 976 solo LTRs, with 91 of these known HML-2 proviruses present in the hg19 human genome build (34-37). Each of these proviruses are known by several different names, but I will refer to them based upon their chromosomal location (34-36). Although ERVs from other species can produce infectious virions, this does not appear to be the case with HERV-K (HML-2) (36).

1.5 Epigenetic Regulation

HERV expression is regulated like every other gene within the human genome. Beyond the role of repressors, cells regulate HERV expression through epigenetics Epigenetics received its name in 1942 from Conrad Waddington who was interested in the mechanism that connected genetic code and adult phenotype (73). Since then the field has evolved significantly from its definition to the inclusion of several mechanisms that collaborate to regulate genetic expression. The modern definition of epigenetics refers to inheritable differences from cellular or parental origin in cellular phenotype due to changes in gene regulation that do not affect Watson-Crick base pairing. Time has shown that epigenetics is critical in normal development - such as pluripotency and pregnancy - and plays a large role in diseases - such as depression and various cancers (66, 67, 71). Inheritance of epigenetic modifications - in particular, DNA methylation on CpG dinucleotides - is strongly conserved and maintained over several generations. For example, inheritance of methylated and unmethylated regions of DNA can be propagated upwards of 100 cell divisions *in vitro* (67, 74).

Epigenetic modifications regulate gene expression by inducing conformational changes to chromatin structure that allow or restrict access of transcriptional machinery through many complex mechanisms. The most well studied mechanisms involve DNA methylation of CpG dinucleotides and histone post-translational modifications with nucleosome rearrangement and RNA-mediated gene silencing being relatively new to the field (66, 67, 69, Figure 1.15). It should be emphasized that none of these modifications act in isolation: each modification relies on other modifications to create a phenotype. Furthermore, while there are general understandings of how these modifications work there are some exceptions. Both exceptions and examples of epigenetic mechanisms working collaboratively will be specified throughout this section.

DNA methylation is primarily used to silence genes by restricting transcription machinery accessibility. It results from the covalent addition of a methyl group to the 5-carbon position of cytosine in CpG dinucleotides by one of three DNA methyltransferases (DNMTs): DNMT1, DNMT3a, and DNMT3b (67). DNMT1 is responsible for “maintenance” methylation by targeting regions of hemimethylation to add methyl groups to the unmethylated daughter strand during replication while DNMT3a and 3b are responsible for *de novo* methylation during early development (67).

DNA methylation plays an important role in cellular functions, such as the silencing of transposable elements (71, 75-79), embryonic development (80), imprinting (81), and X-chromosome inactivation (82) (67, 69). Typically, regions that are heavily methylated are low in transcriptional activity while those that are low in methylated CpGs are high in transcriptional activity (71). As a result, methylated CpG dinucleotides are not found in promoter regions of genes active in healthy cells (67). This decrease in transcriptional activity is thought to be due to blockage of transcriptional machinery from binding to DNA promoters and due to recruitment of histone modifying enzymes that silence portions of chromatin (66, 67). However, there are instances where

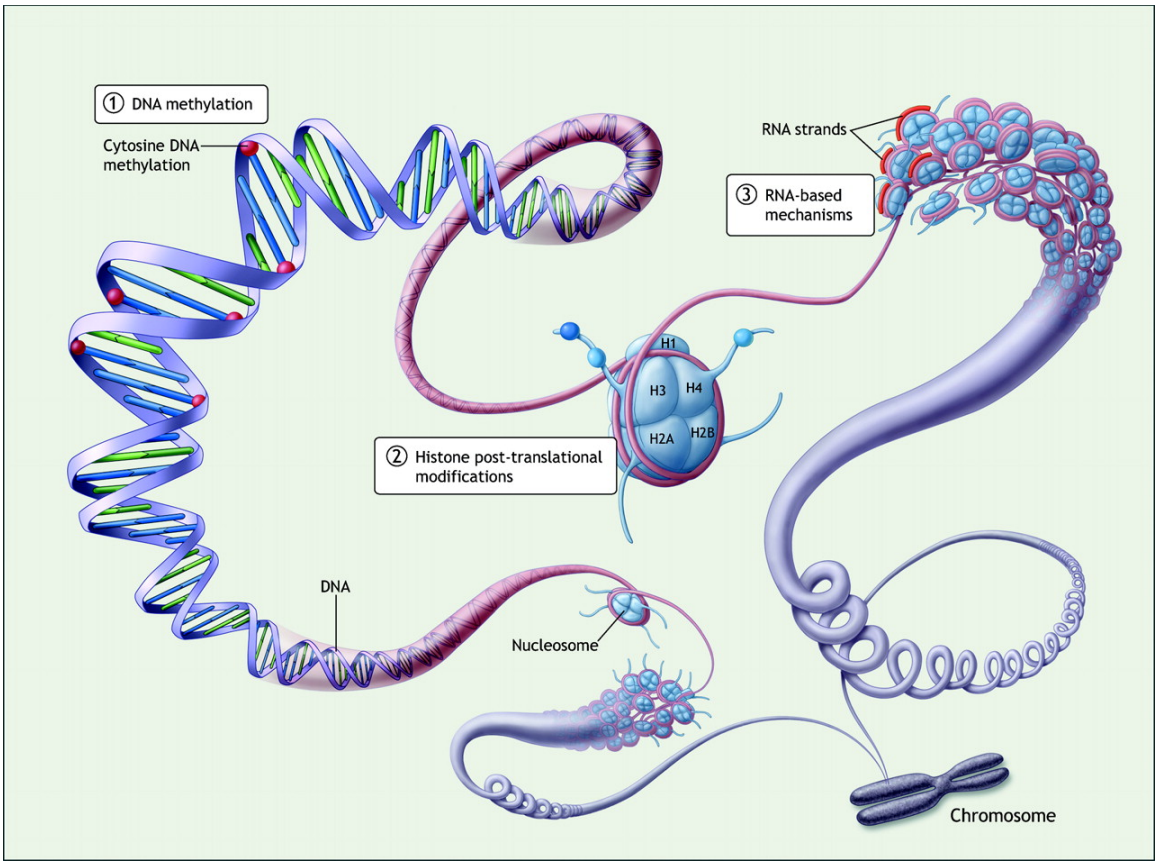


Figure 1.15: Overview of epigenetic mechanisms.

Diagram representing three of the most commonly explored mechanisms behind epigenetic regulation overlayed with different structures of DNA packaging: DNA methylation (1), histone post-translational modification (2), and RNA-based mechanisms (3). (1) DNA is subjected to DNA methylation (red ball) at CpG dinucleotides to restrict transcription machinery binding. (2) DNA wound around histone proteins (H2A, H2B, H3, and H4), and neighboring nucleosomes are subject to further modifications through histone post-translational modification and nucleoside rearrangement. (3) RNA - mediated silencing is a more recent discovery in epigenetic regulation through site specific binding of non-coding RNAs to inhibit the transcription or translation of gene products. Reprinted with permission from Circulation Research [12] (154).

[12] Matouk, C.C., and Marsden, P.A. Epigenetic Regulation of Vascular Endothelial Gene Expression. Circulation Research. 2008;102:873-887. DOI: https://doi-org.ezproxy.library.tufts.edu/10.1161/CIRCRESAHA.107.171025

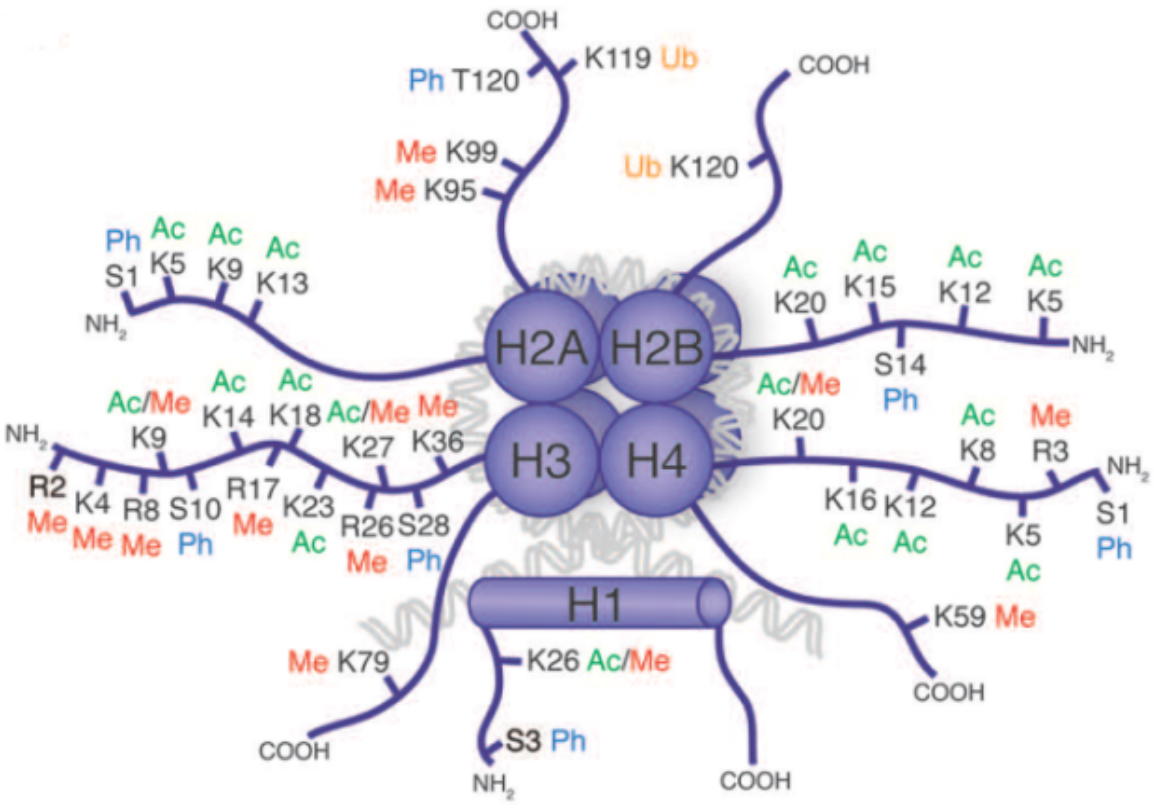


Figure 1.16: Histone post-translational modifications:

A schematic of a nucleosome displaying histone proteins H2A, H2B, H3, and H4 in their dimeric octomer. DNA (grey) is shown wound around the octomer and bound by histone H1 to form a nucleosome. Histone tails are extended into the nuclear lumen where modifications such as acetylation (green), methylation (red), phosphorylation (blue), and ubiquitination (yellow) are shown. Adapted from Epigenetics [13] (68). Adapted to include only panel A. Paper is open source, so permission was not required.

---

[13] Tollervey, J.R., and Lunyak V. 2012. Epigenetics: judge, jury, and executioner of stem cell fate. Epigenetics. 7(8): 823-840. DOI: 10.4161/epi.21141

hypermethylation over a repressor binding site, hypermethylation on promoters for repressors, and transcription factors capable of binding to methylated sites can allow for increased expression of a nearby gene product (69, 70).

Histone post-translational modifications are more diverse and can have a wide range of results on chromatin structure (66, 68, Figure 1.16). Histone proteins - H2A, H2B, H3, and H4 - assemble into a dimeric octomer and wind DNA into a spool-like structure which is secured by the H1 histone protein to form a nucleosome - the basic unit of eukaryotic DNA packaging (Figure 1.16). Each protein within the dimeric octomer extends a tail into the nuclear lumen which can be modified by a diverse set of enzymes (69, Figure 1.16). Known modifications consist of acetylation, methylation, phosphorylation, and ubiquitination which affect transcriptional machinery access to gene promoters through structural modifications of chromatin (Figure 1.16). For example, histone acetylation of H3K9 or H3K14 results in gene activation through relaxation of chromatin structure to allow transcriptional machinery access to DNA (67, 71).

Acetylation of histones is straight forward in that once acetylated the neighboring genomic landscape relaxes around the histone to allow for transcriptional machinery binding. In this respect, acetylation is commonly referred to as an activating modification. However, once you start to consider other histone modifications such as methylation things get more complicated (69). In the case of histone methylation, these modifications can be either repressive or permissive and the modified sites can be mono, di, or tri-methylated (69). For example, tri-methylation of H3K27 or H3K9 is associated with gene silencing while trimethylation of H3K4 and H3K36 are associated with gene activation (69). This brings up an interesting point where residues subject to more than one modification can have two different results. For example, acetylation of H3K9 results in activation while methylation of the same residue results in inactivation.

There are two other mechanisms of epigenetic modification that I would like to mention but to go into detail would go beyond the scope of this dissertation: non-coding RNAs and nucleosomal rearrangement. Some genes when transcribed produce short RNA transcripts that will bind to DNA or protein to prevent the expression of gene products (66). Non-coding RNAs can be segmented into small (under 200 nucleotides) and large (> 200 nucleotides) but are assigned into subgroups such as microRNAs (miRNAs), long non-coding RNAs (lncRNAS), small nucleolar RNAS (snoRNA), PIWI-interacting RNA (piRNA), and small interfering RNA (siRNA) (69, 71). Each of these non-coding RNAs has been increasingly recognized as being vital to normal cellular function and have been found to be aberrantly expressed in cancer (69, 71). For example, there is an inverse correlation between expression of the histone methyltransferase EZH2 and expression of miRNA-101. In breast cancer cell lines and prostate cancer tumors, and overexpression of EZH2 is correlated with increased histone methylation and a decrease in miRNA-101 expression (66, 83). Finally, nucleosomal rearrangement can block or permit access of transcriptional machinery to DNA depending on their placement throughout the genome.

To get effective regulation of the genome, all these modifications work together (66, 67, Figure 1.17, Figure 1.18). For example, silencing a gene through CpG methylation of DNA can recruit binding of methylation binding proteins (MBPs) which can trigger the recruitment of histone deacetylases (HDACs) (66, 67, Figure 1.17). This recruitment will remove acetyl groups on histones, triggering a conformational change in chromatin, restricting transcriptional machinery access to DNA. Furthermore, DNMTs can be recruited to DNA by histone modifying proteins like heterochromatin protein 1 (HP1) to silence genes (Figure 1.17).

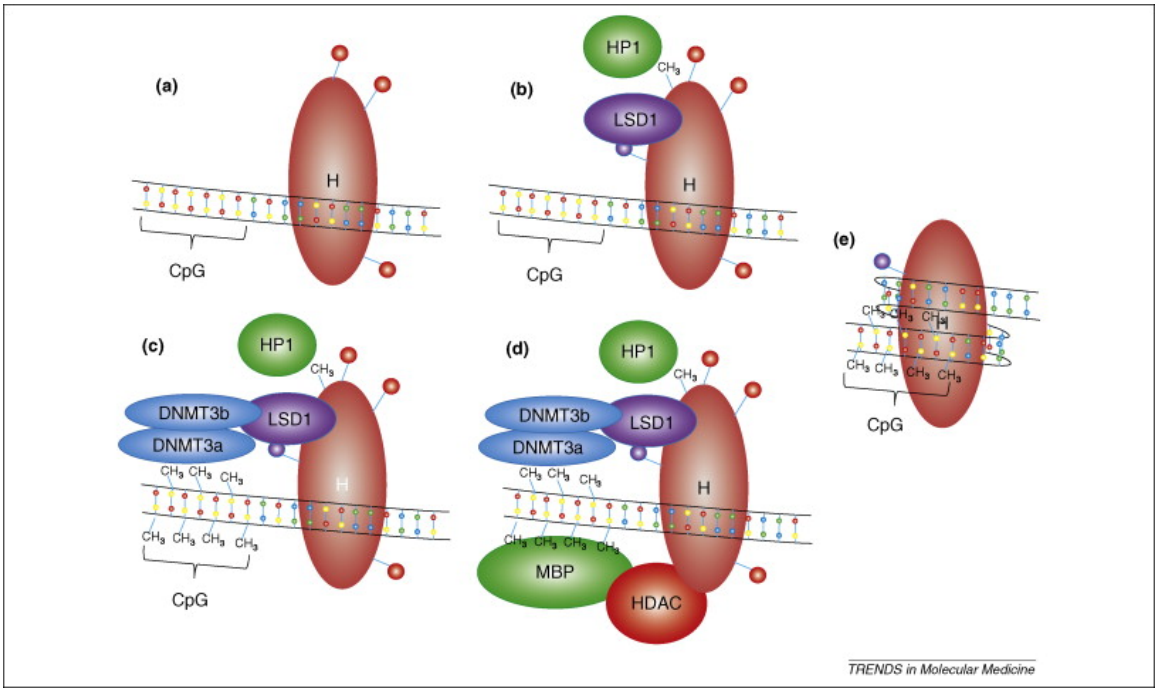


Figure 1.17: Interplay between histone modifying and DNA methylating enzymes.

The above schematic is included to illustrate one example of the interplay between histone modifying and DNA methylating enzymes. (A) a stretch of unmethylated CpGs are located near an acetylated histone octomer (red oval). (B) HP1 and LSD1 bind to the histone octomer for (C) recruitment of DNMT3A/B to methylate CpGs. (D) Methylation binding proteins (MBP) bind to methyl groups and attract histone deacetylases (HDACs) to remove acetyl groups (red circles) from histone tails, resulting in (E) supercoiling of DNA around histones and silencing of neighboring genes. Reprinted with permission from Trends in Molecular Medicine [14] (66).

[14] Handel, A.E., Ebers, G.C., and Ramagopalan, S.V. 2010. Epigenetics: molecular mechanisms and implications for disease. Trends in Molecular Medicine. 16(1): 7-16. DOI: https://doi.org/10.1016/j.molmed.2009.11.003

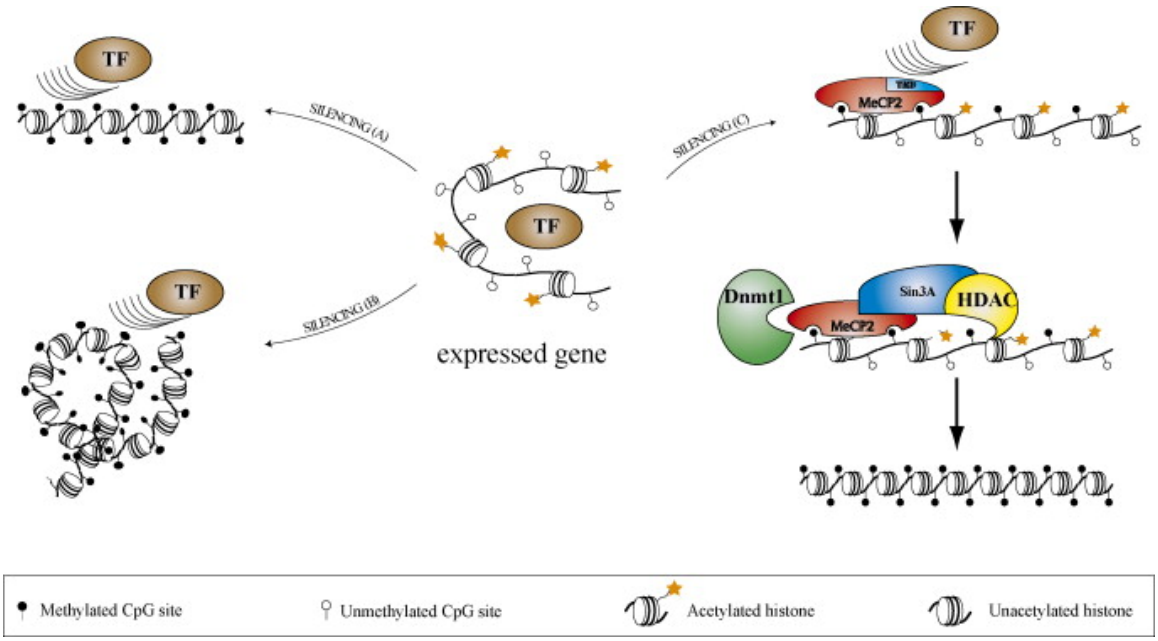


Figure 1.18: Blocking of transcription factors by various mechanisms.

The above schematic is included to illustrate the different mechanisms used to block transcription factor binding to a promoter, enhancer, or repressor element. Transcription factor binding can be blocked through heavy methylation of CpGs (black circles) (A), chromatin structure (B), or through binding of methylation binding proteins (MBPs) such as MeCP2 (C). MBPs such as MeCP2 can then recruit HDACs and DNMT1 to remove acetylation on histones (yellow star) and add methyl groups to unmethylated CpG sites (white circles) Reprinted with permission from Mutation Research [15] (67).

[15] Vaissiere, T., Sawan, C., and Herceg, Z. 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research. 659(1-2): 40-48. DOI: 10.1016/j.mrrev.2008.02.004

As already alluded to, when epigenetic mechanisms go awry disease can occur (66, 67). Epigenetic dysregulation has been reported in mood disorders (84, 85), arthritis (86), asthma (87), and especially in cancer (66). Two mechanisms of epigenetic dysregulation in cancer are aberrant global DNA methylation and histone acetylation, yet as previously mentioned other mechanisms such as non-coding RNAs also play a role (66, 67, Figure 1.19). For example, hypermethylation of *MLH1* - which encodes a DNA mismatch repair enzyme - predisposes multiple generations to colon cancer due to its inactivation (66, 88). Furthermore, aberrant DNA methylation is characteristic of leukemias and lymphomas, and genome-wide screens of such cancers revealed several genes that exhibited differential DNA methylation compared to their non-cancerous counterparts (66, 89).

One characteristic of cancer is global hypomethylation relative to their non-cancerous counterparts was first established in 1983 (90), and since then genome-wide DNA methylation dysregulation (i.e. genome-wide DNA hypomethylation with site-specific DNA hypermethylation) has been reported in several cancers (71, 91-96). Site-specific DNA hypermethylation typically involves CpG islands in promoter regions and results in transcription inactivation (71). This type of modification affects genes within major cellular pathways such as DNA repair, cell cycle control, Ras signaling, apoptosis, detoxification, hormone response, and p53 network, resulting in a greater growth advantage (71).

As for histone modifications in cancer, one of the most well explored is changes in acetylation. In various cancer types, such as prostate, colon, lung, liver, and stomach HDACs are over-expressed resulting in loss of acetylation and a decrease in local gene expression (71, 97-102). Furthermore, aberrant expression of histone methyltransferases (HMTs) and histone demethylases (HDMs) have been reported in

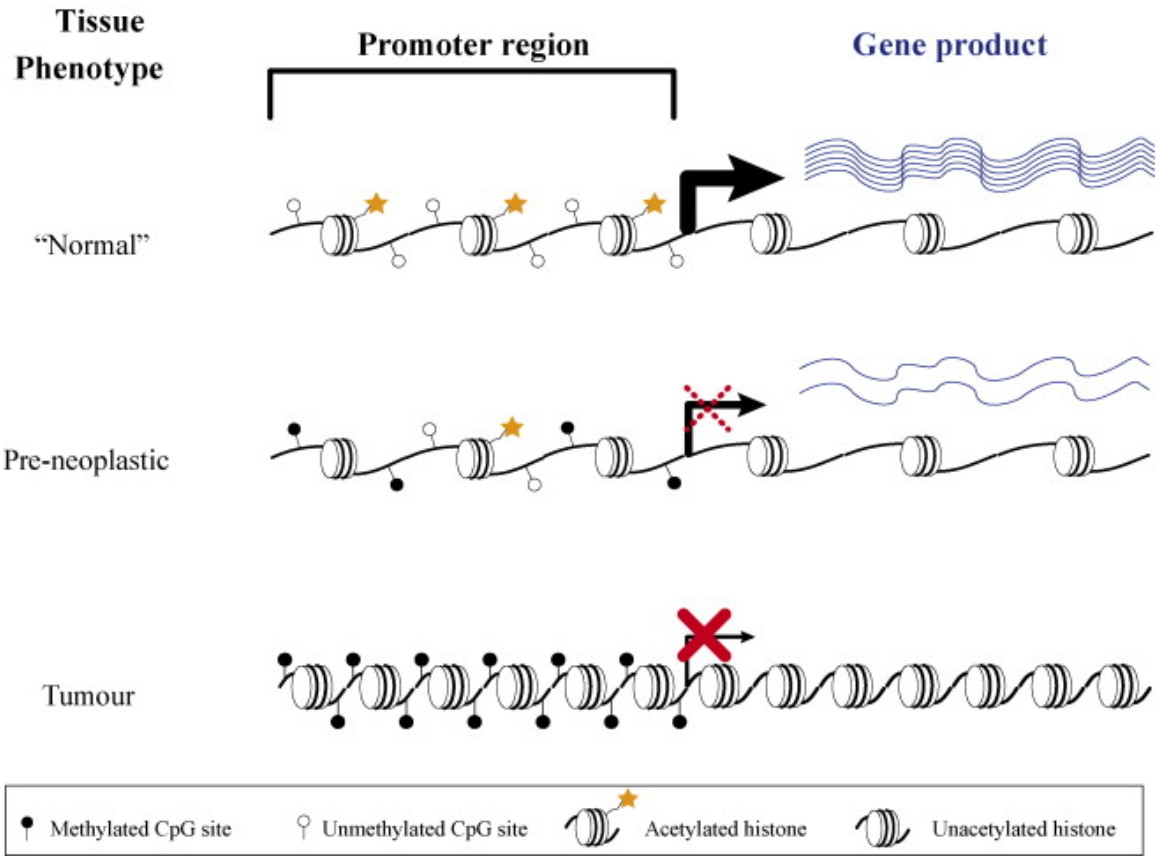


Figure 1.19: Epigenetic steps involved in gene product inhibition in cancer.

The above schematic is included to illustrate how changes in DNA methylation and histone acetylation can affect gene expression (blue). In "normal" state, active promoter regions are characterized by unmethylated CpGs (white circles) and acetylated histones (yellow stars). In the pre-neoplastic state, these modifications are altered (reduced histone acetylation and/or increase in DNA methylation (black circles)) and reduce the production of downstream gene products. In the cancerous state, acetyl groups are removed and CpG dinucleotides are heavily methylated, blocking production of the downstream gene product. Reprinted with permission from Mutation Research [15] (67).

[15] Vaissiere, T., Sawan, C., and Herceg, Z. 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research. 659(1-2): 40-48. DOI: 10.1016/j.mrrev.2008.02.004

various cancer types (71). For example, inactivating mutations in the histone methyltransferase *SETD2* and the histone demethylase *UTX* and *JARID1C* were reported in clear cell renal carcinoma (71, 103). Finally, *EZH2* over-expression in melanoma, prostate, breast, and endometrial cancer is correlated with malignancy, and knockdown of EZH2 in ovarian cancer induced apoptosis and suppressed invasion (71, 104, 105).

HERVs – which are regulated by mechanisms used for regulation of protein coding genes and being transposable elements - are naturally subject to epigenetic regulation. Through reinfection of the host organism, transposable elements can create genomic instability by spreading throughout the host genome (72). Interestingly, DNA hypomethylation in tumor cells also affects repetitive regions of the genome which results in genome instability (71). In the case of retroviruses, LTRs - which contain several transcription factor binding sites (TFBS) - are problematic for host cells because they can promote the expression of both ERVs and downstream sequences (72). This can lead to neoplasia in infected cells due to aberrant expression of proto-oncogenes.

It is important to note that LTRs can serve as promoters for downstream sequences (72). It is also important to note that these LTRs can still function as enhancer elements when present in sense and antisense orientation to the rest of the genome (72). Therefore, placement of LTRs around proto-oncogenes or insertion of ERVs into tumor suppressor genes - intronic or exonic - can be detrimental for host health. Additionally, LTRs can affect malignancy by interfering with epigenetic regulation of non-coding RNAs. For example, one report displayed a correlation between the number of very long non-coding RNAs (vlincRNAs) expressed from ERV promoters and the severity of malignancy (106). Thus, HERVs are not only capable of affecting the expression of protein-coding genes but also the expression of lincRNAs to the detriment of the host (72, 106).

Typically, epigenetic regulation of HERVs can be achieved through DNA methylation and histone deacetylation (72). In healthy tissue, HERVs are heavily methylated and this is lost in cancerous tissue (72, 107, 150). For example, HML-2 expression in melanoma cell lines is correlated with the amount of methylation present on their 5' LTR, and their expression can be induced with treatment of 5-aza-2′-deoxycytidine, a DNA methyltransferase inhibitor that incorporates into DNA upon replication which can then irreversibly bind to DNMTs (150). Furthermore, the lack of methylation of HML-2 5' LTRs correlates with HML-2 expression in the germ cell tumor cell line Tera-1 and methylation of 5' LTRs *in vitro* correlates with decreased expression (72, 108). This form of epigenetic regulation seems to be the most important in HERV regulation as treatment with HDAC inhibitors alone in HIV-infected cells and 293s did not significantly up-regulate their expression (72, 109).

Due to global methylation dysregulation in cancer, it is possible that increased HERV expression in cancer is related to hypomethylation around HERV LTRs (72). For example, global hypomethylation of HERV-W and LINE-1 were reported in ovarian cancer and hypomethylation of HERV-K 5' LTRs are seen in melanoma (72, 110, 111). However, upregulation of all HERVs does not seem to be the case for all cancers (72, 112). Treatment of neuroblastoma cell lines with 5'-azacytidine induced expression of HERV-W loci, indicating that there are some cancer cell lines where HERVs are suppressed by CpG methylation (72, 112).

Additionally, histone methylation or deacetylation in cancer can contribute to HERV activation (72). This was described recently in cancer cell lines where an increase in HERV-Fc1 expression was associated with active histone methylation (72, 113). Some HERV LTRs like that of ERV9 are easily induced by HDACi and are highly expressed in testes (72, 114, 115). This LTR regulates the expression of a tumor suppressor protein called GTAp63, which is absent in germ cell tumors but when cells

are treated with HDACis GTAp63 expression is induced which triggers apoptosis (72, 114, 115). Interestingly, this effect was not seen with 5'-aza treatment so CpG methylation is not required for this LTR regulation (72).

1.6 Viruses and Cancer

As was previously mentioned, the early days of retrovirology helped shape oncology (116). The discovery of RSV led to the detection of the first proto-oncogene which was later found to originate from host cells (116). This led to an understanding of molecular mechanisms in all cancers with and without a retroviral origin by studying how proto-oncogenes cause cancer (116). Through the detection of proto-oncogenes which are often involved in cell cycle regulation and growth, retroviral carcinogenesis has led to a greater understanding in normal cellular mechanisms and growth pathways (116).

Also mentioned earlier in this chapter was the connection of retroviruses to cancer. This was first established with Ellerman, Bang, and Rous' discoveries of ASLV, isolated from avian sarcomas and leukosis (1, 116). Retroviruses isolated from other mammalian cancers such as mouse and – eventually - primate were isolated between the 1950s and 1980s. MLVs and FeLVs exist in both endogenous and exogenous forms and are capable of inducing leukemia (116). JSRV and MMTV also exist in exogenous and endogenous forms, but enJSRV is protective of the host against its exogenous cancer-causing counterpart while MMTV can cause mammary carcinoma in both its endogenous and exogenous form (116).

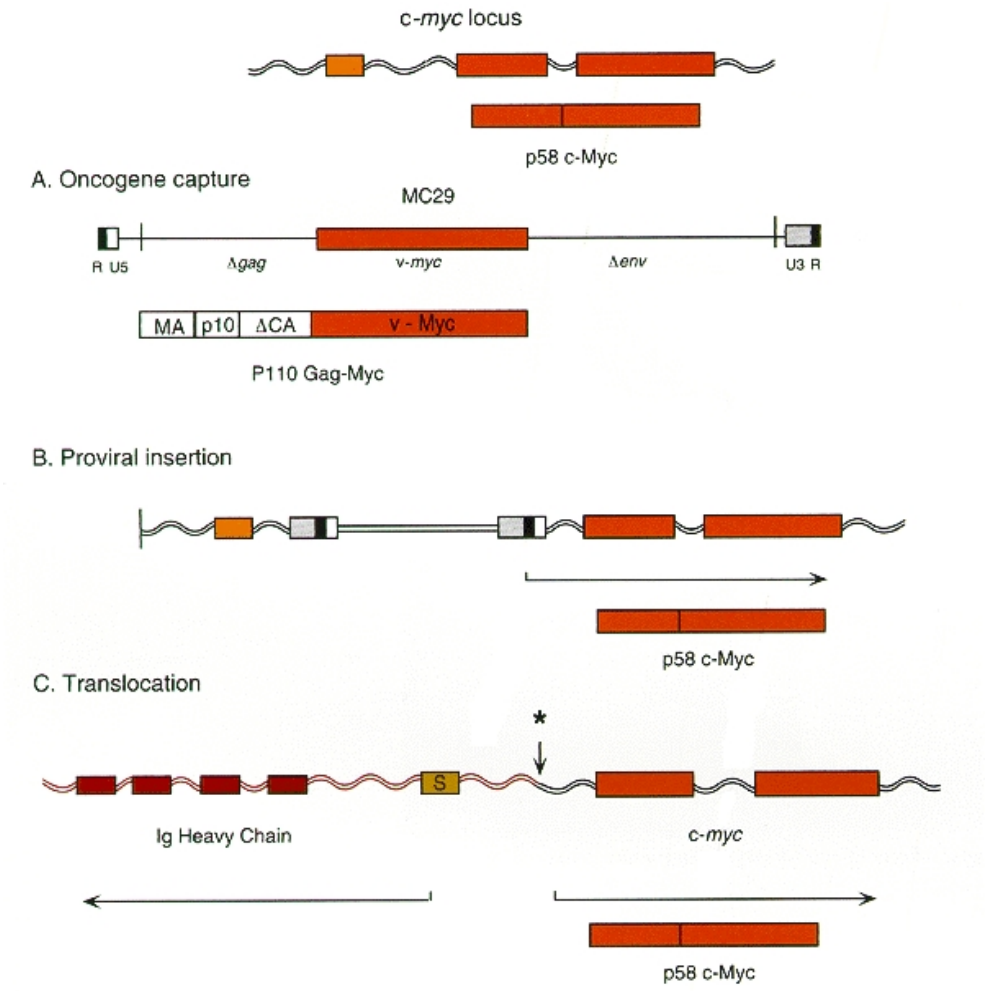


Figure 1.20: Cancerous mechanisms behind proto-oncogene expression.

This example focuses on the *c-myc* locus (top) with its protein product displayed beneath it. This illustration mentions one of several possible outcomes for each mechanism. (A) Oncogene capture by a retrovirus, such as in the case of the MC29 virus, can lead to overexpression of the oncogene through retroviral replication. (B) Proviral insertion upstream of a proto-oncogene or within an intron (as seen above) can drive expression of the proto-oncogene, resulting in overexpression. The above case illustrates ALV integrating into an intron of *c-myc* in sense orientation, which occurs in bursal lymphomas. (C) Translocation of *c-myc* to a region around the immunoglobulin heavy-chain locus switch, which can also result in overexpression of *c-myc* as seen in Burkitt's lymphoma. Reprinted with permission from Cold Spring Harbor Laboratory Press [16] (116).

---

[16]Rosenberg, N., and Jolicoeur, P. 1997. Retroviral Pathogenesis. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press

Retroviruses such as ALV can cause cancer in host cells through the acquisition of an oncogene from the cell or through proviral insertion adjacent or within a proto-oncogene or tumor suppressor (116). In this respect, the replication pathway of a retrovirus has the potential to be oncogenic. This can involve the removal of all ORFs, some of them, or a small fraction of one. To continue replicating, this virus will need another replication competent virus that will provide the necessary core and enzymatic proteins for transportation to a new host. This virus is known as a "helper" virus and has been reported with ALV strains such as AMV, E26-AMV, and AEV-ES4 (116, Figure 1.22).

In terms of their carcinogenic potential, both the helper and proto-oncogene - containing (also known as "transforming") provirus can transform cells. If the transforming provirus contains a proto-oncogene, this virus upon infection of a host will drive expression of their proto-oncogene to induce transformation, especially considering that after a few rounds of passage these viruses have evolved to be more efficient in their transforming capabilities. This mechanism typically has a quick onset in *in vitro* and *in vivo* models (116). The helper virus has a longer latency period and transforms cells through proviral insertion upstream of a proto-oncogene or within a tumor suppressor (116, Figure 1.18, Figure 1.21). This method results in a longer incubation period in that a provirus must insert into the "right" site in addition to the presence of other mutations that are required for cellular transformation. A retrovirus that inserts neighboring of proto-oncogenes could activate their expression through their 5' LTR either directly upstream or from within an intron. In either situation, it will lead to overexpression of the proto-oncogene within the host, yet insertion within an intron will lead to an altered form that is missing one or more exons. Expression of a proto-oncogene through a provirus' 5' LTR could also lead to alternate splicing into a hyperactive or inactive product. It is also

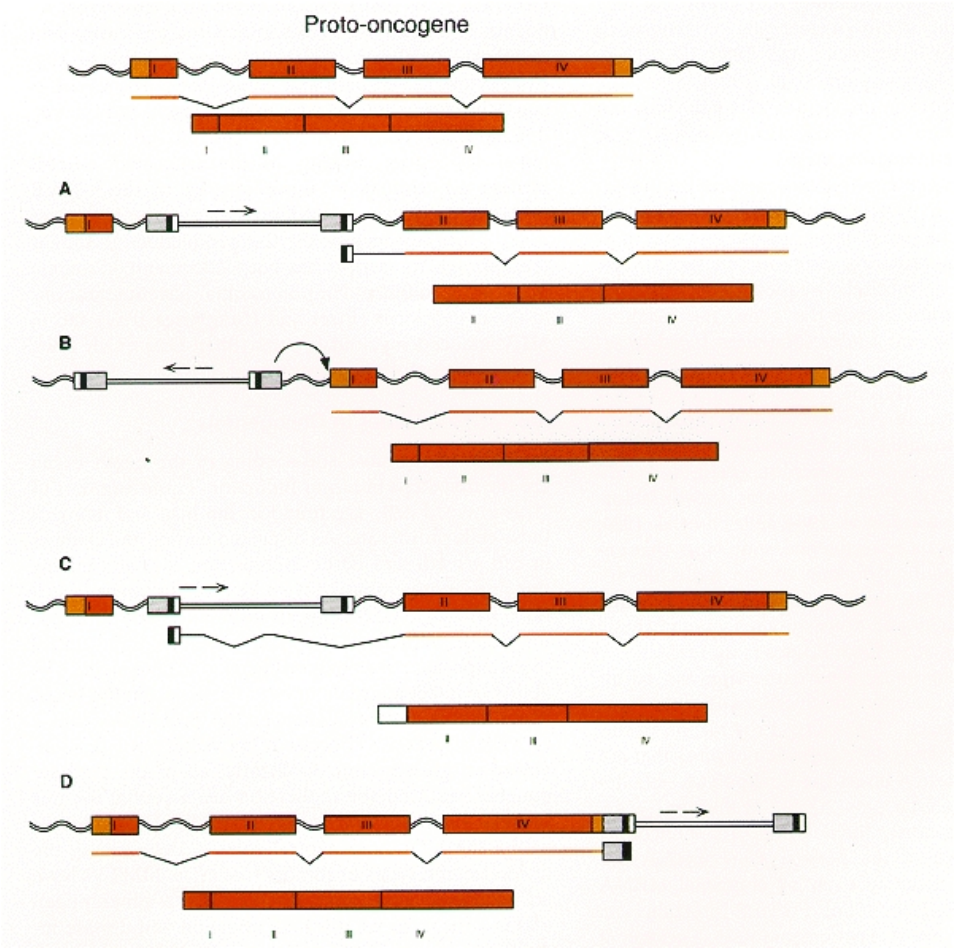


Figure 1.21: Mechanisms behind retroviral-induced oncogenesis via proviral insertion.

A representative proto-oncogene with regulatory elements (light orange), exons (dark orange), and introns (double lines) shown on top with its spliced and translated product underneath. The provirus is represented by grey boxes with an arrow to show the direction of its transcription. (A) Proviral insertion in between exons and the natural promoter. Transcription is driven by the 3' LTR of the provirus. (B). Activation by enhancer sequences within the 5' LTR. Even in this situation where the provirus is antisense to the proto-oncogene, it can still activate the proto-oncogene. (C) Activation by read through. Expression of the proto-oncogene is driven by the 5' LTR. (D) Activation by insertion of the provirus into regulatory elements of the proto-oncogene that stabilize mRNA. Reprinted with permission from Cold Spring Harbor Laboratory Press [16] (116).

---

[16]Rosenberg, N., and Jolicoeur, P. 1997. Retroviral Pathogenesis. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press

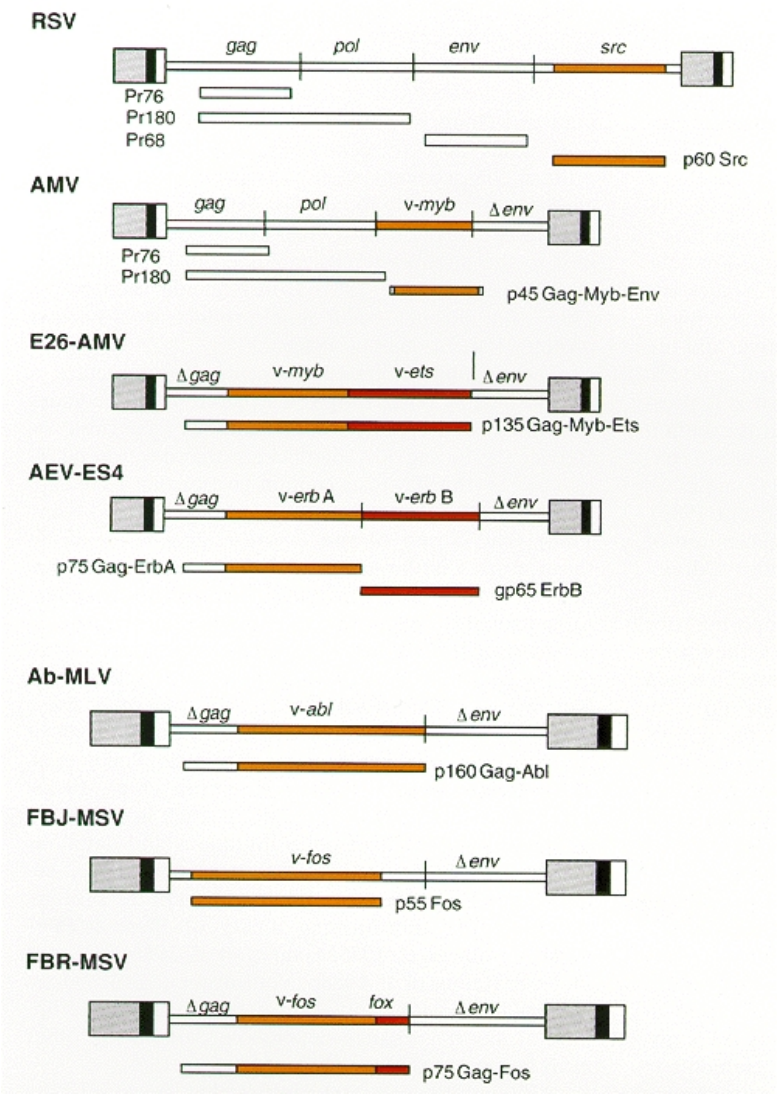


Figure 1.22: Structure of select oncogene-containing retroviruses and their protein products.

The genetic structure of oncogene-containing retroviruses and their representative protein products. Oncogenes are shown in color. Note that in cases where the same oncogene is captured, their placement in the retroviral genome varies. The top four sequences are derived from ALV while the bottom three are derived from MLV. Reprinted with permission from Cold Spring Harbor Laboratory Press [16] (116).

[16]Rosenberg, N., and Jolicoeur, P. 1997. Retroviral Pathogenesis. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY). Cold Spring Harbor Laboratory Press

possible that a retrovirus integrated antisense to a proto-oncogene could allow for enhancer binding to their LTR, allowing for increased expression of the proto-oncogene.

1.7 Melanoma

Skin cancer can be divided into melanoma and non-melanoma skin cancer. Altogether, skin cancer is segmented into basal and squamous cell carcinoma - the most common form of skin cancer -, Merkle cell, lymphoma of the skin, Kaposi's sarcoma, and melanoma (118). Thought the true number is difficult to estimate, skin cancer is one of the most diagnosed cancers in the United States. It was estimated in 2012 that the number of newly diagnosed cases of basal and squamous cancers alone was 5.4 million, which was a difficult estimation to make considering that reporting of this statistic is not required (117). Comparatively, melanoma is certainly not one of the more common forms of skin cancer - accounting for an estimation of 87,110 newly diagnosed melanomas in the United States in 2017 (119). The problem lies in its mortality rate, where it is responsible for the most skin cancer related deaths at an estimation of 9,730 for 2017 compared to basal and squamous carcinoma which is estimated to account for 2,000 (119, 120). Unlike basal and squamous carcinoma, melanoma incidence has also risen over the past 30 years (119, 120).

Early detection of melanoma has significantly improved the 5-year survival rate where patients with stage I have a greater than 90% survival compared to patients with metastatic disease who have less than 20% (122, 124, 128). Screening will typically evaluate factors that relate to the Asymmetry, Border, Color, Diameter, and Evolution of atypical moles. Screening will also look for an increase in the number of new moles or the total number of moles (>100) a patient has at the time of referral, as these patients are typically more at risk than others (122, 124, 125). If a mole has multiple colors - or

simply odd colors such as blue or white -, has asymmetrical borders, is larger than six millimeters, or is changing it is flagged for closer inspection. This ABCDE rule has become a common and easy way to remember these diagnostic criteria, allowing for patients with a family history of melanoma to monitor themselves at home (122).

Melanoma - sometimes referred to as malignant melanoma or cutaneous melanoma when cancer arises from the epidermis - arises from melanocytes, which we often think of as residing in the epidermal layer of skin (121). Melanocytes are derived from neural crest cells which migrate throughout the body to take residence in the epidermis and to a lesser extent eyes, hair, mouth, gastrointestinal tract, and genital mucosa (122, 123). Once there, they produce melanin to protect deeper layers of tissue from UV damage, producing more as an individual is increasingly exposed. Melanomas can continue to produce melanin - taking on a dark brown or black appearance -, but there are some instances where they do not and display a white, blue, or pink color as is the case with nodular melanoma (121, 122).

A typical way that melanoma can develop is through repeat exposure to UV radiation from either the sun or tanning beds (123). Melanomas that arise from chronic sun damage (CSD) are often found on the face, arms, and legs. Meanwhile, melanoma can also develop on areas that are routinely covered by the sun in individuals with a family history or predisposition (ex. red hair, blue eyes, fair skin, or more than 100 moles) (123). As a result, differences between CSD and non-CSD melanomas can be divided into placement on the body, accumulated DNA damage, host age, exposure to UV radiation, and types of mutations (123). CSD melanomas appear in older populations (> 55 years) where tumors contain a high mutation burden in key signaling pathways such as in the proliferation and growth factors *NF1, NRAS, BRAF,* or *KIT* (123). Meanwhile, non-CSD melanomas can arise in younger individuals (< 55 years) on areas that are routinely covered. Interestingly, approximately 50% of both CSD and non-CSD

melanomas contain a BRAF mutation, where over 90% of BRAF mutations occur at the V600 residue, changing the valine to a glutamic acid. This substitution residue occurs within the kinase active domain, and mutations at this site will often increase kinase activity, resulting in transformation *in vitro*. Non-CSD melanomas typically contain the BRAF ^{V600E} mutation, while CSD melanomas will have a BRAF ^{V600} variant (123, 126). As melanomas progress, they accumulate multiple mutations in not only growth and proliferation pathways, but also apoptosis resistance, cell cycle control, and lifespan (123).

The progression of melanoma is typically depicted as a linear path from normal skin to malignant melanoma that have spread to sites such as lungs, liver, bone and brain (123, 127). This is typically an oversimplification of the process, where primary melanomas can co-exist with benign nevi, for example. However, melanomas can arise from new moles that developed quickly over time, or from old moles that have started to change. Evolution of a pre-existing mole is rare as most will fade or cease growing, so typically only new and rapidly evolving moles are monitored by dermatologists (124).

1.8 HERV-K (HML-2) expression in melanoma

The first reported cases of ERV expression associated with melanoma occurred in the 1970s, where viral-like particles (VLPs) - some containing packaged RNA and RT - were detected in patient melanoma samples (143-145). More recently, an antigen was detected in a patient's melanoma sample that was recognized by cytotoxic T cells (142). The gene in question had high sequence similarity with HERV-K: it contained R-U5 and U3-R sequences flanking the internal sequence, it contained a binding site for lysine tRNA, and it contained a polypurine tract on the 3' end of the cloned sequence (142).

This gene - though riddled with mutations - contained a fragment that shared 78% sequence identity to HERV-K (HML-6)'s *env* gene, further confirming that this antigen was a HERV-K element (142). This provirus - dubbed HERV-K MEL - was expressed in melanoma and benign nevi, but not in healthy tissue except for testis and some skin samples (142).

One year after the HERV-K MEL work was published, it was reported by Thomas Muster that melanoma cell lines produce retroviral particles that package HML-2 sequences 7p22.1 and 19p12b (138). They also showed that melanoma cell lines - in this case, SK-Mel-28 and 518A2 - and melanoma patient samples produced HERV-K specific Env and Gag while non-cancerous counterparts did not (138). Expression of HERV-K Env and Gag in melanoma cell lines and melanoma tissue was also confirmed by Kristina Buscher by immunofluorescence, immunohistochemistry, and Western blot (146). She was also able to detect the presence of VLPs by looking for RT activity and HERV-K viral RNA in the supernatant of SK-Mel-28, but they were not infectious and appeared to be defective by electron microscopy (146). Finally, it was shown that high intensity UVB irradiation of melanoma cell lines trigger significant transcriptional expression of HERV-K *pol* and *env* as well as the release of VLPs into supernatant *in vitro* compared to their non-cancerous counterpart (148). Which also triggered expression of HERV-K Env in melanoma cell lines with no expression detected in non-cancerous cells (148).

In addition to elevated HERV-K RNA transcripts, anti-HERV-K antibodies have been detected in patient sera but not in healthy controls (146, 147). Buscher noticed that ~20% of her melanoma patient sera samples contained antibodies to HERV-K Env TM, while healthy controls did not (146). Likewise, Silvia Hahn detected antibodies to HERV-K Gag and Env in ~20% of melanoma sera with an increasing number of antibody producing patients as the stage of the cancer progressed (147).

Melanoma is not the first cancer to shown upregulation of HERV-K RNA transcripts, protein, or VLPs. Prostate cancer has been shown through Sanger sequencing, amplicon sequencing, and qRT-PCR to upregulate the HML-2 proviruses 22q11.23, 3q12.3, and 7q34 (130). This same method confirmed that germ cell tumors, such as the cell line Tera-1, express 22q11.21 which has been confirmed elsewhere (130, 135). Through RNA sequencing, Neeru Bhardwaj has also shown that Tera-1 cells express a diverse set of proviruses such as but not limited to 22q11.21, 22q11.23, and 19p12c (137). Furthermore, cloning frequency analysis by Klemens Ruprecht of Tera-1 cells and VLPs showed expression of 5q33.3, 7p22.1, and 22q11.21 (150). Patients with germ cell tumors have been reported to express antibodies to HML-2 Env and Gag, suggesting their value as diagnostic tools (134). This increase in HML-2 RNA in patient sera was also shown in breast cancer and lymphoma, along with the presence of HML-2 VLPs in lymphoma patient sera (136). In hepatocellular carcinoma, it was shown by qRT-PCR that patients with increasingly poor prognosis had a slight but significant increase in overall expression levels of HML-2 RNA compared to adjacent non-cancerous tissue (132). Finally, an elevation in HML-2 - specific mRNA along with antibodies was seen in breast cancer patients when compared to healthy controls (133). The Coffin lab was able to confirm this increase in HML-2 RNA through RNA sequencing, where Meagan Montesion showed that 1q21.3 and 3q12.3 were abundantly expressed in the breast cancer cell line HCC1954, differing from non-transformed mammary epithelium which primarily expressed 1q22 (140).

The overexpression of HML-2 in cancer generated some interesting questions. For instance, their role in cancer has been questioned since their detection. HML-2 has two additional splice variants around *env* known as Rec and NP9. Depending on the presence (type 1) or absence (type 2) of a 292 bp deletion at the junction of *pol* and *env*, one or the other will be present. Rec is found within type 2 HML-2s and is responsible for

nuclear export of unspliced mRNA, while NP9 is found in type 1. Expression of the HML-2 protein Rec has been suggested to play a role in cancer etiology *in vitro* in collaboration with NP9 through binding to promyolytic zinc finger (PLZF) and *in vivo* by causing carcinoma in situ in transgenic mice (151, 152). Furthermore, expression of HML-2 Env seems to have some oncogenic properties. Recently, it was shown that expression of HML-2 Env is effective at inducing the MAPK ERK1/2 pathway by inducing expression of downstream transcription factors (149). It has also been questioned if HML-2 expression plays a role in cancer etiology or are simply expressed due to dysregulation. As previously mentioned, epigenetic dysregulation is a common occurrence in diseases such as cancer. It is possibly for this reason that an increase in HML-2 expression has been detected in cancer cell lines and patient tumor tissue (130-140, 142-150). This is supported from the observation that HML-2s are overexpressed in pluripotent cells and are down-regulated as cells differentiate, owing to DNA hypomethylation at LTRs and transactivation of pluripotency factors such as OCT4 (129, 141).

Since little was known about the HML-2 transcriptome in melanoma, it was impossible to determine which specific HML-2 proviruses were expressed. If specific proviruses could be identified, one could determine what is activating their expression and identify potential targets for either drug targeting or screening. A better exploration of the HML-2 transcriptome in melanoma would be published by Katja Schmitt (139). Using cloning frequency, they detected the expression of 21 HML-2 proviruses across melanoma cell lines, melanoma patient samples, and primary melanocytes (139). The provirus 7p22.1 was expressed in melanoma cell lines and patient samples only, while other proviruses such as 3q12.3, 1q22, and 7q22.2 were expressed in melanoma and melanocytes (139). While this was a good beginning, there are a few concerns with this approach. Cloning frequency does not consider sense vs antisense transcripts, therefore

some of these reads could be essentially nonsense reads that code for nothing HML-2 related. While cloning frequency provides a good initial way to detect specific expressed proviruses, it is not as sensitive as other methods such as RNA sequencing for transcript detection. Finally, this information does not explain what caused their activation in melanoma.

A few years ago, Neeru Bhardwaj and Meagan Montesion developed an RNA-sequencing pipeline using Tera-1 cells to better understand the HML-2 transcriptome in Tera-1s and breast cancer, respectively (137, 140). This pipeline needed to meet a few requirements: 1) it needed to be sensitive, 2) it needed to be able to distinguish individual proviruses from each other, 3) it needed to be able to determine the difference between sense and antisense transcripts, and 4) it needed to detect nonreference provirus expression. For these purposes, the Illumina MiSeq system was ideal: there were sense stranded kits, the system can produce reads up to 300 bps, and the kits are capable of producing paired end reads to make longer reads for more specific detection of proviruses. In addition to stringent sequencing analysis guidelines, they were able to detect multiple sense stranded proviruses in both Tera-1 cells and VLPs (137). This pipeline was then used by Meagan Montesion to analyze the HML-2 transcriptome in breast cancer (140). In addition to identifying HML-2 specific transcripts, she determined that some proviruses were capable of LTR-driven expression and that transcription factors such as HOX-PBX were critical in this activation (140, unpublished). Finally, she also conclusively showed that the HML-2 transcriptional profile of non-cancerous mammary epithelial cells changed over the course of transformation (140).

1.9 Thesis Objectives

I had three goals for this work: 1) identify individual HML-2 proviruses that were expressed in melanoma cell lines and primary melanocytes, 2) determine the causative factors that drove HML-2 expression in melanoma cell lines and primary melanocytes, and 3) look at the effect of 5' LTRs in HML-2 expression. The purpose of doing this is to see which proviruses were specifically expressed and in what orientation for identifying proviruses that could be implicated in cancer. Furthermore, I wanted to explore the HML-2 transcriptome in cancer and non-cancerous cells to gain a better understanding of expression. Finally, I wanted to compare the causes of HML-2 activation in melanoma to breast cancer. I went into this hopeful that I would be able to better understand the HML-2 transcriptome in cancer to help determine their role in cancer and their potential as therapeutic target or diagnostic tool.

Chapter 2: Materials and Methods

2.1 Cell Culture

I selected five human melanoma cell lines (A375, Malme3M, SKMel28, Steele, and WM164), and three human primary melanocyte populations (HEMN-LP, HEMN-MP, and NHEM-Neo) formywork. The human melanoma cell lines A375 (ATCC, Cat#: CRL-1619), WM164, and Steele were grown in DMEM with Glutamax (Gibco, Cat #: 10569010) supplemented with 10% FBS (Gibco, Cat#: 16000044) and 1% Pen/Strep (Gibco, Cat# 15140-122). The human melanoma cell line SKMel28 (ATCC, Cat#: HTB-72) was grown in EMEM (ATCC, Cat#: 30-2003) and supplemented with 10% FBS and 1% Pen/Strep. The human melanoma cell line Malme3M (ATCC, Cat#: HTB-64) was grown in IMDM (ATCC, Cat#: 30-2005) and was supplemented with 20% FBS and 1% Pen/Strep. While A375, Malme3M, and SKMel28 were purchased from ATCC, WM164 and Steele were a gift from Philip Hinds. The breast cancer cell line, HCC1954 (ATCC, Cat#: CRL-2338), was a gift from Charlotte Kupperwasser, and was grown in RPMI (Thermo Fisher, Cat#:11875-093) supplemented with 10% FBS and 1% Pen/Strep. The human primary melanocyte populations HEMN-LP (Gibco, Cat#: C0025C) and HEMN-MP (Gibco, Cat#: C1025C) were grown in Medium 254 (Gibco, Cat#: M254500) supplemented with HMGS-2 (Gibco, Cat#: S0165) and 1% Pen/Strep. The human primary melanocyte population NHEM-Neo was grown in the MGM-4 BulletKit (Lonza, Cat#: CC3249), which contains MBM-4 Basal Medium and MGM-4 SingleQuot Kit Supplement and Growth Factors and was supplemented with 1% Pen/Strep. All cells were grown in 10 cm dishes at 37C with 5% CO2.

2.2 RNA Extraction and DNase Treatment

Human melanoma cell lines and primary melanocytes were grown in duplicate to ~80% confluency in a 10 cm dish. RNA was extracted using the Qiagen AllPrep DNA/RNA/Protein kit (Qiagen, Cat#: 80004) according to the manufacturer's protocol and eluted in RNA-safe buffer (1U/uL RNase Out (Invitrogen, Cat#: 10777019) and 1mM DTT in 5 mM Tris-HCL). Cells were harvested by scraping (Corning, Cat#: 3008), and lysate was homogenized using QIAshredder (Qiagen, Cat#: 79654). Eluted RNA was diluted in RNA-safe buffer to ~250 ng/uL before treatment with six units of Turbo DNAse (Invitrogen, Cat#: AM1907) for an hour and a half at 37C.

2.3 RNA Quality Control

Prior to submitting RNA for sequencing, I confirmed that detectable DNA was removed using a quantitative qPCR protocol that has previously been described (155), where removal of detectable DNA is shown by the absence of amplification in samples that were not reverse transcribed (RT - samples). This qRT-PCR was performed in triplicate using SuperScript III Platinum SYBR Green One-Step qRT-PCR Kit with ROX (Invitrogen, Cat#: 11746100) with primers specific to a conserved domain within HML-2 *env* TM (For: 5' CTAACCATGTCCCAGTGATG 3', Rev: 5' GGAGACAGACTCATGAGCTTAGAA 3'). Both these primers and the cycling protocol have been previously described (155). Briefly, 2 uL of DNase treated RNA was combined with 12.5 uL of 2X SYBR Green Reaction Mix with ROX (final concentrations: 0.2 mM of each dNTP, 3 mM MgSO4, and 0.5 μM ROX), forward primer (final concentration: 300 nM), reverse primer (final concentration: 600 nM), additional MgSO4 (final concentration: 1 mM), and either SuperScript III RT/Platinum Taq Mix (RT+) or Platinum Taq (final concentration: 2 units) (RT-). Reverse transcription and amplification were performed in the same plate. The reverse transcription parameters were the

following: 50^0C for three minutes, and 95^0C for five minutes. The cycling parameters were the following for 40 cycles: 95^0C for 15 seconds, and 60^0C for 30 seconds followed by a read step. The parameters for the melting curve were as follows: 95^0C for 15 seconds, 40^0C for one minute, and 40-95^0C for 15 seconds followed by a read step. This last step increased at intervals of 0.3^0C after each reading. qRT-PCR was performed on a StepOnePlus Real-Time PCR System (Applied Biosystems: 4376600).

2.4 RNA Sequencing Library Preparation and Sequencing

Once detectable DNA was removed, RNA was submitted with biological replicates to the Tufts University Core Facility for sequencing. A stranded multiplexed library was built using an average of 600 ng per sample of total RNA using TruSeq Stranded Total RNA kit with RiboZero Gold (Illumina, Cat#: RS-122-2301). Total RNA was depleted of rRNA through a combination of biotinylated rRNA-specific oligos and RiboZero rRNA removal beads. RNA transcripts are purified prior to first strand cDNA synthesis with reverse transcriptase and random hexamers. Second strand synthesis is performed shortly after with DNA Polymerase I and RNase H. A single 3' A-overhang is added to cDNA fragments, which is followed by the ligation of barcoded sequence adaptors to the end of fragments. Primers specific to these adaptors were used to amplify the library for 15 cycles of PCR. Up to four libraries were multiplexed and sequenced on the MiSeq benchtop sequencer using the v3 kit to produce paired end reads up to 301 bps long (Illumina, Cat#: MS-102-3001).

2.5 RNA Sequencing Alignment

Reads were filtered through Trimmomatic to remove low quality reads (phred < 20), poor quality ends (phred <20), residual adaptors, and reads shorter than 100 bps. Read quality pre- and post-trimming was confirmed using FastQC to ensure that I had

high quality reads formyalignment. Since HML-2 proviruses contain highly conserved sequences, surviving reads were merged with FLASH to improve mapping quality. Paired end merged RNA-seq libraries were aligned (stranded and unstranded) to hg19 using Tophat (v. 2.0.10) and the underlying aligner Bowtie (v. 2.1.0), then filtered using Samtools (v. 0.1.19) to isolate unique reads. Libraries were also aligned (stranded and unstranded) to a synthesized HML-2 specific genome, which has been described previously (137) and was used to ensure that I was not missing any rare expressed proviruses inmyhg19 alignment. The filtered and unfiltered as well as the stranded and unstranded alignment files were analyzed for HML-2 expression. All libraries were normalized using CuffDiff (v. 2.2.1) with default parameters. Expression of HML-2 proviruses was determined using the Cuffdiff output file in terms of fragments per kilobase of transcript per million mapped reads (FPKM). In addition to total expression, relative abundance of HML-2 provirus expression was calculated ((provirus FPKM/ total HML-2 FPKM) x 100) and any expressed provirus that constituted >1% of total HML-2 provirus expression per cell line were evaluated further. Expression and abundance data was graphed using Prism 6 (bar graphs) or R-Studio (heatmaps [Pheatmap], dot plot [ggplot2], v. 1.0.143).

2.6 Transcription Methods Analysis

Proviruses that were above my transcription threshold were visualized in Integrative Genomics Viewer (IGV, v. 2.3). .bam and .bai files for each cell line were loaded into IGV, and individual proviruses were visualized by mapping to their position within the hg19 genome. Mechanisms behind provirus expression were determined based upon read placement relative to the provirus - especially its corresponding 5' LTR - and neighboring genomic elements. LTR-driven expression was determined based upon placement of reads in relation to the approximate location of the transcription start

site within the 5' LTR. If reads started nearby the initiation sequence of a provirus' 5' LTR, then the provirus was considered to be potentially LTR-driven. If an abundant number of reads aligned to an upstream or downstream element that appeared to carry through to a provirus, the provirus was considered to be driven by said element. If a provirus exists within an intron of an active gene, the provirus was considered to be driven by read through transcription.

2.7 Detecting Intact Open Reading Frames (ORFs) in Expressed Proviruses

ORFs were detected with NCBI's ORF finder. Sequences corresponding to each provirus were obtained from NCBI's GenBank. Each provirus was compared to the full-length sequence of the HML-2 infectious progenitor, HML-2 K-Con. An ORF was considered intact if it matched the corresponding ORF in K-Con (i.e. approximately the same length with intact coding sequences for all domains within each ORF). ORF status (i.e. the presence or absence of an intact ORF) was confirmed utilizing previous results generated in my lab (34).

2.8 Dual Luciferase Promoter Activity Assay

5' LTRs from selected proviruses were synthesized by Genewiz using their uploaded sequences on GenBank. The sequences are banked as the following: 1q22 (JN675014.1), 3q12.3 (JN675021.1), 7p22.1b (JN675044.1), 7q22.2 (JN675046.1), 11p15.4 (JN675061.1), and 21q21.1 (JN675086.1). These synthesized products were cloned into puc57-Amp and sub-cloned into pGL4.17[luc2/Neo] vector upstream of the Firefly luciferase gene. These vectors were co-transfected into A375, SKMel28, WM164, and HCC1954 (positive control) along with the pRV-SV40 vector as a transfection control at a ratio of 30:1 (luciferase vector to control vector).

Cells were transfected in triplicate using Amaza's Nucleofector 2b device (Lonza, Cat#: AAB-1001), which was loaned to my lab by Karl Munger, and grown in six well

plates. Transfection protocols were optimized using Amaza's Cell Line Optimization Nucleofector Kit (Lonza, Cat#: VCO-1001N). A375 were nucleofected using Kit V (Lonza, Cat#: VVCA-1003), while the remaining three cell lines were nucleofected using Kit L (Lonza, Cat#: VVCA-1005). Media was changed, or cells were passed the night before Nucleofection. 1-2 million cells / sample were nucleofected with 1 ug of the luciferase/control mixture using the most optimal program for each cell line. Luciferase activity was read approximately 48 hours post transfection following Promega's Dual Luciferase Reporter Assay System (Promega, Cat#: E1980) instructions. Specific details for each cell line are listed in Table 2.1.

Cell Line	Program	Kit	# of cells Nucleofected
SKMel28	A033	Solution L	2 million/well
WM164	X005	Solution L	1 million/well
HCC1954	X001	Solution L	1 million/well
A375	E023	Solution V	1 million/well

Table 2.1: Nucleofection protocol for melanoma and control cell lines.

A brief summary of the Nucleofection protocol used for my LTR activity assay. The cell lines used in my luciferase assay are represented alongside their respective Nucleofection program, kit name, and number of cells used for each Nucleofection replicate.

2.9 Identifying Unique Transcription Factor Binding Sites (TFBS) with Genomatix

5' LTRs that showed promoter activity were probed for unique transcription factor binding sites (TFBSs). A TFBS was considered unique if it was identified in only one of the 5' LTRs analyzed. Sequences for each 5' LTR were collected from GenBank and were analyzed using Genomatix's MatInspector (v. 3.9). A consensus sequence of all four active LTRs was generated using MUSCLE (alignment) and SeaView (consensus sequence generation from the alignment, v. 4.6.1). The mutations responsible for these sites were identified using the consensus sequence in MEGA 7. Each mutation was then

reverted to match the consensus, or to remove the TFBS using the recognition sequence detected by MatInspector.

2.10 Q5 Site Directed Mutagenesis

The TFBS potentially responsible for LTR activity were mutated using the Q5 Site-Directed Mutagenesis kit (New England Biolabs, Cat#: E0554S). Primers were designed using NEBaseChanger, which is the recommended online primer design software for this kit. LTRs were mutated according to manufacturer's protocol. Primer sequences are in Table 2.2. The only exception is the mutated 7p22.1b vector, which was synthesized by Genewiz. Mutagenesis was confirmed using the pGL4.17[luc2/Neo] specific primer (RVPrimer3: 5' CTAGCAAAATAGGCTGTCCC 3') and an internal HML-2 LTR primer provided by Zachary Williams (M7F: 5' AAGCCAGGTATTGTCCAAGG 3').

LTR (5')	Mutation	Forward primer	Reverse primer
1q22	LTATA	AGGATTAGTAAAAGAGGAAGGAATG	CCTCAGCACAGACCCTTT
	ZBRK1	AAGACCTGACCGTCCCCCAGCC	TCCCTTCCCACGAGGCCA
3q12.3	HOX_PBX	GATTGTATGTTCCATCTAC	GGGTTTTATACCGAGACATTC
	RFX3	TCTTGTGACCCTGACACATCC	CAATTGTGGGGAGAGGGT
	E2F	GCGGGATCCTCCATATGCTGAAC	CAGCCTCTGAGTTCCCTTAGTATTTATTG
	EGR1	AAACACCCACAGATGATCAATAAATACTAAG	CTCTGAGAGGGGGATGTG
	SOX10	ATTGATTGTATGCTCCATCTACTG	CGGGTTTTATACCGAGAC
7q22.2	CEBP	CTTTTTCTTTTCCAAATCTCTCGTTC	ACACAGAGACAAAGTATAG
	IR1_NGRE	TCCCCCTCTTCGAGAAACACC	TGTGTCAGGGTCACAAGAC

Table 2.2 Mutagenesis primers used to mutate 5' LTRs.

A list of primers used for site directed mutagenesis to remove unique transcription factor binding sites within my 5' LTRs of interest. Primers were designed using NEBaseChanger, which is the recommended software for the Q5 Site Directed Mutagenesis kit.

2.11 Western blot for HML-2 Env

Melanoma cell lines and primary melanocytes were grown in 10 cm dishes until they were approximately 70 - 80% confluent. Cells were washed with 1x PBS, then treated with NP40 lysis buffer containing Halt protease inhibitor. Cells were harvested by scraping on ice and lysate was transported into microfuge tubes where they were incubated on ice for 30 minutes. After incubation, cell lysate pellets were collected by centrifugation at room temperature for one minute at 12,000 RPM on a tabletop centrifuge. Clarified lysate was moved to a new chilled tube. As a control, 293T were transfected with 28 ug of HERV-K Con Env using the Lipofectamine 2000 standard protocol. These transfected cells recovered for 48 hours prior to lysate collection. Protein concentration was calculated using the Quick Start Bradford Protein Assay kit (Bio-Rad, Cat #: 5000201) following their microplate standard assay protocol.

30 ugs from each sample were mixed with 6X Laemmli buffer (Alfa Aesar, Cat#: J61337) and boiled for 10 minutes. Lysate was cooled and centrifuged to pool condensation. All samples were loaded into a 10% SDS-PAGE gel and run until the dye ran off. Proteins were transferred (100 volts) onto a PVDF membrane for one hour at room temperature with an ice pack. Membranes were blocked in 5% nonfat milk containing 0.2% TBS-T at room temperature for one hour, then probed overnight in blocking buffer with either Anti-Env TM (Austral, CAT#: HERM-1811, 1:2000) or Anti-Beta Actin (1:5000), then for one hour with secondary horseradish peroxidase-conjugated antibody. Blots were visualized using Novex ECL (Invitrogen, Cat#: WP20005) and a Syngene G:Box imager.

Chapter 3: Detecting the HML-2 Transcriptome in Melanoma Cell Lines and Primary Melanocytes using Next Generation Sequencing

3.1 Total HML-2 Expression in Melanoma Cell Lines and Primary Melanocytes is Low Compared to Housekeeping Genes

HML-2 expression in the form of protein and RNA transcripts have been detected in a variety of cancer cell lines and patient tissue samples, including melanoma, breast cancer, and germ cell tumors (130, 132 - 137, 139, 140, 150). Previously, the Coffin lab has studied the HML-2 transcriptome in a breast cancer cell line – Hcc1954 – and during transformation of human mammary epithelial cells (HuMECs) and the teratocarcinoma cell line Tera-1, revealing the diversity of transcribed HML-2 proviruses as well as the difference in total HML-2 expression level to different cells as measured by FPKM (137, 140). Tera-1 cells were shown to have the highest total expression of HML-2 transcripts at approximately 225 FPKM and predominantly expressed proviruses including 22q11.21 and 22q11.23, to name a few of the several found (137). Breast cancer cell lines ranged from approximately 25 FPKM for Human Mammary Epithelial cells overexpressing Ras (HMLE-Ras) to as low as approximately five FPKM for Human Mammary Epithelial cells overexpressing Her2 (HMLE-Her2) (140). Interestingly, Montesion et al were able to show a change in the HML-2 transcriptome which correlated with the transition from non-cancerous to cancerous mammary cells: non-cancerous cells primarily expressed 1q22 antecedent to transformation into a cancerous state where they primarily expressed 3q12.3 and 1q21.3.

At this point, the HML-2 transcriptome had only been measured in such depth in these two cancer types. Considering that HML-2 expression has been detected in several different cancers, I thought it wise to study other cancers to see how similar the HML-2 transcriptome is in both diversity and expression level. I found ourselves well positioned to do this since Montesion and Bhardwaj developed a robust method for

measuring HML-2 transcription, and so I chose to study HML-2 expression in melanoma - one of the most studied cancers for HML-2 expression after breast cancer and germ cell tumors. To begin, I selected five melanoma cell lines (A375, Malme3M, SKMel28, Steele, and WM164) all of which contained the BRAF activating mutation $BRAF^{V600E}$ and three primary melanocyte populations (NHEM-Neo, HEMN-LP, and HEMN-MP) for my RNA-sequencing analysis. I selected melanoma cell lines that contained $BRAF^{V600E}$ because roughly half of all melanoma tumors contain a mutation in BRAF, and of those with BRAF mutations approximately 90% contain the V600E activating mutation. RNA was harvested from two biological replicates and DNase treated until I could no longer detect DNA via qRT-PCR. After submission for Illumina MiSeq Paired End 300 library prep, I removed residual adaptor sequences and low-quality ends, merged reads, and aligned them to the human genome assembly hg19. Once aligned, I normalized my data with CuffDiff and expression was measured as FPKM.

I started my analysis by comparing total HML-2 expression to housekeeping genes to get perspective on their expression in my model system (Figure 3.1). For my comparison, I selected three housekeeping genes previously chosen by Bhardwaj whose expression ranged from high to low: GAPDH, ACTB, and RAB7A. To these genes I compared total HML-2 expression in my unstranded hg19 alignment ("Total Unstranded Accepted Hits HML-2"), uniquely aligning hits from my sense stranded alignment ("Total Sense Accepted Hits Unique HML-2"), and the most abundantly expressed provirus from my unique hits sense stranded alignment ("7q22.2 Sense Accepted Hits Unique HML-2") to see how increased filtration affected HML-2 expression in my data set. I emphasized the sense stranded alignment since I wished to focus on potentially protein producing reads for the purpose of examining HML-2's role in melanoma etiology and to understand the transcription mechanisms behind their

activation. Furthermore, I chose to study uniquely aligning reads since HML-2 proviruses have high sequence similarity.

On average, HML-2 expression in my unstranded alignment was 36X less than their corresponding RAB7A expression with a range of 17X less in HEMN-MP to 71X less in A375. Two cell lines appeared to have a higher HML-2 expression than the rest: SKMel28 and Malme3M. I thought this interesting as SKMel28s have been previously shown to express Viral Like Particles (VLPs) in conditioned supernatant and I was initially hopeful to isolate VLPs for RNA-sequencing to detect packaged HML-2 transcripts (146). Filtering my data set to isolate unique sense stranded reads specific to HML-2 showed a slight drop in expression. This was expected since these proviruses have high sequence similarity, yet since the decrease was not substantial this suggests that there is still sufficient expression to examine HML-2 expression post-filtering. When I briefly looked at the HML-2 transcriptome in my model system, I noticed that most of HML-2 expression in my unique hits sense stranded alignment data corresponded to the provirus known as 7q22.2 (Figure 3.1, Figure 3.3, Figure 3.4). Transcripts specific for this provirus essentially comprises half of the primary melanocyte HML-2 transcriptome (average: 48.9%), while it accounts for approximately one third of the melanoma cell line HML-2 transcriptome (average: 35.8%). This result is similar to what Montesion et al have shown in that primary cells predominantly express one provirus while cancerous cell lines express a range of different proviruses.

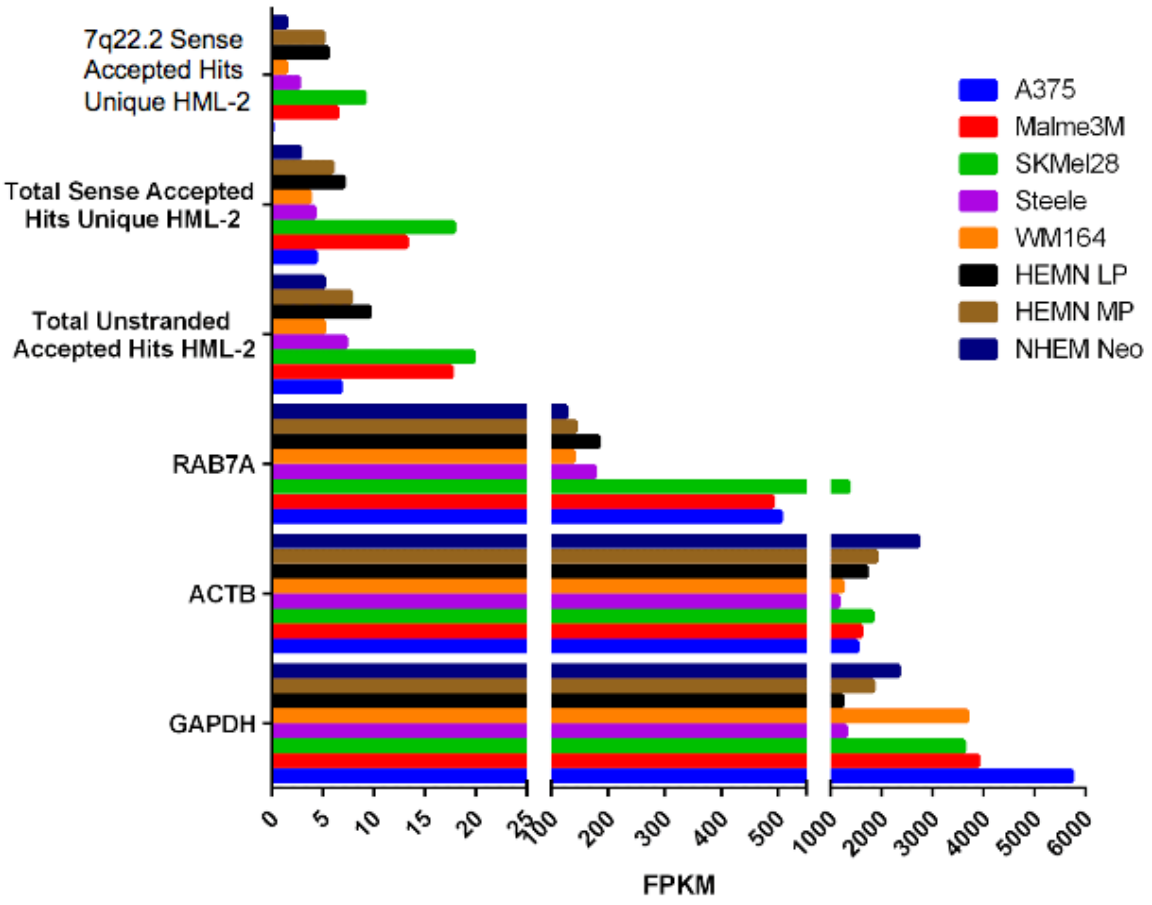


Figure 3.1: Total HML-2 Expression in Melanoma Cell Lines is Low Compared to Housekeeping Genes Expression.

RNA was harvested from melanoma cell lines (A375, Malme3M, SKMel28, Steele, and WM164) and primary melanocytes (HEMN LP, HEMN MP, and NHEM Neo) using Qiagen's AllPrep kit and eluted in RNA safe buffer. RNA was DNase treated with Turbo DNase until DNA was undetectable by qRT-PCR. DNase-treated RNA was submitted for paired end Illumina MiSeq 300, and sequenced reads were processed to remove residual adaptors, low quality reads/ends, and to merge reads. Merged and unmerged paired reads were aligned (unstranded and sense stranded) using TopHat to the hg19 human genome build. CuffDiff was then used to both normalize and calculate expression by FPKM. Presented above is the comparison of total unstranded HML-2 reads, total uniquely aligning HML-2 reads from the sense stranded alignment, and the highest expressed provirus (7q22.2, unique stranded reads only) in the sense stranded alignment to three housekeeping genes within all surveyed melanoma and melanocytes.

3.2 HML-2 Expression in Melanoma Cell Lines and Primary Melanocytes is Due to Sense Stranded Transcription

2013 saw the publication of a paper by Katja Schmitt that reported on the HML-2 transcriptome in melanoma and primary melanocytes using cDNA cloning frequency (139). my work aimed to look for additional expressed proviruses that may have been missed by this method while comparing the melanoma HML-2 transcriptome to breast cancer and Tera-1 cells. To compare what I detected to Schmitt et al, I looked at the HML-2 transcripts that were detected in my unstranded alignment and grouped them based on proviruses that were expressed in both studies, in my study alone, or in their study alone (Figure 3.2). This comparison showed significant overlap where most of the expressed proviruses they detected by cloning frequency were also detected by RNA-sequencing. Furthermore, I was able to detect an additional 31 expressed proviruses that were previously not detected due to either insensitivity on the behalf of cloning frequency or because Schmitt et al samples simply were not expressing them.

There were three proviruses that were not expressed in my model system: 1p31.1, 12q14.1, and 19q12. These three proviruses were expressed in patient tumor samples (i.e. 1p31.1 and 19q12 in a melanoma from one patient, and 12q14.1 in a lymph node metastasis from another patient). Since their expression was detected in tissue samples and not cell lines, it is quite possible that I did not detect expression of 1p31.1, 12q14.1, and 19q12 due to sample origin. Considering that these proviruses are polymorphic, this hypothesis is further supported. Regardless, my RNA-sequencing approach was sensitive enough to overlap well with what was previously reported on while expanded on the known HML-2 transcriptome in melanoma cell lines and primary melanocytes.

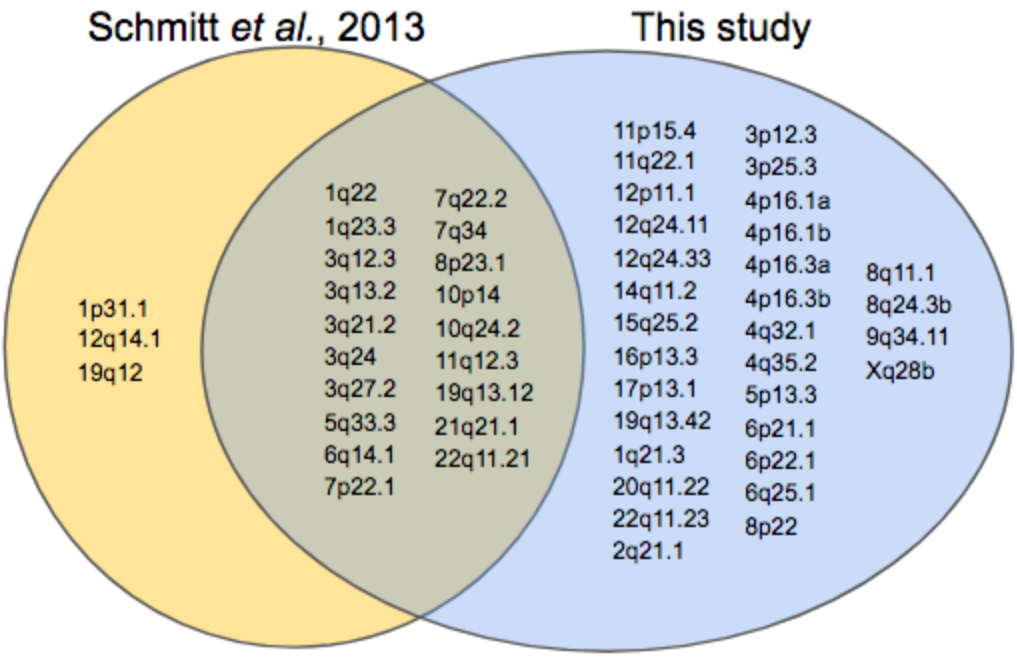


Figure 3.2: The HML-2 Transcriptome Detected by Cloning Frequency and RNA-Sequencing Overlap Significantly.

Post - hg19 alignment, a list of expressed proviruses (i.e. FPKM > 0) from my unstranded alignment (blue) was compiled, containing the names of expressed proviruses from all of my melanoma cell lines and primary melanocytes. To compare the results of my unstranded alignment with Schmitt et al 2013 paper, a list of expressed proviruses (cloning frequency > 0%) was compiled from all of their cell lines, melanocytes, and tissue samples (yellow). Overlap between the two studies is represented above.

I next wanted to examine the diversity of HML-2 expression in each of my samples to see how this expression was divided across my model system (Figure 3.3). At first, I noticed that all but one of my cell lines (A375) expressed 7q22.2. 7q22.2 is within an intron of *LHFPL3*, which is associated with deafness in humans and mice (156). Expression of *LHFPL3* does not appear to account for expression of 7q22.2, as seen in Figure 3.3: Though *LHFPL3* is not expressed in A375 cells, there is also no expression of it in Steele despite the detection of 7q22.2 transcripts in Steele cells (Figure 3.3). This lack of 7q22.2 expression in A375 cells could be due to not only a lack of *LHFPL3* expression but to enhancer or transcription factor expression found in Steele cells but not in A375 cells or the expression of repressors found in A375 cells but not in Steele cells. Additionally, I noticed that all my melanoma cell lines and primary cells expressed 3q12.3, corresponding well with Montesion and Bhardwaj's breast cancer and Tera-1 data where 3q12.3 expression was detected in all of Montesion's and Bhardwaj's cells. Finally, I noticed that two of my melanoma cell lines - SKMel28 and WM164 - expressed transcripts for the polymorphic provirus 7p22.1. This provirus is the result of a duplication event where a copy of this provirus was inserted immediately next to the original sequence, and therefore these proviruses are referred to as 7p22.1a and 7p22.1b.

There were many proviruses that were expressed in my data set for each of my cell lines and primary cells, yet upon viewing most of these proviruses in the Integrative Genomics Viewer (IGV) I observed that most were poorly expressed (i.e., < 1% of the total HML-2 transcriptome) or were poorly covered (i.e. <50% of the provirus was covered by reads). This left me wondering if I could reliably consider such proviruses as expressed. To solve this problem, I established a threshold where transcripts comprising less than 1% of the total HML-2 transcriptome were combined with transcripts from

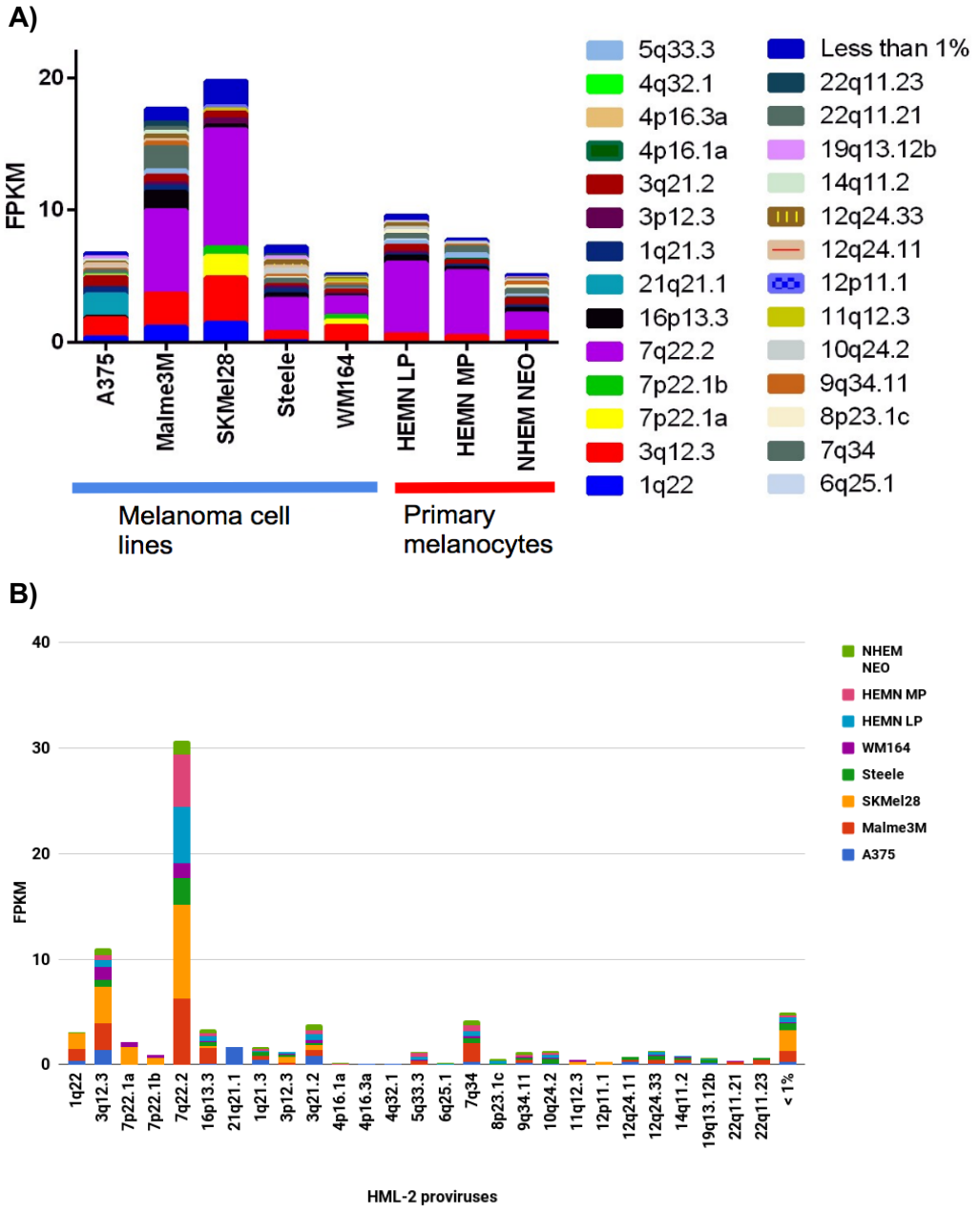


Figure 3.3: Unstranded HML-2 Expression in Melanoma Cell Lines and Primary Melanocytes.

After generating the unstranded hg19 alignment, my output was filtered to group expressed proviruses that constituted > 1% of the HML-2 transcriptome in a given cell line into a “less than 1%” group. Abundance was calculated for each cell line as (FPKM of HML-2 provirus X / total HML-2 FPKM) * 100. HML-2 expression from the unstranded non-unique hits hg19 alignment is separated (A) by cell line, and (B) by individual proviruses for easier visualization. Expression is reported in FPKM. In panel A, melanoma cell lines are underlined in blue while primary melanocytes are underlined in red.

proviruses that had poor coverage (< 50% of the provirus) into a “below threshold” group. Furthermore, I observed that similar to Montesion et al work there were proviruses that were expressed in both sense and antisense orientation. As previously stated, I chose to analyze a stranded library to determine transcription mechanisms behind HML-2 expression, so I performed separate hg19 alignments to separate expression of unique sense transcribed proviruses (Figure 3.4) from unique antisense transcribed proviruses (Figure 3.5).

The number of expressed proviruses within my uniquely aligned sense stranded alignment decreased significantly (Figure 3.4). 7q22.2 is the only provirus expressed above my threshold in primary melanocytes, while it still constitutes most expressed proviruses in four of my five melanoma cell lines. Interestingly, 3q12.3 was only expressed in my melanoma cell lines suggesting that 3q12.3 was either expressed in antisense orientation or fell below threshold in my primary cells. The expression of 3q12.3 in melanoma cell lines only stands in contrast with Montesion et al breast cancer work where 3q12.3 transcripts were detected in both a breast cancer cell line an in HuMECs (140). Since I did not detect 3q12.3 expression above threshold in my uniquely aligning antisense hits, it appears that 3q12.3 fell below my threshold (Figure 3.5). Since my model system consists of different cell types, its presence in Montesion et al breast cancer data could be due to differences in cell type transcription.

A closer inspection of uniquely aligning antisense reads revealed that most expression fell below my threshold (Figure 3.5). Cell lines such as SKMel28 and WM164 did not express any proviruses in this orientation above my threshold, indicating that HML-2 transcription in these cell lines is mainly due to *cis*-acting elements such as LTRs. There were a few cell lines such as A375 and Steele that did express proviruses antisense to their genomic orientation. Overall, while there is some antisense transcription in my model system most HML-2 expression is due to sense transcription.

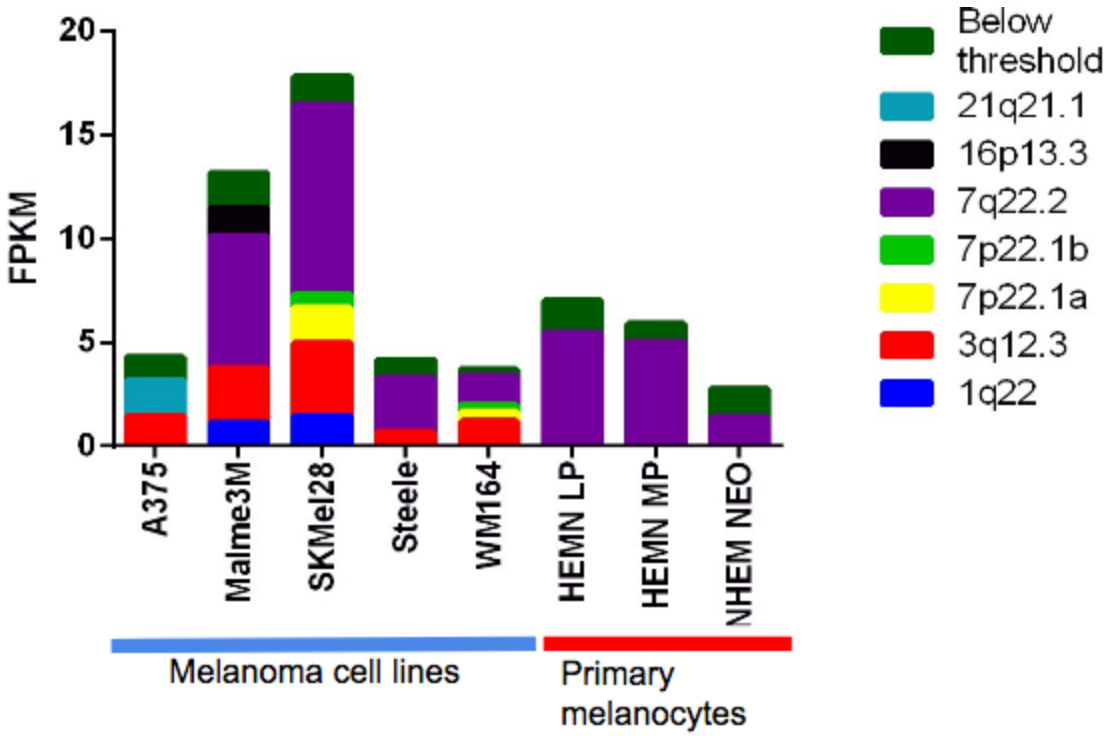


Figure 3.4: Sense Stranded HML-2 Expression in Melanoma Cell Lines and Primary Melanocytes.

RNA-sequencing reads were re-aligned to hg19 to isolate uniquely aligned sense stranded reads. The output of this alignment was filtered using Samtools to remove non-uniquely aligned reads, and to group poorly expressed (<1% of the HML-2 transcriptome) and poorly covered (<50% of the provirus) proviruses into the “Below threshold” group. Expression is represented in FPKM. Melanoma cell lines are underlined in blue while primary melanocytes are underlined in red.

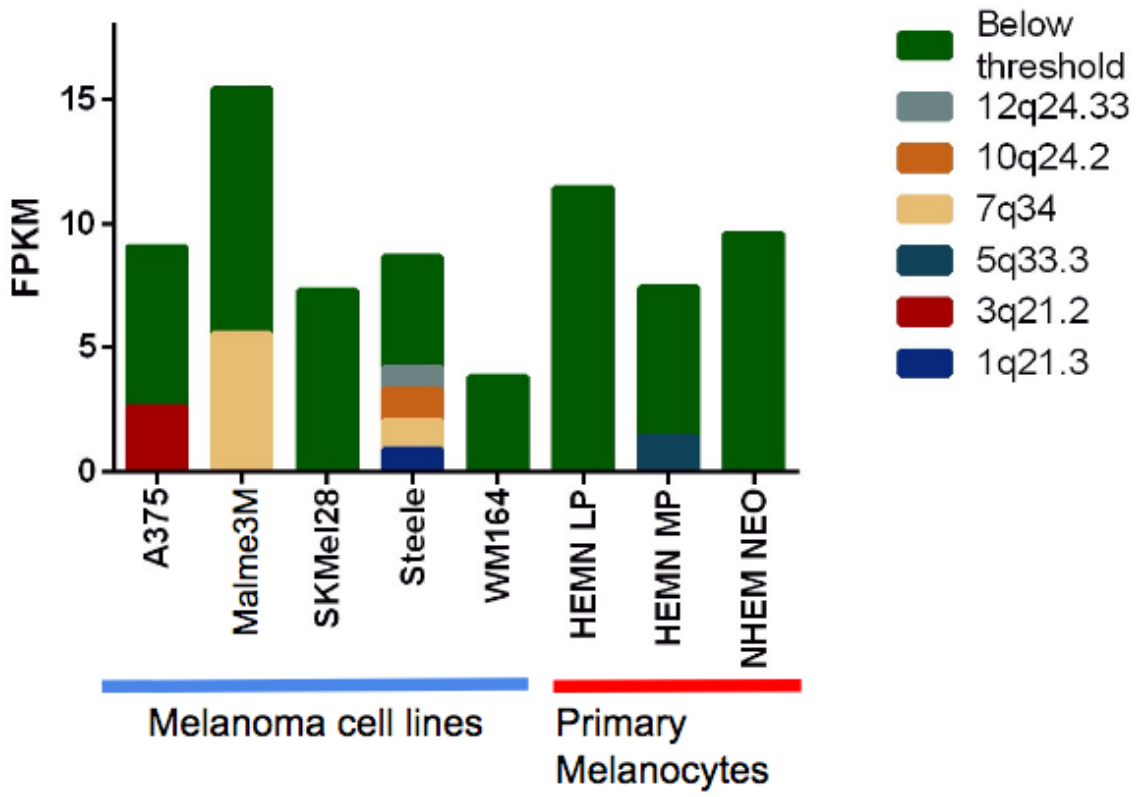


Figure 3.5: Antisense Stranded HML-2 Expression in Melanoma Cell Lines and Primary Melanocytes.

RNA-sequencing reads were re-aligned to hg19 to isolate uniquely aligned antisense stranded reads. The output of this alignment was filtered using Samtools to remove non-uniquely aligned reads, and to group poorly expressed (<1% of the HML-2 transcriptome) and poorly covered (<50% of the genome) proviruses into the “Below threshold” group. Expression is represented in FPKM. Melanoma cell lines are underlined in blue while primary melanocytes are underlined in red.

3.3 No rare provirus transcripts are detected within my model system

The hg19 genome build contains 91 annotated HML-2 proviruses, of which 30 are human-specific and 11 are insertionally polymorphic (34, 35). Recent work within the Coffin lab has shown that the hg19 build contains at least 36 non-reference HML-2 proviruses (36). In particular, four solo LTRs within hg19 are actually full-length proviruses, while there were five non-reference proviruses annotated as pre-integration sites (36). Since the Coffin lab has access to these additional sequences, Montesion and Bhardwaj built an HML-2 “genome” containing all the known HML-2 - related sequences to use as reference. My melanoma and primary melanocyte libraries were aligned to this synthetic genome and I compared sense stranded HML-2 transcripts of uniquely aligned reads to sense stranded uniquely aligned HML-2 reads from my hg19 alignment (Figure 3.6). The output of these separate alignments was normalized with CuffDiff, with expressed proviruses that contributed less than 1% of the total HML-2 transcriptome or had poor coverage grouped into the “below threshold” group. By comparing these two alignments, I saw minimal difference in terms of which HML-2 proviruses were expressed. No transcripts for the non-reference proviruses previously corrected by Wildschutte and Williams were detected, suggesting that rare proviruses are either not present or are not expressed in my model system.

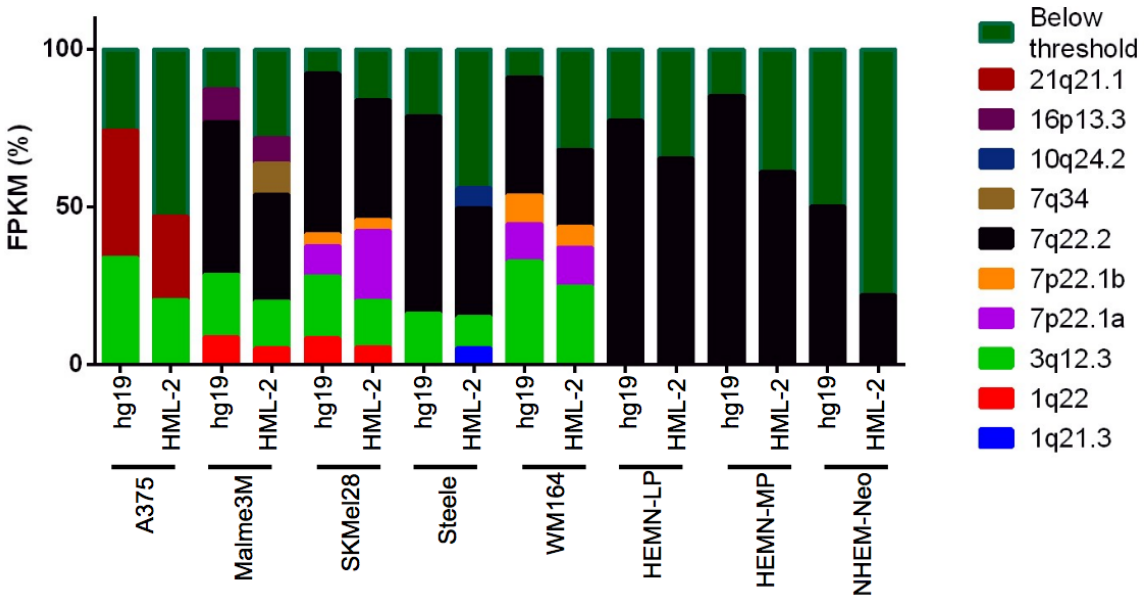


Figure 3.6: Comparison of Unique Sense Stranded hg19 Alignment to Unique Sense Stranded HML-2 Alignment Shows Little Difference in HML-2 Transcriptome.

To look for the expression of rare proviruses, RNA-sequencing reads from melanoma cell lines and primary melanocytes were aligned to a synthetic HML-2 genome using Tophat to isolate uniquely aligned sense stranded reads. CuffDiff was used to normalized expression data and to calculate expression data in terms of FPKM. Abundance of the total HML-2 transcriptome for each alignment is represented (FPKM %), where FPKM % = (FPKM of HML-2 provirus X / total HML-2 FPKM) * 100 for each cell line. Hg19 refers to the human genome 19 build, while HML-2 refers to the custom built HML-2 genome.

3.4 When Comparing HML-2 Expression across Cancerous and Non-Cancerous Cell Lines, 3q12.3 Expression is Detected in Cancer Cell Lines Only

During their time in the Coffin lab, Montesion and Bhardwaj expanded upon the known HML-2 transcriptome in breast cancer cell lines and Tera-1 cells through the development of a sensitive RNA-sequencing pipeline (137, 140). Comparison of their findings indicated that the HML-2 transcriptome differed in HML-2 expression level and provirus diversity. While Tera-1 cells expressed a wide range of proviruses at high levels, breast cancer cell lines expressed only a few proviruses. The benefit of having these data sets and my pipeline is I could compare the transcriptome of other cell lines to see how HML-2 expression changes across cell and cancer type. To see how melanoma cell lines and primary melanocytes compare, I contrasted their expression to their prior sequencing data. Since Montesion was only able to sequence one breast cancer cell line and one immortalized HuMEC, I submitted four additional breast cancer cell lines (BT20, BT474, MDAMB231, and T47D) and two additional primary human mammary epithelium populations (HuMECs from Thermo Fisher and ATCC) for RNA-sequencing. The same pipeline was followed for these samples as before.

As previously stated, Tera-1 cells strongly expressed HML-2 transcripts (Figure 3.7). Melanoma and primary melanocytes had approximately equivalent HML-2 expression levels as breast cancer cell lines and HuMECs. Interestingly, there were distinct differences in HML-2 expression in cancerous cell lines and their non-canceorus counterparts. Primary melanocytes only express 7q22.2, while HuMECs primarily express 1q22 with some expression of 4p16.3a. My melanoma cell lines express 7q22.2, 1q22, and 3q12.3, while my breast cancer cell lines predominantly express 3q12.3. Finally, my Tera-1s express an array of different HML-2s with the majority of their expression derived from 22q11.21, 22q11.23, and 19p12c. HuMEC expression differs slightly here compared to what was previously shown, in that expression of 3q12.3,

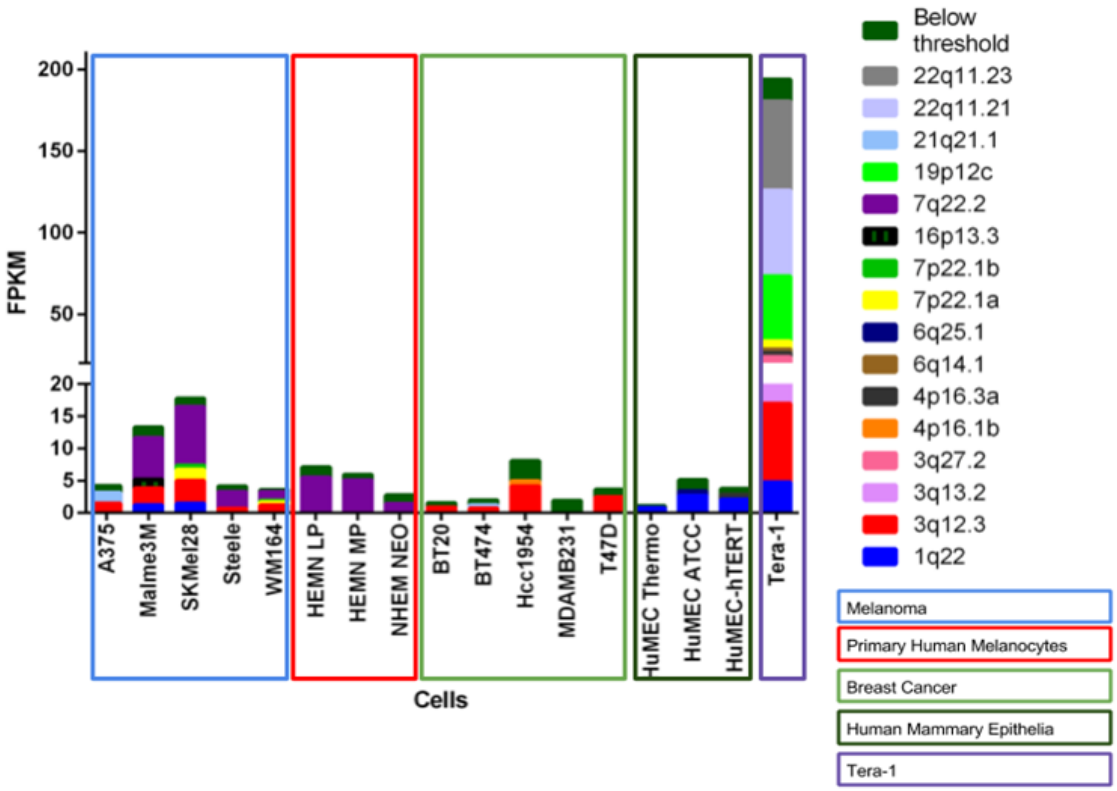


Figure 3.7: Comparison of Cancer Cell Lines and Primary Cells Show How Diverse HML-2 Expression is Across Different Cancer and Cell Types.

HML-2 expression in three cancer types and their corresponding non-cancerous counterparts was compared using CuffDiff. Additional breast cancer and mammary epithelium cells were sequenced and aligned with the same protocol. A color-coded key to connect cell name to phenotype is present in the bottom right corner. Expression level in FPKM was determined from the unique sense hg19 alignment.

1q21.3, and 4p16.1b is missing. Expression of HML-2s in Hcc1954 also differs from Montesion et al work in that 4p16.3a and – surprisingly - 1q21.3 are missing. Strikingly, I observed an interesting change where cells of a certain origin express a distinct provirus(es) which change between healthy and cancerous states. Overall, I observed distinct patterns of expression where cells of a certain type express specific proviruses.

Interestingly, all of my cancer cell lines expressed 3q12.3. This provirus is older than most of the proviruses expressed in melanoma and melanocytes (> eight million years old), contains an "intact" ORF for *gag* (more on this later), and Montesion et al were able to show that it's 5' LTR is active (34, 140). Analysis of its 5' LTR sequence showed the presence of a duplication event that generated a new transcription factor binding site for HOX_PBX that appeared important in LTR activity in HME-Ras and HME-hTERT with minimal effects in Hcc1954. Cancers such as breast and melanoma are known to overexpress HOX and PBX family transcription factors, and the expression of 3q12.3 appears to be partially tied to HOX_PBX expression in my cell lines.

3.5 Characteristics of Sense Transcribed HML-2 Proviruses

For 16 years, reports of melanoma cell lines and patient samples producing HML-2 - specific transcripts, proteins, VLPs, and antibodies have slowly surfaced. Since there are several HML-2 proviruses that could be responsible for this expression, finding the proviruses responsible was difficult until the transcriptome was investigated. It was hoped that doing so would not only determine the identity of expressed proviruses but would also allow us to study the role in disease etiology and explore the possibility of their use in therapeutics and as biomarkers. Using previous work from the Coffin lab as a guide, I examined the characteristics of my uniquely aligned sense stranded expressed proviruses to see if they contained any interesting characteristics (34).

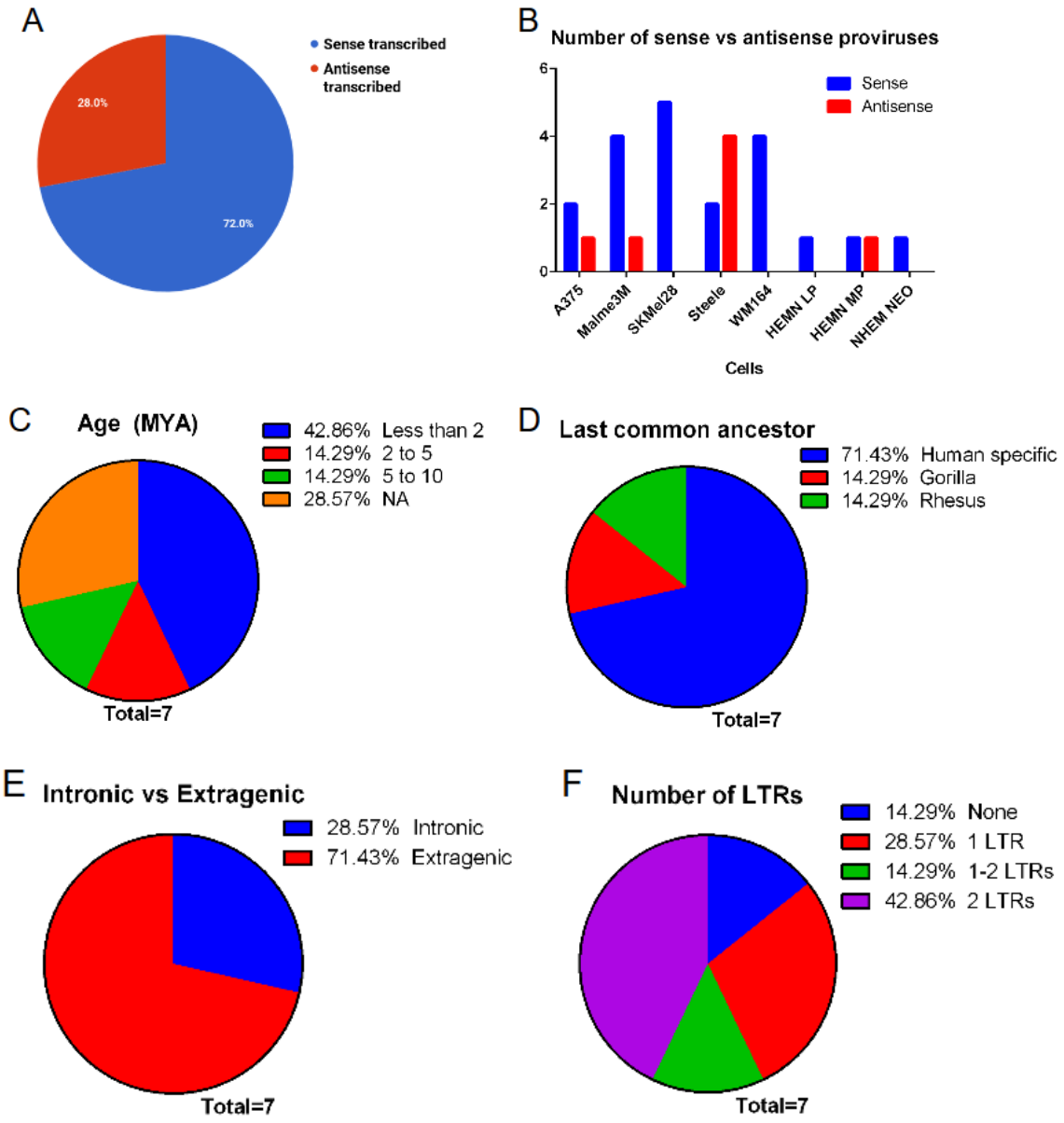


Figure 3.8: Characteristics of Sense Stranded HML-2 Proviruses Expressed in Melanoma Cell Lines and Primary Melanocytes.

The average abundance (A) and total number (B) of sense vs antisense proviruses was determined from unique sense and unique antisense hg19 alignment data. Information on provirus age (C), last common ancestor (D), and number of LTRs (F) was collected from Subramanian et al (34). (E) The number of intronic and extragenic proviruses was determined from unique sense hg19 alignment data.

Provirus	Expressed in…	Alternative Names	Co-ordinates	Orientation	Age (MYA)	Last Common Ancestor	ORFs	Type 1 or Type2
1q22	Malme3M, SKMel28	K102, K(C1b), K50a, ERVK-7	Chr1: 155596457 - 155605636	(-)	<2	Human specific		Type 1
3q12.3	A375, Malme3M, SKMel28, Steele, WM164	K(II), ERVK-5	chr3: 101410737 - 101419859	(+)	5.51 - 9.98	Gorilla	gag	Type 1
7p22.1a	SKMel28, WM164	K108L, K (HML.2-HOM), K(C7), ERVK-6	chr7: 4622057 - 4631528	(-)	<2	Human specific	pol, env	Type 2
7p22.1b	SKMel28, WM164	K108R, ERVK-6	chr7: 4630561 - 4640031	(-)	<2	Human specific	pol, env	Type 2
7q22.2	Malme3M, SKMel28, Steele, WM164, HEMN-LP, HEMN-MP, NHEM Neo	ERVK-14	chr7: 104388369 - 104393266	(-)	N/A	Human specific		NA
16p13.3	Malme3M	K(OLDAC004034	chr16: 2976160 - 2977661	(+)	N/A	Rhesus		NA
21q21.1	A375	K60, ERVK-23	chr21: 19933916 - 19941962	(-)	3.46 - 6.27	Human specific		Type 1

Table 3.1: Expressed proviruses: Sense Stranded.

Summary of sense expressed proviruses along with characteristic information on each, which was collected from Subramanian et al (34). MYA = million years ago.

As previously mentioned, most of the proviruses that are expressed are expressed in the sense orientation except for Steele cells which have more antisense expressed proviruses (Figure 3.8 A and B). Slightly more than half of all sense stranded expressed proviruses are less than five million years old (MYO) and most are human specific (Figure 3.8 C and D). In Montesion et al breast cancer data, expressed proviruses were at least five MYO or older and were primarily expressed by *trans*-activating factors.

Interestingly, most of my sense stranded expressed proviruses were found within the extragenic region of the genome (Figure 3.8 E). Considering these proviruses are expressed, that most of these are extragenic indicates that they are driven by either upstream elements (neighboring genes or repetitive elements) or they are capable of driving their own expression via their 5' LTRs. Since at least 40% of them contain two LTRs, this possibility is quite likely to be the case with at least one provirus (Figure 3.8 F). These and other characteristics are summarized in Table 3.1.

3.6 Despite Intact ORFs, HML-2 - Specific Protein and VLPs Not Detected

Considering that HML-2 protein expression has been detected in patient tissue samples and cell lines, I chose to look for the presence of intact ORFs within my uniquely aligned sense stranded expressed proviruses (Figure 3.9). To do so, I used NCBI's ORF Finder and the banked sequences for each of my expressed sense stranded proviruses of interest from Genbank. The accession number for these sequences are located in the Materials and Methods section.

In 2006 and 2007, two teams led by Thierry Heidmann and Paul Bieniasz each synthesized an infectious HML-2 virus (160, 161). These viruses - named Phoenix and HERV-K Con, respectively - were created *in silico* by building consensus sequences from human specific proviruses. the Coffin lab has acquired constructs for HERV-K Con

for the production of virus and viral - specific proteins. Therefore, I used the HERV-K Con sequence to identify intact ORFs.

Whole sequences for HERV-K Con and my proviruses of interest were run through NCBI's ORF Finder, and the length and amino acid sequence of each potential ORF were compared to HERV-K Con. If an ORF was identified as an HML-2 - specific protein by BLAST, and if the length and sequence were off by no less than a few amino acids from the equivalent HERV-K Con ORF, the ORF was considered "intact" for my analysis. By doing so, I was able to identify three proviruses that contained intact ORFs by my standard: 3q12.3, and 7p22.1a and b (Figure 3.9). 3q12.3 contains an intact ORF for *gag*, yet premature stop mutations and other deletions interrupt other ORFs. Therefore, most of my melanoma cell lines should produce detectable Gag. On the other hand, 7p22.1a and b contain intact ORFs for *pro, pol*, and *env*. Interestingly, both of these proviruses are expressed in SKMel28 and WM164. While no data suggests that WM164 produces VLPs, it should be able to due to expression of 3q12.3 transcripts. Furthermore, it has been previously reported that SKMel28 cells release VLPs. Therefore, it's possible that not only do these cell lines express VLPs, but it's possible that these are the proviruses responsible for their production.

To look for the expression of VLPs and HML-2 - specific protein production, I collected both whole cell lysate and conditioned supernatant at a series of time points to detect VLPs and either Env or Gag (Figure 3.10). Searching filtered supernatant left us empty handed: through a combination of longer time points and different virus concentration and pelleting techniques including ultracentrifugation and sucrose gradients I was unable to detect VLPs by qRT-PCR for HML-2 - specific transcripts (data not shown). The use of conditioned supernatant from Tera-1 cells showed HML-2 transcripts from pelleted virions using these methods, so their lack was not attributed to failure of my methodology. Furthermore, I was unable to detect Env expression by

Western blot in any of my melanoma cell lines or primary melanocytes (Figure 3.10). Though I made numerous rounds of troubleshooting, I could not get my anti-Gag antibody (HERM-1831 and HERM-1841 from Austral Biologicals) to detect anything, including my positive control. It should be mentioned that while 3q12.3 *gag* appears to be intact by my standards that there are other mutations outside of *gag* within 3q12.3 that affect Gag production of which I did not account for. Therefore, the lack of VLPs or detectable Gag - and for that matter, Env - could also be attributed to such mutation events. While this does not mean that these cell lines do not produce HML-2 - specific protein, in the next chapter I chose to move on to other aspects of this work.

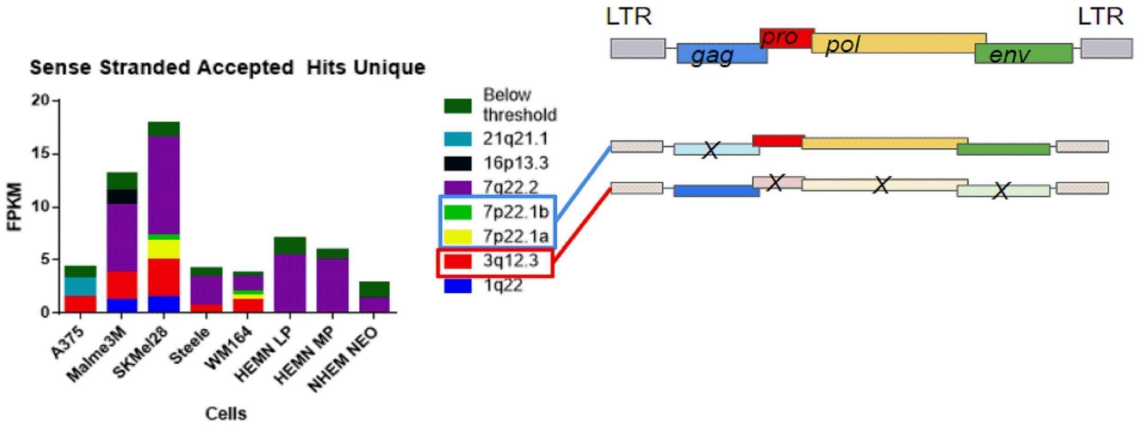


Figure 3.9: Three Proviruses Contain Intact ORFs for Gag, Pol, and/or Env.

Provirus sequences were compared to the consensus sequence HERV-K Con using NCBI ORF Finder to look for the presence of intact ORFs. Full length sequences for the above proviruses were collected from GenBank (see methods for accession numbers), and the full length sequence for HERV-K Con was taken from the Lee et al 2007 paper (160). ORFs were considered “intact” if the length and sequence of an ORF detected with NCBI ORF Finder was > or = 99% identical. Expression data (left) was reprinted to show which cell lines expressed proviruses with intact ORFs. A cartoon representation of a provirus (right, top) is present to act as a guide. A faded box with an X through it indicates an ORF that did not match HERV-K Con by my standards, while a bold colored box represents an ORF that did match my standards.

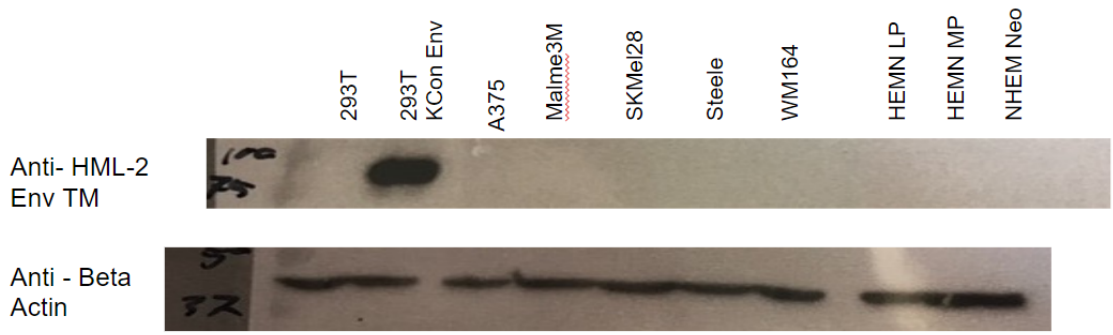


Figure 3.10: HML-2 Env is Undetectable in Anti-Env Western Blot.

Whole cell lysate from melanoma cell lines and primary melanocytes were collected with NP40 buffer, and 30 ugs was loaded onto a 10% SDS-PAGE gel. Whole cell lysate was also collected from 293Ts transfected with empty vector or with HERV-K Con Env (28 ug with Lipofectamine 2000), and 30 ugs of each sample was also loaded. Transfer was performed on PVDF and blocked with 5% nonfat milk in 0.2% TBS-T. Blots were probed overnight in nonfat milk with either Anti-Env (Austral, HERM-1811, 1:2000) or Anti-Beta Actin (1:5000), then for one hour with secondary HRP-conjugated antibody. Blots were visualized using Novex ECL.

Chapter 4: HML-2 Expression in Melanoma is Partially Regulated by Transcription Factor Binding

4.1 HML-2 Expression in Melanoma Cell Lines and Primary Melanocytes is Driven by Three Different Mechanisms

To this point, my RNA-sequencing approach has helped us dissect the HML-2 transcriptome within melanoma cell lines and primary melanocytes. While knowing which proviruses are expressed is important for the development of potential diagnostic tools or therapeutic targets, it is equally important to know if these proviruses are involved in disease etiology. Potentially, provirus expression could affect proper host cell functioning through activation of oncogenic pathways, disruption of a tumor suppressor gene through insertional mutagenesis, or activation of neighboring proto oncogenes (116, 149). Furthermore, understanding the mechanisms behind HML-2 transcription in disease would help explain why they are expressed to begin with. Therefore, I next chose to analyze the mechanisms surrounding HML-2 expression in melanoma.

It should first be mentioned that HML-2 proviruses have a few different mechanisms for their activation. Montesion et al showed this by analyzing their RNA-sequencing data in IGV. Based upon the placement and pattern of reads within the alignment, they noticed that HML-2 proviruses can be expressed by at least four different mechanisms: lncRNA-Associated Transcription, Intronic Transcription, LTR-Driven Transcription, and Read-Through Transcription (Figure 4.1). My criteria for each of these different mechanisms is explained in the Materials and Methods section. For example, in her breast cancer data set the provirus at 1q22 was determined to be expressed due to lncRNA-Associated Transcription. Similar circumstances were also seen for other proviruses such as 12q24.33 driven by Intronic Transcription and 7q34 driven by Read-Through Transcription. In some instances, there was a gap between transcription of an upstream element and a provirus, such as the case with 3q12.3 and

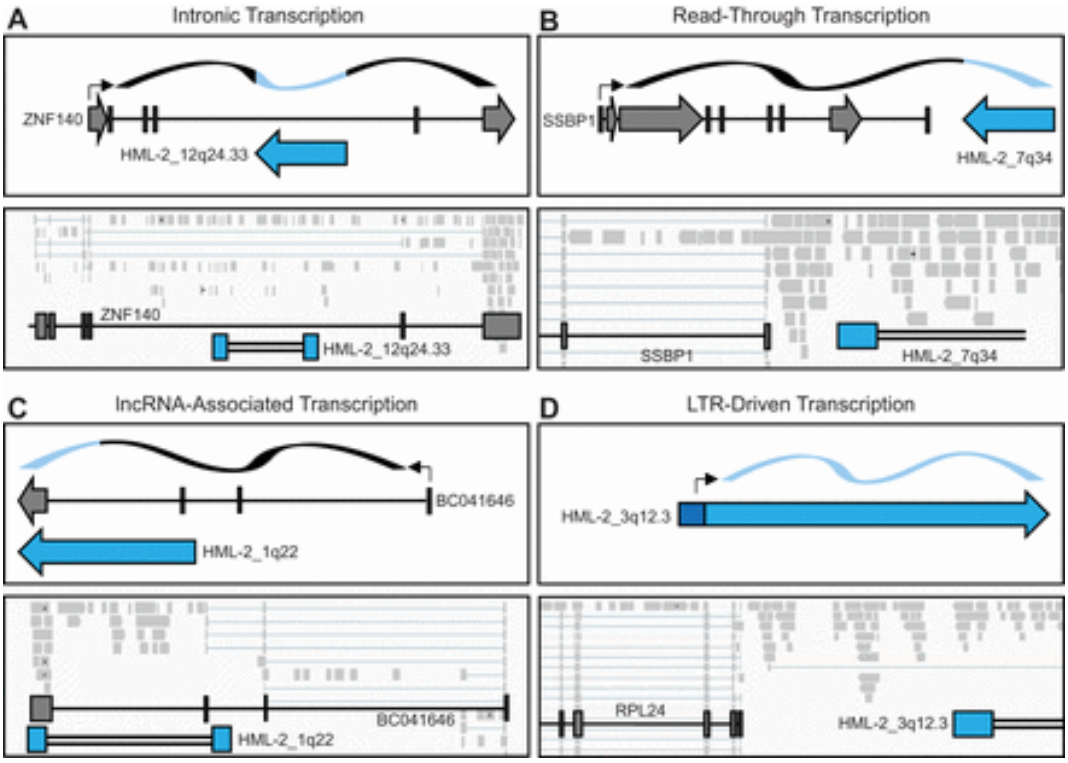


Figure 4.1: Different Mechanisms that Drive HML-2 Expression.

Elements that flank the provirus can promote the expression of a provirus through (A) lncRNA-Associated Transcription, (B) Intronic Transcription, or (D) Read-Through Transcription. In some instances, a provirus is capable of promoting its own expression via it's 5' LTR (C). Blue arrows indicate the direction HML-2 proviruses (blue boxes (LTRs) and double lines) are transcribed. Nearby genes or genetic elements are represented by grey boxes (promoter elements) and a single black line. Names of these elements and proviruses are printed near each element. Grey arrows indicate the direction of transcription for genetic elements. Ribbon in top segment for all panels represents the mRNA transcript with the genetic element (black) sometimes spliced with the HML-2 provirus (blue). The black arrow indicates the location of the transcriptional start site. Grey boxes in the bottom segments for each panel are transcripts aligned to each region. Alignment data is visualized with the Broad Institute's Integrative Genomics Viewer (IGV). Reprinted from Journal of Virology [17] (140). For publication in a non-commercial thesis, permission is not needed.

[17] Montesion, M., Bhardwaj, N., Williams, Z., Kuperwasser, C., and Coffin, J.M. 2018. Mechanisms of HERV-K (HML-2) transcription during human mammary epithelial cell transformation. Journal of Virology. 92(1); e01258-17. DOI: 10.1128/JVI.01258-17

RPL24. In this instance, where reads align to the 5' LTR and there are no reads immediately upstream of the LTR, a provirus is marked as being LTR-Driven. Figure 4.1 summarizes these mechanisms.

Montesion's work led to a few interesting discoveries (140). Most expressed proviruses within her breast cancer model system were transcribed by Read-Through or Intronic Transcription. These proviruses especially included the antisense transcribed proviruses, but there were also a few within her sense transcribed alignment that were expressed by these mechanisms. Additionally, two proviruses within her sense alignment - 3q12.3 and 4p16.1b - appeared to be LTR-Driven. Interestingly, only transformed cells such as the breast cancer cell line Hcc1954 and HMLE-Ras showed this LTR-Driven activity (140). Finally, while neighboring genetic elements had an effect on HML-2 expression there was no indication that HML-2 expression had any effect on neighboring gene activation (140).

Using the guidelines that Montesion established, I looked at the mechanisms driving HML-2 expression in melanoma cell lines and primary melanocytes (Figure 4.2). I found three different mechanisms that drove HML-2 expression in my model system: LTR-Driven Transcription, Intronic Transcription, and Read-Through Transcription. Similar to what Montesion observed, all of the antisense transcribed proviruses were driven by either Read-Through or Intronic Transcription (Figure 4.2, Table 4.1). Read-through and intronic transcription was not restricted to antisense alignment as Read-Through transcribed or Intronic transcribed proviruses were also found within sense alignment transcripts (Table 4.1). For example, 7q22.2 appeared to be Intronic transcribed within *LHFPL3*. Yet further analysis revealed that the proviruses 1q22, 3q12.3, and 7p22.1a/b appeared to be LTR driven in Malme3M, SKMel28, and WM164 (Figure 4.2, Figure 4.3, Table 4.1). Within the antisense transcribed alignment, I noticed the expression of proviruses in the 3' – 5' orientation rather than from the 5' – 3'

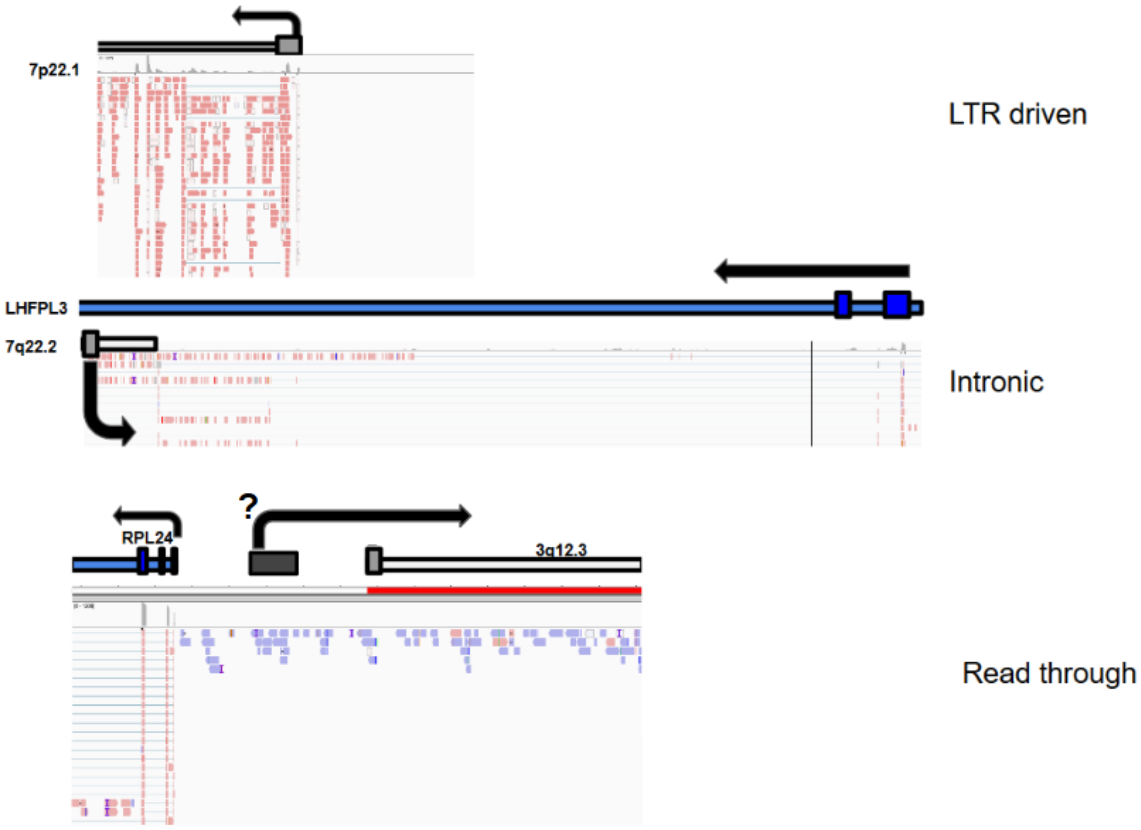


Figure 4.2: Transcription Mechanisms of HML-2 in Melanoma Cell Lines.

RNA-seq data was visualized in the Broad Institute's Integrative Genomics Viewer (IGV) to determine modes of transcription for sense and antisense stranded expressed HML-2 proviruses. The three main modes of transcription I detected in my dataset are represented, with the provirus highlighted in grey and neighboring genes in blue. Transcripts (below cartoon of each genetic element) are colored based on their orientation, where red transcripts are in minus sense to the genome and blue are in positive sense. The dark grey box with a question mark above it represents an unknown element responsible for transcribing 3q12.3 in some cell lines.

Sense transcribed

Cell line	Provirus	Intronic/ Extragenic	Transcription method
A375	3q12.3	Extragenic	Possibly read through (unknown element)
	21q21.1	Intronic (MIR548XHG)	Intronic
Malme3M	1q22	Intronic (BC041646)	**LTR driven**
	3q12.3	Extragenic	**LTR driven**
	7q22.2	Intronic (LHFPL3)	Intronic
	16p13.3	Intronic (FLYWCH1)	Intronic
SKMel28	1q22	Intronic (BC041646)	**LTR driven**
	3q12.3	Extragenic	**LTR driven**
	7p22.1	Extragenic	**LTR driven**
	7q22.2	Intronic (LHFPL3)	Intronic
Steele	3q12.3	Extragenic	Read through
	7q22.2	Intronic (LHFPL3)	Intronic
WM164	3q12.3	Extragenic	Read through
	7p22.1	Extragenic	**LTR driven**
	7q22.2	Intronic (LHFPL3)	Intronic
HEMN LP	7q22.2	Intronic (LHFPL3)	Intronic
HEMN MP	7q22.2	Intronic (LHFPL3)	Intronic
NHEM Neo	7q22.2	Intronic (LHFPL3)	Intronic

Antisense transcribed

Cell line	Provirus	Intronic/ Extragenic	Transcription method
A375	3q21.2	Extragenic	Possibly read into (unknown element)
Malme3M	7q34	Extragenic	Read into
SKMel28	**None expressed**		
Steele	1q21.3	Extragenic	Read into (unknown element)
	7q34	Extragenic	Read into (unknown element / SSBP1)
	10q24.2	Intronic (ABCC2)	Intronic
	12q24.33	Intronic (ZNF140)	Intronic
WM164	**None expressed**		
HEMN LP	**None expressed**		
HEMN MP	5q33.3	Intronic (SGCD)	Intronic
NHEM Neo	**None expressed**		

Table 4.1: Transcription Mechanisms of Sense and Antisense Expressed HML-2s.

Summary of transcription mechanisms used to express each provirus in melanoma cell lines and primary melanocytes. Sense transcribed proviruses are located in the top table, while antisense transcribed are in the bottom. Proviruses that appeared to be LTR Driven in a particular cell line are bolded, and does not mean that it is the only mechanism of transcription for said provirus.

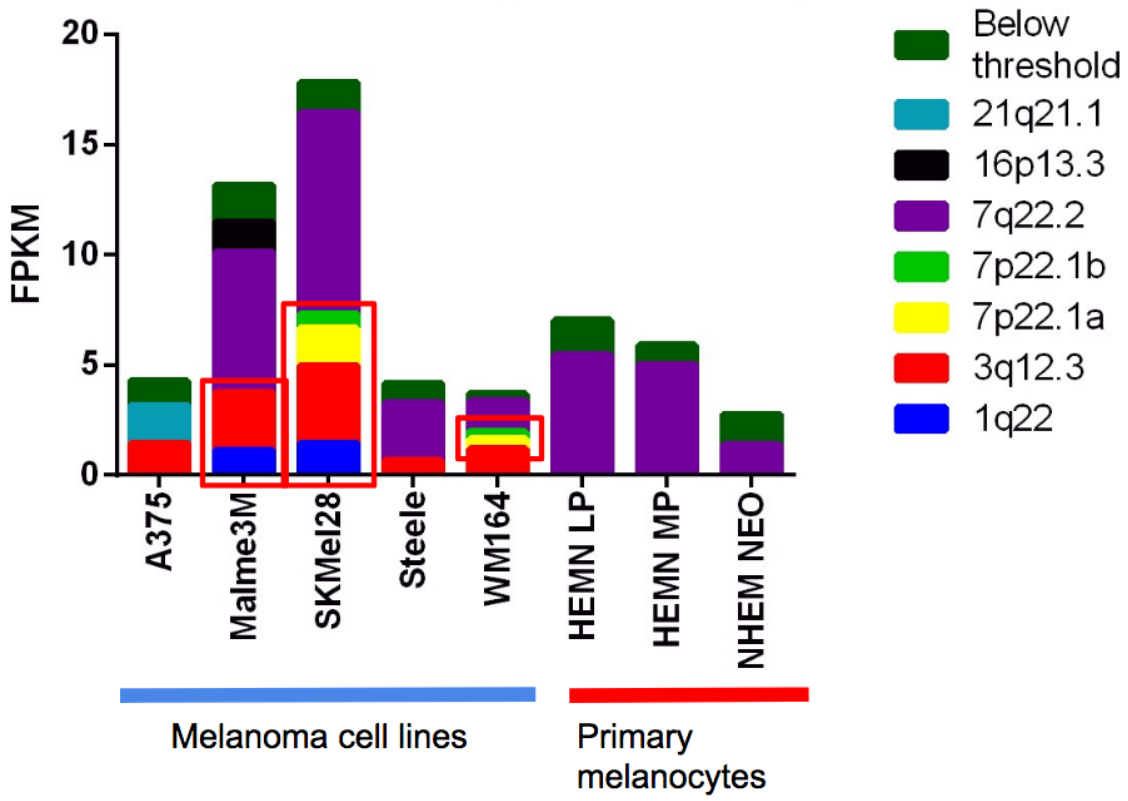


Figure 4.3: LTR-Driven HML-2 Proviruses in Melanoma.

Representation of which proviruses appear to be LTR-driven (red boxes). Figure is reprinted from sense stranded unique only expression data from chapter 3, and is used as a visual representation of where LTR-driven proviruses are expressed in my model system. 3q12.3 - a provirus expressed in all cancer cell lines - appeared to be capable of both LTR-driven transcription and read through transcription depending on the cell line. Cell lines with distinct 3q12.3 LTR-driven transcription were singled out in this analysis, with cell lines that appeared to express 3q12.3 by read through were thrown out of downstream analysis.

orientation from upstream elements, as is the case in Steele cells (Table 4.1). Previously, the Coffin lab has described Read Through transcription as transcribing a genetic element in its 5' – 3' direction. Therefore, expression of proviruses such as 1q21.3 in Steele cells within my antisense alignment are designated as Read Into transcription. Similar to what Montesion reported, I did not see any LTR-Driven activity within my primary cells. Therefore, my IGV data suggests that LTR-Driven activity is restricted to transformed cells within my model systems due to regulatory or expression differences.

Interestingly, proviruses that appeared to be LTR Driven did not appear to use these mechanisms in all of the cell lines they were expressed in, suggesting the involvement of other regulatory mechanisms. For example, 3q12.3 was primarily driven by read through transcription in Steel and WM164 cells while it appeared to be LTR-driven in Malme3M and SKMel28 cells. This appeared to be due to the activity of an upstream unidentified element in Steele and WM164 that does not appear to be active in Malme3M and SKMel28. 3q12.3's use of read through transcription does not mean that 3q12.3 is not capable of LTR-driven transcription in Steele and WM164 cells. Rather read through transcription is more readily apparent in my model system. This could change depending on the lot of Steele and WM164 cells, their passage number, the conditions they are grown in (ex, media type), and the density at the time of harvest.

4.2 Confirming LTR-Driven Proviruses in Melanoma Cell Lines

To confirm that 1q22, 3q12.3, and 7p22.1a/b were LTR-Driven, I used a luciferase activity assay Montesion developed to measure LTR activity *in vitro* (140). LTR activity was measured in four cell lines: SKMel28, WM164, A375, and the breast cancer cell line Hcc1954. I chose these cells since SKMel28 and WM164 appeared to have LTR-Driven

activity, while A375 cells displayed no LTR-Driven activity. Furthermore, I selected Hcc1954 as my control cell line since my chosen LTRs previously exhibited activity in this cell line (data not published). Though Malme3M did appear to have LTR-Driven activity, chemical transfection and electroporation with this assay did not result in detectable luciferase. This was unfortunately the case with most of my cell lines - primary included - so I chose to focus on these four cell lines. To further complicate things, SKMel28s did not transfect easily by chemical means and resulted in low to no detectable luciferase regardless of the transfection agent, the ratio of DNA to transfection agent, or the quantity of plasmid used. Therefore, I elected to use the Nucleofector system – an electroporation type system that requires less DNA while being more sensitive to cells - which has been reported to have great success in transfecting difficult cells.

The 5' LTRs of 3q12.3, 7p22.1b, and 1q22 were synthesized by Genewiz and sub cloned into the Firefly luciferase vector pGL4.17[luc2/Neo] (Figure 4.4, Figure 4.5). I chose to not include the 7p22.1a 5' LTR since transcription starts within the 7p22.1b 5' LTR, though future work should also consider analyzing the 5' LTR activity of 7p22.1a to compare it to 7p22.1b's 5' LTR. The sequences for these LTRs were taken from Genbank, and their accession numbers are located in the Materials and Methods section. As additional controls, I created constructs containing 21q21.1 (a provirus with seemingly no LTR activity in my target cells and only expressed in A375 cells), 11p15.4 (a provirus not expressed in my model system but shows LTR activity in my control cell line), and 7q22.2 (a provirus expressed in all of my melanoma cell lines and appeared to be driven by Intronic Transcription).

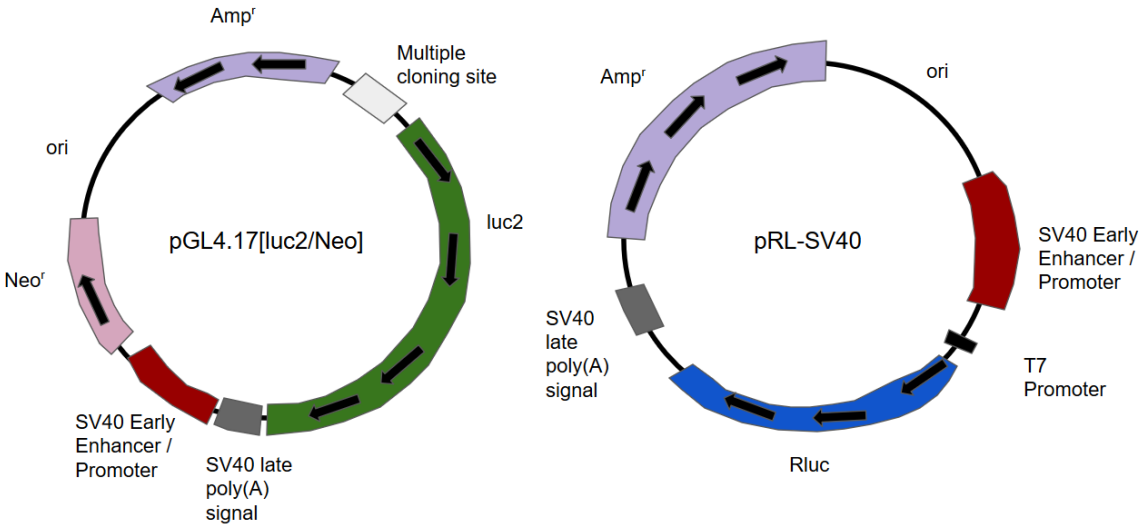


Figure 4.4 Map of Luciferase plasmids used for Dual Promoter Luciferase Assay.

Re-drawn from the Promega manual for respective plasmids. Cartoon representation of the constructs used in the Dual Promoter Luciferase assay, where pGL 4.17 [luc2/Neo] contained the 5' LTR and pRL-SV40 was used to normalize transfection efficiency. 5' LTRs were subcloned into the multiple cloning site of pGL4.17[luc2/Neo] to act as promoters for *luc2*. LTR-pGL4.17[luc2/Neo] constructs were co-transfected with pRL-SV40 to normalize expression of *luc2* to R*luc.*

Our LTR-pGL4.17[luc2/Neo] constructs were co-transfected with the Renilla luciferase construct pGL-SV40 via Nucleofection (Figure 4.4). 48 hours post-Nucleofection, I assayed my cell lines for luciferase activity via the Dual Luciferase Promoter kit's protocol and normalized the expression of Firefly luciferase to Renilla luciferase to determine Relative Light Units (RLUs). An overview of this protocol is presented in Figure 4.5.

I observed that LTR activity in Hcc1954 for 1q22, 3q12.3, 7p22.1b, and 21q21.1 was higher than what was previously reported, however 11p15.4 was consistent with what was previously shown (140, Figure 4.6). 11p15.4 activity was detected in Hcc1954 yet was absent in my melanoma cell lines, which suggests that its absence in my melanoma cell lines is due to a lack of LTR activity. This lack of LTR activity could be attributed to differential expression of cell type - specific transcription factors which are important in 11p15.4 activity between Hcc1954 and my melanoma cell lines. As for the difference in LTR activity between my work and Montesion et al work, this could be due to sequences differences between my LTR constructs and the ones she used as the LTR sequences for Montesion et al work was cloned from Tera-1s.

Measurement of LTR activity within my selected melanoma cell lines revealed that several of them display activity (Figure 4.6). Though 21q21.1 was only found to be expressed in A375, the corresponding LTR did not show activity within this cell line and instead exhibited activity in WM164, SKMel28, and Hcc1954. Within A375, this provirus was shown to be activated by Intronic Transcription, therefore the lack of LTR activity in A375 suggests that this LTR is not active in this cell line and this provirus relies upon its neighboring gene for its expression. Since this LTR displayed activity in SKMel28, WM164, and Hcc1954, it is also quite possible that epigenetic factors such as transcription factor binding, transcription factor expression, or other epigenetic inhibitors are responsible for its activity in my model system.

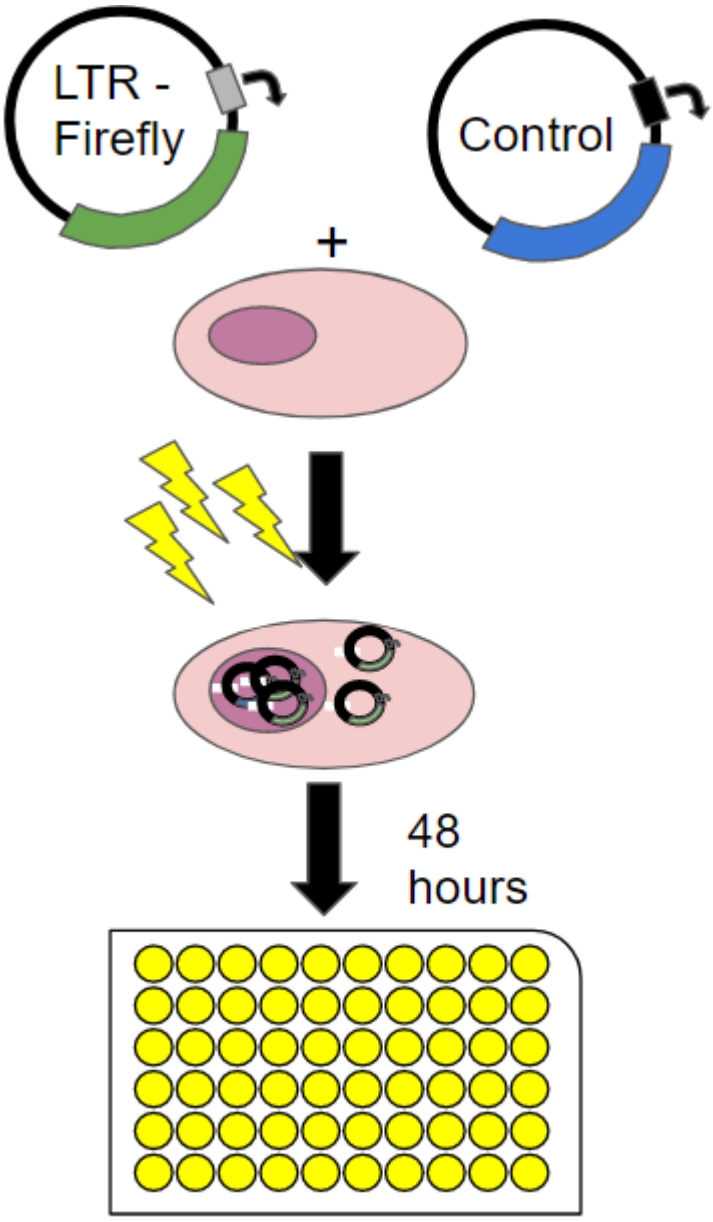


Figure 4.5: Protocol for HML-2 LTR Nucleofection.

Schematic of transfection protocol by Nucleofection. The 5' LTRs (grey box) were synthesized by Genewiz and subcloned into the Firefly luciferase (green box) vector pGL 4.17 [luc2/Neo]. Cells were co-transfected with this construct and pRL-SV40 which contains Renilla luciferase (blue box) driven by the early SV40 promoter (black box). Vectors were co-Nucleofected at a ratio of 30:1 (pGL 4.17 [luc2/Neo] : pRL-SV40). 48 hours post-Nucleofection, cells were harvested and assayed for Firefly and Renilla luciferase expression with the Promega Dual Promoter Luciferase kit.

.

The 5' LTR of 7q22.2 exhibited activity in all of my cell lines (Figure 4.6). This was surprising as my IGV analysis indicated that this provirus relied upon Intronic Transcription (Figure 4.2). Though this LTR has activity, this LTR is not participating in 7q22.2 expression as it is present in reverse orientation to the rest of the provirus (Figure 4.7). It is possible that enhancer elements could still bind to this LTR and affect 7q22.2 expression but considering the apparent effect *LHFPL3* had on its transcription this effect would be minimal. 3q12.3 LTR activity was detected in all of my cell lines, which was anticipated as this LTR was previously shown to be functional in Montesion's breast cancer model and 3q12.3 was expressed in all four of these cell lines. Finally, while 1q22 was only expressed in SKMel28, Malme3M, and Hcc1954 this LTR also displayed activity in all of my transfected cell lines.

The most surprising result originated from 7p22.1b (Figure 4.6). Though I expected LTR-Driven activity from 7p22.1b, I was not expecting to the extent I observed. This LTR showed very high activity in WM164 and SKMel28 with some activity in Hcc1954 and A375. Since A375 did not express 7p22.1 and WM164 showed low expression, the potential role of epigenetic regulation became more possible within my model system. The role of epigenetic regulation was supported by a large difference between LTR activity and provirus expression in my melanoma cell lines (Figure 4.8, Table 4.2). RLUs were consistently higher than FPKM for most cell lines with a few exceptions: 3q12.3 activity in A375, 7q22.2 in SKMel28, and 21q21.1 in A375. 3q12.3 in A375 were equivalent, indicating that this expression is possibly due to regulatory factors. As for 7q22.2 in SKMel28 and 21q21.1 in A375, the higher expression of these proviruses compared to their RLUs is consistent with Read-Through Transcription.

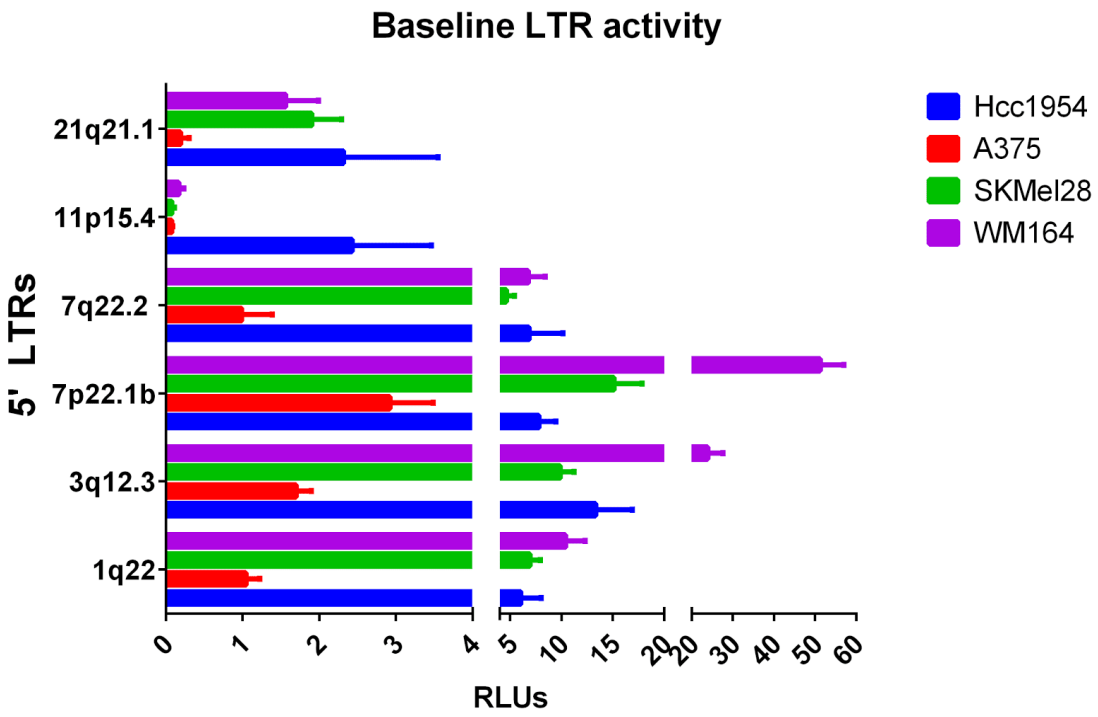


Figure 4.6: Baseline Activity of 5' LTRs within Melanoma Cell Lines.

Luciferase activity of HML-2 5' LTRs within three melanoma cell lines, and one breast cancer cell line (Hcc1954). Raw Firefly luciferase and raw Renilla luciferase activity was normalized to an empty vector control where activity of the empty vector was subtracted from the sample vector. Relative light units (RLUs) were calculated as (normalized Firefly luciferase activity/ normalized Renilla luciferase activity) and averaged across replicates. N = minimum of 9.

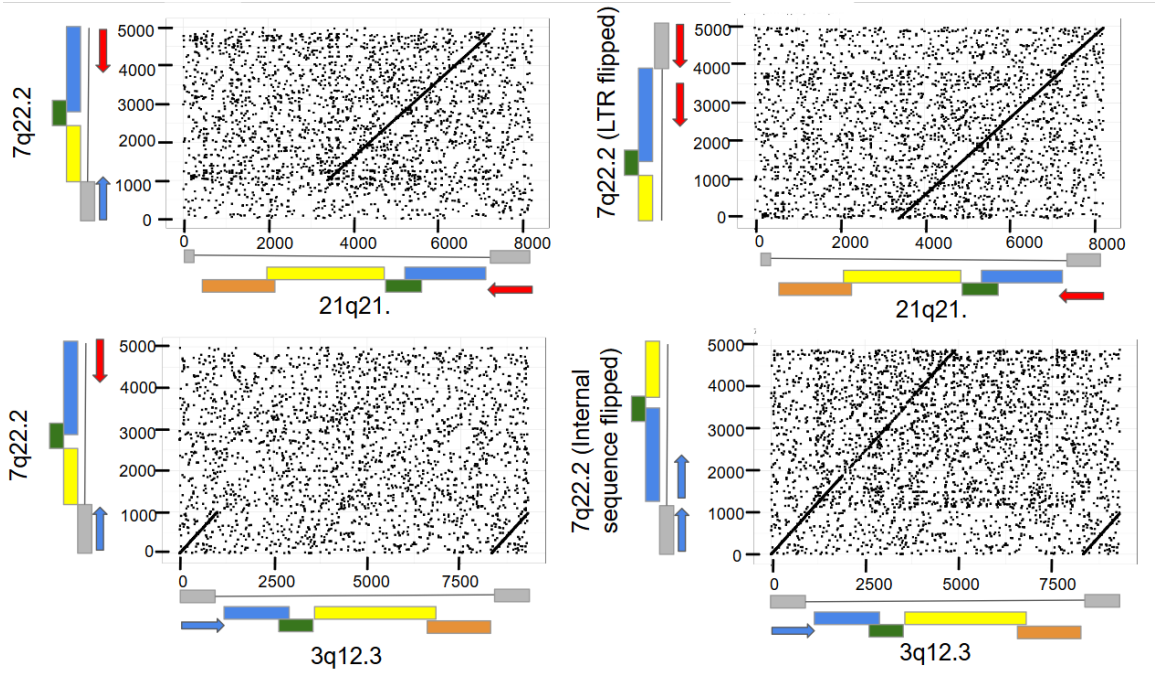


Figure 4.7: The 5' LTR of 7q22.2 is in the opposite orientation of its internal sequence.

To show 7q22.2's 5' LTR orientation to its internal sequence, I aligned 7q22.2 to 3q12.3 and 21q21.1. These proviruses were chosen because they exist entirely on the + strand (3q12.3) or the – strand (21q21.1), and their 5' LTR sequences are closely phylogenetically related to 7q22.2 (34). Sequences for these proviruses were obtained from the UCSC Genome Browser, and the alignment and dot plot were generated in R using ggplots2. Cartoons of these proviruses are shown along the X and Y axis of each dot plot. Grey box = LTR, blue box = *gag*, green box = *pro*, yellow box = *pol*, orange box = *env.* Arrows indicate the strand each provirus is on, where red indicates the elements exist on the – strand and blue indicates elements exist on the + strand.

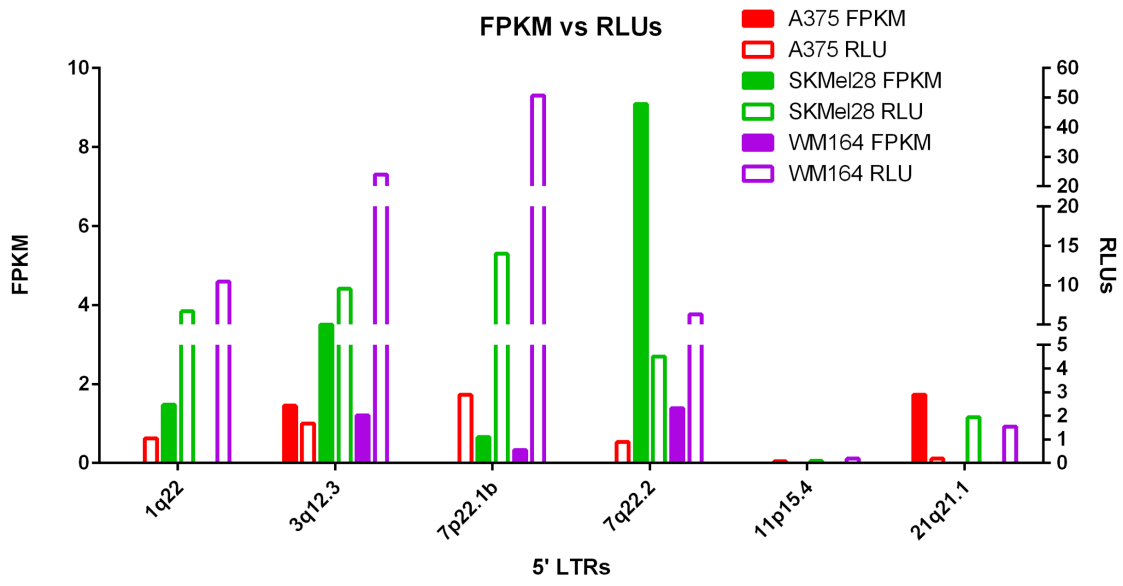


Figure 4.8: Comparison of HML-2 FPKM and RLUs in Melanoma Cell Lines.

Comparison of provirus expression (FPKM, left Y axis) to LTR activity (RLUs, right Y axis) for each 5' LTR tested in my Dual Promoter Luciferase assay. Hollow bars represent RLUs for cell lines, while solid bars represent FPKM for cell lines. Colors are matched for easier comparison.

LTRs	A375 FPKM	A375 RLU	SKMel28 FPKM	SKMel28 RLU	WM164 FPKM	WM164 RLU
1q22	0.000	1.039	1.478	6.700	0.000	10.453
3q12.3	1.454	1.677	3.509	9.543	1.212	23.977
7p22.1b	0.000	2.887	0.655	14.006	0.333	50.652
7q22.2	0.000	0.892	9.095	4.515	1.391	6.335
11p15.4	0.000	0.079	0.000	0.089	0.000	0.190
21q21.1	1.731	0.183	0.000	1.934	0.000	1.541

Table 4.2: Comparison of HML-2 FPKM and RLUs in Melanoma Cell Lines.

Raw values for FPKM and RLU comparison.

When I combined the observed difference in HML-2 expression across my model system, the difference in LTR activity across my transfected cells, and the high RLU activity relative to FPKM it suggested to me that transcription factors could play a large role in LTR – Driven HML-2 transcription in melanoma. As the effect of transcription factor binding on LTR activity was confirmed in Montesion et al's breast cancer work, I chose to pursue this prospect.

4.3 Detection and Evolution of Unique Transcription Factor Binding Sites

Before Montesion began her RNA-sequencing work, Ravi Subramanian - a postdoctoral fellow in the Coffin lab - identified HML-2 transcripts expressed in breast cancer cell lines using single genome sequencing (data not published). Using the top 10 expressed proviruses - of which 3q12.3 and 11p15.4 were among them - Montesion was able to show that several of their LTRs were active in breast cancer cell lines and transformed cells with no activity in human mammary epithelial cells (HMEs). Due to differential HML-2 expression between breast cancer cell lines and HMEs and LTR-driven activity, she hypothesized that differential expression and LTR activity is due to transcription factors expressed in transformed cells but not in non-cancerous cells. Since these proviruses showed both 5' LTR activity and were expressed in breast cancer cell lines, she chose to explore this using the 5' LTRs of 3q12.3 and 11p15.4 (data not published).

She used a combination of unique transcription factor binding sites (TFBS) that she identified using a motif finding software known as Genomatix MatInspector and site directed mutagenesis to remove these sites (data not published). These unique TFBS were not specifically unique to HML-2s but were unique to the subset of LTRs she was analyzing. While there were several unique sites within 3q12.3's and 11p15.4's 5' LTR, she chose to analyze four sites due to significant transcription factor differential

expression between transformed and non-transformed cells: HOX_PBX and RFX3 for 3q12.3, and ATF and RORA for 11p15.4. The removal of these sites did correlate with a significant decrease in LTR activity within transformed cells, yet this decrease was not seen in Hcc1954, possibly due to other transcription factors that compensated for their loss of binding.

To see if unique TFBS could be responsible for HML-2 LTR activity in my model system, I searched for unique motifs within my four LTRs of interest using Genomatix MatInspector (Table 4.3). By doing so, I was able to identify a list of 39 motifs that were unique. Most of these sites were due to base substitutions with the exception of HOX_PBX in 3q12.3 which as Montesion previously showed was created subsequent to integration due to a sequence duplication event (Figure 4.9).

To see if these unique TFBS were acquired pre- or post-insertion, I compared the 5' and 3' LTR sequences within each provirus. Since 7q22.2 does not have two LTRs, I reserved this analysis for 1q22, 7p22.1b, and 3q12.3. Differences between 5' and 3' LTRs were 99.69% (1q22), 98.23% (3q12.3), and 99.79% (7p22.1b). Sequence analysis of proviruses with two LTRs revealed that while most of these motifs were present at insertion due to their presence in both LTRs, there were two motifs that were acquired after insertion: HOX_PBX in 3q12.3 and ZBRK1 in 1q22 (Table 4.4).
Interestingly, I observed that some of the unique sites in 3q12.3 were acquired across primate divergence. The E2F binding site is present within humans yet is absent in chimpanzees and gorillas which indicates that this site was acquired no later than five million years ago. This event occurs after 3q12.3's approximate integration time of more than eight million years ago, which suggests the acquisition of this site is due to human-specific events (34). Additionally, the EGR1 binding site is present in both chimpanzees and humans but not gorillas, suggesting that this site was acquired in the

Provirus	Unique Transcription Factor Binding Site
1q22	**ZBRK1**, POU2F3, **LTATA**
3q12.3	TAL1_E2A, TGIF, TEF_HLF, HOXC9, TR2, MYRF, SIX4, MRG1, P53, PTF1, **HOX_PBX**, TBX20, OC2, HNF3B, SF1, SOX7, **RFX3**, INSM1, **EGR1**, HSF2, TEAD, **E2F**, SZF1, PRDM4, **SOX10**, PROX1, SOX18, CDE
7p22.1b	**GAGA, HNF4, GLI3, EKLF, SIX1**
7q22.2	**IR1_NGRE**, CTCF, **CEBP**

Table 4.3: List of Unique Transcription Factor Binding Sites in 5' LTRs of Expressed HML-2s.

Unique transcription factor binding sites (TFBS) were identified with Genomatic MatInspector using 5' LTR sequences from the above proviruses. LTR sequences were acquired from Genbank, and their accession numbers are located in Section 2.8. Unique TFBS were defined as sites only found in one of the above four LTRs. The motifs for bolded transcription factors were back mutated in the above LTRs with site-directed mutagenesis to eliminate them.

Provirus	Binding site	5’ and 3’ LTR sequences	Binding site evolution
1q22	ZBRK1	5’:GGGAAAGACCTGACTGTCCCCCAGC 3’:GGGAAAGACCTGACCGTCCCCCAGC	**Acquired post insertion**
1q22	LTATA	5’: TAGTATAAGAGGAAGGA 3’: TAGTATAAGAGGAAGGA	Present at insertion
3q12.3	HOX_PBX	5’: AACCCGATTGATTGTAC 3’: AACCC--------GATTGTAC	**Acquired post insertion**
3q12.3	RFX3	5’: CTTGTGACCATGACACATC 3’: CTTGTGACCATGACACATC	Present at insertion
3q12.3	EGR1	5’: GAGAAACACCCACGAATGA 3’: GAGAAACACCCACGAATGA	Present at insertion
3q12.3	E2F	5’: AGACGGCGCGGATCCTC 3’: AGACGGCGCGGATCCTC	Present at insertion
3q12.3	SOX10	5’: GATTGTACGTTCCAT 3’: GATTGTACGTTCCAT	Present at insertion
7p22.1b	HNF4A/ GLI3/ KLF1	5’:GAAACATGTGCTGTGTCCACTCAGG 3’:GAAACATGTGCTGTGTCCACTCAGG	Present at insertion
7p22.1b	SIX1	5’: ATGTGATAGCCTGAA 3’: ATGTGATAGCCTGAA	Present at insertion
7p22.1b	GAGA	5’: GTCCCCTTCTTTCTTTCTCTATACT 3’: GTCCCCTTCTTTCTTTCTCTATACT	Present at insertion

Table 4.4: Evolution of Unique Transcription Factor Binding Sites.

List of differences between the 5’ and 3’ LTRs for my proviruses of interest. 5’ and 3’ LTRs for 1q22, 7p22.1b, and 3q12.3 were aligned in Clustal Omega to determine sequence similarity between LTRs of the same provirus. Motifs for the above transcription factors present in both LTRs were present at the time of insertion, while motifs only present in the 5’ LTR were acquired. Differences between the two sequences are highlighted in red.

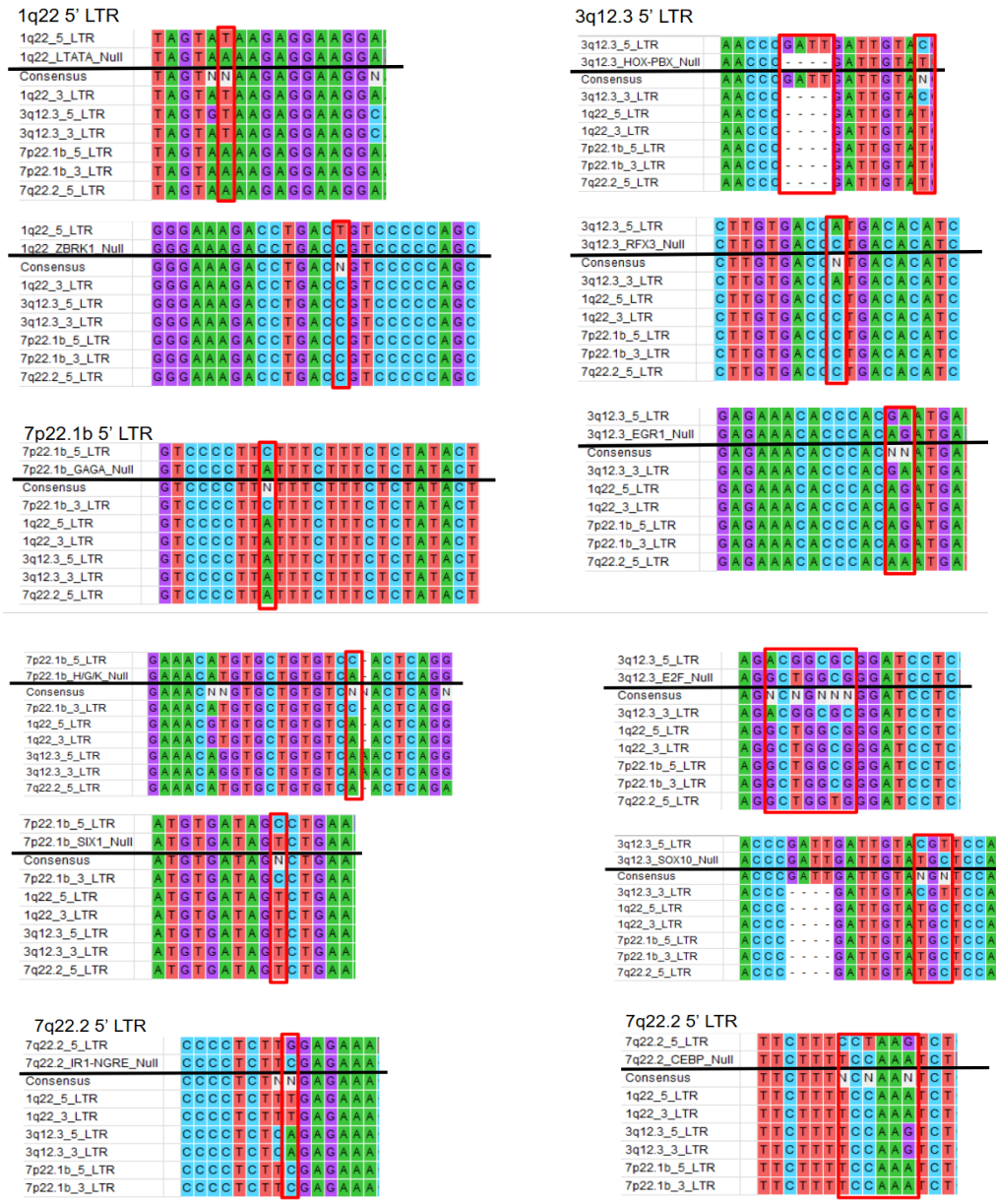


Figure 4.9: Mutations Responsible for Unique Transcription Factor Binding Sites in 5' LTRs.

Comparison of LTRs to identify mutations responsible for the generation of unique TFBS. Line separates 5' LTR of interest from 5' and 3' LTRs used in analysis. Consensus sequence was used to back mutate the 5' LTR to remove the unique TFBS. Above the black line is the 5' LTR sequence with the unique site removed to show which changes were made to remove the motif, also known as the "Null" sequence. Changes are highlighted in red.

human-chimpanzee common ancestor after their divergence from gorillas approximately six million years ago. Finally, while SOX10 is a unique TFBS to 3q12.3 compared to my other LTRs of interest, this site is present in humans, chimpanzees, and gorillas and was therefore not acquired by 3q12.3 post-integration into the primate lineage.

4.4 Removal of Unique Transcription Factor Binding Sites by Site Directed Mutagenesis Reduces but Does Not Eliminate LTR Activity

The identification of unique TFBS helped us to identify a few candidates for the one(s) responsible for differential expression in my model system. To identify which unique TFBS I would target, I looked for transcription factor expression that roughly correlated with provirus expression (Figure 4.10). Doing so, I identified nine unique sites to target (Figure 4.10, Table 4.3).

Most of these unique sites were identified in the 3q12.3 LTR with a few sites in the 5' LTRs for 1q22 and the melanocyte-origin expressed provirus, 7q22.2, that appeared to be of interest. Most TF that correlated with HML-2 differential expression was located within 3q12.3 due to the numerous isoforms of HOX_PBX and E2F. Furthermore, TF such as CEBP in 7q22.2 also showed a correlation between HML-2 differential expression. Interestingly, I could not identify any unique TFBSs that correlated with 7p22.1 expression, so I chose to target all of the unique TFBSs within this LTR.

To test if these unique TFBSs were critical for HML-2 activity in my model system, I used the Q5 Site-Directed Mutagenesis kit from NEB to revert the sequence in the target LTR to a consensus sequence developed among my four LTRs of interest (Figure 4.9). Point mutations were reverted to the consensus sequence and insertions were deleted. In the instance where there was no consensus, the sequence was changed to something of minimal difference between the other sequences while ablating

the unique TFBS. Once these constructs were synthesized, they along with their wild type counterpart were transfected into my target cell lines as before (Figure 4.11). Interestingly, mutagenesis of unique sites in 1q22 did not reduce LTR activity but rather increased it. In this case, it was possible that the mutation of unique TFBS allowed for the binding of other transcription factors that resulted in an increase in LTR activity. When I analyzed the mutated TFBSs for the creation of new TFBSs, I noticed that the T-A mutation used to remove the LTATA binding site resulted in the creation of a site for BARBIE (Barbituate-Inducible Element), while the T-C substitution to remove the ZBRK1 binding site resulted in the removal of all TFBS that overlapped with the ZBRK1 site. Interestingly, I learned mid-writing that ZBRK1 is a transcriptional repressor so the acquisition of this site post-insertion in 1q22 is interesting. Therefore, the increase in 1q22 LTR activity upon its removal should not come as a surprise.

Mutagenesis of unique TFBSs for 7q22.2 and 7p22.1b resulted in a slight decrease in LTR activity but did not result in the elimination of LTR activity as was previously seen for other LTRs in transformed mammary epithelium cell lines. Mutagenesis of unique TFBSs in 3q12.3's 5' LTR also resulted in lower activity but did not eliminated LTR activity as was previously shown in transformed mammary epithelium cell lines. Interestingly, the removal of HOX_PBX and RFX3 did result in a steep decrease in LTR activity - especially in Hcc1954 which was not seen in Montesion's work - yet the removal of these sites did not eliminate expression of this LTR as was previously seen in transformed mammary epithelium cell lines. This lack of strong difference in LTR activity between mutant and wild type sequences suggests that while these transcription factors do play a role in LTR activity in my model system, they are not the most important regulatory factors in the expression of these HML-2 proviruses. It is possible that there are other transcription factors that are more

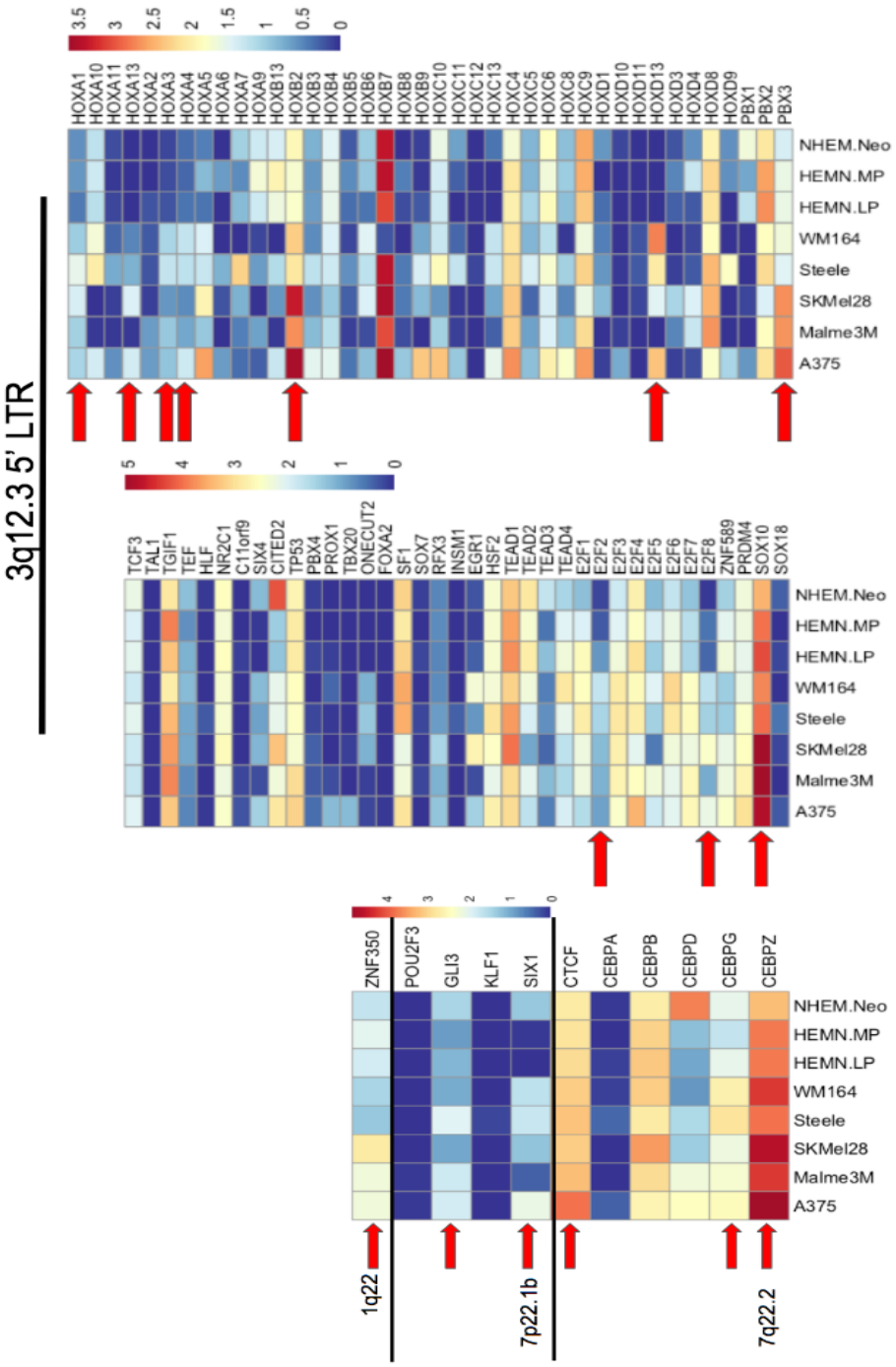


Figure 4.10: Transcription Factor Expression in Melanoma and Primary Melanocytes.

Heatmap of transcription factor expression in melanoma cell lines and primary melanocytes. Transcription factors with available expression date in my alignment are plotted above using pheatmaps in RStudio and separated by their LTR of origin. Red arrows indicate transcription factors where there were notable expression differences between melanoma cell lines and primary melanocytes.

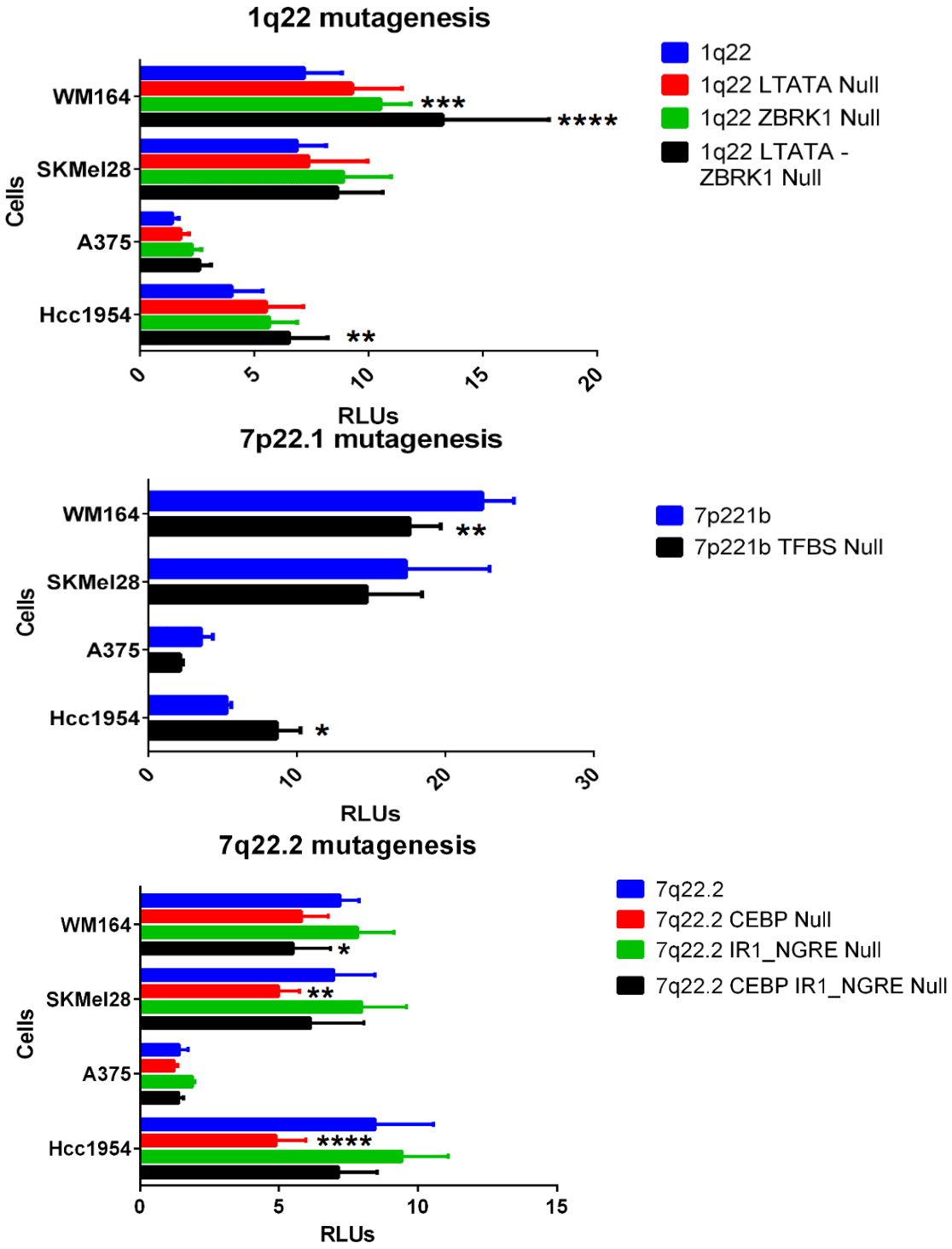


Figure 4.11: LTR Activity is Reduced but Not Eliminated by Site Directed Mutagenesis of Unique Transcription Factor Binding Sites.

Continued on next page along with figure legend.

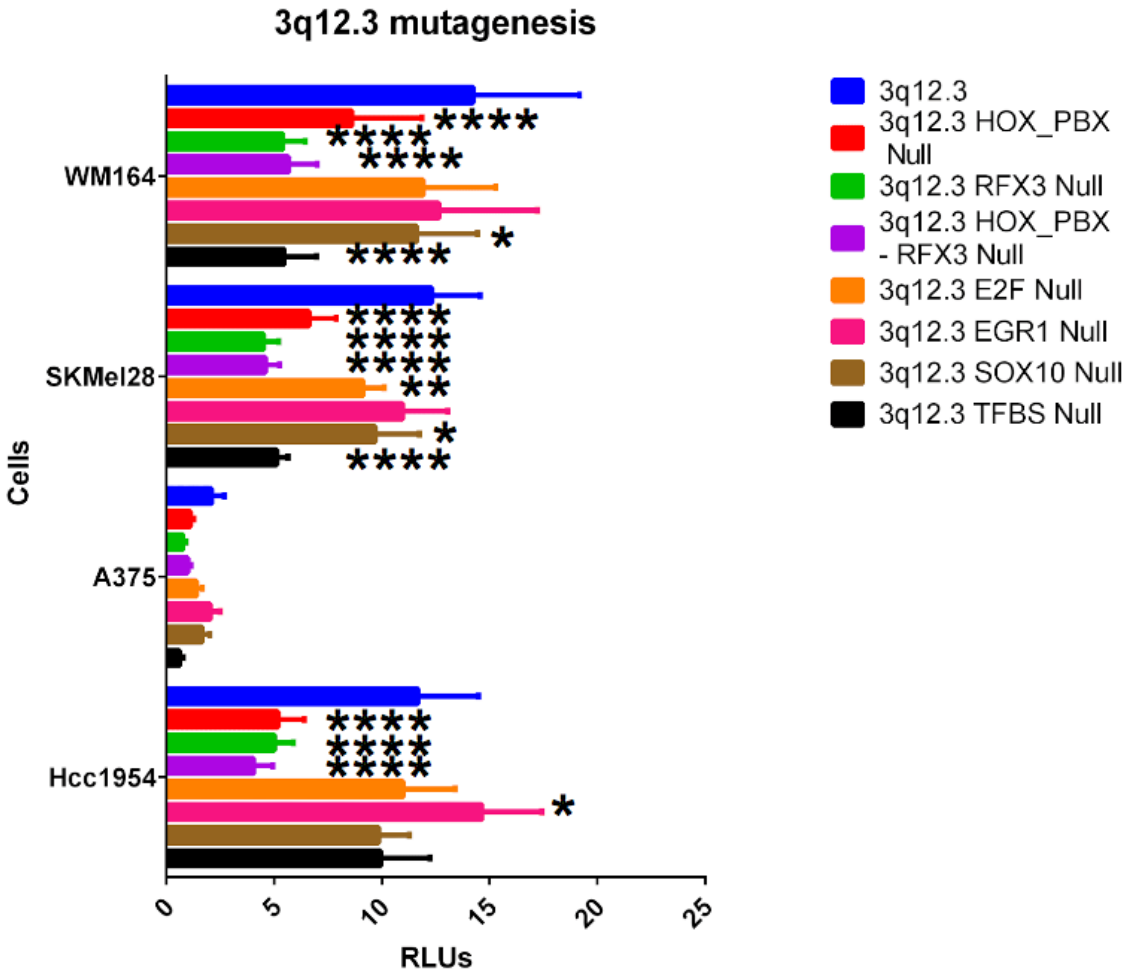


Figure 4.11: LTR Activity is Reduced but Not Eliminated by Site Directed Mutagenesis of Unique Transcription Factor Binding Sites.

5' LTRs underwent site directed mutagenesis to remove unique transcription factor binding sites of interest and were nucleofected into melanoma cell lines to measure their activity with Promega's Dual Promoter Luciferase Kit. Relative light units (RLUs) were calculated as (normalize Firefly luciferase activity/ normalized Renilla luciferase activity) and averaged across replicates. N = minimum of 9. Empty vector signal and background subtracted to normalize RLUs. Significance analysis was performed in Prism GraphPad by two way ANOVA with Bonferroni's multiple comparisons. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.

responsible for HML-2 activity, and quite possible that other epigenetic regulators such as DNA methylation is involved in their activity.

Chapter 5: Discussion

Human endogenous retroviruses (HERVs) are remnants of exogenous retrovirus infection into the germ line of primates (20, 22, 37). Most HERV proviruses date back tens of millions of years ago and contain mutations that range from point mutations that affect protein production to large truncations and recombination events that result in the formation of solo LTR sequences. In this respect, the HERV-K (HML-2) subclade is unique: integration time of these proviruses range from millions of years to a few hundred thousand years ago, earning them the title of the clade with the youngest, most human specific proviruses in which some sequences contain intact ORFs for *gag, pro, pol*, and/or *env* (34 - 36, 129, 133, 134, 137, 140).

Both HML-2 - specific RNA transcripts and protein have been detected in several cancer *in vitro* models (130, 132 -138, 140, 142-148, 150). From prostate cancer to germ cell tumors, HML-2 specific RNA transcripts have been detected by Sanger sequencing, amplicon sequencing, and qRT-PCR. In some instances, specific provirus expression has been isolated such as the detection of 22q11.21 by both Ruprecht and Bhardwaj in HML-2 - specific particles budding from Tera-1s (135, 137). In other cases, HML-2 expression has been shown to change between non-cancerous and cancerous cells. For example, Montesion showed that specific HML-2 proviruses were expressed in human mammary epithelium and that the process of transforming to a cancerous state shifted the HML-2 transcriptome from predominantly expressing 1q22 to expressing 3q12.3 and 1q21.3 (140).

This shift in provirus transcription introduces an interesting possibility. The shift in transcription suggests that the process of transformation results in a change of HML-2 transcription due to a shift in the gene expression profile from non-cancerous to cancerous. These changes probably involve the expression of enhancer or repressor transcription factors, the expression of pioneer transcription factors, and epigenetic markers, to name a few. In my thesis and in Montesion's thesis, we were only able to look at the effect of unique transcription factor binding in HML-2 LTR activity - what I assumed would have correlated with HML-2 expression. To an extent it did: the removal of unique transcription factor binding sites did have an effect to some degree on LTR activity in my transfection experiments, a result that Montesion also observed. However, the effect of other transcription factors were overlooked. These other non-unique transcription factors including those that were shared by my LTRs of interest could have had an effect on LTR activity, which might be able to explain the continued presence of signal for the HOX-PBX and RFX3 mutants in 3q12.3. Furthermore, I was unable to investigate the epigenetic modifications affecting these proviruses which could have further explained the differences in HML-2 expression and LTR activity through regulating access of the transcriptional machinery to the provirus. For example, histone modification or DNA methylation around 7p22.1 in WM164 could explain the dramatic difference between FPKM and RLUs. However, this simplifies the true situation in that it overlooks pioneer transcription factors, such as FOX and Gal4. Since all of my proviruses of interest contain at least one FOX site in their 5' LTR, it is also possible that expression of pioneer transcription factors in melanoma would also affect HML-2 expression.

Considering how connected epigenetic modifications, expression, and transcription factors are, I would hypothesize that each have their respective role to play in HML-2 expression. The benefit of exploring their regulation further would allow for a

better understanding of the interplay between epigenetic modifications and transcription factors. I would suggest future work explore the effect of non-unique transcription factor binding, focusing on the factors that correlate the best with provirus expression in tissue states, while exploring the epigenome around HML-2 proviruses. Since the effect of DNA methylation of HERV activity has been shown before, I would recommend beginning with this modification before exploring the positioning of nucleosomes and histone modifications. Finally, I would recommend focusing on a few key pioneer transcription factors that are differentially expressed in cancer vs non-cancerous cells to see what effect their expression has on HML-2 activity. Forming these three facets into one complete mechanisms would be difficult, but it would also help us understand the different factors involved in their activity.

HERV-K - specific transcripts, viral - like particles (VLPs), and structural proteins have been detected in melanoma cell lines as well as in patient samples since the 1970s (138, 139 142-148). Reports of VLPs produced in melanoma were first detected in tumor tissue and have been detected in a select group of melanoma cell lines as well (138, 143-145). For example, the melanoma cell lines SKMel28 and 518A2 have been shown to produce retroviral-like particles that are HERV-K in origin (138). Though not infectious, they were capable of packaging 7p22.1 and 19p12b RNA transcripts, which makes sense considering 19p12b contains an intact ORF for *gag*. In addition to transcripts, Env and Gag were shown to be expressed in melanoma cell lines and patient tissue samples yet were not found in non-cancerous cells (138, 146). Possibly owing to their expression in tumors, anti-HERV-K Env TM antibodies were detected in approximately 20% of melanoma patient sera with a correlation between increased antibody detection and cancer progression (146, 147). Between their increased expression in melanoma and the existence of cancer-associated ERVs in other mammals, HERV-K (HML-2) finds itself in the crosshairs of virologist and oncologists alike.

Increased HML-2 expression in several types of cancer relative to healthy cells is established yet detecting their expression does not begin to dissect their function in disease. There are at least 96 proviruses within the human genome, of which some are exceedingly rare in the human population (34-36). Their rarity could be attributed to their detrimental effect on host health and would be selected against over time, presenting in only the smallest subsets of humans. Alternatively, their rarity could be due to their relatively recent integration time, at which point they might not be fixed in the human population. Furthermore, expression of proviruses near proto oncogenes could also affect host health through dysregulation of gene expression. Finally, if one wished to use these sequences to develop a new diagnostic tool for early detection or as a new therapeutic target, specific proviruses expressed in diseased vs non-diseased cells must be identified. Therefore, identifying specific HML-2 sequences expressed in disease is critical to determining their pathogenic role and potential for patient care.

Recently, the HML-2 transcriptome in melanoma tumor and cell lines was inspected using cDNA cloning frequency (139). From this study, the authors observed 24 expressed HML-2 proviruses in both melanoma and primary melanocytes (cell line and patient samples alike). While informative, cloning frequency is not as sensitive or quantitative as other transcriptomic analysis methods, rendering some proviruses as undetectable due to their low expression level. Proviruses with low expression that may play a role in disease would go undetected and therefore HML-2's role in disease could go unnoticed. Cloning frequency also prevents the determination of their effect on neighboring gene activation and grasping the mechanisms behind their activation would be nearly impossible. Finally, since HML-2 proviruses are very similar to each other – especially the relatively recently integrated proviruses – reads could easily map multiple times to one site. Due to the high sequence similarity between HML-2 RNA transcripts,

not accounting for multimapping would result in an overrepresentation of expressed proviruses and would lead to confusion over expression of specific proviruses.

Montesion and Bhardwaj took issue with the scarcely illuminated HML-2 transcriptome and the scant knowledge surrounding their expression mechanisms in illness, so they developed an RNA-sequencing pipeline to solve these problems (137, 140). Using Tera-1s and a series of breast cancer cell lines, they identified a series of HML-2 proviruses that were expressed in each cell type. For example, Bhardwaj was able to show that 22q11.21, 22q11.23, and 19p12c were predominantly expressed in Tera-1 cellular RNA, while HML-2 RNA packaged into VLPs predominantly originated from 22q11.21. Such work not only exposes the HML-2 transcriptome in Tera-1s, but also allows virologists to examine HML-2 packaging into virions.

The RNA-sequencing pipeline that Bhardwaj and Montesion developed has numerous benefits. Their pipeline allows for sensitive detection of HML-2 sequences while elucidating the cause of their expression in cell lines (140). For example, Montesion was able to show that while most HML-2 proviruses expressed in breast cancer were driven by *trans*-acting elements, there were some capable of activation by *cis*-acting elements such as their corresponding 5' LTR. Furthermore, she was able to show the role transcription factor binding has on HML-2 LTR activity, such as the binding of transcription factors from the HOX/PBX gene family to their canonical motif in 3q12.3's 5' LTR dramatically affects its LTR activity.

Additionally, the software used in my pipeline is well established and has been used in other transcriptomic analysis. While there are several aligners available, Tophat2 allows for both swift alignment and mapping of spliced reads to reduce read loss. Compared to other sequencers such as HiSeq, MiSeq provides long reads (up to 3x longer than HiSeq) which reduces the incidence of incorrect mapping and multimapping. At the time Montesion and Bhardwaj developed their pipeline, Illumina MiSeq was the

sequencing platform of choice. They chose this system because at the time it was the only tabletop sequencer I had available that had the highest fidelity and produced the longest reads. Currently, PacBio's Nanopore sequencer has lapped MiSeq in terms of read length. Therefore, future sequencing could be attempted on this platform to get reads of higher fidelity and longer length, which is critical for sequences that are highly similar to each other. Finally, to reduce the incidence of false positives, the expression data from unique alignments were used to measure HML-2 expression intensity. While focusing on uniquely aligning reads reduces the incidence of false positives, it also underestimates the true expression level of these proviruses. Future work such as HML-2 expression in other model systems should involve improved mapping of multi aligning reads to the genome to get a better sense of total HML-2 expression.

Since the Coffin lab has developed a pipeline well suited to analyze the HML-2 transcriptome, I chose to study melanoma - a cancer studied for its HML-2 expression - to see how it relates to Tera-1 cells and breast cancer. I selected a series of melanoma cell lines and primary melanocyte populations where cell lines all contained the $BRAF^{V600E}$ activating mutation, and primary melanocytes were derived from single donor tissue. I detected HML-2 transcripts from 51 proviruses, and comparison to Schmitt's work displayed high similarity between my results and theirs (Figure 3.2, Figure 3.3). I detected the expression of all but three proviruses they reported as transcribed: 1p31.1, 12q14.1, and 19q12 (Figure 3.2). Since RNA transcripts for these three proviruses were detected in patient tumor samples, I probably did not detect transcripts specific to them because they were not expressed in my cell lines. Due to their polymorphic nature, it is possible that they simply are not found within my model system. I was also able to detect the expression of an additional 31 proviruses missed in the cloning frequency study. This is most likely due to the higher sensitivity afforded us through RNA-sequencing and possible transcriptional differences between my chosen cells. Fortunately, there was

significant overlap with some variation, indicating that my work expanded upon the known HML-2 transcriptome in melanoma.

The presence of such an overlap is interesting. Though I was unable to detect three proviruses, these proviruses - of which are polymorphic - originated from patient samples. The inclusion of patient samples in my data set could have confirmed their expression or could have introduced the expression of other polymorphic proviruses that have yet to be detected. Furthermore, the overlap between my data set and the cDNA cloning frequency data set indicate that these proviruses are certainly expressed in cells of melanocyte origin. Finally, the abundance of other proviruses that I detected as expressed using my RNA sequencing library justifies my work and indicates how diverse HML-2 expression is in cell lines. This also justifies the use of cloning frequency as a mile high view of HML-2 expression in cell lines, while showing how much can be missed due to cDNA cloning frequency's detection limit.

Comparison of the HML-2 transcriptome between melanoma and primary melanocytes indicated that the HML-2 transcriptome is much more diverse in melanoma. This difference in transcription is not surprising considering epigenetic modifications - such as DNA methylation - are often aberrantly modified in cancer, resulting in aberrant expression of otherwise silenced genes. This difference is also potentially due to differential expression in transcription factors between primary cells and cancerous cell lines. Most expressed proviruses were either poorly expressed or poorly covered, leaving only a few proviruses that were expressed at a high enough level for further analysis (i.e. above my threshold) (Figure 3.4, Figure 3.5). To isolate proviruses that could potentially express HML-2 - specific proteins while also understanding the mechanisms behind HML-2 expression in my model system, I divided my proviruses into sense and antisense transcribed sequences. Doing so revealed that most provirus expression was from sense transcribed proviruses, while there were a few HML-2

proviruses expressed in antisense orientation. The presence of more sense transcribed proviruses compared to antisense transcribed proviruses stands in contrast with what Montesion found in breast cancer, where most proviruses were expressed in antisense orientation. This difference between sense transcribed being most common in melanoma cell lines and antisense transcribed being most common in breast cancer cell lines could be due to a difference in the transcriptional profile of these cells, where cellular requirements for growth and needed cellular functions vary.

All of my melanoma and primary melanocytes expressed RNA transcripts from the 7q22.2 provirus with the exception of A375 (Figure 3.4). I initially thought that the absence of 7q22.2 transcripts were due to a lack of *LHFPL3* expression, yet 7q22.2 is expressed in Steele where *LHFPL3* expression is also absent. This provirus appears to rely upon Intronic transcription, so nonexistence in A375 is partially due to the lack of *LHFPL3* expression. Developing a potential explanation for its expression in Steele where *LHFPL3* is also not expressed required some digging into the structure of 7q22.2. This provirus contains a 5' LTR, yet this LTR is in reverse orientation to the proviral internal sequence (Figure 4.7). As can happen when a provirus inserts near a proto oncogene, it's possible that 7q22.2's 5' LTR has a small effect on 7q22.2 expression through the binding of an enhancer element to its LTR, however its effect on 7q22.2 expression would be minimal compared to *LHFPL3*'s effect on 7q22.2's expression. Furthermore, I was unable to tell what promoter was responsible for 7q22.2 expression in Steele cells so I would need to determine what that is to conclude the role an enhancer element would have on 7q22.2 expression.

Expression of other proviruses such as 3q12.3 was reserved for my melanoma cell lines. In regard to 3q12.3, Montesion observed low level expression in human mammary epithelium (HME) while the breast cancer cell line Hcc1954 predominantly expressed 3q12.3. Comparison between melanoma cell lines, primary melanocytes,

breast cancer cell lines, HME, and Tera-1 reveal that 3q12.3 is expressed in my cancer cell lines only (Figure 3.7). What may account for 3q12.3's expression is the presence of a HOX_PBX binding site within 3q12.3's 5' LTR which was generated from a duplication event after insertion into the primate genome and divergence of human-chimpanzee lineages. HOX and PBX gene families encode transcription factors that have become associated with a wide range of cancers due to their dysregulation and have been shown to play a role in cancer development. Montesion's work into the relationship between transcription factor binding and LTR activity showed the importance of this HOX_PBX binding motif in 3q12.3's 5' LTR activity (data not published). Therefore, 3q12.3s activity in cancer is simultaneously interesting and not surprising: the binding of the pro-oncogenic HOX/PBX transcription factors has been shown by Montesion to affect 3q12.3 LTR activity, opening the possibility of its use as a biomarker or a therapeutic target.

Akin to Montesion's work in breast cancer, I did not observe the expression of rare proviruses within my model system through comparison of the hg19 alignment to the custom built HML-2 genome derived from a screen of over 2500 individuals (Figure 3.6). The lack of rare provirus expression not only suggests that the hg19 alignment was sufficient for my purposes, but that there is no detectable expression of rare proviruses that could play a role in pathogenesis.

A deeper dive into HML-2 differential expression among cells in my cancer data set revealed a diverse profile where each cancer and cell type have their own signature. For example, breast cancer cell lines predominantly express 3q12.3 while melanoma cell lines express an array of proviruses from 3q12.3 to 7p22.1 (Figure 3.7). Furthermore, primary cells each expressed their own unique HML-2 profile: while melanocytes primarily expressed 7q22.2, primary breast epithelium expressed 1q22. These differences are possibly due to cell - specific signaling as it has been previously shown

that epigenetics and transcription factors play a large role in HML-2 expression (72, 111, 137, 140). Alternatively, neighboring gene activity could be responsible for HML-2 expression.

Since all of my melanoma samples were cell lines, it would be interesting to look into patient samples to see how the HML-2 transcriptome differs. Cell lines themselves are drastically different from what they were when they were initially collected, considering they have been grown in a synthetic environment for decades at a time and became highly aneuploid. While using cell lines is still incredibly valuable, it can be widely different from the true disease occurring in individuals. There has been a fair amount of work done with patient tissue samples, but none analyze the HML-2 transcriptome. Therefore, it would be interesting to use a similar analysis pipeline to see what the HML-2 transcriptome is in tissue samples. Finally, while there are differences in HML-2 expression between cancerous and non-cancerous cells, it is still unclear if this difference is sufficient for therapeutic targeting or diagnostics. In this regard, proviruses with intact ORFs that express HML-2 specific proteins could be useful as they would provide a therapeutic target. In cases such as melanoma tissue and cell lines where HML-2 Env and Gag have been detected, these could be a good target. To clarify this, it would be necessary to amass a larger set of patient samples with patient matched controls across a series of cancers.

Within my melanoma samples, most expressed proviruses were expressed in sense orientation, were young, and were human specific. Furthermore, I identified the expression of two proviruses that contain an intact ORF for Pro, Pol, and Env (7p22.1) (Figure 3.8, Figure 3.9). By my screening methods, 3q12.3 also technically contains an intact ORF for Gag but due to other mutations outside of this ORF it is incapable of producing Gag protein. The presence of intact ORFs stands in contrast with breast cancer, where most of the expressed proviruses were old and did not contain intact

ORFs. Despite previous reports showing the detection of VLPs in melanoma cell lines and patient tissue samples, I was unable to detect expression of another provirus that contains an intact ORF for Gag. This would explain the absence of VLPs in conditioned supernatant and Gag in whole cell lysate. As for the expression of Env, I was also unable to detect it by Western blot (Figure 3.10). SKMel28 was previously shown to produce VLPs and Env, so their absence in my work was noticeable. It's possible that they are undetectable in my samples due to differences in lot number, handling, antibody, or extraction techniques. One way to test for it would be qPCR to look for the corresponding mRNA. Regardless, I suggest that expression of Env is due to 7p22.1.

Montesion was able to show that HML-2 proviruses are expressed by at least four different mechanisms: lncRNA-Associated, Intronic, LTR-Driven, and Read-Through Transcription (Figure 4.1). HML-2 proviruses expressed in melanoma rely on three of these processes: Read-Through, Intronic, or LTR-Driven (Figure 4.2). Similar to what Montesion showed, antisense transcribed proviruses are expressed through read through or intronic transcription. Most proviruses expressed in sense orientation depend on intronic or read through transcription, yet a few proviruses - 1q22, 3q12.3, and 7p22.1 - appeared to rely on LTR-driven mechanisms (Table 4.1, Figure 4.3). Interestingly, LTR-driven proviruses were only found in my melanoma cell lines. This coincided with Montesion's work where she only saw the expression of LTR-driven proviruses in her transformed cell lines (140). LTR-driven activity being detectable only in melanoma cell lines is probably due to transcriptional differences between non-cancerous and cancerous cell lines. Interestingly, none of my sense transcribed proviruses appeared to drive the expression of neighboring genes, an event also seen in her breast cancer work. The lack of HML-2 activity not driving neighboring gene expression does not rule

out the role their RNA or protein expression could have in cancer etiology, however this potentially rules out their role in affecting nearby oncogenic target gene expression.

To test their activity, the 5' LTRs of potentially LTR-Driven proviruses were cloned into a Firefly luciferase vector and co-transfected into a subset of melanoma cell lines (Figure 4.6). They proved to be very active, especially in WM164 cells. Interestingly, HML-2 expression and LTR activity were dramatically different within WM164, where LTR activity was exceptionally high for proviruses with relatively low expression (Figure 4.8). For example, the expression of 7p22.1b was low in WM164 yet its 5' LTR activity was the highest recorded. Based on this result, I first thought that this divergence could be due to differential transcription factor expression or binding. This was a likely explanation since Montesion uncovered the importance of transcription factor binding to LTRs, where their binding significantly affected their activity. Due to their differential expression across cell lines, I began looking at unique transcription factor binding sites (TFBS). I identified 39 motifs, of which 14 were selected for further screening owing to disparate transcription factor expression across melanoma and primary melanocytes (Table 4.3).

The origin of these motifs is mostly due to point mutations between different proviruses except for the HOX_PBX site which was created from a post-integration duplication event found only in humans. Site-directed mutagenesis of unique TFBS indicated that transcription factor binding does play a small role in HML-2 expression in melanoma, but not as large of a role as it does in breast cancer (Figure 4.11). Interestingly, 1q22 seems to have acquired a TFBS for ZBRK1, which is a transcriptional inhibitor. Considering that removal of this site in my site-directed mutagenesis experiments increased LTR activity for 1q22, it is possible that 1q22 acquired this mutation because its expression was detrimental to the host. The acquisition of transcriptional inhibition motifs could help explain the lack of circulating HML-2 virions

and their poor expression. Therefore, it would be interesting to screen other 5' LTRs for acquired transcriptional repressor sites.

By now, retrovirologists, oncologists, and molecular biologists alike have amassed a large data set showing expression of HML-2s in various cancers such as breast, melanoma, germ, liver, and teratocarcinoma to name a few. Since there is at least some effect on HML-2 expression derived from epigenetic changes which are often seen in cancer, I would hypothesize that HML-2 expression would be detected in more cancers than what have currently been tested. In this respect, I would like to suggest that more cancers are investigated for HML-2 expression and studied to determine the underlying causes of their expression, of which I would hypothesize further has to do with epigenetic modifications involving disease and regulatory elements such as transcription factors. By doing so, I would expect to identify a range of HML-2s that are expressed in cancer tissues where there would be some overlap based on my work analyzing the differences between melanoma, breast cancer, and Tera-1s. How extensive this overlap will be is unknown, though I would expect that the more cancers that are surveyed the more overlap one would see in their expression due to common expression of transcription factors and other pathways that are shared between cancers. Rather than focusing only on cancerous tissues, I would also suggest studying the HML-2 transcriptome in their healthy counterparts to identify proviruses that are expressed in cancerous and not non-cancerous cells. In this way, the community could better evaluate the use of these proviruses for either diagnostic or therapeutic purposes.

Speaking of their potential use in medicine, based upon my data showing the differences and similarities in HML-2 expression between breast cancer cell lines, melanoma cell lines, and Tera-1s I would suggest that it could be possible to use provirus expression as a marker for cancer diagnosis. Since these three cancers expressed 3q12.3, it's possible that this could be a shared cancerous marker. However,

this data was accumulated from a series of cell lines which have evolved to rely on an artificial system and therefore could have different expression profiles compared to cancerous tissue. In the future, I would suggest using a pipeline similar to the one Montesion and Bhardwaj developed to study the HML-2 transcriptome in cancerous tissues. This would require patient samples that contain a sufficient number of cells since HML-2 proviruses are expressed in such low abundance compared to the rest of the cellular genome. To also investigate their potential purpose as a therapeutic target, I would also suggest obtaining patient-matched non-cancerous control tissue.

To determine if HML-2 provirus expression could be used as a diagnostic, I would also recommend further investigation into HML-2 expression within cancer patient serum. In this regard I can imagine two different scenarios: free floating RNA within patient blood, which has been observed to exist within cancer patients, and in circulating cancerous cells. Looking for their presence within patient blood where the potential of liquid biopsies are being pursued could improve patient diagnosis as an inclusion of such a test in a routine complete blood count.

In conclusion, HML-2 expression in melanoma cell lines is primarily due to a handful of proviruses that are young, sense transcribed, and contain intact ORFs except for 3q12.3 which is more than eight MYO. Comparison of the HML-2 transcriptome across breast cancer, HMEs, melanocytes, melanoma, and Tera-1 cells show that most cancerous cell lines express 3q12.3 while non-cancerous cells typically express one provirus that differs between cell types possibly due to transcription factor expression. Proviruses expressed in melanoma and melanocytes are mostly driven by read through or intronic transcription, but there are three proviruses - 1q22, 3q12.3, and 7p22.1 - that are LTR driven in melanoma cell lines. The 5’ LTRs for 1q22, 3q12.3, and 7p22.1b appear to be active partially due to unique TFBS, but the removal of these motifs does not completely eliminate LTR activity. Therefore, there are other elements that contribute

to their activity in cancer. my data identify a few targets for downstream analysis to determine the possibility of using expressed HML-2 proviruses as biomarkers or therapeutic targets. Future prospects for determining the HML-2 transcriptome in cancer would result in new detection methods and therapies, especially if any expressed proviruses contain intact ORFs. Furthermore, studying the mechanisms of expression behind HML-2 activity could help us better understand how HML-2s are regulated, resulting in a better understanding of how retroviruses are regulated in humans.

References

1. **Vogt, P.K.**, 1997. Historical Introduction to the General Properties of Retroviruses. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY).
2. **Vogt, V.M.**, 1997. Retroviral Virions and Genomes. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY).
3. **Johnson, W.E., and Coffin, J.M.** 1999. Constructing primate phylogenies from ancient retrovirus sequences. PNAS. 96(18): 10254-10260.
4. **Vogt, P.K.** 2012. Retroviral Oncogenesis: A Historical Primer. Nature Reviews Cancer. 12(9):639-648.
5. **Weiss, R.A. and Vogt, P.K**. 2011. 100 years of Rous sarcoma virus. Journal of Experimental Medicine. 208 (12): 2351-2355.
6. **Rubin. H.** 2011. The early history of tumor virology: Rous, RIF, and RAV. PNAS. 108(35): 14389-14396.
7. **Coffin, J.M. and Fan, H.** 2016. The Discovery of Reverse Transcriptase. Annual Review of Virology. 3:29-51.
8. **Vahlne, A**. 2009. A historical reflection on the discovery of human retroviruses. Retrovirology. 6:40.
9. **Coffin. J.M.** 2015. The discovery of HTLV-1, the first pathogenic human retrovirus. PNAS. 112(51):15525-15529.
10. **Montagnier, L.** 2002. A History of HIV Discovery. Science. 298:1727-1728.
11. **Gallo, R.C.** 2002. The Early Years of HIV/AIDS. Science. 298:1727-1728.
12. **Junqueira, D.M. and Almeida, S.E.d.M**. 2016. HIV-1 subtype B: Traces of a pandemic. Virology. 495: 173-184.
13. **Junqueira. D.M., de Medeiros, R.M., Matte, M.C.C., Araujo, L.A.L., Chies, J.A.B., Ashton-Prolla, P., Almeida, S.E.d.M.** 2011. Reviewing the History of HIV-1: Spread of Subtype B in the Americas. PLoS ONE. 6(11):e27489.
14. **Centers for Disease Control and Prevention.** (2017, September, 27th). *HIV Among Gay and Bisexual Men*. Retrieved on: 12.12.2017. Retrieved from:
15. **Centers for Disease Control and Prevention**. (1981, June, 5th). *Pneumocycstis Pneumonia - Los Angeles*. Retrieved on: 12.12.2017. Retrieved from: https://www.cdc.gov/mmwr/preview/mmwrhtml/june_5.htm
16. **Altman, L.K**. (1981, July, 3rd). *Rare Cancer Seen in 41 Homosexuals*. Retrieved on: 12.12.2017. Retrieved from: http://www.nytimes.com/1981/07/03/us/rare-cancer-seen-in-41-homosexuals.html
17. **Centers for Disease Control and Prevention.** (1981, July 3rd). *Kaposi's Sarcoma and Pneumocycstis Pneumonia Among Homosexual Men - New York City and California*. Retrieved on: 12.12.2017. Retrieved from: https://stacks.cdc.gov/view/cdc/1265
18. **Coffin, J.M., Haase, A., Levy, J.a., Montagnier, L., Oroszlan, S., Teich, N., Temin, H., Toyoshima, K., Varmus, H., Vogt, P., and Weiss, R.** 1986. Human Immunodeficiency Viruses. Science. 232(4751): 697.
19. **Weiss, R.A**. 2006. The Discovery of Endogenous Retroviruses. Retrovirology. 3:67.
20. **Boeke, J.D., and Stoye, J.S.** 1997. Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY).

21. **Dewannieux, M., and Heidmann, T.** 2013. Endogenous retroviruses: acquisition, amplification, and taming of genome invaders. Current Opinion in Virology. 3: 646-656.
22. **Bannert N., and Kurth R.** 2006. The Evolutionary Dynamics of Human Endogenous Retroviral Families. Annu. Rev. Genomics Hum. Genet. 7:149-173.
23. **Tarlintona,R., Meersb, J., and Young, P.** 2008. Biology and evolution of the endogenous koala retrovirus. Cell. Mol. Life Sci. 65: 3413 – 3421
24. **Xu,W., Stadler, C.K., Gorman, K., Jensen, N., Kim, D., Zheng,H.,Tang, S., Switzer, W.M., Pye, G.W., and Eidena, M.V.** 2013. An exogenous retrovirus isolated from koalas with malignant neoplasias in a US zoo. Proc Natl Acad Sci U S A. 110(28): 11547–11552.
25. **Anai, Y., Ochi, H., Watanabe, S., Nakagawa, S., Kawamura, M., Gojobori, T., and Nishigakia, K.** 2012. Infectious Endogenous Retroviruses in Cats and Emergence of Recombinant Viruses. J. Virol. 86(16): 8634-8644
26. **Denner, J.** 2016. Transspecies Transmission of Gammaretroviruses and the Origin of the Gibbon Ape Leukaemia Virus (GaLV) and the Koala Retrovirus (KoRV). Viruses. 8(12): 336.
27. **Denner. J., and Young, P.R.** 2013. Koala retroviruses: characterization and impact on the life of koalas. Retrovirology. 10:108.
28. **Kassiotis, G.** 2014. Endogenous Retroviruses and the Development of Cancer. Journal of Immunology. 192:1343-1349.
29. **Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A,**

Batzoglou S, **Birney E**, **Bork P**, **Brown DG**, **Burge CB**, **Cerutti L**, **Chen HC**, **Church D**, **Clamp M**, **Copley RR**, **Doerks T**, **Eddy SR**, **Eichler EE**, **Furey TS**, **Galagan J**, **Gilbert JG**, **Harmon C**, **Hayashizaki Y**, **Haussler D**, **Hermjakob H**, **Hokamp K**, **Jang W**, **Johnson LS**, **Jones TA**, **Kasif S**, **Kaspryzk A**, **Kennedy S**, **Kent WJ**, **Kitts P**, **Koonin EV**, **Korf I**, **Kulp D**, **Lancet D**, **Lowe TM**, **McLysaght A**, **Mikkelsen T**, **Moran JV**, **Mulder N**, **Pollara VJ**, **Ponting CP**, **Schuler G**, **Schultz J**, **Slater G**, **Smit AF**, **Stupka E**, **Szustakowki J**, **Thierry-Mieg D**, **Thierry-Mieg J**, **Wagner L**, **Wallis J**, **Wheeler R**, **Williams A**, **Wolf YI**, **Wolfe KH**, **Yang SP**, **Yeh RF**, **Collins F**, **Guyer MS**, **Peterson J**, **Felsenfeld A**, **Wetterstrand KA**, **Patrinos A**, **Morgan MJ**, **de Jong P**, **Catanese JJ**, **Osoegawa K**, **Shizuya H**, **Choi S**, **Chen YJ**, **Szustakowki J**; **International Human Genome Sequencing Consortium**. 2001. Initial sequencing and analysis of the human genome. Nature. 409:860-921.

30. **Armezzani, A., Varela, M., Spencer, T.E., Palmarini, M., and Arnaud, F.** 2014. “Menage a Trois”: The Evolutionary Interplay between JSRV, enJSRV and Domestic Sheep. Viruses. 6(12), 4926-4945.
31. **Arnaud, F., Varela, M., Spencer, T.E., and Palmarini, M.** 2008. Coevolution of endogenous Betaretroviruses of sheep and their host. 65(21): 3422–3432.
32. **Varela, M., Spencer, T.E., Palmarini, M., and Arnaud, F.** 2009. The Special Relationship between Endogenous Retroviruses and Their Host. Ann. N.Y. Acad. Sci. 1178: 157–172.
33. **Tandon, R., Cattori, V., Willi, B., Lutz, H., and Hofmann-Lehmann, R.** 2008. Quantification of endogenous and exogenous feline leukemia virus sequences by real-time PCR assays. Veterinary Immunology and Immunopathology 123 (2008): 129–133.
34. **Subramanian R., Wildschutte J., Russo C., and Coffin J.M.** 2011. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. Retrovirology. 8:90.
35. **Wildschutte, J. H., Ram, D., Subramanian, R., Stevens, V. L., and Coffin, J. M.** 2014. The distribution of insertionally polymorphic endogenous retroviruses in breast cancer patients and cancer-free controls. Retrovirology. 11:62.
36. **Wildschutte, J. H., Williams, Z. H., Montesion, M., Subramanian, R. P., Kidd, J. M., and Coffin, J. M.** 2016. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. PNAS. E2333.
37. **Hohn, O., Hanke, K., and Bannert, N.**, 2013. HERV-K(HML-2), the best preserved family of HERVs: endogenization, expression, and implications in health and disease. Frontiers in Oncology. Vol 3, Article 246.
38. **Bannert, N., and Kurth, R.** 2004. Retroelements and the human genome: New perspectives on an old relation. PNAS. 101: 14572–14579.
39. **de Parseval, N., and Heidmann, T.** 2003. Human endogenous retroviruses: from infectious elements to human genes. Cytogenetic and Genome Research. 110: 318-332.
40. **Coffin, J.M.** 2003. Evolution of Retroviruses: Fossils in my DNA. Proceedings of the American Philosophical Society. 148(3): 264-280.
41. **Jern, P. and Coffin, J.M.** 2008. Effects of Retroviruses on Host Genome Function. Annual Review of Genetics. 42(2008): 709-732.
42. **Belshaw, R., Dawson, A.L.A., Woolven-Allen, J., Redding, J., Burt, A., and Tristem, M**. 2005. Genomewide Screening Reveals High Levels of Insertional Polymorphism in the Human Endogenous Retrovirus Family HERV-K(HML2): Implications for Present-Day Activity. Journal of Virology. 79(19):12507-12514.

43. **Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A., and Tristem, M.** 2004. Long-term reinfection of the human genome by endogenous retroviruses. PNAS. 101(14): 4894-4899.
44. **Deininger, P.L., and Batzer, M.A.** 2002. Mammalian Retroelements. Genome Research. 12: 1455-1465.
45. **Belshaw, R., Katzourakis, A., Paces, J., Burt, A., and Tristem, M.** 2005. High Copy Number in Human Endogenous Retrovirus Families is Associated with Copying Mechanisms in Addition to Reinfection. Molecular Biology and Evolution. 22(4):814-817.
46. **Pavlicek, A., Paces, J., Elleder, D., and Hejnar, J.** 2002. Processed pseudogenes of human endogenous retroviruses generated by LINEs: their integration, stability, and distribution. Genome Research. 12: 391-399.
47. **Mang, R., Maas, J., van der Kuyl, A.C., and Goudsmit, J**. 2000. *Papio cynocephalus* Endogenous Retrovirus among Old World Monkeys: Evidence for Coevolution and Ancient Cross-Species Transmissions. 74(3): 1578-1586.
48. **Tarlinton, R.E., Meers, J., and Young, P.R.** 2006. Retroviral invasion of the koala genome. Nature. 442:79-81.
49. **Tristem, M.** 2000. Identification and Characterization of Novel Human Endogenous Retrovirus Families by Phylogenetic Screening of the Human Genome Mapping Project Database. Journal of Virology. 74(8): 3715-3730.
50. **Hughes, J.F., and Coffin, J.M.** 2005. Human Endogenous Retroviral Elements as Indicators of Ectopic Recombination Events in the Primate Genome. Genetics. 171(3): 1183-1194.
51. **Sverdlov, E. D.**, 2000 Retroviruses and primate evolution. *BioEssays* 22**:** 161–171.
52. **Lavie, L., Medstrand, P. Schempp, W., Meese, E., and Mayer, J.** 2004 Human endogenous retrovirus family HERV-K(HML-5): status, evolution, and reconstruction of an ancient betaretrovirus in the human genome. *J. Virol.* 78**:** 8788–8798.
53. **Vogt, P.K., and Friis, R.R.** 1971. An avian leukosis virus related to RSV(0). Properties and evidence for helper activity. Rirology. 43: 223-234.
54. **Coffin, J.M., Champion, M.A., and Chabot, F.** 1978. Nucleotide sequence relationships between the genomes of an endogenous and exogenous avian tumor virus. Journal of Virology. 28: 972-991.
55. **York, D.F., Vigne, R., Verwoerd, D.W., and Querat, G.** 1992. Nucleotide sequence of the jaagsiekte retrovirus, and exogenous and endogenous type D and B retrovirus of sheep and goats. Journal of Virology. 66: 4930-4939.
56. **Palmarini, M., Sharp, J.M., de las Heras, M., and Fan H.** 1999. Jaagsiekte sheep retrovirus is necessary and sufficient to induce a contagious lung cancer in sheep. Journal of Virology. 73:6964–6972.
57. **Murcia, P.R., Arnaud, F., and Palmarini, M.** 2007. The Transdominant Endogenous Retrovirus enJS56A1 Associates with and Blocks Intracellular Trafficking of Jaagsiekte Sheep Retrovirus Gag. Journal of Virology. 81(4): 1762-1772.
58. **Bittner, J.J**. 1936. Some possible effects of nursing on the mammary gland tumor incidence in mice. Science. 84: 162.
59. **Bentvelzen, P., and Daams, J.H**. 1969. Hereditary Infections with Mammary Tumor Viruses in Mice. Journal of the National Cancer Institute. 43(5):1025-1035.

60. **Bentvelzen, P., Daams, J.H., Hageman, P., and Calafat, J**. 1970. Genetic Transmission of Viruses That Incite Mammary Tumor in Mice. PNAS. 67(1): 377-384.
61. **Stewart, M.A., Warnock, M., Wheeler, A., Wilkie, N., Mullins, J.I., Onions, D.E., and Neil, J.C.** 1986. Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and long terminal repeat and evidence for the recombinational origin of subgroup B viruses. Journal of Virology. 58:825–834.
62. **Jarrett, O., Hardy, W.D. Jr., Golder, M.C., and Hay, D.** 1978. The frequency of occurrence of feline leukaemia virus subgroups in cats. International Journal of Cancer. 21(3): 334-337.
63. **Callahan, R., Drohan, W., Tronick, S., and Schlom, J.** 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. 79(18): 5503-5507.
64. **Westley, B., and May, F.E**. 1984. The human genome contains multiple sequences of varying homology to mouse mammary tumor virus DNA. Gene. 28(2):221-227.
65. **Ono, M**. 1986. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. Journal of Virology. 58(3): 937-944.
66. **Handel, A.E., Ebers, G.C., and Ramagopalan, S.V.** 2010. Epigenetics: molecular mechanisms and implications for disease. Trends in Molecular Medicine. 16(1): 7-16.
67. **Vaissiere, T., Sawan, C., and Herceg, Z.** 2008. Epigenetic interplay between histone modifications and DNA methylation in gene silencing. Mutation Research (659(2008): 40-48.
68. **Tollervey, J.R., and Lunyak V.** 2012. Epigenetics: judge, jury, and executioner of stem cell fate. Epigenetics. 7(8): 823-840.
69. **Handy, D.E., Castro, R., and Loscalzo, J.** 2011. Epigenetic Modifications: Basic Mechanisms and Role in Cardiovascular Disease. Circulation. 123(19): 2145-2156.
70. **Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.D., Kivioja, T., Dave, K., Zhong, F., Nitta, K.R., Taipale, M., Popov, A., Ginno, P.A., Domcke, S., Yan, J., Schubeler, D., Vinson, C., and Taipale, J.** 2017. Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. 356 (6337): eaaj2239.
71. **Kanwal, R., Gupta, K., and Gupta, S**. 2015. Cancer Epigenetics: An Introduction. In: Verma M. (eds) Cancer Epigenetics. Methods in Molecular Biology (Methods and Protocols), vol 1238. Humana Press, New York, NY.
72. **Hurst, T.P., and Magiorkinis, G.** 2017. Epigenetic Control of Human Endogenous Retrovirus Expression: Focus on Regulation of Long-Terminal Repeats (LTRs). Viruses. 9(6): 130.
73. **Waddington C.H.** 1942. The epigenotype. Endeavour 1: 18–20.
74. **Schubeler, D., Lorincz, M.C., Cimbora, D.M., Telling, A., Feng, Y., Bouhassira, E.E., and Groudine, M.** 2000. Genomic Targeting of Methylated DNA: Influence of Methylation on Transcription, Replication, Chromatin Structure, and Histone Acetylation. Molecular and Cellular Biology. 20(24): 9103-9112.
75. **Wolff, E.M., Byun, H.M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A., and Liang, G.** 2010. Hypomethylation of a *LINE-1* promoter activates an alternate transcript of the *MET* oncogene in bladders with cancer. PLoS Genet. 2010;6:e1000917.

76. **Swets, M., Zaalberg, A., Boot, A., van Wezel, T., Frouws, M., Bastiaannet, E., Gelderblom, H., van de Velde, C., and Kuppen, P.** 2016. Tumor LINE-1 methylation level in association with survival of patients with stage II colon cancer. Int. J. Mol. Sci. 2016;18:36.
77. **Harada, K., Baba, Y., Ishimoto, T., Chikamoto, A., Kosumi, K., Hayashi, H., Nitta, H., Hashimoto, D., Beppu, T., and Baba, H.** 2014. *LINE-1* methylation level and patient prognosis in a database of 208 hepatocellular carcinomas. Ann. Surg. Oncol.1:1280–1287.
78. **Anwar, S.L., Krech, T., Hasemeier, B., Schipper, E., Schweitzer, N., Vogel, A., Kreipe, H., and Lehmann, U.** 2015. Loss of DNA methylation at imprinted loci is a frequent event in hepatocellular carcinoma and identifies patients with shortened survival. Clin. Epigenet. 7:110.
79. **Anwar, S.L., Wulaningsih, W., and Lehmann, U.** 2017. Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. International Journal of Molecular Sciences. 18(5): 974.
80. **Li, E., Bestor ,T.H., and Jaenisch, R.** 1992.Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*. 69(6): 915–926.
81. **Li, E., Beard, C., and Jaenisch, R.** 1993. Role for DNA methylation in genomic imprinting. *Nature*. 366(6453): 362–365.
82. **Mohandas, T., Sparkes, R.S., Shapiro, L.J.** 1981. Reactivation of an inactive human X chromosome: Evidence for X inactivation by DNA methylation. Science 211:393–396.
83. **Varambally, S., Cao, Q., Mani, R.S., Shankar, S., Wang, X., Ateeq, B., Laxman, B., Cao, X., Jing, X., Ramnarayanan, K., Brenner, J.C., Yu, J., Kim, J.H., Han, B., Tan, P., Kumar-Sinha, C., Lonigro, R.J., Palanisamy, N., Maher, C.A., and Cinnaiyan, A.M.** 2008. Genomic Loss of microRNA-101 Leads to Overexpression of Histone Methyltransferase EZH2 in Cancer. Science. 322 (5908): 1695-1699.
84. **Short, A.K., Fennell, K.A., Perreau, V.M., Fox, A., O'Bryan, M.K., Kim, J.H., Bredy, T.W., Pang, T.Y., and Hannan, A.J.** 2016. Elevated paternal glucocorticoid exposure alters the small noncoding RNA profile in sperm and modifies anxiety and depressive phenotypes in the offspring. Translational Psychiatry. 6(6): e837.
85. **McGowan, P.O., Sasaki, A., D'Alessio, A.C., Dymov, S., Labonte, B., Szyf, M., Turecki, G., and Meaney, M.J.** 2009. Epigenetic regulation of the glucocorticoid receptor in human brain associates with childhood abuse. Nature Neuroscience. 12: 342-348.
86. **Klein, K., and Gay, S.** 2015. Epigenetics in rheumatoid arthritis. Current Poinion in Rheumatology. 27(1): 76-82.
87. **Adcock, I.M., Ito, K., and Barnes, P.J.** 2005. Histone Deacetylation: An Important Mechanism in Inflammatory Lung Diseases. COPD. 2(2005): 445-455.
88. **Hitchins, M.P., Wong, J.J.L., Suthers, G., Suter, C.M., Martin, D.I.K., Hawkins, N.J., and Ward, R.L.** 2007. Inheritance of a Cancer-Associated *MLH1* Germ-Line Epimutation. The New England Journal of Medicine. 356:697-705.
89. **Martin-Subero, J.I., Ammerphol, O., Bibikova, M., Wickham-Garcia, E., Agirre, X., Alvarez, S., Bruggemann, M., Bug, S., Calasanz, M.J., Deckert, M., Dreyling, M., Du, M.Q., Durig, J., Dyer, M.J.S., Fan, D.B., Gesk, S., Hansmann, M.L., Harder, L., Hartmann, S., Klapper, W., Kuppers, F., Montesinos-Rongen, M., Nagel, I., Pott, C., Richter, J., Roman-Gomex, J., Seifert, M., Stein, H., Suela, J., Trumper, L., Vater, I., Prosper, F., Haferlach,**

C., Cigudosa, J.C., and Siebert, R. 2009. A Comprehensive Microarray-Based, DNA Methylation Study of 367 Hematological Neoplasms. PLoS One. 4(9): : e6986.

90. **Gama-Sosa, M.A., Slagel, V.A., Trewyn, R.W., Oxenhandler, R., Kuo, K.C., Gehrke, C.W., and Ehrlich, M.** 1983. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Research. 11(19): 6883-6894.
91. **De Smet, C., Loriot, A., and Boon, T.** 2004. Promoter-dependent mechanism leading to selective hypomethylation within the 5' region of gene MAGE-A1 in tumor cells. Molecular and Cellular Biology. 24(11): 4781-4790.
92. **Bedford, M.T., and van Helden, P.D.** 1987. Hypomethylation of DNA in pathological conditions of the human prostate. Cancer Research. 47(20): 5274-5276.
93. **Wahlfors, J., Hiltunen, H., Heinonen, K., Hamalainen, E., Alhonen, L., and Janne, J.** 1992. Genomic hypomethylation in human chronic lymphocytic leukemia. Blood. 80(8): 2074-2080.
94. **Lin, C.H., Hsieh, S.Y., Sheen, I.S., Lee, W.C., Chen, T.C., Shyu, W.C., and Liaw, Y.F.** 2001. Genome-wide hypomethylation in hepatocellular carcinogenesis. Cancer Research. 61(10): 4238-4243.
95. **Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C., Schuller, D.E., Yu, L., Bloomfield, C.D., Caligiuri, M.A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N.J., Zhang, X., O'Dorisio, M.S., Held, W.A., Cavenee, W.K., and Plass, C**. 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nature Genetics. 24(2): 132-138.
96. **Esteller, M., Corn, P.G., Baylin, S.B., and Herman, J.G.** 2001. A gene hypermethylation profile of human cancer. 61(8): 3225-3229.
97. **Weichert, W., Roske, A., Gekeler, V., Beckers, T., Stephan, C., Jung, K., Fritzsche, F.R., Niesporek, S., Denkert, C., Dietel, M., and Kristiansen, G.** 2008. Histone deacetylases 1, 2, and 3 are highly expressed in prostate cancer and HDAC2 expression in associated with shorter PSA relapse time after radical prostatectomy. British Journal of Cancer. 98(3): 604-610.
98. **Weichert, W., Roske, A., Niesporek, S., Noske, A., Buckendahl, A.D., Dietel, M., Gekeler, V., Boehm, M., Beckers, T., and Denkert, C.** 2008. Class I histone deacetylase expression has independent prognostic impact in human colorectal cancer: specific role of class I histone deacetylases in vitro and in vivo. Clinical Cancer Research. 14(6): 1669-1677.
99. **Minamiya, Y., Ono, T., Saito, H., Takahashi, N., Ito, M., Mitsui, M., Motoyama, S., and Ogawa, J.** 2011. Expression of histone deacetylase 1 correlates with a poor prognosis in patients with adenocarcinoma of the lung. Lung Cancer. 74(2): 300-304.
100. **Osada, H., Tatematsu, Y., Saito, H., Yatabe, Y., Mitsudomi, T., and Takahashi, T.** 2004. Reduced expression of class II histone deacetylase genes is associated with poor prognosis in lung cancer patients. International Journal of Cancer. 112(1): 26-32.
101. **Rikimaru, T., Taketomi, A., Yamashita, Y., Shirabe, K., Hamatsu, T., Shimada, M., and Maehara, Y.** 2007. Clinical significance of histone deacetylase 1 expression in patients with hepatocellular carcinoma. Oncology. 72(1-2): 69-74.
102. **Weichert, W., Roske, A., Gekeler, V., Beckers, T., Ebert, M.P., Pross, M., Dietel, M., Denkert, C., and Rocken, C.** 2008. Association of patterns of

class I histone deacetylase expression with patient prognosis in gastric cancer: a retrospective analysis. 2008. Lancet Oncology. 9(2): 139-148.

103. **Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., Teague, J., Andrews, J., Barthorpe, S., Beare, D., Buck, G., Campbell, P.J., Forbes, S., Jia, M., Jones, D., Knott, H., Kok, C.Y., Lau, K.W., Leroy, C., Lin, M.L., McBride, D.J., Maddison, M., Maguire, S., McLay, K., Menzies, A., Mironenko, T., Mulderrig, L., Mudie, L., O'Meara, S., Pleasance, E., Rajasingham, A., Shepherd, R., Smith, R., Stebbings, L., Stephens, P., Tang, G., Tarpey, P.S., Turrell, K., Dykema, K.J., Khoo, S.K., Petillo, D., Wondergem, B., Anema, J., Kahnoski, R.J., Teh, B.T., Stratton, M.R., and Futreal, P.A.** (2010) Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature 463:360–363.

104. **Bachmann, I.M., Halvorsen, O.J., Collett, K., Stefansson, I.M., Straume, O., Haukass, S.A., Salvesen, H.B., Otte, A.P., and Akslen, L.A.** 2006. EZH2 expression is associated with high proliferation rate and aggressive tumor subgroups in cutaneous melanoma and cancers of the endometrium, prostate, and breast. Journal of Clinical Oncology. 24(2): 268-273.

105. **Li, H., Cai, Q., Godwin, A.K., and Zhang, R.** 2010. Enhancer of zeste homolog 2 (EZH2) promotes the proliferation and invasion of epithelial ovarian cancer cells. Molecular Cancer Research. 8(12): 1610-1618.

106. **St. Laurent, G., Shtokalo, D., Dong, B., Tackett, M.R., Fan, X., Lazorthes, S., Nicolas, E., Sang, N., Triche, T.J., McCaffrey, T.A., Xiao, W., and Kapranov, P.** 2013. VlincRNAs controlled by retroviral elements are a hallmark of pluripotency and cancer. Genome Biology. 14(7): R73.

107. **Szpakowski, S., Sun, X., Lage, J.M., Dyer, A., Rubinstein, J., Kowalski, D., Sasaki, C., Costa, J., and Lizardi, PM.** 2009. Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. Gene. 448(2): 151-167.

108. **Lavie, L., Kitova, M., Maldener, E., Meese, E., and Mayer, J.** 2005. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). Journal of Virology. 79:876–883.

109. **Hurst, T., Pace, M., Katzourakis, A., Phillips, R., Klenerman, P., Frater, J., and Magiorkinis G.** 2016. Human endogenous retrovirus (HERV) expression is not induced by treatment with the histone deacetylase (HDAC) inhibitors in cellular models of HIV-1 latency. Retrovirology.13:10.

110. **Menendez, L., Benigno, B.B., and McDonald, J.F.** 2004. L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. Molecular Cancer. 3:12.

111. **Stengel, S., Fiebig, U., Kurth, R., and Denner, J.** 2010. Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation. Genes Chromosomes and Cancer.49:401–411.

112. **Hu, L., Uzhameckis, D., Hedbor,g F., and Blomberg, J.** 2016. Dynamic and selective HERV RNA expression in neuroblastoma cells subjected to variation in oxygen tension and demethylation. APMIS Journal of Pathology, Microbiology and Immunology. 124:140–149.

113. **Criscione, S.W., Zhang, Y., Thompson, W., Sedivy, J.M., and Neretti, N.** 2014. Transcriptional landscape of repetitive elements in normal and cancer human cells. BMC Genomics. 15:583.

114. **Beyer, U., Krönung, S.K., Leha, A., Walter, L., and Dobbelstein, M.** 2016. Comprehensive identification of genes driven by ERV9-LTRs reveals TNFRSF10B as a re-activatable mediator of testicular cancer cell death. Cell Death and Differentiation. 23:64–75.
115. **Beyer, U., Moll-Rocek, J., Moll, U.M., and Dobbelstein, M.** 2011. Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes. PNAS.108:3624–3629.
116. **Rosenberg, N., and Jolicoeur, P.** 1997. Retroviral Pathogenesis. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY).
117. **American Cancer Society.** Cancer Facts and Figures 2017. http://www.cancer.org/acs/groups/content/@editorial/documents/document/acspc-048738.pdf. Accessed January 2, 2018
118. .**American Cancer Society.** Skin Cancer. https://www.cancer.org/cancer/skin-cancer.html Accessed January 2, 2018.
119. **American Cancer Society.** Key Statistics for Melanoma Skin Cancer.. https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html Accessed January 2, 2018.
120. **American Cancer Society.** Key Statistics for Basal and Squamous Cell Skin Cancers. https://www.cancer.org/cancer/basal-and-squamous-cell-skin-cancer/about/key-statistics.html Accessed January 2, 2018.
121. **American Cancer Society.** What is Melanoma Skin Cancer? https://www.cancer.org/cancer/melanoma-skin-cancer/about/what-is-melanoma.html Accessed January 2, 2018.
122. **McCourt, C., Dolan, O., and Gormley, G.** 2014. Malignant Melanoma: A Pictorial Review. The Ulster Medical Journal. 83(2): 103-110.
123. **Shain, A.H., and Bastian, B.C.** 2016. From melanocytes to melanomas. Nature Reviews Cancer. 16: 345-358.
124. **Bishop, J.N., Bataille, V., Gavin, A., Lens, M., Marsden, J., Mathews, T., and Wheelhouse, C**. 2007. The prevention, diagnosis, referral and management of melanoma of the skin: concise guidelines. Clinical Medicine 7:283-290.
125. **Kvaskoff, M., Pandeya, N., Green, A.C., Perry, S., Baxter, C., Davis, M.B., Mortimore, R., Westacott, L., Wood, D., Triscott, J., Williamson, R., and Whiteman, D.C.** 2013. Site-specific determinants of cutaneous melanoma: a case-case comparison of patients with tumors arising on the head or trunk. Cancer Epidemiology, Biomarkers and Prevention. 22(12): 2222-2231.
126. **Silva, Idos, S., Higgins, C.D., Abramsky, T., Swanwick M.A., Frazer, J., Whitaker, L.M., Blanshard, M.E., Bradshaw, J., Apps, J.M., Bishop, D.T., Newton-Bishop, J.A., and Swerdlow, A.J.** 2009. Overseas sun exposure, nevus counts, and premature skin aging in young English women: a population-based survey. Journal of Investigative Dermatology. 129(1): 50-59.
127. **Sundararajan, S., and Badri, T.** Cancer, Melanoma, Metastatic. [Updated 2017 Dec 18]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2017 Jun-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470358/
128. **Heistein JB, and Acharya U.** Cancer, Melanoma, Malignant. [Updated 2017 Nov 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2017 Jun-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK470409/

129. **Fuchs, N.V., Loewer, S., Daley, G.Q., Izsvak, Z., Lower, J., and Lower, R.** 2013. Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. Retrovirology. 10:115.

130. **Goering, W., Schmitt, K., Dostert, M., Schaal, H., Deenen, R., Mayer, J., and Schulz, W.A.** 2015. Human Endogenous Retrovirus HERV-K (HML-2) Activity in Prostate Cancer Is Dominated by a Few Loci. The Prostate 75:1958-1971.

131. **Flockerzi, A., Ruggieri, A., Frank O., Sauter, M., Maldener, E., Kopper, B., Wullich, B., Seifarth, W., Muller-Lantzsch, N., Leib-Mosch, C., Meese, E., and Mayer, J.** 2008. Expression patterns of transcribed human endogenous retrovirus HERV-K(HML-2) loci in human tissue and the need for a HERV Transcriptome Project. BMC Genomics. 9:354.

132. **Ma, W., Hong, Z., Liu, H., Chen, X., Ding, L., Liu, Z., Zhou, F., and Yuan, Y.** 2016. Human Endogenous Retroviruses-K (HML-2) Expression Is Correlated with Prognosis and Progress of Hepatocellular Carcinoma. BioMed Research International. Article ID 8201642.

133. **Wang-Johanning, F., Li, M., Esteva, F.J., Hess, K.R., Yin, B., Rycaj, K., Plummer, J.B., Garza, J.G., Ambs, S., and Johanning, G.L.** 2014. Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer. International Journal of Cancer. 134, 587–595.

134. **Kleiman, A., Senyuta, N., Tryakin, A., Sauter, M., Karseladze, A., Tjulandin, S., Gurtsevitch, V., and Mueller-Lantzsch, N.** 2004. HERV-K(HML-2) Gag/Env Antibodies and Indicator for Therapy Effect in Patients with Germ Cell Tumors. Int. J. Cancer. 110, 459-461.

135. **Ruprecht, K., Ferreira, H., Flockerzi, A., Wahl, S., Sauter, M., Mayer, J., and Mueller-Lantzsch, N.** 2008. Human Endogenous Retrovirus Family HERV-K(HML-2) RNA Transcripts Are Selectively Packaged into Retroviral Particles Produced by the Human Germ Cell Tumor Line Tera-1 and Originate Mainly from a Provirus on Chromosome 22q11.21. Journal of Virology. 82(20); 10008-10016.

136. **Contreras-Galindo, R., Kaplan, M.H., Leissner, P., Verjat, T., Ferlenghi, I., Bagnoli, F., Giusti, F., Dosik, M. H., Hayes, D. F., Gitlin, S.D., Markovitz, D.M.** 2008. Human Endogenous Retrovirus K (HML-2) Elements in the Plasma of People with Lymphoma and Breast Cancer. Journal of Virology. 82(19): 9329-9336.

137. **Bhardwaj N, Montesion M, Roy F, and Coffin JM.** 2015. Differential expression of HERV-K (HML-2) proviruses in cells and virions of the teratocarcinoma cell line Tera-1. Viruses. 7(3):939-68.

138. **Thomas Muster, Andrea Waltenberger, Andreas Grassauer, Sonja Hirschl, Peri Caucig, Ingrid Romirer, Dagmar Fo¨dinger, Heide Seppele, Oliver Schanab, Christine Magin-Lachmann, Roswitha Lo¨wer, Burkhard Jansen, Hubert Pehamberger, and Klaus Wolff.** 2003. An Endogenous Retrovirus Derived from Human Melanoma Cells. Cancer Res. 63:8735-8741.

139. **Schmitt, K., Reichrath, J., Roesch, A., Meese, E., and Mayer, J.** 2013. Transcriptional Profiling of Human Endogenous Retrovirus Group HERV-K(HML-2) Loci in Melanoma. Genome Biology and Evolution. 5(2):307-328.

140. **Montesion, M., Bhardwaj, N., Williams, Z., Kuperwasser, C., and Coffin, J.M.** 2017. Mechanisms of HERV-K (HML-2) transcription during human mammary epithelial cell transformation. Journal of Virology. JVI.01258-17.

141. **Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., Reijo Pera, R.A., and Wysocka, J.** 2015. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 522: 221-225.
142. **Schiavetti, F., Thonnard, J., Colau, D., Boon, T., and Coulie, P.G.** 2002. A Human Endogenous Retroviral SEquence Encoding an Antigen Recognized on Melanoma by Cytolytic T Lymphocytes. Cancer Research. 62(19): 5510-5516.
143. **Balda, B.R., Hehlmann, R., Cho, J.R., and Spiegelmann, S.** 1975. Oncorna-like particles in human skin cancers. PNAS. 72: 3697-3700.
144. **Birkmayer, G.D., Balda, B.R., Miller, F., and Braun-Falco, O.** 1972. Virus-like particles in metastases of human malignant melanoma. Naturwissenschaften. 8:369-370.
145. **Birkmayer, G.D., Balda, B.R., and Miller, F.** 1974. Oncorna-viral information in human melanoma. European Journal of Cancer. 10: 419-424.
146. **Buscher, K., Trefzer, U., Hofmann, M., Sterry, W., Kurth, R., and Denner, J.** 2005. Expression of Human Endogenous Retrovirus K in Melanomas and Melanoma Cell Lines. Cancer Research. 65(10): 4172-7180.
147. **Hahn, S., Ugurel, S., Hanschmann, K-M., Strobel, H., Tondera, C., Schadendorf, D., Lower, J., and Lower, R.** 2008. Serological Response to Human Endogenous Retrovirus K in Melanoma Patients Correlates with Survival Probability. AIDS Research and Human Retroviruses. 24(5): 717-723.
148. **Schanab, O., Humer, J., Gleiss, A., Mikula, M., Sturlan, S., Grunt, S., Okamoto, I., Muster, T., Pehamberger, H., and Waltenberger, A.** 2011. Expression of human endogenous retrovirus K is stimulated by ultraviolet radiation in melanoma. Pigment Cell and Melanoma. 24(4): 656-665.
149. **Lemaitre, C., Tsang, J., Bireau, C., Heidmann, T., and Dewannieux, M.** 2017. A human endogenous retrovirus-derived gene that can contribute to oncogenesis by activating the ERK pathway and inducing migration and invasion. PLoS Pathogens. 13(6): e1006451.
150. **Ruprecht, K., Ferreira, H., Flockerzi, A., Wahl, S., Sauter, M., Mayer, J., and Mueller-Lantzsch, N.** 2008. Human endogenous retrovirus family HERV-K (HML-2) RNA transcripts are selectively packaged into retroviral particles produced by the human germ cell tumor line Tera-1 and originate mainly from a provirus on chromosome 22q11.21. Journal of Virology. 82 (20): 10008-10016.
151. **Denne, M., Sauter, M., Armbruester, V., Licht, J.D., Roemer, K., and Mueller-Lantzsch, N.** 2007. Physical and Functional Interactions of Human Endogenous Retrovirus Proteins Np9 and Rec with the Promyelocytic Leukemia Zinc Finger Protein. Journal of Virology. 81(11): 5607–5616.
152. **Galli, U. M., M. Sauter, B. Lecher, S. Maurer, H. Herbst, K. Roemer, and N. Mueller-Lantzsch.** 2005. 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 24**:**3223-3228.
153. **Dimitrov, D.S.** 2004. Virus entry: molecular mechanisms and biomedical applications. Nature Reviews Microbiology. **2**:109–122.
154. **Matouk, C.C., and Marsden, P.A.** 2008. Epigenetic Regulation of Vascular Endothelial Gene Expression. Circulation Research. 102:873-887.
155. **Telesnitsky, A., and Goff, S.P.** 1997. Reverse Transcriptase and the Generation of Retroviral DNA. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY).

156. **GeneCards.** LHFPL3 Gene. http://www.genecards.org/cgi-bin/carddisp.pl?gene=LHFPL3 Accessed February 11, 2018.
157. **Rabson, A.B., and Graves, B.J.** 1997. Synthesis and Processing of Viral RNA. In Coffin JM, Hughes SH, Varmus HE (ed), Retroviruses, Cold Spring Harbor (NY).
158. **Eckwahl, M.J., Telesnitsky, A., and Wolin, S.L.** 2016. Host RNA Packaging by Retroviruses: A Newly Synthesized Story. mBio. 7(1), e02025–15. http://doi.org/10.1128/mBio.02025-15
159. **Jern, P., Sperber, G.O., and Blomberg, J.** 2005. Use of Endogenous Retroviral Sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. Retrovirology. 2:50. https://doi.org/10.1186/1742-4690-2-50
160. **Lee, Y.N., and Bieniasz, P.D.** 2007. Reconstitution of an infectious human endogenous retrovirus. PLoS Pathogens. Jan;3(1):e10.
161. **Dewannieux, M., Harper, F., Richaud, A., Letzelter, C., Ribet, D., Pierron, G., and Heidmann, T.** 2006. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. Genome Research. 16(12): 1548–1556.
162. **Hindmarsh, P., and Leis, J.** 1999. Retroviral DNA Integration. Microbiology and Molecular Biology Reviews. 63(4):836-843