J Genet Resour2025; 11(1): 50-59 *Homepage: http://sc.journals.umz.ac.ir/*

REVIEW ARTICLE *DOI: 10.22080/jgr.2025.28114.1411*

Transposable Elements as Regulators of Gene Expression, Evolutionary Forces, and Disease Contributors

Francis Ushie Ebuara[1], Chinyere Mary-Cynthia Ikele[2*], and Darin R. Rokyta[1]

[1] Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA
[2] Integrated Germline Biology Group Laboratory, Osaka University, Osaka, Japan

ARTICLE INFO

Article history:
Received 23 November 2024
Accepted 25 December 2024
Available 20 January 2025

Keywords:
Disease
Evolution
Gene expression
Gene regulation
Transposable elements

***Corresponding authors:**
✉ C.M. Ikele
chinyere.ikele.191144@unn.edu.ng

p-ISSN 2423-4257
e-ISSN 2588-2589

ABSTRACT

Transposable elements (TEs), once considered "junk DNA," are now recognized as significant players in gene regulation, genome evolution, and disease development. These mobile genetic sequences act as enhancers, promoters, or silencers, influencing gene expression in various species. This review explores the multifaceted roles of TEs in gene regulation, focusing on organisms such as maize, Drosophila, and mice. The present study closely examined the evolutionary impact of TEs, highlighting how they contribute to genetic diversity and innovation through chimeric transcripts and exaptation. This research also discussed the involvement of TEs by exertion of genomic instability and oncogenic activation in diseases like cancer, neurological disorders, and autoimmune conditions. Finally, it discussed experimental approaches for analyzing TE-fusion transcripts, providing insights into their evolutionary and pathogenic potentials. This comprehensive overview underscores the dual nature of TEs as drivers of genetic innovation and contributors to disease. It further highlights the need to study TE mechanisms to fully understand the complex roles of TEs in biology and disease.

Please cite this paper as: F.U. Ebuara, C.M. Ikele, & D.R. Rokyta (2025). Transposable elements as regulators of gene expression, evolutionary forces, and disease contributors. *Journal of Genetic Resources*, 11(1), 50-59. doi: 10.22080/jgr.2025.28114.1411

Introduction

Transposable elements (TEs) are repetitive DNA sequences capable of moving within the genome, often referred to as "mobile elements" or "jumping genes" (Wells and Feschotte, 2020). They constitute a significant portion of eukaryotic genomes; approximately 45% of the human genome comprises TEs (Lander *et al*., 2001). Initially dismissed as "junk DNA," TEs are now recognized as key regulators of gene expression and key drivers of genomic evolution (Chung and Feschotte, 2017).

McClintock's pioneering discovery of TEs in maize (*Zea mays*) in the 1940s and 1950s set the stage for our current understanding of their roles in gene regulation, adaptation, and disease (McClintock, 1956). McClintock proposed that these "controlling elements" could regulate neighboring genes, a hypothesis that was revolutionary at the time. Later, Britten and Davidson (1971) expanded on this idea, suggesting that repetitive DNA sequences could spread regulatory elements throughout the genome, influencing gene expression and evolution. TEs are broadly classified into two main classes based on their transposition mechanisms: Class I elements (retrotransposons) mobilize via an RNA intermediate and reverse transcription, while Class II elements (DNA transposons) move directly through a "cut-and-paste" mechanism (Wells and Feschotte, 2020). Each class is further divided into subclasses and superfamilies based on structural and functional characteristics.

Modern research has revealed that TEs can act as enhancers, promoters, and silencers, shaping complex gene regulatory networks across species (Chung and Feschotte, 2017). They influence gene expression through various mechanisms, including epigenetic modifications,

transcriptional activation or repression, alternative splicing, and the formation of chimeric transcripts (Slotkin and Martienssen, 2007). Moreover, TEs contribute to genomic innovation and diversity by creating new regulatory elements and exons, a process known as exaptation or domestication (Cosby *et al*., 2021). While exaptation refers to the co-option of TEs into novel functions beneficial to the host, domestication involves adapting a TE's original function for essential roles within the host organism. However, the mobility and mutagenic potential of TEs also pose risks to genomic integrity. Dysregulation of TEs has been implicated in various diseases, including cancer, neurological disorders, and autoimmune conditions (Burns, 2017). In cancer, TEs can activate oncogenes or disrupt tumor suppressor genes through insertional mutagenesis or by providing cryptic promoters leading to oncogenesis (Jang *et al*., 2019). An example is the activation of the LIN28B oncogene by an AluJb insertion, which creates a new promoter region and enhances tumor growth (Jang *et al*., 2019). The reactivation of LINE-1 (L1) retrotransposons has also been frequently observed in various cancers, contributing to genomic instability and mutations (Babaian and Mager, 2016a, 2016b). Advancements in high-throughput sequencing technologies and bioinformatics have facilitated the study of TEs and their impact on genomes. Techniques such as cap analysis gene expression (CAGE) and RNA annotation and mapping of promoters for the analysis of gene expression (RAMPAGE) have enabled precise mapping of TE-driven transcription start sites (Batut and Gingeras, 2013).

This review aims to provide a comprehensive overview of the multifaceted roles of TEs in gene regulation, evolution, and disease and answer whether TEs are always disruptive. This study will explore specific examples from maize, Drosophila, mice, and other organisms to illustrate how TEs shape genomes and influence biological systems. These organisms are model species, and their limited size and genome make it possible to easily understand TE manipulation through insertional mutagenesis. Additionally, the research will discuss experimental approaches for analyzing TE-fusion transcripts and consider future directions in TE research.

1. Mechanisms of TE-mediated gene regulation

TEs influence gene expression through several mechanisms, including epigenetic regulation, transcriptional control, and post-transcriptional modifications. These mechanisms allow TEs to either promote or repress gene expression, depending on their genomic context and interaction with other regulatory elements.

1.1. Epigenetic regulation

TEs play a crucial role in shaping the epigenetic landscape of genomes. Epigenetic modifications such as DNA methylation and histone modifications can silence or activate TEs, depending on the needs of the host organism. DNA methylation often works alongside histone modifications, such as H3K9 and H3K27 methylation, to maintain TE silencing. These modifications can persist even when DNA methylation is lost, indicating a complex interplay between different epigenetic marks (Guo *et al*., 2021; Liu *et al*., 2020). DNA methylation is essential for TE silencing, often in conjunction with other pathways. For instance, MBD2 acts as a methyl reader that silences TEs during male gametogenesis in Arabidopsis, functioning downstream of DNA methylation and in redundancy with other silencing mechanisms (Wang *et al*., 2924). In many species, TEs are heavily methylated to prevent their transposition and maintain genomic stability (Slotkin and Martienssen, 2007). However, when these silencing mechanisms fail, TEs can escape repression and become active (Fig. 1), influencing the expression of nearby genes (Lisch, 2013).

In maize (*Zea mays*), McClintock's discovery of "controlling elements" showed how TEs can influence plant development by modifying chromatin structure and gene expression (McClintock, 1956). TEs tend to cluster in heterochromatic regions, where epigenetic marks often silence them, but stress or environmental factors can cause their reactivation, leading to dynamic changes in gene expression (Makarevitch *et al*., 2015).

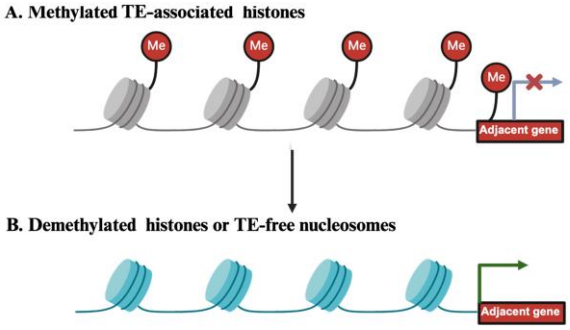

Fig. 1. Chromatin states of nucleosomes: A) Nucleosomes associated with TEs are epigenetically modified by the methylation of associated histones. This modification signals a transcriptionally repressive chromatin, thereby shutting down gene expression; B) TE-free nucleosomes and/or TE-associated demethylated histones have transcriptionally active chromatin, allowing for gene expression.

In mice, repressive marks such as histone H3K9 methylation prevent TE activation in germline cells, but during early embryonic development, many TEs are demethylated, allowing their expression (Flemr *et al*., 2013). This transient expression of TEs during development can result in epigenetic reprogramming that influences gene regulatory networks.

1.2. Transcriptional control

Some researchers reported that TEs can act as enhancers**,** promoters, or silencers, directly regulating gene transcription. These elements contribute to genome plasticity by providing transcription factor binding sites that modulate the transcriptional activity of nearby genes (Chung *et al*., 2007). For example, in maize, a TE called "Hopscotch" inserted near the *teosinte branched 1* (tb1) gene enhances its expression, which is associated with changes in plant architecture and maize domestication (Studer *et al*., 2011).

In Drosophila, the insertion of an Accord retrotransposon into the promoter region of the insecticide resistance gene *Cyp6g1* enhances its expression, leading to resistance against Dichlorodiphenyltrichloroethane (Chung and Feschotte, 2017). This demonstrates how TEs can provide regulatory elements that contribute to the adaptation of organisms in response to environmental pressures (Fig. 2).

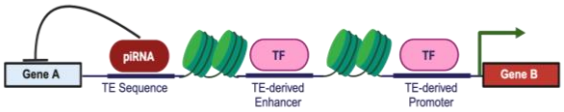

Fig. 2. Transcriptional control: TEs can be inhibited by piRNAs, the collateral effect of which can cause the silencing of an adjacent gene (Gene A). On the other hand, TEs can be co-opted as enhancers or promoters of an adjacent gene (Gene B) by providing transcription factor binding sites (TFBS).

TEs can also act as transcriptional silencers, inhibiting the expression of nearby genes. For instance, in mice, TEs can repress gene expression through mechanisms such as chromatin remodeling or the recruitment of transcriptional repressors (Bai and Brutnell, 2011). Although TEs drive impact transcriptionally, it has been reported to cause deleterious effect, genome instability, or diseases (Slotkim and Martienssen, 2007; Colonna and Fanti, 2022).

1.3. Post-transcriptional control

However, TEs influence post-transcriptional regulation by affecting RNA splicing, stability, and decay. TEs can introduce alternative splicing sites or act as polyadenylation signals, producing novel mRNA isoforms (Rebollo *et al*., 2011). This phenomenon is known as exonization, where TEs are incorporated into protein-coding regions, creating new exons that diversify the transcriptome (Ma *et al*., 2022). In mice, a TE-derived long terminal repeat (LTR) serves as an oocyte-specific promoter to produce the *DicerO* isoform, essential for RNA interference during oogenesis Flemr *et al*., 2013). Deleting this LTR results in the loss of *DicerO* expression and leads to female sterility, highlighting the critical role of TE-derived promoters in gene regulation (Fig. 3).

2. TEs as evolutionary innovators

Additionally, TEs are not only regulators of gene expression but also powerful drivers of genomic evolution by providing new transcription factor binding sites and influencing gene expression. Through processes such as exaptation and chimeric transcript formation, TEs contribute to the development of new gene functions and regulatory networks.

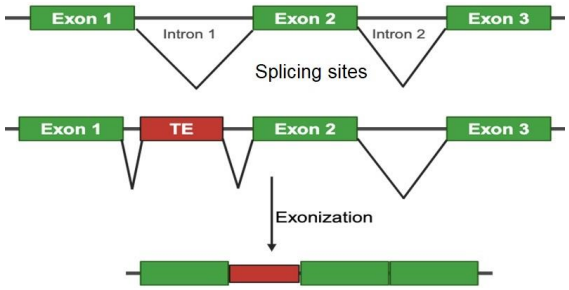

Fig. 3. Post-transcriptional control: TE insertion in intron 1 results in alternative splicing or exonization. This can in turn result in a protein isoform or nonsense-mediated decay (NMD).

TEs enhance regulatory genomes by supplying cis-regulatory sequences that function as enhancers, promoters, and silencers. These elements can swiftly facilitate changes in gene regulatory networks, resulting in species- and cell-type-specific regulatory innovations.

2.1. Chimeric transcripts and genome innovation

Chimeric transcripts arise when TEs integrate into gene regions, fusing their sequences with those of host genes (Fig. 4).

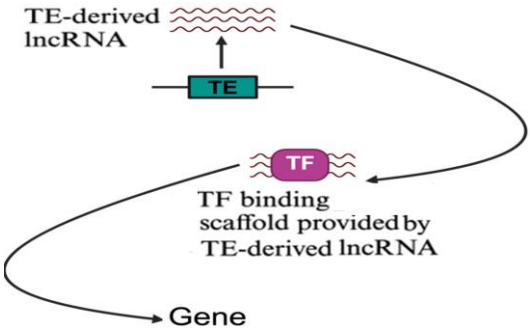

Fig. 4. Regulatory role of TE-derived lncRNAs: TE-derived long non-coding RNAs (lncRNAs) serve as a scaffold for a transcription factor that controls the expression of another gene in trans (could also occur in cis).

This fusion can result in the creation of novel regulatory elements, protein isoforms, or non-coding RNAs (Cordaux *et al*., 2006). TEs play an essential role in genomic innovation by introducing new gene regulatory features and contributing to phenotypic variation. For example, TEs affect traits like flowering time in *Capsella rubella* and phenotypic diversity in rice. In *C. rubella*, TE insertions have been shown to affect flowering time by altering gene regulatory networks (Niu *et al*., 2019). Similarly, in rice, TEs contribute to phenotypic diversity by providing epigenetic regulation through DNA methylation and histone modifications, impacting gene expression in response to environmental stimuli (Song and Cao, 2017). In Drosophila**,** it has been shown that approximately 1.6% of the developmental transcriptome originates from TE-driven transcription initiation (Batut and Gingeras, 2013). These chimeric transcripts shape the developmental gene expression profile and contribute to the evolution of new traits. Another example is in mice, where the insertion of an LTR retrotransposon within the *Cdk2p1* gene creates a chimeric transcript that produces an N-terminally truncated protein essential for preimplantation embryo development (Modzelewski *et al*., 2021). The formation of these chimeric transcripts demonstrates the role of TEs in generating new protein isoforms that contribute to species-specific traits. TEs can benefit hosts through exaptation and domestication.

2.2. Exaptation and TE domestication

Exaptation refers to the co-option of TEs for novel functions that benefit the host organism, while domestication involves adapting a TE's original function for essential roles within the host (Cosby *et al*., 2021). TEs have been domesticated across various species to serve in critical biological processes, such as DNA repair, gene regulation, and immune responses.
One of the most well-known examples of exaptation is the evolution of the SETMAR gene in primates. This gene is a fusion between a SET histone methyltransferase and the transposase domain of the *Hsmar1* transposon. The transposase component of SETMAR retains its DNA-binding ability, allowing the gene to play a role in DNA repair (Cordaux *et al*., 2006). Furthermore, in maize, TEs have contributed to plant evolution by introducing regulatory elements that control stress responses and developmental processes. For example, the insertion of TEs into the non-coding regions of the maize genome has been shown to regulate flowering time and other agronomically important traits (Salvi *et al*., 2007).

3. TEs and disease development

While TEs play essential roles in gene regulation and evolution, their reactivation and uncontrolled transposition can contribute to various diseases. TE dysregulation has been implicated in cancer, neurological disorders, autoimmune diseases, and age-related pathologies. This section explores the mechanisms through which TEs contribute to disease development.

3.1. TEs in Cancer

One of the most well-documented roles of TEs in disease is their involvement in cancer. TEs can drive oncogenesis by acting as cryptic promoters for oncogenes or by disrupting tumor suppressor genes through insertional mutagenesis (Burns, 2017). This phenomenon, known as onco-exaptation, involves repurposing TE-derived sequences to promote tumor growth. A study by Jang *et al*. (2019) identified over 100 TE-mediated cryptic promoter activation instances in various cancer types. In particular, the activation of the *LIN28B* oncogene by an AluJb insertion highlights the role of TEs in driving oncogenic expression. The insertion of this TE creates a new promoter region, leading to the overexpression of *LIN28B*, which is known to enhance cell proliferation and tumorigenesis. Moreover, retrotransposons such as LINE-1 (L1) have been implicated in cancer progression. L1 reactivation is often observed in various cancers, contributing to genomic instability by inserting into critical genes, leading to mutations (Babaian and Mager, 2016a and b). In some cases, L1 insertions disrupt DNA repair pathways. This disruption essentially leads to genomic instability that further promotes cancer development and progression.

3.2 TEs in Neurological Disorders

TEs are also implicated in neurological disorders, where their reactivation can lead to neuroinflammation, genomic instability, and neuronal death. Neuroinflammatory diseases are explained by the reactivation of L1 elements, which triggers neuroinflammation by causing DNA damage and activating innate immune responses (Saleh *et al*., 2019). Furthermore, studies have shown that stress and chronic alcohol exposure can activate retrotransposons in the brain, contributing to conditions such as post-traumatic stress disorder (PTSD) and alcoholism (Reilly *et al*., 2013). In addition, TEs have been linked to neurodegenerative diseases such as Alzheimer's and Huntington's disease. The reactivation of L1 in neuronal cells has been observed in aging brains and neurodegenerative diseases, contributing to neuronal death and cognitive decline (Saleh *et al*., 2019). In Alzheimer's disease, L1 activation has been associated with increased DNA damage and neuroinflammation, further exacerbating disease progression.

3.3. TEs in autoimmune diseases

TEs can trigger autoimmune diseases by activating the innate immune response. The immune system can recognize TE-derived sequences as foreign, producing inflammatory cytokines and autoantibodies (Crow and Rehwinkel, 2009). This immune activation can contribute to autoimmune conditions such as lupus, where L1 reactivation accumulates retroelement-derived nucleic acids that trigger inflammation. In some cases, impaired degradation of TE-derived RNA can result in chronic immune activation. For instance, patients with systemic lupus erythematosus (SLE) exhibit elevated levels of TE-derived RNA in their cells, which may contribute to the development of autoimmunity (Payer and Burns, 2019).

3.4. TEs in aging and cellular senescence

TEs are also involved in cellular senescence and aging-related disorders. As cells age, the mechanisms that repress TEs, such as DNA methylation and histone modifications, become less effective, leading to the derepression of TEs. This derepression results in genomic instability, DNA damage, and activation of inflammatory pathways (De Cecco *et al*., 2019). A study by De Cecco *et al*. (2019) revealed that the derepression of L1 in senescent cells triggers a type-I interferon response, contributing to the senescence-associated secretory phenotype (SASP)**.** The SASP is characterized by the secretion of pro-inflammatory cytokines, growth factors, and proteases, which promote chronic inflammation and tissue damage, contributing to age-related diseases (Fig. 5). Key therapeutics strategies targeted at TEs for disease

management and prevention include cancer immunotherapy, where the SB transposon system has been utilized to engineer T cells with tumor-specific T-cell receptors, offering a personalized approach to cancer treatment by targeting unique tumor neoantigens (Deniger *et al*., 2016). Transposons are also used to generate iPSCs by delivering reprogramming factors, which can be applied in regenerative medicine (Vanden Drissche *et al*., 2009).

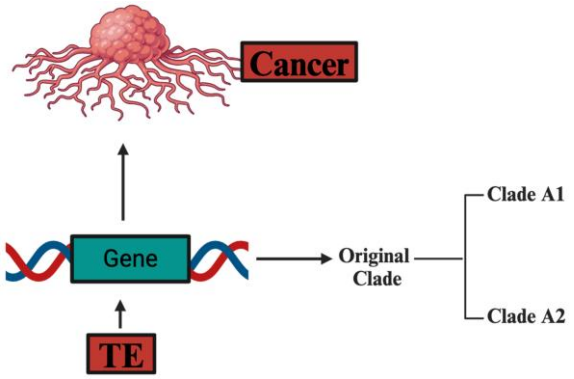

Fig. 5. Summary of the dual role of TEs as contributors to disease development and drivers of genomic evolution.

4. Experimental approaches for TE-fusion transcript analysis

Advancements in high-throughput sequencing technologies and bioinformatics have enabled researchers to more precisely study TEs and their impact on gene expression. This section highlights the key experimental techniques for analyzing TE-fusion transcripts and TE-driven gene regulation.

4.1. Expressed sequence tag sequencing

Expressed sequence tag (EST) sequencing was one of the earliest techniques used to study gene expression by identifying transcribed regions of the genome. While next-generation sequencing methods have largely replaced EST sequencing, it remains a valuable tool for identifying chimeric transcripts involving TEs.

In a study by De Cecco *et al*. (2019), EST sequencing was used to examine the prevalence of chimeric gene-TE transcripts in Drosophila melanogaster. The authors found that 33.5 percent of genes harbored chimeric transcripts derived from TE insertions, highlighting the role of TEs in shaping the transcriptome.

4.2. Cap analysis gene expression

Cap analysis gene expression (CAGE) is a technique that focuses on mapping transcription start sites (TSSs), making it ideal for studying TE-derived promoters and enhancers (Faulkner *et al*., 2009). By capturing the 5' ends of mRNA molecules, CAGE allows researchers to identify both abundant and rare transcripts, including those driven by TEs.

CAGE has been instrumental in studying the role of TEs in gene regulation. For instance, a study by Faulkner *et al*. (2009) used CAGE to demonstrate that TEs contribute significantly to transcription initiation in the human genome. The authors found that many TSSs are located within TE sequences, indicating that TEs serve as regulatory elements that shape the transcriptome.

4.3. RNA annotation and mapping of promoters for the analysis of gene expression

RNA annotation and mapping of promoters for the analysis of gene expression (RAMPAGE) builds upon CAGE by not only identifying TSSs but also quantifying gene expression levels (Batut and Gingeras, 2013). Comparatively, CAGE identifies TSSs, captures both abundant and rare transcripts, and provides a broad view of TE-driven transcription, while RAMPAGE offers a higher sensitivity and base-resolution accuracy of TSSs through an in vitro elongation step. Fundamentally, CAGE provides an initial landscape of TSSs, while RAMPAGE refines this view by accurately quantifying expression levels from specific promoters. In a study by Batut and Gingeras (2013), RAMPAGE was used to map TSSs in Drosophila embryos, revealing a direct causal relationship that TEs play a significant role in shaping the developmental transcriptome (Batut and Gingeras, 2013).

The technique's high sensitivity makes it particularly useful for studying the dynamic transcriptional landscape driven by TEs in various biological systems.

4.4. Bioinformatics pipelines for analysis of TE-associated transcripts

The analysis of TEs requires sophisticated bioinformatics pipelines to accurately identify

TE-derived sequences and their regulatory effects.

Tools such as RepeatModeler2, RepeatMasker, and TEtoolkit are commonly used to annotate TEs in genomic sequences, while programs like Homer and FIMO are used to identify transcription factor binding sites within TE sequences. Several pipelines have been developed to detect TE-driven transcriptional activity from RNA-seq data to analyze chimeric transcripts. Some of these tools are briefly discussed below. Fundamentally, these tools provide insights into the regulatory roles of TEs and their contributions to gene expression in various contexts, including development, disease, and evolution.

4.4.1. Chimeric LIne finder

Chimeric line finder (CLIFinder) [2018] has been developed to identify chimeric transcripts, specifically RNA molecules that arise from chromosomal rearrangements or TE insertions, particularly those involving long interspersed nuclear elements (LINEs). The tool is optimized for analyzing stranded paired-end RNA-seq data, providing crucial insights into the read orientation to accurately detect chimeric events. While CLIFinder can also process non-stranded data, this approach is prone to a higher rate of false positives, particularly from L1 promoter sequences. Despite its strength in detecting chimeric transcripts, CLIFinder requires manual validation by researchers to minimize false positives, as noted by Pinson *et al*. (2018). Additionally, the tool depends on a reference genome, limiting its utility for non-model organisms with incomplete genomic resources.

4.4.2. Library of information for operon analysis in NGS data sets

Library of Information for Operon Analysis in NGS data sets (LIONS) [2019], introduced by Babaian *et al*. (2019), offers a robust platform for detecting and quantifying TE-associated transcripts in RNA-seq data. It classifies TE-exon pairs into categories such as TE-initiation, TE-exonization, and TE-termination, providing a comprehensive view of TE-driven transcription. Known for its sensitivity in capturing low-abundance TE-fusion transcripts, LIONS also features a user-friendly interface streamlining the analysis process. However, it is limited by its exclusive focus on TE-initiated transcripts, potentially overlooking other TE contributions, such as internal exonization or TE-mediated terminations. Furthermore, like many TE-focused tools, LIONS requires a well-annotated reference genome, which may restrict its application to species with comprehensive genomic resources.

4.4.3. TEchim

TEchim [2020] is a specialized toolkit for exploring TE-fusion transcripts, particularly effective in analyzing transcripts linked to polymorphic TEs absent from the reference genome. Introduced by Treiber and Waddell (2020), TEchim has proven useful in studies such as investigating dynamic TE expression in the Drosophila brain. Its ability to detect chimeric transcripts independent of a complete reference genome makes it an invaluable tool for non-model organisms with incomplete genomic data. However, TEchim's limitation lies in its lack of full automation, requiring manual customization for new datasets, which can hinder its broader applicability.

4.4.4. ChimeraTE

ChimeraTE [2023], developed by Oliveira *et al*. (2023), represents a significant advancement in TE-fusion transcript analysis. This pipeline is designed to operate both with and without a reference genome, offering flexibility in detecting chimeric transcripts. ChimeraTE excels in identifying transcripts associated with polymorphic TE insertions that are not present in reference genomes. In one study, the tool was used to analyze RNA-seq data from Drosophila melanogaster ovaries, revealing that 1.12% of genes harbor chimeric transcripts, with 88.97% of them being TE-exonized. Furthermore, ChimeraTE’s unique ability to detect previously hidden polymorphic insertions underscores its capacity to provide novel insights into TE involvement in transcript structure. However, the computational demands of this tool are significant, and careful planning is necessary when utilizing it in large-scale studies.

Conclusion

Transposable elements are integral to the regulation of gene expression, genomic evolution, and disease development. While they contribute to genetic diversity and innovation, their dysregulation can lead to serious pathologies such as cancer, neurological disorders, and autoimmune diseases. Building on this understanding, TE research holds immense potential for novel therapeutic interventions targeting diseases associated with TE dysregulation. For example, TE silencing through CRISPR-based epigenome editing or RNA interference may offer new avenues for treating TE-driven cancers and neurodegenerative disorders. Moreover, the roles of TEs in genome dynamics and evolution are complex, reflecting their dual nature as both drivers of genetic innovation and agents of genomic instability.

Advancements in sequencing technologies and bioinformatics have greatly enhanced our ability to study TEs, providing insights into their roles in shaping gene regulatory networks across diverse species. The benefits of using these technologies and pipelines in the context of both disease and evolutionary biology underscore the importance of enhanced technological training and the development of open-source bioinformatics tools. We hope that future research will continue to explore the therapeutic potential of targeting TEs in disease treatment and the evolutionary implications of their activity in non-model organisms.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work reported in this paper**.**

Authors' Contributions

FUE Conceptualization and methodology; **FUE and CMI** Writing; **FUE**, **CMI,** and **DRR** Validation; **FUE** Writing-original draft preparation; **CMI** Writing-review and editing; **DRR** Supervision; **FUE**, **CMI,** and **DRR** Project administration; All authors have read and agreed to the published version of the manuscript.

References

Wells, J.N., & Feschotte, C. (2020). A field guide to eukaryotic transposable elements. *Annual Review of Genetics*, 54, 539-561. https://doi.org/10.1146/annurev-genet-040620-022145

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Dewar, K., … International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409 (6822), 860-921. https://doi.org/10.1038/35057062

Chung, E.B., Elde, N.C., & Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, 18(2), 71-86. https://doi.org/10.1038/nrg.2016.139

McClintock, B. (1956). Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology*, 21, 197-216. https://doi.org/10.1101/SQB.1956.021.01.017

Britten, R.J., & Davidson, E.H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *The Quarterly Review of Biology*, 46(2), 111-138. https://doi.org/10.1016/j.dci.2025.105358

Slotkin, R.K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, 8(4), 272-285. https://doi.org/10.1038/nrg2072

Cosby, R.L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E.J. and Feschotte, C., (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*, 371(6531), eabc6405. https://doi.org/10.1126/science.abc6405

Burns, K.H. (2017). Transposable elements in cancer. *Nature Reviews Cancer*, 17(7), 415-424. https://doi.org/10.1038/nrc.2017.35

Jang, H. S., Shah, N. M., Du, A. Y., Dailey, Z. Z., Pehrsson, E. C., Godoy, P. M., ... & Wang, T. (2019). Transposable elements drive widespread expression of oncogenes in human cancers. *Nature Genetics*, 51(4), 611-617. https://doi.org/10.1038/s41588-019-0373-3

Batut, P., & Gingeras, T.R. (2013). RAMPAGE: promoter activity profiling by paired-end sequencing of 5′ complete cDNAs. *Current Protocols in Molecular Biology*, 104(1), 25B- 11. https://doi.org/10.1002/0471142727.mb25b11s104

Lisch, D. (2013). How important are transposons for plant evolution? *Nature Reviews Genetics*, 14(1), 49-61. https://doi.org/10.1038/nrg3374

Makarevitch, I., Waters, A. J., West, P. T., Stitzer, M., Hirsch, C. N., Ross-Ibarra, J., & Springer, N. M. (2015). Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genetics*, 11(1), e1004915. https://doi.org/10.1371/journal.pgen.1005566

Flemr, M., Malik, R., Franke, V., Nejepinska, J., Sedlacek, R., Vlahovicek, K., & Svoboda, P. (2013). A retrotransposon-driven dicer isoform directs endogenous small interfering RNA production in mouse oocytes. *Cell*, 155(4), 807-816. https://doi.org/10.1016/j.cell.2013.10.001

Chung, H., Bogwitz, M. R., McCart, C., Andrianopoulos, A., Ffrench-Constant, R. H., Batterham, P., & Daborn, P. J. (2007). Cis-regulatory elements in the Accord retrotransposon result in tissue-specific expression of the *Drosophila melanogaster* insecticide resistance gene *Cyp6g1*. *Genetics*, 175(3), 1071-1077. https://doi.org/10.1534/genetics.106.066597

Studer, A., Zhao, Q., Ross-Ibarra, J., & Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics*, 43(11), 1160-1163. https://doi.org/10.1038/ng.944

Bai, L., & Brutnell, T.P. (2011). The Activator/Dissociation transposable elements comprise a two-component gene regulatory switch that controls endogenous gene expression in maize. *Genetics*, 187(3), 749-759. https://doi.org/10.1534/genetics.110.124149

Rebollo, R., Karimi, M. M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., ... & Mager, D. L. (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genetics*, *7*(9), e1002301. https://doi.org/10.1371/journal.pgen.1002301

Ma, G., Babarinde, I. A., Zhou, X., & Hutchins, A. P. (2022). Transposable elements in pluripotent stem cells and human disease. *Frontiers in Genetics*, 13, 902541. https://doi.org/10.3389/fgene.2022.902541.

Cordaux, R., Udit, S., Batzer, M.A., & Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. *Proceedings of the National Academy of Sciences*, 103(21), 8101-8106. https://doi.org/10.1073/pnas.0601161103

Modzelewski, A. J., Shao, W., Chen, J., Lee, A., Qi, X., Noon, M., ... & He, L. (2021). A mouse-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for development. *Cell*, 184(22), 5541-5558. https://doi.org/10.1016/j.cell.2021.09.021

Salvi, S., Sponza, G., Morgante, M., Tomes, D., Niu, X., Fengler, K. A., ... & Tuberosa, R. (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proceedings of the National Academy of Sciences*, 104(27), 11376-11381. https://doi.org/10.1073/pnas.0704145104

Babaian, A., & Mager, D. L. (2016a). Endogenous retroviral promoter exaptation in human cancer. *Mobile DNA*, 7 (1), 24. https://doi.org/10.1186/s13100-016-0081-9

Babaian, A., & Mager, D. L. (2016b). Endogenous retroviruses in development and disease. *Genome Biology*, 17 (1), 258. https://doi.org/10.1186/s13059-016-1124-8

Reilly, M. T., Faulkner, G. J., Dubnau, J., Ponomarev, I., & Gage, F. H. (2013). The role of transposable elements in health and diseases of the central nervous system. *Journal of Neuroscience*, *33* (45), 17577–17586. https://doi.org/10.1523/JNEUROSCI.3964-13.2013

Saleh, A., Macia, A., & Muotri, A. R. (2019). Transposable elements, inflammation, and neurological disease. *Frontiers in Nneurology*, 10, 894. https://doi.org/10.3389/fneur.2019.00894

Crow, Y. J., & Rehwinkel, J. (2009). Aicardi-Goutieres syndrome and related phenotypes: linking nucleic acid metabolism with

autoimmunity. *Human Molecular Genetics*, 18(R2), R130-R136. https://doi.org/10.1093/hmg/ddp293

Payer, L. M., & Burns, K. H. (2019). Transposable elements in human genetic disease. *Nature Reviews Genetics*, 20(12), 760-772. https://doi.org/10.1038/s41576-019-0165-8.

De Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., … Sedivy, J. M. (2019). LINE-1 derepression in senescent cells triggers interferon and inflammaging. *Nature, 568* (7752), 405-409. https://doi.org/10.1038/s41586-019-1087-5

Lipatov, M., Lenkov, K., Petrov, D. A., & Bergman, C. M. (2005). Paucity of chimeric gene-transposable element transcripts in the *Drosophila melanogaster* genome. *BMC Biology*, 3, 1-18. https://doi.org/10.1186/1741-7007-3-24.

Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., ... & Carninci, P. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nature Genetics*, 41(5), 563-571. https://doi.org/10.1038/ng.368

Pinson, M.E., Pogorelcnik, R., Court, F., Arnaud, P. and Vaurs-Barrière, C., (2018). CLIFinder: identification of LINE-1 chimeric transcripts in RNA-seq data. *Bioinformatics*, 34(4), 688-690. https://doi.org/10.1093/bioinformatics/btx671

Babaian, A., Thompson, I. R., Lever, J., Gagnier, L., Karimi, M. M., & Mager, D. L. (2019). LIONS: analysis suite for detecting and quantifying transposable element initiated transcription from RNA-seq. *Bioinformatics*, 35(19), 3839-3841. https://doi.org/10.1093/bioinformatics/btz130

Treiber, C. D., & Waddell, S. (2020). Transposon expression in the Drosophila brain is driven by neighboring genes and diversifies the neural transcriptome. *Genome Research*, 30 (11), 1559-1569. https://doi.org/10.1101/gr.265938.120

Oliveira, D. S., Fablet, M., Larue, A., Vallier, A., Carareto, C. M., Rebollo, R., & Vieira, C. (2023). ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements. *Nucleic Acids Research*, 51(18), 9764-9784. https://doi.org/10.1093/nar/gkad671

Wang, S., Wang, M., Ichino, L., Boone, B. A., Zhong, Z., Papareddy, R. K., ... & Jacobsen, S. E. (2024). MBD2 couples DNA methylation to transposable element silencing during male gametogenesis. *Nature Plants*, 10, 13-24. https://doi.org/10.1038/s41477-023-01599-3

Guo, W., Wang, D., & Lisch, D. (2021). RNA-directed DNA methylation prevents rapid and heritable reversal of transposon silencing under heat stress in *Zea mays*. *PLoS Genetics*, 17(6), e1009326. https://doi.org/10.1371/journal.pgen.1009326

Liu, S., De Jonge, J., Trejo-Arellano, M., Santos-González, J., Köhler, C., & Hennig, L. (2020). Role of H1 and DNA methylation in selective regulation of transposable elements during heat stress. *New Phytologist*, 229(4), 2238-2250. https://doi.org/10.1111/nph.17018

Colonna Romano, N., & Fanti, L. (2022). Transposable elements: major players in shaping genomic and evolutionary patterns. *Cells*, 11(6), 1048. https://doi.org/10.3390/cells11061048

Niu, X. M., Xu, Y. C., Li, Z. W., Bian, Y. T., Hou, X. H., Chen, J. F., ... & Guo, Y. L. (2019). Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proceedings of the National Academy of Sciences*, 116(14), 6908-6913. https://doi.org/10.1073/pnas.1811498116

Song, X., & Cao, X. (2017). Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Current Opinion in Plant Biology*, 36, 111-118. https://doi.org/10.1016/j.pbi.2017.02.004

VandenDriessche, T., Ivics, Z., Izsvák, Z., & Chuah, M. K. (2009). Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. *Blood*, 114(8), 1461-1468. https://doi.org/10.1182/blood-2009-04-210427

Deniger, D. C., Pasetto, A., Tran, E., Parkhurst, M. R., Cohen, C. J., Robbins, P. F., … Rosenberg, S. A. (2016). Stable, nonviral expression of mutated tumor neoantigen-specific T-cell receptors using the *Sleeping Beauty* transposon/transposase system. *The Journal of the American Society of Gene Therapy*, 24 (6), 1078-1089. https://doi.org/10.1038/mt.2016.6