HERVs, Transposons and Human Diseases – Part I

ALFRED GRECH
SANDRA BALDACCHINO

Abstract

It has been found that the human genome is full of relic retroviral DNA sequences called HERVs (Human Endogenous RetroViruses). A HERV is a type of a transposon, the latter being a piece of DNA sequence that can move from one position to another position in the genome, hence its other name of 'jumping gene'. HERVs and other transposons are held in check from doing havoc in the genome by several mechanisms, one of which is epigenetic in nature (namely DNA methylation and histone modifications). HERVs and other transposons are being implicated to have physiological and pathological functions in the genomes of the cells that host them. Accumulating evidence is showing that they may be associated with certain human diseases, specifically in some autoimmune diseases (e.g. rheumatoid arthritis, psoriasis, systemic lupus erythematosus, insulin-dependent diabetes mellitus), neurological diseases (e.g. schizophrenia, multiple sclerosis, motor neuron disease) and cancer. Understanding how these relic viruses and other jumping genes bring about these human diseases could help in their prevention and treatment.

Definitions

A **transposable element** (**TE**) is a DNA sequence that can move and change its position (transpose) within the same chromosome or from one chromosome to another. These mobile DNA sequences are also called **mobile elements** and have been discovered single-handedly way back in 1956 by **Barbara McClintock** in her work on maize.[1] Since then, similar mobile elements have been also shown to exist in mammalian genomes, including that of humans, and some of them are exclusive to our own species.[2] Indeed, almost half of the genome in mammals consists of transposable elements that have gained access to the genome by infecting the germline in the distant evolutionary past (millions of years ago!). Moreover, these transposable elements have been found to exist in almost all living species, including bacteria (here they are called **Insertion Sequences (IS)**).[3]

These transposable elements were regarded as 'selfish DNA parasites' or 'junk DNA'. But these 'terms' for these transposable elements are no longer suitable since their role in the genome is now being uncovered. Infact, one could say that they might have been parasitic when they were integrated into the genome but with time they entered into a kind of symbiosis with the host. Also, for some transposable elements, 'junk' is inappropriate because they are being found to have physiological and pathological significance on cell processes and functioning. Moreover, some believe that these transposable elements had and still have a role in the advancement and structure of the genome and hence to evolution itself.[4]

The focus of this paper is on such transposable elements (with specific emphasis on **human endogenous retroviruses (HERVs))** and their implications in some human diseases.

History

It was believed that genes, likened to 'beads on a string', were static and that they were passed from one generation to the next without being changed. This notion prevailed until Barbara McClintock, studying the **mosaic colouration in maize**, single-handedly found out that pieces of DNA, which she called **Dissociator** and **Activator** 'mutable loci', were capable of moving around in the genome.[1,3] She described them as **'controlling elements'** since they appeared to regulate the expression of certain genes. Her idea was not received well, and it was only when transposable elements were discovered in plants and bacteria that biologists started to acknowledge her findings. Those biologists that did not recognize McClintock notion were responsible for the era of such terms like 'selfish DNA' and 'junk DNA'. They envisaged such TEs as molecular parasites that seize and take over the cellular mechanisms for their own propagation.

But evolution biologists pointed out that the processes of evolution dispose of that which is useless or harmful for a species and the fact that many species harbour so much 'junk' DNA in their genome was surely an implication of a very valid reason. It is now believed that **genomes have co-evolved with TEs** and have devised ways to control them from running out of control while at the same time developed biological functions from their presence.

With the advent of **genomic sequencing studies**, TEs were found to be present in abundance in eukaryotic genomes. Indeed, they are a major determinant of the genome size (Table 1).

It is a known fact that the mining of the data in genome databases (**computational analysis**) has led to the discovery of new genes. But not only this, it also led to the finding of TEs and has proved useful to explore their function in the genome.

Classifications

The two main classes of transposable elements are **DNA-transposons** and **retrotransposons** (also called **retroelements**).[5] This classification is based on whether an RNA intermediate is involved during transposition.[6] Indeed, the main difference between a retrotransposon and a DNA-transposon is the way they amplify in the genome. A retrotransposon uses an RNA intermediate that is retro-transcribed into DNA using reverse transcriptase. A DNA-transposon does not use an RNA intermediate.

Table 1: Percentages of TEs sequences in eukaryotic genomes

Plants	80%
Fungi	3-20%
Metazoans	3-52%

Table 2: Main classes or types of transposable elements

Transposable Element	% in Mammalian Genome
Class I Retrotransposons (or type 2 TEs)	42.2%
Class II DNA-Transposons (or type 1 TEs)	2.8%

Table 3: Classification of retrotransposons

Non-LTR retrotransposons	LTR retrotransposons
SINE Alu repeats MIR repeats LINE L1 autonomous sequence L2 autonomous sequence	ERV (classes I, II, or III) MST MLT

Abbreviations:
LTR: Long Terminal Repeats, SINE: Short INterspersed Elements
LINE: Long INterspersed Elements, ERV: Endogenous RetroVirus

Figure 1: Typical DNA sequence of a HERV

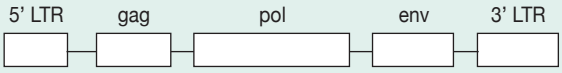

Table 4: Classification of DNA-transposons

TIR DNA transposons	The classical 'cut and paste' terminal inverted repeat (TIR) transposons
Cryptons Helitrons	Lack terminal inverted repeats
Mavericks (also known as Polintons)	The largest and most complex transposons; have terminal repeats

Table 5: Typical DNA sequence of infectious retroviruses

Sequence	Function
LTR (long terminal repeats) sequence	Contain promotors, enhancers, and regulatory sequences
gag gene	Codes for structural proteins
env gene	Codes for surface envelope proteins
pol gene	Codes for viral enzymes, including reverse transcriptase

In mammals, only less than 0.05% of retrotransposons have the ability to transpose and the LINE-1 and SINE subfamilies are mainly implicated. Again, in mammalian genomes LINE-1 accounts for about 20% of the genome.[7,8] The *Alu* elements belong to the SINE subfamily of retrotransposons. These *Alu* elements are the most abundant elements in the human genome reaching more than one million copies. LINEs are autonomous[9] i.e. they can self-propagate and transpose. SINEs like *Alus* are not autonomous and can only transpose using LINE's machinery.[10]

The DNA structure of an exogenous infectious retrovirus is shown in Figure 1 and the functions of the same sequential parts of an infectious retrovirus are explained in Table 5.

Knowing this sequence, researchers started to find similar sequences in the genomes of many species, including our own. Focusing on HERVs, their classification (Table 6) was based on the similarity of the sequence of the *pol* gene to that of the exogenous retroviruses.

Researchers use the divergence of HERVs LTR sequences from those of the exogenous counterpart retrovirus to estimate the age of HERVs in the genome. Thus LTRs act as a 'molecular clock'.[11] Class I and Cass III HERVs appear to be the oldest ones, while class II includes HERVs that have been most recently active.

So the question arises. Is the human germline still being infected? Compared to the evolutionary past or to the rate of infection in other mammals,[12] the rate of new human germline infection with evident insertions is extremely low. Indeed, presently only a small percentage of the 'youngest Alu elements and LINE-1 are still transposing in humans'.[6]

(to be continued)

References may be accessed at www.thesynapse.net

Table 6: The three classes of HERVs

HERV class	HERV example	Class of the related (exogenous) infectious retrovirus	Example of related exogenous retrovirus
Class I	HERV-W HERV-H	Gamma-retroviruses	Murine leukemia virus
Class II	Several types of HERV-K	Beta-retroviruses	Mouse mammary tumour virus
Class III	HERV-L HERV-S	Spuma-retroviruses	Primate foamy virus