The Effect of Human Endogenous Retroviral DNA on Human Embryonic Development

Is Human Endogenous Retroviral DNA, once classified as junk DNA, essential to human embryonic development, and if so to what extent does it influence embryonic development?
Abstract

In this scientific based research paper I researched the effect on retroviral DNA on human embryonic development, especially during the blastocyst stage. This is a current and rapidly expanding area of science, in which there is still a lot to be discovered. Of particular interest nowadays is the presence of active viral DNA, HERV. It has recently been found to play a large role in pluripotency as well as protection from other viruses.

Most of my research was from academic journals, in order to find reliable information as well as quantitative observations. Part of my information is from scholarly websites with articles about their research findings, while these include limited information, they helped me to get a better understanding of the intricate science behind my topic. I chose to focus on just human embryonic development and the role of viruses during this stage to narrow down my research. Through background research I early one decided to narrow it even further to the role of HERVK and HERVH during the earlier days of embryonic development.

From my research, I found that viral DNA plays a large role in early human development. My findings suggest that that retroviral DNA has entered human genomes from the beginning of evolution and is generally inactive, except in the early stages of embryonic development. My paper also looks at the DNA’s origin and its active role in embryonic development. As I concluded my research I found that HERV induces pluripotency and differentiation, while also protecting embryos from other viruses.
Introduction

Everyone’s DNA is unique, it is the blueprint of who they are. But as researchers have recently found, some of the human DNA was originally part of retroviruses. For a long time this DNA was considered “junk” DNA because it showed no apparent role, my question is; Is human endogenous retroviral DNA, once classified as junk DNA, essential to human embryonic development, and if so to what extent does it influence embryonic development? Since 1972 large parts of the genome have been classified as “junk” DNA, due to their high rates of mutation and lack of expression. (Loomis). However, modern advances in technologies such as polymerase chain reaction (PCR) and sequence homology profiling, have suggested that this DNA might actually play an essential role in mammalian development. I think that research into this new field of study will lead to a deeper understanding of retroviral DNA in human embryonic development, but also the concepts of pluripotency and stem cells, which can be used to help understand human evolution as well as help solve medical challenges.

Background

All complex life starts to develop immediately after conception. The first week is arguably the most important week for cell division and cell development. We go from fertilization to a “zygote”: the stage where the two parent genomes merge; to replication of the zygote into two cells, 4 cells and then 8 cells. The zygote is said to be totipotent, which means it can give rise to any type of cell. In the next stage of development, the collection of cells is called the morula; cells in this stage still have a spherical shape and are identical. Subsequently, a so-called “Blastocyst” is formed that consists of two parts: a clump of cells forming the outer layer called the trophoblast, and of a cavity called the blastocoel, and a cluster of cells on the interior called the epiblast (Edgar R). At this
stage, the epiblast cells are called naïve stem cells. This means that they have the ability to divide into more stem cells and are pluripotent. By the time the blastocyst is formed the epiblast is also said to be pluripotent, meaning it can give rise to all cell types except those that have formed the trophoblast (New York State Stem Cells Science). It is at this stage it is hypothesized that “Junk” DNA plays an important role in the development.

Not all viruses can alter the DNA of a host cell, only specific viruses can incorporate their DNA into a host’s. Viruses can enter our body at any moment in time. In most cases when our cell or cells get infected, our immune system wipes them out with time. But in rare cases the virus inserts itself into the host’s DNA, by cutting the host’s DNA and “melt[ing] with the genome of their hosts” (Zimmer). An important group of viruses that is known to possess this property are called retroviruses: human endogenous retroviruses (HERV) are an example of this. Usually, retroviruses behave like other viruses, infecting a population by person-to-person transmission (Stoye). What makes these retroviruses different is that they can insert their DNA into their host’s by using the proviral pol gene. The pol gene codes for reverse transcriptase. Retroviruses can use reverse transcriptase and integrase, to integrate a DNA copy of their DNA into the host cell’s DNA (Stoye). When these viruses infect a germ cell (egg or sperm cell) something unique can occur: the viral DNA can be incorporated into the DNA of the germ cell, and it is copied into all the cells that form the embryo. We, as humans, carry as much as 100,000 small pieces of endogenous retroviral DNA. As a result, retroviral DNA makes up an estimated 5% to 8% of our entire genome (Izsvák, Wang and Singh). It took millions of years for this 5 to 8% viral DNA to accumulate. The
question now remains whether these pieces of viral DNA have a function inside our cells or whether the viral DNA is indeed just “junk” DNA. In biology we usually see that function and role are connected so it seems logical to assume that HERV is still at least partly active must infer some kind of a function. However, this DNA was seen as “junk” DNA for a long time and therefore not actively researched, so in this paper I will discuss recent research on the topic of HERV and its role in embryonic development.

**HERV Expression**

In most adult or somatic cells HERV, human endogenous retrovirus, is silenced, or not actively translated, however, new research by Coger and others have shown that HERVK and HERVH, two HERV families, are expressed in early human development (Coger). As seen in figure 1b, the expression of HERVK increases significantly during the 8-Cell stage. This continues into the morula stage, three to four days after fertilization, in which the embryo assumes a spherical shape (The Endowment for Human Development). The largest amount of viral HERVK expression is found during the blastocyst stage (figure 1c), especially in the epiblast region, the cluster of cells on the interior of the embryo.
HERV expression is measured by the detection of proteins needed for production of a viral capsid and electron microscopy “which reveals the presence of virus-like particles similar to those found in reconstructed HERVK particles” (Racaniello). While this method experimentally works, there are limitations. The detection of proteins associated with the viral capsid of a protein might not mean the retrovirus is active. During evolution, it could be possible for these parts to have split or due to mutation, the retroviral DNA might no longer be active.

More recently researchers, Santoni et al, did a similar experiment by focusing on HERV-H expression and differentiation status: differentiation status is how far the cells have specialized towards a particular cell line. These researchers were able to gather more reliable data than the data recorded in the previous experiment. Due to the advancement of technology and scientific understanding, the researchers were able to analyze transcriptome changes that occur during the differentiation of human embryonic stem cells. The transcriptome is a collection of all messenger RNA produced by an organism and over time, the transcriptome changes as the cells start to take on specific roles. Detecting a change is useful as the transcriptome can show researchers whether a gene is turned on or off in specific cells or tissue, depending on if the specific mRNA is produced from the coding strand of DNA (Transcriptome fact sheet). This can elucidate function of a gene, by comparing its expression in healthy cells versus specimens lacking a portion of a gene. Another application of this method is to detect why a gene is expressed at certain times but not in others. In the experiment done by Santoni et al, 4 stages of differentiation were chosen to track the expression of genes through differentiation; undifferentiated stem cells (hESC), N1 cells (early initiation), N2
cells (neural progenitor), and N3 cells (early gliallike). The transcriptome was then analyzed using single read, paired-end read, and long read RNA sequencing.

The experiment showed new transcripts and splicings of HERV, specific to each step during differentiation. Looking at the data, diversity of the transcriptome is highest in undifferentiated cells and then decreases as the cell goes down a differentiation pathway. This can be explained by the fact that a more specified cell no longer has to be able to express its entire genome and now starts to act as part of a tissue. The data shows, as summarized in figure 2, that the levels of transcription of HERV decline very rapidly as differentiation progresses.

The relation between expression of retroviral DNA and a possible role in embryonic development starts here. As is common in biology and evolution the fact that HERV has continued to be expressed is probably not without a reason. In the case of HERV it is not only expressed but also remains intact enough to perform a role. Both types of HERVs, K and H, contain long terminal repeat sequences (LTRs) at both their 5’ and 3’ end (Haase, Mosch and Frishman). These LTRs have a strong promoter function, which means that they can increase the transcription level of themselves as well as neighboring genes. This is a clear sign of “usefulness” of HERV for the host.
HERV-H/L1HS, one of the intact LTRs found in HERV, has been found to make up as much as 99.8% of human transcription factor binding sites in the genome of Embryonic Stem Cells (ESC) (Schon, Diem and Leitner). This large number of LTRs point towards a very vital role of HERV. Transcription factor binding sites are sites for transcription factors (TF) to bind to the DNA, these TFs stimulate transcription speeding up the process of transcription (Compton). HERV seems to be essential in this stage of embryonic development. This discovery started the search for its specific role in human development.

Another discovery found a significant decrease in HERV at the blastocyst stage. Figure 3 summarizes HERVK transcriptional regulation in human embryos and in vitro cultured pluripotent cells. The dashed lines indicate inference because of the absence of data from actual post implantation human embryos. Some hypothesize this is because of the DNA Methylation. DNA methylation is an epigenetic mechanism used by cells to control gene expression. DNA methylation is the addition of a methyl group to the DNA itself, usually to the fifth carbon of a cytosine ring (Grow, Flynn and Bayless).

![Figure 3; HERV and DNA methylation expression timeline (Grow, Flynn and Bayless)](image-url)
Methylation essentially condenses DNA so that polymerase and other proteins can’t get to the genes in order to transcribe the gene. By keeping HERV from being translated it is silenced, usually for the rest of one’s life, except in specific cases of cancer.

**Role in pluripotency/differentiation**

**TEs**

HERV and other retroviruses, as well as some human parts of DNA, are called transposable elements (TEs). Transposable elements are pieces of genes that can move around in the genome, they can attach and detach themselves in different places of the genome. This was what many scientists originally classified as junk DNA, however, more recent studies have found that it might play a role in determining what genes are turned on or off (Pray). TE’s make up about 50% of the human genome and are thought to have entered through viral invasions of HERV and similar retroviruses. (Pavlicev, Hiratsuka and Swaggart). HERV is not only classified as a retrovirus but also a type of transposable element that has become fixed in the genome (Subramanian, Wildschutte and Russo).

A study by Izsvák et al. found that TE activation occurs in well-defined waves during human embryogenesis, involving a distinct set of TE and that “At [the] morula and blastocyst stages, most of the TE-derived transcripts drive from HERVH.” (Izsvák, Wang and Singh). In other words, HERV acts as the transposable element for expression of certain genes during embryonic development.
When HERVs are active they are marked with transcriptionally active histone marks such as H3K4me1/2/3, H3K9ac, H3K36me3, and H3K79me2 (Shlyueva, Stampfel and Stark). This means that the chromatin, the proteins around which DNA is wrapped, is folded in a way the genes are accessible and the protein complex is marked with a “tag” in order for easy recognition for proteins involved in transcription. These “tags” are also called epigenetic markers because, while they are not DNA, they are attached to it and can be passed on to offspring.

Originally this is where the role of HERV became apparent. Early research in this field found a “high-level association between gamma-retrovirus integration sites [of which HERV is an example] and particular epigenetic markers” (Santoni, Guerra and Luban). This means that in

Figure 4; levels of association between HERB and active marker H3K4me3 (Santoni, Guerra and Luban)
places they found HERV they also found chromatin with the tags for easy recognition. H3K4me3 was the first histone marker, tag, to be used for detection of active HERV because of the large amounts of cells carrying this tag. Experiments by Santoni et al. were done to show the association of HERV with H3K4me3. 40 different cell types were used to find any possible correlation. As figure 4 shows the highest association, shown in red, was between embryonic stem cells (H1, H9 and I3) and certain induced pluripotent stem cells (iPs-15b and iPS-11a). Medium associations, shown in green, include breast cancer cells as well as different types of fetal cells. Blue indicates a barely significant association between the histone marker and location of HERV, interestingly a few induced pluripotent stem cells fall in this group, showing that there is a significant difference in the two types of stem cells regarding epigenetic markers. The presence of these tags indicate that the close by HERV is active.

**TFs**

Detection of TEs, HERV, and the active histones associated with it are an important step in understanding how HERV may induce pluripotency. In order for pluripotency to start the TE, in our case HERV as a whole, has to contain transcription factors (TF) binding sites. Transcription factors are proteins that initiate and regulate the transcription from DNA to RNA. Binding sites are the place in the DNA where the TFs attach (Transcription factor/ transcription factors). HERV carries key pluripotent TF binding sites in its LTR. This allows for binding of pluripotency factors NANOG, OCT4, KLF4, and LBP9/Tfcp2l1 (figure 4). These pluripotency factors are what control the process of pluripotency. Not only does the LTR sequence allow for different factors to attach to the DNA, it also plays a large role in controlling the RNA that is synthesized. LTR7/HERVH (the exact site where the pluripotency factors attach) can take control
away from cellular promoters and enhancers, the parts of DNA where RNA polymerase binds to transcribe the DNA into RNA. Therefore controlling when and how much RNA is made. Secondly, LTR7/HERVH can create new transcriptional units by providing alternative splicing as seen in figure 5. Splicing is the synthesis of alternate proteins from same genetic code or polyadenylation signals, addition of only adenine bases added to mRNA, usually AAUAAA. (Premier Biosoft; Accelerating Research in Life Science). This allows for many variations of proteins to help the cell leave the naïve pluripotent state and start differentiating into all the different cell types needed.

**Oct4**

One of the before mentioned pluripotency factors that binds to the HERV/LTR is Oct4. What first signaled the involvement of Oct4 in pluripotency is the drop in expression of Oct4 after fertilizations, after which expression become restricted to the ICM (inner cell mass), and then the PGCs (primordial germ cells). As has recently been discovered, Oct4 plays a role in determining the duration of the cell cycle. This is important as in the first few cell divisions mitosis is slightly different in order for all the cells to remain totipotent (Yeom, Ha and Balling).
Some argue that the role of Oct4 might, however, be limited. In mouse embryonic cells where Oct4 synthesis was clinically blocked, the next stage of development was still reached. However, these embryos could not maintain pluripotency and started differentiating (Yeom, Ha and Balling).

The statement that the role of Oct4 is limited is refuted by others as Oct4 is said to repress factors such as Wnt/b-catenin, which promote differentiation (Davidson, Adams and Goodson). This could also be why the mice would still differentiate as Oct4’s role is to inhibit differentiation. The role of Wnt/β-catenin itself was debated for a long time, mainly because researchers weren’t sure whether it promoted self-renewal or differentiation. The study done by Davidson et al. used a sensitive reporter to “establish that Wnt/β-catenin signaling is not active during hESC [human Embryonic Stem Cell] self-renewal.” In their study, the blockage of the signaling pathway had no real effect of hESC maintenance, but when the signaling pathway was activated there was a loss in the ability to self-renew and a start in mesoderm cell-specific expression. This discovery led the researchers to hypothesize that Wnt/β-catenin must be repressed during self-renewal in order to maintain pluripotency. In another experiment, using embryonic stem cells from preimplantation-stage mammalian embryos, the researchers used previous research to determine that with active Oct4 present the signaling of Wnt/β-catenin went down. With that knowledge they turned off Oct4 in the study group, subsequently, they used a fluorescent reporter of β-catenin to observe that the hESC with higher β-catenin signaling expressed higher levels of differentiation markers (Davidson, Adams and Goodson).
NANOG

Oct 4 is not the only pluripotency marker. NANOG is very similar to Oct4 and acts as a transcriptional activator or repressor. NANOG is known to prevent differentiation “towards extraembryonic endoderm and trophoderm lineages” (NANOG Gene (Protein Coding)). It is hypothesized that overexpression of the NANOG gene can lead to S phase and proliferation (R. J. Duronio). S phase is a stage in cell division in which DNA is duplicated, while proliferation is “the process that results in an increase of the number of cells, and is defined by the balance between cell divisions and cell loss through cell death or differentiation” which is very great in the early days of embryonic development (Cell proliferation). While all cells go through these stages, embryonic stem cells might be a little different also in the way they divide and replicate. They have cleavage cycles that produce a lot of cells in a relatively small amount of time in order to grow. During this cleavage cycle there is an S phase and mitosis but no G1 and G2 cycles, cycles during which the cell produces proteins to help with physically dividing the cells. (R. Duronio)

In conclusion as a human embryo goes through the stages of development, and the stem cells move down a

Figure 6; number of total HERV reads and number of hESC TF reads (Santoni, Guerra and Luban)
specific differentiation pathway of expression, the number of HERV reads goes down dramatically as well as the number of SOX2, OCT4 and NANOG reads, this is summarized in figure 6. N0 indicates undifferentiated embryonic cells, N1 early initiation stage of differentiation and N2 neural progenitor stage (Santoni, Guerra and Luban). As can be seen from the figure as we get further into the life of the embryo expression of HERV goes down, this allows for the start differentiation. This supports the later hypothesize, discussed above, regarding Oct4’s role in withholding differentiation as well as NANOG’s role in duplication of pluripotent stem cells.

**Role in protection**
HERV might not only have a role in pluripotency, but also a role in keeping the embryo safe from other viruses. It does this by using “nuclear export of viral RNAs… [which] triggers the innate antiviral immune response and prevents retroviral (re-)infection of the host cell” (Human Preimplantation Embryos Go Retroviral). What this means is that the virus places a marker on a cell signaling that it has already been infected. Through evolution, this had evolved in order to prevent re-entry of the virus that had already infected the cell.

During the stage HERV is expressed in the blastocysts, antiviral protein IFITM1 [interferon induced transmembrane protein 1] is also synthesized

![IFITM1 RNA level (RNA-seq)](image)

*Figure 7; IFITM1 RNA levels in primed versus naïve stem cells in two studies (Grow, Flynn and Bayless)*
(Racaniello) (figure 7). IFITM1 and other IFN proteins are known anti-viral proteins, meaning they stop viruses from entering a cell’s cytoplasm through the cell wall (IFITM1 Gene (Protein Coding)). The cell catches the virus through endocytosis and lets the virus enter the cell in a vesicle rather than letting viral fusion occur. In viral fusion the virus membrane combines with the host cell and lets the virus into the cell. By using endocytosis it prevents viral particles from being released into the cytoplasm and so blocks possible infection from the virus as it can not enter the nucleus of the cell.

The expression of the IFITM1 gene is highly important for the protection of embryonic stem cells and is expressed because of HERV, it is different from pluripotency factors like Oct4 because continues to be expressed throughout human life in almost all cell types.

Another important effect of HERV is the HERVK, protein Rec. Rec is produced more specifically by HERV-K. HERV-K has remained so intact that it is technically still able to produce viral like particles. However, restriction factors including the APOBEC family restrict the production of viral particles in our cells today. This is an example of the co-evolution between retroviruses and the host cell. HERV-K rec closely resembles that of HIV rev and HTLV rex proteins (Hohn, Hanke and Norbert). Rec is only found in type II proviruses, as type I has a deletion that prevents expression. This is important because Rec “binds to a variety of cell mRNAs and either increases or decreases their association with ribosomes.” (Racaniello) and so helps or interferes with the translation of the mRNA into proteins.
Other roles of HERV

HERV is also impactful outside embryonic development, especially in medical fields. An example of this is the culturing of ESCs, it has been found that ESCs have the ability to self-renew almost indefinitely, however, induced pluripotent stem cells (IPSCs), somatic cells turned into stem cells, have been found to undergo spontaneous differentiation. This means that the IPSCs are not able to hold their pluripotent state over generations or prolonged amounts of time and start to differentiate, this is possibly due to differences in levels of Oct 4 and NANOG expression.

A second role for HERV has been found with detection of Oct 4 and other pluripotency factors in several types of cancer. Cancer is described as unusual rapid and uncontrolled cell division and the factors that drive this division are the same factors as in embryonic development. Because of this cancerous cells can be used for testing or can even be reintroduced to the blastocyst, where they can be incorporated in the fetus and contribute normally to the fetal tissue (Chambers and Smith).

Conclusion

While there is yet to be much to discovered about HERV, research is compiling to disprove the original assumption that HERV is nothing more than just “junk” DNA. Through my evaluation of the current research on viral DNA, I found its most important roles to be in differentiation and protection of the embryo. First of all HERV and its associated TE and TF’s are instrumental to the transcription of all kinds of genes and without HERV differentiation could likely not be possible or happen too fast not allowing for proper growth. The second most important role I found was HERVs role in protection
of the embryo. The RNA HERV produces marks on the outside of the cell which stops many viruses like the original HERV virus from entering the cell as the cell is already ‘infected’. I believe that in the future research in this area will rapidly expand and possibly more uses of HERV will be discovered as well as further applications in both embryonic development and cancer.
Works Cited


Coger, K. Viral Proteins may regulate human embryonic development. 20 Apr 2015. Website. 28 December 2015.


