Transposons: Properties, Past, and Perspectives

Jiucheng Ding*

School of Arts & Science University of Miami Coral Gables, FL, USA

*Corresponding author: jxd1164@miami.edu

Keywords: Transposable elements, jumping genes, suppressor mutator, genome, transposase, retrotransposons, DNA transposons, autonomous, transposition, TSD, Tc1/*mariner*, carcinogenesis, neurodegeneration, bioengineering tools.

Abstract: Constructing more than 50% of the entire genome of most mammals, transposable elements (TEs), also named as the "jumping genes", were discovered more than half-century ago for their position changing - property within the genome. Overthrew the "genes as still chain of codes" ideology, this finding manifested that the genome is not an immobile set of DNA; deletions, insertions, and translocations of DNA sequences are constantly taking place inside the genome. During the next several decades after transposable elements were discovered, TEs were gradually recognized and found in almost all living creatures. Since that, TEs were no longer considered purely "junk genes", but an outstanding component in the interpretation of convergence and divergence in evolution. They are also tremendously responsible for a lot of tumorigenesis due to their lability. Researchers have found that transposons can be detrimental to the genome, therefore lead to diseases. Until today, a considerable amount of research has been conducted on this topic, including their fundamental biochemical principles, pathogenesis, and future perspectives. Hence, the major focus of this review is to introduce the basic background and mechanism of TEs with a few examples of tools developed based on the best-investigated Tc1/*mariner* system, so that the readers may have a brief understanding of these sequences and how they may impact the genome of the biosphere.

1. Background

Barbara McClintock was the discoverer of transposons that were also known as the jumping genes. For most of the 20th century, genes were considered to be fixed to their order in the chromosomes. She began her research life back in the 1930s at Cornell University, pioneered cytogenetics by crossing-over maize to study their phenotypes and results of genetic recombination.

When Barbara McClintock found that some genes might be mobile in the late 1940s, she shattered preconceptions about what the fundamental nature of genes. Her research on maize chromosomal breakage led to the discovery of a chromosome-breaking site that could move around inside a chromosome. Such locus was named dissociation locus (Ds), which could be activated by a locus activator (Ac) even from a distant location within the genome. McClintock continues to study such entities in chromosomes, now named "transposons". She also discovered that mobile elements may reversibly change the expression of other genes depending on where they were introduced into a chromosome. She documented and made conclusion about her findings on the first transposable elements she came across, suggesting the dislocation and transposition of Ac and Ds could lead to unstable and unpredictable mutations [1]. TEs were first referred to as "Suppressor Mutator" (Spm) by McClintock, who described such elements switching back and forth between active and inactive states. Although maize geneticists accepted their presence very quickly, the ubiquitous nature of mobile genetic elements and the consequences of McClintock's discovery took decades to be commonly acknowledged, since her transposition theory was a fundamental challenge to the genetic interpretation of the time [3]. In 1983, after the scientific community discovered and testified the omnipresence of TEs within all living organisms, the significance of McClintock's findings was eventually recognized, and soon she was granted the Nobel Prize. Nowadays, with a wide range of selections of genetic detection and sequencing approaches, scientists have concluded that more than

50% of the human genome and 90% of the maize genome have been proven to be composed of TEs [2]. The study of TEs will not only provide the scientific community more hints in elaborating evolutionary events, but also help to understand how and what mutational diseases may be caused, or possibly learn how to harness TEs for bioengineering uses.

2. Classification

Transposable elements (TEs) are DNA sequences that move from one location to another location on the chromosome. Scientists have categorized TEs in different ways, one of the most common definition methods is by whether the TEs require reverse transcription (i.e., transcribing RNA to DNA) or not. The former type of TEs is also known as retrotransposons **(**Class 1 TEs), whereas the other one is characterized "DNA transposons", or class 2 TEs. Different classes of TEs present in the genomes of different eukaryotes in different proportions. Both Class 1 and 2 TEs may also be classified as either autonomous or nonautonomous as Figure 1 manifests.

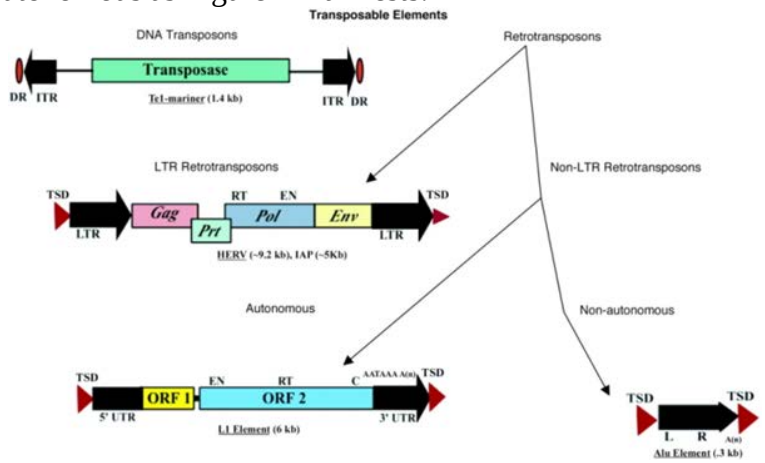

Figure 1. Classification of TEs.

Inverted terminal inverted repeats (ITRs) and a single open reading frame (ORF) that encodes a transposase are found in DNA transposons. Short direct repeats surround them on both sides (DRs). Retrotransposons are classified by autonomousness based on whether or not they have ORFs that encode retrotransposition proteins. LTRs and non-LTRs are two types of autonomous retrotransposons.

2.1 Retrotransposons

Class 1 TEs, retrotransposons do not encode transposase; instead, they move through the action of RNA intermediaries. They produce RNA transcripts and then reverse transcribe the RNA sequences back into DNA by relying upon reverse transcriptase. Interestingly, retrotransposons compose around 90% of overall human TEs (Figure 2).

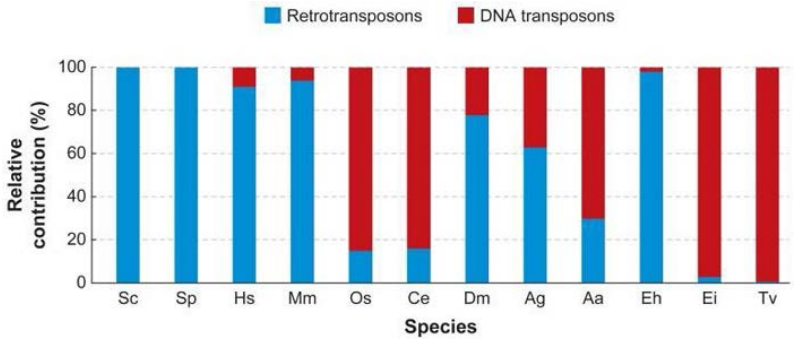

Figure 2. The relative amount of retrotransposons and DNA transposons in diverse eukaryotic genomes.

The proportional contribution of DNA transposons and retrotransposons to the overall number of transposable elements in each species is shown in this graph. Hs: Homo sapiens [5].

2.2 DNA Transposons

Unlike retrotransposons, class 2 TEs (DNA transposons) are known for their 9-40 bp terminal inverted repeats on both ends. The terminal inverted repeats are reversed complements of each other (i.e., the inverted repeat on the right side, ACGCTA, has a complement of TGCGAT at the left side of the sequence). The recognition by transposase is the function that inverted repeats play in the entire "jumping" mechanism.

2.3 Autonomousness

TEs in both classes 1 and 2 can be autonomous or nonautonomous. The translocations of non-autonomous elements require the presence of other TEs, whereas TEs categorized as autonomous can move on their own. This is due to the fact that non-autonomous elements lack the gene for the transposase or reverse transcriptase necessary for transposition, they must "borrow" these proteins from another element. [5].

2.4 Superfamilies

Based on sequence similarity and/or distinctive signatures, DNA transposases are further divided into superfamilies [6]. Several superfamilies contain members from two or three different eukaryotic kingdoms, showing that they separated early in the evolution of eukaryotes or before. According to the findings of a functional analysis of a small number of individual genetic entities, eukaryotic DNA transposons adopt a "cut-and-paste" approach of transposition. Transposase molecules attach to the terminals of their respective TEs and catalyze both of the transposition operation's DNA cleavages and strand transfer procedures of the transposition operation. Transposon integration causes a brief host sequence to be duplicated at the insertion site, also known as target-site duplication (TSD). The enzymatic characteristics of each transposase dictate the length of the TSD. As a result, elements that react to the same transposase superfamily produce TSD of the same length and have TIR sequences that are comparable.

2.4.1 Tc1/mariner.

The catalytic domains of the Tc1/mariner TEs are used to split the Tc1/mariner superfamily. It generally uses a DDE (Asp-Asp-Glu) or DDD catalytic triad. Because numerous Tc1/mariner elements are widely used as genetic tools, the Tc1/mariner superfamily is the best-known superfamily of TEs.The motif employed in theTc1/mariner system is a "TA" dinucleotide that is duplicated on both ends after insertion. It is the most discovered type of transposons found across all organisms, therefore the best-characterized one.

Tc1/mariner TEs range in size from 1 to 5 kb and encode a transposase with 282 to 345 amino acids flanked by two TIRs varying in size from 17 to 1100 bp. Although the sequences of transposase proteins from different Tc1/mariner elements differ, they all have two distinct domains: an amino-terminal region containing the helix-turn-helix (HTH) motif required for TIR recognition and binding, and a carboxy-terminal domain containing the catalytic motif DDD, or a tri-aspartic acid domain. [7].

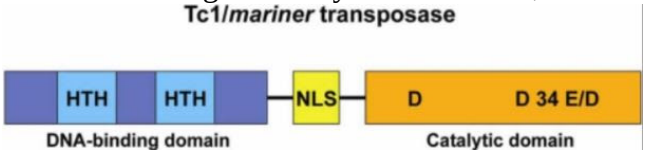

Figure 1. Structure of Tc1/mariner TE systems.

It constitutes of flankingDNA-binding domains with the *Helix-Turn-Helix* motif (HTH), a *Nuclear Localization Signal* (NLS) and a catalytic domain with the DDE or DDD motif [7].

Due to the omnipresence of Tc1/mariner system, many artificial systems have been developed with them as bioengineering tools. Sleeping Beauty, for example, is a synthetic TE rebuilt from several dormant fish Tc1-like transposon sequences and frequently utilized in genetic engineering for somatic gene delivery and genome sequencing. (e.g., gene discovery**).** Likewise, a Tc1-like element termed Frog Prince was discovered in the Rana pipiens genome and reassembled for gene-trapping in

amphibians, fish, and mammals. Hsmar1, which was reconstructed from the human genome, was the first mariner element to be employed as a genetic tool. Attributing to its capability in it transposing in vertebrate cells and has been linked to the development of non-autonomous MITE elements, Hsmar1 is a useful system for studying the transposition patterns and evolution of mariner elements in primate genomes. [8].

3. Genetic Dysregulation and Pathogenesis

Active TEs are characterized as highly mutagenic and have been linked to various stages of cancer genesis and progression. TEs have been shown to have a part in regulating the human genome by modulating endogenous gene expression and establishing new genetic loci. The insertion of TEs into genes that are responsible for DNA repair can cause the disruption of gene expression and further affect genome instability. In bladder cancer, demethylation of a particular LINE1 promoter can activate an aberrant transcript that encodes a shortened and constitutively active MET protein. Alternative transcription of a novel isoform in the ALK (anaplastic lymphoma kinase) gene is caused by de novo insertions of LTR and LINE sequences. Around 11% of melanomas express the novel ALK isoform, which has a particular sensitivity to the ALK inhibitor. Multiple LTR transcripts have been discovered in various genes that are known to elicit lymphomagenesis using a holistic approach in diffuse large B-cell lymphomas [9].

The presence of TEs is also demonstrated in disrupting transcriptions. The insertions of TEs may lead to a series of defects taking place in important DNA regions, such as promoters or silencers [10]. The insertion of TE within exons has the potential to change the open reading frame (ORF) and cause missense or nonsense mutations, which can damage transcription factor binding sites in a sequence. Insertion into gene regulatory and coding areas, on the other hand, might result in the creation of novel splice sites, the perturbation of normal splice sites, or the creation of new signal peptides. TE insertions into introns can also result in physiologically functioning alternatively spliced exons.TE insertions in 3′UTRs and introns also have an effect on mRNA stability, localization, and translation, resulting in a reduction in gene expression. Alu elements have binding sites for transcription factors involved in carcinogenesis, such as FOXA1 or P53, as well as retinoic acid receptors. Furthermore, various LTR- and non-LTR elements have been identified to interact with p53 binding sites. The transcription factor p53 is thought to be involved in more than half of all human malignancies due to its dysregulation, which influences the expression of many other genes [11].

Transposons are also reported to be associated with inflammation. The immune system defends against viral infections by coordinating innate and adaptive immune responses. While the role of innate immunity in antiviral defenses has been extensively studied, little is known about the role of transposable elements in immune responses that are not infected with viruses. The cGAS-STING pathway is triggered when viral DNA is detected in the cytoplasm, leading in the generation of interferons, which induces an inflammatory response. The interferon responses activated during the targeting of virus-infected cells may be linked to the deregulation of retroelement synthesis. Endogenous retrotransposons are increasingly being investigated as possible causes of inflammatory and neurological disorders. As significant inflammatory effectors, retrotransposon intermediates have been linked to illnesses such as MS and Aicardi-Goutières syndrome (AGS) [12].

The onset of aging is another possible consequence of transposons' unique behaviour [13]. In a recent study, the researchers discovered increased amounts of LINE-1 (Long-Interspersed Element - 1, capable of autonomous retrotransposition) mRNA in senescent cells. The accumulation of cytoplasmic LINE-1 cDNA drives the development of the senescence-associated secretory phenotype (SASP). A type-I interferon (IFN) response is prevalent with age-related inflammation in many organs. LINE-1 mRNA expression was significantly higher in the liver and adipose tissue of 26-month-old mice as compared to animals at 5 months. This finding adds another notorious record to transposons, in terms of contributing negative factors to health.

A variety of neurodegenerative and immunological disorders have been associated to abnormal TE activation. Because TEs damage coding regions, disrupt transcriptional networks, and alter epigenetic

and post-transcriptional gene expression regulation, retroelement activation promotes genomic and cellular instability. Retrotransposons have eluded evolutionary influence and are involved in somatic mosaicism. In the pathogenic state, when retrotransposon repression or control is reduced, endogenous nucleic acid expression is increased, promoting a host response similar to that seen in response to a viral infection or environmental stimuli. In most cases, the cell will initiate an interferon response. We hypothesize that retroelement misregulation has a bigger influence on human etiology than previously thought because persistent inflammation leads to functional abnormalities and disease phenotypes. As a result, developing a method to manage TEs' undesirable behavior will be beneficial **[**12].

4. Future and Potential

Our genome is filled with TEs, and the complement of mobile element-derived structural variations as well as retrotransposition-capable elements each individual inherits is unique to each other. These are not without risk. De novo and inherited TE insertions, for example, can disrupt key gene sequences and cause monogenic disorders. Although this mutational mechanism has been thought to be rare, as clinical exome-sequencing analyses adopt procedures to find TE insertions and as additional long-read sequencing technologies are used clinically, we will be able to better estimate its prevalence in the future years.

TEs appear to play a function in diseased tissues, in addition to their impact on genes. We now know that some sequences are transcriptionally activated in malignancies, and that cancer genomes accumulate additional somatically acquired L1 insertions. There's still a lot to understand about how this dysregulation affects cancer cell biology and whether similar alterations in TE expression are responsible for other diseases. Although there are examples of novel insertions causing cancer, most acquired L1 insertions in malignancies are broadly distributed in the non-coding genome, and additional research is needed to determine whether they have any functional impact.

Due to Transposons (TEs)' capability of self-excision and reinsertion within the genome at DNA sequence sites, TE systems have been selected as powerful tools for editing gene expression on a genome. TE libraries have been extensively used to discover genes that confer fitness under negative selection as in signature tagged mutagenesis, transposon site hybridization, and drug-dependent and harness the potential of transposon mutagenesis for antibacterial target identification and assessment. Lack of transposon insertions or selective depletion from a particular population identify genes critical for survival or growth, and these methods have contributed significantly to our understanding of microbial physiology and pathogenicity [14]. Traditional techniques like plasmid-based gene overexpression and controlled promoter-antisense underexpression libraries, for example, take a long time and a lot of work to determine antibacterial mechanism of action since each construct has to be pre-calibrated for the appropriate growing circumstances. The usage of TEs, on the other hand, is a considerably more hopeful option. A TE can generate a gradient of target gene expression levels, from upregulation to knockout, when coupled with an outward-facing promoter, depending on the insertion location and orientation. Because the TE is chromosomally integrated as a single copy, TE-modulated gene overexpression is appropriately bounded within a physiologically relevant region, and observed changes in resistance are thus more reliably associated with compound-target interactions. Thus, more and more studies are being conducted to examine the viability of TEs as bioengineering tools. ctive TEs are characterized as highly mutagenic and have been linked to various

5. Conclusion

The scientific community has never ceased the studies on transposons since it was discovered by McClintock. Back to the early phase of the research, when we did not have a vast selection of sequencing tools and screenings as we do today, studies were somehow limited to brief functionality and structure prediction through epigenetic approaches, such as contrasting the results of crossovers. Due to the significant composition in the human genome (more than half), the dysregulation of TEs may have serious effects on one's genotype, thus leading to notorious diseases, such as carcinomas,

innate neurodegenerative disorders, inflammations and more. Meanwhile, as more studies demonstrate, science is gradually converting labile instability into powerful tools when it comes to the introduction of DNAs and other bioengineering needs. More studies are still required to better understand the properties of TEs, infer the risks as well as developing them into capable scientific approaches. Countless benefits shall be anticipated when TEs are eventually harnessed.

References

[1] McCLINTOCK B. (1950). The origin and behavior of mutable loci in maize. Proceedings of the National Academy of Sciences of the United States of America, 36(6), 344–355.

[2] SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z., & Bennetzen, J. L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science (New York, N.Y.), 274(5288), 765–768.

[3] Ravindran S. (2012). Barbara McClintock and the discovery of jumping genes. Proceedings of the National Academy of Sciences of the United States of America, 109(50), 20198–20199.

[4] Kazazian H. H., Jr (2004). Mobile elements: drivers of genome evolution. Science (New York, N.Y.), 303(5664), 1626–1632.

[5] Pray, L. (2008) Transposons: The jumping genes. Nature Education 1(1):204

[6] Craig, N. L., R. Craigie, M. Gellert, and A. M. Lambowitz. (2002). Mobile DNA II. American Society for Microbiology Press, Washington, D.C.

[7] Muñoz-Lópe, M., & García-Pérez, J. L. (2010). DNA transposons: nature and applications in genomics. Current genomics, 11(2), 115–128.

[8] Dupeyron, M., Baril, T., Bass, C., & Hayward, A. (2020). Phylogenetic analysis of the Tc1/mariner superfamily reveals the unexplored diversity of pogo-like elements. Mobile DNA, 11, 21.

[9] Anwar, S. L., Wulaningsih, W., & Lehmann, U. (2017). Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. International journal of molecular sciences, 18(5), 974.

[10] Rebollo, R., Romanish, M. T., & Mager, D. L. (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annual review of genetics, 46, 21–42.

[11] Menendez, D., Inga, A., & Resnick, M. A. (2009). The expanding universe of p53 targets. Nature reviews. Cancer, 9(10), 724–737. [12] Ian Editor (Ed.). 2018. *The title of book two* (2nd. ed.). University of XXX Press, City, Chapter 100. DOI: http://dx.doi.org/10.1000/0-000-00000-0.

[12] Saleh, A., Macia, A., & Muotri, A. R. (2019). Transposable Elements, Inflammation, and Neurological Disease. Frontiers in neurology, 10, 894.

[13] De Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., Caligiana, A., Brocculi, G., Adney, E. M., Boeke, J. D., Le, O., Beauséjour, C., Ambati, J., Ambati, K., Simon, M., Seluanov, A., Gorbunova, V., Slagboom, P. E., Helfand, S. L., Neretti, N., … Sedivy, J. M. (2019). Author Correction: L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature, 572(7767), E5.

[14] Meredith, T. C., Wang, H., Beaulieu, P., Gründling, A., & Roemer, T. (2012). Harnessing the power of transposon mutagenesis for antibacterial target identification and evaluation. Mobile genetic elements, 2(4), 171–178.