Review paper

Journal of High School Science


# Contextualizing transposable elements and genetic disorders in cystic fibrosis

Gooty A[1], Dinakaran S[2]

Submitted: May 1, 2023, Revised: version 1, May 31, 2023
Accepted: May 31, 2023

**Abstract**

Transposable elements are mobile segments of DNA that can integrate themselves into different parts of the genome through various mechanisms. Retrotransposons are a type of transposable element found in humans that transpose *via* a copy-paste mechanism. Several mutations have been linked to retrotransposon insertions, resulting in genetic disorders, including cystic fibrosis. This review examines the role of retrotransposons in human genetic diseases like cystic fibrosis. Approximately 15% of the CFTR gene mutations can be attributed to retrotransposon insertions. Since there is an understandable propensity to focus on the most common CFTR mutation (F508del, in 70% of CF cases), there is a pronounced gap in the field of cystic fibrosis research involving transposable elements; which may also be present in conjunction with other mutations. Furthermore, anecdotal evidence is emerging that CRISPR targeted, therapeutic intervention sites may subsequently be more prone to retrotransposon insertions. This review aims to bridge these gaps. By uncovering the existence of, and function of retrotransposons in gene regulation networks, their roles in genetic disease can be better understood. We describe the various types of human retrotransposons, their role in gene regulation, the two major disease-causing insertions of retrotransposon elements in the CFTR gene, and the overall clinical implications of transposable elements in the context of human disease in general and for cystic fibrosis in particular. In combination with inventions in molecular biology and advancements in bioinformatic approaches, in an ironic twist, it has emerged that retrotransposons may even be harnessed for their therapeutic effects in cystic fibrosis and other genetic disorders.

**Keywords**

Transposable elements, Retrotransposons, Cystic fibrosis, *Alu* insertions, Retrotransposon insertions, CFTR mutation, Bioinformatics, Genetic therapy, Genetic diseases, Mutagenesis, Single gene disorder

---

[1]Corresponding Author: Anindita Gooty, Irvington High School, 41800 Blacow Rd, Fremont, CA 94538, USA. aninditagooty@gmail.com
[2]Mentor: Sajeth Dinakaran, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell,500 Hofstra Blvd, Hempstead, NY 11549, USA. sajethdinakaran@gmail.com

## Introduction

Transposable Elements (TEs), or "jumping genes", are mobile genes that can integrate themselves into other parts of the genome (1-3). TEs make up nearly half the human genome, and while they have been widely known to be junk DNA, recent studies have uncovered their functional capabilities, especially in areas involving gene expression. TEs can be classified into two main categories: DNA transposons, which transpose themselves directly, and retrotransposons, which transpose themselves via an RNA intermediate (3). Of these, retrotransposons are the only TEs found in the human genome.

Retrotransposons use a copy and paste mechanism to replicate, in which reverse transcriptase transcribes the RNA intermediate, and the cDNA that is produced gets inserted into another part of the genome (1, 2). Retrotransposons can be classified into two main categories: those with Long Terminal Repeats (LTR), and those without. Non LTR retrotransposons can be further classified into long interspersed elements (LINEs) and short interspersed elements (SINEs) (2-4). LINEs and SINEs play important roles in gene regulation. Their functions include regulating chromatin boundaries, transcription, and RNA processing. They also have the potential to suppress the expression of nearby genes depending on epigenetic factors, like DNA methylation (1, 3).

Some retrotransposon insertions have the potential to cause diseases in humans. Many of these are single gene disorders, and include diseases like Hemophilia, Breast Cancer, and Menkes Disease (2). Cystic fibrosis is a notable inclusion among the single gene disorders linked to these insertions, and this review will focus on how retrotransposons impact this disease.

It should be noted that, while retrotransposon mediated indels have been shown to be sufficient to cause cystic fibrosis, there are many other components that contribute to the disease's manifestation. There are several molecular and cell signaling cofactors which manifest on the translational level that contribute to CFTR mutations. These include, but are not limited to, defective protein processing, dysregulation of upstream promoter elements, and even defective post-translational protein processing (5, 6). In some cases, there is variance among the tissues of individual cystic fibrosis patients. For instance, Pagani et. al. reported that patients with cystic fibrosis have different levels of retrotransposon mediated exon 9 skipping among different tissues (7). This illustrates variance in the ability of retrotransposon mediated indels to cause cystic fibrosis, and suggests that other factors operate in conjunction with the polymorphic locus to regulate the amount of exon 9 skipping. Overall, cystic fibrosis still remains a prominent inclusion in the breadth of retrotransposon linked diseases, and this review will focus on the role of retrotransposons in cystic fibrosis origin and management.

## Methods

Pubmed (https://pubmed.ncbi.nlm.nih.gov/) was utilized as the literature search tool for this review. The search query "retrotransposon genetic disease" yielded 1,199 results. The query results were restricted to a range of 2000-2022, and the keywords "cystic fibrosis" and "single gene disorders" were applied. After further filtering based on relevance and sufficiency, 50 results remained. These results revealed two major retrotransposon insertions in the CFTR gene that cause cystic fibrosis, bioinformatic approaches to retrotransposon insertions in genetic disease, and the role of various transposable elements in gene regulation pathways. 26 results were ultimately

included in the review. A flowchart presenting the literature search method is shown in Figure 1.

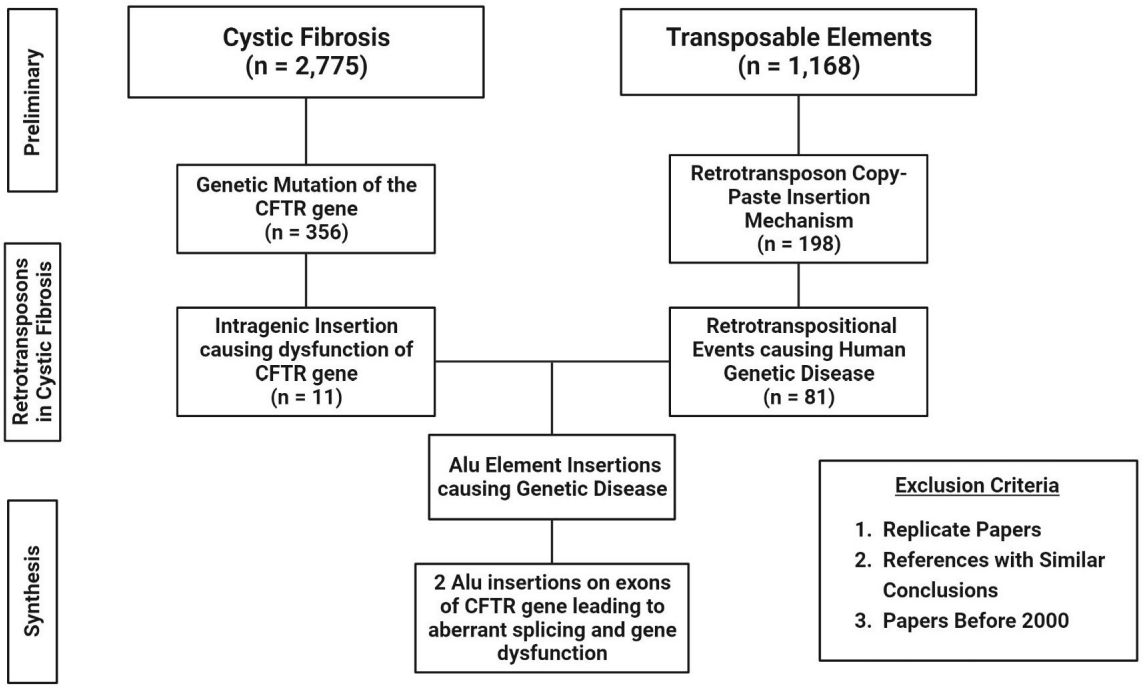


Figure 1: Literature search method

*Cystic fibrosis*

Cystic fibrosis is a life-restricting autosomal recessive disorder that primarily affects the lungs and pancreas (8-10). This disease is characterized by infection prone airways, resulting respiratory failure, as well as inadequate exocrine activity from the pancreas, resulting in malnutrition and other related diseases (8, 10). Cystic fibrosis results from mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Malfunction of this gene leads to dysregulation of ion flow across plasma membranes. This causes those afflicted to be particularly vulnerable to infections due to mucus buildup in the lungs, and malabsorption due to mucus blockage of the pancreatic ducts (8, 9).

Cystic fibrosis is most prevalent in Caucasians, with an incidence of 1:3,200. Caucasians also have the highest carrier frequency, at 1/29. Those of Hispanic descent have an incidence between 1:10,000 and 1:14,000. Those of African descent have an incidence of 1:15,000. Finally, Asians are the racial group least susceptible to cystic fibrosis, with an incidence of 1:35,000 and a carrier frequency of 1/90 (8). Currently, there are an estimated 70,000 cases of cystic fibrosis worldwide, with the distribution being most concentrated in Europe, North America, and Australia (10).

Over the past decade, the median survival age has increased significantly, from around 36, to about 43 (8). Additionally, improvements in treatment for the disease has shifted cystic fibrosis from a disorder primarily found in children, to one found mostly in adults, with the number of adults with the disorder being predicted to increase by nearly 70% by 2025 in Europe (9).

Cystic fibrosis has multiple clinical manifestations that contribute to the disease's progression, several being fatal. The ones that result most frequently in mortality are bronchiectasis and small airway obstruction, as they often lead to respiratory failure. Cystic fibrosis' impact on epithelial cells also causes accompanying diseases like biliary cirrhosis, malabsorption, and heat shock (9). Clinical approaches involving the development and administration of drugs to clear mucus from the airways and treat infections have substantially improved over the years. Additionally, therapies that correct the basic defect in the CFTR gene present a novel solution to the disease (8, 9). Currently, the two major approaches are the use of small molecules to regulate and restore function to the CFTR protein, and the use of molecular or gene therapies to correct the CFTR gene mutation and produce functional CFTR proteins (9). Though innovative therapies like these are being presented, there is an overwhelming lack of solutions that consider the role of retrotransposons. Out of over 1,300 studies documented on http://clinicaltrials.gov about cystic fibrosis, not a single one considered the retrotransposon insertions that impact the disease. The lack of exploration of this avenue may limit the development of solutions, and should as such be considered when developing newer treatments. This review will focus on the retrotransposon insertions that affect cystic fibrosis, how they link to the underlying cause of the disease, and their clinical implications.

*Conventional disorders linked to retrotransposon function*

In humans, the primary source of continuous retrotransposition is the LINE-1 family, a group of retrotransposons that contribute to protein coding activities. Because TE expression is regulated so rigidly, most elements are silenced in somatic cells, usually through heterochromatinization or DNA methylation. Because of this, TE expression in the genome is generally limited to intron sequences and long noncoding RNAs (lncRNAs). However, when these regulatory mechanisms are bypassed, patterns of modified TE expression emerge, impacting many disorders (11). For instance, Hemophilia A and B, Breast Cancer, Menkes Disease, and Dent's disease have all been linked to retrotransposon insertions (2).

Several types of cancer, including lung, colon, and pancreatic cancer, are closely associated with deregulation of L1 promoters and elements. This occurs through retrotransposition of L1 elements in somatic cells, hypomethylation, and expression of L1 synthesized proteins like L1 ORF1p. Additionally, though rare, some L1 insertions have the capability to cause cancerous mutations, as seen in the mutation of the APC tumor suppressor gene of early-stage colon cancers (11).

*Detection of first Alu insertions in the CFTR gene in cystic fibrosis*

Furthermore, some categories of non-autonomous retrotransposons are also linked to diseases. The most notable of these in humans are *Alu* elements, a class of SINE retrotransposons found abundantly in primates. They have a high capacity for polymorphism in their insertions, with their presence varying among individuals. Additionally, they have many mechanisms for disease causing insertions, the most common of which include integration of retrotransposed *Alu* elements and *Alu-Alu* recombination events (12). These insertions contribute to disease by disrupting coding regions of the genome or intercepting a splice signal (13-15). Overall, this makes them a prime source of genetic disease.

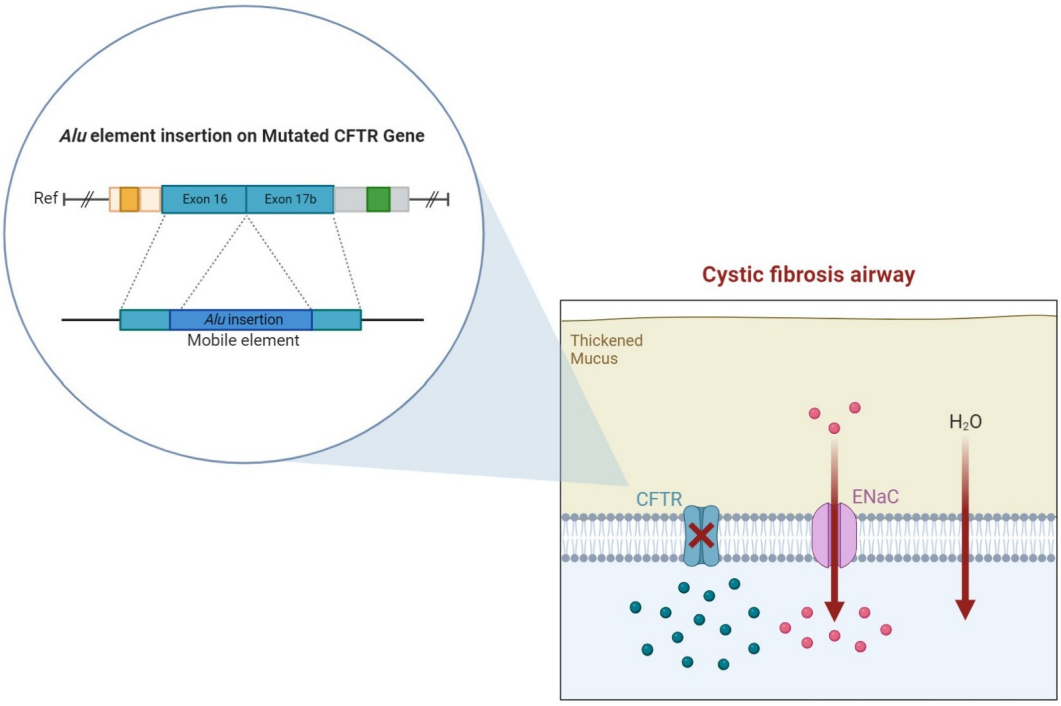


Figure 2: Alu insertion in the CFTR gene influences protein function by a disruption of transcription, leading to cystic fibrosis phenotypes.

Integration of *Alu* elements in the human genome has been linked to several single gene disorders, like Hemophilia, XLA, and HIGM. *Alu* elements have also been revealed to play a role in cystic fibrosis as shown in Figure 2 (2). Notably, two *Alu* insertions have been detected in the CFTR gene, a gene whose mutation is instrumental in the development of cystic fibrosis (12, 16). Specifically, a study performed to identify *Alu* sequence linked mutational events used quantitative high-performance liquid chromatography (QHPLC) to reveal two simple, distinct *Alu* insertions on the CFTR gene within exons 16 and 17b (16). Although many CFTR mutations have been found using QHPLC methods, there exists varying literature surrounding the best method of detection. Prior literature, for instance, also utilizes quantitative multiplex PCR of short fluorescent fragments (QMPSF) to find certain alleles not characterized by QHPLC (17, 18). This indicates that a combination of the two methods may be the most efficacious method for identifying retrotransposon indels, as certain indels may not be characterized through use of either method alone.

The first insertion on exon 16 was discovered in a 24-year-old French woman. When performing gene analysis to identify the mutation in the CFTR gene, a large deleted genomic section was expected to be found, but the researchers failed to identify any such deletion. However, studying events in which retrotransposition led to human genetic disease prompted the researchers to instead look for an intragenic insertion. Further QHPLC

analysis indicated the possibility of an insertional allele across from the unmutated allele. Upon sequencing a PCR amplified exon 16, the first simple *Alu* insertion on the CFTR gene was identified. This likely functional consequence of this insertion is aberrant splicing (16). The discovery of a disease-causing retrotransposon insertion in the CFTR gene frames cystic fibrosis mutations under a newfound lens. By linking transposable element insertions to genetic diseases like cystic fibrosis, a whole new field of research opens up, and novel treatments involving alleviation of these transposable element insertions can then be proposed.

The second insertion instance (insertion two) was found on an exon 17b of a 28-year-old Czech man. Once again, QHPLC analysis was used to identify the exon on which an insertional allele may exist. The exon was amplified and cloned using PCR, to reveal a prominent band of DNA. Sequencing of this band revealed the second simple *Alu* insertion in exon 17b. Insertion two was also been presumed to cause aberrant splicing, a consequence similar to that of the first discovered insertion. The main difference between the first and second *Alu* insertions included their size, with the second insertion demonstrating a far weaker band length (16).

Within all CFTR mutations, up to 0.13% can be accounted for by these two simple *Alu* insertions. With over 1523 different known CFTR mutations, nearly 200 can be accounted for by these retrotransposon insertions (16). By further exploring the effect of these insertions on the CFTR gene and cystic fibrosis as a whole, new solutions can be proposed to combat the disease. Additionally, these two insertions are a part of over 50 retrotransposition events that have been proven to cause human genetic disease, all with similar circumstances involving their transposition (16). By looking at genetic disease as a whole through the lens of retrotransposons, the scientific community can uncover and propose novel therapies that can better address the different mechanisms by which these diseases manifest.

*Therapeutic potential of retrotransposons*

It turns out that retrotransposons may be therapeutically harnessed to treat the very diseases that they cause; in a classic adaptation of the 'killers-as-healers' model. It is thought that insertion of the correct genetic sequence; in the form of a long retrotransposon mediated insertion; may correct for several mutations in one shot. Biotechnology firms are investigating pre-clinical methods of rewriting the CFTR gene by retrotranspositions using a method known as gene writing. This technology utilizes the adeno-associated virus (AAV), which is non-pathogenic to humans, packaged in lipid nanoparticles as a vector for the delivery of gene modifying components. Research indicates that the use of DNA based gene writing allows for a greater durability of longer delivered sequences, making them unlikely to be mutated or lost. This would also allow them to address larger or multi-point sequence mutations, like those characteristic of the CFTR gene, using retrotransposon mediated methods (19).

In particular, retrotransposons provide an avenue to address multiple different mutations with a singular treatment. While current technologies, like CRISPR, do facilitate the editing of mutated sections of the genome, they are often patient specific. This means they require knowledge of the exact mutation present to properly address it (20). Therefore, CRISPR technology has limitations, especially when considering the large gamut of CFTR mutations, many of which may still be

undiscovered. Retrotransposition allows for large scale DNA editing, and could potentially enable the rewriting of a large portion of the CFTR gene to restore its function (19). These qualities allow both a wider range of impact for a single therapy and a higher confidence level for its efficacy, illustrating the advantage of considering retrotransposons as a novel solution to CFTR mutations. Furthermore, recent literature suggests that, when serially performed, retrotransposon mediated mutagenesis could lead to genome domestication (21, 22). In this way, retrotransposon therapy would both help treat cystic fibrosis patients and prevent further post-therapy mutagenesis, providing a protective effect as well as increasing genome fidelity. For these reasons, retrotransposons should; and are being; considered when developing novel therapies for genetic disorders like cystic fibrosis.

*Bioinformatic approaches to identifying various retrotransposons in Alu + CF function*

There is still much to be elucidated with regard to the origin of retrotransposon indels in the varying mutations of cystic fibrosis patients. Bioinformatic approaches present an avenue to address these gaps by identifying the lineages of these mutations. For instance, the Recombination Analysis Tool (RAT) can be used to detect DNA or RNA exchange within the genome (23). Another such approach used genomic DNA screening to identify breakpoints associated with the CFTR gene. It was concluded that there was low correlation between homologous recombination and CFTR deletion (18). Additional findings include a preliminary study that asserted the high allelic heterogeneity in CFTR, submitting that *Alu* mediated homologous recombination does not significantly contribute to retrotransposon mediated indels in the CFTR gene (17). There is also literature that submits that CFTR mutations occur in low complexity parts of the genome (18).

These findings speak to the aptness of bioinformatic approaches in investigating CFTR mutations, as they can help to identify the various specific sequences that underlie different patients, and which of those may be considered at greater risk for retrotransposon mutations. All of these approaches are valuable in validating earlier findings concerning the CFTR gene and the common mutations that occur within it.

Bioinformatics approaches effectively address the high heterogeneity of the CFTR gene by allowing researchers to address millions of sequences at once. Furthermore, they allow for deeper investigation of any similar genes in which retrotransposons interact at a low complexity region of the genome. This would provide a way to characterize CFTR as a representative gene and open a gateway for understanding all retrotransposon-based diseases. Overall, bioinformatics offers a method of collecting and quantifying genomic data on multiple patients and identifying any possible patterns of mutation.

Genome Wide Association Studies (GWAS) have been used in several research studies to analyze various aspects of retrotransposons. These studies have been useful in evaluating both the impacts of retrotransposons on genetic disease, and their role in genome evolution. Through studies using GWAS, valuable discoveries about retrotransposon roles in regulatory networks and their various mechanisms of mutation have been made. For instance, a study on human retrotransposon insertions analyzed their impact on human health and connected phenotypes through using refined GWAS and bioinformatic approaches. Through these, the researchers revealed how disease

associations were mediated by retrotransposons through their gene regulatory properties, like DNA methylation and chromatin boundary regulation (24, 25).

Bioinformatics is a powerful tool in uncovering the interactions between transposable elements and genes. It has helped target and identify specific retrotransposon insertions that cause disease, like that of the *Alu* insertion in the CFTR gene. In the future, bioinformatic studies can allow for the specific screening of transposable elements within select genes of interest, and investigate how the gene regulatory mechanisms within that sequence affect health and the expression of particular phenotypes. These studies would allow researchers to both identify the mechanism by which retrotransposon insertions affect genomic stability, and illuminate their larger role in genome evolution (26). Identification of these aspects of the insertions could then be applied in a clinical context, and allow for a more holistic approach to genetic therapies for certain disorders (18, 19).

**Conclusion**

Retrotransposon mutational insertions in the CFTR gene may be sufficient to cause the production of nonfunctional or partially functional protein, leading to cystic fibrosis. These mutational insertions may be present in conjunction with other, more frequent and visible mutations such as the F508del; and may account for sizable proportion (~15%) of all CFTR mutations. CRISPR based therapies designed to address single mutations hence may not achieve clinical success, if such retrotransposon mediated mutations are also present. Furthermore, recent evidence suggests that CRISPR targeted therapeutic sites are subsequently more prone to retrotransposon insertional mutagenesis. Ironically, retrotransposons can be therapeutically harnessed to re-write entire genes or large sequences in a gene; thereby potentially correcting multiple mutations within a single gene with one large retrotransposon mediated therapeutic insertion. The advantages of such a method, over currently approved 'antisense exon-skipping' therapies is self-evident; especially if no single mutation is more prevalent in a gene with multiple mutations. It is evident that much knowledge remains to be unearthed about retrotransposon insertions and retrotranspositional events in the human genome so that new ways to approach and treat genetic diseases may become evident.

**References**

1. Mita P, Boeke JD. How retrotransposons shape genome regulation. Curr Opin Genet Dev. 2016 Apr;37:90-100. http://doi.org/10.1016/j.gde.2016.01.001.

2. Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. Curr Opin Genet Dev. 2012 Jun;22(3):191-203. http://doi.org/10.1016/j.gde.2012.02.006.

3. Elbarbary RA, Lucas BA, Maquat LE. Retrotransposons as regulators of gene expression. Science. 2016 Feb 12;351(6274):aac7247. http://doi.org/10.1126/science.aac7247.

4. Ichiyanagi K. Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. Genes Genet Syst. 2013;88(1):19-29. http://doi.org/10.1266/ggs.88.19.

5. Chen Q, Shen Y, Zheng J. A review of cystic fibrosis: Basic and clinical aspects. Animal Model Exp Med. 2021 Sep 16;4(3):220-232. http://doi.org/10.1002/ame2.12180.

6. Molinski SV, Shahani VM, Subramanian AS, MacKinnon SS, Woollard G, Laforet M, et. al. Comprehensive mapping of cystic fibrosis mutations to CFTR protein identifies mutation clusters and molecular docking predicts corrector binding site. Proteins. 2018 Aug;86(8):833-843. http://doi.org/10.1002/prot.25496.

7. Pagani F, Buratti E, Stuani C, Romano M, Zuccato E, Niksic M, Giglio L, Faraguna D, Baralle FE. Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. J Biol Chem. 2000 Jul 14;275(28):21041-7. http://doi.org/10.1074/jbc.M910165199.

8. Goetz D, Ren CL. Review of Cystic Fibrosis. Pediatr Ann. 2019 Apr 1;48(4):e154-e161. http://doi.org/10.3928/19382359-20190327-01.

9. Elborn JS. Cystic fibrosis. Lancet. 2016 Nov 19;388(10059):2519-2531. http://doi.org/10.1016/S0140-6736(16)00576-6.

10. Dickinson KM, Collaco JM. Cystic Fibrosis. Pediatr Rev. 2021 Feb; 42 (2): 55-67. http://doi.org/10.1542/pir.2019-0212.

11. Payer LM, Burns KH. Transposable elements in human genetic disease. Nat Rev Genet. 2019 Dec;20 (12) :760-772. http://doi.org/10.1038/s41576-019-0165-8.

12. Callinan PA, Batzer MA. Retrotransposable elements and human disease. Genome Dyn. 2006;1:104-115. http://doi.org/10.1159/000092503.

13. Deininger P. Alu elements: know the SINEs. Genome Biol. 2011 Dec 28;12(12):236. http://doi.org/10.1186/gb-2011-12-12-236.

14. Deininger PL, Batzer MA: Alu repeats and human disease. Mol Genet Metab. 1999, 67: 183-193. http://doi.org/10.1006/mgme.1999.2864.

15. Belancio VP, Hedges DJ, Deininger P: Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res. 2008, 18: 343-358. http://doi.org/10.1101/gr.5558208.

16. Chen JM, Masson E, Macek M Jr, Raguénès O, Piskackova T, Fercot B, Fila L, Cooper DN, Audrézet MP, Férec C. Detection of two Alu insertions in the CFTR gene. J Cyst Fibros. 2008 Jan;7(1):37-43. http://doi.org/10.1016/j.jcf.2007.04.001.

17. Audrézet MP, Chen JM, Raguénès O, Chuzhanova N, Giteau K, Le Maréchal C, Quéré I, Cooper DN, Férec C. Genomic rearrangements in the CFTR gene: extensive allelic heterogeneity and diverse mutational mechanisms. Hum Mutat. 2004 Apr;23(4):343-57. http://doi.org/10.1002/humu.20009.

18. Férec C, Casals T, Chuzhanova N, Macek M Jr, Bienvenu T, Holubova A, et. al. Gross genomic rearrangements involving deletions in the CFTR gene: characterization of six new events from a large cohort of hitherto unidentified cystic fibrosis chromosomes and meta-analysis of the underlying mechanisms. Eur J Hum Genet. 2006 May;14(5):567-76. http://doi.org/10.1038/sj.ejhg.5201590

19. Tessera Therapeutics. (n.d.). http://www.tesseratherapeutics.com

20. Zhang F, Wen Y, Guo X. CRISPR/Cas9 for genome editing: progress, implications and challenges. Hum Mol Genet. 2014 Sep 15;23(R1):R40-6. http://doi.org/10.1093/hmg/ddu125.

21. Comeaux MS, Roy-Engel AM, Hedges DJ, Deininger PL. Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? Genome Res. 2009 Apr;19(4):545-55. http://doi.org/10.1101/gr.089789.108.

22. Carmi S, Church GM, Levanon EY. Large-scale DNA editing of retrotransposons accelerates mammalian genome evolution. Nat Commun. 2011 Nov 1;2:519. http://doi.org/10.1038/ncomms1525.

23. Etherington GJ, Dicks J, Roberts IN. Recombination Analysis Tool (RAT): a program for the high-throughput detection of recombination. Bioinformatics. 2005 Feb 1;21(3):278-81. http://doi.org/10.1093/bioinformatics/bth500.

24. Wang L, Norris ET, Jordan IK. Human Retrotransposon Insertion Polymorphisms Are Associated with Health and Disease via Gene Regulatory Phenotypes. Front Microbiol. 2017 Aug 2;8:1418. http://doi.org/10.3389/fmicb.2017.01418.

25. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet. 2009 Oct;10(10):691-703. http://doi.org/10.1038/nrg2640.

26. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE. Natural mutagenesis of human genomes by endogenous retrotransposons. Cell. 2010 Jun 25;141(7):1253-61. http://doi.org/10.1016/j.cell.2010.05.020.