ADGRE2 Gene Polymorphism and Its Influence on CML Pathogenesis

Submitted by: Harooma Jamshaid (357182)

Ayesha Afzal (357397)

Supervisor

Dr. Yasmin Badshah

Co-Supervisor

Dr. Maria Shabbir

Atta-Ur-Rahman School of Applied Biosciences

National University of Science and Technology

Islamabad, Pakistan

2024

ADGRE2 Gene Polymorphism and Its Influence on CML Pathogenesis

Submitted by:

Harooma Jamshaid (357182)

Ayesha Afzal (357397)

This thesis is submitted to the National University of Sciences and Technology, Islamabad, in partial fulfillment of the requirements for the degree of Bachelor of Sciences in Applied Biosciences

Supervisor

Dr. Yasmin Badshah

Co-Supervisor

Dr. Maria Shabbir

Atta-Ur-Rahman School of Applied Biosciences

National University of Science and Technology

Islamabad, Pakistan

2024

THESIS ACCEPTANCE CERTIFICATE

This document certifies that this is the final copy of BS FYP Thesis written by **Ms. Harooma Jamshaid** (Reg No. 357182) and **Ms. Ayesha Afzal** (Reg No.357397) of **Atta-Ur-Rahman School of Applied Biosciences (ASAB)**. The undersigned has thoroughly reviewed and found that the document complies with all regulations set by the National University of Sciences and Technology (NUST). It is confirmed to be free of plagiarism, errors, and mistakes, and is accepted as a partial fulfillment for the Bachelor of Sciences degree in Applied Biosciences. Additionally, all necessary amendments suggested during the final presentation have been included in the thesis.

Signature: Dr. Yasmin Badshah, Assistant Professor, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Name of Supervisor: **Dr. Yasmin Badshah**

Date: 30-05-24

Signature: Dr. Maria Shabbir, Associate Professor, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Name of Co-Supervisor: **Dr. Maria Shabbir**

Date:

Signature (HOD): Dr. Hussain Mustatab Wahedi, Head of Department (HoD), Department of Healthcare Biotechnology, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Head of Department Biomedicine: **Dr. Hussain Mustatab Wahedi**

Date: 31/05/2024

Signature (Dean/Principal):

Principal ASAB: **Dr. Muhammad Asghar**

Date: 31-05-24

DECLARATION

We affirm that the research work titled "**ADGRE2 Gene Polymorphism and Its Influence on CML Pathogenesis**" authored by Harooma Jamshaid and Ayesha Afzal is our original creation. This work has not been presented elsewhere for evaluation. The research was conducted during our undergraduate studies at Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), under the guidance of Dr. Yasmin Badshah. Any material borrowed from external sources has been appropriately cited and acknowledged.

Ayesha Afzal

Harooma Jamshaid

CERTIFICATE FOR PLAGIARISM

It is certified that this Undergraduate Thesis Titled "***ADGRE2* Gene Polymorphism and Its Influence on CML Pathogenesis**" written by **Ms. Harooma Jamshaid** (Reg No. 357182) and **Ms. Ayesha Afzal** (Reg No. 357397) of **Atta-Ur-Rahman School of Applied Biosciences (ASAB),** has been examined by me.

I undertake that:

a. The thesis contains original contributions and novel insights not found in previously published works or those currently being reviewed for publication elsewhere. No exact sentences, equations, diagrams, tables, paragraphs, or sections have been directly copied from prior sources without proper quotation marks and citations.

b. The work presented is original and own work of the authors i.e., there is no plagiarism.

c. There is no falsification of data or misrepresentation of research findings in the records. The thesis has undergone inspection using TURNITIN (an originality report is attached), confirming compliance with the HEC plagiarism policy and relevant guidelines.

Dr. Yasmin Badshah
Assistant Professor
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST Islamabad

Signature of Supervisor

Dr. Yasmin Badshah

Dr. Maria Shabbir
Tenured Associate Professor
Applied Biosciences (ASAB), NUST Islamabad

Signature of Co-Supervisor

Dr. Maria Shabbir

DEDICATION

“This work is dedicated to our beloved parents (and my uncle), siblings, teachers and friends. Their unwavering support, encouragement and understanding have been our motivation to strive and move forward.

Acknowledgements

In the name of Allah, the Most Merciful, the Most Kind. All praises and thanks to Allah Almighty Who has bestowed His countless blessings and gifts upon us and without His mercy we would not have been able to complete this thesis.

We would like to express our heartfelt gratitude to our supervisor, Dr. Yasmin Badshah, and our co-supervisor, Dr. Maria Shabbir, for their invaluable guidance and encouragement throughout this project.

We would like to thank Sameen Zafar, Amna Hafeez and Aneela Mustafa for their continuous support, encouragement, and assistance throughout the project.

Lastly, a special thanks to our parents and Family members (my uncle) as this journey would not have been possible without their love support and guidance.

Thank you

Table of Contents

List of Figures

List of Tables

Abstract

Chronic myeloid leukemia (CML) is a type of blood cancer that begins in the hematopoietic stem cell (HSC) compartment. It is characterized by a specific chromosomal aberration known as the Philadelphia (Ph) chromosome, which arises from a reciprocal translocation between chromosomes 9 and 22, leading to the union of the BCR and ABL1 genes. While the fusion gene is a major contributor to CML, several other genes are also implicated in the disease's progression. Until recently, little research had been conducted to identify single nucleotide polymorphisms (SNPs) associated with CML. Therefore, this study aims to investigate the influence of missense variants on the structure and function of the ADGRE2 gene and to identify missense variants associated with CML and its clinical and pathological characteristics. Missense variants of ADGRE2 were retrieved from the ENSEMBL genome browser, and the pathogenicity of deleterious variants was assessed using six different consensus tools. Additionally, various *in silico* tools were employed to explore the relationship between damaging SNPs, function, stability, and structure. Genotype analysis was performed on collected blood samples, revealing that the most pathogenic SNP among all missense variants was rs765071211 (D67N). Significant differences were observed between the wild-type and mutant variant (D67N) structures of ADGRE2. Genotype analysis indicated an association between the variant and CML, as genotypes CT and TT were found to have significant P-values. Genotype heterozygous CT has been identified as a risk factor for CML, while TT has a protective role in CML patients. This study suggests that the ADGRE2 SNP rs765071211 could serve as a potential genetic marker for the diagnosis of CML.

Chapter 1: Introduction

Cancer is an illness that is made up of complex changes in genes and the alterations of the cell bodies and non-cell bodies of the host. From the beginning, it was regarded as one of the deadliest diseases of the twentieth century. Its spreading is much more than the other diseases, affecting one in four persons at some time in their lives. Although tumors are curable, their growth to canceric poses difficulties in treatment. Tumor masses becoming into cancer can be dealt with by surgery, chemotherapy, hormone therapy, radiation therapy, and targeted cell destruction.On the contrary, abnormally increasing cells inevitably turn into cancer. Cancer can be any disease, with the growth of small cells no longer being controlled. The transformation of normal cells into cancerous ones is usually caused by a defective immune system that cannot destroy the inadequately few cancer new cells. Many factors can be the cause of the immune system's malfunctioning during cancer, for instance, old age, long-standing diseases, stress, previous chemotherapy, and the use of drugs such as antibiotics, painkillers and corticosteroids.

Cancer is most of the times curable if detected at the initial stage. Different kinds of cancer have been discovered, and for many of them, treatments are available. For adult cancers which can be treated effectively there are Hodgkin's lymphoma, testicular malignancies, and several kinds of blood cancers. Most of the early-stage cancers of skin, cervix, colon, prostate, and breast are usually treatable.

1.1Chronic Myeloid Leukemia

Chronic myeloid leukemia (CML), before that, it was known as chronic granulocytic leukemia, was first described in 1845 by John Bennett and Rudolf Virchow, who independently published case reports of patients showing splenomegaly, liver enlargement, and leukocytosis. Nevertheless, the people knew nothing about the disease at that time. Virchow has named the term "Leukämie," which means "white blood," indicating that the symptoms of the disease were the result of the irregularities in blood cell production. The discovery that leukemia is a bone marrow cancer was made by Ernst Neumann. The first clue that drew the connection between chromosomes and cancer was the discovery in 1960 by Peter Nowell and David Hungerford of the minute chromosome, also known as the Philadelphia chromosome in patients with chronic myeloid leukemia (CML) using the advanced chromosome imaging techniques. In 1973, Janet Rowley saw that this chromosome

was the result of a translocation between chromosomes 9 and 22 (t (9;22)), the first evidence of a chromosomal translocation being responsible for cancer. Hematopoietic stem cells, which are at the beginning of the blood cell creation hierarchy, are usually the places of this translocation. In the end, Nora Heisterkamp and Jim Groffen found the ABL1 gene in the area of chromosome 9 which translocates to chromosome 22, which is the corresponding human Abelson murine leukemia viral oncogene (v-Abl) gene.(Minciacchi et al. 2021).

CML is a leukemia that affects 1-2 persons per 100,000 annually and comprises 15% of adult leukemia cases. The most significant factor in the pathophysiology of CML is the amalgamation of the ABL kinase gene with the BCR gene, which is caused by the interchange of the chromosomes t(9;22), also known as the Philadelphia chromosome. This translocation, t(9;22)(q34;q11. 2), the combination of sequences from the 22q11 BCR gene and the 9q34 ABL1 gene, leads to the formation of the BCR/ABL1 fusion gene, a characteristic feature of CML (Wu et al. , 2019). The BCR is a cytoplasmic protein that is expressed everywhere and is involved in different cell functions. ABL1 is the other ubiquitously expressed gene with broad functions like: cell cycle progression, cell proliferation regulation, integrin signaling, and DNA repair.

Generally, the body has strict regulation of ABL1 kinase activity. The other hand, the chimeric BCR/ABL1 protein that is present in CML that is constitutively active moves to the cytoplasm. This mutant protein due to the improper cell adhesion and migration, decreased apoptosis, and increased cell proliferation, and genetic instability by activating many downstream pathways. Various studies into these signaling pathways have revealed major discoveries on the molecular processes of CML. The BCR/ABL1 fusion protein is a kind of protein that activates several important downstream substrates and pathways, for example, the phosphatidylinositol-3 kinase (PI3K) pathway, the RAS/mitogen-activated protein kinase (MAPK) system, and the Janus kinase (JAK)/Signal transducer and activator of trans Pathology, cytogenetics and the application of fluorescence in situ hybridization (FISH) or reverse transcriptase-polymerase chain reaction (RT-PCR) to detect the BCR-ABL1 transcript or identify the Philadelphia chromosome are of great importance in the diagnosis of chronic myeloid leukemia (Minciacchi et al., 2021).

The previous studies on Subjects with chronic myeloid leukemia (CML) have primarily concentrated on small gene sets, and have used different techniques for that. The result of the meta-analysis showed that ASXL1 and IKZF1 mutations were found in more than 5% of CML in chronic

phase (CML-CP) cases. Moreover, genes such as RUNX1, TET2, DNMT3A, TP53, SETD1B, JAK2, KMT2D, CBL, and EZH2 were found to have recurrent mutations. (Ochi 2023).

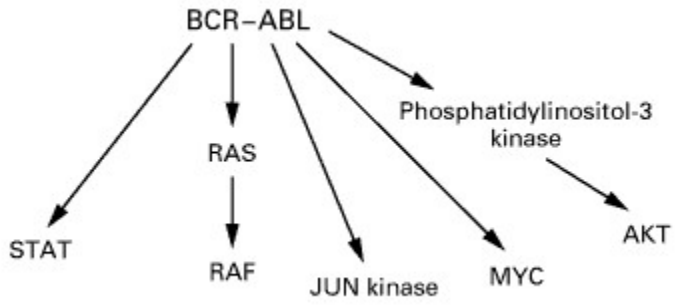


Figure 1: Signal Transduction by the BCR-ABL Protein

1.2 Clinical Features

A form of cancer that originates from the blood-forming cells of the bone marrow, known as chronic myeloid leukemia (CML), is a kind of cancer that starts from this place. It leads to the increase in the white blood cell proliferation in the bone marrow and also to the rise in myeloid, erythroid and platelet cells in the blood. Although CML can trouble any person of any age, it predominantly affects adults over the age of 53. A patient's irregular blood cell count is usually used to find chronic myeloid leukemia (CML), even though many individuals do not at first show the signs **Figure 2.** However, symptoms can include fatigue, loss of appetite, and weight loss. Physical examination may reveal an enlarged spleen in about half of patients.

Within a very short time, CML can change from the somewhat stable phase of the chronic form to the fast progressing blast crisis, if the treatment is not given. The immature cells, like those that are present in acute leukemia, which are found in the bone marrow and blood during the blast crisis, are the immature cells that look like those in acute leukemia. Generally, drugs that target the one genetic error that results in the disease are used to treat CML. (Sawyers 1999).

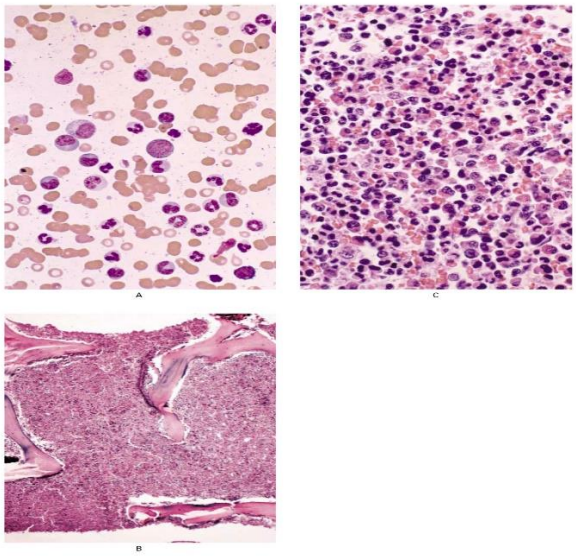


Figure 2: Photomicrographs of a peripheral blood sample and bone marrow samples from a patient with chronic myeloid leukemia.

1.3 Treatment options

The majority of the Chronic Myeloid Leukemia (CML) patients in the chronic phase require the cytoreductive therapy in order to prevent the thrombotic problems that are caused by the high blood neutrophil counts. By fate, some oral chemotherapy drugs are at work on the CML cells. 90% of the patients who are treated with hydroxyurea or busulfan regain their hematological functions. Researches have shown that hydroxyuruea is better than busulfan in the overall survival and the length of the chronic phase. The chief advantage of Hydroxyurea is not its direct effect on the cells of CML, but rather its good toxicity profile. On the other hand, both the response and the rate of progression to blast crisis remain unaffected by either of the drugs. Thus, these treatments are seen as the therapy that is not curing but giving comfort.

1.4 Allogeneic bone marrow transplantation

Hematopoietic stem cells are the ones which are harmed by CML. Although high-dose chemotherapy can, at the same time, kill all the leukemic cells, it also kills the normal bone marrow. Hence, allogeneic bone marrow or stem cell transplantation followed this treatment is the only way. Based on the prolonged data of many institutes, patients with CML in the chronic phase can be cured with high-dose chemotherapy using busulfan and cyclophosphamide, or with

chemotherapy plus cyclophosphamide and fractionated total-body irradiation followed by allogeneic bone marrow translation (Sawyers 1999).

1.5 Interferon Alfa

Since allogeneic transplantation is not a viable solution for most CML patients, researchers have turned to other treatments. Interferon alfa is a treatment that has been thoroughly researched as an agent of this disease. Research has proven that interferon alfa can be the agent of cytogenetic and hematologic remissions in chronic myeloid leukemia (CML) patients when they are given subcutaneously for a long period at a dose of 5 million U/m^2/day of body surface area. The study of patients that are under this treatment has been compared to those that are receiving other treatments. The results of several clinical trials have revealed that the former are having higher survival rates and cytogenetic responses. Though it is effective, the exact manner of interferon alfa's action in CML is still a mystery. Although it is known to inhibit tumor cells by enhancing the immune system, including the increase of HLA class I antigens, it is still unknown whether these actions play a role in its therapeutic benefits (Sawyers 1999).

1.6 Molecular Therapy

The latest breakthroughs in our knowledge of the molecular mechanisms of cancer have changed the direction of treatment from finding the drugs which boost the protein activity responsible for signal transmission by the cancer cells to the ones that target the proteins normally what is happening to the signal transmission by the cancer cells. The principal goal now is to find the drugs that would exactly target the proteins responsible for the signal transmission by cancer cells, for the reason that the recent researches have improved the knowledge about the molecular structure of many types of cancers. The BCR-ABL protein is involved in many ways that can result in leukemia, therefore, there are many possible targets for the development of pharmacological therapy.

A substance has proven to be a promising and effective for therapeutic development by specifically inhibiting BCR-ABL's tyrosine kinase activity in preclinical studies carried out in vitro and in animal models. Also the RAS pathway, the one that is fundamental for the anti-apoptotic and transforming properties of the BCR-ABL protein can be a potential target. At present, clinical

trials are investigating the efficacy of farnesyltransferase inhibitors on patients with different types of malignancies, among which are those that affect the hematological system. This group of drugs hinders a lipid change that is crucial for the RAS to work as a signaling molecule.(Sawyers 1999).

1.7 ADGRE2

Researching the variations that are related to cancer development is the key thing to be done. The discovery of these disparities is a difficult process that takes time and money, but it is very important for the right and fast diagnosis of cancer. ADGRE2 is the gene that is responsible for the encoding of the epidermal growth factor–like module containing mucin-like hormone receptor–like 2 (EMR2) protein. EMR2 is the one that is produced by myeloid cells which are monocytes, macrophages, dendritic cells and granulocytes and its expression is strictly controlled during the differentiation of monocytes/macrophage (Bhudia et al. 2020). ADGRE2 enhances its expression in macrophages and neutrophils in inflamed tissue and hence it can be used as a biomarker in patients with systemic inflammatory reaction syndrome. ADGRE2, also named EGF-like module containing mucin-like hormone receptor-like 2, is the essential factor for the innate immune system through the control of the movement and activation of neutrophils.

1.8 Function

ADGRE2 is involved in cell adhesion and migration, which helps in the gathering of immune cells at inflamed sites. Besides, it also boosts phagocytosis, which is the process of the removal of the pathogens and cellular debris. Thus, through the recognition of cell surface receptors that initiates intracellular signaling pathways such as the MAPK pathway, ADGRE2 influences different cellular functions like differentiation, apoptosis, and proliferation. Also, ADGRE2 can help in the cellular rearrangement and the healing process, thus it can be used to decrease the inflammation and to keep the tissue structure. The variety of activities of ADGRE2 shows that it is a major element in the immune system, and therefore, it is a reliable factor for inflammation and autoimmune diseases. (Hamann et al. 2016a).

1.9 Pathophysiology

Previously missense mutation of ADGRE2 in which cysteine was substituted with tyrosine at amino acid position 492(C492Y) was identified in vibratory urticaria. An extracellular subunit of ADGRE2 receptor non-covalently binds to a transmembrane unit when ADGRE2receptor undergoes autocatalytic cleavage. The variant probably destabilizes an auto inhibitory subunit interaction, sensitizing mast cells to IgE-independent vibration-induced degranulation (Boyden et al. 2016). In addition, ADGRE2 expression is higher in synovial tissue of rheumatoid arthirits patients compared to a control group (Nijmeijer et al. 2016a).

Examining the link between single nucleotide polymorphisms (SNPs) in the ADGRE2 gene is the first step towards understanding this issue. The SNPs effect is determined by their position in the gene; for instance, the SNPs in the promoter region can affect the gene expression, while the ones in the exonic region can cause the missense or nonsense variations that can affect the protein functionality of the gene it encodes.

1.10 Computational and Experimental analysis

Identifying these SNPs computationally is essential before moving on to experimental research to find the correlation of SNPs with ADGRE2. This computational method has the advantage of being labor and time-efficient, while also being cost-effective. Following variant identification, specific sites within the ADGRE2 gene are determined through experimental analysis using techniques such as polymerase chain reaction and DNA extraction.

1.11 Study Objectives

The aim and objectives of the study are:

1. To study the impact of missense variants on the structure and function of ADGRE2.

2. To detect the association of missense variants with Chronic Myeloid Leukemia

Chapter: 2 Literature Review

2.1 Chronic myeloid leukemia

Chronic myeloid leukemia (CML) is a disorder that impacts various blood cell lineages, including myeloid, monocytes, erythroid, and megakaryocytic, B-lymphoid, and sometimes T-lymphoid cells. It is characterized by the clonal overproduction of primitive hematopoietic progenitor cells. Importantly, CML was the first leukemia to be linked with a specific chromosomal abnormality known as the Philadelphia (Ph) chromosome.(Fausel 2007).

2.2 Philadelphia Chromosome

The Philadelphia (Ph) chromosome is a chromosomal abnormality resulting from a reciprocal translocation, specifically designated as t(9;22)(q34;q11), which involves the shortening of chromosome 22. This translocation affects the long arms of chromosomes 9 and 22. One distinguishing feature of chronic myeloid leukemia (CML) is the presence of the Ph chromosome. The Philadelphia (Ph) chromosome translocation involves the joining of a segment from the ABL gene on chromosome 9q34 to the BCR gene on chromosome 22q11. This fusion event results in the formation of a fusion gene called BCR–ABL, which gives rise to a chimeric BCR–ABL messenger RNA (mRNA) when transcribed **Figure 3**. The BCR ABL gene encodes a non-receptor Tyrosine kinase that promote uncontrolled proliferation of myeloid progenitor cells.

Figure 3: Structure of BCR-ABL Fusion protein

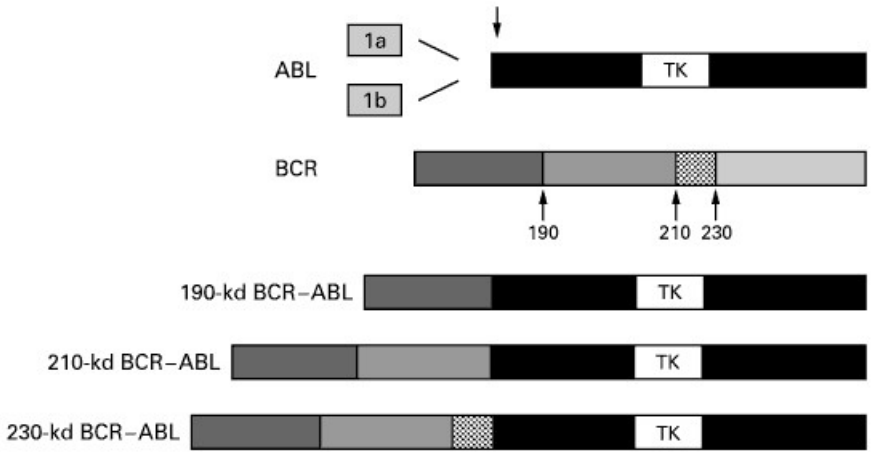


Figure 3: Structure of BCR-ABL Fusion protein

2.3 Epidemiology

CML is 15-20% of the adult leukemia cases. In 2017, the total number of new cases of CML was 34,179 around the world, and there were 24,054 CML-related deaths. Between 1990 to 2017, the ASIR and ASDR of the age-standardized death rate and incidence rate showed a consistent reduction of -2. 58% and -2. 74% per year, respectively.It's worth to note that CML was more common and deadly in males than females all over the world. Throughout different regions, the age-standardized rates (ASRs) had a lot of variability with the highest burden observed in Andean Latin America, Central Sub-Saharan Africa, and Southeast Asia. It is important to note that the ASRs dropped more significantly in the regions with a higher SDI when compared to the regions with a lower SDI.

The EUTOS and SIMPLICITY registries indicate that in Western countries, the average time from which CML is diagnosed is usually between 56 and 57 years old, with 20% of patients being over 70 years old. On the other hand, in underdeveloped countries with younger populations, the median age at which a cancer is diagnosed is under 50 years. The annual incidence of CML per 100,000 people is not the same for all ages, the range is from 1 to 2 depending on the age distribution of the population(Hehlmann 2020).

Furthermore, it was observed that in regions with a lower SDI, a larger proportion of CML-related deaths occurred in younger age groups.

2.4 Clinical Symptoms

Approximately 50% of the found cases are during the regular medical screenings, and most of them are usually found during the chronic stage. On the day of the diagnosis, possible symptoms and clinical indications are fatigue, anemia, bleeding, abdominal pain, splenomegaly, fatigue, unexplained weight loss, purplish skin discoloration (purpura), white blood cell count increase (leukocytosis), and platelet count increase (thrombocytosis).

2.5 Diagnosis

Simple validation of the Philadelphia chromosome abnormalities and unexplained elevated white blood cell (or sometimes elevated platelet) counts are required for the diagnosis of classical CML. This verification can be obtained by standard cytogenetic analysis or by employing fluorescence in situ hybridization (FISH) or molecular testing to identify Ph-related molecular BCR-ABL1 abnormalities.(Jabbour and Kantarjian 2020).

2.5.1 FISH (Fluorescence In Situ Hybridization)

Special genetic markers designed for the BCR and ABL1 genes are employed in FISH testing. A strong correlation is discovered when both bone marrow and blood specimens are analyzed simultaneously using FISH techniques. The false positive rate of a FISH study may vary, between 1% to 5% based on the markers utilized.

2.5.2 RT PCR (Real time PCR)

The reverse transcriptase-polymerase chain reaction (RT-PCR) amplifies the splice junction region of the ABL1 and BCR genes, making it highly sensitive for detecting minimal residual disease. Two PCR testing methods exist: quantitative PCR (QPCR), which measures the amount of BCR-ABL1 transcripts, and qualitative PCR, which confirms the presence of the BCR-ABL1 transcript. While qualitative PCR is useful for identifying CML, quantitative PCR is preferred for monitoring residual disease.

2.6 Treatment

Before imatinib a Tyrosine Kinase Inhibitor (TKI) was developed the primary treatments, for Chronic Myeloid Leukemia (CML) included interferon, hydroxyurea or based chemotherapy and allogeneic stem cell transplantation. Imatinib, which became available in 2001 and gained approval from the FDA and EMA marked a breakthrough in CML treatment. Since then the focus has shifted towards optimizing TKI therapy.

Following imatinib newer generations of TKIs like bosutinib, dasatinib, nilotinib, radotinib and ponatinib have been introduced. These TKIs target the ATP binding site of the BCR ABL onco protein within its kinase domain. However their efficacy varies based on differences in chemical structures and specific mutations in the kinase domain. Researchers are investigating approaches as our understanding of CML biology progresses. These approaches include strategies such as asciminib—an inhibitor that targets the pocket of the BCR ABL protein. Furthermore ongoing research focuses on combination therapies that merge established TKI medications like interferon with new drugs targeting various aspects of CML biology such, as gliptins and thiazolidinedione. (Westerweel et al. 2019).

2.7 ADGRE Gene Family

Adhesion G protein-coupled receptors (GPCRs), commonly referred to as the ADGRE gene family, are a class of cell surface receptors that are vital to many physiological functions.

Numerous physiological processes, including as cell adhesion, migration, immune response control, and tissue homeostasis, are mediated by ADGRE receptors. Their interactions with immune cells and extracellular matrix constituents impact various processes, including inflammation, tissue healing, and leukocyte trafficking.

The dysregulation of the ADGRE gene family has been connected to a number of diseases, including cancer, autoimmune disorders, and inflammatory ailments. Different members of the ADGRE gene family are associated with particular tissues and cell types. The complex roles that ADGRE genes play in cellular physiology and their potential as therapeutic targets in managing disease are still being uncovered by research on these genes.

The human ADGRE (Adhesion G protein-coupled Receptor) gene family comprises 33 known members. The ADGRE2 gene will be our primary concern.

2.8 ADGRE2

ADGRE2, also known as EGF-like module receptor 2 (EMR2) or EGF-like module-containing mucin-like hormone receptor-like 2, is a human protein encoded by the ADGRE2 gene. It is a member of the adhesion G protein-coupled receptor (aGPCR) family, characterized by an extended extracellular region containing N-terminal protein modules. These modules are linked to the GPCR-Auto Proteolysis Inducing (GAIN) domain, which spans the transmembrane region.

2.9 Localization of gene

The ADGRE2 gene, also known as EGF-like module-containing, mucin-like, hormone receptor-like 2 (EMR2), is located on human chromosome 19 at position 19q13.31 on the long arm (q arm) of the chromosome. This gene encodes a G protein-coupled receptor that plays a role in cell adhesion and migration. ADGRE2 is expressed in various tissues, including immune cells and the nervous system, and is classified as a member of the adhesion G protein-coupled receptor (aGPCR) family. While its precise functions are still under investigation, ADGRE2 is known to impact immune responses and may be implicated in various disorders.

2.10 Function of ADGRE2 membrane protein

ADGRE2 has been the subject of a great deal of research lately, which has shown its importance. Interestingly, ADGRE2 and CD97 have an astounding 97% amino acid identity in the EGF-like domains. This protein is mostly found in monocytes/macrophages, dendritic cells, and different kinds of granulocytes. It is a member of the adhesion GPCR family. Its expression has been found in breast and colorectal adenocarcinomas, but it is rarely seen in tumor cell lines and tumors.

The function of ADGRE2 in myeloid cell motility and adhesion during differentiation, maturation, and activation has been the subject of recent studies. According to these findings, EMR2 can promote neutrophil adhesion, migration, and the synthesis of anti-microbial mediators when it is activated, which may lead to a rise in systemic inflammation. Furthermore, several studies have shown a correlation between the severity of organ failure and an increase in EMR2 expression on neutrophils in individuals with Systemic Inflammatory Response Syndrome (SIRS).

2.11 Cellular pathways associated with ADGRE gene family

2.11.1 Signal Transduction

The adhesion GPCRs are the ones which activate the various intracellular signaling pathways through ligand binding. These pathways frequently include the activation of G proteins and then the downstream effectors and so the conversion of the external signal to the physiological response becomes a necessity. They communicate with local ligands which are exemplified by odors, hormones, neurotransmitters, chemokines, and a diversity of molecules such as photons, amines, carbohydrates, lipids, peptides and proteins. (Sassone-Corsi 2012).

2.11.1.1 cAMP/PKA Pathway

The second messenger cyclic adenosine 3′,5′-monophosphate (cAMP) has a major part to play in the cellular responses to different hormones and neurotransmitters. The levels of cAMP inside the cells are controlled by two enzymes - cyclic nucleotide phosphodiesterase (PDE) and adenylyl cyclase (AC). Except for the soluble bicarbonate-regulated ACs, the majority of ACs are activated through the G-protein-coupled receptors (GPCR) such as the beta-adrenoceptor. Activation is the result of the communication between the α subunit (αs) of the Gs protein and the other part of the system. When the agonist ligand binds to the GPCR (e.g., it is activated by the presence of a specific molecule), the GPCR becomes functional thus the receptor can send a signal to the cell. g. It is also worth mentioning the other effects of adrenaline for β-adrenoceptors, the heterotrimeric αβγ G-protein complex releases αs, which then activates AC. Among the most researched effectors of cAMP, one is the cAMP-dependent protein kinase (PKA) which is activated by the cAMP that is produced following AC activation. PKA is a symmetric complex which comprises of two catalytic (C) and two regulatory (R) subunits. Through the process where cAMP binds to specific sites on R subunits, they become separated from the C subunits, resulting in PKA (Sassone-Corsi 2012) activation of PKA phosphorylation of various metabolic enzymes, including glycogen synthase and phosphorylase kinase. This phosphorylation on the other hand, forestalls glycogen synthesis and induces glycogen breakdown. Besides, PKA hampers the lipid synthesis by phosphorylating acetyl CoA carboxylase.(Sassone-Corsi 2012).

2.11.1.2 Phospholipase C (PLC) Pathway

Besides, GPCRs can also activate phosphoinositide of cyclic 3,5-phospho-o-quinolone (PIP3), resulting in the formation of diacylglycerol (DAG) and inositol tris phosphate (IP3). IP3 starts the process of the release of intracellular calcium ions, whereas DAG stimulates protein kinase C (PKC).

2.11.1.3 PI3K/Akt Pathway

GPCRs can activate phosphoinositide 3-kinase (PI3K), leading to the activation of Akt (protein kinase B), which regulates cell survival and growth shown in **Figure 4**.

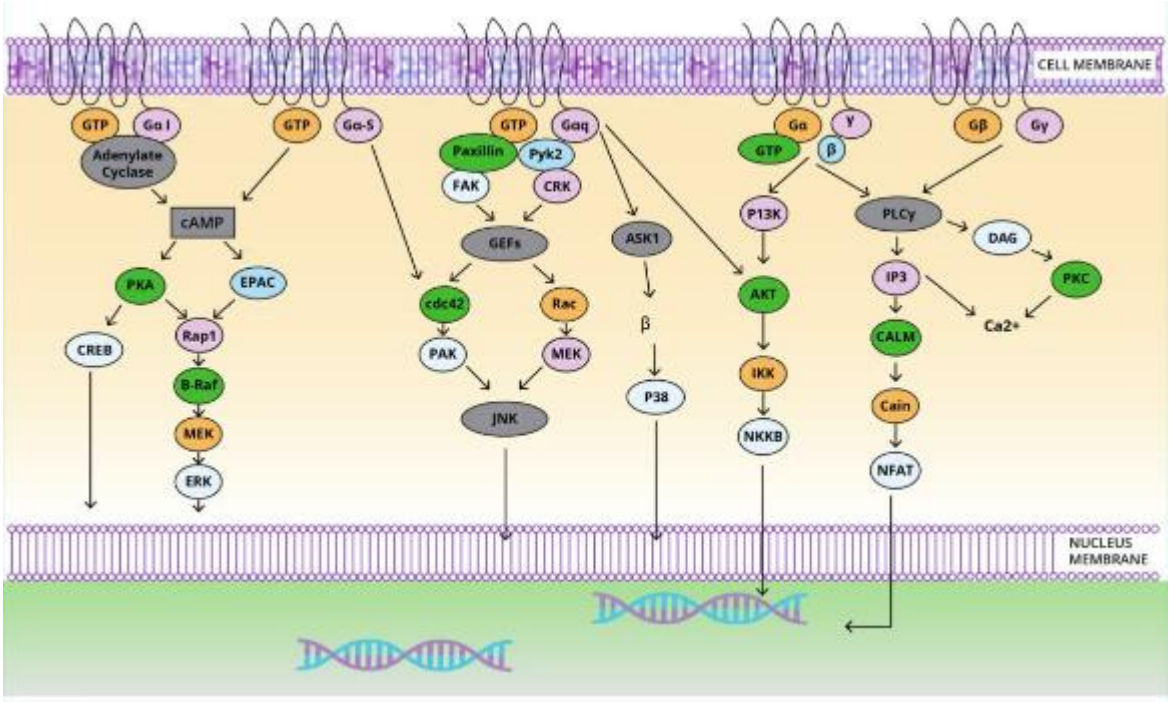


Figure 4: Adhesion GPCRs Signaling Pathways

2.11.1.4 Immune Signaling Pathways

The GPCRs, which are also usually found in immune cells, like the ADGRE2, and they are playing a part in the immune signaling pathways. These routes may include immune cell activation, chemotaxis and the interplay with other immune cells or antigens. (Hamann et al. 2016b).

2.11.1.5 Cytoskeletal Regulation

GPCRs are a protein family which can influence the cell shape, initiate chemotaxis, regulate the process of cell division, differentiation, and secretion.

The catecholamine and corticosteroid signaling through the G protein-coupled receptors (GPCRs), assisted by the different G protein families, is the way of changing the actin cytoskeleton. The GPCR/G12/13 axis seems to be the most powerful signal for the polymerization of actin, whereas the GPCR/Gs axis works as the main inhibitor of the actin polymerization. This control of actin is very dynamic and can, therefore, influence a whole set of cellular processes, such as motility, survival, adhesion, polarity, morphology, and gene expression, which makes cells to respond very precisely to the different stimuli appear in **Figure 5**.(Vázquez-Victorio et al. 2016).

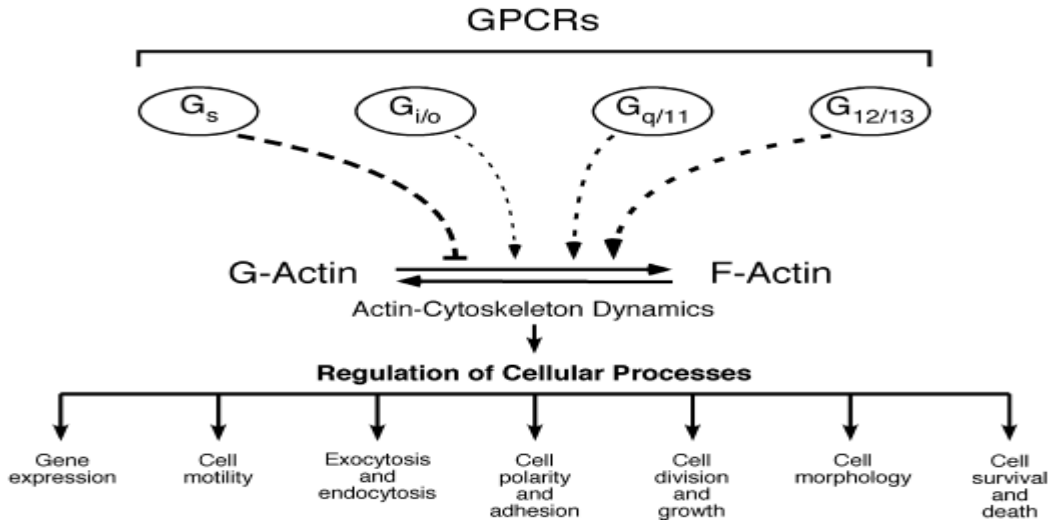


Figure 5: Signaling pathways of GPCRs regulate the dynamics of the actin cytoskeleton, contributing to a variety of physiological processes.

2.11.1.6 Phagocytosis

The cells on communication and the defense mechanisms are the factors that determine the efficient immune responses. Arigently, anADGPCR ADGRB1 and other members of the ADGRE

class are associated with functions such as phagocytosis, activation, and migration of white blood cells.(Nijmeijer et al. 2016b).

2.11.1.7 Tissue Development and Homeostasis

Inside the wide GPR-auto proteolysis inducing (GAIN) domain, the GPCR proteolysis site (GPS) is where the G protein-coupled receptors (aGPCRs) auto proteinulize. The latest developments in the research of aGPCR have greatly contributed to the extension of the knowledge of their functions in organogenesis. These receptors are in charge of controlling cell signaling, intercellular adhesion, cell-matrix interactions, and other important parts of organ development. They are the key elements for the basic functions of the body like the motility promotion, cell adhesion, cell polarity regulation and the spindle orientation and thus are the jealously guarded by the cells for their survival. The evolution of the important body systems including the brain, the musculoskeletal system, the kidneys, the cardiovascular system, the hormone production and the process of the precise control of the immunological activities has been connected to the different aGPCRs. (Sreepada et al. 2022).

2.12 Pathophysiology of ADGRE2 gene

2.12.1 Missense mutation in ADGRE2 gene

The alteration of the ADGRE2 gene, which is also called EMR2, to cysteine for tyrosine at amino acid position 492 (p. C492Y) is the reason for the vibratory urticaria. People with this autosomal dominant disease have localized hives and systemic signs when they touch the skin. The given phenomenon is together with mast cell degranulation and the histamine in the bloodstream is elevated. The ADGRE2 receptor itself is then cut by its self, thus forming an extracellular subunit that binds to a transmembrane subunit. This genetic variant probably is the reason for a break in the regulatory subunit interaction, thus, the mast cells become more sensitive to the vibration-induced degranulation even when there is no IgE stimulation.(Boyden et al. 2016).

Mechanical activatation of the mast cells of human which are expressing the p. C492Y-ADGRE2 mutant and coupled to dermatan sulfate, an ADGRE2 ligand, led to different reactions and intracellular signals. The situation of the presence of p. C492Y-ADGRE2 led to the extension of the degranulation and the number of responsive mast cells of the increase and the lowering of the

activation threshold. Vibration activated phospholipase C which in turn led to transient increases in the cytosolic calcium levels, phosphoinositide3-kinase and extracellular signal-regulated kinases 1 and 2 were then activated. These responses were regulated by Gβγ, Gαq/11, and Gαi/o proteins through different ways. Vibrating wheel of the cell induced degranulation through phospholipase C pathways that are associated with calcium, protein kinase C, and phosphoinositide 3-kinase. On the other hand, extracellular signal-regulated kinases 1/2 pathways were not involved. Dr. J.H. Harrison proved that the Pt-sensitive signals were necessary for the degranulation process. Besides, prostaglandin D2, a mediator not before linked with vibratory urticaria, was secreted and released during the mechanical activation of mast cells. The main adrenaline ones came from the activation of kinases by extracellular signals, as well as calcium, protein kinase C, and to a lesser extent, phosphoinositide 3-kinase. (Naranjo et al. 2020).

2.12.2 Role of ADGRE2 on the neutrophil

New data are showing that the participation of ADGRE2, a protein linked to CD97 and part of the EGF-TM7 family, in the migration and adhesion processes of myeloid cells during their differentiation, maturation, and activation is emerging. When the ADGRE2 becomes bound, it has the ability to improve the adhesion, migration, and the production of the antimicrobial mediators, thus, on top of that, systemic inflammation will be intensified.

Experiments have shown that in patients with SIRS, the ADGRE2 on the neutrophils is expressed and is correlated with the severity of organ failure. Furthermore, cirrhotic patients, especially those with infectious complications, have the noticeably high levels of ADGRE2 expression on neutrophils. Neutrophils that are expressing the protein display an activated phenotype but have impaired functions like an increased resting oxidative burst and reduced phagocytosis ability.

ADGRE2 is also capable of killing off neutrophils, which is an interesting fact. The evidence which has been discovered that shows EMR2 expression on neutrophils is linked to sepsis has been presented showing that ADGRE2 is a ligand of LPS and it inhibits the survival of neutrophils. Thus, taking into account these findings, it becomes clear that ADGRE2 expressing neutrophils have both the roles in inflammation and may be the reason of the immune dysregulation, a very important pathological process in sepsis.(Huang et al. 2016).

2.12.3ADGRE2 role in inflammatory diseases

Through its interaction with FHR1, a regulatory protein of complement Factor H, ADGRE2 has been found to regulate the NLRP3 inflammasome in monocytes. However, the exact functional consequences of EMR2 activation and the signaling pathways it uses to activate the NLRP3 inflammasome remain unclear. In both primary monocytes and the THP-1 monocytic cell line, ADGRE2-mediated signaling provides the second signal for NLRP3 inflammasome activation. Stimulation of ADGRE2 by its agonistic 2A1 monoclonal antibody initiates a Gα16-dependent PLC-β activation pathway. This pathway leads to downstream activities such as Akt, MAPK, NF-κB, and Ca2+ mobilization, culminating in K+ efflux. Intriguingly, the NLRP3 inflammasome, which is associated with inflammatory processes, can be activated by simply reducing the intracellular K^+ content (I et al. 2020).

CHAPTER 3: MATERIALS AND METHODS

3.1 Data retrieval and Processing

Data for ADGRE2 gene ENSG00000127507 was retrieved from the databases including Ensembl (Howe et al. 2021), Catalogue of Somatic Mutations in Cancer (COSMIC) (Tate et al. 2019) and Genome Aggregation Database (gnomAD) (Karczewski et al. 2020). The protein transcripts, genomic sequence, SNP data which includes Variant IDs, amino acid coordinates and alterations in nucleotide base and residues were systematically acquired from each of three databases. The total number of variants from all the databases were combined and total variants of ADGRE2 gene were determined. The data for the missense, coding, frameshift, splice variants and untranslated region (UTR) variants was collected. The missense variants of ADGRE2 were filtered from the rest of the variants from all three databases and merge these variants. The variants that were reported by each database as missense counted them once and graph was plotted by calculating the total number of missense variants for ADGRE2 gene. The number of missense mutations per amino acid for ADGRE2 was calculated. The information of domains of protein encoded by ADGRE2 gene was collected by using protein amino acid sequence The amino acid sequence for the protein encoded by ADGRE2 gene was retrieved from Ensemble database and this sequence was used for the collection of information of domains of protein from InterPro database (Blum et al. 2021). The complete protein domains, including number of amino acids within each domain were calculated. Subsequently, the frequency of missense variants for each amino acid and within each domain was determined. The number of exons within the gene sequence and amino acids within each exons were noted from NCBI database. .

3.2 Prediction of Pathogenicity

For the analysis of pathogenic missense variants data six tools were selected namely SIFT (Sim et al. 2012), Polyphen (Adzhubei et al. 2010) Mutation Assessor (Reva et al. 2011), Meta-Lr, Revel (Ioannidis et al. 2016) and Cadd (Rentzsch et al. 2021) These tools evaluate all variants, assigning each a specific score based on its potential pathogenicity as determined by the algorithm. The potential pathogenicity of all missense variants score from six tools was calculated and the variants with the potential pathogenicity score of 80% and above 80% were filtered out. Polyphen provides

scores ranging from 0 to 1, with classifications of benign (0-0.4), possibly damaging (0.4-0.8), and probably damaging (0.9-1). Variants with PolyPhen scores ≥ 0.999 were chosen. SIFT scores range from 0 to 1, where 0 signifies deleterious and 1 signifies tolerated. Variants with scores ≤ 0.05 were selected. REVEL categorizes variants as benign (<0.5) or likely disease-causing (>0.5), with variants scoring ≥ 0.5 chosen. MetaLR categorizes SNPs as damaging (0.5-0.9) or tolerated (0-0.4), with variants scoring ≥ 0.5 filtered out. Mutation Assessor assigns SNPs into high (>0.9), medium (0.5-0.9), low (0.2-0.4), and neutral (0.0-0.1) groups, with variants scoring ≥ 0.7 selected for processing. CADD utilizes "likely benign" and "likely deleterious" terms for scores ranging from 0 to 30 and 31 to 35, respectively. After stringent filtering, only one missense variant was retained for further analysis.

3.3 Structure Prediction and Validation

The complete structure of protein encoded by ADGRE2 was unavailable in protein data bank, so protein structure prediction was performed through I-TASSER (Iterative Threading ASSEmbly Refinement) (Yang et al. 2015) which employs the threading technique to predict the structure on the basis of peptide sequence, assigning the confidence score ranging from -5 to 2, based on the significance of template alignment. For the subsequent analysis, the model exhibiting the highest score was chosen and visualized its structure using the PyMOL (The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC). The domains of protein structures were validated through InterPro and highlighted in PyMOL.

3.4 Sub-Cellular localization

The Sub-cellular localization of ADGRE2 protein was predicted by using the tool Deeploc1.0 (Almagro Armenteros et al. 2017)which predicts the location of protein on the basis of amino acid sequence information by identifying the regions important for particular cellular location.

3.5 Stability analysis

Genetic variation has effect on the stability of protein encoded by gene. To analyze the effect of SNP on the stability of protein different tools were used like I-Mutant2.0 (Capriotti et al. 2005),

MUpro (Chen et al. 2024), Dynamut (Rodrigues et al. 2018). I-Mutant2.0 predicts the alteration in the stability of protein caused by variation in the terms of DDG (Kcal/mol). A DDG value <0 indicates decrease while >0 indicates increase in the stability of protein as result of variation. MUpro is another software that was used for the stability analysis after the Single Nucleotide Polymorphism. This tool is also presented a results in the terms of DDG values. Increased stability of protein is indicated by values >0 and decreased stability of protein is indicated by values <0. Dynamut evaluates shifts in protein stability, alterations in vibrational entropy (DDS), and changes in molecular dynamic characteristics attributed to SNPs.

3.6 Structural and Functional analysis

Genetic variations also have effect on structure and function of proteins. Different tools were used for these analysis. Project HOPE (Venselaar et al. 2010) was used to analyse the functional impact of the missense variant on the protein. Dynamut (Rodrigues et al. 2018). Was used to anlyze structural flexibility of the variant on protein and lastly, Mutpred2 (Pejaver et al. 2017)was used to understand the functional impact of the selected missense variant.

3.7 Effect of polymorphisms on RNA stability

The effect of missense mutation D67N on the stability of ADGRE2 mRNA secondary structure was evaluated through the RNAFold software (Lorenz et al. 2016)on the basis of thermodynamic parameters. The effect of change in the stability of mRNA secondary structure was predicted from the results of RNAFold software.

3.8 Primer Designing

The Primers for ARMS PCR were designed using the tool **Primer 1**. Two Primers were variation specific and third one primer was common shown in **Table 1**. Primers were designed for investigating the genotype of ADGRE2 gene in the Control and Patient data. The primers were validated by using the UCSC in silico PCR (Kent et al. 2002). Primer optimization was done to establish the condition for the amplification of ADGRE2 gene.

Table 1: ADGRE2 Primer sequences for ARMS PCR

ADGRE2 Primers for ARMS PCR		
D67N (rs765071211)	Common Primer	CCCCATGGAGACTTGTGCCG
	C primer (Wild type)	GCTGCCCTCAAGCCTCTGTCCT
	T Primer (variant)	GCCCAGGCTGGTCTTGAATTCC

3.9 IRB and Sample Collection

The Institutional Research Review Board of the National University of Science and Technology (NUST) gave permission to the study. Only blood specimens were taken from the Chronic Myeloid Leukemia (CML) patients, while the patients with the other morbidities were excluded from the research. Those who did not have any disease like hypertension or diabetes were considered as healthy controls for genotyping purposes. The purpose of the study was to find the pathogenic variants in the CML group and the control group. The study included 50 samples of controls and 58 samples of CML patients.

3.10 DNA Extraction

The DNA extraction was done by the way of the phenol-chloroform method for both the diseased and the control samples. The extraction protocol involved the preparation of four buffer solutions: Four solutions A, B, C, and D. For solution A, precise amounts of 0 are to be determined. 32 mM of sucrose, 10mM of Tris-base, 5mM of MgCl2, and 1% V/V of Triton were mixed with distilled water, and thus, a final volume of 500ml was obtained. Triton has the ability to withstand heat,

whereas Tricyhos was added after the autoclaving because of its heat-sensitivity. The pH of solution A was brought up to 7. 5-8, the mixture is then introduced with the concentrated HCL and NaOH (40%w/v). Thus, this approach makes the cell membrane to break down to get DNA out. Solution B was a mixture of 10mM Tris-base, 400mM Sodium Chloride, and 2mM Ethylene Diamine Tetra Acetic Acid (EDTA) with distilled water, and the volume of the mixture was 500ml. The pH was modified to reach a value of 7. The length of the experiment was from 5 to 8 and the solution was autoclaved. Solution B is a favorable factor for DNA precipitation and protein separation. Solution C, which is based on phenol, was wrapped in aluminum foil for the reason of the protection from the light. Phenol is necessary for the division of the aqueous and organic layers. Solution D was made by mixing chloroform and isoamyl-alcohol with distilled water, which resulted in a volume of 500ml. This answer solves the problem of protein coagulation and foaming hence avoiding the pro time of DNA isolation.

The phenol-chloroform DNA extraction method takes two days. On the first day, 750µl of blood was mixed with an equal volume of solution A in an Eppendorf tube, then after that, for 10 minutes, it was incubated at room temperature. Centrifugation at 13000rpm for 1 minute was followed by the removal of the supernatant, and the pellet was centrifuged again with 750µl of solution A to guarantee the pellet was clean. Next, 400µl of solution B, with the addition of 12µl of 20% SDS and 5µl of proteinase K were mixed to dissolve the pellet, and then centrifuged at 13000rpm for 1 minute, after which the supernatant was discarded. The samples were incubated at 37 degrees Celsius for a whole night.

On the next day, 250µl of both solution C and D were added into the samples which were incubated before. Thereafter, the Eppendorf tubes were spun down at 13000 rpm for 10 minutes to separate the aqueous and organic layers. The aqueous layer, which was filled with the DNA, was carefully moved to a separate Eppendorf tube. To cause the DNA to precipitate, 55µl of sodium acetate and 500µl of chilled isopropanol were poured into the tube, and the tube was then inverted to mix the contents. After centrifugation at 13000 rpm for 10 minutes, the supernatant was drained. Thus, the resulting pellet was re-suspended in 200µl of chilled ethanol and another centrifugation for 8 minutes at 13000 rpm was performed. The supernatant was thrown away again, and the tubes were placed upside down to speed up the evaporation of ethanol. After it dried, the DNA was re-suspended in 200µl of the PCR water.

3.11 ARMS PCR

The amplification-refractory mutation system (ARMS) is a straightforward method for the detection of mutations, mainly of single-base alterations or minor deletions. The technology is based on the PCR primers that are tailor-made for the target sequence thus, it implies that the test DNA gets amplified only if the allele of interest is present. After an ARMS reaction, the PCR product, which indicates the presence or absence of the target allele, is the result of the presence or absence of a PCR product. In the experiment, three Primers were used for ARMS PCR, in which one was common for both alleles and the other two were the ones specific to each allele. The Solis Biodyne FIREpol master mix and The Applied Biosystems™ Veriti™ 96-Well Thermal Cycler were the instruments that were used. For each sample a PCR reaction mixture of 12µl which is composed of 6µl master mix, 1µl common primer, 3µl PCR water, 1µl allele-specific primer and 1µl DNA sample was prepared.

The PCR tubes were centrifuged for a few seconds to get rid of the bubbles and to make sure that the components of the tube were thoroughly mixed before being transferred to the Thermocycler for PCR reaction. At first, primer optimization was done by doing multiple reactions at different temperatures through the Gradient Thermocycler.

The variables that were used for PCR shown in **Table 2**. After the best temperature has been found, PCR reactions for all samples were done at that temperature. The PCR reaction had three steps and 30 cycles as a total. The first step involved the heating of DNA up to 95 degrees Celsius for 10 minutes, which was then followed by 30 cycles with each lasting 45 seconds. After the second step, annealing took place at 60 degrees Celsius for 45 seconds. The last phase comprised of primer extension at 72 degrees Celsius for 45 seconds after the annealing. An end paper was also extended for 7 minutes at 72 degrees Celsius. After PCR tubes were finished, they were frozen in the freezer at 4 degrees Celsius until the gel electrophoresis was done.

Table 2: PCR Conditions

Steps	Temperature	Time	Cycles
Initial Denaturation	95◦C	10 minutes	1
Denaturation	95◦C	45 seconds	30 cycles
Annealing	60◦C	45 seconds	30 cycles

Extension	72◦C	45 seconds	30 cycles
Final extension	72◦C	7 minutes	1 cycle
Storage	4◦C	∞	

3.12 GEL Electrophoresis

The extracted DNA was visualized by the preparation of a 1% w/v agarose gel. This process was done by first dissolving 1 gram of agarose in 50ml of 1x TAE buffer. The mixture was warmed up in a microwave to the point that it became clear, and then 5 µl of ethidium bromide was added for DNA staining. Following the careful mixing, the liquid was put into a gel caster and left to solidify at room temperature. The gel was finally set, and then was placed in the buffer tank of electrophoresis which was filled with 1x TAE buffer. Samples made of the 8µl of DNA mixed with the 1µl of loading dye were filled into the wells of the gel. Electrophoresis was done for 20 minutes at 80V, and the DNA bands were observed under the UV trans-illuminator.

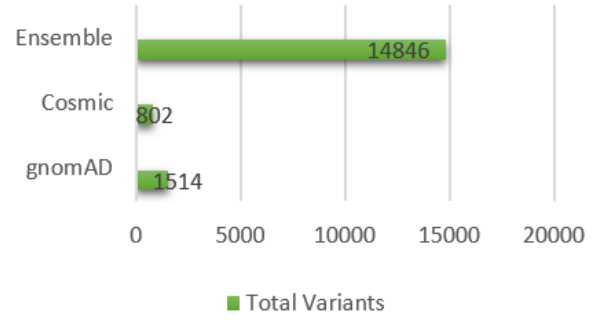
In the case of PCR visualization, a 2% w/v agarose gel was made using the same procedure as that given above. Afterwards, 10µl of the PCR product was placed in the gel wells, and electrophoresis was performed for 40 minutes at 80V. The PCR bands thus obtained were finally seen under UV trans-illumination.

CHAPTER 4: RESULTS

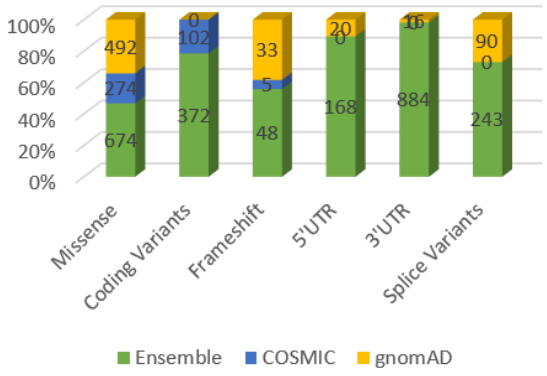
4.1 SNP Identification:

The total 17162 variants were collected from all three databases, of which 1514 were from gnomAD, 802 from COSMIC and 14846 from Ensemble, as shown in **Figure 6**(A). The missense variant data from all three databases, was divided and combined afterwards and the total number of unique missense variations was 921 in **Figure 6**(B). For instance, splice site, missense, frameshift, coding sequence, 5' UTR and 3' UTR variants were plotted in a graph in **Figure 6**(C) coming from the three repositories. The exon data for ADGRE2 gene was obtained from the ensemble, and the number of amino acids per exon was calculated by using the UniProt Kb. As for the number of mutations for each exon, it was calculated from the data that was received from the ensemble in **Figure 6**(D). Exon number 16 has the most number of amino acids while exon number 7 has the most number of mutations.

A



B



C

D

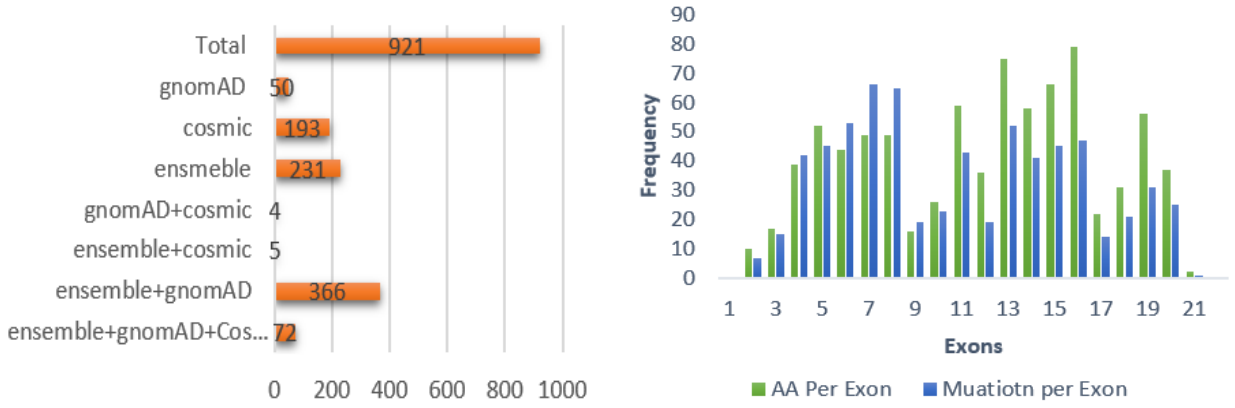
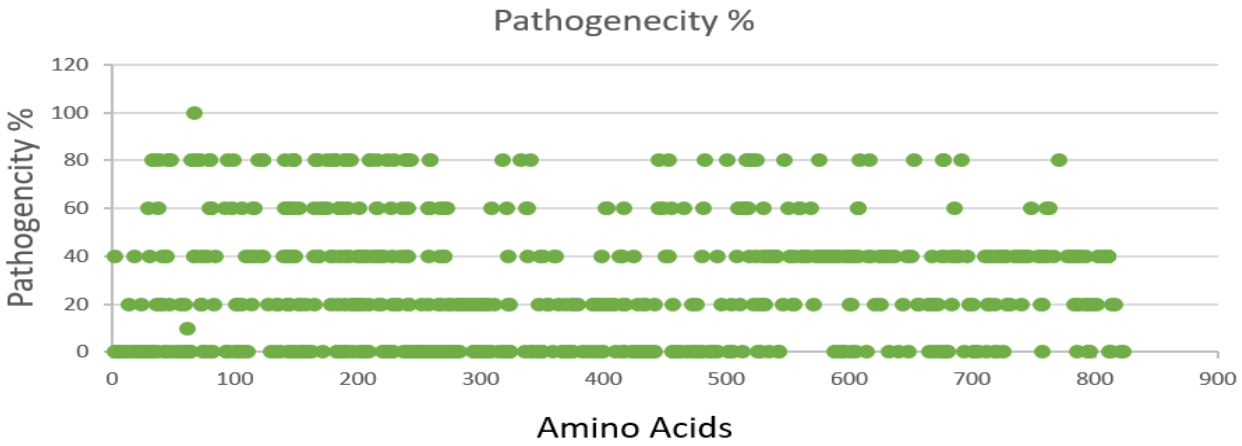


Figure 6: A: Count of KPCI variants from different databases. **B**: Count of total coding variants for ADGRE2. **C:** Count of total missense variants. **D:** Amino Acids and Mutations per Exon for ADGRE2.

4.2 SNP Data Filtration

The missense data was carefully filtered to find only one SNP for the next examination. At first, filtration was the function of the pathogenicity percentages of the SNPs. The chances of these incidents were calculated by the pathogenicity scores from the tools mentioned in the methodology, such as SIFT, PolyPhen, CADD, REVEL, Mutation Assessor, and MetaLR **Figure 7**(A). The SNPs with a pathogenicity percentage of 80% or higher were taken forward **Figure 7**(B). Out of the initial 921 missense mutations, 46 had the pathogenicity percentage that was equal to or above 80% that is listed in **Table 3**.

A



B

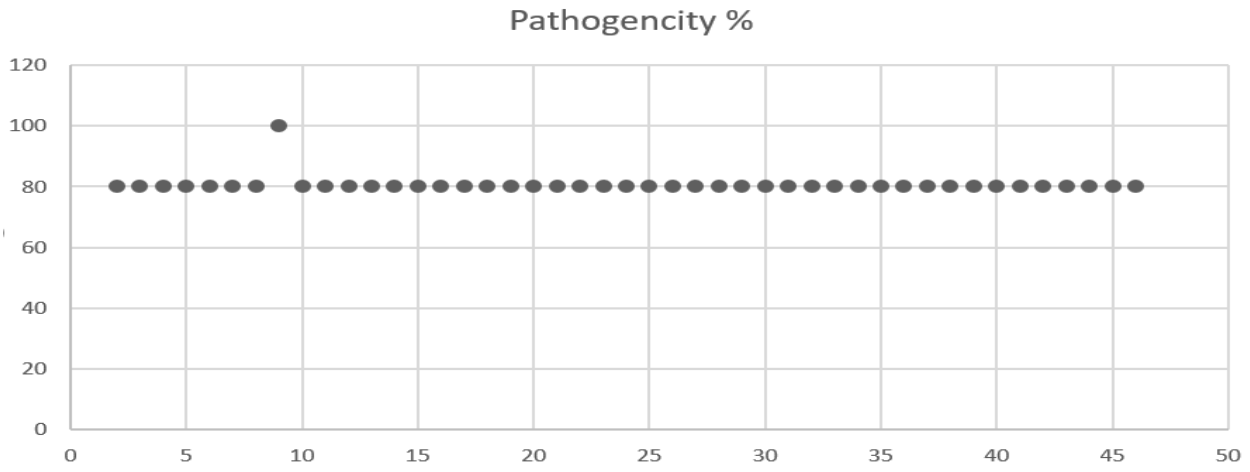


Figure 7: A: Missense variants predicted to be pathogenic by the in Silico tools. **B:** Missense variants with pathogenicity percentage equal to or greater than 80%.

Table 3: Variants with pathogenicity percentage ≥80: 1= Deleterious, 0=Tolerated, Highlighted cells= SNP of interest.

Variant ID	AA coord	Conseq. Type	sift_class	polyphen_	cadd_clas:	revel_clas:	meta_lr_c	Pathogenecity %	polyphen_
rs149725020	33	missense variant	1	1	0	1	1	80	1
rs767699480	39	missense variant	1	1	0	1	1	80	1
rs764094291	45	missense variant	1	1	0	1	1	80	1
rs142343910	46	missense variant	1	1	0	1	1	80	1
rs928575810	46	missense variant	1	1	0	1	1	80	1
rs373171000	47	missense variant	1	1	0	1	1	80	1
rs371292166	65	missense variant	1	1	0	1	1	80	1
rs765071211	67	missense variant~	1	1	1	1	1	100	1
rs747246737	68	missense variant~	1	1	0	1	1	80	1
rs893305633	70	missense variant	1	1	0	1	1	80	1
rs1010286043	70	missense variant	1	1	0	1	1	80	1
rs1215417977	71	missense variant	1	1	0	1	1	80	1
rs747287302	79	missense variant	1	1	0	1	1	80	1
rs747287302	79	missense variant	1	1	0	1	1	80	1
rs1235127973	94	missense variant	1	1	0	1	1	80	1
rs1316691934	99	missense variant	1	1	0	1	1	80	1
rs1354971232	141	missense variant	1	1	0	1	1	80	1
rs778643418	147	missense variant	1	1	0	1	1	80	1
rs763935066	180	missense variant	1	1	0	1	1	80	1

This was the next filtration process which was based on the scores given by the tools to decide if the SNP is deleterious or it is tolerated that is shown in **Table 4**. Thus, SIFT Variants with a score of 0 is one of the examples. The player should have 3 or less, for Polyphen variants with a score of 0. 999 or higher, for REVEL variations with a score of 0. 588 or what you answer gets into the

category, and for MetaLR variants with a score of 0. 5 or more, the students who got such grades were the ones chosen. The thorough filtration of the missense variants by the maxmittd structure allowed the missense variant with the ID rs765071211 to be selected for further in Silico analysis. This mutation is present at the amino acid number 67 in the Egfca_6 domain of ADGRE2 and therefore, it leads to a change of Aspartic Acid to Asparagine at the 67 position.

Table 4: Filtered SNPs on basis of pathogenicity score

Variant ID	vf	Alleles	Conseq. Type	AA	AA coord	SIFT	PolyPhen	CADD	REVEL	MetaLR
rs149725020	2.1E+08	A/G	missense variant	C/R	33	0	0.988	23	0.668	0.948
rs912959685	2.65E+08	G/T	missense variant	P/H	34	0.03	0.672	23	0.539	0.781
rs767699480	2.5E+08	C/T	missense variant	C/Y	39	0.01	0.988	21	0.53	0.933
rs764094291	2.4E+08	C/T	missense variant	C/Y	45	0	0.999	23	0.731	0.951
rs142343910	2.09E+08	C/G/T	missense variant	R/P	46	0	0.985	21	0.566	0.73
rs928575810	2.68E+08	G/A	missense variant	R/C	46	0.01	0.994	24	0.548	0.722
rs373171000	2.14E+08	C/T	missense variant	C/Y	47	0	0.999	23	0.729	0.922
rs371292166	2.14E+08	C/G	missense variant	C/S	65	0.01	0.997	22	0.612	0.829
rs765071211	2.49E+08	C/T	missense variant~sp	D/N	67	0.02	0.967	33	0.585	0.968
rs747246737	2.35E+08	T/A	missense variant~sp	I/F	68	0	0.987	24	0.722	0.879
rs893305633	2.56E+08	C/A/T	missense variant	E/D	70	0.02	0.998	18	0.632	0.965
rs1010286043	2.98E+08	C/T	missense variant	E/K	70	0.04	0.984	21	0.736	0.974
rs1215417977	3.35E+08	A/C	missense variant	C/G	71	0	1	23	0.795	0.993
rs747287302	2.35E+08	A/G/T	missense variant	C/R	79	0	0.999	21	0.705	0.988
rs747287302	2.35E+08	A/G/T	missense variant	C/S	79	0	1	20	0.719	0.985
rs1235127973	3.39E+08	A/G	missense variant	C/R	94	0	0.997	23	0.74	0.988
rs1316691934	3.71E+08	C/T	missense variant	G/R	99	0.01	1	19	0.668	0.949
rs751263352	2.36E+08	C/G	missense variant~sp	D/H	119	0.02	0.892	21	0.666	0.97
rs368628691	2.13E+08	A/C	missense variant	C/W	123	0.01	0.892	22	0.65	0.982
rs764854362	2.49E+08	C/T	missense variant	C/Y	123	0.01	0.65	22	0.767	0.985
rs1354971232	3.9E+08	C/G/T	missense variant	G/R	141	0.03	0.931	21	0.775	0.985

4.3 Structure Prediction and Validation

I-TASSER used the threading technique to predict the structure of ADGRE2, and thus, the five best models were presented based on confidence scores, where the higher scores are the indications of the best quality. The model with the top score of confidence, which specifically is -1, will be the winner. 12 was the only student who got the grade of B+ that was chosen to be examined further. The ADGRE2 culman domain information was validated by InterPro. Every domain of the model was separately colored by PyMol so that the clear visualization of **Figure 8(A)** could be displayed. The parts of the protein depicted in **Figure 8(B).** The EGF_5 domain is represented in raspberry color and goes from 28 to 66 amino acids, thus spanning 39 residues. The egfca_6 (1) domain, which is of marine blue color from residue 67 to 118, is highlighted in this domain and contains a 52 amino acid residues while egfca_6 (2), which is of sand color, contains a 49 amino acid residues. The final egfca_6 (3) domain was colored in magentas from 212 to 260 and was

composed of a 49 amino acids. The EGF_3 domain which is highlighted in cyan color is shown from amino acid residues 119 to 162. GPS_3 domain is from 478 to 529 amino acid residues shown in the orange color and 7tmB2 EMR domain is highlighted in the hot pink color from 533 to 795. Quantity and variety of amino acids in each domain are displayed in Table 6.

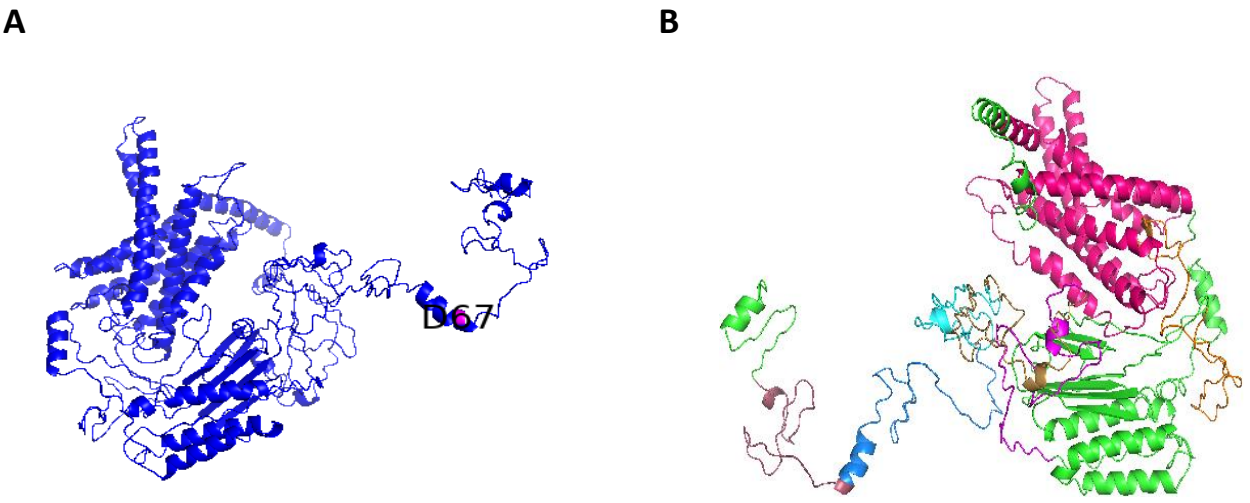


Figure 8: ADGRE2 predicted structure. **A:** ADGRE2 3D Structure **B:** ADGRE2 Domains: egf_5, egfca_6 (1), egfca_6 (2), egfca_6 (3), EGF_3, GPS_3 and 7tmB2 EMR domain.

Table 5: ADGRE2Domains and their corresponding residues

	Domain	Amino acid coordinates	AA number
1.	egf_5	28-66	39
2.	egfca_6 (1)	67-118	52
3.	egfca_6 (2),	163-211	49
4.	egfca_6 (3)	212-260	49
5.	EGF_3	119-162	44
6.	GPS_3	478-529	52
7.	7tmB2 EMR	533-795	262

4.4 Cellular localization

Deeploc 1. 0 examined the subcellular localization of ADGRE2 and gave the potential cellular locations with the likelihood scores in a table format. **Table 6** shows the highest likelihood score allocated to the cell membrane, which is in line with the literature, as ADGRE2 is famous for its significant role in cell-cell interaction. The inner workings of the cell are also shown in **Figure 9**.

Table 6: ADGRE2 Sub-Cellular Localization

Localization	Likelihood
Cell Membrane	0.9995
Endoplasmic Reticulum	0.0002
Lysosome/ Vacuole	0.0002
Golgi Apparatus	0.0001
Mitochondria	0
Peroxisome	0
Nucleus	0
Extracellular	0
Cytoplasam	0
Plastid	0
Type	Likelihood
Soluble	0
Membrane	1

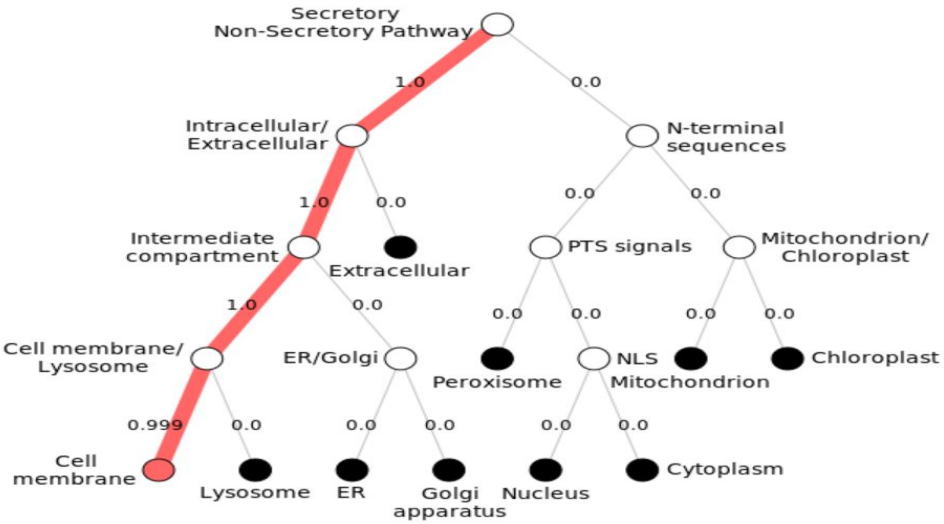


Figure 9: ADGRE2 Sub-Cellular Localization

4.5 Stability analysis

Stability analysis using I-Mutant2.0 indicated a DDG value of -0.44 kcal/mol for the D67N polymorphism, indicating a decrease in protein stability. Similarly, Mupro analysis showed a DDG value of -1.1507558 for D67N, signifying reduced protein stability due to the variation. Dynamut calculated the ΔΔG value to illustrate the impact of the D67N mutation on ADGRE2 free energy, resulting in a ΔΔG value of 0.13 kcal/mol. This indicates a stabilizing effect of the amino acid alteration on the protein structure of ADGRE2.

4.6 Functional analysis

The score that Mutpred2 gives for pathogenicity of missense mutations is a proof of its usefulness. Scores above 0. 5 are viewed as pathogenic, with a level of 0 being the threshold. 68 test subjects were given a 10% false positive rate (FPR), and 0 of them were found to be honest. The percentage of students who did not pass their PAPB from 80 to a 5% FPR. The D67N gene mutation gives a Mutpred2 score of 0. The 095, which is in the pathogenic category, is the one that kills abruptly.

4.7 Effect of polymorphisms on RNA secondary rs765071211 structure

MFE structure is the RNA secondary structure which is associated with the lowest free energy. The lower the thermodynamic energy, the better the predicted secondary structure stability. The RNA secondary structures, which contain the position of the SNP rs765071211 in the 120-

nucleotide sequence, are compared in **Figure 10(A)** where the wild-type RNA had the minimum free energy of -37. The aim is to reformulate the given sentence. In this case, the sentence could be rephrased as: 40 kcal/mol, but the mutant in **Figure 10(B)** had 36 kcal/mol. The 0 kcal/mol, being a sign of the decrease in stability.

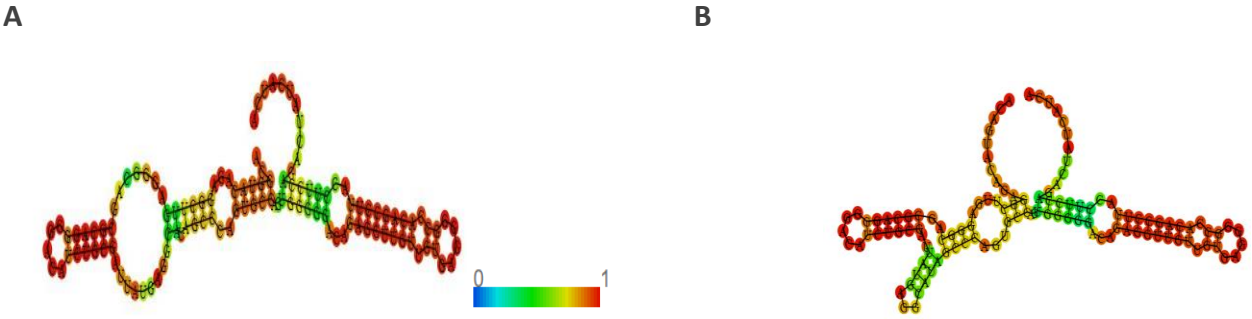


Figure 10: Effect of ADGRE2 variant on its RNA secondary structure. A: Comparison of MFE structure drawing encoding base-pair probabilities between wild type and mutant variant.

4.8 Genotyping

Thus, the statistical analysis of the association between Chronic Myeloid Leukemia and Mutated ADGRE2 was done by the application of the Fisher exact test. Statistical analysis of the ADGRE2 variant rs765071211 (D67N) in **Table 7** revealed the following associations: the homozygous CC genotype was not found to be significantly linked with CML (p=0.). 3879, odds ratio=1. 484, relative risk=1. On the contrary, the homozygous TT genotype had a possible protective effect, which was linked to a lower risk of CML (p<0.0001, odds ratio=0. 1374, relative risk=0. 3272). Furthermore, the CT heterozygous genotype showed the biggest genetic link with the risk of CML (p=<0. 000, odds ratio=8. 000, relative risk= 2. 400).

Table 7: Genotype analysis of rs765071211 D67N.

Genotype	**Patient (n=58)**	**Control (n=50)**	**Odds Ratio**	**95% CI Odds Ratio**	**Relative risk**	**95% CI Relative risk**	**P value**
	n (%)	n (%)					
CC	13 (46.43%)	15 (53.57%)	1.484	0.6150 to 3.370	1.224	1.7717 to 1.819	0.3879

TT	39 (78.00%)	11 (22%)	0.1374	0.05789 to 0.3344	0.3272	0.1851 to 0.5469	<0.0001
CT	6 (20.00%)	24 (80.00%)	8.000	2.880 to 21.70	2.400	1.662 to 3.464	<0.0001

The **Table 8** shows that in genotype analysis of individuals aged, the CC homozygous genotype was not related to CML, while the heterogeneous TT genotype demonstrated a significant association with a higher risk of CML.

Table 8: Genotype analysis on the basis of age below and Above 40.

Genotype	**Below 40 (n=32)**	**Above 40 (n=26)**	**Odds Ratio**	**95% CI Odds Ratio**	**Relative risk**	**95% CI Relative risk**	**P value**
	n. (%)	n.(%)					
CC	9 (64.29%)	5 (35.71%)	1.219	0.3873 to 3.783	1.078	0.6267 to 1.586	>0.9999
TT	15 (42.86%)	20 (57.14%)	0.1800	0.05559 to 0.5678	0.5314	0.3383 to 0.7884	0.0024
CT	8 (87.5%)	1 (12.5%)	4.412	0.6209 to 52.42	1.487	0.8206 to 2.018	0.2306

Statistical analysis of genotypes based on gender in **Table 9** proved that the genotype is not related to any particular gender among the patients.

Table 9: Genotype analysis on the basis of gender.

Genotype	**Male (n=35)**	**Female (n=23)**	**Odds Ratio**	**95% CI Odds Ratio**	**Relative risk**	**95% CI Relative risk**	**P value**
	n. (%)	n.(%)					
CC	8 (61.54%)	5 (38.46%)	1.067	0.3171 to 3.515	1.026	0.5696 to 1.562	>0.9999

TT	24 (61.54%)	15 (38.46%)	1.164	0.4084 to 3.609	1.063	0.6978 to 1.779	>0.9999
CT	3 (50.00%)	3 (50.00%)	0.6250	0.1365 to 2.907	0.8125	0.2994 to 1.43	0.6727

Wide spread health studies indicated that blood group O was the most common among the patients followed by blood group A and B. The statistical analysis of genotypes based on blood group in **Table 10** showed that people with blood group O had the TT genotype and therefore, they were at a higher risk of developing CML as compared to the other blood groups.

Table 10: Genotype analysis on the basis of Blood Group (O and A+B).

Genotype	**O (n=29)**	**A+B (n=29)**	**Odds Ratio**	**95% CI Odds Ratio**	**Relative risk**	**95% CI Relative risk**	**P value**
	n. (%)	n.(%)					
CC	1 (7.69%)	12 (92.31%)	0.05060	0.004586 to 0.3558	0.1236	0.02183 to 0.5502	0.0010
TT	25 (64.10%)	14 (35.90%)	6.696	1.872 to 20.68	3.045	1.400 to 7.719	0.0045
CT	3 (50.00%)	3 (50.00%)	1.000	0.2176 to 4.596	1.000	0.3645 to 1.833	>0.9999

The statistical analysis of the genotypes of the blood group A in **Table 11** showed that the people with this blood group are not associated with any genotype.

Table 11: Genotype analysis on the basis of Blood Group (A and O+B).

Genotype	**A (n=18)**	**O+B (n=40)**	**Odds Ratio**	**95% CI Odds Ratio**	**Relative risk**	**95% CI Relative risk**	**P value**
	n. (%)	n.(%)					

CC	6 (46.15%)	7 (53.85%)	2.357	0.6827 to 8.805	1.731	0.7674 to 3.492	0.1947
TT	10 (25.64%)	29 (74.36%)	0.4741	0.1645 to 1.398	0.6090	1.2952 to 1.310	0.2367
CT	2 (33.33%)	4 (66.67%)	1.125	0.1978 to 5.244	1.083	0.2995 to 2.692	>0.9999

The statistical research of the genotypes based on the blood group in **Table 12** showed that people with blood group B and genotype CC have a risk of association while the other two genotypes were not statistically significant.

Table 12: Genotype analysis on the basis of Blood Group (B and O+A).

Genotype	**B (n=18)**	**O+A (n=40)**	**Odds Ratio**	**95% CI Odds Ratio**	**Relative risk**	**95% CI Relative risk**	**P value**
	n. (%)	n.(%)					
CC	6 (46.15%)	7 (53.85%)	5.829	1.529 to 26.33	3.600	1.325 to 9.448	0.0189
TT	4 (25.64%)	29 (74.36%)	0.2365	0.06950 to 1.027	0.3290	0.1146 to 0.9330	0.0741
CT	1 (16.67%)	5 (83.33%)	0.7200	0.05605to 5.042	0.7667	0.1311 to 3.137	>0.9999

Figure 11 shows a statistical analysis of Platelet count of CML patient and it was found that Platelet count was highest in the patient with TT genotype.

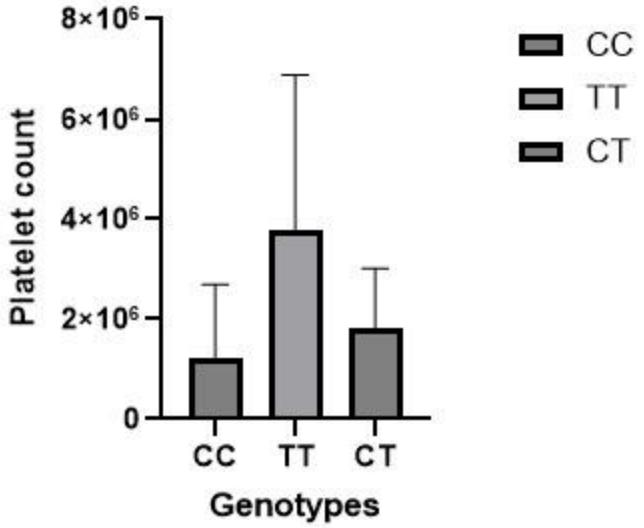


Figure 11: Platelet Count in CML Patient

Chapter 5: Discussion

Single-nucleotide polymorphisms (SNPs) are the most common type of genetic discrepancies in the human genome and are primarily responsible for individual differences. These genetic markers are increasingly important in clinical diagnosis and therapy. SNPs are associated with various harmful conditions, including cancer, heart disease, hereditary metabolic disorders, and autoimmune diseases Wu et al. 2023. When an SNP occurs in a gene involved in regulating cell growth, division, or repair, it can interfere with normal cellular functions and potentially result in uncontrolled cell growth, which is a hallmark of cancer.

The purpose of this research is to study the effect of missense variants on the structure and function of ADGRE2 through flexibility analysis by Dynamut and to detect the missense variants associated with chronic myeloid leukemia and its pathological and clinical parameters. This association of the ADGRE2 with CML can be investigated by the genotype analysis in patient and control samples.

Alternatively referred to as EGF-like module-containing, mucin-like, hormone receptor-like 2 (EMR2), the ADGRE2 gene is located on human chromosome 19 at location 19q13.31 on the long arm (q arm). A G protein-coupled receptor involved in cell adhesion and migration is encoded by this gene. As a member of the adhesion G protein-coupled receptor (aGPCR) family, it is expressed in the brain system and immune cells, among other areas. It is believed to play a role in a number of disorders and is known to participate in immunological responses, while its precise roles are still being investigated. Its expression has been found in breast and colorectal adenocarcinomas, but it is rarely seen in tumor cell lines and tumors.

To study the link of ADGRE2 with CML it is important to find the non-synonymous SNPs on ADGRE2 that have the link in causing this disease. For the analysis of the effect of missense variants on ADGRE2 the three-dimensional structure of the protein is required. The amino acid sequence of ADGRE2 in FASTA format helped in getting the 3-D structure of the protein from alphaFold. The AlphaFold is a bioinformatics tool the helps predict the structure of the protein by directly predicting the coordinated of the heavy atoms through the given amino acid sequence and at the end it align the homologous sequence as the input (Jumper et al. 2021). The confirmation of the structure of ADGRE2 predicted from alphaFold was done through InterPro that helped in the getting the accuracy about the domains and sequence of the protein.

A total of 17,162 variants were gathered from three databases, with 1,514 from gnomAD, 802 from COSMIC, and 14,846 from Ensemble. These variants were evaluated for pathogenicity using six computational tools: Polyphen2, SIFT, REVEL, MeraLR, Mutation Assessor, and CADD. Based on their results, one pathogenic missense variant, D67N, was selected. This variant involves the substitution of aspartic acid with asparagine at position 67 of the ADGRE2 amino acids and is identified by the rsID rs765071211 (C/T). Following these strict selection criteria, this SNP was chosen and further analyzed for its in-silico and genotype link with ADGRE2.

I-Tasser tool was utilized to predict the 3D structure of EMR2 protein. The tool generated five different models, each with a confidence score indicating its quality. The model with the highest confidence score (-1.12) was selected for further analysis. The domains were further identified by using Interproscan. The selected D67N variant of ADGRE2 was located on the Egfca_6 domain of ADGRE2. Tools such as I-Mutant and Project Hope were used to study the structural and functional effects of the D67N missense variant on ADGRE2 which resulted in the change in stability of the protein. The Dynamut tool was used to do flexibility analysis of the D67N variant. The results had shown decrease in the molecular flexibility of the mutant structure because of change in Aspartic acid to Asparagine as compared to the wild type structure of ADGRE2.

In Silico analysis of the D67N (rs765071211) mutation in ADGRE2 was conducted to understand its role in chronic myeloid leukemia (CML). For genotypic analysis of the rs765071211 missense variant in ADGRE2 associated with CML, the tri primer ARMS-PCR method was employed. Primers targeting the variant were designed using the Primer1 tool for PCR amplification.

GraphPad Prism 9 was used to analyze the genotyping data collected statistically. The Chi-square test was utilized for both the patient and control gatherings. Moreover, risk and odd ratios, alongside their determined confidence intervals, were determined utilizing the Fisher Exact test. Measurable importance was thought about when the p-value was beneath 0.05(Khan et al. 2023).

The genotypic analysis revealed that genotypes TT and CT showed significant results with a P-value >0.05, indicating their potential involvement in CML. However, the CC genotype did not show any significant association, excluding it as a risk factor for CML. The heterozygous CT genotype was found to potentially contribute to the development of CML, with some studies suggesting that the TT genotype may be protective while others indicate a risk for the disease.

Furthermore, the association of genotypes of rs765071211 with different risk factors such as age and gender for CML was examined. The heterozygous CT genotype showed significant association with age and CML, being a risk factor for both age groups (below 40 and above 40), while the TT genotype was protective against CML progression.

However, the genotypes CC, CT, and TT did not show significant results in association with gender and CML progression as the p- values were less than 0.05 while some studies show that the incidence rate of CML increase with age and higher in males as compared to females (Rohrbacher and Hasford 2009). Validation of these results is crucial as there is currently no literature linking the rs765071211 variant with CML.

Another association was studied was the blood group and platelet count being the risk factor for the development of the disease. We analyzed that the pateints with blood group O having TT genotype are more prone towards CML as compared to counterpart genotypes. The validations of the results are further needed as till date there is no literature available that links this variant with CML. Therefore, large sample size studies are required to be conducted in order to study the role of rs765071211 variant ADGRE2 and its other variants with CML. Similarly pateints having TT genotype have higher concentration of platelet count in their blood and they are more susceptible to disease as compared to other genotypes.

Large scale studies are required to determine the role of this variation and other ADGRE2 variations in CML. The study on the association of rs765071211 (D67N) variant in ADGRE2 with CML suggests that this variant is a risk factor for the disease. This finding could potentially lead to the development of t for CML targeting this SNP in ADGRE2. However, further large-scale studies are required to fully understand the importance of this SNP and its role in the progression of CML and its diagnosis. SNP analysis may help in developing more affordable and effective diagnostic tool for CML by targeting specific genetic variants. This research could also pave the way for identifying new genetic markers for the prognosis and diagnosis of not only CML but also other types of cancer, by studying genes involved in various signaling pathways related to cancer development.

Conclusion

Previous studies on CML have not provided insights into disease progression, emphasizing the need to identify new pathogenic variants that could potentially serve as novel treatments. This study focused on the SNP rs765071211 (D67N) of the ADGRE2 gene. The impact of this SNP on the structure, function, and flexibility of the ADGRE2 gene was studied, considering the amino acid change at position 67 of the protein. The study also evaluated the pathogenicity of the mutant ADGRE2 gene. To validate the *in silico* analysis, the association of the damaging variant with CML was confirmed through genotype analysis. Large-scale genotype studies are necessary to establish this variant as a viable genetic marker for detection of CML.

Publication bibliography

Minciacchi, Valentina R.; Kumar, Rahul; Krause, Daniela S. (2021): Chronic Myeloid Leukemia: A Model Disease of the Past, Present and Future. In *Cells* 10 (1), p. 117. DOI: 10.3390/cells10010117.

Ochi, Yotaro (2023): Genetic landscape of chronic myeloid leukemia. In *Int J Hematol* 117 (1), pp. 30–36. DOI: 10.1007/s12185-022-03510-w.

Osman, Afaf E. G.; Deininger, Michael W. (2021): Chronic Myeloid Leukemia: Modern therapies, current challenges and future directions. In *Blood Reviews* 49, p. 100825. DOI: 10.1016/j.blre.2021.100825.

Wu, Jiaxin; Wang, Aoli; Li, Xixiang; Chen, Cheng; Qi, Ziping; Hu, Chen et al. (2019): Discovery and characterization of a novel highly potent and selective type II native and drug-resistant V299L mutant BCR-ABL inhibitor (CHMFL-ABL-039) for Chronic Myeloid Leukemia (CML). In *Cancer Biology & Therapy* 20 (6), pp. 877–885. DOI: 10.1080/15384047.2019.1579958.

Fausel, Christopher (2007): Targeted chronic myeloid leukemia therapy: seeking a cure. In Journal of managed care pharmacy : JMCP 13 (8 Suppl A), pp. 8–12. DOI: 10.18553/jmcp.2007.13.s8-a.8.

Hehlmann, Rüdiger (2020): Chronic Myeloid Leukemia in 2020. In HemaSphere 4 (5), e468. DOI: 10.1097/HS9.0000000000000468.

Jabbour, Elias; Kantarjian, Hagop (2020): Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. In American Journal of Hematology 95 (6), pp. 691–709. DOI: 10.1002/ajh.25792.

Westerweel, Peter E.; Te Boekhorst, Peter A. W.; Levin, Mark-David; Cornelissen, Jan J. (2019): New Approaches and Treatment Combinations for the Management of Chronic Myeloid Leukemia. In Frontiers in oncology 9, p. 665. DOI: 10.3389/fonc.2019.00665.

Boyden, Steven E.; Desai, Avanti; Cruse, Glenn; Young, Michael L.; Bolan, Hyejeong C.; Scott, Linda M. et al. (2016): Vibratory Urticaria Associated with a Missense Variant in ADGRE2. In The New England journal of medicine 374 (7), pp. 656–663. DOI: 10.1056/NEJMoa1500611.

Hamann, Jörg; Hsiao, Cheng-Chih; Lee, Chang Sup; Ravichandran, Kodi S.; Lin, Hsi-Hsien (2016): Adhesion GPCRs as Modulators of Immune Cell Function. In Handbook of experimental pharmacology 234, pp. 329–350. DOI: 10.1007/978-3-319-41523-9_15.

Huang, Chien-Hao; Jeng, Wen-Juei; Ho, Yu-Pin; Teng, Wei; Hsieh, Yi-Chung; Chen, Wei-Ting et al. (2016): Increased EMR2 expression on neutrophils correlates with disease severity and predicts overall mortality in cirrhotic patients. In Scientific reports 6, p. 38250. DOI: 10.1038/srep38250.

I, Kuan-Yu; Tseng, Wen-Yi; Wang, Wen-Chih; Gordon, Siamon; Ng, Kwai-Fong; Lin, Hsi-Hsien (2020): Stimulation of Vibratory Urticaria-Associated Adhesion-GPCR, EMR2/ADGRE2, Triggers the NLRP3 Inflammasome Activation Signal in Human Monocytes. In Frontiers in immunology 11, p. 602016. DOI: 10.3389/fimmu.2020.602016.

Naranjo, Andrea N.; Bandara, Geethani; Bai, Yun; Smelkinson, Margery G.; Tobío, Araceli; Komarow, Hirsh D. et al. (2020): Critical Signaling Events in the Mechanoactivation of Human Mast Cells through p.C492Y-ADGRE2. In The Journal of investigative dermatology 140 (11), 2210-2220.e5. DOI: 10.1016/j.jid.2020.03.936.

Nijmeijer, Saskia; Vischer, Henry F.; Leurs, Rob (2016): Adhesion GPCRs in immunology. In Biochemical pharmacology 114, pp. 88–102. DOI: 10.1016/j.bcp.2016.04.013.

Sassone-Corsi, Paolo (2012): The cyclic AMP pathway. In Cold Spring Harbor Perspectives in Biology 4 (12). DOI: 10.1101/cshperspect.a011148.

Sreepada, Abhijit; Tiwari, Mansi; Pal, Kasturi (2022): Adhesion G protein-coupled receptor gluing action guides tissue development and disease. In J Mol Med 100 (10), pp. 1355–1372. DOI: 10.1007/s00109-022-02240-0.

Vázquez-Victorio, Genaro; González-Espinosa, Claudia; Espinosa-Riquer, Zyanya P.; Macías-Silva, Marina (2016): GPCRs and actin-cytoskeleton dynamics. In Methods in cell biology 132, pp. 165–188. DOI: 10.1016/bs.mcb.2015.10.003.

Publication bibliography

Adzhubei, Ivan A.; Schmidt, Steffen; Peshkin, Leonid; Ramensky, Vasily E.; Gerasimova, Anna; Bork, Peer et al. (2010): A method and server for predicting damaging missense mutations. In *Nature methods* 7 (4), pp. 248–249. DOI: 10.1038/nmeth0410-248.

Almagro Armenteros, Jose Juan; Sønderby, Casper Kaae; Sønderby, Søren Kaae; Nielsen, Henrik; Winther, Ole (2017): DeepLoc: prediction of protein subcellular localization using deep learning. In *Bioinformatics* 33 (24), p. 4049. DOI: 10.1093/bioinformatics/btx548.

Bhudia, Nisha; Desai, Sapna; King, Natalie; Ancellin, Nicolas; Grillot, Didier; Barnes, Ashley A.; Dowell, Simon J. (2020): G Protein-Coupling of Adhesion GPCRs ADGRE2/EMR2 and ADGRE5/CD97, and Activation of G Protein Signalling by an Anti-EMR2 Antibody. In *Sci Rep* 10 (1), p. 1004. DOI: 10.1038/s41598-020-57989-6.

Blum, Matthias; Chang, Hsin-Yu; Chuguransky, Sara; Grego, Tiago; Kandasaamy, Swaathi; Mitchell, Alex et al. (2021): The InterPro protein families and domains database: 20 years on. In *Nucleic acids research* 49 (D1), D344-D354. DOI: 10.1093/nar/gkaa977.

Boyden, Steven E.; Desai, Avanti; Cruse, Glenn; Young, Michael L.; Bolan, Hyejeong C.; Scott, Linda M. et al. (2016): Vibratory Urticaria Associated with a Missense Variant in ADGRE2. In *The New England journal of medicine* 374 (7), pp. 656–663. DOI: 10.1056/NEJMoa1500611.

Capriotti, Emidio; Fariselli, Piero; Casadio, Rita (2005): I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. In *Nucleic acids research* 33 (Web Server issue), W306-10. DOI: 10.1093/nar/gki375.

Chen, Yuting; Zhang, Haoling; Wang, Wen; Shen, Yue; Ping, Zhi (2024): Rapid generation of high-quality structure figures for publication with PyMOL-PUB. In *Bioinformatics* 40 (3). DOI: 10.1093/bioinformatics/btae139.

Fausel, Christopher (2007): Targeted chronic myeloid leukemia therapy: seeking a cure. In *Journal of managed care pharmacy : JMCP* 13 (8 Suppl A), pp. 8–12. DOI: 10.18553/jmcp.2007.13.s8-a.8.

Hamann, Jörg; Hsiao, Cheng-Chih; Lee, Chang Sup; Ravichandran, Kodi S.; Lin, Hsi-Hsien (2016a): Adhesion GPCRs as Modulators of Immune Cell Function. In *Handbook of experimental pharmacology* 234, pp. 329–350. DOI: 10.1007/978-3-319-41523-9_15.

Hamann, Jörg; Hsiao, Cheng-Chih; Lee, Chang Sup; Ravichandran, Kodi S.; Lin, Hsi-Hsien (2016b): Adhesion GPCRs as Modulators of Immune Cell Function. In *Handbook of experimental pharmacology* 234, pp. 329–350. DOI: 10.1007/978-3-319-41523-9_15.

Hehlmann, Rüdiger (2020): Chronic Myeloid Leukemia in 2020. In *HemaSphere* 4 (5), e468. DOI: 10.1097/HS9.0000000000000468.

Howe, Kevin L.; Achuthan, Premanand; Allen, James; Allen, Jamie; Alvarez-Jarreta, Jorge; Amode, M. Ridwan et al. (2021): Ensembl 2021. In *Nucleic acids research* 49 (D1), D884-D891. DOI: 10.1093/nar/gkaa942.

Huang, Chien-Hao; Jeng, Wen-Juei; Ho, Yu-Pin; Teng, Wei; Hsieh, Yi-Chung; Chen, Wei-Ting et al. (2016): Increased EMR2 expression on neutrophils correlates with disease severity and predicts overall mortality in cirrhotic patients. In *Scientific reports* 6, p. 38250. DOI: 10.1038/srep38250.

I, Kuan-Yu; Tseng, Wen-Yi; Wang, Wen-Chih; Gordon, Siamon; Ng, Kwai-Fong; Lin, Hsi-Hsien (2020): Stimulation of Vibratory Urticaria-Associated Adhesion-GPCR, EMR2/ADGRE2, Triggers the NLRP3 Inflammasome Activation Signal in Human Monocytes. In *Frontiers in immunology* 11, p. 602016. DOI: 10.3389/fimmu.2020.602016.

Ioannidis, Nilah M.; Rothstein, Joseph H.; Pejaver, Vikas; Middha, Sumit; McDonnell, Shannon K.; Baheti, Saurabh et al. (2016): REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. In *American journal of human genetics* 99 (4), pp. 877–885. DOI: 10.1016/j.ajhg.2016.08.016.

Jabbour, Elias; Kantarjian, Hagop (2020): Chronic myeloid leukemia: 2020 update on diagnosis, therapy and monitoring. In *American Journal of Hematology* 95 (6), pp. 691–709. DOI: 10.1002/ajh.25792.

Jumper, John; Evans, Richard; Pritzel, Alexander; Green, Tim; Figurnov, Michael; Ronneberger, Olaf et al. (2021): Highly accurate protein structure prediction with AlphaFold. In *Nature* 596 (7873), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

Karczewski, Konrad J.; Francioli, Laurent C.; Tiao, Grace; Cummings, Beryl B.; Alföldi, Jessica; Wang, Qingbo et al. (2020): The mutational constraint spectrum quantified from variation in 141,456 humans. In *Nature* 581 (7809), pp. 434–443. DOI: 10.1038/s41586-020-2308-7.

Kent, W. James; Sugnet, Charles W.; Furey, Terrence S.; Roskin, Krishna M.; Pringle, Tom H.; Zahler, Alan M.; Haussler, David (2002): The human genome browser at UCSC. In *Genome research* 12 (6), pp. 996–1006. DOI: 10.1101/gr.229102.

Khan, Naila; Khan, Khushbukhat; Badshah, Yasmin; Trembley, Janeen H.; Ashraf, Naeem Mahmood; Shabbir, Maria et al. (2023): Investigating pathogenic SNP of PKCι in HCV-induced hepatocellular carcinoma. In *Sci Rep* 13 (1), p. 12504. DOI: 10.1038/s41598-023-39804-0.

Lorenz, Ronny; Luntzer, Dominik; Hofacker, Ivo L.; Stadler, Peter F.; Wolfinger, Michael T. (2016): SHAPE directed RNA folding. In *Bioinformatics* 32 (1), pp. 145–147. DOI: 10.1093/bioinformatics/btv523.

Minciacchi, Valentina R.; Kumar, Rahul; Krause, Daniela S. (2021): Chronic Myeloid Leukemia: A Model Disease of the Past, Present and Future. In *Cells* 10 (1), p. 117. DOI: 10.3390/cells10010117.

Naranjo, Andrea N.; Bandara, Geethani; Bai, Yun; Smelkinson, Margery G.; Tobío, Araceli; Komarow, Hirsh D. et al. (2020): Critical Signaling Events in the Mechanoactivation of Human Mast Cells through p.C492Y-ADGRE2. In *The Journal of investigative dermatology* 140 (11), 2210-2220.e5. DOI: 10.1016/j.jid.2020.03.936.

Nijmeijer, Saskia; Vischer, Henry F.; Leurs, Rob (2016a): Adhesion GPCRs in immunology. In *Biochemical pharmacology* 114, pp. 88–102. DOI: 10.1016/j.bcp.2016.04.013.

Nijmeijer, Saskia; Vischer, Henry F.; Leurs, Rob (2016b): Adhesion GPCRs in immunology. In *Biochemical pharmacology* 114, pp. 88–102. DOI: 10.1016/j.bcp.2016.04.013.

Ochi, Yotaro (2023): Genetic landscape of chronic myeloid leukemia. In *Int J Hematol* 117 (1), pp. 30–36. DOI: 10.1007/s12185-022-03510-w.

Pejaver, Vikas; Mooney, Sean D.; Radivojac, Predrag (2017): Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. In *Human mutation* 38 (9), pp. 1092–1108. DOI: 10.1002/humu.23258.

Rentzsch, Philipp; Schubach, Max; Shendure, Jay; Kircher, Martin (2021): CADD-Splice-improving genome-wide variant effect prediction using deep learning-derived splice scores. In *Genome medicine* 13 (1), p. 31. DOI: 10.1186/s13073-021-00835-9.

Reva, Boris; Antipin, Yevgeniy; Sander, Chris (2011): Predicting the functional impact of protein mutations: application to cancer genomics. In *Nucleic acids research* 39 (17), e118. DOI: 10.1093/nar/gkr407.

Rodrigues, Carlos Hm; Pires, Douglas Ev; Ascher, David B. (2018): DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. In *Nucleic acids research* 46 (W1), W350-W355. DOI: 10.1093/nar/gky300.

Rohrbacher, Maren; Hasford, Joerg (2009): Epidemiology of chronic myeloid leukaemia (CML). In *Best practice & research. Clinical haematology* 22 (3), pp. 295–302. DOI: 10.1016/j.beha.2009.07.007.

Sassone-Corsi, Paolo (2012): The cyclic AMP pathway. In *Cold Spring Harbor Perspectives in Biology* 4 (12). DOI: 10.1101/cshperspect.a011148.

Sawyers, C. L. (1999): Chronic myeloid leukemia. In *The New England journal of medicine* 340 (17), pp. 1330–1340. DOI: 10.1056/NEJM199904293401706.

Sim, Ngak-Leng; Kumar, Prateek; Hu, Jing; Henikoff, Steven; Schneider, Georg; Ng, Pauline C. (2012): SIFT web server: predicting effects of amino acid substitutions on proteins. In *Nucleic acids research* 40 (Web Server issue), W452-7. DOI: 10.1093/nar/gks539.

Sreepada, Abhijit; Tiwari, Mansi; Pal, Kasturi (2022): Adhesion G protein-coupled receptor gluing action guides tissue development and disease. In *J Mol Med* 100 (10), pp. 1355–1372. DOI: 10.1007/s00109-022-02240-0.

Tate, John G.; Bamford, Sally; Jubb, Harry C.; Sondka, Zbyslaw; Beare, David M.; Bindal, Nidhi et al. (2019): COSMIC: the Catalogue Of Somatic Mutations In Cancer. In *Nucleic acids research* 47 (D1), D941-D947. DOI: 10.1093/nar/gky1015.

Vázquez-Victorio, Genaro; González-Espinosa, Claudia; Espinosa-Riquer, Zyanya P.; Macías-Silva, Marina (2016): GPCRs and actin-cytoskeleton dynamics. In *Methods in cell biology* 132, pp. 165–188. DOI: 10.1016/bs.mcb.2015.10.003.

Venselaar, Hanka; Te Beek, Tim A. H.; Kuipers, Remko K. P.; Hekkelman, Maarten L.; Vriend, Gert (2010): Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. In *BMC bioinformatics* 11, p. 548. DOI: 10.1186/1471-2105-11-548.

Westerweel, Peter E.; Te Boekhorst, Peter A. W.; Levin, Mark-David; Cornelissen, Jan J. (2019): New Approaches and Treatment Combinations for the Management of Chronic Myeloid Leukemia. In *Frontiers in oncology* 9, p. 665. DOI: 10.3389/fonc.2019.00665.

Wu, Kaimin; Kong, Feizhi; Zhang, Jingjing; Tang, Ying; Chen, Yao; Chao, Long et al. (2023): Recent Progress in Single-Nucleotide Polymorphism Biosensors. In *Biosensors* 13 (9), p. 864. DOI: 10.3390/bios13090864.

Yang, Jianyi; Yan, Renxiang; Roy, Ambrish; Xu, Dong; Poisson, Jonathan; Zhang, Yang (2015): The I-TASSER Suite: protein structure and function prediction. In *Nature methods* 12 (1), pp. 7–8. DOI: 10.1038/nmeth.3213.