https://doi.org/10.48047/AFJBS.6.12.2024.4902-4913

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Prevalence of Mobile Genetic Elements in Cancer-Associated Genes

P.Preeti[1], Chhavi Dudeja[2], Priya Kumari[3], Abhishek Singh[4], Kamal Rawal[5], Tapan Kumar Baral[6], Mihir Nakul[7]

[1]*Centre for Computational Biology and Bioinformatics, AIBAmity University, Noida, India. p.preeti8227@gmail.com*

[2]*Centre for Computational Biology and Bioinformatics, AIBAmity University, Noida, India. chhavidudeja2002@gmail.com*

[3]*Centre for Computational Biology and Bioinformatics, AIBAmity University, Noida, India. priyasinghchas@gmail.com*

[4]*Centre for Computational Biology and Bioinformatics, AIBAmity University, Noida, India. abhisheksingh223amity@gmail.com*

[5]*Centre for Computational Biology and Bioinformatics, AIBAmity University, Noida, India. k.rawal@amity.edu*

[6]*Centre for Computational Biology and Bioinformatics, AIBAmity University,Noida, India tbaral666@gmail.com*

[7]*Centre for Computational Biology and Bioinformatics, AIBAmity University, Noida, India. nakul.mihir@gmail.com*

Article History
Volume 6, Issue 12, 2024
Received: June 10, 2024
Accepted: July 5, 2024
doi:
10.48047/AFJBS.6.12.2024.4902-4913

Abstract-
Mobile genetic elements (MGEs) represent a significant portion of the human genome, contributing to evolutionary processes through their mobility. This mobility, however, also introduces mutations. Some mutations induce exon skipping or premature truncation, resulting in non-functional or impaired proteins that contribute to various human diseases, including cancer. This study aims to explore MGE prevalence within cancer-associated genes in comparison to non-cancer-associated genes. We compiled gene lists for both categories, retrieved nucleotide sequences, and employed RepeatMasker analysis for MGE identification. Our analysis revealed a notable difference: cancer-associated genes exhibit higher MGE counts and greater sequence coverage by MGEs compared to non-cancer-associated genes. This heightened MGE presence may correlate with an increased susceptibility to significant DNA segment deletions/insertions, potentially elevating the risk of cancer initiation. Further investigation is essential to unveil the precise nature of these associations and their implications comprehensively.

Keywords: Mobile Genetic Elements, Transposable Element, RepeatMasker, Cancer Associated Genes, retrotransposon, SINE, LINE, Alu.

1. Introduction

Mobile genetic elements (MGEs) are the DNA segments that can change their location within the genome. MGEs are commonly called Transposable Elements (TEs) or jumping genes due to their ability to move. MGEs are powerful enough to create genetic innovation or evolution as they can create gene modifications due to mutation [1], [2]. They are present in the genomes of prokaryotes as well as eukaryotes.In humans, they occupy a large fraction (>50 %) of the genome [3]. MGEs can be broadly classified into retrotransposons class (Class I) and DNA transposons (Class II), based on their mode of transmission. Out of two classes, the retrotransposon is widely spread MGEs in mammals. Retrotransposons can be further categorized into Long Terminal Repeats (LTRs) and Non-Long Terminal Repeats (non-LTRs) [4]. Autonomous non-long terminal repeat retrotransposons are commonly referred to as long interspersed nuclear elements (LINEs). Short non-autonomous elements that borrow the LINE machinery are called short interspersed nuclear elements (SINES) [5].

The distinguishing characteristic of cancer cells is their dedifferentiation, which results from genetic and epigenetic modifications. Furthermore, the epigenetic changes that occur during dedifferentiation may promote increased expression of developmental TE, leading to even more dedifferentiation [6]. In 2019, Jang and colleagues examined how TEs might contribute to the activation of cancer associated genes in human cancers. Their research showed that TEs could impact cancer associated gene expression in approximately 50% of tumours [7].

Numerous types of cancer have been extensively researched, revealing a lack of effective surveillance mechanisms for TE repressions, such as DNA methylation and histone modification [4]. One example of this is the occurrence of LINE-1 promoter hypomethylation and subsequent retrotransposition across various malignancies [8]. Such transposition has been specifically observed in ovarian cancer patients[9], breast cancer [10], non-small cell lung cancer [11], hepatocellular carcinoma [12], and colon cancer [13], and has been correlated with unfavourable outcomes.

TEs are recognized instigators of genome instability within cancer cells due to their insertion into the genome and subsequent reintegration, along with homologous recombination between TEs. A notable instance is the insertion of LINE-1 into the adenomatous polyposis coli (APC) gene, a pivotal tumour suppressor in colon cancer [14]. Notably, colon cancers display susceptibility to somatic LINE-1 activity, harbouring up to 106 LINE-1 copies in individual cases [15], [16]. LINE-1 somatic insertion is also observed across other cancer types such as hepatocellular carcinoma [17], gastric and small intestine tumours [18], head and neck cancers [19], non-small cell lung cancer [14], and adenocarcinomas [20].

Furthermore, investigations have indicated a link between the expression of other TE families like SVA, SINE, and HERV families and reduced survival rates among patients with colorectal cancer [21], [22].

According to the studies conducted by Nguyen et al. and Sin et al., TE mobilization leads to a reorganization of sequences, changes in gene structure, and modifications in gene expression [23], [24]. This process is responsible for the formation of structural variations that contribute to individual differences and can initiate certain serious diseases, such as CHARGE syndrome, neurofibromatosis, colon cancer, breast cancer, and microdeletion or microduplication disorders [25], [26], [27], [28]. Though the majority of the MGEs are inactive in the human genome, except non-LTRs (like LINEs and SINEs) are capable of generating mutations. These active MGEs have the potential to disrupt genes and play a role in chromosomal rearrangements [4]. In humans, MGE-generated mutations found to be associated with diseases such as Mankes syndrome, X-linked agammaglobulinemia, Dent disease, haemophilia, cholinesterase deficiency, hereditary nonpolyposis colorectal cancer, familial

adenomatous polyposis, leukaemia, and cystic fibrosis [29]. In a study performed by Wimmer K. et al., on neurofibromatosis type 1, 18 unrelated patients were identified with MGE insertions in the NF1 gene in the exonic or intronic region [30]. Studies have also shown that the insertion of mobile DNA in genes might promote exon skipping, resulting in either reduced or no expression of the affected gene(s) [31], [32], [33]. In 1988, Kazazian H. H. et al. discovered that MGE insertion was responsible for the causation of haemophilia A [34]. Large-scale genomic projects, such as the Functional Annotation of Mouse (FANTOM) and ENCyclopedia Of DNA Elements (ENCODE) projects, as well as independent studies, have indicated that MGEs exhibit pronounced activity that is intricately tailored to specific cell types. They intricately regulate not only their own cell-specific transcription but also exert control over the transcription of adjacent genes [35], [36], [37].Newer studies indicate that changes in the regulation of TE and the resulting cellular and immune response to this disruption are linked to several human illnesses, such as age-related conditions and cancer [1], [21], [38], [39].

Previously, our group has developed an integrated system for the analysis and de novo detection of MGEs. ELAN is an efficient suite that detects different MGEs and their distribution on the genome. It was observed that signals present in exonic regions are different from the intronic regions of the genes. All those studies indicated the important role of MGE insertions in human genes[40]. Understanding TEs in human genes is important because they have the potential to disrupt gene function and regulation, leading to genetic diseases. Studying TEs can help improve our understanding of disease mechanisms and aid in the development of new diagnostic and therapeutic approaches targeting TE-related diseases. The objective of this study is to investigate the distribution of MGEs (including the number of MGEs and sequence covered by MGEs) in cancer associated genes which have the potential to cause cancer. The experiment will help us in understanding the distribution of MGEs on cancer associated genes compared to non-cancer genes. This study may shed light on the role of MGEs in the development of cancer and provide insights into the molecular mechanisms underlying cancer pathogenesis and thus may help in the identification of potential therapeutic targets for cancer treatment.

2. Material and Method

A flowchart is given in Fig. 1, summarizing the workflow to achieve the goal.

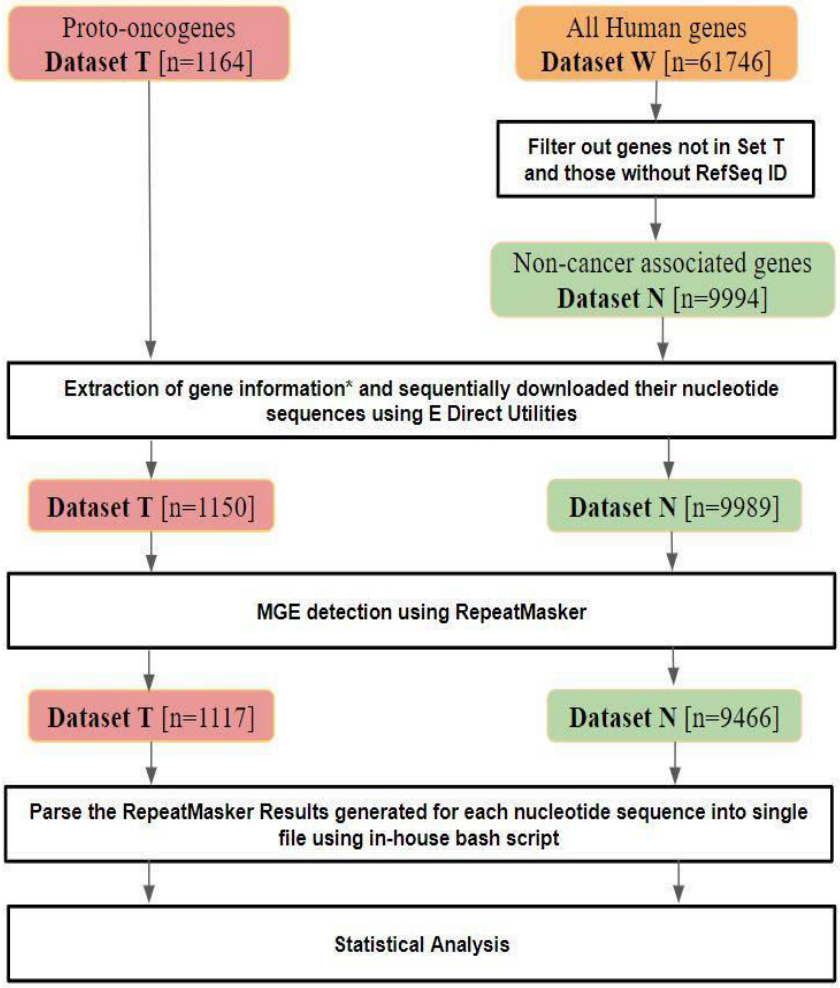

Figure 1: Flow Diagram of the Process of Gene Extraction and MGE Detection in Dataset T and Dataset N

A. Preparation of Gene Dataset

The cancer associated genes were collected and curated from publicly available databases namely OMIM [41], Genetics-MedlinePlus [42] and Cancer Genetics Web [43]. There are 1,164 cancer associated genes collected and this is called the positive dataset (Dataset T). Further, We extracted a complete list of genes from homo sapiens from the NCBI database (Dataset W) [44]. Now, to obtain the negative dataset we compared the dataset T with dataset W and filtered out the non-matching genes thus resulting in a non-cancerous associated genes dataset (Dataset N).

B. Sequence Retrieval of Positive and Negative Dataset

We have used Entrez direct (E-direct) [45] utility which provides access to NCBI allowing the user to get detailed information on that particular gene. Since the E-direct takes the input as a single gene at a time so we developed a bash script to download the nucleotide sequence collectively provided in the gene list. The script is helpful in retrieving all the necessary gene details like NCBI chromosome ID, start and end position, sense/antisense information etc.

C. Detection of Mobile Genetic Elements

To identify and characterize the MGEs in each nucleotide sequence we have usedRepeatMasker version 4.1.2 [46]. RepeatMasker is a widely used open-source application written in Perl language, used for the identification of MGEs in genomic sequences[47], [48], [49], [50]. RepeatMasker identify all the repeat region in a given nucleotide sequence and characterise these repeats into different categories, including, SINEs (subtypes - ALU and

MIRs), LINEs (subtypes - LINE1, L3/CR1 and LINE2), LTRs (subtypes of LTR - ERVL, ERV_ClassI, ERVL-MaLRs and ERV_ClassII), DNA elements (subtypes of DNA elements - TcMar-Tigger and hAT-Charlie), terminal inverted repeats, unclassified repeats, Satellites, Small RNA, simple repeats, and low complexity. The execution of RepeatMasker broadly involves the following steps: i) evaluating the input sequence; ii) checking of RepeatMasker library or user-defined TE library; iii) splitting the sequence into fragments; iv) performing alignment on the fragments; and v) merging of identified TEs in the fragments. RepeatMasker supports four search engine software, AB-BLAST, RMBLAST, CROSS_MATCH and HMMER. Since the RepeatMasker RMBLAST is optimized for TE search, we employed it for our analytical investigation[51]. We used the default parameters to run RepeatMasker. Since we required the MGE results of each gene individually, we developed a BASH script which runs on the Linux platform via the command line interface. This script takes a directory as input that has all the nucleotide sequences in fasta format and an output directory where the output files will be saved. Then each nucleotide sequence saved in the input directory is sequentially run. The resultant files of each run are directed to save in an output folder. Further, we parsed RepeatMasker output files of each sequence of the respective dataset to further perform statistical analysis.

D. MGE Distribution in Well-known Cancer/Non-cancer Associated Genes

We conducted a batch comparative analysis of widely recognized genes categorized as cancer associated genes and non-cancer associated genes. Among the non-cancer associated genes, we identified three well-known genes that have no association with cancer: *SLC26A5* (prestin protein, found in hair cells of the inner ear), *PRPH2* (peripherin 2 protein, found in light-sensitive cells of the retina), and *CRYZ* (quinone reductase protein, found in the lens of the eye)[52]. As for the cancer associated genes, some commonly known ones include *BRAF*, *APC*, *ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *MSH2*, *PMS2* and *TP53* [53], [54].

E. Statistical Analysis

We computed standard statistical measures such as mean and standard deviations for the data obtained from Repeatmasker. Next, we compared these values in different datasets (namely T and N) using the independent T-test (*P* values <0.05).

3. Results

A. Data Collection

Under Dataset T, we curated a list of 1,164 genes from different databases that have been previously reported to be associated with cancer (**see Supplementary Table 1**). We then extracted the complete list of human genes, consisting of 61,746 genes in Dataset W (**see Supplementary Table 2**). To obtain Dataset N, a list of genes not associated with cancer, we removed all genes present in Dataset T from Dataset W. Further we removed genes without RefSeq IDs from Dataset N. The resulting dataset N, comprised 9,994 genes (**see Supplementary Table 3**). This dataset represents a comprehensive list of non-cancer-associated genes and was used for further analysis.

B. Nucleotide Sequence Retrieval

Out of the 1,164 genes in dataset T, we were able to extract the nucleotide sequences for 1,150 genes. The unavailability of nucleotide sequences for a few genes or the use of an alias gene name in E-direct may have contributed to the failure to obtain nucleotide sequences for the remaining 14 genes. Similarly, we were able to download 9,989 nucleotide sequences out of the 9,994 genes present in dataset N. Nucleotide sequences in datasets T and N were further used for the identification of mobile genetic elements.

C. MGE Detection

To identify and characterize MGEs in datasets T and N, all extracted nucleotide sequences were subjected to RepeatMasker analysis. The RMBLAST algorithm was employed to process all the genes. Of the 1,150 genes in dataset T, MGEs were detected in 1,117 genes, given in

Supplementary Table 4. However, RepeatMasker did not detect any MGEs in the remaining 33 sequences. Similarly, in dataset N, MGEs were identified in 9,466 out of 9,989 total sequences, presented in **Supplementary Table 5**.

We compared the occurrence of MGEs in both datasets. The findings revealed a higher prevalence of MGEs in cancer-associated genes compared to non-cancer-associated genes. In dataset T, the average length of the total interspersed repeats was 41,920.68 bp, which was significantly higher than the 32,603.68 bp observed in dataset N. Nex, we analyzed different classes of MGEs, including SINEs, LINEs, LTRs, and DNA elements, in both datasets. The number of each MGE type was significantly higher in dataset T compared to dataset N, as depicted in **Fig. 2.** Additionally, the sequence coverage by these MGEs was also notably greater, as shown in **Table 1**.

In our investigation, we delved into subclasses of SINEs, LINEs, and LTRs within datasets T and N. The results demonstrated a significant difference, indicating a higher abundance of the number of MGEs and the sequence region covered by MGEs in cancer associated genes compared to non-cancer associated genes, as depicted in **Supplementary Fig. 1-4** and **Table 1**.

D. MGE Distribution in Well-known Cancer/Non-cancer Associated Genes

Upon comparison of a few well-known cancer associated genes (*BRAF*, *APC*, *ATM*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *MSH2*, *PMS2* and *TP53*) and non-cancer associated genes (*SLC26A5*, *PRPH2* and *CRYZ*), we found that the sequence covered by MGEs (all subclass combined) are comparatively higher in cancer associated genes (**Fig. 3**).

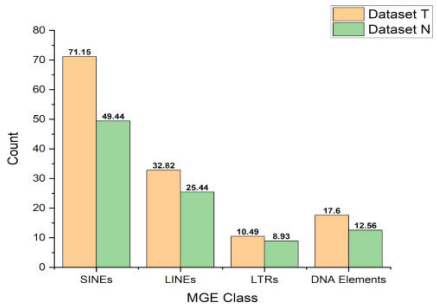

Figure 2. Distribution of the Average Count of Various MGE Classes in Genes Derived from Datasets T and N.

Table 1.The Average Sequence Length (In Base Pairs) Covered by Different MGE Classes in Genes from Datasets T and N

MGE CLASS/DATASETS	Dataset T	Dataset N
Total Sequence Length	1,07,552.28	76,954.21
SINEs	16,571.53	11,385.79
ALUs	13,338.29	8,915.54
MIRs	3,181.14	2,439.51
LINEs	16,031.46	13,827.02
LINE1	11,793.91	10,384.24
LINE2	3,643.37	2,983.38
L3/CR1	441.25	334.02
LTR	4,902.54	4,263.75
ERVL	1,021.04	880.55
ERVL-MaLRs	2,300.86	1,963.76
ERV_classI	1,298.28	1,174.58
ERV_classII	131.87	118.28

DNA_elements	4,137.08	2,932.42
hAT-Charlie	1,928.65	1,346.97
TcMar-Tigger	1,516.05	1,061.30

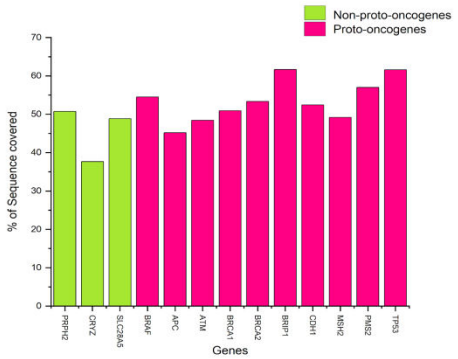

Figure 3. Distribution of the Percentage of Sequence Covered by MGEs in Non-cancer Associated Genes and Cancer Associated Genes

4. Discussion

Mobile genetic elements (MGEs) are prevalent across diverse life forms, spanning bacteria, archaea, and eukaryotes. They often bear the label "selfish genetic elements" due to their inherent lack of advantage to their host organism. In reality, their presence can occasionally yield detrimental effects on the host [55]. Numerous studies reported instances where MGE-induced insertions or deletions contributed to human diseases. For instance, Brooks et al. identified an insertion of LINE1 within the factor IX gene, resulting in hemophilia B in an affected individual [56]. Similarly, Ishihara et al. detailed a 382 bp insertion of the AluY element (a subclass of SINE) in the ZFHX1B gene, associated with Mowat-Wilson syndrome [57]. Another study by Taşkesen et al. highlighted a 333 bp AluY insertion in the coding region of the ALMS1 gene, causative of Alströmsyndrome[58]. Notably, certain human cancers have also been ascribed to MGE-triggered insertions or deletions within cancer associated genes. Wang et al. identified a deletion/insertion mutation within the BRCA2 gene, linked to breast and/or ovarian cancer. This event led to a frameshift and untimely truncation of the BRCA2 protein[59]. Furthermore, van der Klift et al. reported a LINE-1-mediated retrotranspositional insertion within the PMS2 gene, identified in a patient diagnosed with Lynch syndrome [60]. To date, several such studies have been published indicating the role of MGEs in cancer development [61], [62], [63], [64], [65].

Active mobile genetic elements (MGEs) frequently underlie insertion and duplication mutations within the genome. A notable investigation documented 810 somatic MGE insertions across diverse cancer types-such as lung squamous, head and neck, colorectal, and endometrial carcinomas. The researchers postulated a connection between elevated somatic retrotransposition rates in tumours and heightened occurrences of genomic rearrangements and somatic mutations [19].

The focus of this study was to assess the prevalence of MGEs within genes associated with cancer and those unrelated to cancer. Through rigorous comparative analysis, a notable disparity was unveiled. Specifically, within cancer associated genes, genes that exhibit a propensity to contribute to cancer development, the presence of MGEs was strikingly more pronounced when contrasted with their presence within non-cancer associated genes. Moreover, MGEs exhibit a relatively expanded coverage of sequence regions within cancer associated genes. This heightened coverage suggests a potentially pivotal role of MGEs in sculpting the functional and regulatory aspects of cancer associated genes, possibly implicating them in various stages of cancer initiation and progression. These findings

underscore the intricate interplay between MGEs and cancer development, shedding light on their potential contribution to the complex landscape of cancer-associated genes.

SINEs and LINEs constitute a significant portion of many genomes, including human genomes. In humans, Alu elements are the most common SINEs and LINE-1 (L1) elements are the most well-known and active LINEs. Research has shown that active Alu and L1 elements can contribute to somatic genomic alterations in cancer cells, promoting tumorigenesis[19], [61], [62], [64], [66]. While comparing the occurrence of MGEs between cancer associated genes and non-cancer associated genes we identified that there is a significant abundance of Alu and L1 elements in cancer associated genes. Altogether, the findings of this study provided a comprehensive comparison of MGE distribution between cancer and non-cancer-associated genes.Further studies should be performed to determine the contribution of such an abundance of MGEs in cancer development.

5. Conclusion and Future Perspective

In this investigation, we undertook a comparative analysis of the distribution of mobile genetic elements (MGEs) within genes associated with cancer and those unrelated to cancer. The findings have distinctly highlighted a prevalence of MGEs within cancer-associated genes.Though the reason for this abundance is not known, it can be hypothesized that the presence of these MGEs can make these genes amenable to large deletions or duplications/insertions that could lead to pathogenic mutations in the gene, resulting in cancer. further comprehensive research is imperative to elucidate the precise role of MGEs in cancer genomics. These enigmatic elements could potentially serve as therapeutic targets, warranting further exploration. To delve into MGEs' significance within cancer genomics, we have also devised a meticulous pipeline aimed at identifying these mobile genetic elements within genomic datasets[67]. This intriguing avenue holds the promise of therapeutic manipulation, thereby warranting an in-depth investigation of its potential.

6. Contribution of Authors

This study was conducted under the overall guidance of KR.KR and PP contributed to the protocol, critical evaluation of data and manuscript. Data curation, analysis, and cross-checking were done by PP, CD, AS and MN. Data was generaonted by PP and TKB. All the statistical analysis was done by PP, TKB and KR. Manuscript writing was done by PP and PK. All the authors are responsible for the content of the manuscript.

Acknowledgement

We extend our sincere gratitude to Amity University for providing the administrative and technical support required for the conduct of this study.

Financial Support and Sponsorship

Preeti P. has received a research fellowship from SERB, Department of Science and Technology (File Number: CVD/2020/000842) [Feb-2021 to Apr-2022] and ICMR, Indian Council of Medical Research (BMI/12(66)/2021 2021-6442) [Apr-2022 to Present]. We used the computational infrastructure provided by the Department of Biotechnology (DBT), Ministry of Science and Technology Government of India (Grant ID: BT/PRI7252/BID/7/708/2016), Conquer Cancer Foundation of ASCO [2021IIG-5555960128], Pfizer Products India Pvt. Ltd. and Robert J. Kleberg, Jr. and Helen C. Kleberg Foundation, USA, Texas Children's Hospital and Baylor College of Medicine, Houston USA.

References

[1] M. E. Jönsson, R. Garza, P. A. Johansson, and J. Jakobsson, "Transposable Elements: A Common Feature of Neurodevelopmental and Neurodegenerative Disorders," Trends Genet., vol. 36, no. 8, pp. 610–623, Aug. 2020. doi: 10.1016/j.tig.2020.05.004.

[2] C. Feschotte and E. J. Pritham, "DNA Transposons and the Evolution of Eukaryotic Genomes," Annu. Rev. Genet., vol. 41, no. 1, pp. 331–368, Dec. 2007.

doi: 10.1146/annurev.genet.40.110405.090448.
[3] X. Cao et al., "Polymorphic mobile element insertions contribute to gene expression and alternative splicing in human tissues," Genome Biol., vol. 21, no. 1, p. 185, Dec. 2020. doi: 10.1186/s13059-020-02101-4.
[4] S. Anwar, W. Wulaningsih, and U. Lehmann, "Transposable Elements in Human Cancer: Causes and Consequences of Deregulation," Int. J. Mol. Sci., vol. 18, no. 5, p. 974, May 2017, doi: 10.3390/ijms18050974.
[5] A. A. Bakre, K. Rawal, R. Ramaswamy, A. Bhattacharya, and S. Bhattacharya, "The LINEs and SINEs of Entamoeba histolytica: Comparative analysis and genomic distribution," Exp. Parasitol., vol. 110, no. 3, pp. 207–213, Jul. 2005. doi: 10.1016/j.exppara.2005.02.009.
[6] C. F. Lynch-Sutherland, A. Chatterjee, P. A. Stockwell, M. R. Eccles, and E. C. Macaulay, "Reawakening the Developmental Origins of Cancer Through Transposable Elements," Front. Oncol., vol. 10, p. 468, May 2020, doi: 10.3389/fonc.2020.00468.
[7] H. S. Jang et al., "Transposable elements drive widespread expression of oncogenes in human cancers," Nat. Genet., vol. 51, no. 4, pp. 611–617, Apr. 2019, doi: 10.1038/s41588-019-0373-3.
[8] M. Imbeault, P.-Y. Helleboid, and D. Trono, "KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks," Nature, vol. 543, no. 7646, pp. 550–554, Mar. 2017, doi: 10.1038/nature21683.
[9] J. Pattamadilok et al., "LINE-1 hypomethylation level as a potential prognostic factor for epithelial ovarian cancer," Int. J. Gynecol. Cancer Off. J. Int. Gynecol. Cancer Soc., vol. 18, no. 4, pp. 711–717, 2008, doi: 10.1111/j.1525-1438.2007.01117.x.
[10] A. Q. van Hoesel et al., "Hypomethylation of LINE-1 in primary tumor has poor prognosis in young breast cancer patients: a retrospective cohort study," Breast Cancer Res. Treat., vol. 134, no. 3, pp. 1103–1114, Aug. 2012, doi: 10.1007/s10549-012-2038-0.
[11] K. Saito, K. Kawakami, I. Matsumoto, M. Oda, G. Watanabe, and T. Minamoto, "Long Interspersed Nuclear Element 1 Hypomethylation Is a Marker of Poor Prognosis in Stage IA Non–Small Cell Lung Cancer," Clin. Cancer Res., vol. 16, no. 8, pp. 2418–2426, Apr. 2010, doi: 10.1158/1078-0432.CCR-09-2819.
[12] K. Harada et al., "LINE-1 Methylation Level and Patient Prognosis in a Database of 208 Hepatocellular Carcinomas," Ann. Surg. Oncol., vol. 22, no. 4, pp. 1280–1287, Apr. 2015, doi: 10.1245/s10434-014-4134-3.
[13] S. Ogino et al., "A Cohort Study of Tumoral LINE-1 Hypomethylation and Prognosis in Colon Cancer," JNCI J. Natl. Cancer Inst., vol. 100, no. 23, pp. 1734–1738, Dec. 2008, doi: 10.1093/jnci/djn359.
[14] R. C. Iskow et al., "Natural Mutagenesis of Human Genomes by Endogenous Retrotransposons," Cell, vol. 141, no. 7, pp. 1253–1261, Jun. 2010, doi: 10.1016/j.cell.2010.05.020.
[15] S. Solyom et al., "Extensive somatic L1 retrotransposition in colorectal tumors," Genome Res., vol. 22, no. 12, pp. 2328–2338, Dec. 2012, doi: 10.1101/gr.145235.112.
[16] E. Lee et al., "Landscape of Somatic Retrotransposition in Human Cancers," Science, vol. 337, no. 6097, pp. 967–971, Aug. 2012, doi: 10.1126/science.1222077.
[17] R. Shukla et al., "Endogenous Retrotransposition Activates Oncogenic Pathways in Hepatocellular Carcinoma," Cell, vol. 153, no. 1, pp. 101–111, Mar. 2013, doi: 10.1016/j.cell.2013.02.032.
[18] N. Rodić et al., "Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma," Nat. Med., vol. 21, no. 9, pp. 1060–1064, Sep. 2015, doi: 10.1038/nm.3919.

[19] E. Helman, M. S. Lawrence, C. Stewart, C. Sougnez, G. Getz, and M. Meyerson, "Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing," Genome Res., vol. 24, no. 7, pp. 1053–1063, Jul. 2014, doi: 10.1101/gr.163659.113.

[20] T. T. Doucet-O'Hare, R. Sharma, N. Rodić, R. A. Anders, K. H. Burns, and H. H. Kazazian, "Somatically Acquired LINE-1 Insertions in Normal Esophagus Undergo Clonal Expansion in Esophageal Squamous Cell Carcinoma: HUMAN MUTATION," Hum. Mutat., vol. 37, no. 9, pp. 942–954, Sep. 2016, doi: 10.1002/humu.23027.

[21] B. Chénais, "Transposable Elements and Human Diseases: Mechanisms and Implication in the Response to Environmental Pollutants," Int. J. Mol. Sci., vol. 23, no. 5, p. 2551, Feb. 2022, doi: 10.3390/ijms23052551.

[22] X. Zhu, H. Fang, K. Gladysz, J. A. Barbour, and J. W. H. Wong, "Overexpression of transposable elements is associated with immune evasion and poor outcome in colorectal cancer," Eur. J. Cancer, vol. 157, pp. 94–107, Nov. 2021, doi: 10.1016/j.ejca.2021.08.003.

[23] D.-Q. Nguyen, C. Webber, and C. P. Ponting, "Bias of Selection on Human Copy-Number Variants," PLoS Genet., vol. 2, no. 2, p. e20, Feb. 2006, doi: 10.1371/journal.pgen.0020020.

[24] H.-S. Sin et al., "Transcriptional control of the HERV-H LTR element of the GSDML gene in human tissues and cancer cells," Arch. Virol., vol. 151, no. 10, pp. 1985–1994, Oct. 2006, doi: 10.1007/s00705-006-0764-5.

[25] A. J. Iafrate et al., "Detection of large-scale variation in the human genome," Nat. Genet., vol. 36, no. 9, pp. 949–951, Sep. 2004, doi: 10.1038/ng1416.

[26] R. Redon et al., "Global variation in copy number in the human genome," Nature, vol. 444, no. 7118, pp. 444–454, Nov. 2006, doi: 10.1038/nature05329.

[27] K. A. O'Donnell and K. H. Burns, "Mobilizing diversity: transposable element insertions in genetic variation and disease," Mob. DNA, vol. 1, no. 1, p. 21, 2010, doi: 10.1186/1759-8753-1-21.

[28] J. M. Kidd et al., "Mapping and sequencing of structural variation from eight human genomes," Nature, vol. 453, no. 7191, pp. 56–64, May 2008, doi: 10.1038/nature06862.

[29] D. C. Hancks and H. H. Kazazian, "Active human retrotransposons: variation and disease," Curr. Opin. Genet. Dev., vol. 22, no. 3, pp. 191–203, Jun. 2012, doi: 10.1016/j.gde.2012.02.006.

[30] K. Wimmer, T. Callens, A. Wernstedt, and L. Messiaen, "The NF1 Gene Contains Hotspots for L1 Endonuclease-Dependent De Novo Insertion," PLoS Genet., vol. 7, no. 11, p. e1002371, Nov. 2011, doi: 10.1371/journal.pgen.1002371.

[31] L. M. Payer et al., "Alu insertion variants alter mRNA splicing," Nucleic Acids Res., vol. 47, no. 1, pp. 421–431, Jan. 2019, doi: 10.1093/nar/gky1086.

[32] K. Nikopoulos et al., "A frequent variant in the Japanese population determines quasi-Mendelian inheritance of rare retinal ciliopathy," Nat. Commun., vol. 10, no. 1, p. 2884, Jun. 2019, doi: 10.1038/s41467-019-10746-4.

[33] Y. Jourdy et al., "Reccurrent F8 Intronic Deletion Found in Mild Hemophilia A Causes Alu Exonization," Am. J. Hum. Genet., vol. 102, no. 2, pp. 199–206, Feb. 2018. doi: 10.1016/j.ajhg.2017.12.010.

[34] H. H. Kazazian, C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis, "Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man," Nature, vol. 332, no. 6160, pp. 164–166, Mar. 1988, doi: 10.1038/332164a0.

[35] G. J. Faulkner et al., "The regulated retrotransposon transcriptome of mammalian cells," Nat. Genet., vol. 41, no. 5, pp. 563–571, May 2009, doi: 10.1038/ng.368.

[36] R. E. Thurman et al., "The accessible chromatin landscape of the human genome," Nature, vol. 489, no. 7414, pp. 75–82, Sep. 2012, doi: 10.1038/nature11232.
[37] S. Djebali et al., "Landscape of transcription in human cells," Nature, vol. 489, no. 7414, pp. 101–108, Sep. 2012, doi: 10.1038/nature11233.
[38] A. A. De Cubas et al., "DNA hypomethylation promotes transposable element expression and activation of immune signaling in renal cell cancer," JCI Insight, vol. 5, no. 11, p. e137569, Jun. 2020, doi: 10.1172/jci.insight.137569.
[39] K. H. Burns, "Transposable elements in cancer," Nat. Rev. Cancer, vol. 17, no. 7, pp. 415–424, Jul. 2017, doi: 10.1038/nrc.2017.35.
[40] K. Rawal and R. Ramaswamy, "Genome-wide analysis of mobile genetic element insertion sites," Nucleic Acids Res., vol. 39, no. 16, pp. 6864–6878, Sep. 2011, doi: 10.1093/nar/gkr337.
[41] A. Hamosh, A. F. Scott, J. Amberger, D. Valle, and V. A. McKusick, "Online Mendelian Inheritance In Man (OMIM)," Hum. Mutat., vol. 15, no. 1, pp. 57–61, Jan. 2000, doi: 10.1002/(SICI)1098-1004(200001)15:1<57::AID-HUMU12>3.0.CO;2-G.
[42] "MedlinePlus [Internet]. Bethesda (MD). Accessed: Tue Jan 04 2022. Natl. Libr. Med. US, Jun. 2020, [Online]. Available: https://medlineplus.gov/.
[43] SJ. Cotterill, "Home Page. Accessed: Tue Jan 04 2022. Available from:," Cancer Genet. Web, [Online]. Available: http://www.cancer-genetics.org/index.htm
[44] G. R. Brown et al., "Gene: a gene-centered information resource at NCBI," Nucleic Acids Res., vol. 43, no. D1, pp. D36–D42, Jan. 2015, doi: 10.1093/nar/gku1055.
[45] J. Kans, "Entrez direct: e-utilities on the UNIX command line. In Entrez programming utilities help [Internet].," Natl. Cent. Biotechnol. Inf. US, 2023.
[46] S. Tempel, "Using and Understanding RepeatMasker," in Mobile Genetic Elements, vol. 859, Y. Bigot, Ed., in Methods in Molecular Biology, vol. 859., Totowa, NJ: Humana Press, 2012, pp. 29–51. doi: 10.1007/978-1-61779-603-6_2.
[47] G.-H. Nam et al., "Identification of transposable elements fused in the exonic region of the olive flounder genome," Genes Genomics, vol. 40, no. 7, pp. 707–713, Jul. 2018, doi: 10.1007/s13258-018-0676-2.
[48] Z. Liu et al., "Genome-Wide Survey and Comparative Analysis of Long Terminal Repeat (LTR) Retrotransposon Families in Four Gossypium Species," Sci. Rep., vol. 8, no. 1, p. 9399, Jun. 2018, doi: 10.1038/s41598-018-27589-6.
[49] X. Guo et al., "Chloranthus genome provides insights into the early diversification of angiosperms," Nat. Commun., vol. 12, no. 1, p. 6930, Nov. 2021, doi: 10.1038/s41467-021-26922-4.
[50] J.-T. Li et al., "Parallel subgenome structure and divergent expression evolution of allo-tetraploid common carp and goldfish," Nat. Genet., vol. 53, no. 10, pp. 1493–1503, Oct. 2021, doi: 10.1038/s41588-021-00933-9.
[51] T. J. Wheeler et al., "Dfam: a database of repetitive DNA based on profile hidden Markov models," Nucleic Acids Res., vol. 41, no. D1, pp. D70–D82, Nov. 2012, doi: 10.1093/nar/gks1265.
[52] Lowe D. Name of Gene Isn't Associated with Cancer. Accessed: Tue Aug 01 2023. Science, Oct. 2021, [Online]. Available: https://tinyurl.com/non-cancer-genes.
[53] H. B. Erdem and T. Bahsi, "Spectrum of germline cancer susceptibility gene mutations in Turkish colorectal cancer patients: a single center study," Turk. J. Med. Sci., vol. 50, no. 4, pp. 1015–1021, Jun. 2020, doi: 10.3906/sag-2002-46.
[54] L. Valle et al., "Update on genetic predisposition to colorectal cancer and polyposis," Mol. Aspects Med., vol. 69, pp. 10–26, Oct. 2019, doi: 10.1016/j.mam.2019.03.001.

[55] J. H. Werren, “Selfish genetic elements, genetic conflict, and evolutionary innovation,” Proc. Natl. Acad. Sci., vol. 108, no. supplement_2, pp. 10863–10870, Jun. 2011, doi: 10.1073/pnas.1102343108.
[56] M. B. Brooks, W. Gu, J. L. Barnas, J. Ray, and K. Ray, “A Line 1 insertion in the Factor IX gene segregates with mild hemophilia B in dogs,” Mamm. Genome, vol. 14, no. 11, pp. 788–795, Nov. 2003, doi: 10.1007/s00335-003-2290-z.
[57] N. Ishihara, “Clinical and molecular analysis of Mowat-Wilson syndrome associated with ZFHX1B mutations and deletions at 2q22-q24.1,” J. Med. Genet., vol. 41, no. 5, pp. 387–393, May 2004.
doi: 10.1136/jmg.2003.016154.
[58] M. Taşkesen et al., “Novel Alu retrotransposon insertion leading to Alström syndrome,” Hum. Genet., vol. 131, no. 3, pp. 407–413, Mar. 2012, doi: 10.1007/s00439-011-1083-9.
[59] T. Wang et al., “A deletion/insertion mutation in the BRCA2 gene in a breast cancer family: a possible role of the Alu-polyA tail in the evolution of the deletion,” Genes. Chromosomes Cancer, vol. 31, no. 1, pp. 91–95, May 2001, doi: 10.1002/gcc.1110.
[60] H. M. Van Der Klift, C. M. Tops, F. J. Hes, P. Devilee, and J. T. Wijnen, “Insertion of an SVA element, a nonautonomous retrotransposon, in PMS2 intron 7 as a novel cause of lynch syndrome,” Hum. Mutat., vol. 33, no. 7, pp. 1051–1055, Jul. 2012, doi: 10.1002/humu.22092.
[61] Y. Miki et al., “Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer,” Cancer Res., vol. 52, no. 3, pp. 643–645, Feb. 1992.
[62] E. Teugels, S. De Brakeleer, G. Goelen, W. Lissens, E. Sermijn, and J. De Grève, “De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes,” Hum. Mutat., vol. 26, no. 3, pp. 284–284, Sep. 2005.
doi: 10.1002/humu.9366.
[63] Y. Miki, T. Katagiri, F. Kasumi, T. Yoshimoto, and Y. Nakamura, “Mutation analysis in the BRCA2 gene in primary breast cancers,” Nat. Genet., vol. 13, no. 2, pp. 245–247, Jun. 1996, doi: 10.1038/ng0696-245.
[64] M. Kloor et al., “A large MSH2 Alu insertion mutation causes HNPCC in a German kindred,” Hum. Genet., vol. 115, no. 5, pp. 432–438, Oct. 2004, doi: 10.1007/s00439-004-1176-9.
[65] A. Peixoto et al., “Genomic characterization of two large Alu-mediated rearrangements of the BRCA1 gene,” J. Hum. Genet., vol. 58, no. 2, pp. 78–83, Feb. 2013.
doi: 10.1038/jhg.2012.137.
[66] A. Economou-Pachnis and P. N. Tsichlis, “Insertion of an Alu SINE in the human homologue of the Mlvi -2 locus,” Nucleic Acids Res., vol. 13, no. 23, pp. 8379–8387, 1985, doi: 10.1093/nar/13.23.8379.
[67] P. Preeti, R. Sinha, and K. Rawal, “Identification and characterization of mobile genetic elements in high-throughput sequencing data using a computational approach,” presented at the 2ND INTERNATIONAL CONFERENCE ON RECENT ADVANCES IN COMPUTATIONAL TECHNIQUES, Mumbai, India, 2023, p. 020005. doi: 10.1063/5.0148307.