In Silico analysis of SOX2 genes mutations in Leukemia Patients through Ecosystem approach

Muhammad Usman[1], Shafiq ur Rehman[2], Iftikhar Ahmad[3], Nadeem Ullah[4], Sadia Zulfiqar[5], Abdul Hafeez Nasir[6], Muhammad Asam Raza[7], Ali Zafar[8]

PhD[1], PhD[2] ,PhD[3] ,PhD[4], PhD[5] ,PhD[6] ,PhD[7] ,M.Phil[8]

[1*3] Environmental Management Department NCBA& E Lahore, Pakistan

[2*5*8] Department of Basic and Applied Chemistry University of Central Punjab, Lahore Pakistan

[2] Department of Biotechnology University of Central Punjab, Lahore Pakistan

[7] Department of Chemistry University of Gujrat, Pakistan

Corresponding author: Email ID: Shafiq.rehman@ucp.edu.pk

Abstract

Human exposure to SO_2 and bisulfite (HSO^-_3) interact with SOX gene mutations which impact on AML and ALL patients is divided into different parts. AML is a type of leukemia characterized by the rapid growth and accumulation of immature myeloid cells, which are a type of white blood cell. FLT3 is the mutation that occurs commonly in the AML which promotes the growth and proliferation of cells. The prognosis for AML depends on various factors, such as the age of the patient, specific genetic mutations, and overall health. The current study has been performed to identify the novel genetic mutations in the FLT3 protein which is the root cause of leukemia.

The PROVEAN software studies showed 95% single nucleotide polymorphism predicted as a neutral and 5% single nucleotide polymorphism predicted as a deleterious. The polyphen-2 software showed, 86.60 % single nucleotide polymorphism predicted as a benign, 6.60 % SNPs predicted as probably damaging and 6.60 % SNPs predicted as possibly damaging. I-Mutant 11 SNPs study shows that increase stability and 49 SNPs shows that decrease stability. According to SIFT software 48 % single nucleotide polymorphism predicted as a negative and 52 % single nucleotide polymorphism predicted as a positive. With the help of PhD-SNP software, 100 % single nucleotide polymorphism predicted as a deleterious and 0.0 % SNPs predicted as neutral.

Keywords: Gene Mutation, SOX2, In silico

Tob Regul Sci. ™ 2023;9(1): 2609-2620
DOI: doi.org/10.18001/TRS.9.1.180

Introduction

Sulfur dioxide (SO_2) is a unique among various environmental species due to many routes of human exposure to it. The main source of SO_2 is the combustion of coal and oil in urban areas. It is also present in our food, beverages, pharmaceuticals and it is also within our bodies as a result of catabolism of sulfur containing amino acids. These are short lived within the body and rapidly oxidized to sulfate by sulfite oxidase[1]. Sulfur dioxide SO_2 in aqueous medium hydrates into sulfurous acid (H_2SO_3) which is acidic to persist as such within the living cell. At neutral pH, it dissociates into HSO^-_3 and SO^{-2}_3.

The regulation of our exposure to SO_2 and bisulfite (HSO^-_3) is also divided into different parts. Its exposure to atmospheric SO_2 is controlled by Environmental Protection Agency (EPA), which sets permissible level under Government laws. Ingestion of bisulfite in food and pharmaceutical is regulated by other relevant authorities. In this study, we corelate the source of environmental exposure to bisulfite and SO_2 with SOX gene mutations and its health impact on AML and ALL patients through In silico

Leukemia, a complex and life-threatening form of cancer, poses significant challenges in the field of medical research. This hematologic malignancy originates in the bone marrow and disrupts the normal development of blood cells, leading to the uncontrolled proliferation of abnormal cells [2].It can affect both the blood and the bone marrow cells.

Leukemia can be divided into two types acute or chronic. Acute Leukemia is growing very fast while Chronic is a type of leukemia that grows slowly. Leukemia based on its origin is classified into lymphocytic and myelogenous leukemia. In lymphocytic leukemia, defective growth occurs in marrow cells that convert into lymphocytes, while in myelogenous leukemia, defective growth occurs in marrow cells that convert into white blood cells, platelets, and red blood cells [3].

AML is a type of leukemia that is generated by an excess of hematopoietic transit-amplifying cells (progenitor cells) found in the early stages of hematopoiesis' myeloid stem cell (hematocytoblast) differentiation. Leukemia with genetically distinct lesions is included in the classification of AMLs [4]. The majority of them involve both genes that cause signal transduction molecules to be activated all of the time and genes that affect apoptosis [5]. Both children and adults can get AML. It is less common in women as compared to men (those who are over 65 years old). AML has a 29.5 % five-year survival rate [6]. The PMLRAR translocation for acute promyelocytic leukemia and FMS-like receptor tyrosine kinase contains gene duplications that are among the 17 gene mutations found in pediatric AMLs [7].

Acute myeloid leukemia (AML) is described as a highly heterogeneous hematopoietic abnormality, that is caused by excessive proliferation and accumulation of immature myeloid cells in the bone marrow [8]. AML is actually, an uncontrolled proliferation of the clonal neoplastic cells and their accumulation in the bone marrow [9].AML is associated with an impaired differentiation program

blocked at different maturation steps and is resistant to cell death. This disease also includes the presence of malignant cells in the peripheral blood or extramedullary infiltration [10]. According to the definition, AML majorly affects the bone marrow. It is reported that at least 20 % of the nucleated bone marrow cells are found to be immature or undifferentiated leukemia cells. The National Comprehensive Cancer Network (NCCN) have classified patients into three categories on the basis of risk factor [11]. These categories involve patients who are at better risk, intermediate risk, and poor risk. The patients that are categorized as at favorable risk have NPM1 mutation in the absence of FLT3-ITD and CEPBA mutations. And the patients that fall into the poor risk category are FLT3-ITD mutated or have CN-AML and P53 mutations. In AML, the signaling pathways that work to enhance the survival and the proliferation of hematopoietic progenitor cells are deregulated and cooperate with abnormalities in the functioning of transcription factors that are applied during the normal myeloid differentiation, in order to induce leukemia [12].

Over the last decade, DNA sequencing technology has identified AML gene mutations that were previously undetectable by cytogenetic analysis, and which contribute to AML pathogenesis. In acute myeloid leukemia, the mutation that occurs in the first exon of the n-RAS gene is the most common genetic aberration. This mutation is seen in around 25 % of total cases [13]. Activating mutations in the NRAS, KRAS, PTPN1, and NF1 genes cause abnormal proliferative signaling via the Ras/Raf/Mek/Extracellular signal-regulated kinase (ERK) pathway. These mutations can be found in codons I2 or 61, or can also be present in codon 13 of N-RAS genes [4,14]. In at least one case the mutations changed the GTPase activity of the protein product [15]. The leading hypothesis suggests that in response to receptor-mediated signals, the GTP is bound with the normal NRAS protein, which allows the signal to be transduced further into the cell [16]. With the fundamental GTPase activity converting GTP to GDP, the signal transmission would come to an end. The GTPase activity of mutant RAS proteins is diminished, leading to prolonged GTP binding and thus prolonged activation. As a result, the proper working of the RAS protein signaling pathway is disrupted [17].

FMS like tyrosine kinase 3(FLT3) is the mutation that occurs commonly in the AML. It also promotes the growth of cell and cell proliferation same as the other class I mutation. About one-third of all de novo AML patients have FLT3 mutations and are linked with a particularly unfavorable prognosis [18]. The FLT3 gene produces the FLT3 protein, which aids in the growth of white blood cells. This gene mutation promotes the formation of an excessive number of aberrant leukemia cells. The FLT3 mutation causes a particularly aggressive form of leukemia that is more likely to recur after treatment [19].

The FLT3 gene codes for a protein named fms-like tyrosine kinase 3 (FLT3), which belongs to a member of the receptor tyrosine kinase (RTK) family of proteins. Signal transduction is the mechanism by which receptor tyrosine kinases transfer signals from the cell surface into the cell. A particular protein known as FLT3 ligand, may attach to the FLT3 protein, that is located in the outer membrane of several cell types [20].

This binding activates (turns on) the FLT3 protein, which in turn activates a number of intracellular proteins involved in several signaling cascades. The signaling pathways activated by the FLT3 protein regulate a variety of critical physiological functions, including cell proliferation the growth, division, and survival of cells, and the development of early blood cells known as hematopoietic progenitor cells [21,22].

Despite advancements in understanding the genetic and molecular mechanisms underlying leukemia, there is still much to unravel to improve diagnosis, prognosis, and treatment strategies. Ongoing research endeavors focus on identifying novel genetic mutations, unraveling the interactions within the leukemia microenvironment, and exploring targeted therapies and immunotherapies that hold promise in revolutionizing leukemia management. These research initiatives aim to shed light on the intricate nature of leukemia and the genetic mutations which are the major cause of leukemia

Methodology

Data mining

The amino acid sequence for our target protein was obtained by using NCBI (National Center for Biotechnology Information) database (https://www.ncbi.nlm.nih.gov), with accession number/gene id: P36888. NCBI helps to provide complete access to all information related to proteomics, genomics, and bioactive drug candidates [23].

Performance of BLAST

For mutational analysis of FLT3, the protein sequence was aligned to check any single amino acid sequence. For this purpose, an online software, called Basic Local Alignment Search Tool (BLAST) was used. The alignments were obtained and a number of possible changes in the positions were noted.

Mutation Analysis

SIFT

Sorting Intolerant From Tolerant (https://sift.bii.a-star.edu.sg/) is a widely used tool for predicting whether an amino acid substitution in a protein sequence will affect protein function. It assigns a SIFT score to each substitution, indicating the tolerance or intolerance of the change. The scores range from 0 to 1, with lower scores indicating higher predicted deleteriousness or intolerance to the substitution. If the score is < 0.05 the change is considered to affect protein structure and its stability, while a score of 1.0 is considered neutral. The mutations in SIFT will be declared as deleterious if it predicts any missense mutation [24].

SIFT uses sequence homology and the physical properties of amino acids to make predictions. It compares the target sequence to a database of related sequences and calculates a score based on the degree of conservation at the amino acid position of interest.

PolyPhen-2

Polymorphism Phenotyping v2 (http://genetics.bwh.harvard.edu/pph2/) is a trained machine learning online software used for predicting the potential impact of amino acid substitutions or mutations on protein structure and function. It aims to classify variants as either "benign" or "damaging" based on their predicted effect. It is used to calculate position-specific independent count (PSIC) for each SNP in order to predict the structure and function of the protein by the replacement of amino acids [25].

PolyPhen-2 employs a combination of sequence-based and structure-based features to assess the potential functional impact of mutations. It integrates various sources of information, including multiple sequence alignments, protein structural properties, and functional annotations [26]. The PolyPhen-2 score range is between 0.0 and 1.0, where: a score of 0.0 indicates that the substitution is predicted to be benign and is likely tolerated while a score close to 1.0 indicates that the substitution is predicted to be damaging and is likely to have a significant impact on protein function .

PROVEAN

Provean (http://provean.jcvi. org/index.php) is an online software tool used to predict the impact of genetic variations on protein function, Provean provides valuable insights into the potential effects of amino acid substitutions in protein sequences. By analyzing the sequence conservation and physicochemical properties of amino acids, Provean can predict whether a substitution will be detrimental or benign to the protein's function [27].

I-Mutant

I-Mutant2.0 (at http://gpcr2.biocomp. unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) is online software that uses support vector machines (SVMs) to automatically predict the stability of protein upon changes in SNPs. Both the protein structure and more importantly the protein sequence are used as baselines for I-Mutant2.0 predictions. The results obtained are involved in increasing or decreasing the stability of the protein structure [28]. I-Mutant uses algorithms and statistical models to determine and predict the variations in thermodynamic stability caused by these mutations, providing helpful insights into the impact of amino acid substitutions and their potential consequences. It predicts the protein's stability alteration by estimating the free energy change (ΔΔG) associated with the mutation [29].

Results

Sr. no.	Mutations	Provean	PolyPhen-2	I-Mutant	SIFT	PhD-SNP
1	D7G	Neutral	Benign	Decrease	Negative	Deleterious
2	T227M	Neutral	Probably damaging	Decrease	Positive	Deleterious
3	G8A	Neutral	Benign	Decrease	Negative	Deleterious
4	Q10T	Neutral	Benign	Decrease	Negative	Deleterious
5	L11V	Neutral	Possibly damaging	Increase	Negative	Deleterious
6	E346G	Deleterious	Probably damaging	Decrease	Negative	Deleterious
7	Q10R	Neutral	Benign	Decrease	Negative	Deleterious
8	V126I	Neutral	Benign	Decrease	Positive	Deleterious
9	Q363P	Neutral	Benign	Increase	Positive	Deleterious
10	I527M	Neutral	Probably damaging	Decrease	Positive	Deleterious
11	V557I	Neutral	Benign	Decrease	Positive	Deleterious
12	N938H	Neutral	Benign	Decrease	Positive	Deleterious
13	R961C	Neutral	Benign	Increase	Negative	Deleterious
14	L983P	Neutral	Benign	Decrease	Negative	Deleterious
15	L981P	Neutral	Benign	Decrease	Negative	Deleterious
16	S356N	Neutral	Benign	Decrease	Positive	Deleterious
17	F17L	Neutral	Benign	Decrease	Negative	Deleterious
18	M20I	Neutral	Benign	Decrease	Negative	Deleterious
19	N399D	Neutral	Benign	Decrease	Positive	Deleterious
20	S72A	Neutral	Benign	Decrease	Positive	Deleterious
21	F427L	Neutral	Benign	Increase	Positive	Deleterious
22	E480Q	Neutral	Possibly damaging	Decrease	Positive	Deleterious
23	E767D	Neutral	Benign	Decrease	Positive	Deleterious

24	S57M	Neutral	Benign	Decrease	Negative	Deleterious
25	T479P	Neutral	Benign	Increase	Positive	Deleterious
26	A988T	Neutral	Benign	Decrease	Negative	Deleterious
27	M20V	Neutral	Benign	Decrease	Negative	Deleterious
28	I25M	Neutral	Benign	Decrease	Negative	Deleterious
29	D29E	Neutral	Benign	Decrease	Negative	Deleterious
30	N541S	Neutral	Benign	Decrease	Positive	Deleterious
31	Q10R	Neutral	Benign	Decrease	Negative	Deleterious
32	L558F	Neutral	Possibly damaging	Decrease	Positive	Deleterious
33	E598D	Neutral	Benign	Decrease	Positive	Deleterious
34	V557I	Neutral	Benign	Decrease	Positive	Deleterious
35	L560M	Neutral	Benign	Decrease	Positive	Deleterious
36	A97T	Neutral	Benign	Decrease	Negative	Deleterious
37	A80T	Neutral	Benign	Increase	Positive	Deleterious
38	V958M	Neutral	Benign	Increase	Positive	Deleterious
39	C965R	Neutral	Benign	Decrease	Positive	Deleterious
40	A988T	Neutral	Benign	Decrease	Negative	Deleterious
41	V129I	Neutral	Benign	Decrease	Positive	Deleterious
42	H222R	Deleterious	Benign	Decrease	Positive	Deleterious
43	F453S	Neutral	Benign	Decrease	Positive	Deleterious
44	Q903R	Neutral	Benign	Decrease	Negative	Deleterious
45	S840T	Neutral	Possibly damaging	Decrease	Negative	Deleterious
46	L120V	Neutral	Benign	Decrease	Negative	Deleterious
47	R973P	Neutral	Benign	Decrease	Negative	Deleterious
48	S976G	Neutral	Benign	Decrease	Negative	Deleterious
49	M979T	Neutral	Benign	Decrease	Negative	Deleterious

50	E991N	Neutral	Benign	Increase	Negative	Deleterious
51	E477D	Neutral	Benign	Increase	Positive	Deleterious
52	E596D	Neutral	Benign	Decrease	Positive	Deleterious
53	E598D	Neutral	Benign	Decrease	Positive	Deleterious
54	Y597D	Deleterious	Probably damaging	Decrease	Negative	Deleterious
55	N296S	Neutral	Benign	Decrease	Positive	Deleterious
56	N541S	Neutral	Benign	Decrease	Positive	Deleterious
57	V958M	Neutral	Benign	Increase	Positive	Deleterious
58	S59P	Neutral	Benign	Increase	Negative	Deleterious
59	T74A	Neutral	Benign	Decrease	Positive	Deleterious
60	K900R	Neutral	Benign	Decrease	Negative	Deleterious

According PROVEAN software 95% single nucleotide polymorphism predicted as a neutral and 5% single nucleotide polymorphism predicted as a deleterious which is shown in Figure 1.

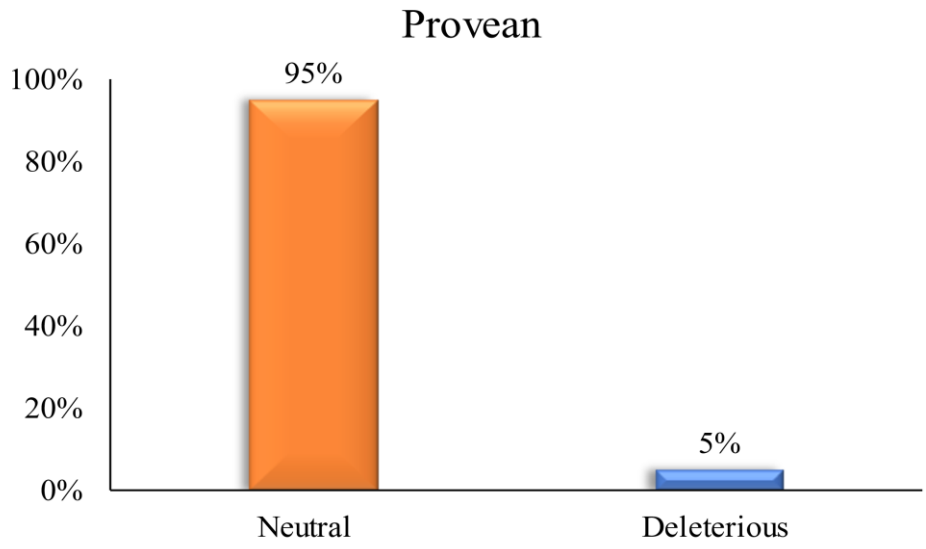

Figure 1 *Proven results of FLT3 protein.*

With the help of polyphen-2 software, 86.60 % single nucleotide polymorphism predicted as a benign, 6.60 % SNPs predicted as probably damaging and 6.60 % SNPs predicted as possibly damaging which is shown in Figure 2.

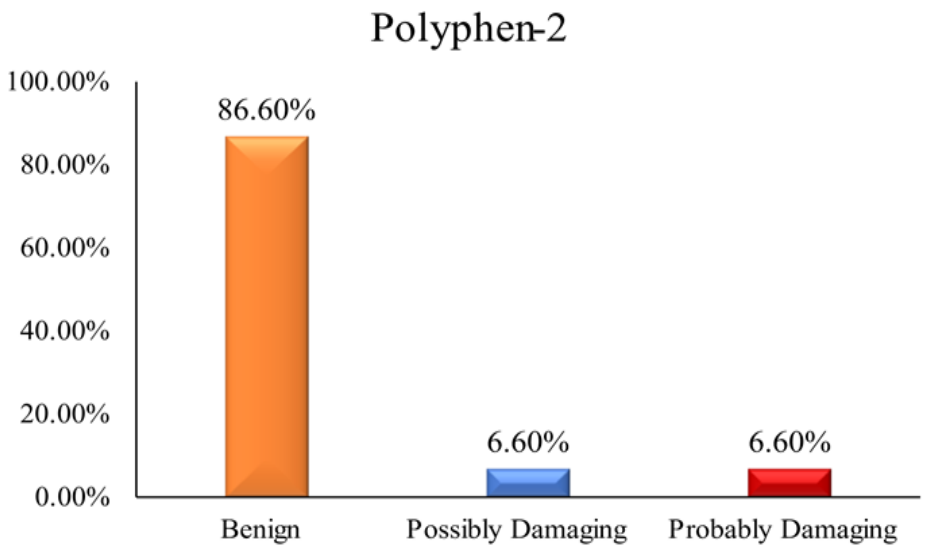

Figure 2. *Polyphen-2 results of FLT3 protein.*

According to I-Mutant 11 SNPs shows that increase stability and 49 SNPs shows that decrease stability which is shown in Figure 3.

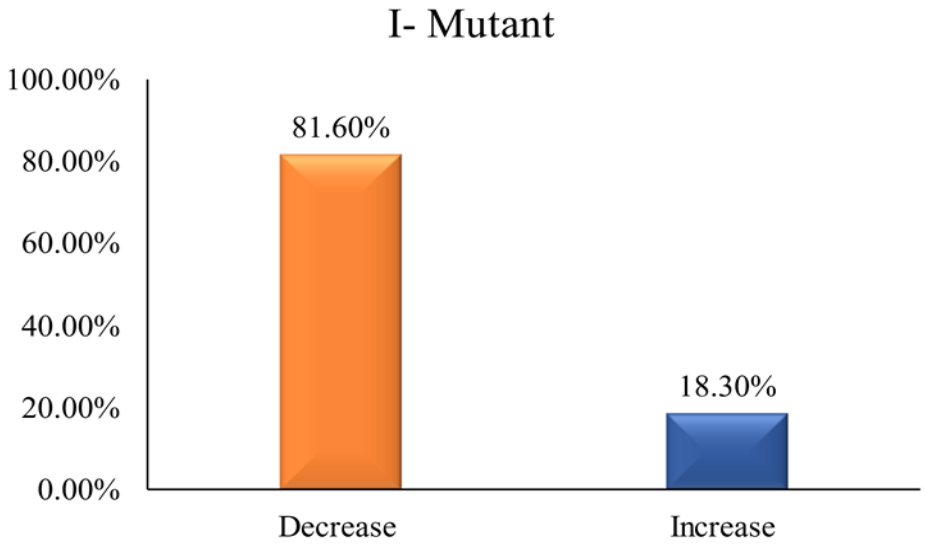

Figure 3. I-Mutant results of FLT3 protein.

According to SIFT software 48 % single nucleotide polymorphism predicted as a negative (affect protein function) and 52 % single nucleotide polymorphism predicted as a positive (Not affect protein function) which is shown in Figure 4.

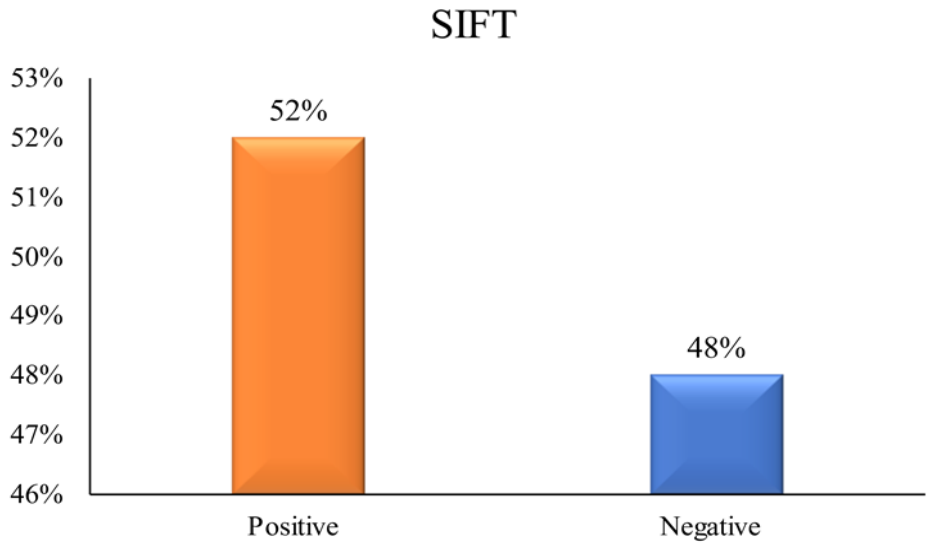

Figure 4. SIFT results of FLT3 protein.

With the help of PhD-SNP software, 100 % single nucleotide polymorphism predicted as a deleterious and 0.0 % SNPs predicted as neutral which is shown in Figure 5.

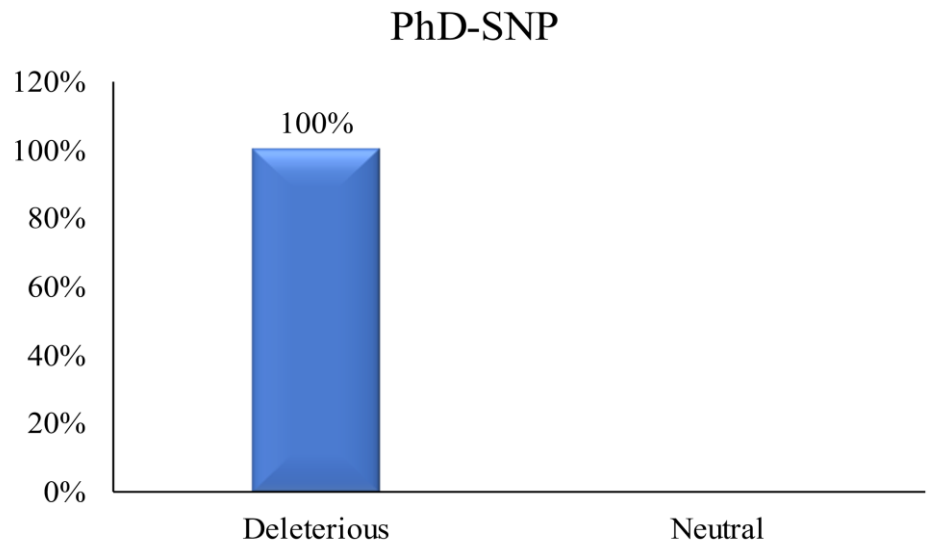

Figure 5. *PhD-SNP results of FLT3 protein.*

Refrences

1. Shapiro, R., Genetic effects of bisulfite: implications for environmental protection. Induced Mutagenesis, 1983: 35-60.
2. Sell S. (2005). Leukemia: stem cells, maturation arrest, and differentiation therapy. Stem Cell Reviews and Reports, **1(3)**, 197–205. https://doi.org/10.1385/SCR:1:3:197
3. Cline, M. J. (1994). The molecular basis of leukemia. New England Journal of Medicine, **330(5)**, 328-336.

4. Varmus, H. E. (1984). The molecular genetics of cellular oncogenes. Annual Review of Genetics, **18(1)**, 553-612.
5. Stone, R. M.; O'Donnell, M. R.; and Sekeres, M. A. (2004). Acute myeloid leukemia. ASH Education Program Book, **2004(1)**, 98-117.
6. Gokhale, P.; Chauhan, A. P. S.; Arora, A.; Khandekar, N.; Nayarisseri, A.; and Singh, S. K. (2019). FLT3 inhibitor design using molecular docking based virtual screening for acute myeloid leukemia. Bioinformation, **15(2)**, 104-115. https://doi.org/10.6026/97320630015104
7. Pui, C. H.; Schrappe, M.; Ribeiro, R. C.; and Niemeyer, C. M. (2004). Childhood and adolescent lymphoid and myeloid leukemia. ASH Education Program Book, **2004(1)**, 118-145.
8. Darici, S., Alkhaldi, H., Horne, G., Jørgensen, H. G., Marmiroli, S., & Huang, X. (2020). Targeting PI3K/Akt/mTOR in AML: rationale and clinical evidence. Journal of Clinical Medicine, 9(9), 2934.
9. Nepstad, I., Hatfield, K. J., Grønningsæter, I. S., & Reikvam, H. (2020). The PI3K-Akt-mTOR signaling pathway in human acute myeloid leukemia (AML) cells. International journal of molecular sciences, 21(8), 2907.
10. Martelli, A. M., Nyåkern, M., Tabellini, G., Bortul, R., Tazzari, P. L., Evangelisti, C., & Cocco, L. (2006). Phosphoinositide 3-kinase/Akt signaling pathway and its therapeutical implications for human acute myeloid leukemia. Leukemia, 20(6), 911-928.
11. Adi, Y. A., Adi-Kusumo, F., Aryati, L., & Hardianti, M. S. (2018). A dynamic model of pi3k/akt pathways in acute myeloid leukemia. Journal of Applied Mathematics, 2018.
12. Park, S., Chapuis, N., Tamburini, J., Bardet, V., Cornillet-Lefebvre, P., Willems, L., ... & Bouscary, D. (2010). Role of the PI3K/AKT and mTOR signaling pathways in acute myeloid leukemia. haematologica, 95(5), 819.
13. Pui, C- H.; Relling, M. V. and Downing, J. R. (2004). Acute lymphoblastic leukemia. New England Journal of Medicine, **350(15)**, 1535-1548.
14. Bos, J. L.; Toksoz, D.; Marshall, C. J.; Verlaan-de Vries, M.; Veeneman, G. H.; van der Eb, A. J. and Steenvoorden, A. C. (1985). Amino-acid substitutions at codon 13 of the N-ras oncogene in human acute myeloid leukaemia. Nature, **315(6022)**, 726-730.
15. McGrath, J. P.; Capon, D. J.; Goeddel, D. V. and Levinson, A. D. (1984). Comparative biochemical properties of normal and activated human ras p21 protein. Nature, **310(5979)**, 644-649.
16. Sweet, R. W.; Yokoyama, S.; Kamata, T.; Feramisco, J. R.; Rosenberg, M. and Gross, M. (1984). The product of ras is a GTPase and the T24 oncogenic mutant is deficient in this activity. Nature, **311(5983)**, 273-275.
17. DiNardo, C. D. and Cortes, J. E. (2016). Mutations in AML: prognostic and therapeutic implications. Hematology 2014, The American Society of Hematology Education Program Book, **2016(1)**, 348-355.

18. Gilliland, D. G.; and Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood, The Journal of the American Society of Hematology, **100(5)**, 1532-1542.
19. Papaemmanuil, E.; Gerstung, M.; Bullinger, L.; Gaidzik, V. I.; Paschka, P.; Roberts, N. D.; and Campbell, P. J. (2016). Genomic classification and prognosis in acute myeloid leukemia. New England Journal of Medicine, **374(23)**, 2209-2221.
20. Kazi, J. U., & Rönnstrand, L. (2019). FMS-like tyrosine kinase 3/FLT3: from basic science to clinical implications. Physiological reviews, 99(3), 1433-1466.
21. Gilliland, D. G., & Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood, The Journal of the American Society of Hematology, 100(5), 1532-1542.
22. K. Bhanumathy, K., Balagopal, A., Vizeacoumar, F. S., Vizeacoumar, F. J., Freywald, A., & Giambra, V. (2021). Protein tyrosine kinases: their roles and their targeting in leukemia. Cancers, 13(2), 184.
23. Elgammal, E. W.; Shawky, H. and Kolomiets, M. V. (2019). In-Silico Analysis, Catalytic Site and Substrate Specificity Prediction of Two Phylogenetically Distinct Zea mays Allene Oxide Synthases. IOSR Journal of Pharmacy and Biological Sciences, **14(1)**, 8-20.
24. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. 2012;40:W452-457.
25. Capriotti E, Altman RB. Improving the prediction of disease-related variants using protein three-dimensional structure. BMC Bioinformatics. 2011;12:S3.
26. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. "A method and server for predicting damaging missense mutations." Nat Methods. 2010 Apr;7(4):248-9. doi: 10.1038/nmeth0410-248.
27. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS ONE. 2012;7:e46688.
28. Capriotti E, Fariselli P, Casadio R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res. 2005;33:W306-W310.
29. Capriotti, E., Fariselli, P., & Casadio, R. (2005). I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic acids research, 33(suppl_2), W306-W310.