IN-SILICO CHARACTERISATION OF PATHOGENIC MISSENSE SNPS IN HUMAN FLT3 GENE

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

MASTER OF SCIENCE

in

Biotechnology

by

Vandana
(Roll No. 24/Mscbio/56)

Under the Supervision of

Prof. Yasha Hasija

Head of Department

Department of Biotechnology

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Delhi, India

May, 2026

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Vandana hereby certify that the work which is being presented in the thesis entitled "**In-silico Characterisation of Pathogenic Missense SNPs in Human FLT3 Gene"** in partial fulfillment of the requirements for the award of the Masters of Science, submitted in the Department of Biotechnology, Delhi Technological University is an authentic record of my own work carried out during the period from 2024 to 2026 under the supervision of Prof Yasha Hasija.

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

Candidate's Signature

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Vandana** (24/Mscbio/56) has carried out their search work presented in this thesis entitled **"In-silico Characterisation of Pathogenic Missense SNPs in Human FLT3 Gene"** for the award of **Master of Science in Biotechnology** from Department of Biotechnology, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself/herself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Signature

Prof. Yasha Hasija

Head, Department of Biotechnology

DTU, Delhi

Date:

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42

ACKNOWLEDGEMENT

I would like to convey my sincere gratitude to Professor Yasha Hasija of the Biotechnology Department, Delhi Technological University for providing me with the chance to conduct this research and for her helpful advice. Her enthusiasm, vision, sincerity, and drive have greatly motivated to me. It was a great honour and joy to be able to work and study under her leadership. I am immensely grateful for everything she has given me. She has always been an inspiration to me. I also like to express my gratitude to all the teaching and support personnel at Delhi Technological University's Biotechnology Department for their direct and/or indirect contributions to the project.

Vandana

IN-SILICO CHARACTERISATION OF PATHOGENIC MISSENSE SNPS IN HUMAN FLT3 GENE

Vandana

Abstract

FLT3 gene (FMS like tyrosine kinase): encodes a receptor tyrosine kinase protein that regulates the differentiation of stem and progenitor cells , and genetic alteration in this gene often causes acute myeloid leukaemia and autoimmune conditions like rheumatoid arthritis. Non-synonymous single-nucleotide polymorphisms (nsSNPs), or missense variants, can critically change the function of the FLT3 protein by keeping it constitutively active, potentially contributing to disease development. This study utilized a in-silico approach to evaluate the functionally deleterious nsSNPs in the human FLT3 gene. A dataset of 1,444 missense variants was sourced from theNCBI dbSNP database, and five tools (SIFT, PolyPhen-2, FATHMM, SNPs3D, RegulomeDB) were used for initial screening. High-risk variants were further validated using MutPred2, CADD, and ClinPred to confirm their deleterious potential. The impact on protein stability and evolutionary conservation was checked using I-Mutant 2.0 and ConSurf. Ten nsSNPs (F349L, A680V, A814S, D829E, D835E, D835V, I836M, Y842C, R849H, I881T) were consistently predicted as deleterious and validated as pathogenic. I-Mutant 2.0 predicted mutations out of which 7 destabilise protein and rest stabilise, and ConSurf analysis revealed that all except F349L are located at highly conserved residues, indicating their critical structural or functional roles. The ten potentially damaging, high-risk nsSNPs in FLT3 identified in this work are likely to alter the functioning and structure of proteins, offering a solid basis for further experimental verification and supporting the creation of tailored treatments for disorders linked to FLT3.

Table of Content

List of Figures

List of Tables

List of symbols and abbreviations

Abbreviation	Full Form
AML	Acute Myeloid Leukemia
CADD	Combined Annotation Dependent Depletion
ClinPred	Clinical Predictor
ConSurf	Conservation Surface
ΔΔG	Change in Gibbs free energy
ERRAT	Error Rating
FATHMM	Functional Analysis Through Hidden Markov Models
FLT3	FMS-like tyrosine kinase 3
GMQE	Global Model Quality Estimation
IKBKG	Inhibitor of NF-κB Kinase Subunit Gamma
IL	Interleukin
ITD	Internal Tandem Duplication
MutPred2	Mutation Predictor 2
NCBI	National Center for Biotechnology Information
NF-κB	Nuclear Factor Kappa B
nsSNP	Non-synonymous Single Nucleotide Polymorphism
PI3K	Phosphoinositide 3-kinase
PROCHECK	Protein Structure Validation Tool
PTM	Post Translational Modification
PyMOL	Python Molecular Graphics
RA	Rheumatoid Arthritis

WT	Wildtype
SNP	Single Nucleotide Polymorphism
SNPs3D	Single Nucleotide Polymorphisms 3D

CHAPTER 1

INTRODUCTION

1.1 Background

Comparative genomics Hematopoiesis is a biological process that allows blood cells to be formed and developed from Hematopoietic Stem cells. Several signaling pathways and growth factor receptors regulate this process and are involved in cell proliferation, differentiation and survival. These signaling mechanisms might be abnormal, resulting in hematological abnormalities and immune related diseases. The FLT3 Gene encodes a receptor tyrosine kinase of class III, which is important in the growth and maintenance of hematopoietic progenitor cells. The activation of the FLT3 receptor triggers several downstream pathways which are related to normal blood cell development. Mutations in the FLT3 gene have been found to be closely linked to AML and are also frequently connected with poorer prognosis and abnormal cell growth. The most common variation in humans is single-nucleotide polymorphisms (SNPs). Of these, non-synonymous SNPs (nsSNPs) are those that cause a change in protein structure and/or function. These differences may be associated with protein stability, post translational modifications, and protein-protein interactions that all can impact the development of disease. As the amount of genomic information grows, computational/in-silico methods have their own value in identifying deleterious mutations. Prediction methods based on bioinformation can aid in the assessment of pathogenic effects of nsSNP and in the understanding of the consequence of its structure and function. So, detailed study of nsSNPs in FLT3 gene could help to get a better understanding of the disease mechanism and help the development of targeted therapies.

1.2 The FLT3 Gene: Structure and Function

The FLT3 (FMS-like tyrosine kinase 3) gene is situated at chromosome 13, q-arm, band 12 (cytogenetic location 13q12) on the long arm of the chromosome and it is approximately 100 kb of genomic DNA which contains 24 exons that code for a protein with a molecular weight of about 113 kDa and 993 amino acids(Rosnet et al., 1991). FLT3 belongs to the class III receptor tyrosine kinases (RTKs), which is a type I membrane receptors. The receptor is consist of five immunoglobulin like domains in the extracellular domain for binding of ligand, a transmembrane domain that tether through the cell membrane, a juxtamembrane domain that regulates autoinhibition, and a tyrosine kinase domain with 2 part (TKD1 and TKD2) that are separated by a kinase insert domain in the intracellular domain(Liu and Gu, 2024). Once FLT3L binds to FLT3, the receptor dimerizes and autophosphorylates at specific tyrosine residues such as Y589, Y591, Y599, Y768, Y955 and Y969, and activates three major downstream signaling pathways: PI3K/AKT pathway leading to cell survival; RAS/MAPK pathway leading to proliferation/differentiation; and STAT5 pathway leading to gene transcription(Rocnik et al., 2006). Biologically, FLT3 has been proven to be important in hematopoiesis, being expressed at highlevels in hematopoietic stem cells, and to promote the development of both B and T cells from lymphoid progenitors, yet dendritic cells and NK cells are also derived from FLT3-positive cells, and mature blood cells do not express FLT3. The activity of FLT3 is tightly regulated by autoinhibition via the FLT3 juxtamembrane domain, ubiquitination and degradation by CBL, dephosphorylation by SHP-1 and SHP-2 phosphatases and negative feedback by SOCS proteins, all of which can be disrupted to create aberrant FLT3 signaling and disease development(Kazi and Rönnstrand, 2019).(Griffith et al., 2004)

1.3 FLT3 in Hematopoiesis and Disease

The expression of FLT3 is highest in the immature hematopoietic cells and is downmodulated by differentiation as mentioned above. This exact expression pattern places FLT3 in a pivotal role in regulating normal blood cell development. In hematopoiesis, FLT3 signals contribute to the survival and expansion of HSCs and HPCs, which generates a continuous production of fully mature blood cells such as red blood cells, leukocytes and platelets. Dysregulation of FLT3, however, is involved in the development of disease. The best established disease that has been linked to FLT3 mutations is acute myeloid leukaemia (AML) where there are tandem repeat duplications in the JM domain or point mutations in the (TKD) of FLT3 that cause the receptor to be constitutively activated without ligand binding. This permanent activation results in the uncontrolled proliferation of immature myeloid cells that become trapped in bone marrow and blood and do not mature to functional cells(Gary

Gilliland and Griffin, 2002). In addition to cancer, FLT3 has also been linked to autoimmune diseases like rheumatoid arthritis. In RA, FLT3 signaling describes its role in providing support to survival and activation of dendritic cells and macrophages, which in turn generate proinflammatory cytokines. These cytokines are destructive to the synovial lining of the joint, leading to chronic inflammation, pain, swelling and ultimately destruction of the joint. Further, the immune cells that express FLT3 infiltrate the synovial tissue of patients with RA, thereby exacerbating the autoimmune response(Ramos et al., 2013). Therefore, although FLT3 is a key factor in normal haematopoietic development, its deregulation, such as through activating mutations in AML or inappropriate expression in autoimmune disorders, can have catastrophic pathological effects.

1.4 Non-synonymous SNPs (nsSNPs) and Their Impact

The most frequent form of genomic variation is a single-nucleotide polymorphism (SNP), which is a variation in a single base pair of the DNA. If this change is located in a coding region of a protein and leads to a change of one amino acid to another, it is called a non-synonymous SNP (nsSNP) or missense variant. Whereas synonymous SNPs do not affect the amino acid sequence, nsSNPs can have a highly significant impact on the resulting protein. An nsSNP could affect protein folding, stability, catalytic activity, substrate binding, interacting properties, depending on the location of the change and what type of change it is. But there are some nsSNPs that have no observed effect, others that may be deleterious resulting in partial or total loss of protein function, and others that may be deleterious resulting in protein gain of function (in this case a toxic function)(Yates and Sternberg, 2013). In the case of the FLT3 gene, harmful nsSNPs could prevent the receptor from being turned off even when its ligand is not present, interfering with normal signaling. It is therefore essential to identify those functionally deleterious nsSNPs in order to understand the mechanism of the disease, predict disease risk and design specific therapies.

1.5 Objectives

This study has specific aims:All missense nsSNPs identified in the human FLT3 gene in the dbSNP database were retrieved and screened.

- Comprehensive identification and validation of high-risk deleterious nsSNPs by using SIFT, PolyPhen-2, FATHMM, SNPs3D, RegulomeDB, MutPred2, CADD and ClinPred.(Bora, Kukreti and Hasija, 2023)
- For high-risk nsSNPs, prediction of stability changes (I-Mutant 2.0) and evolutionary conservation (ConSurf) of proteins.
- MusiteDeep enables the identification of PTM sites in proteins.
- SWISS-MODEL, TM-align, PyMol, ERRAT and PROCHECK were used for modeling and comparative analysis of wild-type and mutant FLT3 proteins.(Bora, Kukreti and Hasija, 2023)
- Exploration of FLT3 gene-gene interaction network using GeneMANIA.

CHAPTER 2

MATERIALS AND METHODS

2.1 Data Collection and Retrieval

2.1.1 dbSNP Database

All single-nucleotide variants (SNVs) related to the human FLT3 gene were collected from National Centre for Biotechnology Information's dbSNP repository. Detailed genomic information, including the SNP identifiers (rsIDs), chromosomal location, amino acid substitution, and global minor allele frequency, was systematically collected for downstream analysis. Additionally, the FASTA format from the NCBI protein database (AccessionID: NP004110.2) was obtained for the primary protein sequence of the human FLT3 protein. All the SNPs reported in the NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) were systematically retrieved in the present study for all the reported SNPs associated to the human FLT3 gene. The non-synonymous Coding SNPs (cSNPs) which result in an amino acid substitution are narrowed down(Smigielski et al., 2000; Bora, Kukreti and Hasija, 2023, Anon., 2026b).(Anon., 2026b)

2.1.2 Dataset of Missense Variants (1,444 nsSNPs)

A Collection of all the reported SNPs of FLT3 gene was combined from the NCBI's dbSNP database. In total, 41,570 SNPs were identified across the human FLT3 gene. Among these, 1,444 SNPs were classified as missense variants, which were further taken for in silico analysis to evaluate their potential structural and functional consequences on the FLT3 protein(Smigielski et al., 2000; Bora, Kukreti and Hasija, 2023).

2.2 In-silico Tools for Pathogenicity Prediction

To understand the functional implications of cSNPs in the FLT3 gene, we designed a stringent in-silico pipeline employing five widely used computational prediction tools: SNPs3D, PolyPhen-2(Polymorphism Phenotyping v2), SIFT (Sorting Intolerant from Tolerant), FATHMM, which utilises the Functional Analysis through Hidden Markov Models [v2.3], and RegulomeDB(Bora, Kukreti and Hasija, 2023). A dataset of 1,444 missense SNPs was collected from the NCBI dbSNP database by applying filters to include only non-synonymous variants altering the amino acid residues. These cSNPs were subjected to individual evaluation using the aforementioned tools. SIFT (Sorting Intolerant From Tolerant) is widely used for the prediction of the consequences of amino acid change on protein. The website used is(https://sift.bii.a-star.edu.sg/). It is based on the assumption that the more similar the amino acid in the two protein sequences, the more important it is for the function of that protein; a substitution at that position is likely to be deleterious. SIFT generates a score of between 0 and 1, a score of 0 or < 0.05 suggests a deleterious substitution and a score > 0.05 suggests a tolerated substitution. The algorithm works by looking for similar sequences, creating a MSA and working out the probability of each amino acid change. Deleterious variants were selected for further analysis if they had a SIFT score ≤ 0.05.(Ng and Henikoff, 2003; Kumar, Henikoff and Ng, 2009). PolyPhen-2 (Polymorphism Phenotyping v2) is a computational program that can analyse the consequence of amino acid changes on protein structure and function. The website used is (https://genetics.bwh.harvard.edu/pph2/). PolyPhen-2 uses both sequence and structure information, such as multiple sequence alignment for evolutionary conservation, 3D protein structure for structural disruption and more, including transmembrane domains, active sites, and interaction interfaces. The tool produces a score between 0 and 1, with higher scores showing a higher risk of damage. Variants are classified into three categories: "probably damaging" (0.909-1.0), "possibly damaging" (0.447-0.908), and "benign" (0.0-0.446).(Xie et al., 2024) The variants were further analyzed based on whether they were probably damaging or possibly damaging.s(Rogers et al., 2018). FATHMM (Functional Analysis Through Hidden Markov Models) is a computational program to access how the coding and non-coding genetic variants inpact protein function. The website used id (https://fathmm.biocompute.org.uk/). FATHMM is unique compared to other sequence conservation prediction programs as it also takes into account protein domain information, evolutionary constraints, and molecular function. The tool analyzes the impact of amino acid changes by estimating the probability of the change being a pathogenic change or a neutral change. The pathogenic/ damaging variant scores above 0.5 and the benign/neutral variant is 0.5 or less. The higher the score above 0.5, the more confident in the prediction of pathogenicity. Furthermore, FATHMM has a weighting system to consider the relative functional importance of protein domains that can be very useful when evaluating variants in known functional domains. Variants having a score >0.5 were considered pathogenic and further analysed.(Rogers et al., 2018). SNPs3D is a computational tool to forecast the biological impact of missense variants (nsSNPs*)* using a combination of structure and sequence

information(http://www.snps3d.org/). It utilizes a support vector machine (SVM) classifier, which uses a variety of features such as evolutionary conservation, protein structure, residue solvent accessibility, and the three-dimensional environment of the amino acid substitution. SNPs3D calculates an SVM profile score indicating a deleterious (or damaging) SNP if the score is negative, or a neutral (or benign) SNP if the score is positive. A higher negative score indicates a greater damaging potential. In addition, SNPs3D offers details of the structural environment of the mutation, including if it is within a conserved domain, active site or protein-protein interaction interface. The variants with a negative SVM profile score were selected for further analysis as deleterious, while those with positive score were excluded as benign.(Yue, Melamud and Moult, 2006). RegulomeDB is a comprehensive database and analysis tool that can predict the functional impact of genetic variation on gene regulation, such as non-coding and coding SNPs.(https://regulomedb.org/regulome-search/) RegulomeDB is a tool that goes beyond the mere analysis of protein structure and sequence conservation, incorporating data from various sources like transcription factor binding sites, DNase hypersensitivity sites, histone modification markers and expression quantitative trait loci (eQTLs) to determine if a variant is likely to affect gene regulation. The tool has a score from 1 to 7 with lower scores representing more evidence for a regulatory function. A score of 1 is the best evidence (such as a known transcription factor binding site, confirmed by direct evidence), and a score of 7 means there is no indication of a regulatory function. The 1-6 scores indicate different types of evidence like transcription factor binding, DNase footprinting or motif changes. Variants ranked 6 and below (scores 1-6 = some regulatory potential) were classified as having some regulatory potential and those ranked 7 (no known regulatory function) were considered as no regulatory potential. Only variants with a high level of evidence of regulation were prioritized for further analysis in the lower ranked variants.s(Cheema, Rosenthal and Ilyas Kamboh, 2017)(Bora, Kukreti and Hasija, 2023)

2.3 Assessment of High-Risk Variants

MutPred2, CADD, and ClinPred were employed to check the deleterious nature of the shortlisted cSNPs. These tools have variable algorithmic techniques and machine learning models to predict pathogenicity of genetic variants, and serve as an independent level of confirmation of the shortlisted variants. MutPred2 is a machine learning based tool that combines information from the genome and the molecular structure to predict the likelihood of pathogenicity of amino acid substitutions. The tool evaluates molecular mechanisms of the mutation, predicting possible changes, including metal binding; loss of allosteric sites; changes in properties of transmembrane proteins; changes in ordered interfaces; and gain or loss of post-translational modification sites. MutPred2 is designed to produce a score of between 0 and 1, with scores above 0.5 being considered disease causing/pathogenic variants, with an associated p-value for each prediction. The higher the score, the more

confidence there will be in the pathogenic classification. MutPred2 web server is (http://mutpred.mutdb.org/). Variants with MutPred2 score > 0.5 were taken as pathogenic variants and analyzed further(Thusberg, Olatubosun and Vihinen, 2011). CADD is a popular method that combines multiple annotations into one sequence score for predicting the deleteriousness of genetic variants throughout the human genome. The tool aggregates data from a variety of sources including evolutionary conservation, regulatory information, expression of transcripts, protein structure and functional genomic data. CADD scores of >15 are generally considered potentially deleterious and scores of >20 are considered as more confident of pathogenicity. The CADD web server is (https://cadd.gs.washington.edu/). Variants with a CADD PHRED score of ≥15 were considered pathogenic and were kept for further analysis(Rentzsch et al., 2019). ClinPred is a classification model that can classify pathogenic and benign variants by genomic and protein features. It is based on a random forest classifier, which was trained with a large collection of known pathogenic and non-pathogenic variants that takes into account evolutionary conservation, protein structure, splicing effects and population frequency data. The score range of ClinPred is from 0 to 1, with a score of ≥0.5 considered "disease-causing" or "pathogenic" and one of <0.5 considered "benign. High accuracy and balanced performance in both Mendelian and complex diseases.The ClinPred web server is (http://biosig.unimelb.edu.au/clinpred/). Any variant that had a score of +0.5 or higher was considered pathogenic(Alirezaie et al., 2018).

2.4 Protein Stability Analysis by I-Mutant2.0

For analysis of the deleterious effects on structural integrity of the FLT3 protein caused by missense SNPs, we utilised I-Mutant2.0, an automated prediction tool designed to estimate changes in protein stability upon single site mutations. I-Mutant 2.0 is rooted in support vector machine (SVM) learning algorithms and aids in further evaluation of the amino acid substitution by calculating Gibbs free energy (ΔΔG), by either structural data or protein sequence as input ..This tool outputs both qualitative and quantitative predictions, indicating whether the mutation increases or decreases the stability, along with a quantitative ΔΔG value. Negative ΔΔG values represent decreased protein stability, and positive values imply an increase in structural stability. In this study, the FASTA format primary protein sequence, together with specific amino acid mutations derived from selected deleterious SNPs, was submitted to the I-Mutant 2.0 server. Mutations predicted to reduce protein stability were prioritised for further analysis, as destabilising effects are more likely to impair protein function and may play a contributory role in disease progression.(Capriotti, Fariselli and Casadio, 2005, Anon., 2026e)

2.5 Evolutionary Conservation Analysis

The conservation profile of amino acids in the human FLT3 protein was determined by the ConSurf.The website used is (https://consurf.tau.ac.il/).The evolutionary relationships among protein sequences are estimated by ConSurf based on an empirical Bayesian inference model. that gives reliable estimates of site-specific substitution rates. The tool will identify the protein homologues in protein databases (UNIREF or SWISSPROT) and generate a multiple sequence alignment to calculate the rate of evolution of each position by using the program Rate4Site. Amino acid residues are given a ConSurf conservation score of 1–9. Residues 1-4 are considered 'variable' and substitutions are evolutionarily tolerated. There is a moderate level of conservation for scores 5 to 6. The scores 7-9 are considered highly conserved residues as mutations at these positions will probably have a deleterious effect. The tool also identifies if the residue is exposed on the protein surface or on the core of the protein. Highly conserved and exposed residues are likely to be disruptive to the protein's functionality, while highly conserved and buried residues are likely to be disruptive to the folding or destabilize the protein. With this analysis, conserved residues of FLT3 were identified and selected to be further analyzed structurally and functionally(Ashkenazy et al., 2016; Ben Chorin et al., 2020).

2.6 Post-Translational Modification Prediction (MusiteDeep)

Post-translational modifications (PTMs) play an important role in regulating various cellular functions including signal transduction, protein stability, and metabolic regulation. The prediction of the PTM sites in the FLT3 protein was carried out on the MusiteDeep web server (http://www.musite.net/)MusiteDeep is a large scale experimentally validated PTM deep learning based tool that is able to capture complex sequence patterns for accurate prediction of PTM sites. Phosphorylation was chosen as the type of PTM because of its importance in FLT3 signalling and activation. The wild type FLT3 protein sequence was submitted to the site and residues with a probability higher than the threshold were listed as possible FLT3 phosphorylation sites. In particular, the authors looked to see if any of the 10 high risk nsSNPs fell in predicted PTM sites(Wang et al., 2020).

2.7 Structural Modelling and Validation

The predicted high-risk nsSNPs were further analyzed for their structural impact using multiple computational tools for 3D modeling, structural comparison, and model validation. SWISS-MODEL is a homology modeling server used to generate three dimensional structures of proteins. It works by identifying suitable template structures from the Protein Data Bank (PDB) based on sequence similarity. The quality of each generated model is assessed using the Global Model Quality Estimation (GMQE) score, which ranges from 0 to 1, where higher values indicate greater reliability. In this study, SWISS-MODEL was used to generate 3D structures for both wildtype and mutant FLT3 proteins(Waterhouse et al., 2018a; 2018b; Wang et al., 2020). TM-align is a structural alignment tool that compares two protein structures and calculates a TM-score (Template Modeling score). TM-score ranges from 0 to 1, where a score close to 1 indicates nearly identical structures, scores between 0.5 and 1.0 suggest the proteins share the same overall fold, and scores below 0.3 indicate random structural similarity. This tool was used to compare structural similarity between wildtype and each mutant FLT3 model(Zhang and Skolnick, 2005)(Anon., 2026d). PyMOL is a molecular visualization system used for structural analysis and comparison. It allows superimposition of multiple protein structures and calculates the Root Mean Square Deviation (RMSD), which measures the average distance between aligned atoms. Lower RMSD values indicate higher structural similarity . PyMOL was used to visualize and calculate RMSD between wildtype and mutant FLT3 models(Anon., n.d.). ERRAT is a protein structure validation tool that analyzes the statistics of nonbonded atomic interactions within a protein model. It calculates an overall quality factor, where scores above 50 indicate high quality models. ERRAT was used to assess the overall structural quality of all generated FLT3 models(Anon., 2026a, Anon., 2026b).PROCHECK validates protein stereochemistry by generating a Ramachandran plot, which shows the distribution of phi and psi backbone angles of amino acid residues. Residues falling in "most favored regions" indicate good stereochemical quality, while residues in "disallowed regions" suggest structural problems. PROCHECK was used to evaluate the backbone conformation quality of all FLT3 models(Anon., 2026c).

2.8 Protein-Protein Interaction Network (GeneMANIA)

Gene-gene interaction analysis was used to analyze the FLT3 functional network with the help of the GeneMANIA prediction server, which is available at (https://genemania.org/) GeneMANIA is a web based application that searches for

genes that are functionally associated to a query gene using several types of biological data. It is based on a vast database of genomics and proteomics information for prediction of gene function, analysis of gene lists and prioritisation of genes for functional assays. It combines several types of interaction data, including co-expression, co-localisation, physical interactions, shared protein domains, genetic interactions, and pathway data. GeneMANIA creates a network of interacting genes and confidence scores for each interaction, using these various data sources. The server then displays the query gene and its neighbors – genes that are related in function – so that the researchers can better understand the "neighborhood" of their gene of interest in the biological context(Warde-Farley et al., 2010).For this study, FLT3 was used as a query gene to search for its potential interacting partners and the biological pathways associated with FLT3. The analysis was targeted in particular to the genes related to immune regulation, signal transduction, and hematopoietic development. A network of interactions was generated to help identify how FLT3 mutations may interact with other pathways in the cell, such as those that may play a role in the pathogenesis of rheumatoid arthritis (RA).

CHAPTER 3

RESULTS

3.1 cSNPs Retrieval

A total of 41,570 cSNPs in the FLT3 gene were extracted from the dbSNP database, including those located in protein-coding, non-coding, and transcriptional regulatory elements. Of these, the majority were mapped to intronic regions (73%), followed by upstream (9%) and downstream (7%) gene variants. A smaller proportion of variants were located within the coding sequence of the gene. Among the coding region SNPs (cSNPs), missense variants constituted the largest category (69%), followed by synonymous (24%), stop-gained (4%), and frameshift (2%) mutations (Fig. 1). In this study, we focused exclusively on non-synonymous variants, particularly the missense substitutions, due to their potential to change the primary protein sequences and impact the structural topology of the FLT3 protein. After applying this filter, 1,444 missense SNPs were shortlisted for further computational analyses

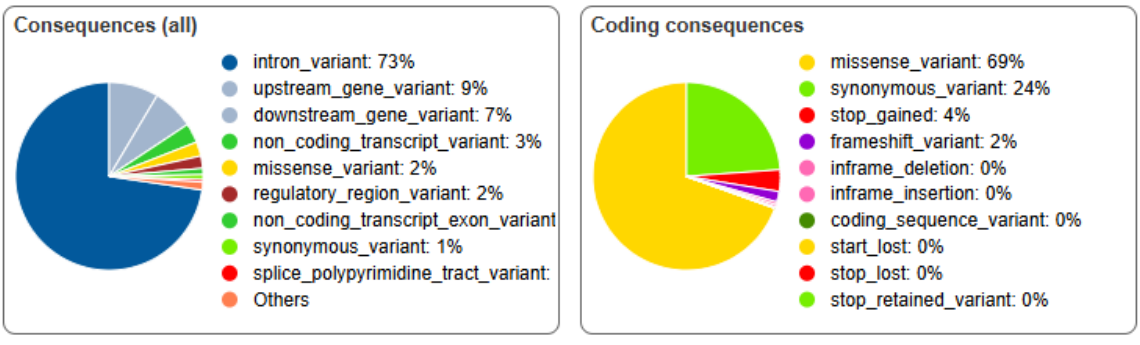


Fig. 1. Mapping Profile of SNPs in the human FLT3 gene regions

3.2 Identification of Deleterious cSNPs

Five SNP computational prediction tools: PolyPhen-2, SIFT, SNPs3D, FATHMM, and RegulomeDB, were employed to analyse deleterious SNPs that may significantly alter the biological structure integrity of the FLT3 protein. Out of 1,444 missense cSNPs analysed, six variants (Y842C, D835E, D835V, D829E, I836M, and I881T) were classified to be pathogenic across all 5 bioinformatics tools. Additionally, four more variants (A680V, F349L, A814S, and R849H) were classified as pathogenic by 4 of the 5 tools. Hence, 10 cSNPs were classified as deleterious and subsequently designated as high-risk deleterious variants (Table I). Notably, kinase domain mutations such as D835E and D835V, previously associated with constitutive FLT3 activation in leukaemias and the activation loop variant I881T were prioritised due to their potential involvement in arthritis-related signalling pathways.

TABLE I. Pathogenics non synonymous SNPs in human FLT3 gene by 5 Bioinformatics Database. (Note: MAF = Minor allele frequency; Cons. = Consequence; MV = Missense variant; Pred. = Prediction; x = Probably damaging; y = Deleterious; z = Pathogenic; N/A = Not available

SNPs ID	Chromosomal position	Nucleotide change	Global MAF	Cons. type	Substitution	PolyPhen-2		SIFT		SNPs3D		FATHMM		RegulomeDB	
						Score	Pred	Score	Pred	SVM Profile	Pred	Score	Pred	Score	Rank
rs376588714	13:28018483	T>A/C	<0.01	MV	Y842C	1	x	0	y	-3.31	y	-1.9	z	0.60906	4
rs372303125	13:28028192	G>A	<0.01	MV	A680V	0.924	x	0	y	-0.91	y	-1.51	z	0.13454	5
rs368432815	13:28015633	A>T	<0.01	MV	D829E	0.999	x	0	y	-1.94	y	-4.7	z	0.13454	5
rs201208287	13:28018568	C>A/G/T	<0.01	MV	A814S	0.998	x	0	y	-2.28	y	-2.5	z	0.60906	4
rs141185874	13:28048435	A>G	<0.01	MV	F349L	0.684	x	0.044	y	-1.52	y	-1.82	z	0.58955	5
rs121913487	13:28018503	A>C/T	N/A	MV	D835E	0.953	x	0	y	-2.1	y	-1.66	z	0.60906	4
rs121913232	13:28018500	G>A/C/T	N/A	MV	I836M	0.997	x	0.004	y	-1.41	y	-1.8	z	0.60906	4
rs121909646	13:28018504	T>A/C/G	N/A	MV	D835V	0.998	x	0	y	-2.78	y	-1.74	z	0.60906	4
rs201923726	13:28015697	C/T	0.01	MV	R849H	0.999	x	0	y	-1.94	y	-2.51	z	0.13454	5
rs370459694	13:28015601	A>G	<0.01	MV	I881T	1	x	0.003	y	-1.55	y	-2.67	z	0.23589	5

3.3 Assessment of Deleterious cSNPs using MutPred2, CADD, and ClinPred

The selected deleterious cSNPs were further analysed using the Mut-Pred2, CADD, and ClinPred servers. Table II shows the score and prediction of their status as determined by these tools. All ten amino acid substitutions (F349L, A680V, A814S, D829E, D835E, D835V, I836M, Y842C, R849H, and I881T) were anticipated to be pathogenic across all three computation tools. Variants such as D835V, R849H, Y842C, and A680V showed high CADD PHRED scores (>28) and ClinPred scores nearing 1, suggesting strong deleterious potential. MutPred2 scores further validated the pathogenic nature of these substitutions, with values ranging from 0.625 to 0.872. This consistent prediction of variants to be damaging or deleterious further validates the objective of this study.

TABLE II. Predicting the impact of Deleterious cSNPs in human FLT3 protein by MutPred2, Clinpred and CADD.

SNPS ID	Substitution	MutPred2				Clinpred		CADD	
		Molecular Mechanism	p-value	MutPred2 Score	Prediction	Score	Prediction	Score	Prediction
rs376588714	Y842C	Ordered interface affected Transmembrane protein affected Decreased Relative solvent accessibility Strand loss Metal binding affected Increased Disulfide linkage at Y842	6.30E-04 7.10E-04 0.03 0.04 0.04 0.04	0.859	Pathogenic	0.829	Pathogenic	29	Pathogenic
rs372303125	A680V	Loss Loop Transmembrane protein affected	0.04 0.03	0.8	Pathogenic	0.542	Pathogenic	28.6	Pathogenic

rs368432815 D829E	Metal binding affected Increased Allosteric site at R834 Decreaded Relative solvent accessibility Transmembrane protein affected DNA binding affected Decreased Disulfide linkage at C828 Increased Catalytic site at D829	2.00E-02 7.80E-03 0.04 4.30E-03 0.02 0.04 0.03	0.791	Pathogenic	0.686	Pathogenic	19.72 Pathogenic
rs201208287 A814S	DNA binding affected Strand gain Increased Allosteric site at R810	6.10E-04 4.10E-03 7.20E-03	0.677	Pathogenic	0.947	Pathogenic	28 Pathogenic
rs141185874 F349L	Strand loss Increased N-linked glycosylation at N351	8.90E-03 5.00E-03	0.759	Pathogenic	0.895	Pathogenic	26.3 Pathogenic
rs121913487 D835E	Decreased Allosteric site at R834 Decreased relative solvent accessibility Transmembrane protein affected Metal binding affected DNA binding affected	0.01 0.03 2.90E-03 0.05 0.05	0.701	Pathogenic	0.994	Pathogenic	19.72 Pathogenic
rs121913232 I836M	Increased Allosteric site at R834 Increased Relative solvent accessibility Ordered interface affected	4.00E-03 0.03 0.05 2.70E-03 0.05 0.04	0.625	Pathogenic	0.966	Pathogenic	17.87 Pathogenic

	Transmembrane protein affected Metal binding affected DNA binding affected					
rs121909646 D835V	Decreased Relative solvent accessibility Ordered interface affected Loop loss Decreased Allosteric site at R834 Transmembrane protein affected Metal binding affected DNA binding affected	4.30E-03 0.03 0.02 8.50E-03 7.70E-04 0.03 0.04	0.872	Pathogenic 0.993	Pathogenic 35	Pathogenic
rs201923726 R849H	Ordered interface affected Transmembrane protein affected Decreased Allosteric site at W854	9.60E-03 2.60E-03 0.03	0.68	Pathogenic 0.995	Pathogenic 33	Pathogenic
rs370459694 I881T	Strand gain Ordered interface affected Transmembrane protein affected Stability affected	0.04 0.02 0.01 0.03	0733	Pathogenic 0,978	Pathogenic 27.4	Pathogenic

3.4 Protein structure's stability assessment by I-Mutant 2.0

The deleterious cSNPs were further evaluated using the I-Mutant 2.0 server, which provides information on stability, the reliability index (RI), alongside net change in the Gibbs free energy change ($\Delta\Delta G = \Delta G_{Mutant} - \Delta G_{Wildtype}$). A negative $\Delta\Delta G$ value indicates reduced stability, whereas a positive value suggests increased stability. In total, out of ten cSNPs, seven mutations (Y842C, A680V, I881T, R849H, F349L, and

I836M, D829E) showed a destabilisation of the FLT3 protein, each showing a ΔΔG value of less than, showing a significant effect on the protein. And the rest cSNPs (A814S, D835E, and D835V) were confirmed to increase or slightly stabilise the protein, with each showing positive ΔΔG values. The results suggest that the majority of the selected high-risk SNPs may destabilise the FLT3 protein, potentially altering its structural conformation and function. The ΔΔG values and predicted stability effects for each variant are summarised in Table III.

TABLE III. I-Mutant 2.0 was employed to quantify the predicted thermodynamic stability and Conservation Profiling of amino acid residues by ConSurf

Note: The Reliability Index is the confidence score for each prediction of a stability change, with higher values indicating greater reliability.

SNPs ID	Substitution	I-Mutant 2.0			ConSurf	
		Stability	Reliability Index	ΔΔG (kcal/mol)	Conservation score	Prediction
rs376588714	Y842C	Unstable (Decrease)	7	-2.45	9	Evolutionarily conserved and core buried (structural stability)
rs372303125	A680V	Unstable (Decrease)	4	-0.99	9	Evolutionarily conserved and core buried (structural stability)
rs368432815	D829E	Unstable (Decrease)	8	-1.5	9	Evolutionarily conserved and surface exposed (protein function)
rs201208287	A814S	Stable (Increase)	2	0.14	9	Evolutionarily conserved and core buried (structural stability)
rs141185874	F349L	Unstable (Decrease)	5	-1.81	2	Predicted as core buried residue by the neural network model.
rs121913487	D835E	Stable (Increase	2	0.08	9	Evolutionarily conserved and core buried (structural stability)
rs121913232	I836M	Unstable (Decrease)	6	-1.01	9	Evolutionarily conserved and core buried (structural stability)
rs121909646	D835V	Stable (Increase	2	0.28	9	Evolutionarily conserved and core buried (structural stability)
rs201923726	R849H	Unstable (Decrease)	8	-2.05	9	Evolutionarily conserved and surface exposed (protein function)
rs370459694	I881T	Unstable (Decrease)	7	-1.64	9	Evolutionarily conserved and core buried (structural stability)

3.5 Evolutionary conservation profile analysis by ConSurf

The residues' evolutionary conservation profiles of FLT3 protein were evaluated by the ConSurf server (Fig. 2). Most of the pathogenic cSNPs were evolutionarily conserved among the ten cSNPs analysed, indicating their potential functional or structural/stability significance. Specifically, Y842C, A680V, I881T, A814S, D835E, I836M, and D835V were confirmed as highly conserved and buried, highlighting their role in structural integrity. In contrast, D829E and R849H were confirmed to be highly conserved and exposed, highlighting their role in the involvement in functional interaction. F349L was identified as a variable and buried residue, suggesting no involvement in the conserved region; therefore, it may influence structural integrity rather than having a direct impact on functional activities of the protein. Notably, the kinase domain mutations D835E and D835V, along with I836M and I881T, were all located at highly conserved buried positions, suggesting that any alteration at these sites could significantly impact protein stability and conformation. Similarly, the exposed and conserved residues R849H were considered functionally important. These observations support the pathogenic potential of these variants, as disruptions at evolutionarily conserved positions are more likely to adversely impact protein functionality(Ashkenazy et al., 2016; Ben Chorin et al., 2020)

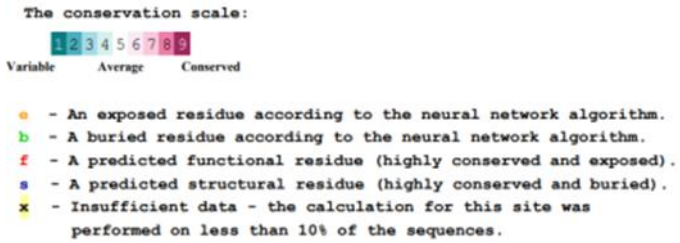


Fig. 2. Results from the Analysis of Human protein by ConSurf

3.6 Post-Translational Modification Prediction (MusiteDeep)

The FLT3 protein was predicted as a post-translational modification (PTM) site by the use of the MusiteDeep server. Ten high-risk cSNPs were assessed and only one was found to be a possible phosphorylation site, Y842C. MusiteDeep was used to predict phosphorylation site at the tyrosine residue 842 in the wild-type sequence and mutation at this site may lead to the disruption of a potential regulatory modification. Not a single other high-risk variants were found to be within the residues that were modifiable as per the chosen PTM type. Y842C was of special interest as it was structurally and functionally important and also predicted as a phosphorylation site. This variant was also subjected to 3D structural modelling and comparative analysis with the other high-risk cSNPs

.3.7 Structural Modeling and Validation

In order to examine whether there is significant structural change in the FLT3 protein of the ten high-risk nsSNPs, the 3-D structures of the wild type FLT3 and the mutant types of the ten nsSNPs were modeled using SWISS-MODEL homology modeling. The FLT3 model was built with FLT3 sequences that showed high identity and GMQE score, as well as high sequence coverage. Mutant models were created for each high-risk nsSNP (F349L, A680V, A814S, D829E, D835E, D835V, I836M, Y842C, R849H, and I881T), and each was subjected to simulation under the same conditions. All generated models were evaluated by GMQE scores, ERRAT overall quality factor and PROCHECK Ramachandran plot analysis. The GMQE score was good for all mutant models, with a score of 0.75 (Table 4.8), and was able to cover the target sequence well, with good reliability. All models have ERRAT overall quality factors above 50, ranging from 90.495 to 91.031, which is an acceptable threshold for models with high structural quality.The structural similarity of the wild-type and mutant models was estimated by TM-align and PyMOL. All

mutant models had a TM-score of 1.0, meaning that the overall fold of all mutants is the same as the wild-type. RMSD values obtained from PyMOL showed that there was no significant difference in the structure of the mutant and the wild-type model (0.001 Å – 0.004 Å). The low RMSD value values demonstrate that no significant global conformational changes take place in the FLT3 protein caused by the 10 nsSNPs.All models were analysed with PROCHECK Ramachandran plot analysis to assess the stereochemical quality. For the wild-type FLT3, 86.5% of residues were found in most favored regions, 10.3% additional allowed regions, 2.1% generously allowed regions and 1.1% disallowed regions. The distributions of the mutant models were very similar with 86.2-86.5% of the residues in most favoured regions, 10.3-10.4% in additional allowed regions, 2.1-2.2% in generously allowed regions and 1.1% in disallowed regions. The values obtained are good backbone conformation and stereochemical quality for all of the models generated. In general, the structural modelling and validation results indicate that none of the mutations of these nsSNPs causes significant change in the overall structure with respect to the wild type FLT3, and thus it is likely that the pathogenic potential of these SNPs is related to minor local perturbations, such as alterations in the stability and/or post-translational modification and/or protein-protein interaction of the protein due to the mutation, but not to the major structural changes.

3.7 Protein-Protein Interaction Network (GeneMANIA)

In order to look for the other functional network of FLT3 and possible interacting partners that could be relevant to RA pathogenesis, gene-gene interaction analysis was carried out by using the prediction server GeneMANIA. Analysis showed that FLT3 did interact with various genes associated with cytokine signaling, signal transduction and immune regulation. Negative regulators of cytokine signaling, such as SOCS1, SOCS2 and SOCS6, have been identified as key interacting genes that are important for modulating inflammatory responses. Also, CBLB was found to interact with each other, and it is an E3 ubiquitin ligase that participates in the degradation of FLT3 and termination of the signal. Another protein, PTPN12, which dephosphorylates signaling molecules, was also discovered to interact with FLT3. IKBKG (also called NEMO), a key part of the NF-κB signaling pathway, was among the interacting genes identified and therefore directly connected to inflammatory pathways central to rheumatoid arthritis, which are linked to FLT3. Other genes that interact were INPP5D (SHIP1), PTEN, and TNK2, which all play roles in PI3K/AKT signaling and regulation of immune cells. The expression of FLT3LG (FLT3 ligand) was also detected, indicating possible positive feedback. Altogether, analysis of the GeneMANIA network places FLT3 in a central position, surrounded by genes involved in inflammation, immune cell survival and production of cytokines, suggesting that FLT3 nsSNPs could play a role in the pathogenesis of RA by disrupting these interactions.

TABLE IV. Structural quality parameters and stereochemical validation of models of FLT3 protein

FLT3 Protein Model	SWISS MODEL		TM Align	PyMol	ERRAT	PROCHECK Ramachandran Statistics			
	GMQE	QSQE	TM score	RMSD (Å)Atomic Displacement	ERRAT Score	Most Favored	Additional allowed	Generously allowed	Outliers
WT	-	-		0.001	90.788	773 (86.5) %	92 (10.3%)	19 (2.1%)	10 (1.1%)
Y842C	0.75	-	1	0.001	90.495	774 (86.5) %	93 (10.3%)	20 (2.1%)	11 (1.1%)
A680V	0.75	-	1	0.004	90.654	771 (86.2%)	93 (10.4%)	20 2.2%	10 (1.1%)
D829E	0.75	-	1	0.001	90.909	773 (86.5) %	92 (10.3%)	19 (2.1%)	10 (1.1%)
A814S	0.75	-	1	0.002	91.031	773 (86.5) %	92 (10.3%)	19 (2.1%)	10 (1.1%)
F349L	0.75	-	1	0.002	91.019	772(86.4) %	93 (10.4%)	19 (2.1%)	10 (1.1%)
D835E	0.75	-	1	0.001	90.667	772(86.4) %	93 (10.4%)	19 (2.1%)	10 (1.1%)
I836M	0.75	-	1	0.002	90.897	772(86.4) %	93 (10.4%)	19 (2.1%)	10 (1.1%)
D835V	0.75	-	1	0.001	90.667	772(86.4) %	93 (10.4%)	19 (2.1%)	10 (1.1%)
R849H	0.75	-	1	0.002	90.495	773 (86.5) %	92 (10.3%)	19 (2.1%)	10 (1.1%)
I881T	0.75	-	1	0.002	90.897	772(86.4) %	93 (10.4%)	19 (2.1%)	10 (1.1%)

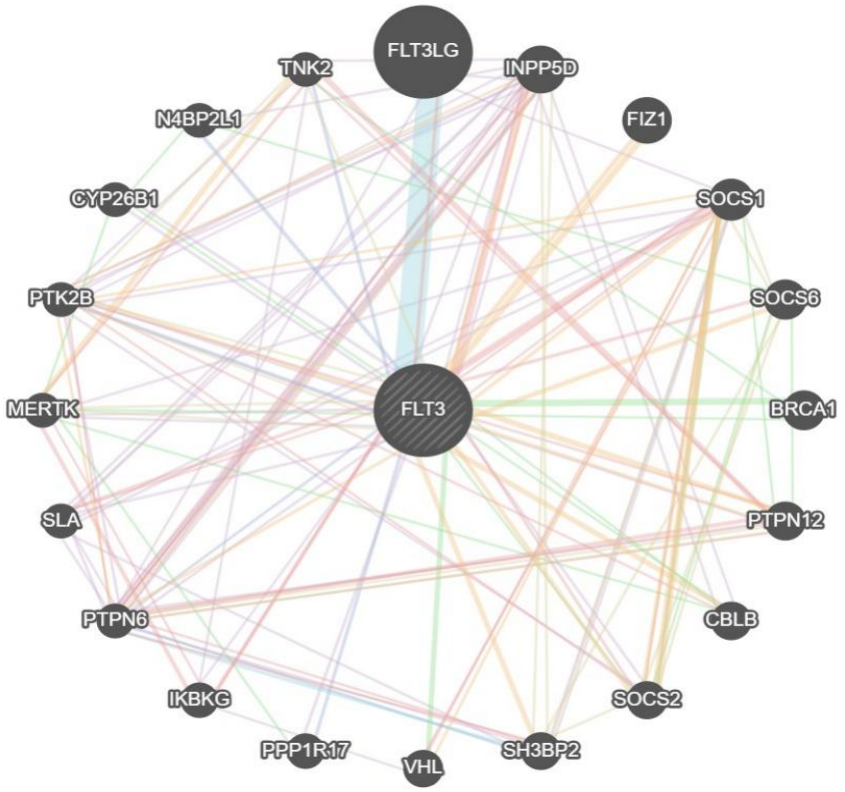


Fig. 3. GeneMANIA interaction network of FLT3

CHAPTER 4

DISCUSSION

A detailed in-silico pipeline was employed in work to identify and characterise deleterious nsSNPs of the FLT3 gene and its potential role in rheumatoid arthritis (RA) and acute myeloid leukaemia. All missense nsSNPs in dbSNP (N=1444) were tested with five prediction programs (SIFT, PolyPhen-2, FATHMM, SNPs3D and RegulomeDB), and those predicted as deleterious by three or more programs were included in the shortlist. This stringent filtering identified ten nsSNPs: F349L, A680V, A814S, D829E, D835E, D835V, I836M, Y842C, R849H, and I881T. All of these ten variants were also predicted pathogenic by MutPred2, CADD and ClinPred and validated by at least two of these tools. D835V, R849H, Y842C and A680V had high CADD and ClinPred scores (> 28 and near 1.0 respectively) indicating high deleterious potential. The molecular mechanisms identified by MutPred2 were changed metal binding, removal of allosteric sites, addition or removal of catalytic sites, changed ordered interfaces, removal of strand, addition of helix, changed transmembrane properties, changed DNA binding, and stability altered. Y842C had changed ordered interface, transmembrane protein, removal of solvent accessibility, removal of strand, changed metal binding, and addition of disulfide linkage. D835E and D835V had removal of allosteric site R834, removal of solvent accessibility, changed transmembrane protein, changed metal binding and changed DNA binding. I836M had addition at R834, gain in solvent accessibility, altered ordered protein interface, altered transmembrane protein, altered metal binding and altered DNA binding. A680V had a loss of loop and changed transmembrane protein. D829E had changed metal binding, addition of allosteric site at R834, loss of solvent accessibility, changed transmembrane protein, changed metal binding and changed DNA binding. I881T had the following changes: addition of strand, change in ordered interface, change in transmembrane protein, and change in stability. These different mechanisms imply that each nsSNP has a different effect on FLT3 function. The I-Mutant 2.0 analysis revealed 7 mutations (Y842C, A680V, I881T, D829E, R849H, F349L, and I836M) which give low stability of the protein (negative $\Delta\Delta G$), which may lead to degradation or misfolding. Three mutations (A814S, D835E and D835V) increase stability that in the context of the kinase domain can lead to gain of function or constitutive activation, a known leukemogenic mechanism. All the mutations except F349L (score 2) are at highly conserved positions (scores 7-9) as per ConSurf evolutionary conservation analysis. Buried residues (Y842C, A680V, A814S, D835E, I836M, D835V, I881T) are very conserved and more likely to have impact on structure, while exposed residues

(D829E, R849H) are also very conserved and likely to have functional impact. This high conservation suggests high mutations rate which are supposed to be deleterious. The only variant was Y842C, at a potential phosphorylation site, whose wild-type tyrosine is expected to be phosphorylated, according to MusiteDeep. Such a mutation to cysteine will remove the regulatory modification, which will disrupt the signaling pathways of proliferation of gene. All the mutant models were structurally modeled with SWISS-MODEL, TM-align and PyMOL, and were found to be highly similar to the wild-type with TM scores of 1.0 and RMSD values of 0.001-0.004 Å. The ERRAT scores of >90 and the Ramachandran plot with 86-87% in most favored regions were an indication of the good model quality and stereochemical reliability. Absence of major changes in the global protein structure suggests that detrimental effects occur from local changes in stability, PTM sites, and protein-protein interaction interfaces rather than large scale conformational changes.The GeneMANIA network analysis indicated that FLT3 associates with several genes that are associated with immune regulation and inflammation. Some of the important interacting genes are the negative regulators of cytokine signaling, SOCS1, SOCS2, and SOCS6, as well as the E3 ubiquitin ligase, CBLB, which facilitates the degradation of FLT3, and the phosphatases PTPN6 and PTPN12, which dephosphorylate FLT3, and importantly, IKBKG (NEMO), which is a component of the NF-κB inflammatory pathway essential to RA pathogenesis. Other interacting genes are INPP5D (SHIP1), PTEN and TNK2, which are involved in PI3K/AKT signaling and immune regulation. FLT3LG is a possible positive feedback loop.

These interactions suggest that the nsSNPs within FLT3 may affect the function of FLT3 as well as its interactions with SOCS proteins, CBLB, PTPN6, and IKBKG. The disruption of FLT3-SOCS interaction may result in dysfunction of the negative feedback mechanism, with continued FLT3 signaling and increased cytokine production. If the binding of FLT3-CBLB is disrupted, ubiquitination and degradation may be impaired leading to receptor accumulation and continued activation. Disruption of FLT3-PTPN6 interaction could lead to failure to terminate the signal, resulting in continued downstream signaling. Most importantly, disruption of the FLT3-IKBKG interaction may cause hyper-activation of NF-κB signaling leading to the excessive production of TNF-α, IL-6 and IL-1β, which are major regulators of synovial inflammation and joint destruction in RA.Therefore, GeneMANIA indirectly shows that FLT3 nsSNPs, by modifying the protein structure and stability, can take part in affecting several protein-protein interactions, which ultimately can result in dysregulation of immune signaling. AML exhibits loss of negative regulation by SOCS/CBL and activating mutations of the kinase domain (D835E, D835V) that add to constitutive activation and uncontrolled proliferation. Alternatively, an overactive NF-κB activation in RA might result from the disruption of FLT3 interactions with IKBKG, SOCS and PTPN6, with the resulting persistent cytokine signaling and chronic inflammation.

Chapter 5

CONCLUSION

This study provides a bioinformatic evaluation of non-synonymous single nucleotide polymorphisms (nsSNPs) of the human FLT3 gene for the first time, with a special focus on their involvement in Acute Myeloid Leukaemia and rheumatoid arthritis. To identify the high-risk missense nsSNPs, a stringent computational pipeline was used that included five missense nsSNP prediction tools and three validation tools (MutPred2, CADD, and ClinPred), which led to the identification of ten high-risk missense nsSNPs: F349L, A680V, A814S, D829E, D835E, D835V, I836M, Y842C, R849H, and I881T. I-Mutant 2.0 was employed to analysies protein stability found that 7 out of 10 mutations were destabilizing, signifying possible increased degradation/misfolding. Conservation analysis by Evolution allowed the use of the ConSurf tool, which indicates that nine mutations out of ten are located at highly conserved residues (scores 7-9), thus confirming they are important for structure or function. With the aid of post-translational modification prediction performed by MusiteDeep, only one variant, Y842C, which was predicted to impact on a potential phosphorylation site was identified and therefore this variant is particularly important to investigate further. Structural analysis of both WT and mutant FLT3 protein using SWISS-MODEL, TM-align and PyMOL showed that there is huge similarity between WT and the mutant proteins with TM-scores of 1.0 and RMSD values between 0.001 – 0.004 Å. Good quality and stereochemical reliability of all models were confirmed by ERRAT scores > 90 and Ramachandran plot analysis with the presence of 87% of residues in the most favoured regions. Considering the lack of major conformational changes at the global scale, the negative consequences of these nsSNPs seem to be the result of minor changes in stability, post translational modifications, and protein protein interaction. GeneMANIA network analysis showed that FLT3 proteins interact with several genes that are related to immune regulation and inflammation such as SOCS1, SOCS2, SOCS6, CBLB, PTPN6 and most notably with IKBKG (NEMO), which is a central component of the NF-kB inflammatory pathway. This is indirect evidence that FLT3 nsSNPs can interfere with these interactions, causing dysregulated cytokine signalling and chronic inflammation. The main findings of this work are that 10 high-risk, deleterious FLT3 nsSNPs were identified that are expected to have an impact on protein stability, evolutionary conservation and post-translational modifications but not on the overall structure of the protein.

FUTURE SCOPE

The results of this in-silico study are significant enough to warrant more in-silico and experimental studies. The following research directions are suggested:

- In-vitro Functionality Checking: The ten high risk nsSNPs, especially Y842C, should be in-vitro functionally validated by site directed mutagenesis and expression studies in appropriate cell lines.
- To gain insight into the dynamic behavior and time-dependent conformational fluctuations of mutant FLT3 proteins, it is necessary to carry out Molecular Dynamics Simulations (100-200 ns).
- Patient Cohort Studies: Engineered variants should be tested in RA patients to confirm genotypic association with susceptibility to and progression of the disease.
- Pathway Analysis: Functional assays to assess the impact of FLT3 mutations on downstream immune signaling should be performed.

Therapeutic Targeting: In the event of their experimental proof of pathogenicity, these FLT3 variants could be used as a marker for the risk assessment of RA and/or as targets for personalized therapeutic interventions

REFERENCES

[1] O. Rosnet, M. G. Matteï, S. Marchetto, and D. Birnbaum, "Isolation and chromosomal localization of a novel FMS-like tyrosine kinase gene," *Genomics*, vol. 9, no. 2, pp. 380–385, 1991, doi: 10.1016/0888-7543(91)90270-O.

[2] J. Liu and J. Gu, "Importance of PTM of FLT3 in acute myeloid leukemia: Importance of PTM of FLT3 in acute myeloid leukemia," *Acta Biochim. Biophys. Sin. (Shanghai).*, vol. 56, no. 8, p. 1199, Jun. 2024, doi: 10.3724/ABBS.2024112.

[3] J. L. Rocnik *et al.*, "Roles of tyrosine 589 and 591 in STAT5 activation and transformation mediated by FLT3-ITD," *Blood*, vol. 108, no. 4, pp. 1339–1345, Aug. 2006, doi: 10.1182/BLOOD-2005-11-011429.

[4] J. U. Kazi and L. Rönnstrand, "FMS-like Tyrosine Kinase 3/FLT3: From Basic Science to Clinical Implications," *https://doi.org/10.1152/physrev.00029.2018*, vol. 99, no. 3, pp. 1433–1466, Jul. 2019, doi: 10.1152/PHYSREV.00029.2018.

[5] J. Griffith *et al.*, "The Structural Basis for Autoinhibition of FLT3 by the Juxtamembrane Domain," *Mol. Cell*, vol. 13, no. 2, pp. 169–178, Jan. 2004, doi: 10.1016/S1097-2765(03)00505-7.

[6] D. Gary Gilliland and J. D. Griffin, "The roles of FLT3 in hematopoiesis and leukemia," *Blood*, vol. 100, no. 5, pp. 1532–1542, Sep. 2002, doi: 10.1182/BLOOD-2002-02-0492.

[7] M. I. Ramos *et al.*, "FMS-related tyrosine kinase 3 ligand (Flt3L)/CD135 axis in rheumatoid arthritis," *Arthritis Research & Therapy 2013 15:6*, vol. 15, no. 6, pp. R209-, Dec. 2013, doi: 10.1186/AR4403.

[8] C. M. Yates and M. J. E. Sternberg, "The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions," *J. Mol. Biol.*, vol. 425, no. 21, pp. 3949–3963, Nov. 2013, doi: 10.1016/j.jmb.2013.07.012.

[9] E. M. Smigielski, K. Sirotkin, M. Ward, and S. T. Sherry, "dbSNP: a database of single nucleotide polymorphisms," *Nucleic Acids Res.*, vol. 28, no. 1, p. 352, Jan. 2000, doi: 10.1093/nar/28.1.352.

[10] S. Bora, R. Kukreti, and Y. Hasija, "An in-silico analysis predicting the impact of coding single nucleotide polymorphisms (SNPs) in the human multidrug pump ABCG2," *Human Gene*, vol. 38, p. 201213, Dec. 2023, doi: 10.1016/j.humgen.2023.201213.

[11] P. C. Ng and S. Henikoff, "SIFT: Predicting amino acid changes that affect protein function," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3812–3814, Jul. 2003, doi: 10.1093/nar/gkg509.

[12] P. Kumar, S. Henikoff, and P. C. Ng, "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm," *Nature Protocols 2009 4:7*, vol. 4, no. 7, pp. 1073–1081, Jun. 2009, doi: 10.1038/nprot.2009.86.

[13] M. F. Rogers, H. A. Shihab, M. Mort, D. N. Cooper, T. R. Gaunt, and C. Campbell, "FATHMM-XF: accurate prediction of pathogenic point mutations via extended features," *Bioinformatics*, vol. 34, no. 3, pp. 511–513, Feb. 2018, doi: 10.1093/bioinformatics/btx536.

[14] P. Yue, E. Melamud, and J. Moult, “SNPs3D: candidate gene and SNP selection for association studies,” *BMC Bioinformatics*, vol. 7, Mar. 2006, doi: 10.1186/1471-2105-7-166.

[15] A. N. Cheema, S. L. Rosenthal, and M. Ilyas Kamboh, “Proficiency of data interpretation: identification of signaling SNPs/specific loci for coronary artery disease,” *Database (Oxford).*, vol. 2017, Jan. 2017, doi: 10.1093/database/bax078.

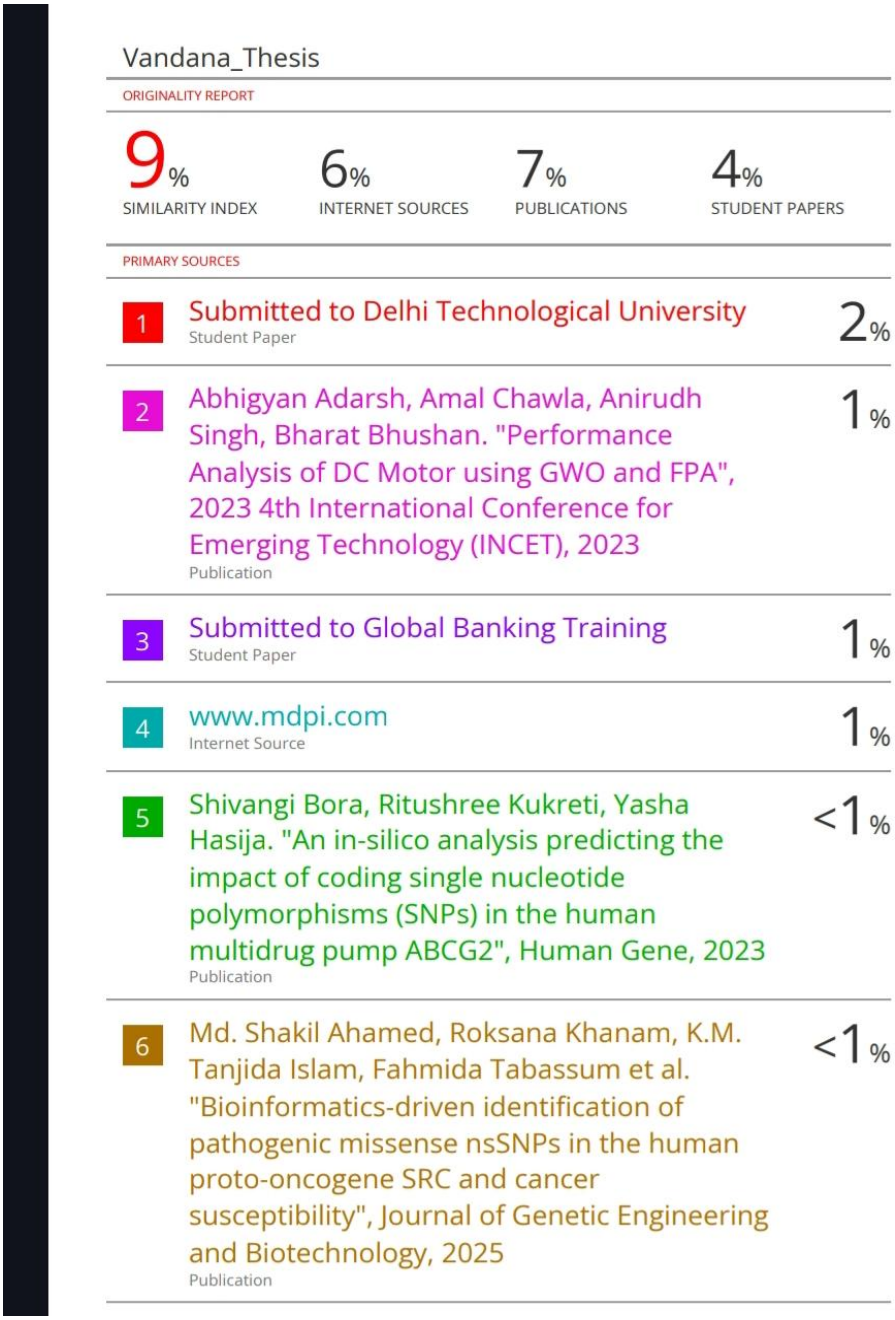
[16] J. Thusberg, A. Olatubosun, and M. Vihinen, “Performance of mutation pathogenicity prediction methods on missense variants,” *Hum. Mutat.*, vol. 32, no. 4, pp. 358–368, Apr. 2011, doi: 10.1002/humu.21445.

[17] P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, and M. Kircher, “CADD: predicting the deleteriousness of variants throughout the human genome,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D886–D894, Jan. 2019, doi: 10.1093/nar/gky1016.

[18] N. Alirezaie, K. D. Kernohan, T. Hartley, J. Majewski, and T. D. Hocking, “ClinPred: Prediction Tool to Identify Disease-Relevant Nonsynonymous Single-Nucleotide Variants,” *Am. J. Hum. Genet.*, vol. 103, no. 4, pp. 474–483, Oct. 2018, doi: 10.1016/j.ajhg.2018.08.005.

[19] A. Ben Chorin *et al.*, “ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins,” *Protein Sci.*, vol. 29, no. 1, pp. 258–267, Jan. 2020, doi: 10.1002/pro.3779.

Plagiarism Report



Vandana_Thesis

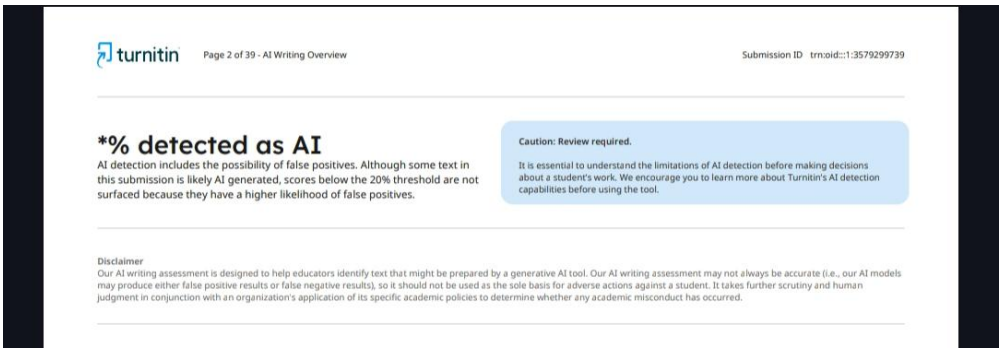
ORIGINALITY REPORT

9%	6%	7%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Delhi Technological University Student Paper	2%
2	Abhigyan Adarsh, Amal Chawla, Anirudh Singh, Bharat Bhushan. "Performance Analysis of DC Motor using GWO and FPA", 2023 4th International Conference for Emerging Technology (INCET), 2023 Publication	1%
3	Submitted to Global Banking Training Student Paper	1%
4	www.mdpi.com Internet Source	1%
5	Shivangi Bora, Ritushree Kukreti, Yasha Hasija. "An in-silico analysis predicting the impact of coding single nucleotide polymorphisms (SNPs) in the human multidrug pump ABCG2", Human Gene, 2023 Publication	<1%
6	Md. Shakil Ahamed, Roksana Khanam, K.M. Tanjida Islam, Fahmida Tabassum et al. "Bioinformatics-driven identification of pathogenic missense nsSNPs in the human proto-oncogene SRC and cancer susceptibility", Journal of Genetic Engineering and Biotechnology, 2025 Publication	<1%

AI Report



turnitin Page 2 of 39 - AI Writing Overview Submission ID trn:oid:::1:3579299739

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.