Functional and Structural Analysis of Single Nucleotide Polymorphisms in NPM1 gene Associated with Human Acute Myeloid Leukemia: An *In Silico* Approach

A Thesis

Submitted to

Biotechnology and Genetic Engineering Discipline, Life Science School, Khulna University for the partial fulfillment of requirements for the degree of Master of Science (M.Sc.) in Biotechnology and Genetic Engineering.

Submitted by

Tamosree Sikder

Student ID: MS-190721

Session: 2018-2019

Biotechnology and Genetic Engineering Discipline

Life Science School, Khulna University

Khulna-9208, Bangladesh

June, 2022

Functional and Structural Analysis of Single Nucleotide Polymorphisms in NPM1 gene Associated with Human Acute Myeloid Leukemia: An *In Silico* Approach

A Thesis

Submitted by

Tamosree Sikder

Student ID: MS-190721

Session: 2018-2019

Approved as to style and content by

.....

Supervisor

Dr. Md. Emdadul Islam

Professor

Biotechnology and Genetic Engineering Discipline

Khulna University

Khulna-9208, Bangladesh

.....

Professor Dr. Md Morsaline Billah

Head

Biotechnology and Genetic Engineering Discipline

Khulna University

Khulna-9208, Bangladesh

June, 2022

Declaration

This thesis report is the original research work of the author, except as otherwise stated. The material presented has not been submitted either in whole or in part for a degree for this or any other university. The findings of this research will not be published anywhere without the concurrence of the concerned supervisor. Legal action would be taken in any case of infringement regarding the declaration.

Tamosree Sikder 12/11/2024

Tamosree Sikder

Student ID: MS-190721

June, 2022

Dedicated

to

My Beloved Husband

Acknowledgment

At the very beginning, the author expresses the deepest gratitude to the almighty God for all his merciful blessings bestowed upon her which enabled her to complete her thesis work.

The author is pleased to express her sincere gratitude and profound appreciation to her Supervisor, Dr. Md. Emdadul Islam, Professor, of Biotechnology and Genetic Engineering Discipline, Khulna University for his inspiration, continuous guidance, valuable suggestions, and sustained support throughout this time.

The author is also grateful to all the teachers of Biotechnology and Genetic Engineering Discipline, Khulna University for their assistance and encouragement during this thesis work.

The author would like to express her heartiest love to her friends for their suggestions.

The Author

June, 2022

Abstract

NPM1 is a protein-coding gene that encodes nucleophosmin1, a shuttling protein that shuttles between the nucleus and cytoplasm and is involved in different cellular processes such as centrosome duplication, protein chaperoning, DNA repair, and regulation of tumor suppressor ARF. Mutations of the NPM1 gene are associated with human acute myeloid leukemia (AML). Acute myeloid leukemia (AML) is a complex hematopoietic cell disorder characterized by excessive proliferation of hematopoietic cells of myeloid lineage in the bone marrow. This study aimed to predict the most damaging missense SNPs in the human NPM1 gene that may be associated with acute myeloid leukemia (AML). In this study, we used various in-silico tools for the analysis of the functional and structural impacts of missense SNPs on the human NPM1 gene. The missense SNPs of the NPM1 gene were retrieved from the Ensembl database. We analyzed functional and structural impacts using in silico tools like SIFT, PROVEAN, PolyPhen-2, I-Mutant, 3.0MUpro, and MutPred2. The secondary structure was predicted and analyzed by PSIPRED. The 3D structure of NPM1 protein was obtained from AlphaFold, visualization along with mutant models was generated using PyMol, and all information about physiological properties was taken from project HOPE. Protein-protein interaction of NPM1 protein was investigated using STRING. Protein-protein interaction reveals because of SNPs if any change occurs in NPM1 protein, it may affect the overall protein network interactions among all partner proteins in a cell. The results of in-silico analysis uncover eight missense mutations (K54N, I59T, L79S, P152A, K193R, K193N, A283G, I284F), which altered protein structure and affect protein function. In this study, we analyzed functional and structural impacts of missense SNPs in the human NPM1 gene through different in silico tools which uncover eight mutations that change NPM1 protein structure and may be associated with human acute myeloid leukemia. The findings from the clinical study from different literature searches also support the missense SNPs of NPM1 as the key causing agent for acute myeloid leukemia. Further clinical and wet-lab studies regarding our identified eight mutations are necessary.

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Elaboration
NPM1	Nucleophonmin1
SNPs	Single Nucleotide Polymorphisms
AML	Acute Myeloid Leukemia
NCBI	National Center for Biotechnology Information
SIFT	Sorting Intolerant From Tolerant
PROVEAN	Protein Variation Effect Analyzer
PolyPhen-2	Polymorphism Phenotyping v2
CSV	Comma Separated Value
SVM	Support Vector Machine
PSIPRED	PSI-blast based secondary structure PREDiction
HOPE	Have Y(Our) Protein Explained]
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
COSMIC	Catalogue for Somatic Mutations in Cancer
UTR	Untranslated Region

Table of Contents

CHAPTER ONE
INTRODUCTION

1. Introduction

Different types of public databases contain a large number of single nucleotide polymorphisms (SNPs) and this amount rapidly increased in the past few years because of high throughput next-generation sequencing (NGS) technologies[9,10] In the human population single nucleotide polymorphisms (SNPs) are the most frequent type of DNA variation which express about 90% of DNA sequence dissimilarity[1]. SNPs are found in the different locations of a gene such as promoters, exons, introns, 3 prime untranslated regions, and 5 prime untranslated regions, etc. When SNPs located in the protein-coding region of genes that cause amino acid change result in functional differences in the protein products of the genes called missense SNPs or non-synonymous[2].

Single nucleotide polymorphisms (SNPs) of the NPM1 gene have been associated with acute myeloid leukemia (AML)[3]. Nucleophosmin 1 (NPM1) is a protein-coding gene that encodes nucleophosmin protein. NPM1 is also known as B23, NPM. It belongs to the nucleophosmin family of nuclear chaperones[4]. Nucleophosmin protein continuously shuttles between the nucleus and cytoplasm and is involved in distinct cellular processes such as centrosome duplication, protein chaperoning, DNA repair, and regulation of tumor suppressor ARF. The NPM1 protein has two isoforms, they are predominant isoform B23.1 which is located in the nucleolus, and abundant isoform B23.2 located in the nucleoplasm. Several structural domains are found in the NPM1 protein, and each domain plays specific functions[5]. According to the NCBI database, the NPM1 gene is located on chromosome 5q35.1 in humans and contains 13 exons.

Acute myeloid leukemia (AML) is a hematopoietic cell disorder that mainly occurs in the bone marrow and is defined by an excess proliferation of myeloid or hematopoietic cells[6]. About 40%-60% Of adult acute myeloid leukemia with normal karyotype (AML-NK) occurs because of the mutations of the human NPM1 gene[7].

Recently, Dai et al.[8] used Next Generation Sequencing (NGS) data of 112 genes from 238 acute myeloid leukemia patients with NPM1 mutations, they demonstrated among the patients 10 (4.2%) were missense mutations.

We thus tried to identify the SNPs responsible for key causing elements of human acute myeloid leukemia. Hence we planned to use the Ensembl database for retrieving the NPM1 gene and different in silico tools namely SIFT, PROVEAN, PolyPhen-2, I-Mutant 3.0, MUpro, MutPred2, and PSIPRED, for the analysis of the NPM1 gene.

The objectives of this research are-

i) To identify key missense SNPs responsible for human acute myeloid leukemia.

ii) To explore the functional and structural impacts on NPM1 protein due to missense SNPs.

CHAPTER TWO

REVIEW OF LITERATURE

2. Review of literature

Different types of public databases contain a large number of single nucleotide polymorphisms (SNPs) and this amount rapidly increased in the past few years because of high throughput next-generation sequencing (NGS) technologies[9,10]. Single nucleotide polymorphisms occur when a single DNA base pair (A, C, G, T) is altered. In human population shows 90% sequence differences because of single nucleotide polymorphisms (SNPs)[1]. SNPs are found in the different locations of a gene such as promoters, exons, introns, 3 prime untranslated regions, and 5 prime untranslated regions, etc. When SNPs occur within a coding region of a gene and affect protein function, which is called non-synonymous SNPs or missense SNPs. Non-synonymous SNPs are associated with different types of cancer in humans[11]. In the past few years, with the help of different bioinformatics tools functional and structural impacts of deleterious SNPs of IGFIR gene[12], ABLI gene[13], BRCA1 gene[14], IL8 gene[15], NY-BR-1 gene[16], GalNAc-T1 gene[17], IL10 gene[18], MYB oncoproteins gene[19], TRX gene[20] have been identified.

Acute myeloid leukemia is a hematopoietic cell disorder that is clinically, cytogenetically, and molecularly heterogeneous. NPM1 gene mutation is the most common cause of acute myeloid leukemia in adults. About 50%-60% of adult acute myeloid leukemia with normal karyotypes show NPM1 gene mutations[7].

The Npm1 gene is also called B23, numatrin, or NO38 which is a protein-coding gene. The NPM1 gene is located on chromosome 5q35.5.[5]. NPM1 genes encode Nucleophosmis protein a shuttle protein that has a distinct cellular function. NPM1 protein is involved in ribosome biogenesis, histone chaperoning, centrosome duplication, DNA repair, and tumor suppressor ARF[6]. There are many structural domains present in NPM1 protein and they perform multiple biochemical functions. The conserved N-terminal domain is involved in oligomerization and is important for chaperoning, it protects NPM1 protein from being misfolded[5].

CHAPTER THREE
MATERIALS AND METHODS

3. Materials and Methods

3.1 Retrieval of SNPs dataset

The SNPs data of the human NPM1 gene was retrieved from the Ensembl database (https://www.ensembl.org/index.html) for our computational analysis. The Ensembl database integrates different data from several sources such as dbSNP, HapMap, the 1000 genome project, etc., and has a collaboration with major bioinformatics databases and resources like NCBI, GENCODE, and UCSC21. We used Ensembl transcript ID ENST00000296930.5 for retrieving missense or non-synonymous SNPs of the NPM1 gene for our investigation.

3.2 Retrieval of amino acid sequence and wild structure

The amino acid sequence of the NPM1 gene was retrieved from the UniProtKB database (https://www.uniprot.org/uniprot/) (UniProt ID: P06748) in FASTA format. The pdb file of the NPM1 protein was downloaded from the AlphaFold protein structure database (https://alphafold.ebi.ac.uk/) (AlphaFold ID: AF-P06748-F1).

3.3 Prediction of functional consequences of missense SNPs by SIFT and PROVEAN

SIFT (Sorting Intolerant From Tolerant) is a prediction program that predicts protein function consequences affected by amino acid change by using a sequence homology-based algorithm. SIFT prediction is conservation-based if any well-conserved amino acid changes SIFT predicted it as deleterious[22]. SIFT predicts substitution consequences as damaging or tolerated if the substitution affects protein function SIFT predicts it as damaging and if the substitution does not affect protein function SIFT predicts it as neutral. In the SIFT program if the SIFT score is less than 0.05 is predicted to be damaging and if the SIFT score is greater than 0.05 is predicted to be tolerated[22].

PROVEAN (Protein Variation Effect Analyzer) is a sequence homology-based prediction tool that predicts the consequences of single or multiple amino acid changes or indel mutations on protein function[23]. PROVEAN default prediction score is -2.5, if the score is greater than -2.5 the variation is predicted as neutral, and if the score is less than or equal the variation is predicted as deleterious[24].

We used PROVEAN (http://provean.jcvi.org/index.php) Human Genome Variants which provides PROVEAN and SIFT prediction together. We submitted a CSV file containing chromosome, position, reference allele, and variant allele as input for our investigation.

3.4 Prediction of missense SNPs by PolyPhen-2

PolyPhen-2 (Polymorphism Phenotyping v2) is a structural homology-based prediction tool that predicts the amino acid substitution on the structure and function of a protein[25]. PolyPhen-2 is available online at http://genetics.bwh.harvard.edu/pph2/. We used NPM1 protein sequence FASTA format and amino acid substitution with position independently as input for the prediction of missense SNPs. PolyPhen-2 prediction score ranging from 0.0 to 1.0 by default. PolyPhen-2 predicts amino acid substitution as probably damaging, possibly damaging, and benign. PolyPhen-2 is predicted as probably damaging when the amino acid substitution shows damaging with a high confidence score, predicted as possibly damaging when the confidence score is low, and benign means not damaging.

3.5 Protein stability prediction by I-Mutant3.0 and MUpro

We used I-Mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) and MUpro (http://mupro.proteomics.ics.uci.edu/) for the prediction of protein stability change of NPM1 protein due to amino acid change. I-Mutant 3.0 is a support vector machine (SVM) based tool, it predicts protein stability change and prediction consequences either increase stability or decrease stability by predicting the value of free energy (DDG)[26]. Free energy value DDG <0 predicts decreased stability of protein and DDG > 0 predicts increased stability of the protein.
MUpro is also used in SVM-based toots to predict protein stability change for SNPs[27].

3.6 Prediction of mutational impacts on functional and structural properties of NPM1 protein by MutPred2

We used MutPred (http://mutpred.mutdb.org/) web server for the prediction of amino acid substitutions impact of NPM1 protein on its functional and structural properties. MutPred is a machine learning-based tool that predicts pathogenicity and molecular causes of disease of amino acid substitution by integrating the genetic and molecular data[28]. MutPred input was NPM1 protein sequence and a list of amino acid substitutions with space.

3.7 Evolutionary conservation analysis of missense SNPs

We used the Consurf (https://consurf.tau.ac.il/) web server for the prediction of the evolutionary conservation of amino acids in human NPM1 protein. The Consurf web server provides Consurf results with colour-coded conservation scale scores ranging from 1 to 9. The conservation score of 7-9 indicates the highly conserved region of a protein[29]. We used the NPM1 protein sequence FASTA format as input for analysis of the evolutionary conservation of NPM1 human protein.

3.8 Secondary structure prediction using PSIPRED

PSIPRED is a highly accurate secondary structure prediction tool for a protein that integrates two feed-forward neural networks[30]. We used PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/) for the prediction of human NPM1 protein. The input was the protein sequence FASTA format.

3.9 Mutant 3D structure prediction, visualization, and structural impact of missense SNPs on human NPM1 protein

The three-dimensional (3D) structure of human NPM1 protein was downloaded from AlphaFold DB (https://alphafold.ebi.ac.uk/). AlphaFold DB is a protein structure database that predicts protein structure from the amino acid sequence of a protein[31].

The 3D structure of NPM1 mutant proteins was visualized and generated by PyMol (https://pymol.org/2/).

HOPE [Have Y(Our) Protein Explained] is a web server that collects data from several data sources and analyzes the physicochemical changes caused by mutation[32]. It was used to analyze the physicochemical changes caused by a mutation in human NPM1 protein. It is available online at (https://www3.cmbi.umcn.nl/hope/). For HOPE we used protein sequence and single mutation as an input.

3.10 Analysis of protein-protein interaction

STRING is a protein-protein interaction analysis tool that interprets all functional interactions in the cell proteins[33]. STRING (https://string-db.org/) was used to analyze protein-protein interactions of human NPM1 protein. The input was the NPM1 protein sequence.

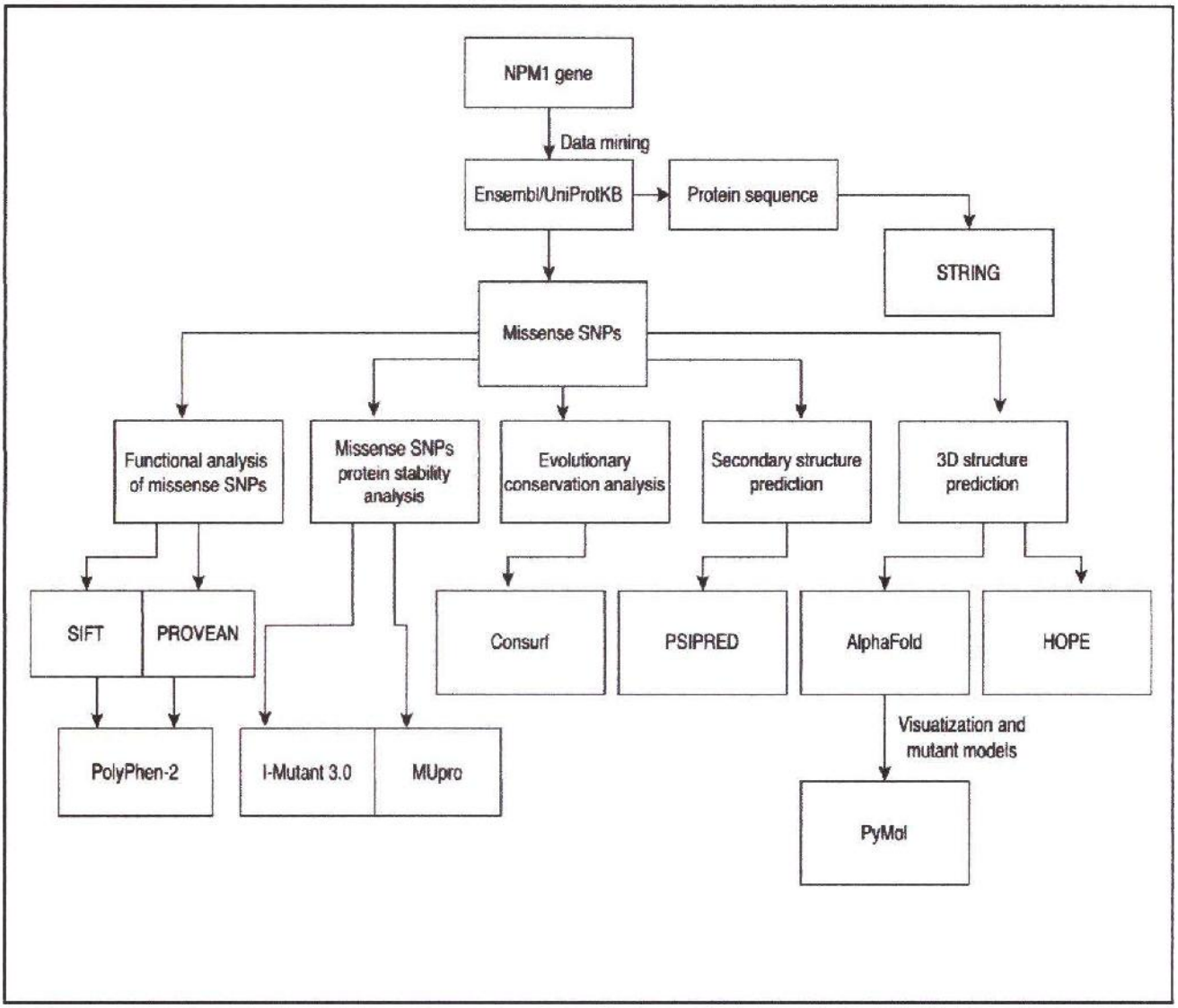


Figure 1: Flowchart of in silico tools for the analysis of the NPM1 gene.

CHAPTER FOUR

RESULTS

4. Results

4.1 SNP dataset

Single Nucleotide Polymorphism of the NPM1 gene investigated in this work was retrieved from the Ensembl database. In the Ensembl database, 8845 SNPs were recorded for the NPM1 gene under the transcript ID ENST00000296930.5. Out of which, 8388 were imported from dbSNP, 456 were imported from COSMIC, and 1 was imported from HGMD-PUBLIC.

Out of all 8845 single nucleotide variants, 112 were synonymous variants, 145 were missense variants, 260 were 5 prime UTR variants, 214 were 3 prime UTR variants, 8 were splice acceptor variants, 5 were splice variants, 15 were frameshift variants, 1 stop variant, 244 were coding sequence variants, 7827 were in intronic variants and rest were other variants of NPM1 gene. We selected 145 missense variants for our analysis.

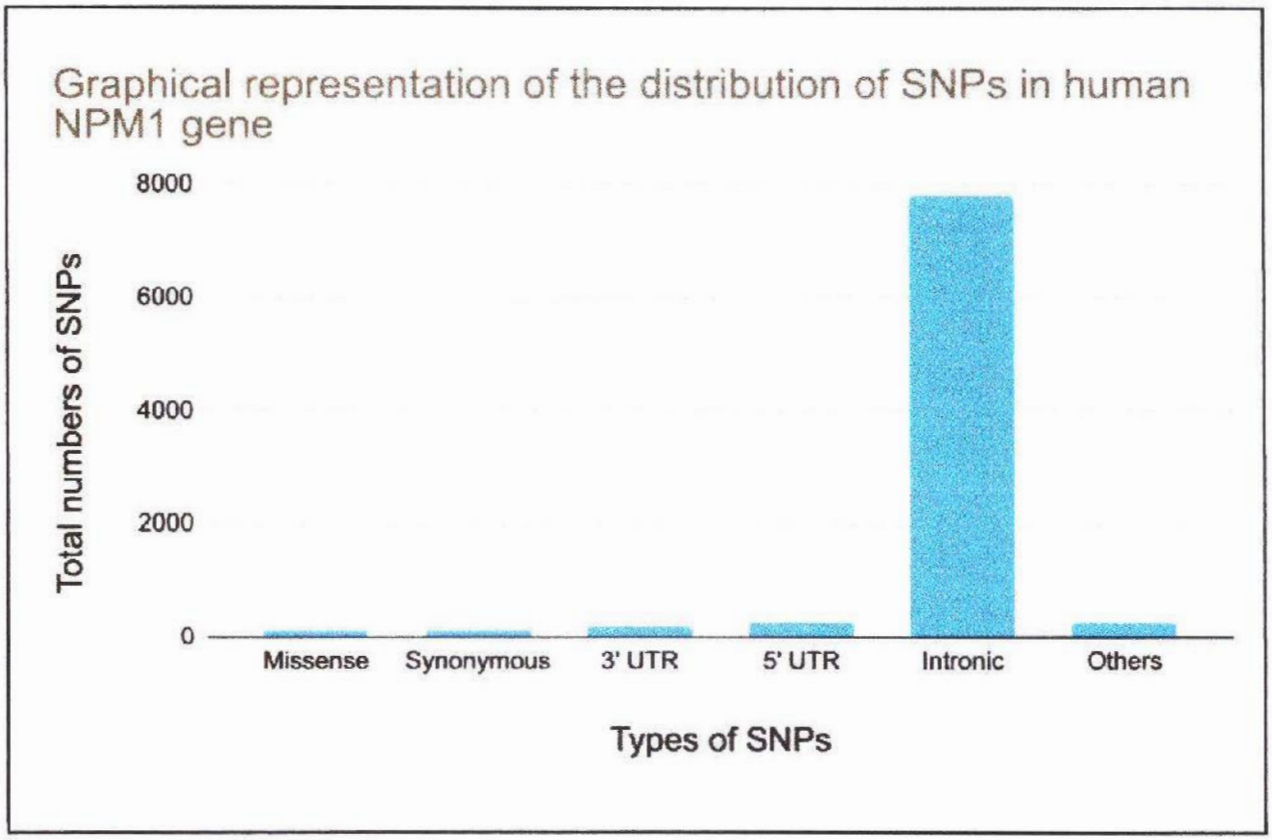


Figure 2: Graphical representation of the distribution of SNPs in the human NPM1 gene.

4.2 Prediction of missense SNPs using PROVEAN and SIFT

The retrieved 145 missense SNPs were submitted as a batch query for PROVEAN prediction. PROVEAN search results provide PROVEAN and SIFT prediction scores and consequences together. Among the 145 analyzed missense SNPs PROVEAN predicted 47 as deleterious and 98 as neutral, and SIFT predicted 62 as damaging with a score >= 0.05 and 83 as tolerated. Among 62 damaging SNPs predicted by SIFT, 4 missense SNPs showed a highly damaging tolerance index score of 0.00 which is rs11551525, rs750633648, rs371956477, rs17850940. Eight showed a score of 0.001, nine showed a score of 0.002, and five showed a score of 0.003. Ten missense SNPs appear with a score between 0.004 and 0.009. The remaining 26 missense SNPs contained tolerance index scores of between 0.01 and 0.05.

Among the 47 deleterious SNPs predicted by PROVEAN 9 showed a highly deleterious score <-4.0 which were rs372208958 (-5.71), rs1288532134 (-4.15), rs11551525 (-5.09), rs954187506 (-4.49), rs868364493 (-4.4), rs1053217290 (-5.54), rs756636316 (-4.21), rs760598971 (-5.24), and rs17851431 (-5.33). 38 missense SNPs showed a deleterious score <=-2.5.

4.3 Prediction of missense SNPs using PolyPhen-2

For high accuracy and a large number of finding we selected 36 SNPs that predicted damaging or deleterious by both SIFT and PROVEAN. Some SNPs that had high confidence scores by at least SIFT or PROVEAN tools were also selected for further analysis by PolyPhen-2. A total of 45 missense SNPs were analyzed by the PolyPhen-2 server. Out of the 45 submitted missense SNPs PolyPhen-2 predicted 9 missense SNPs as probably damaging which means damaging with a high confidence score, 13 missense SNPs predicted as possibly damaging which means damaging but low confidence score, and 23 missense SNPs predicted as benign.

Table 1: List of missense SNPs in the human NPM1 gene analyzed by PolyPhen-2

SNPs	Amino Acid Change	Prediction	Score	Sensitivity	Specificity
rs1363897984	S4L	Benign	0.016	0.95	0.79
rs1278111812	P11S	Possibly damaging	0.691	0.86	0.92
rs1446401368	L18F	Benign	0.449	0.89	0.9
rs372208958	G52D	Benign	0.293	0.91	0.89

rs372208958	G52V	Possibly damaging	0.826	0.84	0.93
rs774730997	A53V	Benign	0.226	0.91	0.88
rs1288532134	**K54N**	**Probably damaging**	**0.999**	**0.14**	**0.99**
rs1561864425	**I59T**	**Probably damaging**	**0.985**	**0.74**	**0.96**
rs11551526	A64V	Possibly damaging	0.566	0.88	0.91
rs11551525	**L79S**	**Probably damaging**	**1**	**0**	**1**
rs17851944	K80E	Benign	0.326	0.9	0.89
rs954187506	V99G	Possibly damaging	0.819	0.84	0.93
rs1467958614	S112G	Benign	0.355	0.9	0.89
rs1308264050	D127N	Benign	0.254	0.91	0.88
rs1177473523	P145L	Benign	0.064	0.94	0.84
rs868364493	**P152A**	**Probably damaging**	**1**	**0**	**1**
rs1454882046	K155N	Possibly damaging	0.879	0.82	0.94
rs1207998618	D167H	Possibly damaging	0.916	0.81	0.94
rs1207998618	D167Y	Benign	0.418	0.89	0.9
rs1449984649	E169K	Benign	0.076	0.93	0.84
rs750633648	D171V	Benign	0.003	0.98	0.44
rs1187263636	E174G	Benign	0	1	0
rs1554135991	D175H	Benign	0.007	0.96	0.75
rs267600552	D178H	Benign	0.082	0.93	0.85
rs17856513	D178G	Benign	0.016	0.95	0.79
rs17856513	D178V	Possibly damaging	0.886	0.82	0.94
rs1239686401	D179G	Benign	0.028	0.95	0.81
rs1481227568	D183H	Possibly damaging	0.91	0.81	0.94
rs183724988	**K193R**	**Probably damaging**	**0.999**	**0.14**	**0.99**
rs761307473	**K193N**	**Probably damaging**	**0.999**	**0.14**	**0.99**
rs1561871583	**S217L**	**Probably damaging**	**0.966**	**0.78**	**0.95**
rs781171336	T234I	Benign	0.014	0.96	0.79

rs1053217290	P235L	Benign	0.092	0.93	0.85
rs374517078	K236T	Benign	0.303	0.91	0.89
rs1246599848	T237I	Possibly damaging	0.558	0.88	0.91
rs756636316	P238T	Possibly damaging	0.62	0.87	0.91
rs756636316	P238A	Possibly damaging	0.943	0.8	0.95
rs771340939	P241S	Benign	0.024	0.95	0.81
rs760598971	D246Y	Benign	0.405	0.89	0.9
rs377348662	A249V	Possibly damaging	0.691	0.86	0.92
rs766808881	K267R	Benign	0.136	0.92	0.86
rs17851431	Y271C	Benign	0.276	0.91	0.88
rs371956477	**A283G**	**Probably damaging**	**0.996**	**0.55**	**0.98**
rs1220932669	**I284F**	**Probably damaging**	**0.971**	**0.77**	**0.96**
rs17850940	L287F	Possibly damaging	0.558	0.88	0.91

We selected nine missense SNPs which are predicted as highly damaging at prediction tools SIFT and PROVEAN both and either SIFT or PROVEAN, and PolyPhen-2. These nine shortlisted missense SNPs were further analyzed by I-Mutant 3.0 and MUpro.

Table 2: Nine Shortlisted missense SNPs predicted by SIFT, PROVEAN, and PolyPhen-2.

Serial No.	**SNP ID**	**Amino Acid Change**
1.	rs1288532134	K54N
2.	rs1561864425	I59T
3.	rs11551525	L79S
4.	rs868364493	P152A
5.	rs183724988	K193R
6.	rs761307473	K193N
7.	rs1561871583	S217L

8.	rs371956477	A283G
9.	rs1220932669	I284F

4.4 Prediction of protein stability using I-Mutant 3.0 and MUpro

For protein stability analysis 9 selected missense SNPs were submitted to I-Mutant 3.0 and MUpro tools. I-Mutant 3.0 predicted 8 mutations as decreased stability and 1 mutation namely S217L predicted increased stability and MUpro also predicted eight mutations as decreased stability and 1 mutation namely S217L as increased stability. Out of the nine mutations, eight mutations (K54N, I59T, L79S, P152A, K193R, K193N, A283G, I284F) were predicted to decrease stability by both I-Mutant 3.0 and MUpro tools. These mutations may alter the protein structure by decreasing the stability as a result the protein function will change also.

Table 3: prediction of protein stability change using I-Mutant 3.0 and MUpro.

Amino Acid Mutation	**I-Mutant 3.0**			**MUpro**	
	Prediction	**DDG(kcal/mol)**	**RI**	**Prediction**	**DDG(kcal/mol)**
K54T	Decrease	-0.45	2	Decrease	-o.87519817
I59T	Decrease	-1.63	6	Decrease	-1.4993491
L79S	Decrease	-1.98	8	Decrease	-2.3760755
P152A	Decrease	-1.16	6	Decrease	-1.2910132
K193R	Decrease	-0.26	4	Decrease	-0.38565313
K193N	Decrease	-0.52	4	Decrease	-0.51229964
S217L	Increase	0.04	3	Increase	0.38124489
A283G	Decrease	-1.18	7	Decrease	-1.4988142
I284F	Decrease	-1.16	8	Decrease	-0.98943866

DDG value: Free energy changes value; RI: Reliability Index.

4.5 Prediction of mutational impacts on functional and structural properties of NPM1 protein by MutPred2

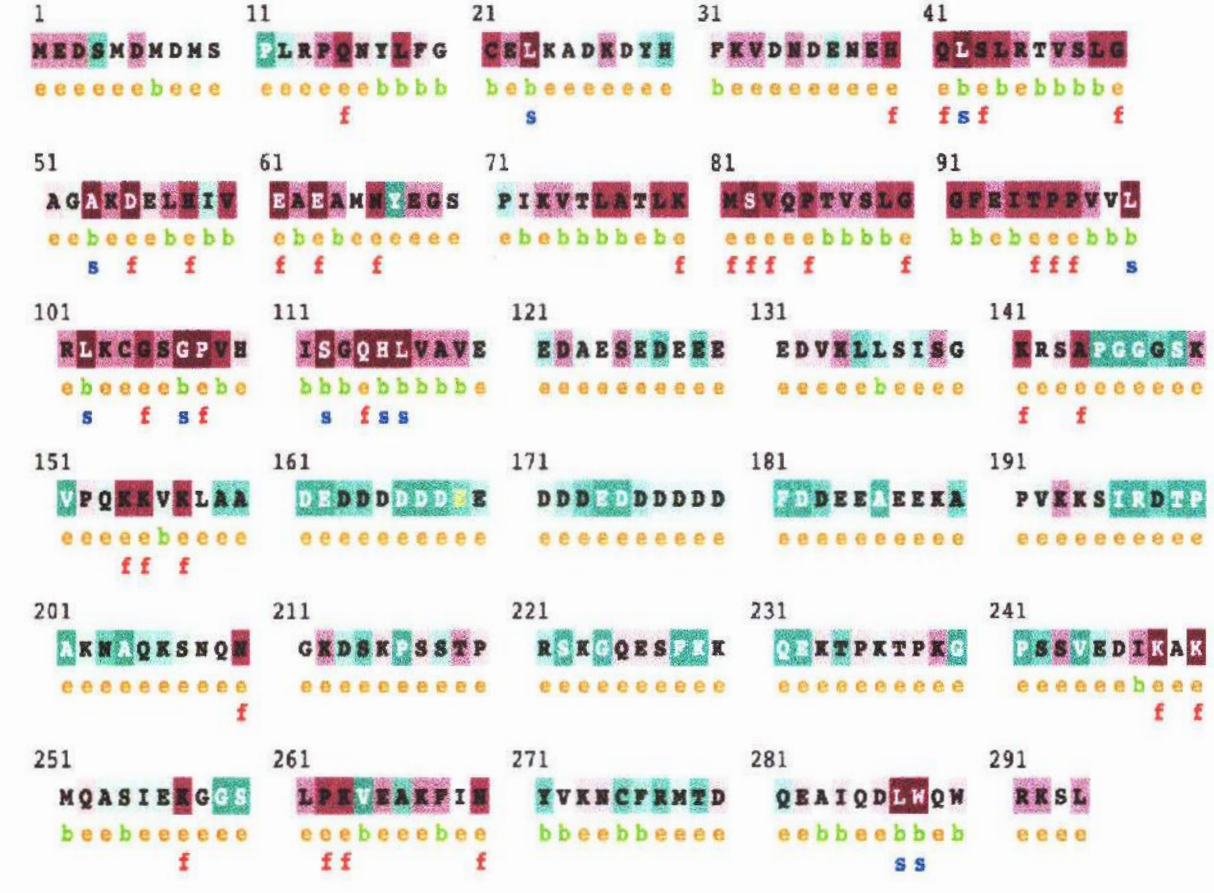
Eight missense SNPs were predicted to decrease protein stability by I-Mutant 3.0 and MUpro was submitted to the MutPred2 web server. MutPred2 predicts molecular mechanism changes induced by mutation with its probability and P-value score. MutPred2 predicts five mutations (K54N, I59T, L79S, A283G, I284F) with a high MutPred2 confidence score greater than or equal to 0.5 which altered various molecular mechanisms of NPM1 protein.

Table 4: Prediction of Molecular Mechanisms of NPM1gene using MutPred2.

Amino Acid Change	Molecular Mechanisms	Probability	P-value	MutPred Score
K54N	Altered metal-binding	0.22	0.02	0.747
	Loss of ubiquitylation at K54	0.15	0.05	
I59T	Altered metal-binding	0.26	0.02	0.635
	Altered stability	0.22	0.01	
L79S	Altered stability	0.58	1.10E-03	0.931
P152A	-	-	-	0.38
K193R	-	-	-	0.131
K193N	-	-	-	0.284
A283G	Altered disordered interface	0.29	0.03	0.568
	Altered ordered interface	0.25	0.03	
	Altered coiled-coil	0.18	0.03	
	Altered transmembrane protein	0.13	0.02	
I284F	Altered disordered interface	0.31	0.02	0.849
	Altered ordered interface	0.24	0.03	
	Altered coiled-coil	0.18	0.02	
	Altered transmembrane protein	0.15	0.01	

4.6 Evolutionary conservation analysis of missense SNPs

We used the Consurf web server to analyze the evolutionary conservation of the NPM1 protein. Consurf web server predicted that out of the 9 shortlisted missense SNPs of NPM1 protein, four missense SNPs (K54N, L79S, K193R, K193L) were located in the highly conserved region ranging 7-9 in the conservation scale. The remaining five missense mutations (I59T, P152A, S217L, A283G, I284F) were located in the conservation region scale 4-6.



e - An exposed residue according to the neural-network algorithm.
b - A buried residue according to the neural-network algorithm.
f - A predicted functional residue (highly conserved and exposed).
s - A predicted structural residue (highly conserved and buried).
? - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure 3: Analysis of evolutionary conservation of human NPM1 protein using Consurf.

4.7 Secondary structure analysis of NPM1 protein

The secondary structure of NPM1 protein was predicted by PSIPRED

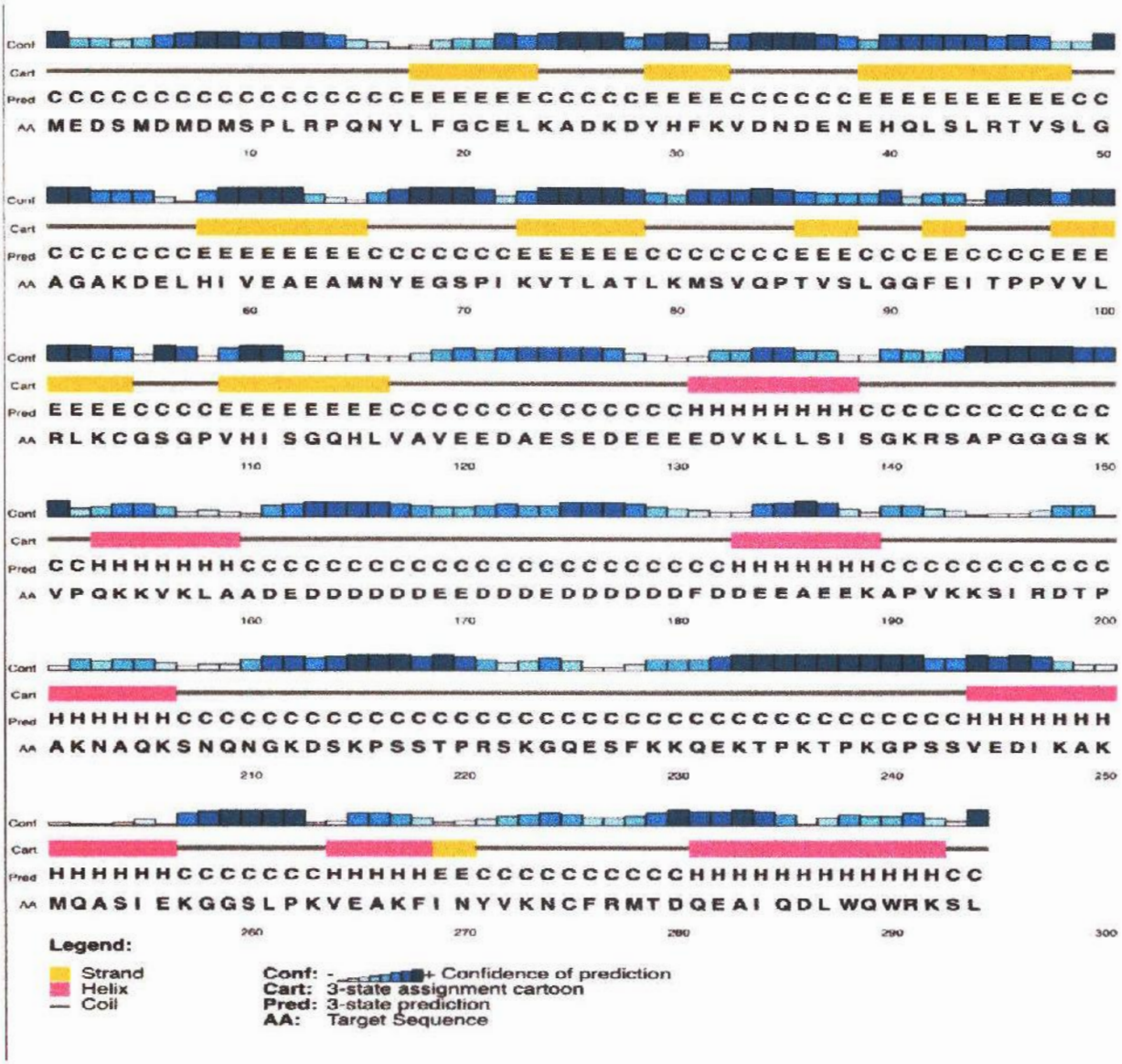


Figure 4: The secondary structure of NPM1 protein predicted by PSIPRED

4.8 3D structure analysis of NPM1 protein

The 3D structure of the NPM1 protein was downloaded from AlphaFold DB. PyMol was used to visualize the amino acid change and the 3D model structure of the eight mutated NPM1 proteins was also generated by PyMol. Project HOPE server was used to study the physicochemical properties

such as size, charge, and hydrophobicity of these mutated proteins compared to wild-type NPM1 protein.

In (K54N): The amino acid Lysine changes to Asparagine at position 54. The wild-type residue forms a hydrogen bond with Glycine at position 52. Because of the size difference (the mutant residue is smaller than the wild-type residue) between wild type and mutant type, it may not make hydrogen bond like wild type residue did and it will affect protein function. The wild-type residue forms a salt bridge with Glutamic acid at position 56. The difference in charge (the wild-type residue is positively charged and the mutant residue is neutral) will disturb the ionic interaction and will affect the structure and function of the NPM1 protein.

In (I59T): The amino acid Isoleucine changes to Threonine at position 59. The wild-type residue was found to be involved in a multimer contact. However, the mutant residue is too small for the wild-type residue to make multimer contacts. The mutant residue is less hydrophobic than the wild-type residue. The difference in hydrophobicity of the Isoleucine and Threonine might lead to a loss of hydrophobic interactions in the core of the proteins. The wild residue is buried in the core of a domain. The difference between the wild residue and mutant residue might disturb the core structure of this domain and it will affect protein function.

In (L79S): The amino acid Leucine changes to Serine at position 79. The mutant residue is smaller than the wild-type residue, this difference will cause an empty space in the core of the protein and might affect protein function. The Serine residue is less hydrophobic than the Leucine residue. The difference in hydrophobicity will cause a loss of hydrophobic interactions in the core of the protein. The wild-type residue is buried in the core of a domain. This mutation might disturb the core structure of the domain and will change protein structure and function.

In (P152A): The amino acid Proline changes to Alanine at position 152. Prolines are very rigid and induce a special backbone conformation. The amino acid changes from Proline to Alanine may disturb this special conformation and will affect protein structure. The mutant residue is smaller than the wild-type residue which might lead to a loss of interactions. As a special motif, the mutation is located within a stretch of residues. The differences in amino acid properties can disturb this region and its function.

In (K193R): The amino acid changes from Lysine to Arginine at position 193. The mutation is located within a stretch of residues as a special motif. The differences between wild and mutant-type residues may disturb the motif and probably affect its function. The mutant residue is bigger than the wild residue, which might lead to bumps.

In (K193N): The amino acid changes Lysine to Asparagine at position 193. The wild residue is positively charged and the mutant residue is neutral. Because of the mutation, the charge of the

wild-type residue will be lost which can cause loss of interactions with other molecules or residues. The mutant residue is smaller than the wild-type residue which will also lead to loss of interactions.

In (A283G): The amino acid Alanine changes to Glycine at position 283. The mutated residue is located in a domain that is important for the binding of other molecules. Residue Alanine changes to Glysine at this position might disturb its function. The mutation introduces Glycine at position 283. Glycines are very rigid it might fail to fulfill the required rigidity of the protein at this position. It will affect protein structure and function. The mutant residue is smaller than the wild-type residue, this might lead to a loss of interactions. The mutant residue is less hydrophobic than the wild residue. Because of hydrophobicity, the hydrophobic interactions either in the core of the protein or on the surface will be lost.

In (I284F): The amino acid Isoleucine changes to Phenylalanine at position 284. The mutated residue is located in a domain that is important for the binding of other molecules. Mutation of this residue might disturb its function. The mutant residue is bigger than the wild residue which might lead to bumps.

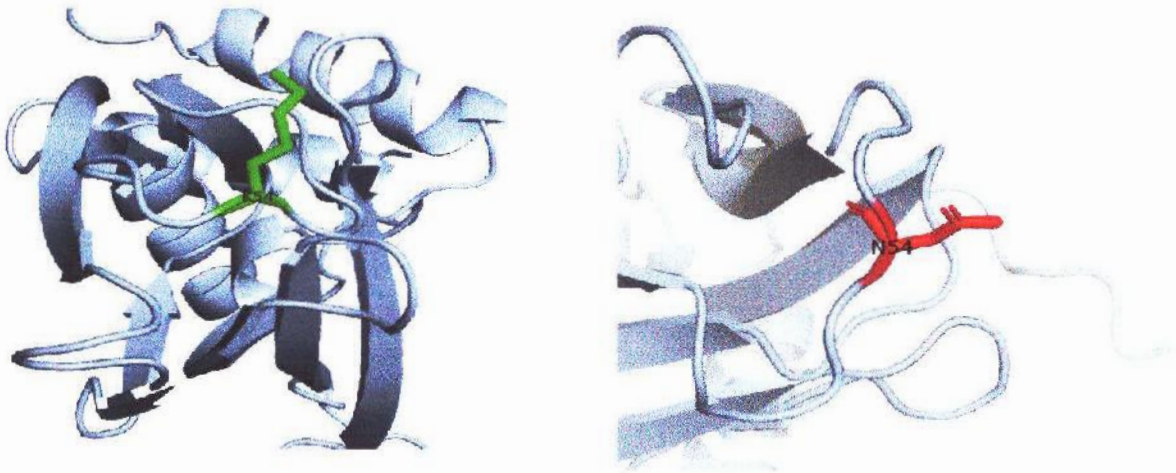


Figure 5: (K54N): The amino acid Lysine changes to Asparagine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, the red color shows the mutant one, while the light purple color shows the rest of the NPM1 protein structure.

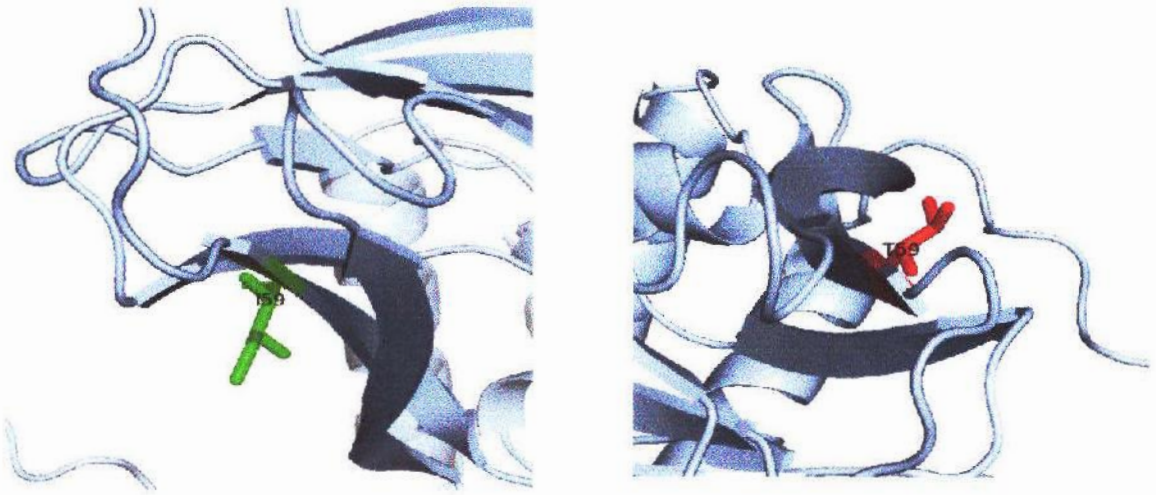


Figure 6: (I59T): The amino acid Isoleucine changes to Threonine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, the red color shows the mutant one, while the light purple color shows the rest of the NPM1 protein structure.

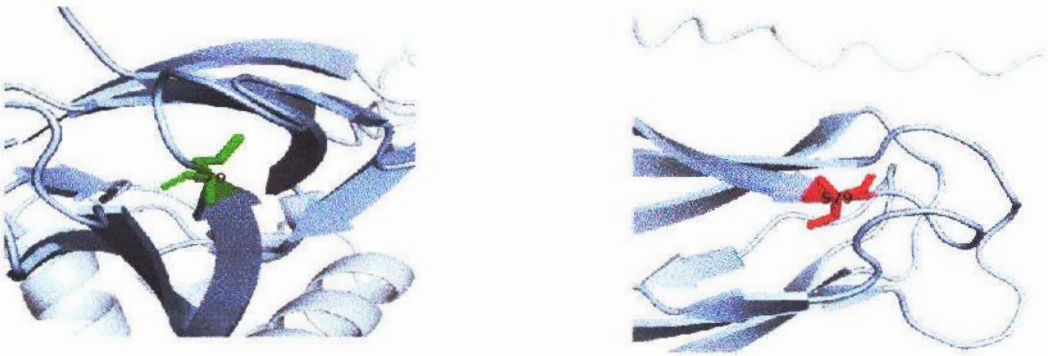


Figure 7: (L79S): The amino acid Leucine changes to Serine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, the red color shows the mutant one, and the light purple color shows the rest of the NPM1 protein structure.

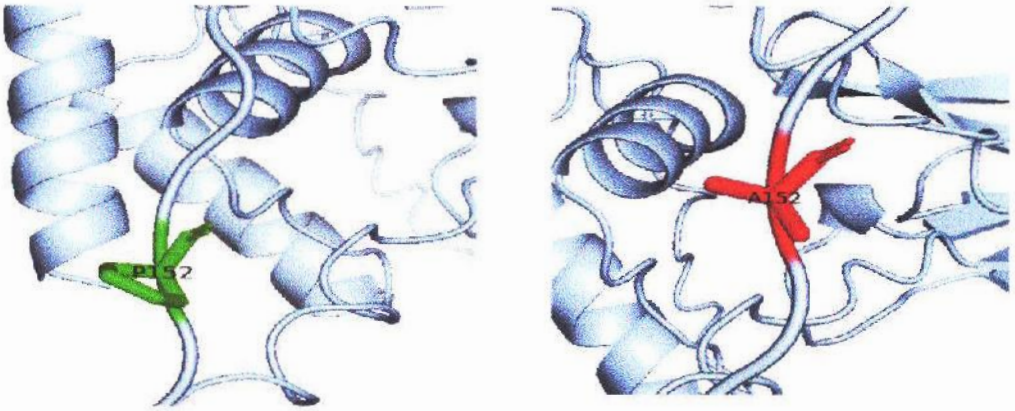


Figure 8: (P152A): The amino acid Proline changes to Alanine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, and the red color shows the mutant one, while the light purple color shows the rest of the NPM1 protein structure.

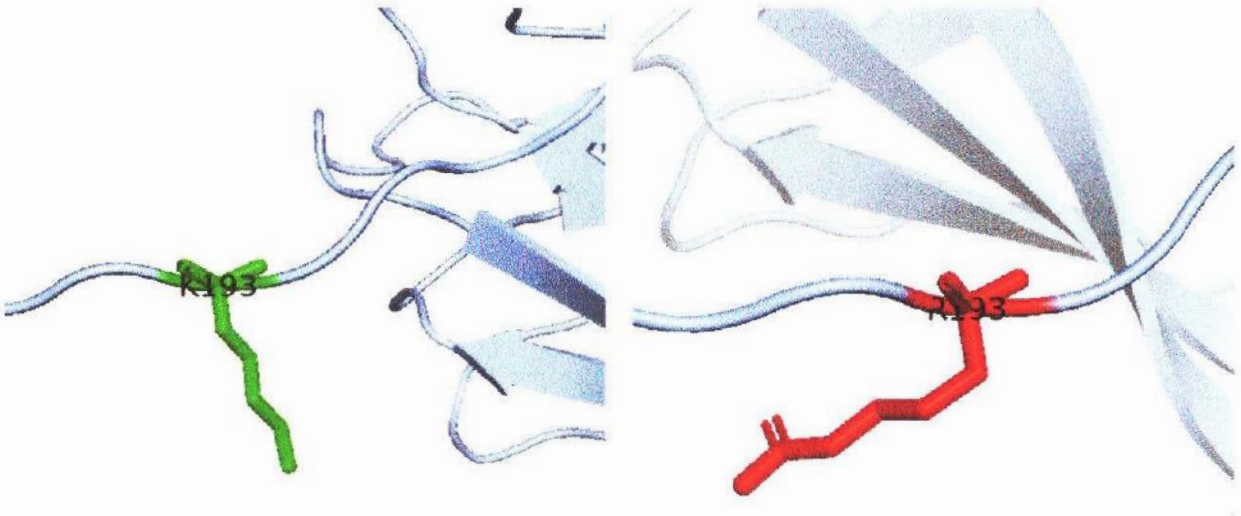


Figure 9: (K193R): The amino acid Lysine changes to Arginine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, and the red color shows the mutant one, while the light purple color shows the rest of the NPM1 protein structure.

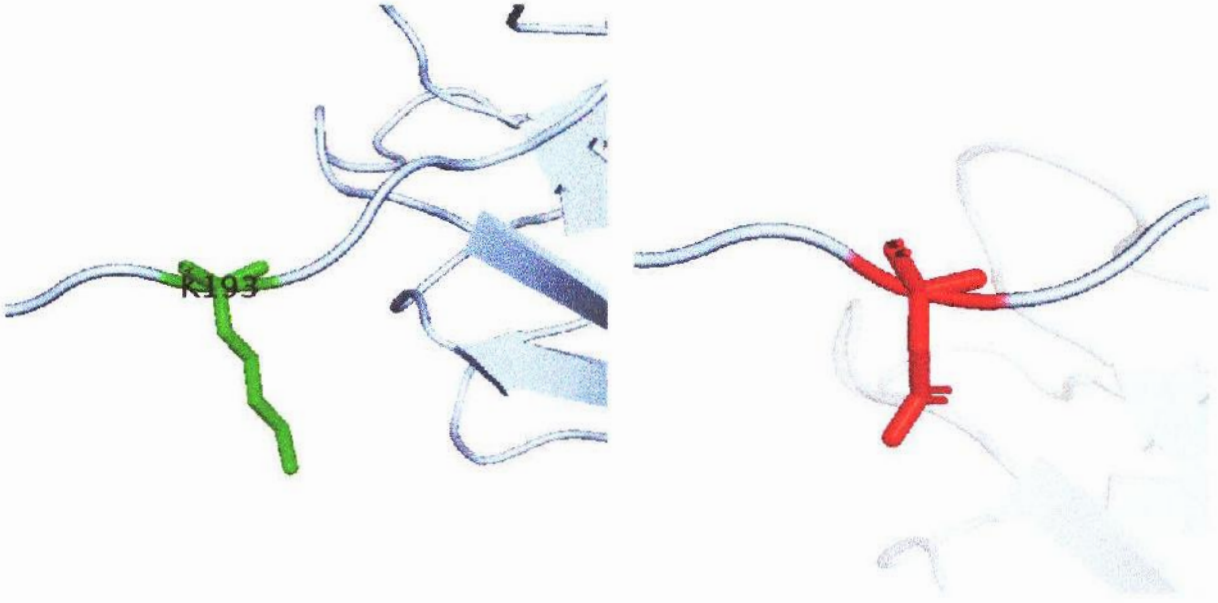


Figure 10: (K193N): The amino acid Lysine changes to Asparagine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, the red color shows the mutant one, while the light purple color shows the rest of the NPM1 protein structure.

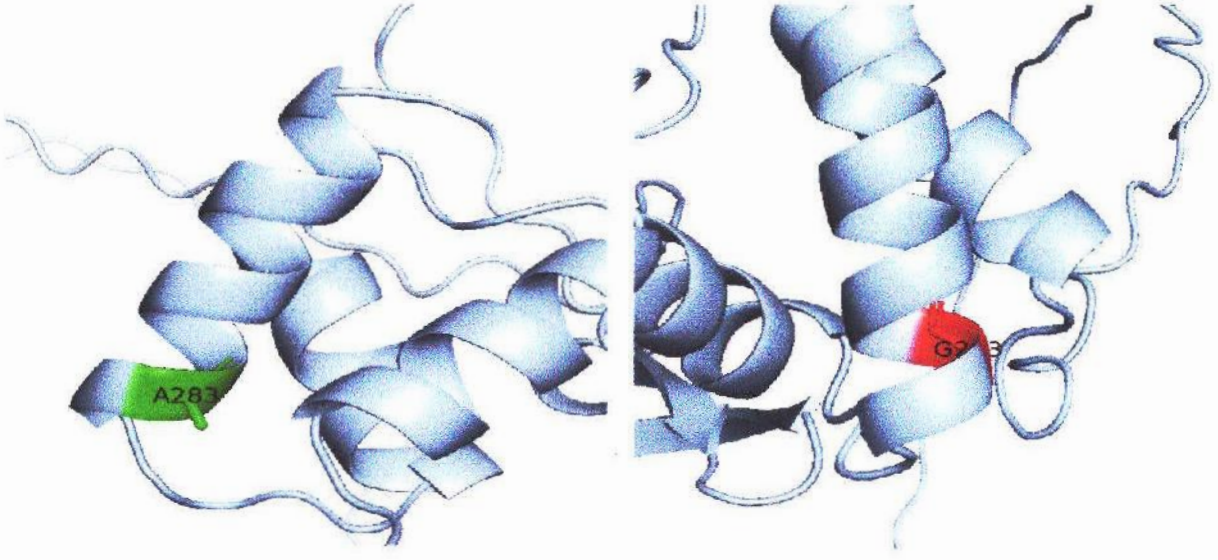


Figure 11: (A283G): The amino acid Alanine changes to Glycine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, and the red color shows the mutant one, while the light purple color shows the rest of the NPM1 protein structure.

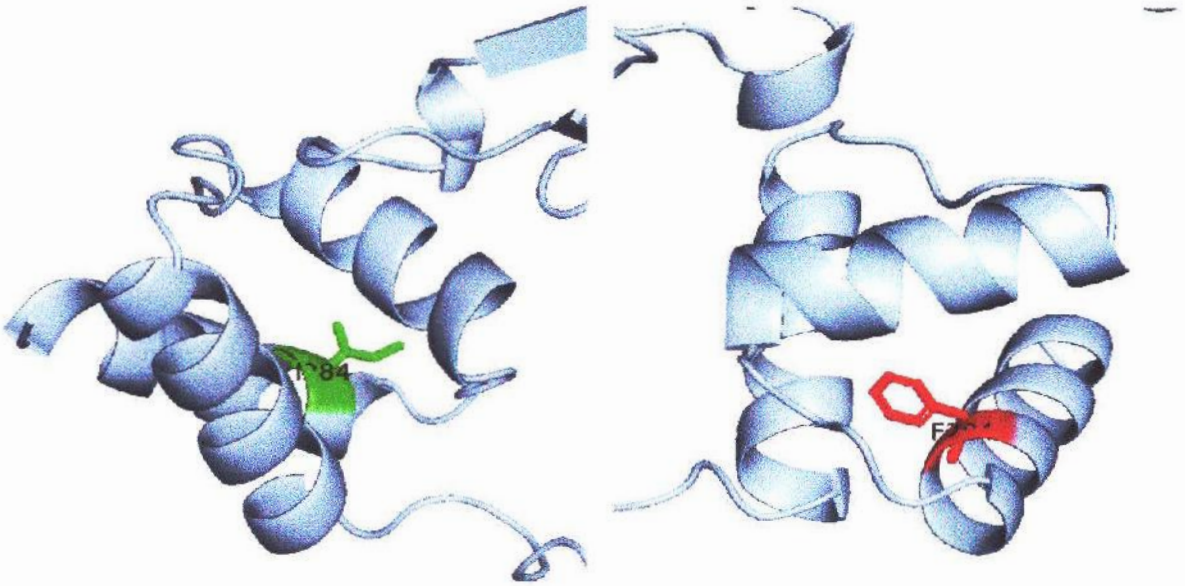


Figure 12: (I284F): The amino acid Isoleucine changes to Phenylalanine (the close-up views of residue). This visualization was done using PyMol. The green color shows the wild amino acid, the red color shows the mutant one, and the light purple color shows the rest of the NPM1 protein structure.

4.9 Protein-protein interaction analysis

For analysis of protein-protein interaction, we used the STRING database. The STRING database predicted that the human NPM1 protein interacts with ten proteins in a cell. Namely Cellular tumor antigen p53 (TP53), Nucleolin (NCL), Cyclin-dependent kinase inhibitor 2A (CDKN2A), Transcriptional repressor CTCF (CTCF), E3 ubiquitin-protein ligase Mdm2 (MDM2), DNA-(apurinic or apyrimidinic site) lyase (APEX1), ALK tyrosine kinase receptor (ALK), NOP2/Sun RNA methyltransferase family member 2 (NSUN2), Histone H3-like centromeric protein A (CENPA), Aurora kinase B (AURKB). All predicted functional partners showed a high interaction score with the NPM1 protein and all partner proteins showed the first shell of interactions. If any change occurs in this NPM1 protein, it may affect the overall protein network interaction among all those partner proteins.

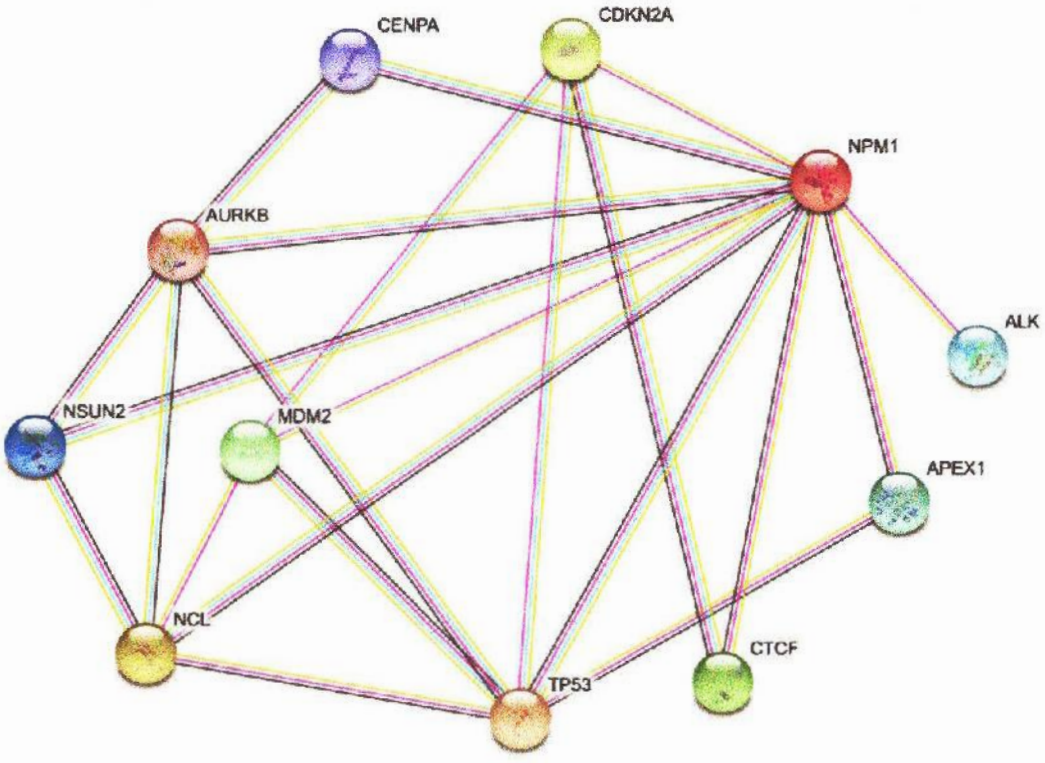


Figure 14: Protein-protein interaction network of NPM1 protein generated by STRING.

CHAPTER FIVE
DISCUSSION

5. Discussion

The human nucleophosmin1 (NPM1) gene is located on chromosome 5q35.1 and encodes 294 amino acid long residues of nucleophosmin1 protein which is a family member of nuclear chaperones[4]. Single nucleotide polymorphisms (SNPs) occur when single DNA base pairs (A, C, G, T) are altered and if this phenomenon happens in the coding region and causes amino acid changes as a result of changes in protein function this is called missense SNPs or non-synonymous[2]. SNPs in the NPM1 gene are associated with acute myeloid leukemia a type of blood cancer[3].

In our current investigation, we used different types of in-silico tools like SIFT, PROVEAN, PolyPhan-2, MutPred2, MUpro, and Consurf for the analysis of missense SNPs of the human NPM1 gene.

Firstly, we analyzed the downloaded 145 missense SNPs using sequence homology-based prediction tools SIFT and PROVEAN. Among the analyzed 145 missense SNPs 36 missense SNPs were predicted damaging or deleterious by both SIFT and PROVEAN and also nine SNPs were predicted damaging or deleterious by either SIFT or PROVEAN. A total of 45 predicted damaging or deleterious missense SNPs were further analyzed by the structure homology-based tool PolyPhen-2. PolyPhen-2 predicted nine SNPs as probably damaging means damaging effects with a high confidence score. We selected these nine SNPs for advanced analysis.

Nine shortlisted SNPs (k54N, I59T, L79S, P152A, K193R, K193N, S217L, A283G, I284F) were analyzed by I-Mutant 3.0 and MUpro for the prediction of protein stability. I-Mutant 3.0 and MUpro both predicted eight missense SNPs (k54N, I59T, L79S, P152A, K193R, K193N, A283G, I284F) as decreased stability. Therefore, we proposed that this mutation can affect protein function and may promote acute myeloid leukemia.

For the analysis of evolutionary conservation, we used the Consurf web server. The Consurf server predicted that almost all mutations were located in the highly conserved region. If any change occurs in the conserved region of a protein it could highly affect the protein structure and as well as protein function.

We used the MutPred2 server for the analysis of molecular mechanisms and their consequences on protein structure. MutPred server predicted five mutations (K54N, I59T, L79S, A283G, I284F) with the high predicted scores. These mutations cause altered metal binding, altered stability, altered disordered interface, altered order interface, altered coiled-coil, and altered transmembrane protein, and because these molecular changes may affect protein function

For the prediction of secondary structure, we used the PSIPRED tool. PSIPRED predicted most of the mutation distributed alpha-helix, beta-sheet, and coil. These can cause a huge change in protein 3D structure and ultimately affect protein function.

We predicted the NMP1 protein 3D structure by using AlphaFoldDB. AlphaFold predicted the NPM1 crystal structure of 294 amino acid residues. All physicological properties were obtained from project HOPE. project HOPE reveals because of these eight mutations size, charge, and hydrophobicity of the NPM1 protein changes between the wild structure and mutant structure of the human NPM1 protein. This will affect protein functions of the NPM1 protein and may cause acute myeloid leukemia.

From our investigation, we obtained eight missense SNPs ((k54N, I59T, L79S, P152A, K193R, K193N, A283G, I284F) of the NPM1 gene which has a deleterious effect on the human NPM1 gene and may cause acute myeloid leukemia.

CHAPTER SIX
CONCLUSION

6. Conclusion

In this study, we investigated functional and structural consequences of missense SNPs in the human NPM1 gene using in silico prediction tools. We identified eight [rs1288532134 (K54N), rs1561864425 (I59T), rs11551525 (L79S), rs868364493 (P152A), rs183724988 (K193R), rs761307473 (K193N), rs371956477 (A283G), rs1220932669 (I284F)] deleterious missense SNPs that may be associated with acute myeloid leukemia. Among these five SNPs (K54N, I59T, L79S, A283G, I284F) are highly deleterious. We strongly believe these results would be helpful for further wet-lab studies of the human NPM1 gene. These mutations can be used as diagnostic and therapeutic markers for acute myeloid leukemia and potential to develop possible drugs.

CHAPTER SEVEN

REFERENCES

7. References

1. Wang, Z. & Moult, J. SNPs, protein structure, and disease. *Hum. Mutat.* **17**, 263–270 (2001).
2. Ramensky, V. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* **30**, 3894–3900 (2002).
3. Falini, B. *et al.* Acute myeloid leukemia with mutated nucleophosmin (NPM1): is it a distinct entity? *Blood* **117**, 1109–1120 (2011).
4. Falini, B. Acute Myeloid Leukemia with Mutated Nucleophosmin (NPM1): Molecular, Pathological, and Clinical Features. in *Acute Myelogenous Leukemia* (ed. Nagarajan, L.) vol. 145 149–168 (Springer New York, 2009).
5. Meani, N. & Alcalay, M. Role of nucleophosmin in acute myeloid leukemia. *Expert Rev. Anticancer Ther.* **9**, 1283–1294 (2009).
6. Hindley, A., Catherwood, M. A., McMullin, M. F. & Mills, K. I. Significance of NPM1 Gene Mutations in AML. *Int. J. Mol. Sci.* **22**, 10040 (2021).
7. Falini, B., Nicoletti, I., Martelli, M. F. & Mecucci, C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood* **109**, 874–885 (2007).
8. Dai, L. *et al.* [The Mutated Gene Sequenced by NGS and the Correlation of Their Coexistent Mutual Exclusiveness in NPM1 Mutated Acute Myeloid Leukemia]. *Zhongguo Shi Yan Xue Ye Xue Za Zhi* **29**, 1733–1740 (2021).
9. The International SNP Map Working Group *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
10. Kumar, S., Banks, T. W. & Cloutier, S. SNP Discovery through Next-Generation Sequencing and Its Applications. *Int. J. Plant Genomics* **2012**, 1–15 (2012).
11. Deng, N., Zhou, H., Fan, H. & Yuan, Y. Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget* **8**, 110635–110649 (2017).
12. de Alencar, S. A. & Lopes, J. C. D. A Comprehensive *In Silico* Analysis of the Functional and Structural Impact of SNPs in the *IGF1R* Gene. *J. Biomed. Biotechnol.* **2010**, 1–8 (2010).
13. George Priya Doss, C. *et al.* Identification and structural comparison of deleterious mutations in

nsSNPs of ABL1 gene in chronic myeloid leukemia: A bio-informatics study. *J. Biomed. Inform.* **41**, 607–612 (2008).

14. Rajasekaran, R., Sudandiradoss, C., Doss, C. G. P. & Sethumadhavan, R. Identification and in silico analysis of functional SNPs of the BRCA1 gene. *Genomics* **90**, 447–452 (2007).
15. Dakal, T. C. *et al.* Predicting the functional consequences of non-synonymous single nucleotide polymorphisms in IL8 gene. *Sci. Rep.* **7**, 6525 (2017).
16. Kosaloglu, Z. *et al.* In silico SNP analysis of the breast cancer antigen NY-BR-1. *BMC Cancer* **16**, 901 (2016).
17. Ali Mohamoud, H. S. *et al.* First Comprehensive *In Silico* Analysis of the Functional and Structural Consequences of SNPs in Human *GalNAc-T1* Gene. *Comput. Math. Methods Med.* **2014**, 1–15 (2014).
18. Das, S. C., Rahman, Md. A. & Das Gupta, S. In-silico analysis unravels the structural and functional consequences of non-synonymous SNPs in the human IL-10 gene. *Egypt. J. Med. Hum. Genet.* **23**, 10 (2022).
19. Lim, S. W. *et al.* Functional and structural analysis of non-synonymous single nucleotide polymorphisms (nsSNPs) in the MYB oncoproteins associated with human cancer. *Sci. Rep.* **11**, 24206 (2021).
20. Mustafa, M. I., Murshed, N. S., Abdelmoneim, A. H. & Makhawi, A. M. In silico analysis of the functional and structural consequences of SNPs in human ARX gene associated with EIEE1. *Inform. Med. Unlocked* **21**, 100447 (2020).
21. Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
22. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* **4**, 1073–1081 (2009).
23. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747 (2015).
24. Choi, Y., Sims, G. E., Murphy, S., Miller, J. R. & Chan, A. P. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE* **7**, e46688 (2012).
25. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, (2013).

26. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**, W306–W310 (2005).
27. Cheng, J., Randall, A. & Baldi, P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins* **62**, 1125–1132 (2006).
28. Pejaver, V. *et al.* Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **11**, 5918 (2020).
29. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. & Ben-Tal, N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.* **38**, W529–W533 (2010).
30. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
31. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* **50**, D439–D444 (2022).
32. Venselaar, H., te Beek, T. A., Kuipers, R. K., Hekkelman, M. L. & Vriend, G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* **11**, 548 (2010).
33. Szklarczyk, D. *et al.* STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

Annexure

Raw data missense SNPs downloaded from Ensembl database.

Variant ID	Chromosome	Location on Chromosome	Alternative Alleles	Alleles	Variant Class	Source	Consequence Type	Amino Acids	AA Coordinates
rs136389784	5	170814963	T	C/T	SNP	dbSNP	missense variant	S/L	4
rs129902051	5	170814965	G	A/G	SNP	dbSNP	missense variant	M/V	5
rs772365842	5	170814971	G	A/G	SNP	dbSNP	missense variant	M/V	7
rs138975390	5	170814974	A	G/A	SNP	dbSNP	missense variant	D/N	8
rs925409148	5	170814977	G	A/G	SNP	dbSNP	missense variant	M/V	9
rs134968067	5	170814979	A	G/A/C	SNP	dbSNP	missense variant	M/I	9
rs134968067	5	170814979	C	G/A/C	SNP	dbSNP	missense variant	M/I	9
rs123554236	5	170814980	G	A/G	SNP	dbSNP	missense variant	S/G	10
rs138421481	5	170814981	C	G/C	SNP	dbSNP	missense variant	S/T	10
rs127811112	5	170814983	T	C/T	SNP	dbSNP	missense variant	P/S	11
rs768919750	5	170814990	A	G/A	SNP	dbSNP	missense variant	R/K	13
rs144640136	5	170815004	T	C/T	SNP	dbSNP	missense variant	L/F	18
rs116220460	5	170817072	A	G/A	SNP	dbSNP	missense variant	D/N	26
rs995892226	5	170817078	A	G/A	SNP	dbSNP	missense variant	D/N	28
rs759228926	5	170817093	A	G/A	SNP	dbSNP	missense variant	V/M	33

rs766975341	5	170817100	G	A/G	SNP	dbSNP	missense variant	N/S	35
rs1324989400	5	170817110	G	T/G	SNP	dbSNP	missense variant	N/K	38
rs1283447658	5	170817133	T	C/T	SNP	dbSNP	missense variant~splice region variant	T/M	46
rs372208958	5	170818325	A	G/A/T	SNP	dbSNP	missense variant	G/D	52
rs372208958	5	170818325	T	G/A/T	SNP	dbSNP	missense variant	G/V	52
rs774730997	5	170818328	T	C/T	SNP	dbSNP	missense variant	A/V	53
rs1288532134	5	170818332	C	G/C	SNP	dbSNP	missense variant	K/N	54
rs1003292380	5	170818345	G	A/G	SNP	dbSNP	missense variant	I/V	59
rs1561864425	5	170818346	C	T/C	SNP	dbSNP	missense variant	I/T	59
rs142102985	5	170818347	G	T/C/G	SNP	dbSNP	missense variant	I/M	59
rs11551526	5	170818361	T	C/T	SNP	dbSNP	missense variant	A/V	64
rs11551576	5	170818363	G	A/G	SNP	dbSNP	missense variant	M/V	65
rs763566520	5	170818372	A	G/A	SNP	dbSNP	missense variant	E/K	68
rs11551525	5	170818406	C	T/C	SNP	dbSNP	missense variant	L/S	79
rs178519444	5	170818408	G	A/G	SNP	dbSNP	missense variant	K/E	80
rs1438414399	5	170818417	A	G/A	SNP	dbSNP	missense variant	V/I	83
rs147806769	5	170818427	T	C/T	SNP	dbSNP	missense variant~splice region variant	T/M	86

rs756323205	5	170818743	T	G/T	SNP	dbSNP	missense variant	V/L	98
rs954187506	5	170818747	G	T/G	SNP	dbSNP	missense variant	V/G	99
rs989977568	5	170818782	G	A/G	SNP	dbSNP	missense variant	I/V	111
rs1467958614	5	170818785	G	A/G	SNP	dbSNP	missense variant	S/G	112
rs1315698538	5	170819716	A	G/A	SNP	dbSNP	missense variant~splice region variant	V/M	119
rs941656814	5	170819721	T	G/T	SNP	dbSNP	missense variant	E/D	120
rs587778580	5	170819723	G	A/G	SNP	dbSNP	missense variant	E/G	121
rs1366001140	5	170819728	C	G/C	SNP	dbSNP	missense variant	A/P	123
rs1425509549	5	170819729	T	C/T	SNP	dbSNP	missense variant	A/V	123
rs1030167696	5	170819731	C	G/C	SNP	dbSNP	missense variant	E/Q	124
rs1308264050	5	170819740	A	G/A	SNP	dbSNP	missense variant	D/N	127
rs1360689654	5	170819755	A	G/A	SNP	dbSNP	missense variant	D/N	132
rs751882247	5	170819764	T	C/T	SNP	dbSNP	missense variant	L/F	135
rs747866671	5	170819773	C	A/C/G/T	SNP	dbSNP	missense variant	I/L	138
rs747866671	5	170819773	G	A/C/G/T	SNP	dbSNP	missense variant	I/V	138
rs747866671	5	170819773	T	A/C/G/T	SNP	dbSNP	missense variant	I/L	138
rs769397977	5	170819775	G	A/G	SNP	dbSNP	missense variant	I/M	138
rs1417200112	5	170819780	C	G/C	SNP	dbSNP	missense variant	G/A	140

rs1177989112	5	170819788	C	T/C	SNP	dbSNP	missense variant	S/P	143
rs375558588	5	170819789	G	C/G	SNP	dbSNP	missense variant	S/C	143
rs934166729	5	170819792	T	C/T	SNP	dbSNP	missense variant	A/V	144
rs1208296966	5	170819794	T	C/T	SNP	dbSNP	missense variant	P/S	145
rs1177473523	5	170819795	T	C/T	SNP	dbSNP	missense variant	P/L	145
rs770789271	5	170819798	A	G/A	SNP	dbSNP	missense variant	G/E	146
rs868364493	5	170819815	G	C/G	SNP	dbSNP	missense variant	P/A	152
rs571071150	5	170819819	G	A/G	SNP	dbSNP	missense variant~splice region variant	Q/R	153
rs1454882046	5	170819923	C	A/C	SNP	dbSNP	missense variant	K/N	155
rs1561865755	5	170819930	G	C/G	SNP	dbSNP	missense variant	L/V	158
rs775291597	5	170819933	C	G/C	SNP	dbSNP	missense variant	A/P	159
rs748317624	5	170819937	T	C/T	SNP	dbSNP	missense variant	A/V	160
rs770013862	5	170819939	A	G/A	SNP	dbSNP	missense variant	D/N	161
rs762935997	5	170819947	G	T/C/G	SNP	dbSNP	missense variant	D/E	163
rs146936311	5	170819953	A	C/A/G/T	SNP	dbSNP	missense variant	D/E	165
rs146936311	5	170819953	G	C/A/G/T	SNP	dbSNP	missense variant	D/E	165
rs774697729	5	170819954	A	G/A	SNP	dbSNP	missense variant	D/N	166
rs147629377	5	170819956	G	T/C/G	SNP	dbSNP	missense variant	D/E	166

rs1207998618	5	170819957	C	G/C/T	SNP	dbSNP	missense variant	D/H	167
rs1207998618	5	170819957	T	G/C/T	SNP	dbSNP	missense variant	D/Y	167
rs1449984649	5	170819963	A	G/A	SNP	dbSNP	missense variant	E/K	169
rs767945148	5	170819965	T	A/T	SNP	dbSNP	missense variant	E/D	169
rs1561865927	5	170819968	C	G/C	SNP	dbSNP	missense variant	E/D	170
rs1381683051	5	170819969	A	G/A	SNP	dbSNP	missense variant	D/N	171
rs750633648	5	170819970	T	A/T	SNP	dbSNP	missense variant	D/V	171
rs753126733	5	170819971	G	T/G	SNP	dbSNP	missense variant	D/E	171
rs1187263636	5	170819979	G	A/G	SNP	dbSNP	missense variant	E/G	174
rs1554135991	5	170819981	C	G/C	SNP	dbSNP	missense variant~splice region variant	D/H	175
rs1476856200	5	170819982	C	A/C	SNP	dbSNP	missense variant~splice region variant	D/A	175
rs747005016	5	170827157	A	T/A	SNP	dbSNP	missense variant~splice region variant	D/E	175
rs754744493	5	170827160	A	T/A	SNP	dbSNP	missense variant	D/E	176
rs1028395248	5	170827161	A	G/A	SNP	dbSNP	missense variant	D/N	177
rs267600552	5	170827164	A	G/A/C	SNP	dbSNP	missense variant	D/N	178
rs267600552	5	170827164	C	G/A/C	SNP	dbSNP	missense variant	D/H	178
rs17856513	5	170827165	G	A/G/T	SNP	dbSNP	missense variant	D/G	178

rs1431020185	5	170827887	C	G/C	SNP	dbSNP	missense variant	Q/H	209
rs373068340	5	170827900	C	T/C	SNP	dbSNP	missense variant	S/P	214
rs1561871583	5	170827910	T	C/T	SNP	dbSNP	missense variant	S/L	217
rs1437600710	5	170832307	C	G/C	SNP	dbSNP	missense variant~splice region variant	G/A	224
rs758630739	5	170832310	G	A/G	SNP	dbSNP	missense variant	Q/R	225
rs1458436735	5	170832312	A	G/A	SNP	dbSNP	missense variant	E/K	226
rs766707224	5	170832314	C	A/C/G	SNP	dbSNP	missense variant	E/D	226
rs1388726393	5	170832315	G	T/G	SNP	dbSNP	missense variant	S/A	227
rs755065911	5	170832336	G	A/G/T	SNP	dbSNP	missense variant	T/A	234
rs755065911	5	170832336	T	A/G/T	SNP	dbSNP	missense variant	T/S	234
rs781171336	5	170832337	T	C/T	SNP	dbSNP	missense variant	T/I	234
rs1053217290	5	170832340	T	C/T	SNP	dbSNP	missense variant	P/L	235
rs374517078	5	170832343	C	A/C	SNP	dbSNP	missense variant	K/T	236
rs1246599848	5	170832346	T	C/T	SNP	dbSNP	missense variant	T/I	237
rs756636316	5	170832348	A	C/A/G	SNP	dbSNP	missense variant	P/T	238
rs756636316	5	170832348	G	C/A/G	SNP	dbSNP	missense variant	P/A	238
rs771340939	5	170832357	T	C/T	SNP	dbSNP	missense variant	P/S	241
rs138586088	5	170832360	G	A/G	SNP	dbSNP	missense variant	S/G	242

rs7715684 44	5	170832366	A	G/A/T	SNP	dbSNP	missense variant	V/I	244
rs7715684 44	5	170832366	T	G/A/T	SNP	dbSNP	missense variant	V/L	244
rs1432096 17	5	170832369	C	G/C	SNP	dbSNP	missense variant	E/Q	245
rs7605989 71	5	170832372	T	G/T	SNP	dbSNP	missense variant	D/Y	246
rs3773486 62	5	170832382	T	C/T	SNP	dbSNP	missense variant	A/V	249
rs1412238 021	5	170832391	G	A/G	SNP	dbSNP	missense variant	Q/R	252
rs1167243 022	5	170832397	A	G/A	SNP	dbSNP	missense variant	S/N	254
rs1155153 2	5	170832398	A	T/A	SNP	dbSNP	missense variant	S/R	254
rs7767802 44	5	170832399	G	A/G	SNP	dbSNP	missense variant	I/V	255
rs7656546 91	5	170834720	G	A/G	SNP	dbSNP	missense variant	K/R	263
rs1351738 253	5	170834728	A	G/A	SNP	dbSNP	missense variant	A/T	266
rs7668088 81	5	170834732	G	A/G	SNP	dbSNP	missense variant	K/R	267
rs9445758 47	5	170834739	G	C/G/T	SNP	dbSNP	missense variant	I/M	269
rs7547007 41	5	170834741	G	A/G	SNP	dbSNP	missense variant	N/S	270
rs1785143 1	5	170834744	G	A/G	SNP	dbSNP	missense variant	Y/C	271
rs7476667 52	5	170834746	T	G/T	SNP	dbSNP	missense variant	V/L	272
rs7775501 67	5	170834762	A	G/A	SNP	dbSNP	missense variant	R/Q	277
rs7579637 72	5	170834764	G	A/G	SNP	dbSNP	missense variant	M/V	278