**Entangled Remnants: Endogenous Retroviruses**

It is often said that one cannot know the sweet without the bitter and so it is with viruses. These tiny packets of nucleic acids and proteins are seen as deadly, pathogenic enemies to the rest of the biological world most of the time. Viruses should not automatically be equated to as enemies of life, however. They have many roles in the biological world besides acting as infectious agents, including influencing the climate and controlling populations of deadly bacteria. Often viruses are considered separate from other biological organisms. It has been discovered over the past several decades, however, that viruses are actually incorporated into the genomes of humans as well as other vertebrates. The human genome is riddled with thousands of these viral fragments that influence the body in numerous ways. This viral genetic material comprises approximately an astonishing 8% of the human genome (De Parseval, N., et al., 2003). These viruses that are found within the human genome are referred to as human endogenous retroviruses (HERVs) or simply endogenous retroviruses (ERVs).

Endogenous retroviruses were first discovered in the late 1960s and early 1970s. Three different ERVs were discovered during that time period. The avian leukosis virus was found in domesticated fowl and the murine leukemia and murine mammary tumor virus were found in laboratory mice (Weiss, R., 2006).

Endogenous retroviruses belong to a larger group of diverse genetic material known as retroelements. Retroelements (also called transposable elements) are genetic elements within the genome that are considered mobile- they have the ability to move throughout the genome itself via an RNA intermediate. There are three significant types of retroelements: long terminal repeat elements (LTRs), long interspersed elements (LINEs), and short interspersed elements (SINEs). Human endogenous retroviruses (HERVs) belong specifically to the LTR group. LTRs comprise approximately 8% of the human genome and these endogenous retroviruses make up the bulk of this percentage (Khodosevich, et al., 2002). Over 98,000 HERVs have been discovered within the published human genome (Belshaw, R., et al., 2005). HERVs consist of genetic information that has been inserted into the germ line of the genome by exogenous retroviruses specifically. They are merely genetic remnants left behind in the germ line by previous retroviral infections (Bannert, N, et al., 2004). During the lifecycle of an exogenous retrovirus, the virus integrates its genetic information into the host cell's genome. The RNA of the retrovirus is reverse transcribed by the enzyme reverse transcriptase into DNA. This DNA is then integrated into the host cell's genome by the enzyme integrase. During this integration, a provirus is formed (Coffin, J., 1997). The provirus, since it is integrated into the germ line of the host cell genome, is passed on to subsequent generations and gives rise to the endogenous retroviruses. Most of the time these endogenous retroviruses have undergone many mutations and/or deletions and are missing open reading frames (ORFs) needed to code for proteins and

have thus been rendered useless and without function. There are exceptions to this generalization, however.

HERVs have been organized into three different classes: I, II, and III. These three separate classes are based on sequence similarity. Type A retroviruses are nonenveloped and are visible only inside of cells. Types B, C, and D retroviruses are enveloped and produce extracellular particles of different sizes and shapes (Nelson, P., et al, 2003). The HERVs within class I are closely related to mammalian type C viruses. There are six subgroups of class I HERVs. Group I consists of: HERV –HF, HERV-H, and HERV-F. Group II consists of: HERV-RW, HERV-W, HERV-R, and HERV-P. Group III consists of: HERV-ER1, HERV-E, HERV-R, and RRHERV-I. Group IV consists of HERV-T. Group V consists of: HERV-IP and HERV-I. Group VI consists of ERV-FRD. The class II HERVs share homology with type A, B and D mammalian viruses. Class II HERVS are divided into 10 subgroups of HERV-K. Endogenous retroviruses belonging to class II are considered the most biologically active. Class III HERVs are related to foamy viruses (a type of retrovirus belonging to the genera Spumavirus) and consist of only one major subgroup- HERV-L (Nelson, P., et al., 2003).

To understand how HERVs function, an understanding of retroviruses and how they function is needed. Retroviruses can be classified as either having simple or complex genomes. Retroviruses with simple genomes include: the alpha, beta, gamma, and epsilon. Retroviruses with complex genomes include: deltaviruses, lentiviruses, and spumaviruses. With the exception of spumaviruses, retroviruses with simple genomes are the only ones that have become endogenous retroviruses (Weiss, R., 2006). Retroviruses capable of infection contain at least the three following genes: Gag, Pol, and Env. The Gag gene codes for structural proteins of the viral core. The Pol gene codes for necessary viral enzymes such as reverse transcriptase. The Env gene codes for surface glycoproteins found on the viral envelope (Griffiths, D., 2001). At each end of the retrovirus genome are areas dubbed long terminal repeats (LTRS). LTRS are merely exact DNA sequences that repeat hundreds or thousands of times. These DNA sequence repeats are created during reverse transcription and are used by the retrovirus to insert its genetic material into the host genome. Within the LTRs lie promoters, polyadenylation signals, and enhancer elements that control the expression of retroviral proteins (Griffiths, D., 2001).

HERVs are very similar in structure to retroviruses; however, they have often lost a portion of the aforementioned structure due to point and frameshift mutations as well as deletions. Typically, the entirety of the gene Env is no longer present (although it has been found competent in some cases) and only one LTR remains (rather than two on each end of the genome) due to homologous recombination. Many HERVs are incapable of function due to premature stop codons within their open reading frames (Nelson, P., et al., 2003). Because of these mutations and deletions most HERVs are rendered inactive and are incapable of causing viral infection within the host. There are exceptions, however. In some cases, LTRs are still active. Transcription of HERVs is sometimes present in certain inflammatory diseases as well as certain cancers. HERVs have also been found to be active in fetal tissues (Griffiths, D., 2001).

The mechanisms behind the inactivation of ERVs are not completely known. Little is also known about the mechanism responsible for the HERVs' ability to copy themselves within the genome. There are hypotheses for this mechanism, however. It has been proposed that ERVs might undergo retrotransposition within germ-line cells by two separate pathways. One pathway is referred to as *cis*. When an ERV undergoes this pathway the encoded proteins used for transport are supplied by the virus itself. The other pathway is referred to as *trans*. Under this pathway the proteins the ERV needs for transport are supplied either by another endogenous virus or by an exogenous virus. In the *cis* pathway, the Env gene does not have to be intact. In the *trans* pathway, no functional genes (Env, Gag, or Pol) are needed (Belshaw, R., et al., 2004).

Although HERVs are no longer capable of functioning as viruses on their own it is possible for scientists to take these sequences and revert them back to fully competent viruses. Heidmann and his colleagues took different sequences of a specific ERV and compared them in order to determine the original sequence. Once they found the original sequence they were able to synthesize a piece of the DNA and insert it into human cells. Once the DNA was in the cells new viruses were produced. Some of these new viruses even had the ability to infect other cells (Zimmer, C., 2015).

Endogenous retroviruses have been around for millions of years. One particular study that was done at the University of Michigan by David Markovitz and his colleagues led to the discovery of a HERV known as K11- it belonged specifically to the HERV-K family. K11 was discovered when the blood of HIV patients was analyzed and the viral genes were sequenced. They ended up looking for K11 in the genomes of chimpanzees and were surprised to find it. They then looked in the genomes of other primates for K11 but this time they were unable to find it. Finding the K11 endogenous retrovirus in the chimpanzees' genomes but not in the genomes of other primates led to the conclusion that the exogenous virus that led to integration into the genome had caused infection in an ancestor right before the split between humans and chimpanzees occurred. K11 demonstrated the elusiveness of endogenous retroviruses by evading discovery for six million years. The viral sequence was able to do this because it hid out in the centers of chromosomes within the genome (Zimmer, C., 2013).

As previously discussed, HERVs have been a part of the human genome for quite some time and this fact warrants the question why they have persisted within the human genome over evolutionary time. It has been hypothesized that HERVs have survived so long within the genome because they offer benefits to the organism. These hypotheses include: immunosuppressive peptides, long terminal repeats, and plasticity (Coffin, J., 1997). It has also been discovered that HERVs might have the capability to act as mediators in regulation of gene expression and cellular differentiation in humans. Some endogenous retroviruses are thought to offer protection against exogenous retrovirus infections by inducing receptor interference or by interfering with the preintegration complex (Blomberg, J., et al., 2000).

One viral protein called Rec is thought to specifically play a role in protecting the body from exogenous viral infections. This viral protein is coded by HERVs within the HERVK family. Rec is normally responsible for binding to viral RNA transcripts in order to escort them to a ribosome where they will undergo translation. It is also capable of binding to RNA transcripts produced by a human host. Interestingly, when Rec is readily present in human cells there is also an increased amount of the human protein IFITM1. This protein is responsible for protecting the body from viral infection (Conger, K., 2015.) In other words, the viruses that are melded into the human genome actually aid in protect the body from being invaded by other viruses- coevolution at its finest!

Many endogenous retrovirus sequences are located in areas of the genome that are rich with promoters and enhancers. Promoters are sequences of DNA that are located upstream near start sites that serve as initiators of transcription. Enhancers are also regions of DNA that aid in activating transcription. One HERV in particular is located on the promoter region of the cbf2 gene. The cbf2 gene regulates the expression of genes that are involved in many cellular regulatory processes including the regulation of heat shock proteins. Another HERV is upstream of the start site of the fntb gene. This gene is responsible for protein farnesylation (facilitates protein-protein interactions and protein-membrane association). There are approximately 30 genes that are co-localized with endogenous retroviral sequences (Khodosevich, K., 2002).

Even more remarkable is the fact that endogenous retroviruses actually play a significant role during embryogenesis. During the process of embryogenesis, the cells of the embryo are totipotent. Totipotency is the ability for a cell to differentiate into any type of cell. Totipotency is controlled by numerous genes. These genes are regulated by a number of promoters. It turns out that these promoters are actually derived from viruses. When these virus-controlled genes are turned on the embryonic cells are totipotent. When the genes are switched off by the viral derived promoters the embryonic cells are no longer totipotent but pluripotent (pluripotent cells can only differentiate into certain cell types not all cell types). In other words, viruses are literally responsible for the differentiation of embryonic cells during embryogenesis. These specific viral promoters are necessary for the formation of human beings (Zimmer,C.,2012).

Another major benefit that is provided by HERVs is the presence of a protein called syncytin in placental mammals. If it were not for this protein, life as we know it would not exist. Syncytin is a HERV-W Env protein (Khodosevich, et al., 2002). The Env protein normally codes for surface glycoproteins found on the viral envelope. In placental mammals, however, the expression of syncytin causes the formation of the syncytiotrophoblast. The syncytiotrophoblast secretes progesterone and leptin during pregnancy in mammals. It is formed by the fusion of cytotrophoblast cells. The fusion of these cells is directly controlled by the viral protein syncytin. Later research by Vargas and her colleagues demonstrated that there are actually two separate syncytin proteins: syncytin-1 and syncytin-2. Both of these proteins collaborate to form the placenta. Syncytin-1 is expressed early in the fusion of trophoblast cells while syncytin-2 is gradually expressed in higher quantities as the formation of the placenta

proceeds (Vargas, et al., 2009). The presence of these viral sequences within the human genome literally allow for the creation of life.

Although HERVs cannot be transmitted between humans by fomites or bodily fluids they are still capable of playing a significant role in causing certain human diseases. As previously discussed, HERVs have the ability to transpose themselves within the genome. It has been discovered that in certain cases this transposition of genetic information can lead to certain types of autoimmune diseases as well as certain cancers.

The HERV-K family contains endogenous retroviral sequences that have the fewest mutations and deletions and are, therefore, the HERVs that are the most similar to the original exogenous retroviral infections. Of all the other HERV families they come the closest to being able to produce an infectious virion. Certain HERV-K viral sequences still have bits of the LTRs and ORFs that are necessary for coding the Env, Gag, and Pol genes. The only known HERV that contains ORFs for all three genes is K113 (Belshaw, R., et al., 2005). In addition all HERV-K viral sequences, also encode for a small protein that shares similar characteristics with the lentivirus protein Rev. The Rev protein acts as a regulator for post transcriptional transport of viral mRNAs (Coffin, J., 1997). This HERV-K family does not give rise to viral particles that are capable of infection between humans; however, they do play a role in certain cancers and autoimmune disorders. The one possible exception to this tendency is the HERV-K (HML2) family, which has been active and infectious for the past 30 million years. This particular family makes up less than 1% of HERVs (Belshaw, R., et al., 2005).

Cancer is one of the most significant detriments caused by HERVs. The transposition of genetic material within the genome can lead to various mistakes in mechanisms that control the cell cycle. One study demonstrated that the Env gene within the HERV-W family has the ability to activate the small conductance $Ca^{2+}$ activated $K^{+}$ channel protein 3 (SK3) through a CREB-dependent pathway. When the Env gene is overexpressed the levels of SK3 increase in human neuroblastoma cells (Li, S., et al., 2013). This is a prime example of how HERVs can affect the expression of certain proteins in the human body. Another HERV that plays a role in cancer in humans is HERV-K. It was discovered that in patients with chronic lymphocytic leukemia the np9 gene within HERV-K was overexpressed. On the other hand, however, the Gag gene was not overexpressed- as is often the case in other cancers (Fischer, S., et al., 2014).

Although HERVs are most often associated with cancer they can also play a role in autoimmune disorders such as multiple sclerosis, schizophrenia, lupus, and diabetes (Blomberg, J., et al., 2000). One study showed the relationship between polymorphisms in HERV-K18 and an increased likelihood of patients with schizophrenia also getting type 2 diabetes. HERV-K18 is located on chromosome 1 in the CD48 signaling lymphocyte activating gene. When there were polymorphisms within the envelope region of HERV-K18 an increase in type 2 diabetes was seen among the patients that were studied (Dickerson, F., et al., 2008).

ERVs can also cause disease in other vertebrates such as mice. One study looked specifically at the events that had to occur in order for leukemia to be induced in AKR mice. The following events were observed: expression of an endogenous ecotropic virus, recombination of this virus with the Bxvl gene to produce a recombinant virus with a U3 alteration, recombination with an endogenous polytropic virus, and then finally duplication of the Bxvl-derived enhancer sequences in U3 (Coffin, J., 1997).

Humans are both benefitted and damaged by endogenous retroviruses. On one hand, these viral sequences have the deleterious effect of causing cancer as well as other auto immune disorders. On the other hand, they have the magnificent ability of aiding in the creation of life as we know it as well as regulating gene expression throughout life. As much as we would like to think that we are separate from viruses and that they are our enemy the truth is that they are as much of a part of us as our own cells. The human genome is merely a complex tapestry with the remnants of lingering viral infections intimately entangled with our own DNA.

**References:**


Bannert, N., and R. Kurth. "Retroelements and the Human Genome: New Perspectives on an Old Relation." *Proceedings of the National Academy of Sciences* (2004): 14572-4579. Print.

Belshaw, R., A. L. A. Dawson, J. Woolven-Allen, J. Redding, A. Burt, and M. Tristem. "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* (2005): 12507-2514. Print.

Belshaw, R., V. Pereira, A. Katzourakis, G. Talbot, J. Paces, A. Burt, and M. Tristem. "Long-term Reinfection of the Human Genome by Endogenous Retroviruses." *Proceedings of the National Academy of Sciences* (2004): 4894-899. Print.

Blomberg J, Ushameckis D, Jern P. Evolutionary Aspects of Human Endogenous Retroviral Sequences (HERVs) and Disease. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-. Available from: http://www.ncbi.nlm.nih.gov/books/NBK6235/

Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997. Structural Classes of Retroelements and Replication Strategies. Available from: http://www.ncbi.nlm.nih.gov/books/NBK19412/

Conger, Krista. "Viral Proteins May Regulate Human Embryonic Development." *News Center*. 2015. Web. 25 Nov. 2015.

De Parseval, Nathalie et al. "Survey of Human Genes of Retroviral Origin: Identification and Transcriptome of the Genes with Coding Capacity for Complete Envelope Proteins." *Journal of Virology* 77.19 (2003): 10414–10422.*PMC*. Web. 6 Nov. 2015.

Dickerson, Faith, Elizabeth Rubalcaba, Raphael Viscidi, Shuojia Yang, Cassie Stallings, Anne Sullens, Andrea Origoni, Flora Leister, and Robert Yolken. "Polymorphisms in Human Endogenous Retrovirus K-18 and Risk of Type 2 Diabetes in Individuals with Schizophrenia." *Schizophrenia Research* (2008): 121-26. Print.

Fischer, Sabrina, Natalia Echeverría, Gonzalo Moratorio, Ana Inés Landoni, Guillermo Dighiero, Juan Cristina, Pablo Oppezzo, and Pilar Moreno. "Human Endogenous Retrovirus Np9 Gene Is over Expressed in Chronic Lymphocytic Leukemia Patients." *Leukemia Research Reports* (2014): 70-72. Print.

Griffith, David. "Endogenous Retroviruses in the Human Genome Sequence." *Genome Biology*. 2001. Web. 6 Nov. 2015.

Khodosevich, Konstantin, Yuri Lebedev, and Eugene Sverdlov. "Endogenous Retroviruses and Human Evolution." *Comparative and Functional Genomics*: 494-98. Print.

Li, S., Z.c. Liu, S.j. Yin, Y.t. Chen, H.l. Yu, J. Zeng, Q. Zhang, and F. Zhu. "Human Endogenous Retrovirus W Family Envelope Gene Activates the Small Conductance Ca2 -activated K Channel in Human Neuroblastoma Cells through CREB." *Neuroscience*: (2005) 164-74. Print.

Nelson, P N. "Demystified . . . Human Endogenous Retroviruses." *Molecular Pathology* (2003): 11-18. Print.

Vargas, Amandine, et al. "Syncytin-2 Plays an Important Role in the Fusion of Human Trophoblast Cells." *Syncytin-2 Plays an Important Role in the Fusion of Human Trophoblast Cells*. 18 Sept. 2009. Web. 25 Nov. 2015.

Zimmer, Carl. "The Lurker: How A Virus Hid In Our Genome For Six Million Years." *Phenomena The Lurker How A Virus Hid In Our Genome For Six Million Years Comments*. 10 May 2013. Web. 6 Nov. 2015.

Zimmer, Carl. "Our Inner Parasites." *A Planet of Viruses*. Chicago: U of Chicago, 2015. Print.

Zimmer, Carl. "We Are Viral From the Beginning - The Loom." *The Loom*. 14 June 2012. Web. 6 Nov. 2015.