

INTEGRATED TRANSCRIPTOMIC ANALYSIS REVEALS POTENTIAL TRANSPOSABLE ELEMENT ASSOCIATION IN ALZHEIMER'S DISEASE

**A Thesis Submitted
in Partial Fulfillment of the Requirements
for the Degree of**

MASTER OF TECHNOLOGY

in

BIOINFORMATICS

by

ANANYA SINGH

(24/BIO/05)

Under the Supervision of

Prof. PRAVIR KUMAR



Department of Biotechnology

**DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India**

ACKNOWLEDGEMENTS

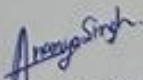
I would like to express my sincere gratitude to my project guide and mentor, Prof. Pravir Kumar, for accepting me in his lab and allowing me to carry out my final semester project under his mentorship. His constant support, encouragement, invaluable guidance and knowledgeable insights has significantly shaped the course of this research.

I would also like to extend my gratitude to the Department of Biotechnology, Delhi Technological University for allowing me to undertake this project and for providing all the essential facilities and resources required for successful completion of the project.

I would also take the opportunity to thank the Ph. D. Scholars Ms. Shrutikirti Vashishth, Ms. Apurva Lakra and Dr. Neetu Rani, for their availability and guidance during this project. Their guidance has been extremely instrumental towards this work.

Finally, I would like to thank my family and friends who have always believed in me and have been by my side offering their unwavering support.

I acknowledge the contribution of all those who have extended their timely help and support for the completion of my project.


Ananya Singh
(24/BIO/05)



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daultpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I **ANANYA SINGH** hereby certify that the work which is being presented in the thesis entitled "**Integrated Transcriptomic Analysis Reveals Potential Transposable Element Association in Alzheimer's Disease**" in partial fulfilment of the requirements for the award of the Degree of Master of Technology, submitted in the Department of Biotechnology, Delhi Technological University is an authentic record of my own work carried out during the period from January 2026 to May 2026 under the supervision of **Prof. Pravir Kumar**.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

A handwritten signature in blue ink that reads "Ananya Singh". The signature is written in a cursive, flowing style.

Candidate's Signature



DELHI TECHNOLOGICAL UNIVERSITY
 (Formerly Delhi College of Engineering)
 Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Ananya Singh** (enrolment no. 24/BIO/05) has carried out their search work presented in this thesis entitled **“Integrated Transcriptomic Analysis Reveals Potential Transposable Element Association in Alzheimer’s Disease”** for the award of **Master of Technology** from Department of Biotechnology, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student herself, and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/ Institution.

P.K.
 14/05/2026

Prof. Pravir Kumar

Supervisor

Department of Biotechnology

Delhi Technological University

YH
 29.05.26

Prof. Yasha Hasija

Head of Department

Department of Biotechnology

Delhi Technological University
 Head of the Department
 Department of Biotechnology
 Delhi Technological University
 (Formerly Delhi College of Engg.)
 Bawana Road, Delhi-110042

Date: 29-05-2026

INTEGRATED TRANSCRIPTOMIC ANALYSIS REVEALS POTENTIAL TRANSPOSABLE ELEMENT ASSOCIATION IN ALZHEIMER'S DISEASE

Ananya Singh

ABSTRACT

Alzheimer's disease (AD) is a progressive neurodegenerative disorder which is characterized by gradual cognitive decline, neuronal dysfunction, extracellular amyloid plaques and intracellular tau tangles. Increasing evidence suggests that transposable elements (TEs), also known as “jumping genes”, become aberrantly activated during aging and can contribute to genomic instability and altered gene regulation in neurodegenerative diseases. However, the relationship between dysregulated TEs and Alzheimer's disease-associated genes remains insufficiently explored at present. The present study is therefore aimed to investigate the differential expression of TEs in AD and identify TE-associated genes and pathways that are potentially involved in disease pathology through integrated transcriptomic and bioinformatics approach.

Publicly available RNA-sequencing data from post-mortem Alzheimer's disease and control brain samples were analyzed. Sequence reads were checked for adapter sequences and Phred quality scores, sequences with low Phred score and adapter sequences were trimmed. The reads were then aligned to the human reference genome, and quantified using the TEtranscripts package. Differential expression analysis was performed using DESeq2 to identify significantly dysregulated TEs. Significant TEs were mapped to overlapping genes using BEDTools, followed by KEGG pathway enrichment, Cytoscape network analysis, and TE–gene correlation analysis.

Differential expression analysis identified significantly upregulated TEs belong to LINE, ERV and DNA transposons family. Genomic mapping analysis identified TE-associated genes, some of which were enriched in neuronal related pathways including calcium signaling, cAMP signaling and Cholinergic synapse pathways. Network analysis identified hub genes. Further correlation analysis revealed strong negative association between dysregulated TEs and GSK3B and PRKCA genes, suggesting possible association between TE activity and neuronal pathways in AD.

The results of the study are in line with the growing evidence that TE dysregulation is associated with neurodegenerative diseases. The study also highlights that integrative TE expression profiling along with pathway enrichment, and network biology approach has the potential to unravel the complex and multifaceted molecular mechanisms underlying neurodegeneration.

Keywords: Alzheimer's Disease, Transposable Elements, Differential Expression Analysis, Pathway Enrichment, Transcriptomics

LIST OF PUBLICATIONS

1. **Poster Presentation** in SNCI (Society for Neurochemistry, India) 2026, Delhi Chapter at Jamia Hamdard University on “Integrative transcriptomic analysis reveals the role of Transposable Elements in Alzheimer's Disease through Differential Expression and Pathway Enrichment”

TABLE OF CONTENTS

Title	Page No.
Acknowledgement	ii
Candidate's Declaration	iii
Certificate by the Supervisor	iv
Abstract	v
List of Publication	vi
Table of Contents	vii
List of Tables	x
List of Figures	xi
List of Symbols, Abbreviations and Nomenclature	xii
 CHAPTER 1	
INTRODUCTION.....	1
 CHAPTER 2	
LITERATURE REVIEW.....	11
2.1. Historical Perspective of Transposable Elements.....	11
2.2. Emergence of Transposable Elements in Human Disease Research.....	12
2.3. Age associated derepression of Transposable Elements.....	14
2.3.1. Transposable Element Activation in Cellular Senescence.....	14
2.3.2. Loss of Heterochromatin Integrity During Aging.....	14
2.3.3. Transposable Element Mediated Inflammatory Signaling.....	15
2.3.4. Experimental Models Demonstrating TE Derepression During Aging.....	15

2.4. Role of Transposable Elements in Neurodegenerative Disorders...	16
2.4.1. Alzheimer's Disease.....	16
2.4.2. Parkinson's Disease.....	17
2.4.3. Huntington's Disease.....	17
2.4.4. Amyotrophic Lateral Sclerosis.....	18
2.5. Molecular Mechanisms associated with TE reactivation.....	19
2.5.1. Genomic Instability Induced by Transposable Elements.....	19
2.5.2. Neuroinflammation and Innate Immune Activation.....	19
2.5.3. Epigenetic Dysregulation and Chromatin Remodelling.....	20
2.5.4. Transcriptional Dysregulation and Regulatory Rewiring.....	21
2.6. Bioinformatics Approaches used in TE Analysis.....	21
2.7. Current Research Gap.....	23

CHAPTER 3

METHODOLOGY.....	24
3.1. Data Acquisition.....	24
3.2. Quality Check and Trimming.....	24
3.3. Alignment to Reference Genome.....	24
3.4. Sorting and Indexing Alignment File.....	25
3.5. Transposable Element Quantification.....	25
3.6. Differential Expression Analysis.....	25
3.7. KEGG Pathway Enrichment.....	25
3.8. Hub Gene Identification.....	25
3.9. Gene-Transposable Element Correlation Analysis.....	26

CHAPTER 4

RESULTS.....	27
4.1. Data Acquisition and Preprocessing.....	27
4.2. Transposable Element Quantification.....	27
4.3. Differential Expression Analysis.....	28
4.4. KEGG Pathway Enrichment.....	28

4.5. Hub Gene Identification.....	33
4.6. Gene-Transposable Element Correlation Analysis.....	33
 CHAPTER 5	
DISCUSSION.....	35
 CHAPTER 6	
CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT.....	37
6.1. Conclusion.....	37
6.2. Future Scope.....	37
6.3. Social Impact.....	38
 REFERENCES.....	39
APPENDICES.....	45
LIST OF PUBLICATION AND THEIR PROOF.....	52
PLAGIARISM VERIFICATION.....	53

LIST OF TABLES

	Title	Page No.
1.1	Overview of RNA transposon superfamilies	4
1.2	Overview of DNA transposon superfamilies	6
2.1	Transposable Element Reactivation in Neurodegenerative Diseases	18
2.2	Comparison of Bioinformatics Tools Used in TE Expression Studies	23
4.1	Overview of dataset used for the study	27
4.2	List of 5 upregulated TEs obtained from Differential Expression Analysis	28
4.3	KEGG Pathway Enrichment of TE associated Genes (filtered by $p\text{-adj} < 0.05$)	29
4.4	Node Table for gene-gene network	33
4.5	Significant correlation identified between TE and TE-linked Genes	34

LIST OF FIGURES

	Title	Page No.
1.1	Classification of Repetitive DNA on the basis of frequency.	2
1.2	Classes of transposons: (a) Retrotransposons mobilize by a 'copy-paste' mechanism. Retrotransposons are first transcribed into RNA, which is subsequently reverse transcribed into cDNA. The cDNA then integrates itself into new positions in the genome. (b) DNA transposons mobilize by 'cut-paste' mechanism. Transposase binds to both ends of DNA transposons and cleaves it from the donor site. The excised DNA transposons are reintegrated into the target DNA.	7
3.1	An overview of the methodology followed for the study	26
4.1	Heatmap showing the normalized expression of upregulated TEs in 16 samples (8 AD; 8 Control)	27
4.2	Dot plot showing KEGG pathway enrichment results, where dot size indicates gene count and color denotes p-value significance.	31
4.3	Calcium Signaling Pathway, the genes highlighted in red are associated with transposable elements linked to Alzheimer's Disease	31
4.4	cAMP Signaling Pathway, the genes highlighted in red are associated with transposable elements linked to Alzheimer's Disease	32
4.5	Cholinergic Synapse Pathway, the genes highlighted in red are associated with transposable elements linked to Alzheimer's Disease	32
4.6	Gene-Gene Network of TE associated genes dysregulated in significant neuronal pathways.	33

LIST OF SYMBOLS, ABBREVIATIONS AND NOMENCLATURE

AD	Alzheimer's Disease
ALS	Amyotrophic lateral sclerosis
A β	Amyloid Beta
cGAS	Cyclic GMP-AMP synthase
CREB	cAMP Response Element Binding Protein
DNA	Deoxyribonucleic Acid
DNMT	DNA methyltransferase
ERV	Endogenous Retro Virus
FTD	Frontotemporal dementia
FUS	Fused in Sarcoma (gene)
GEO	Gene Expression Omnibus
GSK3B	Glycogen Synthase Kinase 3 Beta (Gene)
HD	Huntington's disease
HERV	Human Endogenous Retro Virus
ITR	Inverted Terminal Repeat
KEGG	Kyoto Encyclopaedia of Gene and Genome
LINE	Long Interspersed Nuclear Elements
LOAD	Late onset Alzheimer's Disease
LTR	Long Terminal Repeats
MDA5	Melanoma Differentiation-Associated Gene 5
NGS	Next Generation Sequencing
ORF	Open Reading Frame
PD	Parkinson's Disease

PRKCA	Protein Kinase C alpha (Gene)
RIG-I	Retinoic Acid-Inducible Gene I
RNA	Ribonucleic Acid
RTEs	Retrotransposons
satDNA	Satellite DNA
SINE	Short Interspersed Nuclear Elements
SQuIRE	Software for Quantifying Interspersed Repeat Expression
SSR	Simple Sequence Repeat
STING	Stimulator of Interferon Genes
STR	Short Tandem Repeat
SVA	SINE-VNTR-Alu Like region
TE	Transposable Elements
TIR	Terminal Inverted Repeat
TPD 43	Transactive response DNA binding protein 43 kDa
TR	Tandem Repeats
VNTR	Variable Number Tandem Repeat

CHAPTER 1

INTRODUCTION

Repetitive DNA is sequence of DNA that is found within the genome in multiple repeated copies. Both prokaryotic and eukaryotic organisms contain a certain proportion of repeats in the genome. For instance, approximately 45% of human genome is composed of repetitive DNA [1]. Despite their abundance, repetitive DNA was once dismissed as 'junk' or 'non-functional' DNA as they are mostly located within the introns [2]. However, increasing research highlights the structural and functional importance of repetitive sequences, repetitive DNA sequences in both coding and non-coding region have significant impact on evolution gene expression regulation and variation indication. Not only do the repeats in coding region get translated canonically, the repeats in non-coding region and even telomeric repeat RNAs get translated non chronically [3].

Repetitive DNA sequences on the basis of their frequency can be divided into two classes. The first class of repetitive DNA comprises of satellite DNA sequences (satDNAs) which are high frequency repeats found in various chromosomal regions. satDNAs sequences generally comprise of the constitutive block of heterochromatin that forms essential structures such as centromere and telomere. However, satDNA Restricted to heterochromatin region and is also found in euchromatin region. The length of repeating motif of satDNA can range from few base pairs to over one kilo base pair and can be repeated for over 10^6 times, creating arrays that may extend up to 100 megabases [4]. satDNAs constitute approximately 8 to 10% of the entire human genome [5]. Moderate frequency repeats compromise the second class of repetitive elements. The length of the repeats motifs in these can be between 500-300000 bp, which can be repeated for $10-10^5$ times. Approximately 30% of repetitive DNA in the genome are moderate frequency reads. These repeats can further be divided into two subclasses, the first class consists of microsatellites and minisatellites, and the second class primarily consists of dispersed repeats made of transposable elements. Moderate frequency repeats are known to alter gene expression regulations. These repeats can affect gene expression through various mechanisms including chromatin structure modification transcriptional regulation and post transcriptional processing for instance these repeats may act as an enhancers or silencers of gene expression additionally they can form secondary structure that interfere with transcription or RNA processing leading to changes in gene expression level [4].

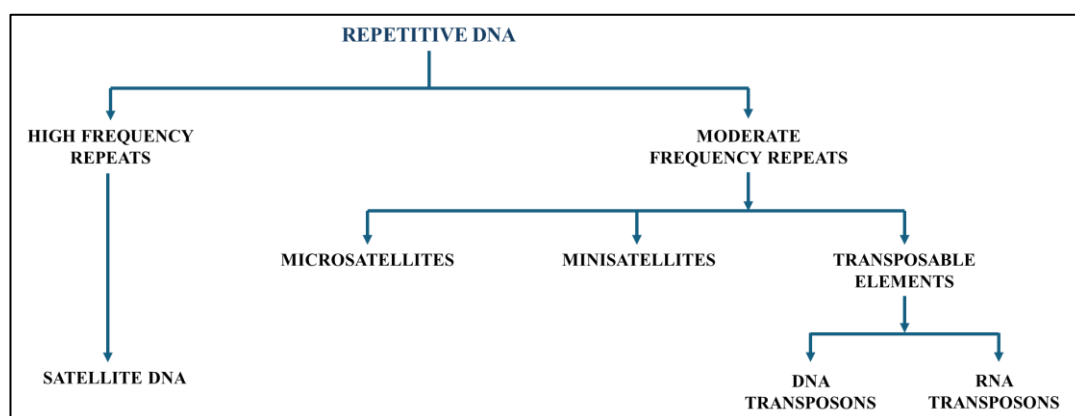


Figure 1.1: Classification of Repetitive DNA on the basis of frequency.

Further, on the basis of arrangement, repetitive DNA is classified into tandem repeats (TRs) and transposable elements (TEs) [4].

Tandem Repeats

Tandem repeats refer to an array of sequences formed by repeated sequences of basic repeat motifs placed one after another in the genome. TRs often tend to aggregate in particular chromosomal locations such as the centromeres and telomeres. TRs are crucial in many cellular processes including chromosomal segregation, genome organization and telomere protection. TRs can be further classified into microsatellites and minisatellites and centromeric satellites. Microsatellites also known as SSRs (simple sequence repeats) or STRs (short tandem repeats) are 1 to 6 basepair repeat motifs that are typically repeated 5 to 50 times [6]. Microsatellites also known as variable number of tandem repeats (VNTRs) are 10-60 basepairs DNA motifs that are repeated two to several hundred times [7]. In humans, centromeric DNA can be categorized into: (a) alpha satellite, (b) satellite II, and (c) satellite III. Satellite II and satellite III are composed of variations on ATTCC motif. Telomeric satellites consist of 300-8000 repeats of CCTAA/ TTAGGG motifs that can stretch up to ~2 to 50 kilobases on the telomeres. Subtelomeric satellite repeats are found between telomere and rest of the chromosomes and span for about 100-300 Kb [4].

Transposable Elements

Transposable elements (TEs) are dynamic DNA sequences that are capable of mobilizing within the genome and are also therefore known as 'viral elements' or 'jumping genes'. TEs were first discovered in maize by Barbara McClintock. They are now recognized in all prokaryotic and eukaryotic organisms and constitute a significant proportion of the species genome. For example, TE comprises approximately 45% of human genome, 37% of mouse genome, and more than 80% of maize genome [60]. Sequence homology was identified between satDNAs and transposons in many species, this suggests the existence of an intimate evolutionary relationship between TEs and emergence of repetitive sequences [8]. TEs were initially

considered ‘junk’ but are now recognized essential in regulating genome functions and are also key contributors in disease pathogenesis [9]. TEs are classified into two major categories: Class I transposable elements commonly known as RNA transposons or retrotransposons (RTEs) and Class II elements commonly known as DNA transposons. TEs can either be non-autonomous or autonomous based on their capability to encode the proteins required for transposition. Autonomous TEs can encode for all the proteins required for transposition on their own and therefore have their own mobility, while non-autonomous transposons can only mobilize if the required proteins are provided by a nearby compatible autonomous TE [10; 11].

RNA Transposons

RNA transposons or Retrotransposons are found exclusively in eukaryotes and mobilize via “copy-and-paste” mechanisms (*figure 1.2*). Based on their structure and transposition mechanisms, retrotransposons are categorized into two groups: the long terminal repeats (LTRs) and the non-LTRs retrotransposons. Both these groups are comprised of several subfamilies of RNA transposons [10].

LTRs are characterized by identical long terminal repeats (LTRs) present at both their ends. These LTRs can be short (about 100 bp) or very long (over 5000 bp). Structurally LTR transposons are very similar to retroviruses, the main difference being that retrovirus have extra genes for envelope proteins which allow them to form infectious particles while LTR transposons do not. The first LTR retrotransposon discovered was Ty1 in *Saccharomyces cerevisiae* in 1979. It was further experimentally demonstrated that these elements transpose via an RNA intermediate. There are two major superfamilies of retrotransposons viz Ty1/Copia and Ty3/Gypsy, which are widely represented in yeast and *Drosophila*. Both share the same essential structural and functional domains, though they differ in the order of encoded proteins. Generally, LTR retrotransposon transcripts are translated into polyproteins, which are then processed by an endogenous protease into capsid protein (gag), integrase, and reverse transcriptase/ribonuclease H [10].

Another significant superfamily of LTR transposons is endogenous retroviruses (ERV), which are thought to have originated from retroviruses. Few ERVs have retained their envelope protein (env) genes, whereas some have lost them. Human endogenous retrovirus (HERV) consists of 8% of total genome and are mostly inactive in the germline. However increasing number of studies suggest the involvement of HERVs in cancer and other diseases which still remains debatable.

Non-LTR RTEs are broadly classified into two major classes: long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). LINEs are autonomous elements, often several kilobases in length, that can encode proteins like reverse transcriptase, endonuclease and an RNA binding protein required for their own mobilization. In contrast SINEs are short, 50 to 500 bp, non-autonomous elements that are derived from cellular RNA polymerase III transcripts. Their proliferation depends on enzymatic machinery provided by LINEs. Both LINEs and SINEs transpose via

target primed reverse transcription, a machinery that occasionally can capture host mRNAs leading to the generation of pseudogenes [10].

The LINE1 (L1) family is the most extensively studied in humans. More than 500,000 L1 copies account for approximately 17% of human genome, although majority of these are truncated or mutated and therefore inactive. Nevertheless, a small number of functional L1 elements remain, and their activity has significant biological consequences. More than 65 hereditary disorders have been associated with L1 insertions, and estimates indicate that every 20th to 200th birth carries a novel germline L1 insertion [10].

SINE elements, particularly the Alu family, are also prominent in humans. Alu elements are approximately 300 bp sequences that are derived from 7SL-RNA. There are around one million copies of Alu elements that constitute approximately 11% of the genome. A majority of these elements are present in the intergenic, intronic and 3' untranslated regions [10]. These can disrupt the gene function through insertional mutagenesis or unequal homologous recombination. It is estimated that Alu elements contribute roughly 0.4% of all human genetic disorders.

A further composite non-LTR family is SVA that is specific to hominids. SVAs consist of a retroviral-derived SINE-R sequence, variable number tandem repeats (VNTRs), and an Alu-like region. Mobilization of SVAs is dependent on L1-encoded proteins. Although relatively rare with approximately 2800 copies in human genome, SVAs have been associated with several diseases [10].

Table 1.1: Overview of RNA transposon superfamilies

Class	Superfamily	Description	Reference
LTR Retrotransposons	Ty1/Copia	It is the most studied group among the LTR RTEs. It encode for four protein domains which are essential for its transposition.	[10]
	Ty3/Gypsy	It is similar to Ty1/Copia and also encodes for the four protein domains required for its transposition, the difference being in the order of coded proteins.	[10]
	ERVs (Endogenous Retrovirus)	LTR retrotransposons derived from ancient retroviral infections; some retain envelope protein (env) genes; ~8% of human genome.	[10]

Non-LTR Transposons	LINEs	Autonomous; encode for [10] proteins required for its mobilization, which include reverse transcriptase, endonuclease and RNA binding protein.
	SINEs	Nonautonomous; derived [10] from RNA Polymerase III transcripts and are dependent on LINE provided transposition machinery.
	SVA	Composite; hominids [10] specific; contains SINE, VNTRs, Alu like region.

DNA transposons

DNA transposons use a “cut-and paste” mechanisms to self-amplify (*figure 1.2*). DNA transposons have been pivotal during primate evolution, although it is now believed that most of them are now inactive within the genome [60]. A DNA transposon typically consist of DNA transposase coding sequence which are flanked by inverted terminal repeat (ITR) on either side. The DNA transposase bin to the ITR and cleaves it to excise the transposon facilitating its reintegration to a new location within the genome. DNA transposons are classified into three major superfamilies: (i) terminal-inverted repeat (TIR) transposon (ii) Helitrons and (iii) Polinton/ Maverick elements.

The TIR transposon is further divided into nine major and several rarer superfamilies, that are differentiated on the basis of length of TIR. Autonomous members of TIR transposon family encode for transposase that bind to the end of elements and catalyzes transposition. Amongst the most studied groups is hAT superfamily which is characterized by short TIRs (10-28 bp), 8 bp TSDs, and a single ORF encoding Tpase, often forming hairpin intermediates during excision. The CACTA superfamily is widespread in plants, especially grasses, with longer TIRs (13-54 bp), though its transposition mechanism remains less understood. The Mutator (Mu) superfamily is highly mutagenic, with ~220 bp TIRs and 9-11 bp TSD, and is notable for preferential insertion near gene [12].

Helitrons were first identified in 2001 in *Arabidopsis thaliana*, rice, and *C. elegans*, and later found across planta, fungi, insects, and vertebrates. Unlike other DNA transposons, Helitrons lack terminal inverted repeats (TIRs) and do not generate target site duplication upon insertion. The defining structural features of Helitron include a TC dinucleotide at 5' end, an internal GC-rich inverted repeat of 6-11 bp, and a conserved 3' CTRR motif (commonly CTAG). Helitron transposition proceeds via a

rolling circle replication mechanism rather than double stranded cleavage mediated by a large Rep/helicase protein (1000-3000 amino acids) containing both replication initiator and helicase domains [13].

Polintons or Mavericks are the largest and most complex DNA transposons, typically 15-20 kilobase pairs in size. They are found in eukaryotic genomes and contain genes that are homologous to viral proteins [14].

Table 1.2: Overview of DNA transposon superfamilies

Superfamily	Description	Reference
Terminal-inverted repeat transposon	These elements are excised by double strand cleavage from their donor site and reintegrate as double stranded DNA elsewhere in the genome.	[12]
Helitrons	Helitrons are newfound transposons found in many eukaryotes they propagate through rolling circle mechanisms.	[13]
Polintons/ Maverics	Mavericks are Large DNA transposons found commonly in eukaryotic genome they have six base pair target size duplication sequence and genes that are homologous to viral proteins.	[14]

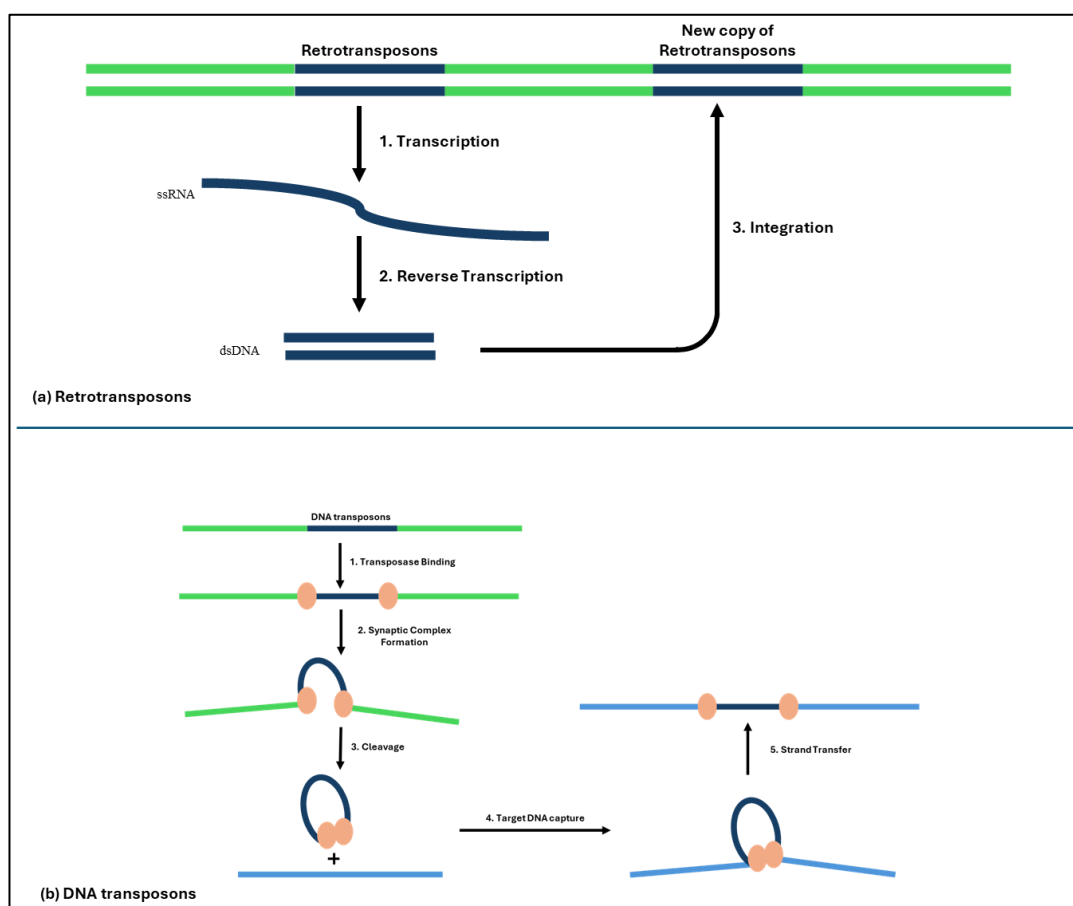


Figure 1.2: Classes of transposons: (a) Retrotransposons mobilize by a ‘copy-paste’ mechanism. Retrotransposons are first transcribed into RNA, which is subsequently reverse transcribed into cDNA. The cDNA then integrates itself into new positions in the genome. (b) DNA transposons mobilize by ‘cut-paste’ mechanism. Transposase binds to both ends of DNA transposons and cleaves it from the donor site. The excised DNA transposons are reintegrated into the target DNA.

Transposons pose a fundamental ‘genomic paradox’ while on one side there are powerful drivers of genetic diversification and evolution, on the other side uncontrolled activity of transposons compromise the genetic integrity and stability. Consequently, the activity of TEs are regulated by several epigenetic modifications such as DNA methylation and histone modification that favors heterochromatic formation and regulates the transposition. The mechanisms that the host genome uses to epigenetically regulate TEs are known as ‘genome defense’ mechanisms [15].

Epigenetic modification can be grouped into two types: direct modification of DNA or modification of histone protein that constitutes the core of nucleosome. The use of these processes to check the activity of TEs is well documented in plants and yeast [16].

DNA methylation is believed to be the most widespread mechanism of transposable element regulation in mammalian genome. DNA methylation involves the addition of methyl group to cytosine residues particularly inside CpG dinucleotides which are

abundant in TEs, such as LINE-1 and ERVs. DNA methylation is mediated by DNA methyltransferases, such as DNMT1 and DNMT3B for de-novo methylation and DNMT1 for maintenance of methylation pattern during DNA replication. Methylation of TE sequences facilitate transcriptional silencing of TEs in several ways. First, it directly inhibits the binding of transcription factors essential for TE expression. Second, methylated DNA is identified by methyl-CpG-binding proteins which recruit more repressive complexes to strengthen transcriptional repression. This creates compact chromatin structures that are inaccessible to transcriptional machinery. DNA methylation is significantly correlated with TE suppression and is required to minimize their genotoxic potential, while also maintaining the genomic stability by preventing new TE insertions [17].

In addition to DNA methylation, histone modification influence TE activity by altering the chromatin structure. Histone proteins undergo a variety of post translational changes including methylation, acetylation, phosphorylation, and ubiquitination, all of which alter the chromatin accessibility. Amongst these histone H3 lysine 9 trimethylation (H3K9me3) is crucial for TE silencing. This repressive mark is typically enriched in TE and is linked to formation of heterochromatin which is a highly condensed and transcriptionally inactive form of chromatin [18].

However, as we age these regulatory mechanisms break down resulting in reactivation of TEs, this is ‘transposon theory of aging’ ([19]). Considering that age is the major risk factor for neurodevelopmental disorders, there are increasing number of studies that have explored the relation between TEs and neurodegeneration. Age-associated transposable element derepression has been identified as a common characteristic of neurodegenerative illnesses, suggesting that TE activity may be a shared molecular vulnerability of the aging nervous system. Neurons are post-mitotic cells with limited capacity to withstand genomic damage by cell division and, as such, are particularly sensitive to cumulative damages generated by transposition events. TE expression has been reported to be upregulated in aging neural tissues in several model organisms, often in the context of heterochromatin integrity loss, decrease of DNA methylation at repetitive loci and breakdown of RNA surveillance pathways [19; 20]. These results suggest that TE reactivation is not a pathogenic event, but rather an integral part of the intrinsic features of brain aging that may contribute to neuronal degeneration.

Induction of genomic instability is one of the most notable processes by which transposable elements contribute to neurodegeneration. Retrotransposition events lead to DNA double-strand breaks during insertion or as a result of the endonuclease activity linked to LINE-1 mobilization. Such breaks accumulate and activate DNA damage response pathways, which are continuously activated in old neurons and in states of neurodegenerative illness. It has been established that chronic DNA damage signaling impairs synapse function, changes neuronal transcriptional programs, and eventually promotes cell death. Inhibition of activity of LINE-1 reduces DNA damage and extends neuronal lifespan in aging mice models emphasizing a causal link between TE activity and neurodegenerative symptoms. Besides the direct damage to the genome, transposable elements induce neurodegeneration also through the activation of innate immune mechanisms. RNA and DNA intermediates produced during

retrotransposition can accumulate in the cytoplasm and be recognized as danger-associated molecular patterns by cellular sensors such as cGAS, RIG-I, and Toll-like receptors. Activation of these pathways results in type I interferon responses and chronic neuroinflammation, a hallmark of neurodegenerative disorders including Alzheimer's disease. It is also reported that the reactivation of LINE-1 elements in senescent cells induces interferon signaling leading to expression of inflammatory genes. This emphasizes the molecular connection between TE activity, inflammation and aging brain. [20].

Alzheimer's disease is globally recognized as the most common form of dementia. AD is characterized by gradual decline in cognition, synaptic loss and neuroinflammation along with accumulation of extracellular amyloid- β plaques and intracellular neurofibrillary tangles of hyperphosphorylated tau. Epigenetic dysregulation caused by aging is increasingly recognized as one of the major players in pathogenesis of the disease. Several studies identified changes in DNA methylation patterns in human brains, these include hypomethylation at several repetitive loci such as LINE-1 and ERV. Since DNA methylation is a key mechanism to regulate the TE activity, these alterations indicate towards increased TE transcription, underscoring that loss of TE repression can be an early event in Alzheimer's disease.

Several studies have associated the activation of transposable elements with tau pathology, which is considered a major hallmark of AD. Frost et al. demonstrated that pathogenic tau lowers H3K9me3 levels that causes heterochromatin disorganization. Since the compact structure of heterochromatin prevents TE mobility, the disorganization leads to increased TE activity in neural cells. Restoration of heterochromatin integrity was shown to block tau-induced neurodegeneration in *Drosophila* models, establishing a functional link between heterochromatin integrity, TE activation and neuronal toxicity [21]. These studies imply that loss of heterochromatin can be a biological driver for transcriptional activation of TES leading to genomic instability and neuronal dysfunction.

Transcriptome based study on post-mortem brain samples of AD patients provide additional evidence of TE association. Bulk RNA Sequencing studies show several LINE, SINE and ERV elements were found to be upregulated in brain samples of AD patients than in age-matched control samples. Interestingly the dysregulation of TEs varies amongst different cell types. Single cell RNA sequencing studies have reported different pattern of TE expression in neurons, microglia and oligodendrocyte. Increased level of TE transcripts in microglia is associated with activation of inflammatory gene networks, suggesting the potential role of TEs in microglial activation and neuroinflammation pathways in AD [22]. Additionally, transposable elements can also influence gene regulation in many ways. Most TEs contain promoter or enhancer sequences that influences and regulates the activity of nearby genes. In context of neurodegeneration, derepressed TEs can operate as alternative regulatory elements and influence the expression of neighbouring genes involved in key neuronal pathways such as synaptic transmission, neuronal survival, etc. For instance, some AD-related genes that are involved in lipid metabolism and immune signaling are mapped near to TE-derived regulatory regions [4].

Another possible mechanism linking TEs to Alzheimer's pathogenesis is somatic retrotransposition in neurons. Most neurons actively repress transposition, however, low amounts of somatic LINE-1 insertions have been found in the human brain. Such events may become more frequent with age and disease-associated epigenetic relaxation, leading to neuronal mosaicism. The complete functional ramifications of somatic TE insertions have to be determined, although insertional mutagenesis in genes important for neuronal function could contribute to selective neuronal susceptibility in Alzheimer's disease. Advances in long-read sequencing methods are now allowing more reliable identification of somatic TE insertions in human brain tissue, opening new pathways to investigate their function in neurodegeneration.

RNA-mediated TE control pathways also appear to diminish throughout aging and in neurodegenerative diseases. Components of the piRNA pathway, which is important for germline TE silencing, are expressed at low levels in the brain but may still participate in TE control. Small RNA processing and RNA-binding proteins are impaired with age which could decrease the efficacy of post-transcriptional TE repression and therefore promote TE accumulation. Also, RNA-binding proteins linked in neurodegenerative illnesses such as TDP-43 have been found to bind repeated RNAs, suggesting that disruption of RNA metabolism may overlap with mechanisms of TE control in disease states.

Taken together, these data are consistent with a hypothesis in which age-related epigenetic deterioration leads to transposable element reactivation in the brain, thus providing a permissive environment for neurodegeneration. Transposable elements have evolved as more than just a passive marker of genomic instability and are now actively associated with several disease linked pathways through genomic disruption, transcriptional dysregulation or immunological activation. In Alzheimer's Disease, aberrant TE activation cross plays with other known pathogenic mechanisms such as tau aggregation, amyloid toxicity, neuroinflammation and others to accelerate the cognitive loss and neuronal dysfunction. The current view of the role of TEs and their association with neurodegeneration pathways lays a strong foundation to further study the involvement of TEs in Alzheimer's Disease pathogenesis.

CHAPTER 2

LITERATURE REVIEW

2.1. Historical Perspective of Transposable Elements

Over the last seven decades, our understanding of transposable elements has significantly expanded. From being considered as evolutionary noise, our understanding has now evolved to study TEs as dynamic genomic elements that influence genomic stability and alternatively acts as key regulatory element. This change in perspective is a manifestation of a broader trend in molecular biology, genomics and epigenetics, largely driven by technological advances and large-scale sequencing projects.

The first clues to transposable elements were given by the pioneering work of Barbara McClintock in the 1940s and 1950s, who discovered mobile genetic elements in studies of maize cytogenetics. McClintock showed that some genetic loci might move about the genome, affecting the expression of genes surrounding them. Her proposition that genomes are not static entities but can be internally rearranged contradicted the prevalent idea of static organization of genetic material and was greeted with scepticism for decades. Her work did not acquire wider appreciation until the 1960s and 1970s when the molecular characterization of transposable elements in bacteria, yeast and *Drosophila* took place, culminating in her Nobel Prize in 1983.

Yet despite these early discoveries, for long of the late twentieth century the prevailing view was that transposable elements were largely non-functional or parasitic DNA. With the advent of early genome sequencing initiatives, it became clear that a large fraction of eukaryotic genomes comprised of repetitive sequences produced from transposable elements. In humans, these sequences account for over half of the genome, and the phrase “junk DNA” has been widely used to designate the non-coding, repetitive portions that were assumed to be biologically irrelevant [1]. This view was supported by the discovery that many TEs were altered, truncated or transcriptionally silent, indicating that they were leftovers of ancient genomic incursions rather than active participants in cellular activity.

The early twenty first century strengthened the area of Transposable Elements research as with growing functional studies and comparative genomics, more and more researchers started questioning the hypothesis that TEs are useless i.e. ‘junk’ in the genome. Many studies highlighted the significant role of transposable elements during evolution by gene duplication, chromosomal rearrangements and exon shuffling. Further studies also highlighted the regulatory potential of TEs and how they can act as promoters, enhancers or even insulators affecting the gene expression. These initial

studies were instrumental in defining TEs as important player in expression regulation [2].

The emergence of next-generation sequencing (NGS) technologies is one of the key drivers behind this paradigm shift. NGS facilitated high-throughput, rapid and remarkably precise sequencing of DNA which subsequently enabled gene-wide mapping of Transposable Elements, transcriptional activity and epigenetic states at an unprecedented resolution. Large-scale projects like ENCODE revealed that a large fraction of regulatory elements in human genome actually overlap with Transposable Element derived sequences including ERV and LINE element. This highlighted the dynamic role of TEs to alter gene regulation networks, particularly in a lineage-specific manner at the time of evolution. Simultaneously, developments in epigenomics revealed the complex mechanisms that regulates the TE activity in host genome. DNA methylation, histone modification and small RNA related pathways were identified to be key in repressing TE activity and maintaining genomic integrity.

Lately, advanced long-read sequencing methods and single cell genomics have now paved way for more insightful understanding of biological nature of TEs. These approaches have enabled accurate resolution of repeat areas within the genome, that was previously not possible with short-read sequencing. These approaches have also allowed identification of cell-specific TE patterns and retrotransposition events. Overall, the historical trajectory of TE research reflects a broader trend in genomics, from the perspective of repetitive DNA as superfluous to an understanding of it as a major contributor to genome control and plasticity. This changing perspective has had obvious implications for current studies of the role of transposable elements in human disease.

2.2. Emergence of Transposable Elements in Human Disease Research

As the regulatory and evolutionary significance of transposable elements became more appreciated, attention was also turned to their possible role in human disease. TEs were no longer considered as neutral genetic components, but as conditional risk factors that, if dysregulated, may contribute to disease states. The change led to the establishment of transposable elements as an independent field of biological research, especially in the context of genomic instability, cancer, immunological disorders, aging and degenerative illnesses.

Cancer research offered one of the earliest and most convincing linkages between transposable elements and disease. Many studies have shown that worldwide DNA hypomethylation, a feature of many malignancies, leads to the re-activation of usually silenced transposable elements, including LINE-1 and endogenous retroviruses. Activation of TEs has been related with insertional mutagenesis, chromosomal rearrangements and altered gene expression, which could all contribute to carcinogenesis. Furthermore, the RNAs and proteins produced from TEs can trigger immunological responses, leading to the inflammatory tumor microenvironment.

These results identified TEs as both a driver of and a marker for genomic instability in cancer.

In addition to cancer, transposable elements are also linked to other autoimmune and inflammatory diseases. Aberrant expression of endogenous retroviral elements has been reported in diseases such as multiple sclerosis and systemic lupus erythematosus. In such cases, nucleic acids derived from TEs are the endogenous danger signals that activate innate immunological pathways and disrupt immunological tolerance. This provides an early insight into how TE dysregulation can induce systemic damage beyond localized genomic damage.

A notable achievement in TE-disease related research has been the revelation of the significance of TE activity during aging. During aging, epigenetic modifications are common and these include DNA hypomethylation, histone modifications and loss of heterochromatin. Such alterations affect repetitive sections of the genome disproportionately and contribute to enhanced transcriptional activity of transposable elements. The “transposon theory of aging” suggests that the re-activation of TEs at older age causes functional decline by increasing genomic instability, chronic inflammation, and cellular senescence [19]. This notion has received support from experimental research in several organisms, in which inhibition of TE activity was beneficial for the improvement of age-associated traits.

Importantly, deregulation of TEs with aging has suggested a conceptual link between TEs and age-related illnesses. Given that aging is the major risk factor for a number of chronic diseases such as neurodegenerative disorders, cardiovascular diseases and metabolic syndromes, TE reactivation has been hypothesized as a common biological mechanism underpinning age-related pathology. This theory has been particularly impactful in re-conceptualizing neurodegeneration as a disorder of epigenetic and genomic instability rather than only a disease of protein aggregation.

Another emerging theme in TE disease research is their impact on transcriptional control. Many transposable elements carry regulatory sequences which, when derepressed, can impact on the expression of neighboring genes. In diseased conditions with altered epigenetic regulation, TE generated promoters or enhancers can be activated causing inappropriate activation of stress response genes, immunological pathways or oncogenes. Such regulatory reconfiguration hinders the differentiation between direct and indirect effects of TE activation and highlights the necessity for integrative analyses taking into account both TE expression and host gene networks.

Together, these studies have definitively positioned transposable elements as complex contributors to human disease, with no one mechanism to rule them all, but rather multiple mechanisms through which TEs can contribute to pathology, including insertional mutagenesis, epigenetic disruption, immune activation, and transcriptional remodeling.

2.3. Age associated De-repression of Transposable Elements

Aging is accompanied by substantial epigenetic and molecular changes that collectively determine genomic stability. Transposable element derepression is a particularly noteworthy change associated with aging. As we age the TE activity increases, leading to increased levels of TE transcripts in many tissues. It is also noteworthy that TE activity is usually kept in check by cross play of DNA-methylation, histone modification and several RNA mediated pathways. Their derepression over aging is therefore indicative of a more general loss of genome protection systems. This process has been described as the “transposon theory of aging” proposing that cumulative TE activation contributes directly to age-related cellular malfunction and organismal degeneration [19].

2.3.1. *Transposable Element Activation in Cellular Senescence*

Cellular senescence is a state of stable cell-cycle arrest that increases with age and is marked by substantial chromatin reorganization, transcriptional reprogramming, and secretion of pro-inflammatory substances collectively called the senescence-associated secretory phenotype (SASP). Several studies have indicated significant overexpression of transposable elements, especially LINE-1 (L1) retrotransposons, in senescent cells. In seminal work, De Cecco et al. showed that L1 elements are transcriptionally active in senescent human fibroblasts, creating cytoplasmic DNA intermediates that activate innate immune signaling pathways [20]

Mechanistically, TE activation in senescent cells is associated with heterochromatin loss and decreased expression of epigenetic repressors. Senescence is characterized by chromatin modifications, including loss of H3K9me3 and redistribution of heterochromatin protein 1 (HP1), both of which are crucial for the silencing of TEs. This exposes the TE regions that were silenced because of the transcriptional machinery. Importantly, we found that pharmacological or genetic inhibition of LINE-1 reverse transcriptase activity attenuated interferon signaling and inflammatory responses in senescent cells, suggesting that TE activation is not merely correlative but functionally contributes to the senescent phenotype [20].

These findings have general implications for aging tissues in which senescent cells increase with age and potentially affect surrounding cells through paracrine inflammatory signals. Senescence-associated TE derepression may therefore constitute a source of persistent inflammation and tissue failure in aging.

2.3.2. *Loss of Heterochromatin Integrity During Aging*

One of the molecular hallmarks of ageing is the progressive loss of heterochromatin, especially at repetitive DNA-rich regions. Early studies in yeast and *Drosophila* have correlated global chromatin decondensation and redistribution of heterochromatic markers with ageing. Later studies have confirmed these findings in mammalian

systems, demonstrating an age-dependent reduction of constitutive heterochromatin at pericentromeric regions and transposable element loci [23]. Loss of heterochromatin directly impacts TE regulation, as repressive histone modifications such as H3K9me3 and H4K20me3 are required for TE silencing. In aged cells, de-repression of retrotransposons occurs due to reduced activity of histone methyltransferases and changed nuclear architecture. Experimental evidence for a causal association between chromatin integrity and aging comes from studies in *Drosophila* where restoring heterochromatin components can decrease age-associated TE activation and lengthen lifespan [24].

Genome-wide epigenetic profiling has shown that the loss of DNA methylation with age in mammals is predominantly occurring at repetitive sequences. LINE-1 hypomethylation has been suggested as a molecular characteristic of aging and is often utilized as a proxy marker for global methylation status. These epigenetic alterations make the TE loci more susceptible to transcriptional activation, especially in stress circumstances [17].

2.3.3. Transposable Element–Mediated Inflammatory Signaling

One of the most relevant outcomes of age-associated TE derepression is activation of innate immune pathways. Retrotransposon activity leads to the generation of aberrant RNA and DNA species, which accumulate in the cytoplasm and are detected by pattern recognition receptors, such as cGAS, RIG-I and MDA5. Activation of these sensors leads to downstream signaling pathways such as the STING pathway, which leads to the generation of interferon and chronic inflammation. This pathway has been firmly implicated in “inflammaging,” the continuous low-grade inflammatory condition observed in aging organisms. LINE-1-derived cDNA has been demonstrated to trigger the cGAS-STING pathway in aged cells causing production of interferon-stimulated genes and pro-inflammatory cytokines [25]. It is important to note that inhibition of reverse transcription decreases the inflammatory signals. This further supports TEs are directly involved in immunological signal activation. Since chronic inflammation is a shared risk factor in several age related diseases such as dementia, cancer and cardiovascular diseases, it can be assumed that TE-mediated immune activation is a common biological mechanism that connects genomic instability to aging.

2.3.4. Experimental Models Demonstrating TE Derepression During Aging

Experimental evidence for the transposon theory of aging has been found in a variety of model species, including yeast, *Drosophila*, mice and human cell systems. Loss of chromatin silencing at repeated loci in yeast results in increased retrotransposition and decreased longevity. In *Drosophila*, age-related derepression of a subset of retrotransposons corresponds with neural dysfunction, and genetic augmentation of heterochromatin suppresses TE activation and extends lifespan [30].

Mice lacking for epigenetic regulators such as DNMTs have premature aging characteristics with increased LINE-1 expression and genomic instability. These animal models provide strong evidence that inactivation of TEs is required for longevity and genome integrity. Additionally, reverse transcriptase inhibitor reduced the inflammatory signals and restored tissue homeostasis in aged mice, this highlights the relevance and positive impact of TE-targeted therapies [20]. Collectively, these experimental systems suggest that TE derepression is an evolutionarily conserved characteristic of aging across species, and that loss of TE regulation is not a passive consequence of aging but an active contributor to age-related decline.

2.4. Role of Transposable Elements in Neurodegenerative Disorders

Neurodegenerative disorders are characterized by progressive loss of neuronal structure and function, resulting in cognitive and motor deficits. Mechanistic models have been dominated by disease-specific protein aggregates such as amyloid- β , tau, α -synuclein and mutant huntingtin. However, there is a growing body of evidence indicating genomic instability and epigenetic dysregulation are key players in neurodegeneration. Loss of transposable element (TE) repression is a particularly significant source of neuronal vulnerability, given that neurons are post-mitotic cells with little capability to repair accumulated genomic damage. More recently, several independent investigations have showed abnormal TE activation in different neurodegenerative diseases, indicating TEs could be drivers and not passive by-products of disease.

2.4.1. Alzheimer's Disease

The most thoroughly researched neurodegenerative pathology in the context of Alzheimer's disease (AD) is transposable element dysregulation. Genome-wide transcriptome and epigenomic studies of post-mortem brains of AD have repeatedly found upregulation of numerous TE families including LINE-1 (L1), Alu elements and human endogenous retroviruses (HERVs). Sun et al. performed one of the first large-scale investigations analyzing TE expression in AD brain tissue comprehensively and found considerable elevation of LINE-1 and HERV transcripts in several cortical areas of AD patients compared with age-matched controls. The observed modifications in TE expression was associated with dysregulated DNA methylation suggesting that epigenetic repression is a major cause of aberrant TE expression.

Further studies of tau protein functionality has provided a major molecular breakthrough by connecting TE activity to AD pathogenesis. Frost et al., highlighted that pathogenic tau disrupts the heterochromatin organization, which results in chromatin decondensation and aberrant activation of transposable elements [21]. In *Drosophila* models, expressing tau, scientists have revealed that TE activation is a contributing factor in neuronal death and restoring heterochromatin integrity reduces neurodegeneration. These findings provided a concrete evidence that TE activation is directly connected to neuronal toxicity in tauopathies.

A more recent study by Guo et al., in mammalian systems revealed that tau aggregation in mice models and human AD brain samples is linked to activation of ERV elements, notably HERV-K elements. HERV activation was found to induce neuronal stress responses and increase inflammatory signalling, further highlighting the contribution of TEs in AD pathology. Most importantly, the retrotransposon activity was inhibited by reverse transcriptase inhibitors and neurodegeneration was reduced in mice model expressing tau, this emphasises a possible therapeutic approach [26].

Recent advancements in single-cell and single-nucleus RNA sequencing approaches have offered greater resolutions to study the cell-specific TE expression in Alzheimer's disease. Single-nucleus transcriptomic studies have revealed that TEs are selectively upregulated in neurons, microglia and oligodendrocytes [22]. Additionally, the upregulated expression of TE in microglia was associated with activation of immunological pathways, suggesting towards the active role of TEs in neuroinflammation, which is a key hallmark of Alzheimer's Disease and other neurodegenerative disorders.

2.4.2. *Parkinson's Disease*

Compared to AD, where extensive research has directly associated TE activation with disease pathology, there are fewer studies that have directly investigated TE dysregulation in PD pathology. However, there are evidence that point towards a similar dysregulation mechanism in PD. Parkinson's disease is essentially characterized by the loss of dopaminergic neurons in the substantia nigra region of the brain and α -synuclein aggregates within affected neurons. In PD pathology, disruption of epigenetic mechanisms used for regulating TE expression is associated with oxidative stress related damage and mitochondrial dysfunction.

Cellular and animal model studies have implicated that oxidative stress increases the expression of LINE-1 elements by altering DNA methylation patterns and histone modifications. Increased activity of LINE-1 in dopamine-producing neurons is also linked DNA damage and activation of apoptotic pathways. It is also demonstrated that α -synuclein alters histone acetylation to potentially relax the heterochromatin, creating a permissive environment for TE activity.

These studies are preliminary in implying TE dysregulation to Parkinson's disease pathology and further investigations are necessary to gain insights on the molecular mechanisms that relates TE activity of PD pathogenesis [27].

2.4.3. *Huntington's Disease*

The role of TEs in molecular pathology of Huntington's disease is extensively studied using many genetic and animal models. HD is caused by expansion of CAG repeats in Huntington gene, which encodes for a mutant Huntington protein which in turn affects transcriptional regulation and chromatin organization.

In mouse and *Drosophila* models used in HD studies, a generalized loss of heterochromatin is observed which aids transposable element activation. Mutant huntingtin protein is linked to loss of H3K9me3 at TE loci which increases the expression of retrotransposons, leading to genomic instability and neuronal death [30]. Inhibition of reverse transcriptase activity or restabilization of heterochromatin structure was shown to reverse neurodegeneration in TE-associated toxicity in HD [28]. Increased TE expression pattern was also observed in post-mortem human brain samples of HD patients. Together these studies, strongly associate TE dysregulation in HD pathology related molecular mechanisms.

2.4.4. Amyotrophic Lateral Sclerosis

Amyotrophic lateral sclerosis (ALS) also known as Lou Gehrig's disease is a progressive neurodegenerative disease that affects motor neurons controlling voluntary muscles and thus impairing movement. ALS is often linked to malfunctioning of RNA-binding proteins including TDP-43 and FUS. Dysfunction of TDP-43 is reportedly linked with accumulation of TE-derived exons in tissues of ALS patients. Additionally, TE expression was strongly linked with expression of nearby genes and several TE-linked gene changes affect crucial pathways like RNA metabolism, extracellular matrix organization and neuronal signaling in ALS. The study strongly indicates that TE can impact ALS pathology [61]. Elevated expression of HERV-K elements was also reported in brain and spinal cord tissue samples of ALS patients, additionally the envelope proteins of HERV-K has neurotoxic effects in neuronal cells [29].

These studies support the hypothesis that dysregulation of TE activity is a common across different neurological pathways in neurodegenerative diseases.

Table 2.1: Transposable Element Reactivation in Neurodegenerative Diseases

Neurodegenerative Disorder	TE Families Implicated	Experimental Evidence	Reference
Alzheimer's	LINE-1, HERVs	RNA-seq, epigenetics, tau models	[21]
Parkinson's	LINE-1, HERVs	Oxidative stress models	[27]
Huntington's	LINE-1, LTRs	<i>Drosophila</i> & mouse models	[30]
ALS/FTD	HERV-K	Patient tissue, RNA-binding proteins	[31]

2.5. Molecular Mechanisms associated with TE reactivation

Numerous studies have associated the reactivation of TEs in aging and other neurodegeneration related disorders. However the dysregulated mechanisms through which aberrantly expressed TEs contribute towards disrupted cellular function are very complex and diverse. However, the activity of transposable elements is not converged to a single molecular pathway that affects disease progression and pathology but is rather a complex interplay of various biological pathways such as immunological signaling, epigenetic mechanisms, RNA-mediated processes and transcriptional processes that causes genome instability and contribute towards disease progression. Increasing evidence highlights that together these biological processes increases the susceptibility of neuronal pathways to various endogenous and exogenous risk factors in aging and age-related diseases.

2.5.1. Genomic Instability Induced by Transposable Elements

Genomic instability induced by TE activation is one of the most well-established process that links TE-activity to neurodegeneration. RNA transposons such as LINE mobilise through an RNA intermediate which is then reverse transcribed into DNA that is inserted at new target positions. This mechanism is fundamentally associated with DNA damage as dsDNA breaks are created either during endonuclease mediated cleavage of elements or during aberrant DNA repair process during insertion. These DNA damages pose a serious threat to post-mitotic neural cells that have limited DNA repair ability. Experimental studies have demonstrated that LINE-1 activity increases the γ H2AX foci, which is a highly sensitive marker for double stranded DNA breaks in neuronal and senescent cells [25]. Lack of TE repressors like SIRT6-associated TE activation is linked with higher rates of chromosomal aberrations in aging phenotypes [32].

Mutations caused by aberrant insertions of TEs at new genomic location is also a key threat for genome stability and integrity. While somatic retrotransposition is relatively infrequent, even infrequent insertions into genes essential in synaptic function, neuronal survival or DNA repair could lead to selective neuronal susceptibility. Advances in long-read sequencing have started to uncover somatic TE insertions in human brain tissue, hinting that the accumulation of insertional events could play a role in neuronal mosaicism and functional variability in aged brains.

2.5.2. Neuroinflammation and Innate Immune Activation

Besides direct damage to the genome, transposable elements have major impacts on cellular homeostasis via activating innate immune pathways. Retrotransposon transcription and reverse transcription produce abnormal RNA and DNA species accumulating in the cytoplasm. These nucleic acids are recognized by pattern recognition receptors including cyclic GMP–AMP synthase (cGAS), retinoic

acid-inducible gene I (RIG-I), and melanoma differentiation-associated protein 5 (MDA5) and activate downstream inflammatory signaling cascades.

Cytosolic DNA from TEs activates the cGAS–STING pathway, resulting in type I interferon and pro-inflammatory cytokine production. De Cecco et al. showed that activation of LINE-1 in senescent cells is sufficient to trigger interferon signaling, contributing to chronic inflammation in aging. Similar pathways have been identified in neurodegenerative illnesses where prolonged neuroinflammation exacerbates neuronal damage and accelerates disease development [20].

Microglia appear to be especially susceptible to immunological impulses originating from TE in the brain. Single-cell transcriptome investigations in Alzheimer's disease have indicated that microglia have higher TE expression and activation of interferon-responsive genes indicating that TE activity might be an endogenous activator of neuroimmune responses. Continued activation of these pathways can change synaptic pruning, increase oxidative stress, and impair neuronal support functions, thus linking TE derepression to neuroinflammatory disease.

2.5.3. Epigenetic Dysregulation and Chromatin Remodelling

Epigenetic dysregulation causes, and results from, transposable element reactivation. Under normal conditions TEs are silenced by DNA methylation, repressive histone modifications (e.g. H3K9me3) and heterochromatin-associated proteins (e.g. HP1, KAP1). Heterochromatin relaxation and greater accessibility of TE loci by loss of these repressive markers are a hallmark of aging and neurological disorders.

An example of this mechanism, which has been well-characterised, is chromatin remodelling mediated by Tau. Frost et al., demonstrated that pathogenic tau disrupts heterochromatin integrity that results from the loss of H3K9me3 methylation. Since DNA methylation is a well-established epigenetic modification known to suppresses expression, loss of methylation results in increased expression of TEs in neurons [21]. Similar dysregulation of epigenetic mechanism is reported in Huntington's disease model where mutant Huntington protein alters the activity of histone methyltransferase that resulted in increased TE expression.

Interestingly, disrupted epigenetic control systems can propagate through a feed-forward cycle. Normally epigenetic mechanisms help keep in check the activity and expression of TEs, however when this regulation is disrupted, TEs become active. Once active the TEs can disrupt chromatic structure and cause DNA damage. The cell then recruits DNA repair machinery for repair, but aberrant DNA repair further disrupts the epigenetic balance. As a result, the instability increased in a self-reinforcing loop. This bidirectional association between TE activity and disruption of epigenetic mechanisms complicate the understanding of cause-effect relationship between them and in addition highlights the key role of heterochromatin integrity in suppression of TE activity.

2.5.4. *Transcriptional Dysregulation and Regulatory Rewiring*

In addition to affecting the genomic integrity and contributing towards inflammation related immunological signaling, transposable elements also contribute to underlying disease-related pathways by disrupting transcriptional mechanism. Many transposable elements have promoter, enhancers and transcription binding sites within them, which are normally suppressed by adequate control mechanisms. However, in case of TE activation, these influence the transcription and expression of neighbouring genes.

The ENCODE project was a major breakthrough that provided an evidence that many regulatory elements within the human genome overlap with TE sequences, majority with ERV elements. These regulatory elements may become abnormally active under diseased conditions, disrupting the normal gene regulatory mechanism and influence the expression of nearby genes. This affects key cellular pathways involved in stress response, neural signaling and immunological signaling [33].

In case of neurodegenerative diseases, this transcriptional rewiring may disturb genes expression patterns related to synaptic pathways, mitochondrial activity, neuroinflammation and neuronal cell death. It is noteworthy that these effects can be observed even in cases where there is no active transposition, that emphasise that TE regulatory elements in itself can have significant effect on cellular function.

2.6. **Bioinformatics Approaches used in TE Analysis**

Bioinformatic study of transposable elements is fraught with problems that are fundamentally different from those of traditional gene-centric transcriptome investigations. Transposable elements are repetitive and extremely identical at the sequence level, and due to their dispersed location in the genome, short-read sequencing data frequently yield ambiguously mapped reads that cannot be uniquely ascribed to a specific genomic locus. Early RNA-sequencing studies commonly either discarded multi-mapping reads or assigned them randomly, leading to systematic under-estimation of transposable element expression. The growing awareness of the biological relevance of TEs in aging and disease made apparent this restriction and led to the creation of dedicated computational tools that explicitly account for the repeating character of these sequences.

One of the first commonly used approaches was to measure TE expression at the family level, rather than at individual loci. Tools such as TETranscripts were created to account for multi-mapping reads by probabilistically assigning them to genes and families of transposable elements using expectation-maximization algorithms. This allowed researchers to keep repetitive readings without biased elimination, which was particularly valuable in determining global variations in TE expression across conditions. TETranscripts has been used in several investigations of aging, cancer and neurodegeneration showing widespread overexpression of LINE-1 and endogenous retroviral families. However, this method sums expression over all copies of a TE

family, and hence cannot differentiate between transcriptionally active and inactive insertions, nor can it identify disease-specific loci generating the observed signal [34].

As needs changed for tracking jumping genes more precisely, analysis tools evolved slowly. A big step forward came through programs like SQuIRE - built to catch repeat activity at very fine detail, even individual sites. Rather than general trends, it leans on reads that map clearly across complex areas, combining strategies meant for tricky repeats. This blend helps uncover faint but real TE movements linked to specific conditions. Most useful where only a few misbehaving sequences trigger disease in body cells. Out of nowhere, active transposable element sites tend to cluster rather than scatter wide. These bursts emerge in distinct regions once signs of brain disorders appear through SQuIRE analysis. Heavy computational power drives this approach - without it, things fall apart. Results swing wildly whenever sequence reads run short, mapping grows unclear, or data depth drops low. Unclear genomes add further strain. So reliability cracks under light sequencing setups [35].

Sorting fuzzy genetic signals often leans on odds, assigning them to precise jumping gene positions. Rather than assuming, programs such as Telescope weigh probable sources through Bayesian math, pinning expression down to specific DNA coordinates, even within near-identical repeats like human endogenous retroviruses. Thanks to that sharpness, scientists frequently turn to it when probing HERVs in diseases including ALS and Alzheimer's, where spotting exactly which location fires up can tie rogue elements to altered gene behavior or toxic neural buildup. Yet because Telescope relies on preset genomic maps, flaws or blanks there may hide real signals, while outcomes can waver based on configuration choices or how tidy the raw sequence data happens to be [36].

An overview of various TE-analysis techniques reveals major differences in outcomes, despite the same dataset being used. When alignment of reads changes - or overlap treatment varies - plus tweaks to gene databases, the final interpretation can flip entirely, evidence suggests. Small changes in gene activity hold weight in neurological disorders, so getting it right becomes essential. Recently, a few scientists have combined wide-scale familial patterns with exact positioning verification, supported by multiple software tools, followed by hints from external sources such as DNA methylation marks or deep sequencing data to back what they see [37].

Scientists and researchers rely on bioinformatic based approaches for the identification and interpretation of TE activity. These approaches shape conclusions about TE activity levels, locations, their presence at all in conditions like aging or neurological disorders. Given that impact unfolds behind every result, selection of tools and method of TE analysis should be made very cautiously. Comparing all available techniques and approaches when uncovering such complex biological patterns is very crucial.

Table 2.2: Comparison of Bioinformatics Tools Used in TE Expression Studies

Tool	Resolution	Strength	Limitation
Tetrascripts	Family-level	Robust, simple	No locus specificity
SQuIRE	Locus subfamily	& High resolution	Computationally intensive
Telescope	Locus-specific	Resolves ambiguous reads	Annotation-dependent

2.7. Current Research Gap

Although there has been significant progress in the understanding of biological and pathogenic importance of TEs, the exact mechanisms still remain unclear. Though TEs are implied in several studies, the integrated analysis that systematically relates TE activity to downstream molecular implication are still lacking at large. A major disconnect is the lack of integration of TE expression and its impact on nearby gene expression that may relate to a significant biological pathway. Most transcriptomic studies focus on TEs as distinct entity and do not involve coordinated TE-gene interactions. The downstream implication of TEs in genes related to pathways linked with neurodegeneration are only minorly characterised [33]. Another major concern is that most studies that reported increased levels of TEs in several neurodegenerative diseases but did not explore the causative connections. This also restricts the exploratory studies that can focus in targeting transposable elements as a novel therapeutic approach [19].

CHAPTER 3

METHODOLOGY

3.1. Data Acquisition

Public RNA-sequencing data was obtained from the NCBI Gene Expression Omnibus (GEO). Dataset was selected with the inclusion criteria of transcriptomic profiles of post-mortem human brain tissue of Alzheimer's disease cases and neurologically healthy cases [38]. RNA sequencing allows quantification of genome wide transcription at read count resolution, enabling robust differential expression analysis [39]. Given the increasing evidence of extensive TE alterations in neurodegeneration, it is important to investigate the transcriptomic profile for determining altered TE expression and its role in disease-associated transcriptional dysregulation in AD cases.

The raw FASTQ files were obtained using the SRA Toolkit and processed for downstream analyses.

3.2. Quality Check and Trimming

The raw reads were then assessed using FastQc to check for low quality reads and adapter sequences. FastQC [40] is a quality control tool for high throughput sequence data that evaluates base quality scores, adapter contamination and sequence overrepresentation.

Trim Galore [41] was then used for trimming. Trimming removed the adapter contamination, filtered low-quality reads and reads below a length threshold ensuring high-quality reads that are suitable for further analyses.

3.3. Alignment to Reference Genome

Bowtie2 [42] was used to align the trimmed reads to GRCh38/hg38 human reference genome. Bowtie2 is a fast and memory-efficient tool for aligning sequencing reads to reference genome sequence. Default alignment parameters were used to perform the alignment.

Prior to alignment, Bowtie2 requires a prebuilt index of the reference genome; for this study, the hg38 index was obtained from the publicly available repository maintained by Ben Langmead at <https://benlangmead.github.io/aws-indexes/bowtie>.

3.4. Sorting and Indexing Alignment File

SAM files obtained from alignment were converted to compressed, computer readable Binary Alignment/Map (BAM) format. After that, the BAM file was sorted by genomic coordinates and indexed using Samtools [43]. This allowed final BAM files to be optimized for computational efficiency and fast data retrieval. Annex A describes the commands used for data acquisition through preprocessing and alignment steps.

3.5. Transposable Element Quantification

Tetrascript was used to quantify TEs [34]. Tetrascript creates a combined count table by concurrently counting reads that map to transposable elements and annotated genes. The corresponding GTF files were provided for TE and Gene Quantification. Appendix B describes the command and flags used in analysis.

3.6. Differential Expression Analysis

To study the altered TE expression in AD and control cases, differential expression analysis was performed in R using DESeq2 Package (Appendix C). DESeq2 use median-of-ratios method to normalize the feature quantification and models expression using a negative binomial distribution. This enables reliable estimation of differential expression while also accounting for biological variability [44].

3.7. KEGG Pathway Enrichment

To identify genes associated with dysregulated TEs, genomic mapping analysis was performed using BEDTools [45] and *awk* scripts. Coordinates of significant TEs were extracted from TE annotation GTF file and converted into BED format. Similarly, gene coordinates were extracted from gene annotation file and converted into BED format. BED files have chromosome-wise genomic coordinates, which enabled mapping of TEs to their nearby genes. The TE coordinates were compared with genes coordinates to identify TEs that are located within the gene region. Genes with TEs within them were extracted and combined into a final gene list.

Further, KEGG Pathway Enrichment was performed using ClusterProfiler [46] in R (Appendix D) to get neurodegenerative pathways involving the extracted genes.

3.8. Hub Gene Identification

A gene network was created for the genes that were involved in significant KEGG pathways. The gene network was constructed on Cytoscape (ver 3.10.3) [47] using STRING PROTEIN [48]. The network based analysis was performed to identify 'hub

genes', which are highly connected genes and are considered biologically significant due to their key involvement in multiple neurodegenerative pathways.

3.9. Gene-Transposable Element Correlation Analysis

To understand the relationship between hub genes and dysregulated TEs, correlation analysis was performed. The analysis was performed using DESeq2 normalized expression. Pearson correlation analysis was performed to evaluate expression patterns between hub genes and dysregulated TEs (Appendix E). Further, correlation coefficients and p-values were calculated for each TE-Gene pair.

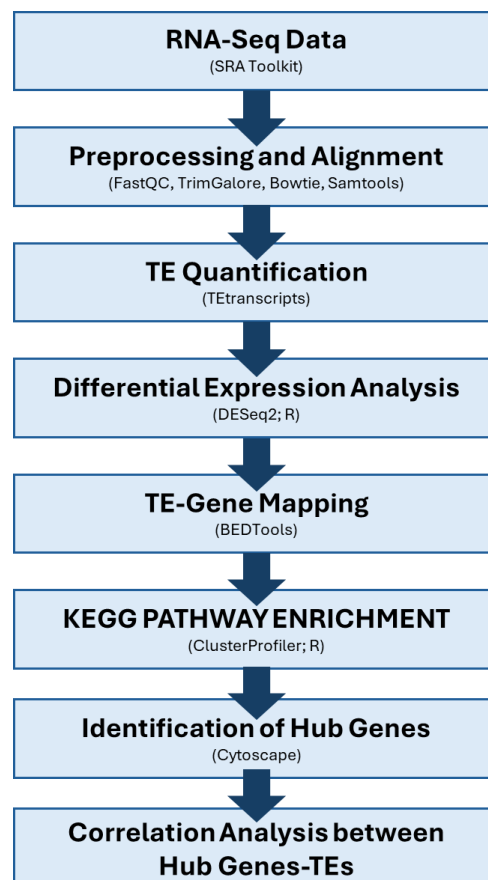


Figure 3.1: An overview of the methodology followed for the study

CHAPTER 4

RESULTS

4.1. Data Acquisition and Preprocessing

The dataset used in this study was retrieved from GEO. A total of 17 samples (9 AD cases and 8 controls) were obtained. However, one AD sample (SRR2422931) is not used in the study as it did not have quality scores needed for methodology.

Table 4.1: Overview of dataset used for the study

GEO Accession	BioProject Accession	SRA Accession	Sample Size	AD Samples	Control Samples	Tissue
GSE53697	PRJNA232669	SRP034831	17	9	8	Dorsolateral Prefrontal cortex

The dataset was carefully selected to include samples from brain region that highly correlated to AD .

Fastq files were downloaded for each sample, and an initial quality check and trimming were performed to remove low-quality reads and adapter sequences. Trimming removed any reads with Phred quality score lower than 20, reads shorter than 20bp and adapter sequence from the reads, ensuring sequence integrity. The trimmed sequences were then aligned to the human reference genome hg38. The alignment achieved high mapping rates across AD and control sequences. To ensure efficient computational handling the SAM files were converted to BAM files. The BAM files were then sorted and indexed enabling fast and random access for further analyses.

4.2. Transposable Element Quantification

Tecount within the Tetrascript package was used to quantify TEs and gene expression using the BAM files. Count table was generated separately for each sample, given the computationally extensive nature of the tool. The resulting count table for each of the sample was then merged on the basis of common features (GeneID and TE-ID) to generate a combined count matrix, which was then used for downstream analysis.

4.3. Differential Expression Analysis

Differential Expression Analysis was performed on the combined count matrix, using DESeq2 package in R. The results were filtered by p adj-values less than 0.05 and log-2-fold change greater than 1 to obtain significantly upregulated TEs. A total of 5 TEs were found to be significantly upregulated (Table 4.2). Figure 4.1 below shows the expression pattern of these TEs in different samples.

Table 4.2: List of 5 upregulated TEs obtained from Differential Expression Analysis

TE	Log2Fold Change	p adj value
FordPrefect:hAT-Tip100:DNA	1.59	0.018
HERV-Fc2-int:ERV1:LTR	1.18	0.026
L1HS:L1:LINE	1.07	0.014
LTR12B:ERV1:LTR	2.04	0.008
MER45C:hAT-Tip100:DNA	1.56	0.008

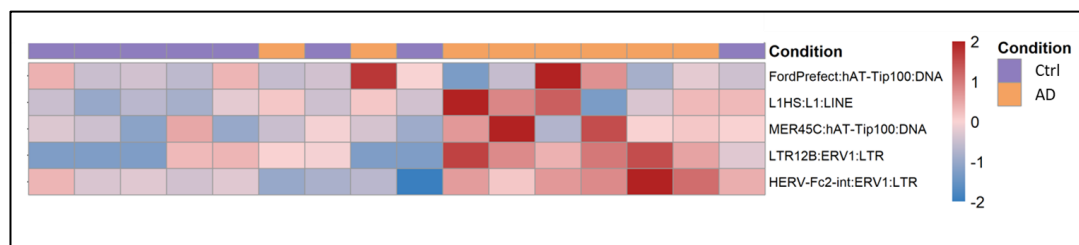


Figure 4.1: Heatmap showing the normalized expression of upregulated TEs in 16 samples (8 AD; 8 Control)

4.4. KEGG Pathway Enrichment

Significantly upregulated TEs identified through differential expression analysis were mapped to the overlapping genes using genomic coordinate-based analysis. Mapping the TEs to the overlapping genes revealed that TEs fall in the genic region of multiple genes. A total of 284 associated genes were extracted to generate a TE-associated gene list which was then used for downstream KEGG enrichment.

KEGG pathway analysis was performed on the TE-associated to identify biologically relevant pathways that may be dysregulated by transposable elements. KEGG enrichment revealed significant pathways are associated with genes linked to transposable elements. Pathways involved in nervous system, signal transduction, development and regeneration were considered for further analysis. Further, to identify

significant pathways a threshold of p-adj value less than 0.5 was applied. Table 4.3 provides a brief overview of KEGG enrichment.

Table 4.3: KEGG Pathway Enrichment of TE associated Genes (filtered by p-adj < 0.05)

ID	Description	Subcategory	Gene Ratio	p adj value	Genes
hsa04725	Cholinergic synapse	Nervous system	15/284	0.0002	ADCY1, ADCY2, CACNA1S, CAMK4, CHRM3, CHRNA7, CREB3L2, CREB5, GNG2, GNGT1, ITPR2, KCNQ1, KCNQ5, MAP2K1, PRKCA
hsa04020	Calcium signaling pathway	Signal transduction	23/284	0.0002	ADCY1, ADCY2, CACNA1S, CAMK1D, CAMK4, CD38, CHRM3, CHRNA7, GRIN3A, GRM1, HTR2C, HTR4, ITPKB, ITPR2, PDE1A, PDE1C, PHKB, PRKCA, PTGFR, RYR2, STIM2, TGFA, TRDN
hsa04727	GABAergic synapse	Nervous system	12/284	0.0009	ADCY1, ADCY2, CACNA1S, GABRA3, GABRB1, GABRG1, GABRG3, GNG2, GNGT1, GPHN, PLCL1, PRKCA

hsa04024	cAMP signaling pathway	Signal transduction	18/284	0.0053	ADCY1, ADCY2, ATP1B2, CACNA1S, CAMK4, CFTR, CREB3L2, CREB5, FSHR, GRIN3A, HTR1F, HTR4, MAP2K1, PDE4B, PDE4D, RAP1A, RYR2, TSHR
hsa04726	Serotonergic synapse	Nervous system	11/284	0.0122	CACNA1S, GABRB1, GNG2, GNGT1, HTR1F, HTR2C, HTR4, ITPR2, KCNN2, MAP2K1, PRKCA
hsa04724	Glutamatergic synapse	Nervous system	11/284	0.0122	ADCY1, ADCY2, DLGAP1, GNG2, GNGT1, GRIN3A, GRM1, GRM7, HOMER2, ITPR2, PRKCA
hsa04360	Axon guidance	Development and regeneration	14/284	0.0189	ABLIM2, EPHA3, EPHB1, GSK3B, LRRC4C, NRP1, PAK5, PRKCA, ROBO1, SEMA3C, SEMA3E, SEMA5A, TRPC6, UNC5C

Further, the three most significant KEGG pathways were analyzed considering both Gene Ratio and p-adj value.

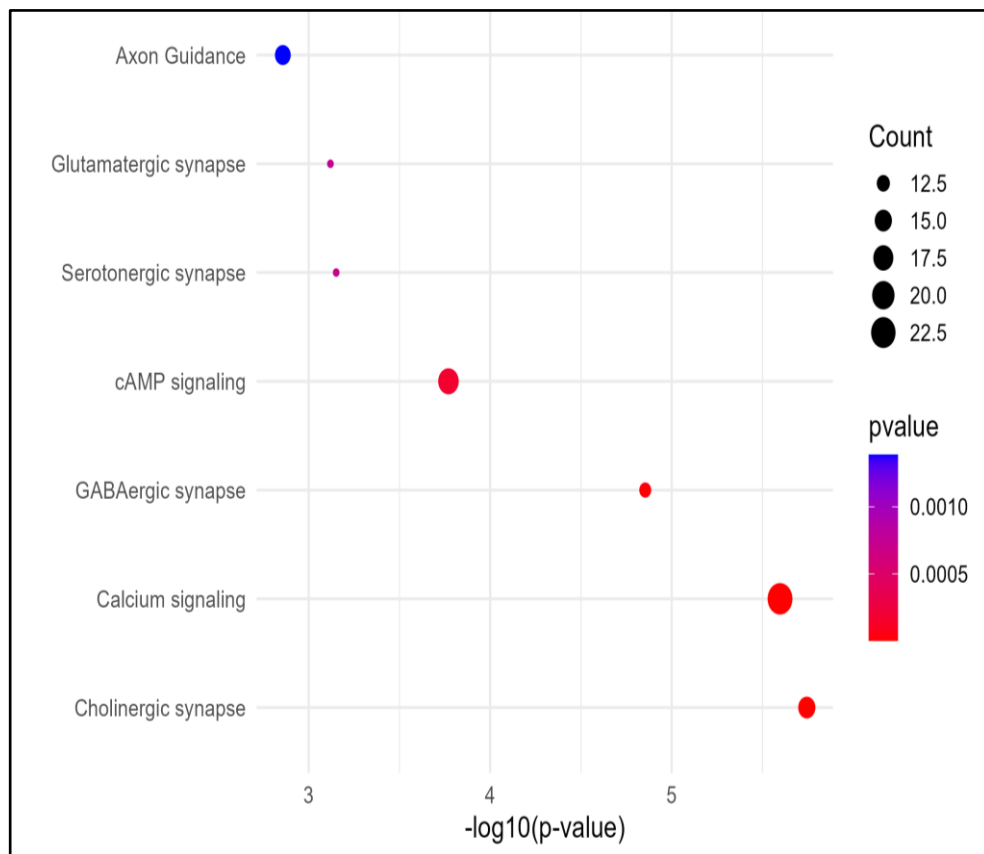


Figure 4.2: Dot plot showing KEGG pathway enrichment results, where dot size indicates gene count and color denotes p-value significance.

Based on the above analysis, calcium signaling pathway, cAMP signaling pathway and cholinergic synapse pathway were identified as top 3 most significant pathways.

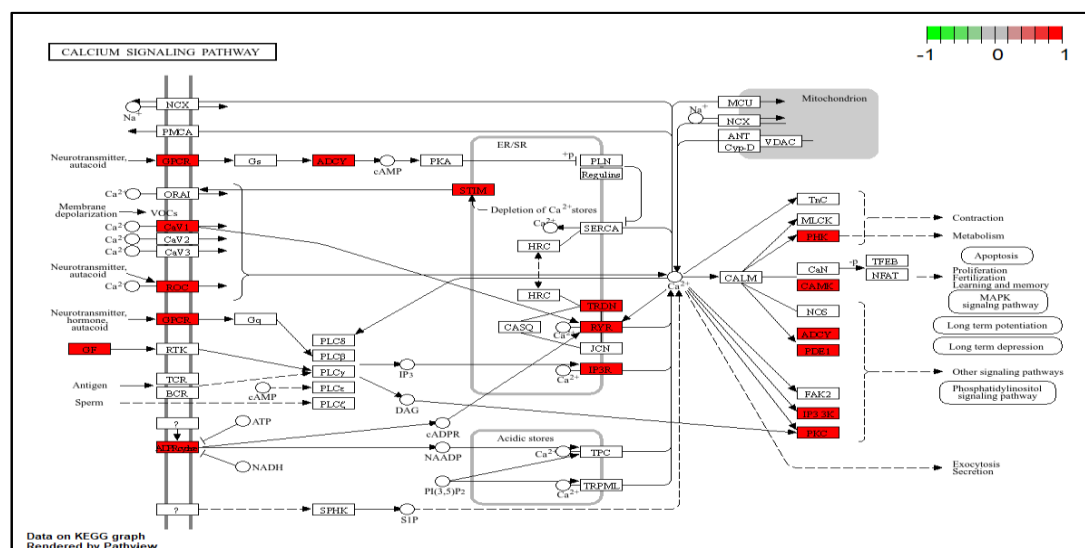


Figure 4.3: Calcium Signaling Pathway, the genes highlighted in red are associated with transposable elements linked to Alzheimer's Disease

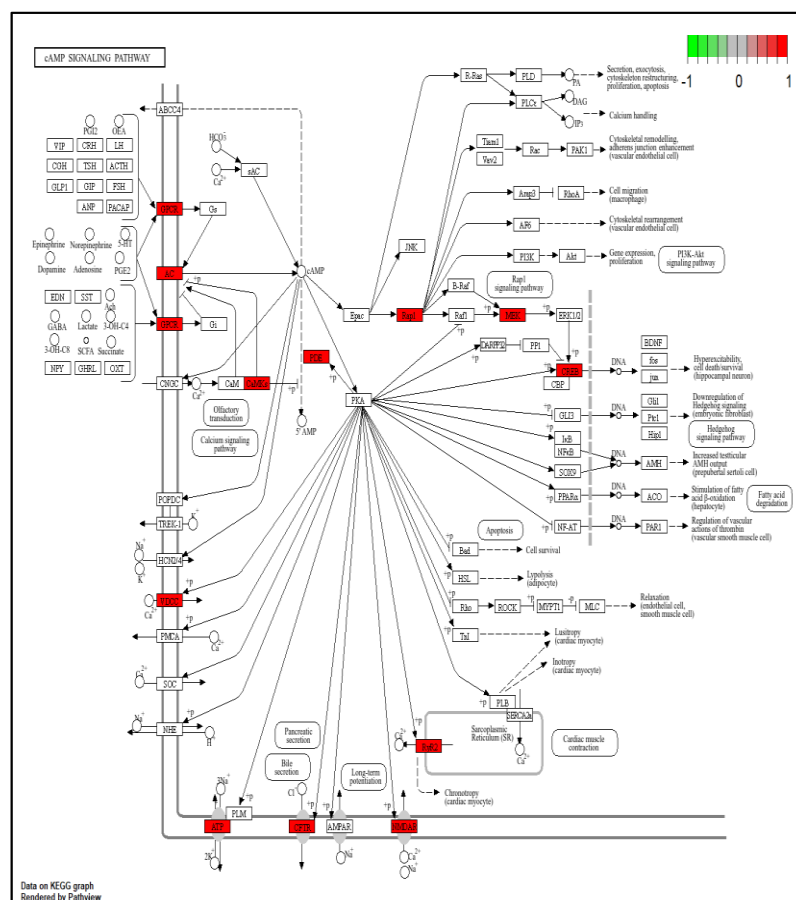


Figure 4.4: cAMP Signaling Pathway, the genes highlighted in red are associated with transposable elements linked to Alzheimer's Disease

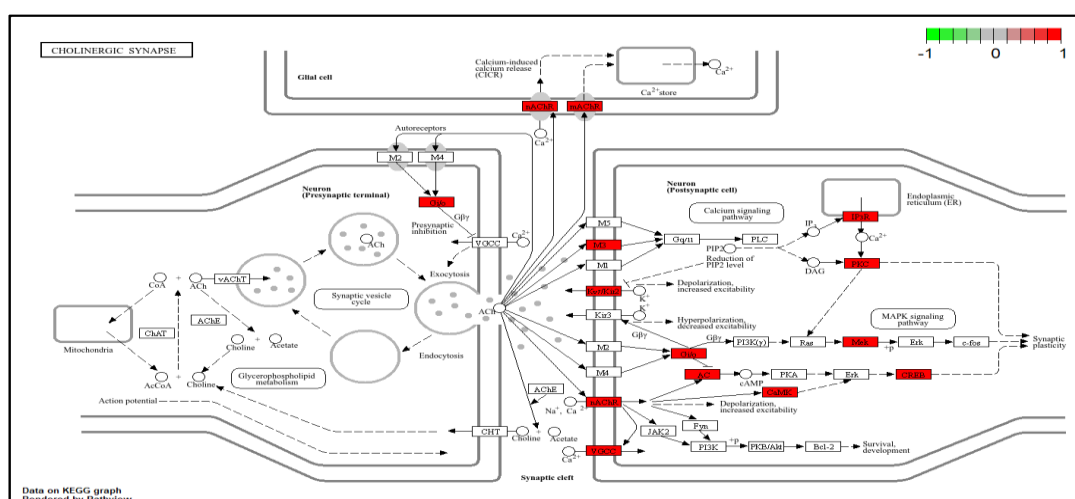


Figure 4.5: Cholinergic Synapse Pathway, the genes highlighted in red are associated with transposable elements linked to Alzheimer's Disease

Pearson correlation analysis was performed to understand the association between expression patterns of upregulated TEs and TE-linked hub genes associated with significant neuronal pathways. To maintain statistical significance the correlation

results were filtered using p-value threshold of less than 0.05. Amongst the identified association, TE families LINE, ERV and DNA transposons- hAT showed a highly significant negative correlation with *GSK3B* gene; LINE also showed a highly significant negative correlation with *PRKCA* gene (Table 4.5).

Table 4.5: Significant correlation identified between TE and TE-linked Genes

TE	Gene	Correlation	p-value
FordPrefect:hAT-Tip100:DNA	GSK3B	-0.94	0.0005
HERV-Fc2-int:ERV1:LTR	GSK3B	-0.77	0.02
L1HS:L1:LINE	GSK3B	-0.77	0.02
L1HS:L1:LINE	PRKCA	-0.71	0.04
LTR12B:ERV1:LTR	GSK3B	-0.77	0.02
MER45C:hAT-Tip100:DNA	GSK3B	-0.96	0.0001

CHAPTER 5

DISCUSSION

The present study aimed to investigate the transcriptomic alterations in post-mortem Alzheimer's disease (AD) brain tissue with emphasis on transposable element (TE) dysregulation and its association with disease-relevant genes. Differential expression analysis of TEs among control and AD test groups identified five significantly upregulated TEs, namely FordPrefect:hAT:DNA-Tip100:DNA, MER45C:hAT-Tip100:DNA, HERV-Fc2-int:ERV1:LTR, L1HS:L1:LINE, and LTR12B:ERV1:LTR. The LINE-1 upregulation observed in our sample aligns with several studies that demonstrate that LINE-1 retrotransposition contributes to genomic instability and affects neuronal plasticity and behavior [50; 51]. Roy et al., experimentally demonstrated elevated immunoreactivity of ORF1p (open reading frame 1 protein) in late-onset Alzheimer's Disease (LOAD) patients, which is encoded by LINE-1 in microglia. They also reported that upregulation of LINE-1 in iMG (iPSC-derived microglia) cause transcriptomic alterations in genes associated with lipid metabolism and antigen presentation, also impacting the expression of many known AD-relevant genes including Tau [52].

Similarly, LTR retrotransposons particularly ERV have also shown altered expression patterns in AD cases. HERV RNA have upregulated expression in samples collected from cerebrospinal fluid of AD patients [53]. An increase in LTR retrotransposition have been linked to AD. Differential expression of several retrotransposons associated with neurofibrillary tangle burden in human AD brains were observed in RNA sequencing data. This highlights the transcriptional activation of TEs among the LINE-1 and ERV clades. Tau-associated, active chromatin signatures were detected at multiple HERV-Fc1 loci. The results underline that TE activation is associated with tau neuropathy [26].

In addition to LINE and ERV retrotransposons, two DNA transposon elements were also significantly upregulated. Although DNA transposons are considered to be inactive in the human genome, growing evidence suggest that age related dysregulation of epigenetic mechanisms can lead to transcriptional derepression of DNA transposons. Studies on transposable element activity in neurodegeneration highlight the loss of heterochromatin and DNA methylation in both retrotransposons and DNA transposons families [54].

Further, KEGG enrichment analysis revealed significant neuronal pathways such as calcium signaling, cAMP signaling and cholinergic synapse pathways among the TE associated genes. These pathways are critical for neuronal communication, synaptic plasticity, and learning and memory, which are severely affected in AD pathogenesis. The calcium signaling pathway emerged as the most significant pathway in this study.

It was hypothesized in 1980s that imbalance in Ca^{2+} homeostasis, disrupts normal neuronal function and may lead to neurodegenerative diseases. Ca^{2+} homeostasis is known to play an important role in many aspects of neuronal regulation including development of learning and memory process, which is most severely affected in AD patients. Dysregulated Ca^{2+} levels can cause dysfunctional autophagy and apoptosis. Aging is a significant factor in LOAD, with age, the functionality of intercellular Ca^{2+} channels can decline resulting in Ca^{2+} overload within cells. This worsens the neuronal hyperexcitability, making neurons more susceptible to damage and ultimately leads to synaptic dysfunction which can cause $\text{A}\beta$ aggregation and hyperphosphorylation of tau. This highlights that imbalance of Ca^{2+} homeostasis may be linked to AD pathogenesis [55].

The cAMP signaling pathway is crucial for several processes, particularly in age-related disorders, where dysregulated pathway is associated with progressive decline of cell function eventually leading to loss of cellular integrity. During the last decades, it was established that changes in cAMP signaling pathway in brain regions can cause dementia-related pathologies which primarily affects the memory. Memory relies on synaptic plasticity, a process which is largely regulated by PKA which is fundamental in maintenance of synaptic plasticity and memory as it regulates transcription through CREB (cAMP Response Element Binding Protein) [56].

Cholinergic Synapses play a crucial role in various cognitive processes such as learning, memory and attention. The cholinergic hypothesis assumes that progressive loss of cholinergic neurons in brain significantly contributes to Alzheimer's Disease pathogenesis. Increase in levels of $\text{A}\beta$ is associated with disruption of cholinergic neurotransmission in AD patients [57].

Further network analysis and correlation analysis identified GSK3B and PRKCA genes as significantly correlated to upregulated TEs. GSK3B (Glycogen Synthase Kinase 3 Beta) is a ubiquitously expressed and active gene that regulates several key biological pathways, some of which are also implicated in neurodegeneration. Dysregulation of GSK3B is observed to affect $\text{A}\beta$ as well as tau metabolism in both in-vitro and in-vivo AD models. *GSK3B* activity is linked to key processes such as neurogenesis, synaptic plasticity and memory consolidation [58]. Further, PRKCA was also identified as a significant gene with strong negative correlation with L1HS:L1:LINE element expression. The gene encodes for Protein Kinase C (PKC) alpha, which is involved in diverse cellular signaling processes. In brain PKC alpha is prominent for learning and memory processes. Activation of PKC alpha strengthens synaptic connections and is believed to support learning and memory formation. Dysregulation of PKC is therefore associated with several neurological disorders including AD [59].

CHAPTER 6

CONCLUSION, FUTURE SCOPE AND SOCIAL IMPACT

6.1. Conclusion

The present study reported the dysregulation of transposable elements (TEs) in Alzheimer's disease with bioinformatics analysis based on RNA-sequencing. Differential expression analysis revealed five significantly increased TEs, including LINE, ERV and DNA transposon families, indicating altered repeating element activity in AD brain tissue.

Further pathway enrichment analysis showed that TE-associated genes were considerably enriched in calcium signaling, cAMP signaling and cholinergic synapse pathways, all of which play a vital role in neural transmission, synaptic plasticity, learning and memory. Network and association analyses further revealed that GSK3B and PRKCA are significant hub genes related to dysregulated TEs. The high correlations discovered between these genes and TE families suggest a probable association between TE activation and neurodegenerative signalling processes in Alzheimer's disease.

In conclusion, the study suggests that TE dysregulation may contribute to molecular changes related to AD and highlights the importance of TE based analysis in studies of neurodegenerative disease.

6.2. Future Scope

Future studies should include larger datasets and more brain regions for further validation of the found TE expression patterns in Alzheimer's disease. Experimental confirmation via qRT-PCR and protein level research may strengthen the findings obtained by computational analysis.

Combining epigenetic and single-cell transcriptome data can further aid to understand the mechanisms responsible for TE activation in neuronal cells. Moreover, investigating TE-gene regulatory networks and therapeutically targeting TE-associated pathways may provide more insights into the course of neurodegenerative diseases.

6.3. Social Impact

Alzheimer's disease remains the leading cause of dementia, accounting for 60% to 80% of dementia cases worldwide and a significant health problem in developed countries. Alzheimer's disease affect around 41 million people around the globe, as reported by World Health Organization. The disease not only impacts the patients but also their families and caregivers. It leads to significant economic cost, including healthcare and long-term personal care of the patients, which increases the emotional toll on loved ones and families. This study contributes in understanding of underlying molecular mechanisms involved in AD pathogenesis by highlighting the role of transposable elements in neurodegeneration.

In future, the findings of the study can support identification of TE-based biomarkers and therapeutic strategies for early diagnosis and management of Alzheimer's disease. In addition, the study also underlines the importance of bioinformatics-based approaches in advancing neurological research and understanding the complex and multifaceted pathogenesis of human diseases.

REFERENCES

- [1] Zepeda-Mendoza, C. J., Lemus, T., Yáñez, O., García, D., Valle-García, D., Meza-Sosa, K. F., Gutiérrez-Arcelus, M., Márquez-Ortiz, Y., Domínguez-Vidaña, R., Gonzaga-Jauregui, C., Flores, M., & Palacios, R. (2010). Identical repeated backbone of the human genome. *BMC Genomics*, 11, 60. <https://doi.org/10.1186/1471-2164-11-60>
- [2] Gemayel, R., Cho, J., Boeynaems, S., & Verstrepen, K. J. (2012). Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes*, 3(3), 461–480. <https://doi.org/10.3390/genes3030461>
- [3] Alexander, R. P., Fang, G., Rozowsky, J., Snyder, M., & Gerstein, M. B. (2010). Annotating non-coding regions of the genome. *Nature Reviews. Genetics*, 11(8), 559–571. <https://doi.org/10.1038/nrg2814>
- [4] Liao, X., Zhu, W., Zhou, J., Li, H., Xu, X., Zhang, B., & Gao, X. (2023). Repetitive DNA sequence detection and its role in the human genome. *Communications Biology*, 6(1), 954. <https://doi.org/10.1038/s42003-023-05322-y>
- [5] Liehr, T. (2021). Repetitive elements in humans. *International Journal of Molecular Sciences*, 22(4), 2072. <https://doi.org/10.3390/ijms22042072>
- [6] Sawaya, S., Bagshaw, A., Buschiazzi, E., Kumar, P., Chowdhury, S., Black, M. A., & Gemmell, N. (2013). Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PloS One*, 8(2), e54710. <https://doi.org/10.1371/journal.pone.0054710>
- [7] Richard, G. F., & Pâques, F. (2000). Mini- and microsatellite expansions: the recombination connection. *EMBO Reports*, 1(2), 122–126. <https://doi.org/10.1093/embo-reports/kvd031>
- [8] Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko, T., Šatović, E., & Plohl, M. (2015). Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, 23(3), 583–596. <https://doi.org/10.1007/s10577-015-9483-7>
- [9] Hong, Y., & Liu, N. (2025). Transposable elements in health and disease: Molecular basis and clinical implications. *Chinese Medical Journal*, 138(18), 2220–2233. <https://doi.org/10.1097/CM9.0000000000003775>
- [10] Wang, Z., & Kunze, R. (2015). Transposons in Eukaryotes (Part A): Structures, Mechanisms and Applications. In *eLS* (pp. 1–13). Wiley. <https://doi.org/10.1002/9780470015902.a0026264>

- [11] Perween, S., Kumar, D., & Kumar, A. (2020). A review on transposons and its utilization as genetic tool. *International Journal of Current Microbiology and Applied Sciences*, 9(2), 1874–1884. <https://doi.org/10.20546/ijcmas.2020.902.214>
- [12] Lazarow, K., Doll, M.-L., & Kunze, R. (2013). Molecular biology of maize Ac/Ds elements: an overview. *Methods in Molecular Biology (Clifton, N.J.)*, 1057, 59–82. https://doi.org/10.1007/978-1-62703-568-2_5
- [13] Xiong, W., He, L., Lai, J., Dooner, H. K., & Du, C. (2014). HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(28), 10263–10268. <https://doi.org/10.1073/pnas.1410068111>
- [14] Kapitonov, V. V., & Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12), 4540–4545. <https://doi.org/10.1073/pnas.0600833103>
- [15] Yoder, J. A., Walsh, C. P., & Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends in Genetics: TIG*, 13(8), 335–340. [https://doi.org/10.1016/s0168-9525\(97\)01181-5](https://doi.org/10.1016/s0168-9525(97)01181-5)
- [16] Huda, A., Bowen, N. J., Conley, A. B., & Jordan, I. K. (2011). Epigenetic regulation of transposable element derived human gene promoters. *Gene*, 475(1), 39–48. <https://doi.org/10.1016/j.gene.2010.12.010>
- [17] Jansz, N. (2019). DNA methylation dynamics at transposable elements in mammals. *Essays in Biochemistry*, 63(6), 677–689. <https://doi.org/10.1042/EBC20190039>
- [18] Huda, A., Mariño-Ramírez, L., & Jordan, I. K. (2010). Epigenetic histone modifications of human transposable elements: genome defense versus exaptation. *Mobile DNA*, 1(1), 2. <https://doi.org/10.1186/1759-8753-1-2>
- [19] Ravel-Godreuil, C., Znaidi, R., Bonnifet, T., Joshi, R. L., & Fuchs, J. (2021). Transposable elements as new players in neurodegenerative diseases. *FEBS Letters*, 595(22), 2733–2755. <https://doi.org/10.1002/1873-3468.14205>
- [20] De Cecco, M., Ito, T., Petrashen, A. P., Elias, A. E., Skvir, N. J., Criscione, S. W., Caligiana, A., Broccoli, 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., ... Sedivy, J. M. (2019). L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature*, 566(7742), 73–78. <https://doi.org/10.1038/s41586-018-0784-9>
- [21] Frost, B., Hemberg, M., Lewis, J., & Feany, M. B. (2014). Tau promotes neurodegeneration through global chromatin relaxation. *Nature Neuroscience*, 17(3), 357–366. <https://doi.org/10.1038/nn.3639>
- [22] Philip, V. M., Anderson, D., Ramirez, P., Frost, B., & Kaczorowski, C. C. (2021). Disentangling the determinants of transposable element expression in cognitive aging and Alzheimer's disease deficits using the AD-BXD mouse population. *Alzheimer's & Dementia*:

The Journal of the Alzheimer's Association, 17 Suppl 3, e053048.
<https://doi.org/10.1002/alz.053048>

- [23] Villeponteau, B. (1997). The heterochromatin loss model of aging. *Experimental Gerontology*, 32(4–5), 383–394. [https://doi.org/10.1016/s0531-5565\(96\)00155-6](https://doi.org/10.1016/s0531-5565(96)00155-6)
- [24] Wood, J. G., Hillenmeyer, S., Lawrence, C., Chang, C., Hosier, S., Lightfoot, W., Mukherjee, E., Jiang, N., Schorl, C., Brodsky, A. S., Neretti, N., & Helfand, S. L. (2010). Chromatin remodeling in the aging genome of *Drosophila*: Chromatin remodeling in aging. *Aging Cell*, 9(6), 971–978. <https://doi.org/10.1111/j.1474-9726.2010.00624.x>
- [25] Simon, M., Van Meter, M., Ablaeva, J., Ke, Z., Gonzalez, R. S., Taguchi, T., De Cecco, M., Leonova, K. I., Kogan, V., Helfand, S. L., Neretti, N., Roichman, A., Cohen, H. Y., Meer, M. V., Gladyshev, V. N., Antoch, M. P., Gudkov, A. V., Sedivy, J. M., Seluanov, A., & Gorbunova, V. (2019). LINE1 derepression in aged wild-type and SIRT6-deficient mice drives inflammation. *Cell Metabolism*, 29(4), 871–885.e5. <https://doi.org/10.1016/j.cmet.2019.02.014>
- [26] Guo, C., Jeong, H.-H., Hsieh, Y.-C., Klein, H.-U., Bennett, D. A., De Jager, P. L., Liu, Z., & Shulman, J. M. (2018). Tau activates transposable elements in Alzheimer's disease. *Cell Reports*, 23(10), 2874–2880. <https://doi.org/10.1016/j.celrep.2018.05.004>
- [27] Kõks, S., Pfaff, A. L., Singleton, L. M., Bubb, V. J., & Quinn, J. P. (2022). Non-reference genome transposable elements (TEs) have a significant impact on the progression of the Parkinson's disease. *Experimental Biology and Medicine (Maywood, N.J.)*, 247(18), 1680–1690. <https://doi.org/10.1177/15353702221117147>
- [28] Casale, A. M., Liguori, F., Ansaloni, F., Cappucci, U., Finaurini, S., Spirito, G., Persichetti, F., Sanges, R., Gustincich, S., & Piacentini, L. (2022). Transposable element activation promotes neurodegeneration in a *Drosophila* model of Huntington's disease. *iScience*, 25(1), 103702. <https://doi.org/10.1016/j.isci.2021.103702>
- [29] *Beyond “junk” DNA: Targeting transposable elements in ALS*. (2025, July 21). Target ALS. <https://targetals.org/news/beyond-junk-dna-targeting-transposable-elements-in-als/>
- [30] Li, Wanhe, Prazak, L., Chatterjee, N., Grüninger, S., Krug, L., Theodorou, D., & Dubnau, J. (2013). Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nature Neuroscience*, 16(5), 529–531. <https://doi.org/10.1038/nn.3368>
- [31] Li, Wenxue, Lee, M.-H., Henderson, L., Tyagi, R., Bachani, M., Steiner, J., Campanac, E., Hoffman, D. A., von Geldern, G., Johnson, K., Maric, D., Morris, H. D., Lentz, M., Pak, K., Mammen, A., Ostrow, L., Rothstein, J., & Nath, A. (2015). Human endogenous retrovirus-K contributes to motor neuron disease. *Science Translational Medicine*, 7(307), 307ra153. <https://doi.org/10.1126/scitranslmed.aac8201>
- [32] Pouchelon, G., Gambino, F., Bellone, C., Telley, L., Vitali, I., Lüscher, C., Holtmaat, A., & Jabaudon, D. (2014). Modality-specific thalamocortical inputs instruct the identity of postsynaptic L4 neurons. *Nature*, 511(7510), 471–474. <https://doi.org/10.1038/nature13390>

- [33] Chuong, 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>
- [34] Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). Tetrascripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics (Oxford, England)*, 31(22), 3593–3599. <https://doi.org/10.1093/bioinformatics/btv422>
- [35] Yang, W. R., Ardeljan, D., Pacyna, C. N., Payer, L. M., & Burns, K. H. (2019). SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Research*, 47(5), e27. <https://doi.org/10.1093/nar/gky1301>
- [36] Bendall, M. L., de Mulder, M., Iñiguez, L. P., Lecanda-Sánchez, A., Pérez-Losada, M., Ostrowski, M. A., Jones, R. B., Mulder, L. C. F., Reyes-Terán, G., Crandall, K. A., Ormsby, C. E., & Nixon, D. F. (2019). Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Computational Biology*, 15(9), e1006453. <https://doi.org/10.1371/journal.pcbi.1006453>
- [37] Goerner-Potvin, P., & Bourque, G. (2018). Computational tools to unmask transposable elements. *Nature Reviews. Genetics*, 19(11), 688–704. <https://doi.org/10.1038/s41576-018-0050-x>
- [38] Scheckel, C., Drapeau, E., Frias, M. A., Park, C. Y., Fak, J., Zucker-Scharff, I., Kou, Y., Haroutunian, V., Ma'ayan, A., Buxbaum, J. D., & Darnell, R. B. (2016). Regulatory consequences of neuronal ELAV-like protein binding to coding and non-coding RNAs in human brain. *eLife*, 5, e10421. <https://doi.org/10.7554/eLife.10421>
- [39] Chowdhury, H. A., Bhattacharyya, D. K., & Kalita, J. K. (2020). Differential expression analysis of RNA-seq reads: Overview, taxonomy, and tools. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(2), 566–586. <https://doi.org/10.1109/TCBB.2018.2873010>
- [40] Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data. (n.d.). Babraham.ac.uk. Retrieved May 11, 2026, from <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- [41] Krueger, F. (2019). Trim Galore: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files, with some extra functionality for MspI-digested RRBS-type (reduced representation bisulfite-seq) libraries. *www.Bioinformatics. Babraham. Ac. Uk/Projects/Trim_galore/*.
- [42] Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- [43] Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>

- [44] Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- [45] Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- [46] Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass.))*, 2(3), 100141. <https://doi.org/10.1016/j.xinn.2021.100141>
- [47] Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504. <https://doi.org/10.1101/gr.1239303>
- [48] Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131>
- [49] Huang, Z.-H., Wang, H., Wang, D.-M., Zhao, X.-Y., Liu, W.-W., Zhong, X., He, D.-M., Mu, B.-R., & Lu, M.-H. (2022). Identification of core genes in prefrontal cortex and hippocampus of Alzheimer's disease based on mRNA-miRNA network. *Journal of Cellular and Molecular Medicine*, 26(23), 5779–5793. <https://doi.org/10.1111/jcmm.17593>
- [50] Singer, T., McConnell, M. J., Marchetto, M. C. N., Coufal, N. G., & Gage, F. H. (2010). LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends in Neurosciences*, 33(8), 345–354. <https://doi.org/10.1016/j.tins.2010.04.001>
- [51] Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., Morell, M., O'Shea, K. S., Moran, J. V., & Gage, F. H. (2009). L1 retrotransposition in human neural progenitor cells. *Nature*, 460(7259), 1127–1131. <https://doi.org/10.1038/nature08248>
- [52] Roy, N., Haq, I., Ngo, J. C., Bennett, D. A., Teich, A. F., De Jager, P. L., Olah, M., & Sher, F. (2024). Elevated expression of the retrotransposon LINE-1 drives Alzheimer's disease-associated microglial dysfunction. *Acta Neuropathologica*, 148(1), 75. <https://doi.org/10.1007/s00401-024-02835-6>
- [53] Kozubek, P., Kuźniar, J., Czaja, M., Sitka, H., Kochman, U., & Leszek, J. (2024). Human endogenous retroviruses and their putative role in pathogenesis of Alzheimer's disease, inflammation, and senescence. *Biomedicines*, 13(1), 59. <https://doi.org/10.3390/biomedicines13010059>

- [54] Deniz, Ö., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. *Nature Reviews. Genetics*, 20(7), 417–431. <https://doi.org/10.1038/s41576-019-0106-6>
- [55] Wang, M., Zhang, H., Liang, J., Huang, J., Wu, T., & Chen, N. (2025). Calcium signaling hypothesis: A non-negligible pathogenesis in Alzheimer's disease. *Journal of Advanced Research*, 77, 513–534. <https://doi.org/10.1016/j.jare.2025.01.007>
- [56] Di Benedetto, G., Iannucci, L. F., Surdo, N. C., Zanin, S., Conca, F., Grisan, F., Gerbino, A., & Lefkimmatis, K. (2021). Compartmentalized signaling in aging and neurodegeneration. *Cells (Basel, Switzerland)*, 10(2), 464. <https://doi.org/10.3390/cells10020464>
- [57] Hampel, H., Mesulam, M.-M., Cuello, A. C., Khachaturian, A. S., Vergallo, A., Farlow, M. R., Snyder, P. J., Giacobini, E., & Khachaturian, Z. S. (2019). Revisiting the Cholinergic Hypothesis in Alzheimer's disease: Emerging evidence from translational and clinical research. *The Journal of Prevention of Alzheimer's Disease*, 6(1), 2–15. <https://doi.org/10.14283/jpad.2018.43>
- [58] Lauretti, E., Dincer, O., & Praticò, D. (2020). Glycogen synthase kinase-3 signaling in Alzheimer's disease. *Biochimica et Biophysica Acta. Molecular Cell Research*, 1867(5), 118664. <https://doi.org/10.1016/j.bbamcr.2020.118664>
- [59] Singh, N., Nandy, S. K., Jyoti, A., Saxena, J., Sharma, A., Siddiqui, A. J., & Sharma, L. (2024). Protein kinase C (PKC) in neurological health: Implications for Alzheimer's Disease and chronic alcohol consumption. *Brain Sciences*, 14(6), 554. <https://doi.org/10.3390/brainsci14060554>
- [60] Hong, Y., & Liu, N. (2025). Transposable elements in health and disease: Molecular basis and clinical implications. *Chinese Medical Journal*, 138(18), 2220–2233. <https://doi.org/10.1097/CM9.0000000000003775>
- [61] Valdebenito-Maturana, B., Rojas-Tapia, M. I., Carrasco, M., & Tapia, J. C. (2022). Dysregulated expression of transposable elements in TDP-43M337V human motor neurons that recapitulate amyotrophic lateral sclerosis in vitro. *International Journal of Molecular Sciences*, 23(24), 16222. <https://doi.org/10.3390/ijms232416222>

APPENDICES

Appendix A

Command set for data acquisition, preprocessing and alignment

```
#Data Acquisition

prefetch SRRXXXXXXXX

#SRA to fastq

fasterq-dump SRRXXXXXXXX.sra --split-files --threads n -outdir ./

#Quality Check

fastqc SRRXXXXXXXX_1.fastq SRRXXXXXXXX_2.fastq

#Trimming

trim_galore -paired SRRXXXXXXXX_1.fastq SRRXXXXXXXX_2.fastq --cores n

#Bowtie grch38 indexing

bowtie2-build grch38.fa grch38_bt2

#Bowtie2 Alignment

bowtie2 -x <Bowtie_indexed_files> -1 SRRXXXXXXXX_1.fastq -2
SRRXXXXXXXX_2.fastq -S sample_xx.sam -p n

#sam to bam

samtools view -bS sample_xx.sam > sample_xx.bam

#bam sorting

samtools sort sample_xx.bam -o sample_xx.sorted.bam

#indexing

samtools index sample_xx.sorted.bam
```

Appendix B

Command set for TE count quantification using TE

#TE Quantification

```
TEcount -format BAM -mode multi -b sample.sorted.bam --GTF
gene_annotation.gtf --TE TE_annotation.gtf -sortByPos --outdir
TE_results/
```

Appendix C

R Script for Differential Expression Analysis using DESeq2

```
library(tidyverse)
library(DESeq2)
library(openxlsx)
library(pheatmap)
library(RColorBrewer)
# Working directory (replace with project path)
setwd("/path/to/project")
# Load merged count matrix
merged <- read.csv("merged_count_matrix.csv", row.names = 1,
check.names = FALSE)
merged <- as.matrix(merged)
storage.mode(merged) <- "integer"
# Sample information
sample_info <- data.frame(
  row.names = colnames(merged),
  condition = c(rep("Ctrl", 8), rep("AD", 8))
)
sample_info$condition <- factor(sample_info$condition, levels =
c("Ctrl", "AD"))
# DESeq2 analysis
dds <- DESeqDataSetFromMatrix(countData = merged, colData =
sample_info, design = ~ condition)
dds <- dds[rowSums(counts(dds)) > 10, ]
```

```

dds <- DESeq(dds)
res <- results(dds)
res_df <- as.data.frame(res)
res_df$feature <- rownames(res_df)
write.csv(res_df, "DESeq2_raw_results.csv", row.names = FALSE)
# Filter transposable elements
res_TE <- res_df[!grepl("^ENSG", res_df$feature), ]
res_TE <- res_TE[!is.na(res_TE$log2FoldChange) &
!is.na(res_TE$pvalue), ]
res_TE <- res_TE[res_TE$pvalue <= 0.05 & abs(res_TE$log2FoldChange)
>= 1, ]
res_TE <- res_TE[!grepl("Satellite|satellite", res_TE$feature), ]
res_TE <- res_TE[order(res_TE$pvalue), ]
write.xlsx(res_TE, file = "TE_filtered_results.xlsx", rowNames =
FALSE)
# Select top transposable elements
top_n <- 30
selected_TEs <- head(res_TE$feature, top_n)
# Variance stabilizing transformation
vsd <- vst(dds, blind = FALSE)
data_norm <- assay(vsd)
data_scaled <- t(scale(t(data_norm)))
data_scaled <- data_scaled[complete.cases(data_scaled), ]
selected_TEs <- intersect(selected_TEs, rownames(data_scaled))
heatmap_matrix <- data_scaled[selected_TEs, ]
# Sample annotation
annotation_col <- data.frame(Condition = sample_info$condition)
rownames(annotation_col) <- rownames(sample_info)
annotation_colors <- list(Condition = c("AD" = "#8E7DBE", "Ctrl" =
"#F4A261"))
# Heatmap
heat_colors <- colorRampPalette(c("#3A7FBF", "#FAD4D4",
"#B22222"))(100)

```

```
breaks <- seq(-2, 2, length.out = 101)
png("TE_heatmap.png", width = 4000, height = 6000, res = 300)
pheatmap(
  heatmap_matrix,
  color = heat_colors,
  breaks = breaks,
  cluster_rows = TRUE,
  cluster_cols = TRUE,
  annotation_col = annotation_col,
  annotation_colors = annotation_colors,
  fontsize = 12,
  fontsize_row = 10,
  fontsize_col = 10,
  cellwidth = 35,
  cellheight = 20,
  show_colnames = FALSE,
  show_rownames = TRUE,
  main = "Top Differentially Expressed TEs"
)
dev.off()
```

Appendix D

R Script for KEGG Pathway Enrichment Analysis

```
library(clusterProfiler)
library(org.Hs.eg.db)
library(pathview)
library(ggplot2)
# Working directory (replace with project path)
setwd("/path/to/project")
# Load gene list
```

```

genes <- read.table("gene_list.txt", stringsAsFactors = FALSE)$V1
# Convert gene symbols to Entrez IDs
gene_df <- bitr(
  genes,
  fromType = "SYMBOL",
  toType = "ENTREZID",
  OrgDb = org.Hs.eg.db
)
gene_df <- na.omit(gene_df)
# KEGG enrichment analysis
kegg <- enrichKEGG(
  gene = gene_df$ENTREZID,
  organism = "hsa",
  pvalueCutoff = 0.05
)
kegg_results <- as.data.frame(kegg)
gene_map <- setNames(gene_df$SYMBOL, gene_df$ENTREZID)
kegg_results$Genes <- sapply(
  kegg_results$geneID,
  function(x) paste(gene_map[strsplit(x, "/")[[1]]], collapse = ",
")
)
write.csv(kegg_results, "KEGG_enrichment_results.csv", row.names =
FALSE)

```

Appendix E

R Script for Correlation Analysis

```

library(tidyverse)
library(readxl)
# Working directory (replace with project path)
setwd("/path/to/project")

```

```

# Load expression data

# Input file must have features in rows and conditions in columns,
# with an additional column named 'Type' indicating whether each
# feature is a TE or Gene.

expr <- read_excel("Final_Correlation_data.xlsx")
expr <- as.data.frame(expr)

# Separate TE and Gene data
TE_data <- expr[expr$Type == "TE", ]
Gene_data <- expr[expr$Type == "Gene", ]

# Expression columns (excluding metadata)
expr_cols <- 4:ncol(expr)

# Initialize results dataframe
cor_results <- data.frame()

# Compute Pearson correlations between each TE and Gene
for (i in 1:nrow(TE_data)) {
  for (j in 1:nrow(Gene_data)) {
    te_expr <- as.numeric(TE_data[i, expr_cols])
    gene_expr <- as.numeric(Gene_data[j, expr_cols])
    cor_test <- cor.test(te_expr, gene_expr, method = "pearson")
    cor_results <- rbind(
      cor_results,
      data.frame(
        TE = TE_data$Symbol[i],
        Gene = Gene_data$Symbol[j],
        Correlation = cor_test$estimate,
        Pvalue = cor_test$p.value
      )
    )
  }
}

# Adjust p-values using Benjamini-Hochberg correction
cor_results$FDR <- p.adjust(cor_results$Pvalue, method = "BH")

```

```
# Export all correlation results
write.csv(cor_results, "All_TE_Gene_Correlations.csv", row.names =
FALSE)

# Filter significant correlations
sig_correlations <- cor_results[
  abs(cor_results$Correlation) > 0.7 & cor_results$Pvalue < 0.05,
]

# Export significant correlations
write.csv(sig_correlations, "Significant_TE_Gene_Correlations.csv",
row.names = FALSE)
```

LIST OF PUBLICATIONS AND THEIR PROOF

1. Poster Presentation in SNCI (Society for Neurochemistry, India) 2026, Delhi Chapter at Jamia Hamdard University on “Integrative transcriptomic analysis reveals the role of Transposable Elements in Alzheimer's Disease through Differential Expression and Pathway Enrichment





DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

PLAGIARISM VERIFICATION

Title of the Thesis **“Integrated Transcriptomic Analysis Reveals Potential Transposable Element Association in Alzheimer’s Disease”**

Total Pages **71**

Name of the student: **Ananya Singh**

Supervisor: **Prof. Pravir Kumar**

Department of Biotechnology, Delhi Technological University

This is to report that the above thesis was scanned for similarity detection. Process and outcome is given below:

Software used: **Turnitin**, Similarity Index: **5%**, Total Word Count: **16341**

Date: **29 - 05 - 2026**

Ananya Singh

Candidate's Signature

Ph
14/05/2026

Signature of Supervisor

Pravir Kumar

Ananya_24BIO05

Nishita Singh

Document Details

Submission ID
trnoid::27535:138788693

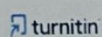
Submission Date
May 13, 2026, 1:39 PM GMT+5:30

Download Date
May 13, 2026, 1:45 PM GMT+5:30

File Name
Ananya_24BIO05.docx

File Size
795.5 KB

37 Pages
12,131 Words
75,434 Characters



Page 2 of 42 - Integrity Overview

Submission ID trnoid::27535:138788693

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text
- Small Matches (less than 10 words)

Match Groups

- 52 Not Cited or Quoted 5%
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%
Matches that are still very similar to source material
- 0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

- 3% Internet sources
- 3% Publications
- 2% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Ananya Singh

Ph
14/05/2026