# Oncogenic Role of KLF8 Variants in Breast Cancer

By

Syeda Amna Ali

(Registration No: 00000399715)

Department of Healthcare Biotechnology

Atta Ur Rahman School of Applied Biosciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

# Oncogenic Role of KLF8 Variants in Breast Cancer


By

Syeda Amna Ali

(Registration No: 00000399715)

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in

Healthcare Biotechnology

Supervisor: Dr. Maria Shabbir

Co Supervisor: Dr. Yasmin Badshah

Atta Ur Rahman School of Applied Biosciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

## THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by **Ms. Syeda Amna Ali, Registration No. 00000399715**, of **Atta-ur-Rahman School of Applied Biosciences** (ASAB), has been vetted by undersigned, found complete in all respects as per NUST Statutes/ Regulations/ Masters Policy, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of master's degree. It is further certified that necessary amendments as pointed out by GEC members and evaluators of the scholar have also been incorporated in the said thesis.

Dr. Maria Shabbir
Tenured Associate Professor
Dept of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

Signature: _____

Name of Supervisor: Dr. Maria Shabbir

Date: 15.08.24

Dr. Sobia Manzoor, PhD
Associate Professor (Tenured)
Dept. of Healthcare Biotechnology, ASAB
National University of Sciences and
Technology (NUST), Islamabad

Signature (HOD): _____

Date: 15.8.2024

Signature (Dean/ Principal): _____

A/Principal & Dean
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST, Islamabad

Date: 15.08.2024

FORM TH-4

# National University of Sciences & Technology

## MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Regn No.) Syeda Amna Ali **Titled**: Oncogenic Role of KLF8 Variants in Breast Cancer be accepted in partial fulfillment of the requirements for the award of degree with ( A grade).

### Examination Committee Members

1. Name: Dr. Saira Justin Signature: Dr. Saira Justin, Assistant Professor, Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

2. Name: Dr. Tougeer Ahmed Signature:

3. Name: Dr. Hashaam Akhter Signature:

Supervisor's name: Dr. Maria Shabbir Signature: Maria Shabbir, Associate Professor, NUST Islamabad

Date: 22/07/2024

Date: 22/07/2024

Head of Department

**COUNTERSINGED**

Date: 08/08/2024

Dean/Principal

A/Principal & Dean
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

# AUTHOR'S DECLARATION

I, Syeda Amna Ali, hereby state that my MS thesis titled " Oncogenic Role of KLF8 Variants in Breast Cancer" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Name of Student: Syeda Amna Ali

Date: _____

# DEDICATION

This thesis is dedicated to my beloved family and my teachers who supported me throughout this process and believed in my capabilities.

# ACKNOWLEDGEMENTS

In the name of Allah, the Most Beneficent and the Most Merciful. All praises be to Allah who made me capable of doing this project.

I would like to acknowledge and give my warmest thanks to my supervisor, Dr. Maria Shabbir and co-supervisor, Dr. Yasmin Badshah who made this project possible. Your guidance carried me through every stage of this project. I would also express gratitude towards my guidance and examination committee members who provided me with comments and suggestions for my work.

I would like to give special thanks to my husband, Mr. Shan ur Rehman and my whole family for their believing in my capabilities, encouraging me and providing their support when undertaking my research project. Your prayers for me sustained me this far.

Further, I thank my seniors and lab fellows for making this project extremely enjoyable.

Thank you,

Syeda Amna Ali

MS HCB,

ASAB, NUST

TABLE OF CONTENTS

# LIST OF TABLES

**Page No.**

# LIST OF FIGURES

**Page No.**

# ABSTRACT

Breast cancer is the most prevalent cancer among females. In Pakistan, late diagnosis of this disease leads to a higher death rate. This type of cancer involves uncontrolled division of epithelial cells surrounding ducts and lobules of breast. This uncontrolled division is mediated by transcription factors. One of the family of transcription factors is the Kruppel-like factors family. These are the zinc-finger proteins that function in cell cycle progression. However, these are also involved in various malignancies where they directly affect cell cycle regulation, apoptosis, metastasis and signaling pathways. Therefore, there is a need to study these proteins. One of the members of this family of proteins is KLF8. This factor is found to be associated with different types of cancers and other disorders. There is not enough evidence of single nucleotide polymorphism in KLF8. This study is aimed at studying the association between SNPs of KLF8 and breast cancer. The SNPs in KLF8 affect the structure and function of the KLF8 protein. In-silico analysis of the SNP rs868781835 revealed that its affect might probably be damaging, while SNP rs1196572444 was found to be damaging. This SNP was rendered to genotype analysis where the genotype GT was associated with metastasis of breast cancer.



**Keywords:** Breast cancer, KLF8, SNPs, zinc-finger proteins, In-silico

# CHAPTER 1: INTRODUCTION

## 1.1 Cancer Overview

Cancer is a disease in which certain cells of the body undergo uncontrolled divisions leading to altered signaling and metabolism. The abnormal mass of cells formed in this way is termed as 'tumor'. This mass establishes its own microenvironment. It is established through research that genetic mutations cause normal cells to transform. A lot of data is available to chalk out its root cause but the quest to find a cure continues (Upadhyay, 2021).

The functional capabilities of cells to grow rapidly are referred to as the hallmarks of cancer. To date, there are eight hallmarks of cancer. These include maintaining signaling pathways for proliferation, deceiving growth suppressors, opposition to apoptosis, unlimited replicative potential, establishment of a vasculature, ability to invade the nearby tissue and metastasis, metabolic reprogramming, and immune evasion (Hanahan et al., 2011).

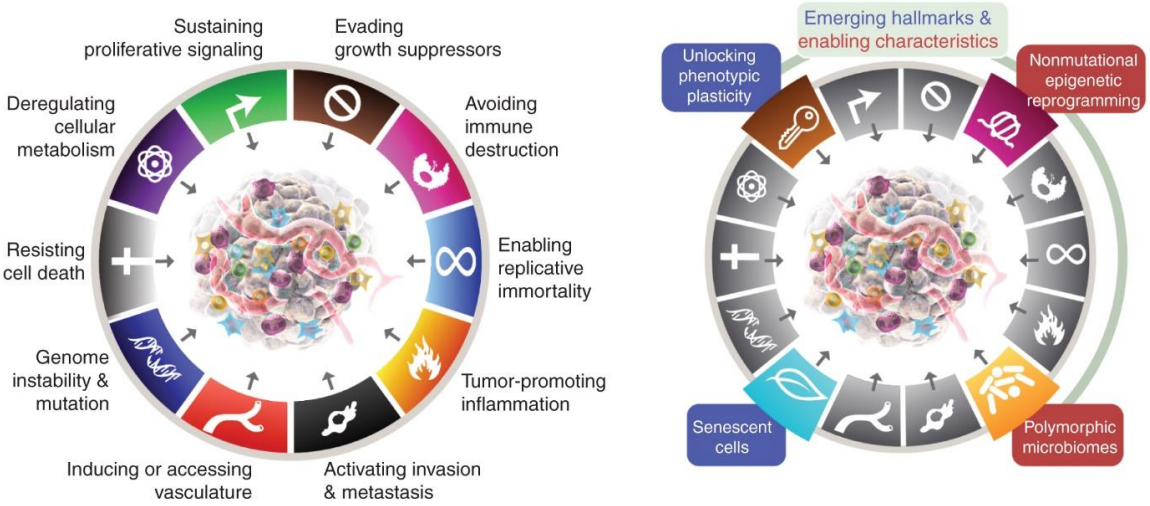



**Figure 1.1** Hallmarks of cancer (Hanahan et al., 2011)

### *1.1.1 Global Cancer Prevalence*

According to the International Agency for Research on Cancer, the disease has been diagnosed in 19.3 million people, and 10 million fell prey to it excluding non-melanoma skin cancer as of the year 2020. Breast cancer held its new place as the most commonly diagnosed cancer type replacing lung cancer and its associated death rate remained high (Sung et al., 2021).

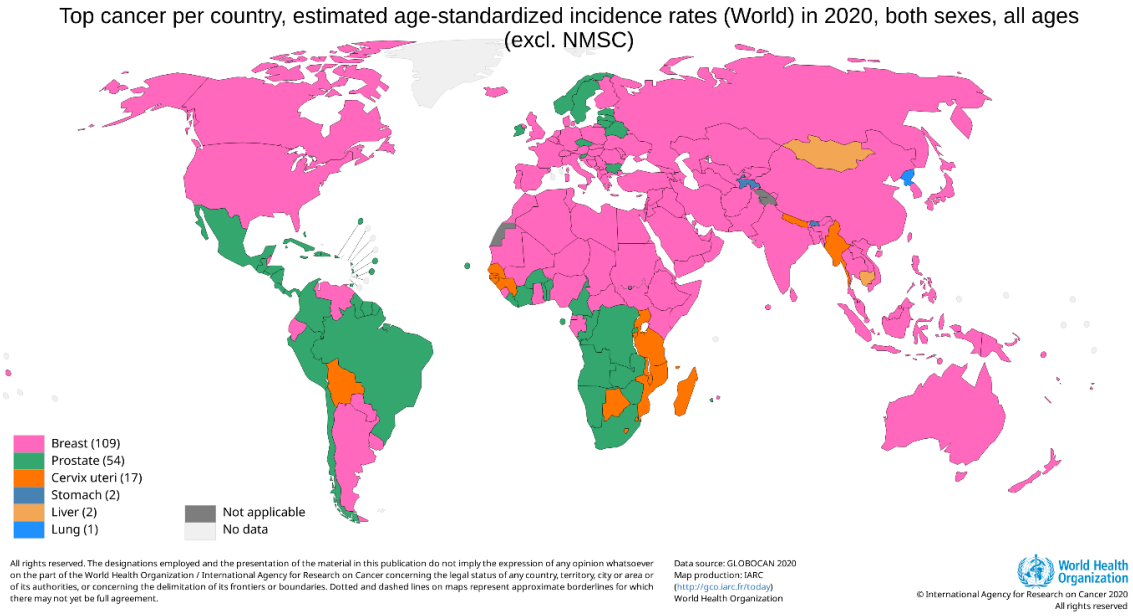



**Figure 1.2 Global Cancer Prevalence (WHO)**

It is evident from the map that breast cancer is the most prevalent one. Our focus will be on this type of cancer.

## 1.2 Breast Cancer

Breast cancer is a multifactorial disease where the epithelial cells lining the ducts and lobules of the breast undergo changes regulated by genes and hormones. Its appearance in an individual involves risk factors that may be modifiable or non-modifiable. The modifiable factors include therapies, alcohol intake, being overweight, exposure to certain chemical and cancer-causing agents (carcinogens), excessive exposure to radiation, lack of physical activity, and insufficient intake of vitamins. The non-modifiable factors include sex, age, ethnicity, family history, previous therapies for cancer that involve radiation, the

hormonal status of the body, density of breast tissue, and genetic mutations (Łukasiewicz et al., 2021a). Of all the factors, we will focus on the genetic basis of breast cancer.

Molecular and histological evidence shows that breast cancer is of three types, one in which hormone receptors (estrogen and progesterone) are expressed, the second that expresses human epidermal receptor (HER2), and the triple-negative breast cancer (TNBC) (Barzaman et al., 2020).

### *1.2.1 Types of Breast Cancer*

According to Hormone Receptors (estrogen and Progesterone)

The hormone receptor-positive breast cancer expresses estrogen and progesterone receptors in approximately 70% of all breast cancers (Brufsky et al., 2018). Estrogen is an important hormone involved in reproduction, development, and growth and is steroidal in nature. They exert their effects through binding with their receptors which belong to the nuclear receptor transcription factor family. These receptors regulate the transition from G0 to the S phase in the breast epithelial cell cycle and so, defects of these receptors lead to breast cancer formation (Zhang et al., 2014). Progesterone is also a steroid hormone with numerous functions in the human body. It specifically functions in the reproductive system in maintaining pregnancy. It is produced by the adrenal cortex (Cable et al., 2023). Progesterone activity is associated with the activation of its receptors. It is a nuclear transcription factor. Its role has been known in estrogen receptor signaling in ER-positive breast cancer (Boland et al., 2020).

HER2

HER2 is a membrane receptor that effects cell proliferation and survival. Its amplification leads to tumor development and invasion in breast cancer. It is amplified in 15-20% of breast carcinomas (Krishnamurti et al., 2014).

Triple-negative Breast Cancer (TNBC)

This type of breast cancer signifies a group of tumors where there is an absence of estrogen receptor, progesterone receptor, and over-expression of the human epidermal growth factor receptor 2 gene. 10-20% of human breast cancers are of this type (P. Kumar & Aggarwal, 2016).

## 1.3 Single Nucleotide Polymorphism (SNP)

A SNP is a position in DNA where there is a change among individuals of a species (Nicholson et al., 2002). This polymorphism is most common, which occurs after every 1000 base pairs in the human genome (Karki et al., 2015). These are responsible for producing variations in the population (Cotton et al., 2018). These SNPs might be linked to phenotypes or disease. For this, genome-wide association studies (GWAS) are carried out. SNPs can alter gene function through affecting mRNA splicing, nucleocytoplasmic export, stability, and translation, thus changing the protein's activity (Robert et al., 2018). SNPs can also lead to disease susceptibility. This is indicated by the fact that a person possessing a certain SNP is more likely to have a disease when compared to a healthy person without that particular SNP. Some examples include, Alzheimer's disease, sickle-cell anemia, schizophrenia and various malignancies (Bessenyei et al., 2004).

### *1.3.1 SNPs in Cancer*

As described earlier, SNPs can cause disease susceptibility. Likewise, these can cause cancer susceptibility if it falls in gene that regulates DNA mismatch repair, cell cycle regulation, metabolic activity, and immunity. These also function as biomarkers in various malignancies as diagnostic and therapeutic agents. If these lie in the promoter region, they modify the promoter activity, binding of transcription factor, DNA methylation and histone modification. If the SNP lies in exonal region, it promotes cancer susceptibility through suppressing gene transcription and translation. In intronic regions, these produce splice variants and either promote or prevent binding of long non-coding RNAs (lncRNAs), while in 5'UTR they affect translation and in 3'UTR, they impact binding of miRNAs (Deng et al., 2017).

### *1.3.2 SNPs in Breast Cancer*

The association between breast cancer and SNPs has been identified through genome-wide association studies. However, not much is known about the disease prognosis (He et al., 2019).

Here, we will study, the impact of single nucleotide polymorphism of a gene belonging to the family of Krupple-like factors (KLFs). We have chosen to take the factor, KLF8. There is not enough evidence of a link between breast cancer and KLF8 polymorphism. It is yet to be discovered whether there is any genetic link between KLF8 SNPs and an increased risk of developing breast cancer. As a result, the primary goal of this research is to identify the possibly harmful SNPs found in KLF8 and link them with breast cancer. This will help us to explore more about KLF8, which will most likely lead to the development of a new possible therapeutic target as well as a prognostic marker for early detection of breast cancer.

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Overview

The oldest depiction of cancer was of bone cancer from ancient Egypt. The term ‘cancer’ originated from a Greek word ‘karkinos’ coined by Hippocrates which depicted carcinoma tumors. It has been established from historical records that there was no treatment procedure for the disease. However, a surgical intervention was performed to clear away the surface tumors as is done in the present day (Sudhakar, 2009).

The process of gaining the ability to grow abnormally is termed ‘neoplasia’. There are two courses of neoplasms, either they are benign, meaning they reside in one place, or they become malignant which refers to invasion of the surrounding tissue and travel to distant organs (W.L., Kemp).

### *2.1.1 Types of Cancer*

There are various types of cancer, based on the organ where it originated. The disease can be classified into five broad classes. These include **carcinoma** (the cancer of body tissue), **sarcoma** (the cancer of connective tissues), **lymphoma** (the cancer of nodes or glands of the lymphatic system), **leukemia** (the cancer of bone marrow), and **myeloma** (the cancer of plasma cells of bone marrow) (Stanford Medicine).

## 2.2 Breast Cancer

According to the World Health Organization, breast cancer is the most prevalent type of carcinoma in females of all ages followed by colorectal cancer and other carcinomas (WHO).

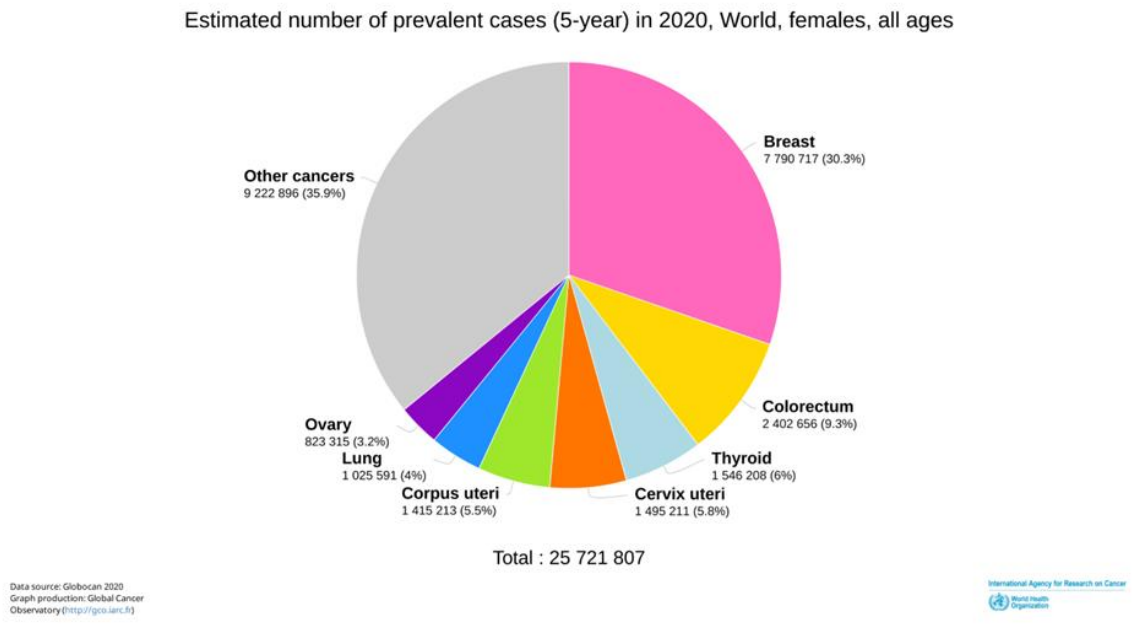



**Figure 2.1** Global Breast Cancer Prevalence (WHO)

Breast cancer is also has a high prevalence in Pakistan to the extent that 1 in every 9 females suffer from the disease at any age in their lives. In Asia, Pakistan holds the first place in breast cancer prevalence where there might be a lot of reasons including poor hygienic conditions and lack of awareness (Menhas, 2020.)

The International Agency of Research on Cancer reports that in the year 2018, the disease had been diagnosed in 34,066 females. Due to lack of proper healthcare facilities and issues related to delayed diagnosis has resulted in a high mortality rate (Khan et al., 2021).

According to the World Health Organization Globocan 2020 data, breast cancer in females remains on top of all new cancer incidences (WHO Globocan 2020).

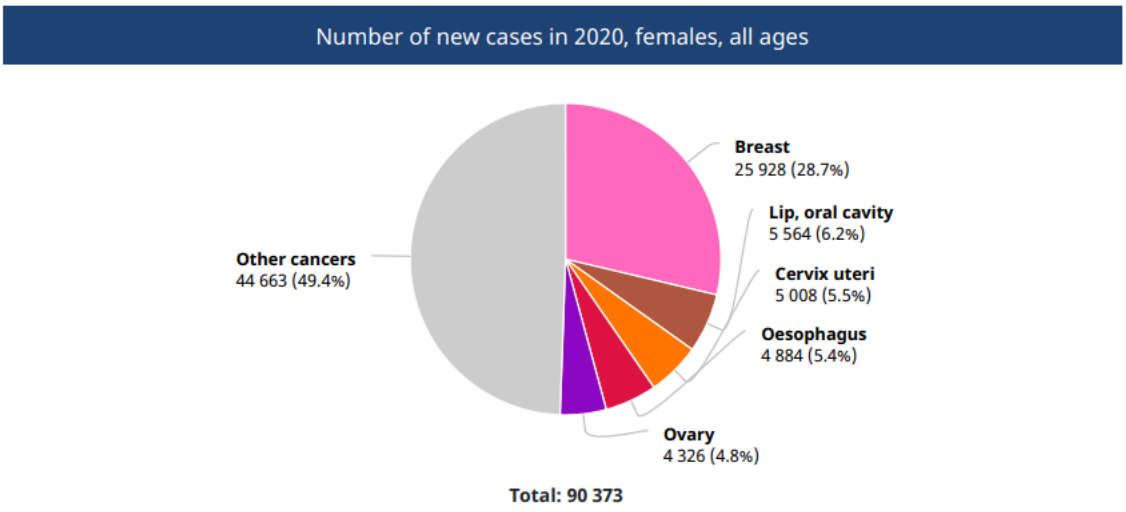



**Figure 2.2** Cancer Incidences in Pakistan (WHO Globocan 2020)

### 2.3 The KLF Gene Family

Gene expression is an important process through which the functioning of a gene is determined. The process primarily involves producing a product from a gene. The product is normally a protein. Gene expression involves the processes of transcription of DNA and translation of mRNA to produce proteins. The process of transcription happens within a nucleus of a cell in which an enzyme uses a coded section of a gene to produce mRNA which is further used by translation process outside of nucleus. During the transcription process the RNA polymerase with the help of proteins called transcription factors binds to a specific section within a gene to produce mRNA. The understanding of gene expression allows us to determine the function of a gene and its purpose through which we can develop a better understanding of the biological processes (Nature Education). Transcription factors that are used to generate mRNA are very important as they act as regulators or controlling factors in the gene expression process. Transcription factors also act as initiators in the gene expression process. The regulation or control of the gene expression process is an important as well as a complex part that ensures the creation of correct proteins (Wakim et al., 2012).

There are different types of transcriptional factors. Krüppel-like factors (KLFs) are transcription factors that bind to DNA and modulate cellular mechanisms and metabolism.

These are zinc finger protein structure-like transcriptional factors that act as regulators in the gene expression process. KLF can enhance as well as repress gene expression. There are 18 types of KLFs which are categorized into 3 groups. This classification is based on their role in regulating the process of transcription and their structural characteristics. Group 1 comprises KLF3, KLF8, and KLF12, Group 2 comprise KLF1, KLF2, KLF4, KLF5, KLF6, and KLF7, and Group 3 comprises KLF9, KLF10, KLF11, KLF13, KLF14, and KLF16. KLF15 and KLF17 have not been included in these groups due to poor knowledge about their association with other factors. Each KLF group has some functional characteristics which are categorized into specific groups. KLFs groups 1 and 3 primarily acts as transcriptional repressors whereas KLFs group 2 act as transcriptional activators (Pollak et al., 2018).

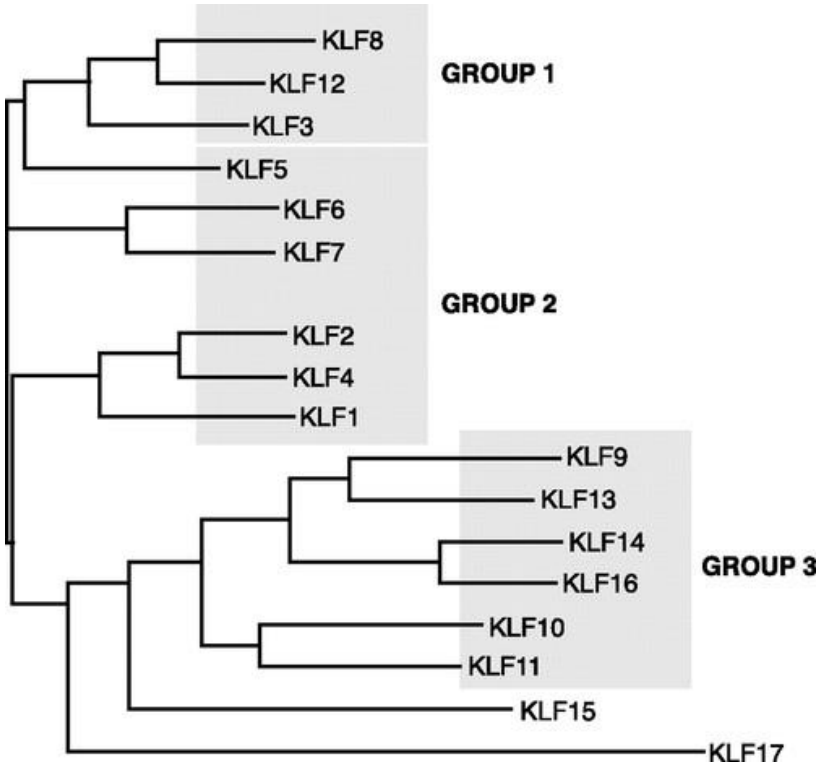



**Figure 2.1:** Phylogenetic Tree of KLF Family (McConnell et al, 2010)

The KLF family members work with other transcriptional regulators and play a vast role in metabolism in approximately every organ of the human body. They work to exert

physiologic effects where required. These are involved in processes such as intake of nutrients, catabolism of amino acids, lipid partitioning, glucose content fluctuation, and energy production through oxidation (Hsieh et al., 2019). These are also involved in important cellular functions including cell proliferation, differentiation, and apoptosis (Bialkowska et al., 2017).

The Kruppel-like factors carry three highly conserved C2H2 zinc finger domains at carboxy-terminals. These facilitate the activation or halting of transcription by interacting with GC-rich consensus sequences including 5'-CACCC-3' sequences of the DNA (Lomberk & Urrutia, 2005). On the other hand, the amino terminus of these factors controls the specific protein-protein and protein-DNA interactions, contrary to the carboxy terminus (Suzuki et al., 2005)

KLFs also share homology with another transcription factor called SP1. This was one of the initially identified and characterized transcription factor which binds to the GC-rich regions of the DNA through three C2H2-type zinc fingers. Due to the structural similarity of KLFs with SP1, these KLFs are categorized among the SP1/KLF family (McConnell et al., 2010).

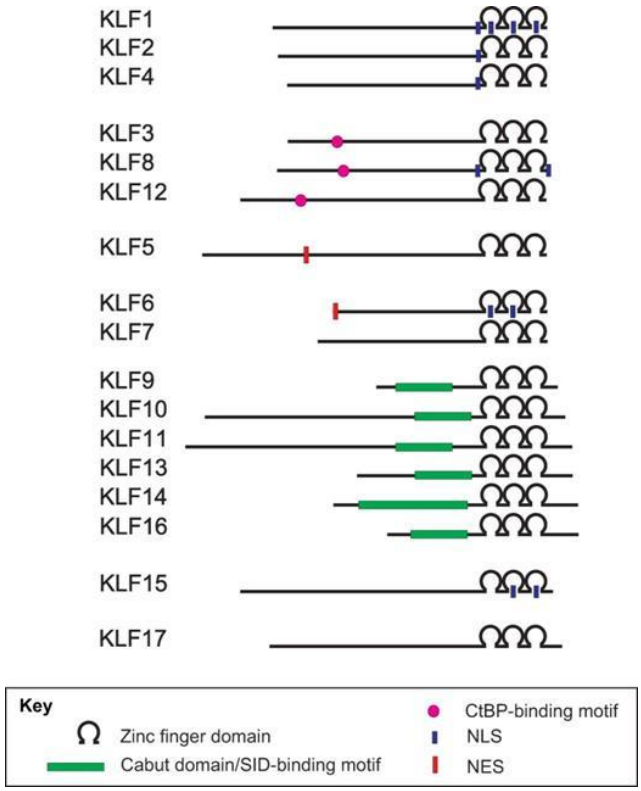



**Figure 2.2:** Domain structure of KLFs in humans (Bialkowska et al., 2017)

The KLF family members work with other transcriptional regulators and play a vast role in metabolism in approximately every organ of the human body. They work to exert physiologic effects where required. These are involved in processes such as intake of nutrients, catabolism of amino acids, lipid partitioning, glucose content fluctuation, and energy production through oxidation (Hsieh et al., 2019). These are also involved in important cellular functions including cell proliferation, differentiation, and apoptosis (Bialkowska et al., 2017).

The Kruppel-like factors carry three highly conserved C2H2 zinc finger domains at carboxy-terminals. These facilitate the activation or halting of transcription by interacting with GC-rich consensus sequences including 5'-CACCC-3' sequences of the DNA (Lomberk & Urrutia, 2005). On the other hand, the amino terminus of these factors controls the specific protein-protein and protein-DNA interactions, contrary to the carboxy terminus (Suzuki et al., 2005)

KLFs also share homology with another transcription factor called SP1. This was one of the initially identified and characterized transcription factor which binds to the GC-rich regions of the DNA through three C2H2-type zinc fingers. Due to the structural similarity of KLFs with SP1, these KLFs are categorized among the SP1/KLF family (McConnell et al., 2010).

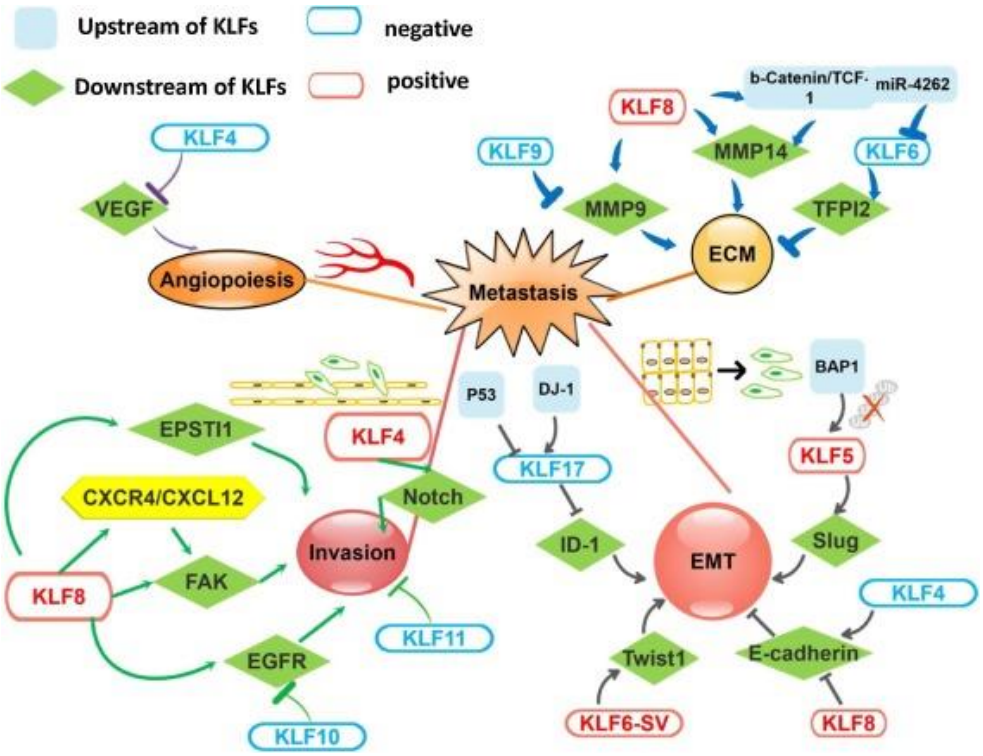



**Figure 2.3:** The Role of KLFs in breast cancer metastasis (Zhang et al., 2020)

## 2.4 KLF8

### *2.4.1 Structure*

KLF8 belongs to the family of KLF (Kruppel-like Factors) which act as transcriptive factors in the regulation of gene expression process. The main purpose of KLF is to regulate or control gene expression by binding to DNA sequences by interacting with promoter regions of genes. KLF8 belongs to group 1 of the KLF family which contains KLF3 and

KLF12 factors too (Satadru et al., 2012.)). In humans, the KLF gene family member, the KLF8 is located on the X-chromosome. It comprises of 359 amino acid residues and was isolated from K562 leukemia cell line for the first time. A transcription factor contains three main domains, a DNA binding domain, which in the case of KLF8 is highly conserved, a transcription regulatory domain, and a nuclear localization signal (Vliet, Turner, & Crossley, 2000).

The DNA-binding Domain of KLF8

This factor contains a DNA binding part which allows it to detect and bind to promoter region of a gene. Its structure allows it to bind to CACCC, a specific pattern in promoter region of a gene (Funnell et al., 2012).

The Transcription Regulatory Domain

Transcription activation part allows KLF8 to regulate or control the transcription of genes. These parts allow transcriptive factors to interact with other factors or proteins to enhance the process of gene transcription. It has been observed that KLF8 acts as a transcriptive activator as well as repressor in certain scenarios (Aiqin Sun & Jie Hao, 2019.).

The Nuclear Localization Signal

Nuclear Localization Signal (NLS) is the process of transporting certain proteins to the nucleus of a cell. This process is very important as regulation of proteins in and out of the nucleus is crucial for gene expression (Cokol, Nair, & Rost, 2000). Since KLF8 is a transcriptive factor and it needs to be transported to the nucleus of a cell in order to regulate gene expression, it contains a nuclear localization signal that helps it to enter into the nucleus. nuclear localization signals of proteins like KLF8 are detected by importins (transport proteins) when formed in the cytoplasm and transported to the nucleus through nucleus pores (Mehta et al., 2009).

### *2.4.2 Functions of KLF8*

KLF8 plays various roles based on its interaction with different genes. KLF8 plays a critical role in different cell processes for example cell proliferation, cell differentiation, and cell

transformation. Cell proliferation is the process by which cells are divided and their numbers are increased. The cell proliferation process plays an important role in tissue development and maintenance in the human body. KLF8 plays its role in cell differentiation in which cells are assigned specialized roles or functions and switching from one cell type to another in order to perform various tasks. Many transcriptive factors play an important role in the cell differentiation process helping a single cell to develop in various tissues and organs. Along with cell proliferation and differentiation, KLF8 also plays its role in the cell transformation process which involves the change in characteristics of a cell (Phenotype) as a result of inserting a new genetic material in a cell (M.-D. Wang et al., 2020).

A dysregulation of KLF8 leads to diseases. For instance, KLF8 has been observed to be involved in Alzheimer's Disease (AD) which is related to the brain affecting brain memory and brain cell structure. KLF8 is expressed in the brain as well and plays its role in neural processes happening in the brain. Several investigations are underway in order to understand the involvement of KLF8 in neural disorders inside the brain (Yi et al., 2014)

### *2.4.3 Oncogenic Transformation and KLF8*

It is the process by which healthy cells are transformed or undergo changes which lead to abnormal or unregulated cell proliferation which in turn lead to cancer cells development. This process involves oncogenes which when activated amplifies the cell proliferation process. KLF8 plays its role in oncogenic transformation as it is an important transcriptive factor involving cell proliferation as described earlier. This leads to tumor progression in the human body (Wang et al., 2007). KLF8's role in the EMT process in tumor progression is the main reason for its involvement in different types of human cancers. Studies have shown KLF8 involvement in different types of cancers in the human body such as breast, gastric, liver etc. KLF8 involvement has been observed in growth of human lung cancer as well. In recent experiments, KLF8 overexpression was involved in growth and proliferation of lung cancer cells however on the contrary the knockdown of KLF8 decreased the lung cancer cell growth (Ma et al., 2019).

## **2.5 KLF8 in Breast Cancer**

KLF8 is a transcription factor which is overexpressed in multiple cancer types including the breast cancer. It plays a significant part in epithelial to mesenchymal transition, proliferation, and DNA damage repair (Tianshu Li et al., 2015a). It has also shown to play a role in breast cancer metastasis (Mukherjee et al., 2016). Another role of KLF8 has been identified to regulate other genes and microRNAs in promoting EMT and stem-cell-like traits to promote invasion (Xianhui et al., 2013)

In EMT

Epithelial to Mesenchymal Transition (EMT) is a process which allows cells to change their characteristics from epithelial (tightly connected cells lying on the surface of organs) to mesenchymal (mobile cells). This process allows the transition of cells from a structured state to a more mobile state. EMT is involved in the beneficial process such as healing wounds, however, it can also play a role in spreading cancer tumor cells on the contrary side (Kalluri et al., 2009).

It has been observed that KLF8 plays its role as an inducer in the EMT process. KLF8 unregulated expression could be a potential reason for spreading cancer cells in the human body. KLF8 facilitates cell migration and invasion which is a core functionality in the EMT process. KLF8 also plays its role in tumor progression which enables cancer cells to acquire mobility in order to spread to different parts to the body. It does so by directly binding and repressing the promoter of E-cadherin independent of E boxes and SNAIL expression. Therefore, more expression of KLF8, lower is the expression of E-cadherin in invasive human breast carcinoma (Xianhui Wang et al., 200(.

In Metastasis

Metastasis is the mobilization of tumorigenic cells from their primary site to the secondary one by rupturing the extracellular matrix. This rupturing of the extracellular matrix for invasion of surrounding cells is brought about by the matrix metalloproteases (MMPs).

MMP9

KLF8 regulates MMPs in breast cancer. This was studied by employing two cell lines, one to check the impact of overexpression of KLF8 on MCF-10A and the second, MDA-MB-231 to check the impact of knocking down the KLF8. High expression of KLF8 enhanced MMP9 expression. Opposite to this, KLF8 knockdown led to reduced expression of MMP9. This signified the role of KLF8 and MMP9 signaling in human breast cancer invasion and metastasis (X Wang et al., 2011).

MMP14

Another role of KLF8 was identified in MMP14 signaling in promotion of invasiveness and metastasis of human breast cancer. At transcriptional level, the KLF8 promotes MMP14 expression and its knockdown showed the opposite effect. This happens through two mechanisms, either by directly affecting the promoter or by indirectly promoting the nuclear translocation of β-catenin, expressing T-cell factor-1 (TCF1) and successively activating the promoter by β-catenin/TCF1 complex. Moreover, by down-regulating focal adhesion kinase (FAK) or its knockdown led to the deduction that the cell surface presentation of active MMP14 is dependent upon FAK expression and activity downstream of KLF8. This signifies the combined role of KLF8 and FAK to regulate the activity of MMP14 for breast cancer metastasis (H Lu et al., 2014).

CXCR4

Another method adopted by KLF8 to induce metastasis is through induction of CXCR4. A cell line MDA-MB-231 was used to carry out experiments to check this effect. Treatment of cells in this cell line with a variant CXCL12, of CXCR4 ligand showed the formation of filopodium-like protrusions. Overexpression of the factor, KLF8 as well as CXCR4 led to aggressiveness of tumor through metastasis in immunocompromised mice. Overall, this pointed towards recurrence of breast cancer malignancy (Mukherjee et al., 2016).

EPSTI1

The Epithelial stromal interaction 1 (EPSTI1) is a newly identified stromal fibroblast-induced gene. It has been shown to play a role in invasive breast cancer malignancy by being overexpressed. Two cell lines were used to check the effect of KLF8 expression on EPSTI1. One cell line was used to check the impact of overexpression of KLF8 on EPSTI1 in MCF-10A and the second cell line, MDA-MB-231 was used to check the impact of knocking down the KLF8. High expression of KLF8 enhanced EPSTI1 expression. Opposite to this, KLF8 knockdown led to its reduced expression. KLF8 attached to the promoter region of EPSTI1. EPSTI1 was explored to interact with the valosin containing protein (VCP), degrading the IκBα which ultimately activated of NF-κB in the nucleus. This however led to the conclusion that KLF8 signaling mechanism along with EPSTI1, VCP, and NF-κB is critical in invasion and metastasis of human breast carcinoma (T Li et al., 2014).

In Drug-resistant Metastasis

KLF8 was observed to be over expressed by treatment of LM2 cells with chemotherapy drugs. This over-induction of KLF8 or CXCR4 is capable of causing increased metastasis to various organs and led to excessive secretion of VEGF, establishment of new vasculature, formation of filopodium-like protrusions, and colonization. Colonization is changing of micro-metastasis to macro-metastasis at distant sites. Such enhancement of KLF8 upon chemotherapy and the association of KLF8/CXCR4 indicated drug-resistant metastasis and can prove to be a new drug target (Hao et al., 2021).

In DNA Repair

A novel role of KLF8 includes protecting the breast cancer cells' from dying and DNA from being damaged by harmful effects of doxorubicin. In this way, it repairs the DNA of breast cancer cells. As soon as the DNA of these cells gets damaged, KLF8 is phosphorylated by DNA-dependent protein kinase's catalytic subunit. It is then sumoylated

by SUMO E3 ligase protein inhibitors of activated STAT. the process of DNA repair then starts as the KLF8 is recruited to the site of DNA damage (Heng Lu et al., 2012).

Stem cell-like Activity

Stem cells are the cells that have the capability to self-renew and to differentiate into mature cell types. These mature cells are specialized to perform specific functions. In case of tumors, the sub-population of stem cells are called the cancer stem cells (CSCs). These illustrate the properties of stem cells as well as cancer cells. The breast cancer stem cells (BCSCs), promote breast cancer invasion and metastasis. The tumor microenvironment thus established, plays a critical role in growth signaling and proliferation (He et al., 2021).

In the case of KLF8, it was observed to be overexpressed in triple-negative breast cancer cells, which further enhanced the population of CSCs. On the contrary, downregulating the expression of kLF8 reduced the CSC population. Recently, it has been observed that a nutrient sensor O-GlcNAc transferase (OGT) and O-GlcNAcylation for increased in the population of cancer stem cells. This promoted stemness and tumorigenesis of breast cancer cells, both in-vivo and in-vitro. This increment was interlinked with increased expression of KLF8 suggesting its potential role in rendering CSC traits (Le Minh et al., 2023).

Invasion through targeting MicroRNAs

MicroRNAs (miRs) are short non-coding regulatory RNAs. They target the 3'UTR (untranslated region) of mRNA. They are tightly regulated in normal cell. However, their role as critical regulators of tumor progression and as cancer stem cells has been explored in the recent past (Mo, 2012).

miRNA-146a

The KLF8 has emerged as a promoter of breast cancer through rendering the cancer stem cell ability to the mammary stem cells. This ability of stem cells can be induced during epithelial to mesenchymal transition (EMT). The miR-146a has been explored

as a target of KLF8 and TGF-β which are responsible for inducing EMT. It was observed that the MCF-10A cells acquired the stem cell properties when KLF8 was over-expressed or treated with TGF-β. This led to overexpression of miR-146a in invasive breast cancer cell lines. On the other hand, if the cells had been treated with miR-146a inhibitors, the MCF-10A cells were incapable of having the stem cell traits. The reason was identified as miR-146a targeting the 3'UTR of Notch signaling inhibitor NUMB for translational inhibition. Hence, the miR-146a potentially activated Notch signaling by being stimulated by KLF8 (Wang et al., 2013).

miRNA-141

The KLF8 is found to target the miR-141. It represses the promoter of miR-141 which targets the 3'-UTR of Epidermal Growth Factor Receptor (EGFR) to stop its translation. EGFR overexpression had been observed in invasive triple-negative breast cancer. Upon upregulating the expression of KLF8 in non-tumorigenic MCF-10A cells, led to overexpression of EGFR whereas knocking it down in another cell line, MDA-MB-231 downregulated it. This shows the impact of KLF8 on expression of miR-141/EGFR signaling pathway in breast cancer malignancy (Tianshu Li et al., 2015).

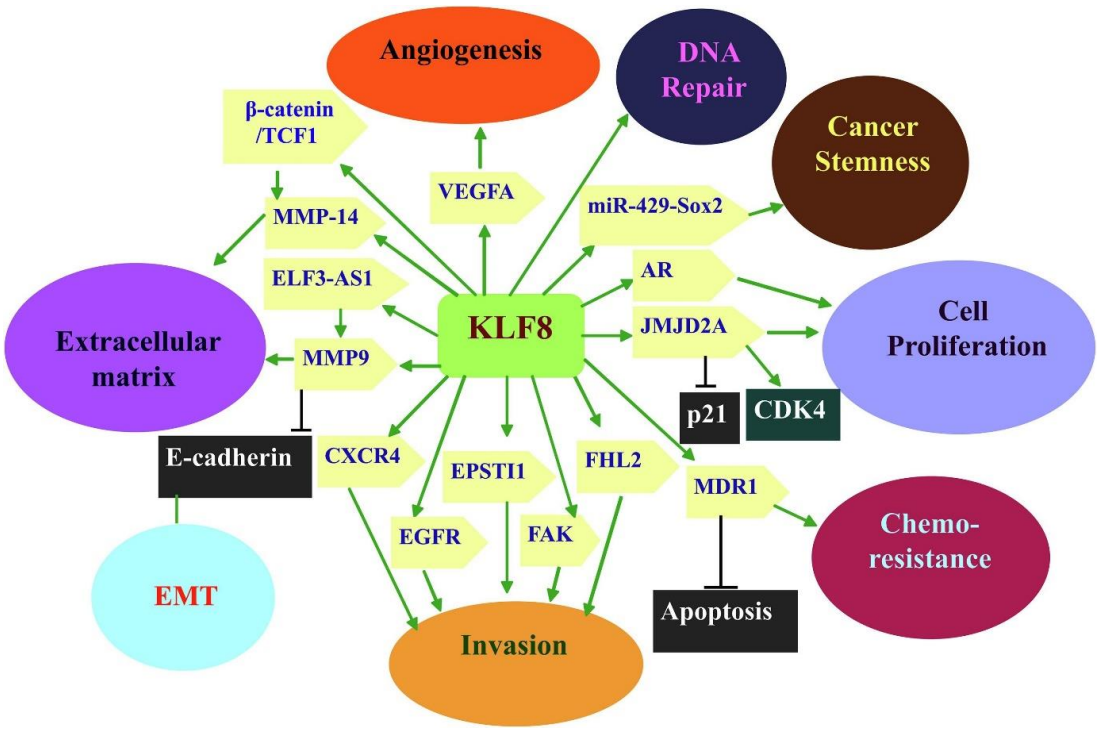



**Figure 2.1:** Role of KLF8 in Cancer (Kumar et al. 2021)

# CHAPTER 3: MATERIALS AND METHODS

## In-Silico Analysis

### *3.1.1 Data Retrieval*

The SNP data for KLF8 was retrieved from the ENSEMBL database under the transcript ID KLF-204-ENST00000468660.6. The information included variant ID, base and residue alterations, genome coordinates and amino acid coordinates. The genome assembly number was GRCh38.p14 (GCA_000001405.29). the transcript had 6 exons is annotated with 24 domains and features, is associated with 17623 variant alleles and maps to 668 oligo probes. This gene has 8 splice variants, orthologues and paralogues.

Further data retrieval was done from the Genome Aggregation Database (gnomAD) and, the Catalogue of Somatic Mutations In Cancer (COSMIC) databases. The KLF8 gene variants were selected excluding the in-frame variants. All the variants were then categorized into two categories. These were named coding variants and regulatory variants. The synonymous, non-synonymous, truncated and frameshift variants were included in the coding variants category while splice site and untranslated region variants (UTRs) were included in regulatory variants. Then the selection criteria was narrowed down to missense variants that give altered protein products. These selected variants were analysed further.

### *3.1.2 Identification of unique missense variants*

From the three databases, ENSEMBL, gnomAD and COSMIC, the number of selected KLF8 variants was summed up. The total number of variants from each database was represented through a graph. These total variants were then compared and the common ones were merged to avoid repetition. All other unique variants were assigned a separate number even if they resided on the same location.

### *3.1.3 Evaluation of deleterious variants*

The missense variants were then filtered using five software tools. These tools include SIFT, PolyPhen, MetaLR, REVEL and Mutation Assessor for determining pathogenicity. These tools give scores to determine the pathogenicity of the variants.

SIFT score predicts whether an amino acid substitution might cause changes in function of the resulting protein. This score ranges from 0.0 to 1.0 where 0.0 to 0.05 score range is predicted deleterious, 0 being the highest value and 0.05 to 1.0 is predicted to be tolerated, 1 being highly tolerated. PolyPhen also applies a score range from 0.0 to 1.0 where 0 indicates benign and 1 indicates deleterious. MetaLR divides the variants into either a 'damaging' or a 'tolerated' category. The scores range from 0 to 1 where a higher value indicates higher chances of being deleterious. REVEL applies a similar score range from 0 to 1 where scores above 0.5 indicate that the variant is 'likely disease-causing' and scores below 0.5 indicate they are 'likely benign'.

After filtering the variants, these were rendered for further analysis. This lead to considering two pathogenic variants of the gene KLF8 that were subjected to wet lab analysis. These variants had rs IDs 'rs1196572444' and 'rs868781835'.

### *3.1.4 Analysis of stability of filtered variants*

For the analysis of stability of filtered variants, different software tools were used. These tools determined the effect of missense mutation on the resulting protein. These softwares include Mupro, Mutpred2 and DynaMut.

For retrieving data from I-Mutant, protein sequence was fed in FASTA format as well as the amino acids and their coordinates. The results were obtained in the form of DDG values (Kcal/mole). A DDG value less than 0 indicates destabilization in protein whereas greater than 0 indicates increasing protein stability.

MUpro is a software that provides a prediction on changes in protein stability due to a variation in a single amino acid. It gives a change in energy score (DDG) which ranges from -1 to 1, where a score less than 0 indicates decreasing stability while greater than 0 indicates increasing stability due to presence of an amino acid variant.

DynaMut tool is responsible for evaluation of the impact of amino acid variants on stability and molecular dynamics of resulting protein. Also, the difference in vibrational entropy (DDS) between mutant and wild type proteins is predicted through this tool.

Another software tool, InterPro was used to gain information on the domains of KLF8, its active site, substrate binding site and the amino acids found in each region. For this, the data was entered in FASTA format.

### *3.1.5 Structural and Functional Analysis of Pathogenic Variants*

The structural and functional analysis of pathogenic variants was done using software tools including MutPred2, DynaMut and HOPE. These tools helped in estimating the effect of these pathogenic variants on structure and function of resultant proteins.

MutPred2 is a software tool that predicts the pathogenic influence of amino acid variants on resulting proteins and their molecular activities through calculating the likelihood of pathogenicity in the form of a score. This probability scores greater than 0.5 indicates pathogenicity but this has a chance of being incorrect. For this, a set criterion has been devised according to which, a score greater than 0.68 indicates 10% likelihood of false positive results. On the other hand, a score greater than 0.80 indicates 5% chance of false positive results.

The DynaMut software tool determines protein flexibility and rigidity. However, this was used to predict the alteration in protein structure.

The HOPE software describes the changed protein domain and charge difference between the wild-type and variant under consideration. It also determines the affect on molecular interaction with surrounding residues. The information regarding change in protein size due to amino acid differences is also provided by HOPE.

### *3.1.6 In-silico mutagenesis*

In-silico mutagenesis was carried out in PyMol, molecular visualization system. The structure of wild KLF8 was visualized and an amino acid in the wild type was changed to

create the mutated variant. The changed amino acid was given a different color from the original one. This was done by selecting the 'Wizard' option > Mutagenesis > Protein and then selecting the residue and converting it to the amino acid as in the variant. The structure was then saved in 'pdb' format. Molecular dynamics simulations were run on wild-type and mutant KLF8 structures.

### *3.1.7 Molecular Dynamics (MD) Simulations*

Molecular dynamics simulations were conducted using GROMACS on a super computer. An open source, terminal emulator, PuTTy was used to give commands to the remote system for performing functions such as cleaning the structure of protein, topology, and labeling of coordinates. Different algorithms performed energy minimization and temperature and pressure equilibrium to carry out 20ns MD. An open source file manager, WinSCP was utilized to transfer files between the local and remote server. When the MD steps were completed in PuTTy, the output data analysis was done to calculate RMSD, RMSF, SASA, Radius of gyration and hydrogen bonds number. The results were depicted graphically.

## 3.2 Primer Designing

Primers were designed to carry out Tetra ARMS PCR. For this, the bioinformatics tool, Primer 1 was used. A total of four, two inner and two outer primers were designed to amplify the gene region containing the variant under study. These primers were then validated through UCSC genome browser in-silico PCR. Here, primer length, temperature, allele difference and SNP position were chosen.

## 3.3 Experimental analysis

### *3.3.1 Sample Collection*

For blood sample collection, an approval was taken from the Institutional Review Board of Atta ur Rahman School of Applied Biosciences. A total of 100 samples each, breast cancer patients and healthy controls were collected. This was done following the

recommendations of the ethical review board. Breast cancer patients without comorbidities such as diabetes and hypertension and pregnant and lactating women were not included in the study. The volunteering patients' consent was taken. Data regarding their medical history, age, tumor grade, cancer type, history of smoking and alcohol consumption and treatment status was also collected. For blood collection, a tourniquet was tied 3cm above the visible vein in the arm or hand and the area was sanitized. Then, using a sterile syringe, 3-5ml of blood was drawn and added to a 5ml EDTA tube.

*3.3.2 DNA extraction*

The organic method was used for DNA extraction from blood samples of experimental and control groups. This method employs phenol, chloroform and isoamyl alcohol. For this, four buffers were prepared, named solution A, B, C and D.

**Solution A:**

0.32M sucrose (54.72g),

10mM Tris-base pH=7.5 (0.602g),

5mM magnesium chloride (0.238g)

} dissolved in autoclaved distilled water to make 500ml of solution

Following this dissolution and autoclaving, 5 mL of 1% V/V Triton X-100 was added and stored at 40º Celsius. This solution breaks cellular membranes to expose the DNA.

**Solution B:**

10mM Tris-base pH=7.5 (0.6057g),

400mM sodium chloride (11.685g),

2mM ethylene diamine tetra acetic acid (0.58g)

} dissolved in distilled autoclaved water with volume raised to 500ml and pH adjusted to 7

**Solution C** is composed of phenol kept in a dark, properly covered reagent container in aluminum foil, in a cold environment

**Solution D** is composed of 24 ml of chloroform and 1 ml of isoamyl alcohol combined to make 50 mL of solution, diluted to 500mL distilled water.

**Procedure:** The process of extracting the DNA takes two days. On the first day, 750µl blood was taken in an Eppendorf tube with an equal volume of Solution A and incubated at room temperature for 10 minutes. It was then centrifuged for 10 minutes at 13000 rpm. Then, the supernatant was discarded. The pallet was resuspended in 500µl Solution A, centrifuged, and the supernatant was discarded. To the supernatant, 400µl of Solution B was added and centrifuged at 13000 rpm for 10 minutes. This caused the proteins to separate and remove from the pallet while precipitating the DNA in the pallet. Then, the pallet was dissolved in 400µl Solution B, 12µl SDS and 5µl proteinase K. The tube was then kept for incubation at 37 ºC for a night.

Next day, the incubated tube was taken, 250 µl of Solution C and the same quantity of Solution D were added and centrifuged at 13000 rpm for 10 minutes. This step was done to create two layers. Solution C separated the aqueous and organic phases in the solution whereas solution D stabilized protein coagulation and aided DNA purification by reducing foam formation. Of these two layers, the upper layer contained DNA and was carefully picked and transferred to a new Eppendorf tube. To this tube, 500µl of isopropanol and 55µl sodium acetate was added and centrifuged at 13000 rpm for 10 minutes. The pallet was suspended in 200µl centrifuged at 13000rpm for 8 minutes. The supernatant was discarded and the tube was inverted for drying. When, dried, 25-30 µl PCR water was added to store the extracted DNA.

*3.3.3 Tetra ARMS PCR*

This procedure was performed to determine the genotype of the samples obtained. This was done using two types of primers, the inner and the outer primers designed according to the variants created by in-silico mutagenesis. A reaction mixture in the quantity of 20µl was prepared. This mixture contained 1 µl of each primer i.e., a total of 4 µl, 2 µl of the

extracted DNA, 2 µl nuclease-free water and 12 µl of SolisFAST master mix. The master mix contained PCR buffer comprising of magnesium chloride, dNTPs and additives that maximize PCR efficiency and DNA polymerase.

The tubes comprising the reaction mixture were centrifuged for a few seconds and placed in the thermal cycler. Primer optimization was carried out at various chemical concentrations and melting temperatures in multiple PCRs. For determining optimum temperature ranges, the gradient PCR technique was applied. The PCR stretched over 30 cycles and three key phases, denaturation, annealing, and extension.

In the first phase, the template DNA was denatured from a double-stranded structure to a single-stranded structure when rendered to 95ºC for 5 minutes. In the next phase, the daughter DNA strands were denatured. This phase stretched to 30 cycles, 30 seconds each at 94ºC. The next phase was primer annealing at 58.5ºC for 30 seconds. Following this phase, was the extension phase carried out at 72ºC for 30 seconds. The final extension period was 7 minutes conducted at 72 ºC. After this, the PCR cycle was paused at 4 ºC to cool the PCR product and temporary storage.

*3.3.4 Agarose Gel Electrophoresis*

Agarose gel electrophoresis was performed to visualize the DNA extracted from the blood samples collected and the PCR products. Agarose gel was made in the concentration of 2% W/V. for this, 2gm of agarose powder was mixed with 1X TAE buffer and the volume was raised to 100ml and heated to ensure complete dissolution. Then, this was slightly cooled and 5-7 µl of ethidium bromide was added for ease of visualization under UV light. The prepared gel was then carefully poured in gel casting tray avoiding bubble formation. A comb was applied to create wells and then the gel was kept to set completely. An electrophoresis tank was filled with 1X TAE buffer. The set gel was taken, comb was removed, and the gel was submerged in the buffer-filled tank. In one of the wells, ladder DNA was injected as a reference in the quantity of 5µl, while in other wells, 8-10 µl of the PCR product was injected. The electrophoresis machine was set at 80V and 500A and the gel was run for 40 minutes. Then the gel was visualized under UV transilluminator.

For visualizing the extracted DNA, the concentration of gel was 1.5% W/V, 8-10 µl DNA sample was combined with 1 µl loading dye and then injected in the wells and visualized under UV light in the similar manner as done for PCR products.

# CHAPTER 4: RESULTS

## 4.1 In-silico Analysis

### *4.1.1 Structure Prediction of KLF8 Protein*

KLF8 protein's structure was predicted by AlphaFold, which used per residue model confidence score (pDDLT) ranging from 0 to 100. Based on this model, a 3D structure was obtained. The interpro analysis provided information on domains of KLF8 protein. There is one N domain from amino acid number 48-216 while three zinc-finger C2H2 domains from 274-303, 304-333 and 334-359. According to AlphaFold, the N domain showed a low confidence score while the other three C2H2 domains showed high and very high confidence score.

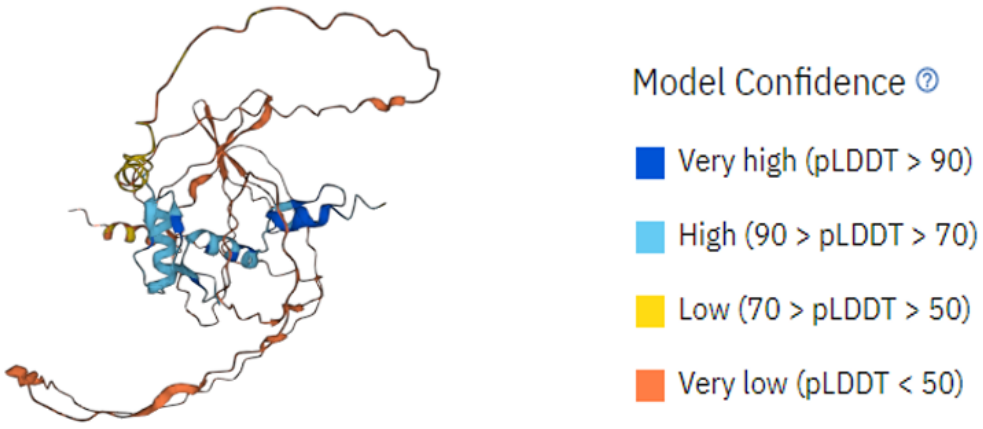



**Figure 4.1:** KLF8 protein's 3D predicted structure, C-terminal region domains showed high confidence score (70-90) of predicted structure.

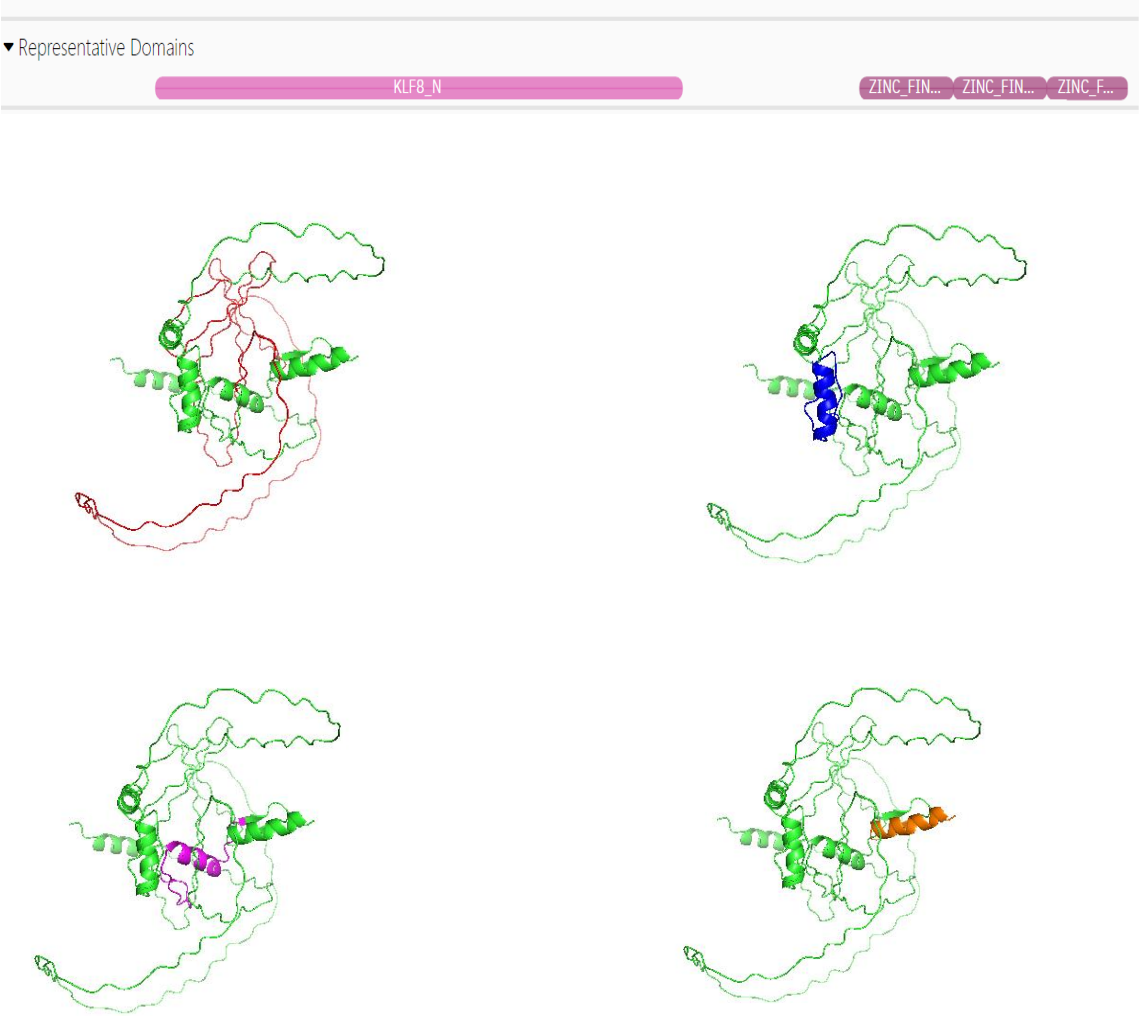



**Figure 4.2:** Representative domains shown in Interpro

The structure was validated in SAVES and PROCHECK analysis revealed that 70.3 percent

of amino acid residues had phi-psi angles in the most preferred regions, 17.3 percent in additionally allowed regions, 8.8 percent in generously allowed regions, and 3.6 percent residue had phi-psi angles in the disallowed regions of the Ramachandran plot

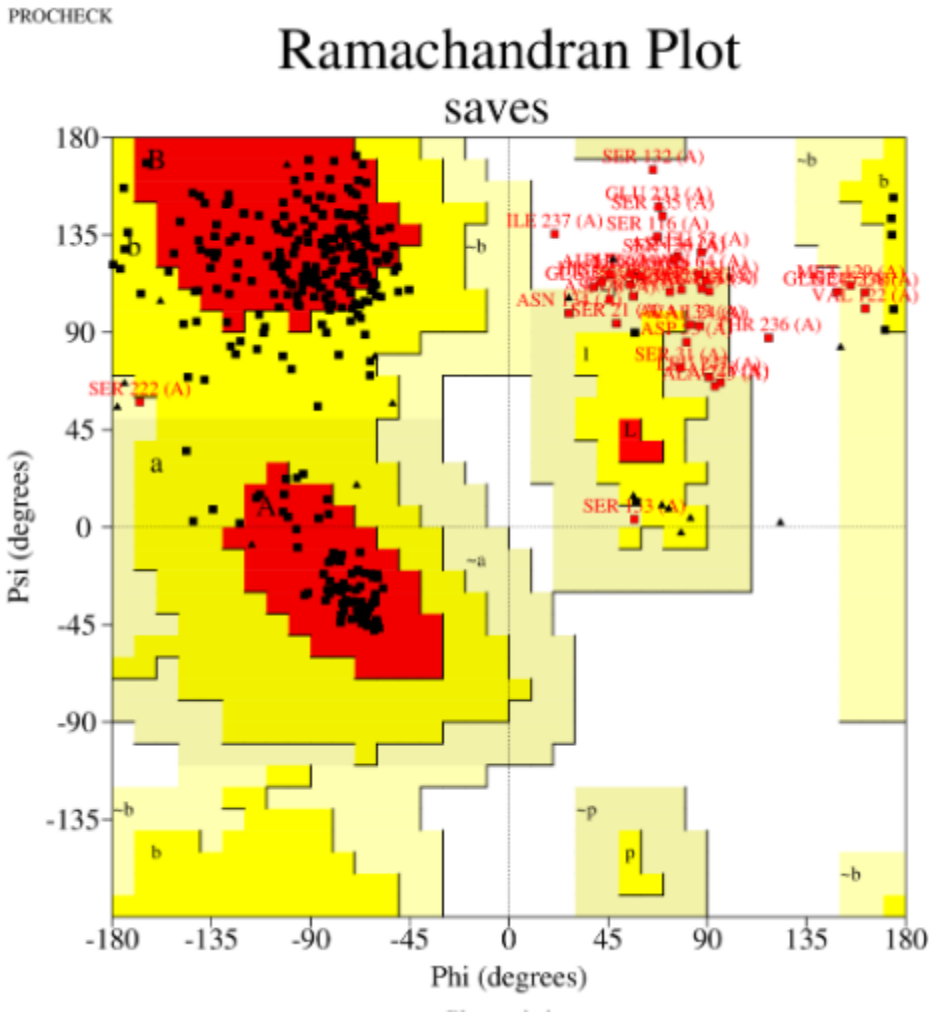



**Figure 4.3** Ramachandran Plot showing the measurements of angles in KLF8

*4.1.2 Subcellular Localization and Phylogenetic Tree*

Deeploc 1.0 predicted the subcellular localization of the KLF8. Figure 4.3 and Table 4.1 below depict the path of KLF8 localization. The path taken by the protein to localize to its compartment within the cell is represented by the red line. The score represents the probability/likelihood of the event. As a result of the score, it's assumed that KLF8 is found in the nucleus.

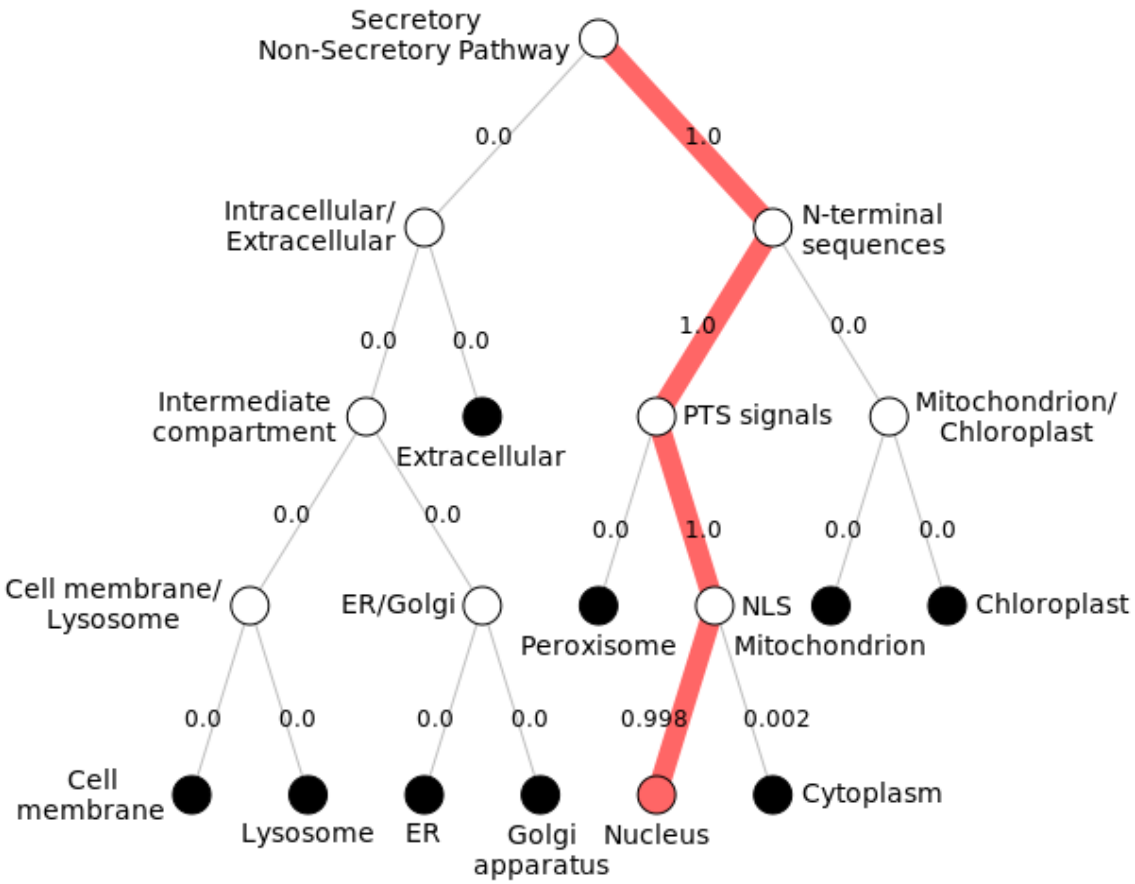



**Figure 4.4:** KLF8 protein's localization route and the likelihood score. The path of localization is shown in red

**Table 4.1** KLF8 Localization and its Likelihood

| **Localization** | **Likelihood** |
|---|---|
| Cytoplasm | 0.0041 |
| Nucleus | 0.9958 |
| Cell membrane | 0 |
| Lysosome/Vacuole | 0 |
| Golgi Apparatus | 0 |
| Peroxisome | 0 |
| Extracellular space | 0 |
| Endoplasmic Reticulum | 0 |
| Mitochondrion | 0 |

Clustal Omega software was used to find the evolutionary relationship between KLF family members. Figure 4.4 depicts the evolutionary profile of KLFs, which reveals that all members of KLF8's descended from a single ancestor protein. The score measures the substitution per site, which indicates how much a specific protein has evolved through time in comparison to its family members. The conserved regions of KLF8 were visualized through ConSurf as shown in the figure below.

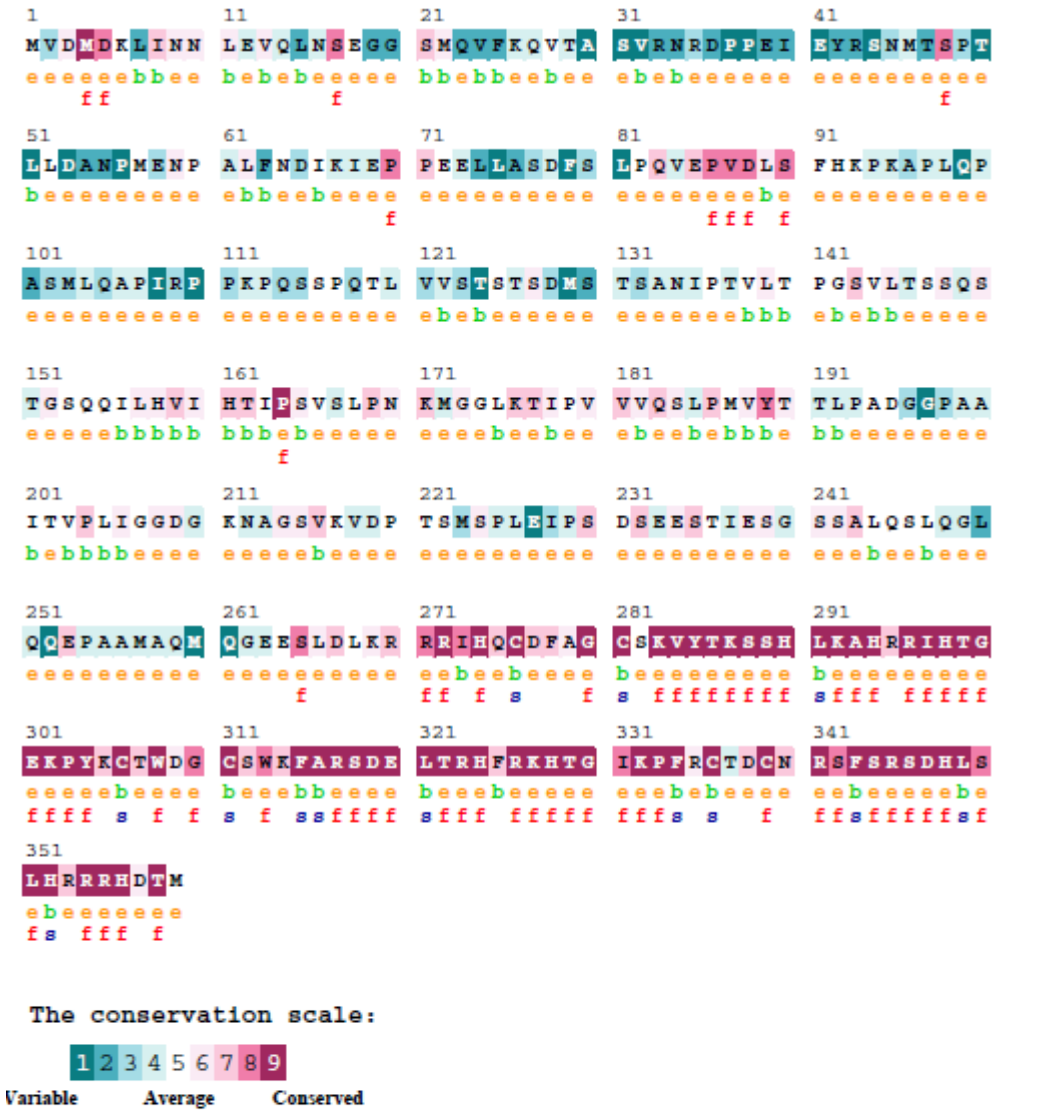



**Figure 4.5:** Conservation Analysis by ConSurf

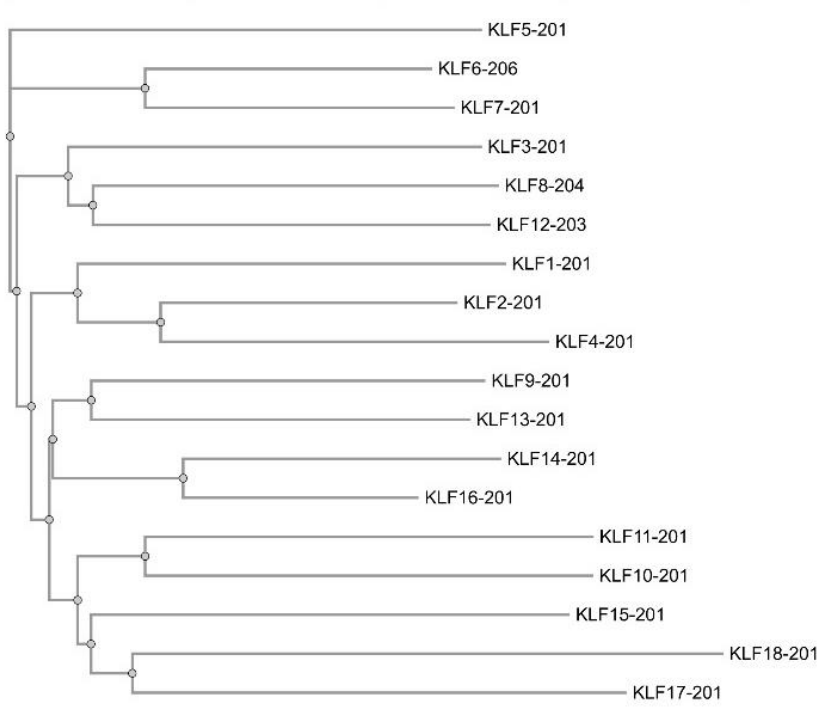



**Figure 4.6** Phylogenetic Tree of the KLF proteins. All the KLFs originated from a common root and then evolve into different classes

*4.1.3 KLF8 Variants Identification*

A total of 486 variants of KLF8 gene, including missense, nonsense, spliced and frameshift variants, were retrieved from multiple databases including Ensembl, COSMIC, and genomAD, containing 251, 85 and 150 variants respectively. Although, Ensembl database included 200 missense, 11 frameshift, 37 spliced variants, 3 nonsense variants, while variants from COSMIC database include 83 missense, 2 frameshift, 0 nonsense and splice variants and genomAD had 131 missense, 5 frameshift, 12 spliced variants and 2 nonsense. Out of 486 variants, 200 were unique and 286 variants were redundant. Unique 200 variants include 159 missense, 16 frameshift, 7 nonsense and 18 spliced variants.

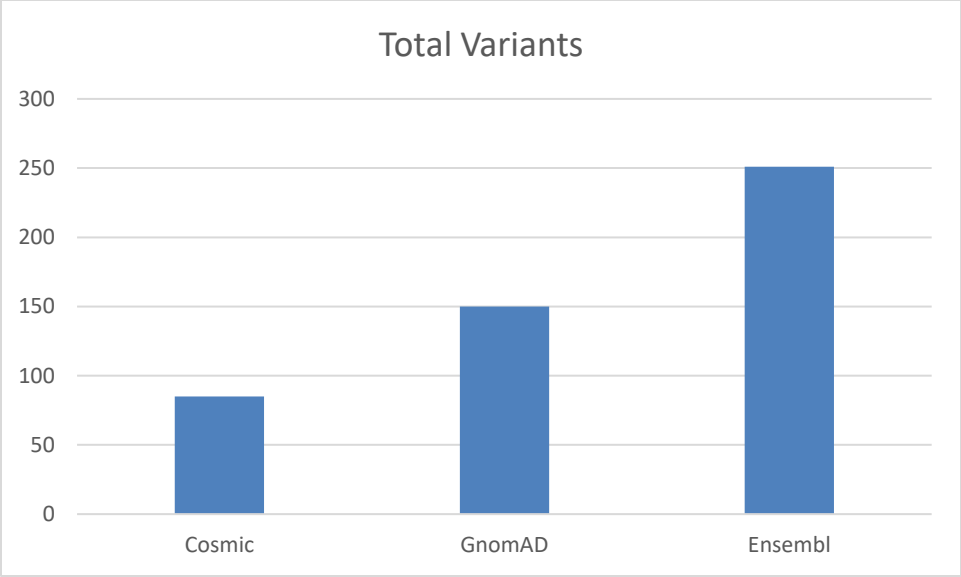



**Figure 4.7** Total variants acquired from all three databases (Cosmic, GnomAD, Ensembl)

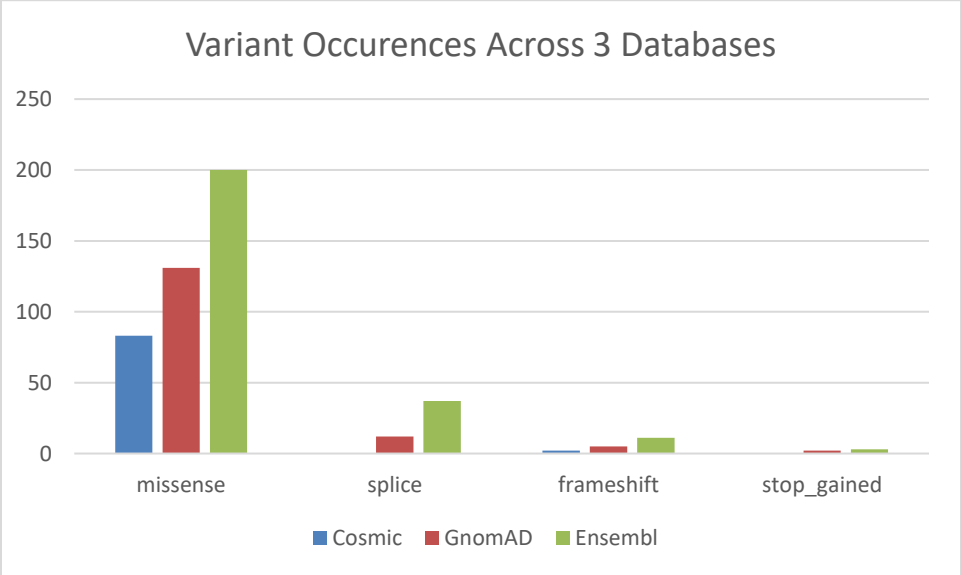



**Figure 4.8** Total Variant Occurrences across all three Databases

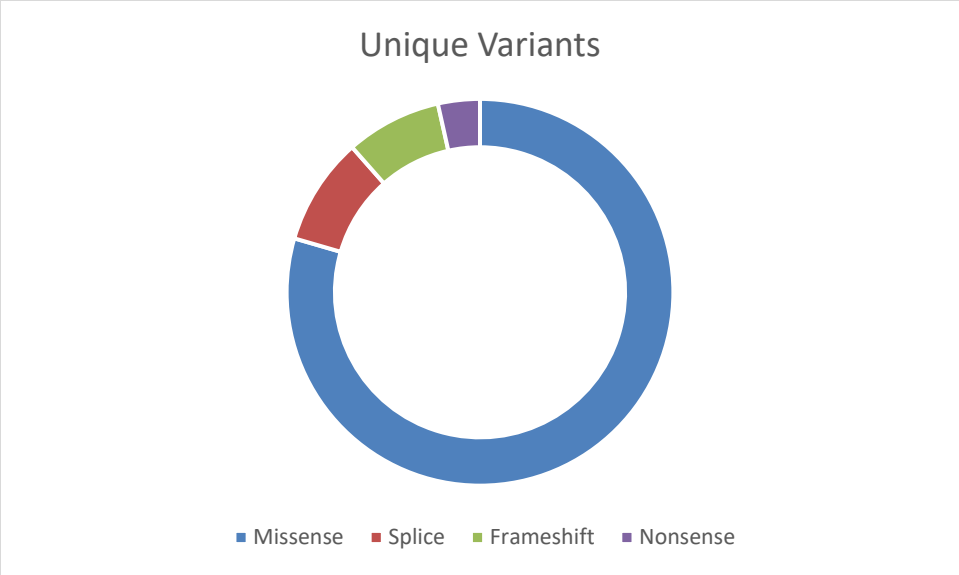



**Figure 4.9** Filtered unique variants from all three databases

Missense variants were filtered from these three databases shown in the figure below.

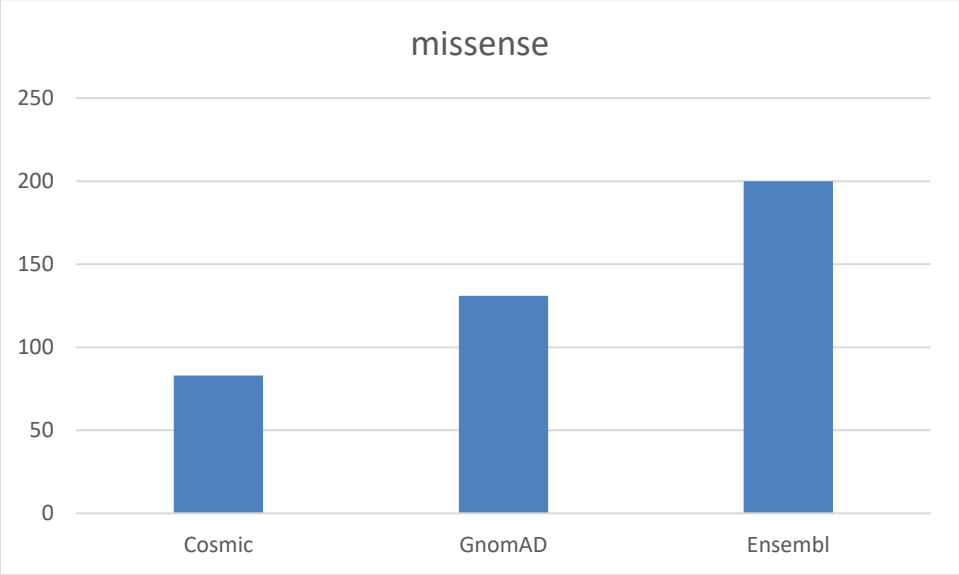



**Figure 4.10** Missense variants filtered from three databases

*4.1.4 KLF8 Missense SNPs Filtration*

The missense data was further filtered to retrieve the SNPs for further investigation. This filtration was based on their pathogenicity percentages. SIFT, PolyPhen, REVEL, Mutation Assessor, and MetaLR tools were used to calculate pathogenicity scores as stated in the methodology. Following table shows the pathogenicity scoring and percentages.

**Table 4.2** Pathogenicity table of missense variants

| **Serial no.** | **Variant ID** | **AA coord** | **sift_class** | **polyphen_class** | **revel_class** | **meta_lr_class** | **mutation_assessor_class** | **Pathogenicity %** |
|---|---|---|---|---|---|---|---|---|
| 1 | rs911255592 | 3 | 0 | 0 | - | - | - | 0% |
| 2 | rs776924073 | 4 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 3 | rs750116979 | 5 | 1 | 0 | 0 | 0 | 0 | 20% |
| 4 | rs1457982343 | 7 | 1 | 0 | 0 | 0 | 0 | 20% |
| 5 | rs759821417 | 9 | 0 | 0 | 0 | 0 | 0 | 0% |
| 6 | rs1602425063 | 15 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 7 | rs765607510 | 17 | 1 | 0 | 0 | 0 | 0 | 20% |
| 8 | rs763410667 | 18 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 9 | rs1409861399 | 19 | 1 | 0 | 0 | 0 | 0 | 20% |
| 10 | rs1329785837 | 19 | 1 | 0 | 0 | 0 | 0 | 20% |
| 11 | rs764459709 | 20 | 1 | 0 | 0 | 0 | 0 | 20% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12 | rs752124107 | 24 | 0 | 0 | 0 | 0 | 0 | 0% |
| 13 | rs1229194237 | 26 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 14 | rs1438658965 | 26 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 15 | rs1438658965 | 26 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 16 | rs1398077363 | 28 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 17 | rs1040382871 | 29 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 18 | rs1340917869 | 29 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 19 | rs368900678 | 33 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 20 | rs144407506 | 33 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 21 | rs144407506 | 33 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 22 | rs774648329 | 34 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 23 | rs762366439 | 37 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 24 | rs762366439 | 37 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 25 | rs768058968 | 38 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 26 | rs759073766 | 40 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 27 | rs1417633235 | 42 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 28 | rs368933949 | 43 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 29 | rs758061219 | 48 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 30 | rs777618241 | 49 | 0 | 0 | 0 | 0 | 0 | 0% |
| 31 | rs751362149 | 50 | 0 | 0 | 0 | 0 | 0.5 | 10% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 32 | rs757188676 | 51 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 33 | rs757188676 | 51 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 34 | rs1253393491 | 52 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 35 | rs899019890 | 53 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 36 | rs745893681 | 53 | 0 | 0 | 0 | 0 | 0 | 0% |
| 37 | rs745893681 | 53 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 38 | rs770010841 | 53 | 0 | 0 | 0 | 0 | 0 | 0% |
| 39 | rs780360757 | 56 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 40 | rs1342034737 | 57 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 41 | rs1245069875 | 58 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 42 | rs139081150 | 61 | 0 | 0 | 0 | 0 | 0 | 0% |
| 43 | rs1050502377 | 62 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 44 | rs768821062 | 62 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 45 | rs774753989 | 69 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 46 | rs887584045 | 69 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 47 | rs1306007343 | 70 | 1 | 1 | 0 | 1 | 0.5 | 70% |
| 48 | rs1200383021 | 70 | 1 | 1 | 0 | 1 | 0.5 | 70% |
| 49 | rs868781835 | 71 | 1 | 1 | 1 | 1 | 0.5 | 90% |
| 50 | rs143969730 | 75 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 51 | rs772525064 | 80 | 0 | 0 | 0 | 0 | 0.5 | 10% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 52 | rs1379874289 | 81 | 0 | 1 | 0 | 0 | 0 | 20% |
| 53 | rs1160583250 | 84 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 54 | rs1361224783 | 87 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 55 | rs1361224783 | 87 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 56 | rs1404365594 | 89 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 57 | rs773726642 | 94 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 58 | rs758988041 | 95 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 59 | rs1451874481 | 98 | 0 | 1 | 0 | 0 | 0 | 20% |
| 60 | rs752213619 | 100 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 61 | rs752213619 | 100 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 62 | rs1330392987 | 102 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 63 | rs1047950271 | 104 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 64 | rs1427759514 | 106 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 65 | rs146429909 | 108 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 66 | rs139841730 | 109 | 0 | 0 | 0 | 0 | 0 | 0% |
| 67 | rs139841730 | 109 | 0 | 0 | 0 | 0 | 0 | 0% |
| 68 | rs1024570270 | 109 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 69 | rs753632321 | 110 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 70 | rs753632321 | 110 | 0 | 0 | 0 | 0 | 0 | 0% |
| 71 | rs756190986 | 111 | 1 | 1 | 0 | 0 | 0.5 | 50% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 72 | rs756190986 | 111 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 73 | rs756190986 | 111 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 74 | rs1239437887 | 114 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 75 | rs780171768 | 115 | 0 | 0 | 0 | 0 | 0 | 0% |
| 76 | rs749373892 | 115 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 77 | rs1383363632 | 117 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 78 | rs374892798 | 118 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 79 | rs1180735183 | 119 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 80 | rs779031240 | 120 | 0 | 0 | 0 | 0 | 0 | 0% |
| 81 | rs748491106 | 120 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 82 | rs1344440945 | 122 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 83 | rs143280924 | 124 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 84 | rs771483872 | 126 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 85 | rs1397842630 | 129 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 86 | rs775036329 | 129 | 0 | 0 | 0 | 0 | 0 | 0% |
| 87 | rs775036329 | 129 | 0 | 0 | 0 | 0 | 0 | 0% |
| 88 | rs1221527601 | 134 | 0 | 0 | 0 | 0 | 0 | 0% |
| 89 | rs1221527601 | 134 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 90 | rs763754140 | 135 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 91 | rs763754140 | 135 | 1 | 0 | 0 | 0 | 0.5 | 30% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 92 | rs757211205 | 138 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 93 | rs36102105 | 139 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 94 | rs1314515814 | 140 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 95 | rs913427803 | 141 | 0 | 0 | 0 | 0 | 0 | 0% |
| 96 | rs767419429 | 142 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 97 | rs750305226 | 143 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 98 | rs756101201 | 145 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 99 | rs200907884 | 148 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 100 | rs755032067 | 150 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 101 | rs1433771646 | 150 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 102 | rs779134931 | 156 | 0 | 0 | 0 | 0 | 0 | 0% |
| 103 | rs1169792053 | 161 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 104 | rs369282415 | 162 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 105 | rs778094890 | 165 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 106 | rs747415809 | 166 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 107 | rs1209125016 | 166 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 108 | rs1569186150 | 167 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 109 | rs1569186150 | 167 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 110 | rs771429775 | 168 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 111 | rs777027560 | 169 | 0 | 1 | 0 | 0 | 0.5 | 30% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 112 | rs1054205139 | 169 | 1 | 1 | 0 | 0 | 0 | 40% |
| 113 | rs748774811 | 174 | 0 | 0 | 0 | 0 | 0 | 0% |
| 114 | rs1220892845 | 180 | 0 | 0 | 0 | 0 | 0 | 0% |
| 115 | rs1274128070 | 182 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 116 | rs1321414525 | 182 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 117 | rs761546600 | 187 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 118 | rs140327143 | 188 | 0 | 0 | 0 | 0 | 0 | 0% |
| 119 | rs1260887085 | 189 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 120 | rs1050451843 | 190 | 1 | 0 | 0 | 0 | 0 | 20% |
| 121 | rs1194175582 | 193 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 122 | rs760574954 | 195 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 123 | rs766287376 | 195 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 124 | rs753727815 | 197 | 0 | 0 | 0 | 0 | 0 | 0% |
| 125 | rs753727815 | 197 | 0 | 0 | 0 | 0 | 0 | 0% |
| 126 | rs1569186275 | 198 | 0 | 1 | 0 | 0 | 0 | 20% |
| 127 | rs755018799 | 199 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 128 | rs1381116096 | 200 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 129 | rs1418504822 | 201 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 130 | rs1298150637 | 205 | 1 | 1 | 0 | 0 | 0 | 40% |
| 131 | rs1359243793 | 206 | 1 | 1 | 0 | 0 | 0 | 40% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 132 | rs1402386561 | 208 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 133 | rs1484964264 | 211 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 134 | rs1484964264 | 211 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 135 | rs376737146 | 213 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 136 | rs754199549 | 214 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 137 | rs1018147187 | 218 | 0 | 0 | 0 | 0 | 0 | 0% |
| 138 | rs1274299131 | 220 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 139 | rs1274299131 | 220 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 140 | rs1220870159 | 221 | 1 | 0 | 0 | 0 | 0 | 20% |
| 141 | rs765267281 | 222 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 142 | rs771485287 | 223 | 0 | 0 | 0 | 0 | 0 | 0% |
| 143 | rs200006019 | 223 | 0 | 0 | 0 | 0 | 0 | 0% |
| 144 | rs1324948642 | 227 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 145 | rs370185272 | 231 | 0 | 1 | 0 | 0 | 0 | 20% |
| 146 | rs145056308 | 235 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 147 | rs1157044023 | 235 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 148 | rs757569499 | 236 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 149 | rs1456723587 | 236 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 150 | rs781401053 | 237 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 151 | rs763246473 | 239 | 1 | 0 | 0 | 0 | 0.5 | 30% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 152 | rs766866162 | 239 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 153 | rs780734959 | 240 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 154 | rs747609626 | 240 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 155 | rs1160586311 | 243 | 1 | 0 | 0 | 0 | 0 | 20% |
| 156 | rs995444916 | 245 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 157 | rs1402780164 | 246 | 0 | 0 | 0 | 0 | 0 | 0% |
| 158 | rs142294522 | 247 | 0 | 1 | 0 | 0 | 0 | 20% |
| 159 | rs1477147807 | 249 | 0 | 1 | 0 | 0 | 0 | 20% |
| 160 | rs1477147807 | 249 | 0 | 1 | 0 | 0 | 0 | 20% |
| 161 | rs1433525053 | 256 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 162 | rs755606346 | 258 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 163 | rs1057099030 | 259 | 0 | 0 | 0 | 0 | 0.5 | 10% |
| 164 | rs1602457545 | 261 | 0 | 0 | 0 | 0 | 0 | 0% |
| 165 | rs1287721402 | 264 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 166 | rs756827184 | 272 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 167 | rs745593957 | 275 | 0 | 1 | 0 | 0 | 0 | 20% |
| 168 | rs1379763969 | 276 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 169 | rs1290497846 | 282 | 1 | 0 | 0 | 0 | 0 | 20% |
| 170 | rs1385427586 | 284 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 171 | rs749213915 | 293 | 1 | 1 | 0 | 0 | 0.5 | 50% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 172 | rs139041929 | 295 | 1 | 1 | 0 | 0 | 0 | 40% |
| 173 | rs774267049 | 296 | 1 | 1 | 0 | 0 | 0 | 40% |
| 174 | rs760877139 | 306 | 1 | 0 | 1 | 1 | 1 | 80% |
| 175 | rs1391902410 | 307 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 176 | rs969886508 | 309 | 1 | 0 | 0 | 0 | 0 | 20% |
| 177 | rs766842755 | 312 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 178 | rs754312008 | 313 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 179 | rs1331025659 | 313 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 180 | rs1322306475 | 317 | 1 | 0 | 0 | 0 | 0 | 20% |
| 181 | rs1405382114 | 319 | 1 | 0 | 0 | 0 | 0.5 | 30% |
| 182 | rs755518394 | 322 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 183 | rs1204570188 | 323 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 184 | rs1286585777 | 326 | 1 | 1 | 0 | 0 | 0 | 40% |
| 185 | rs1380748194 | 331 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 186 | rs1456582819 | 332 | 1 | 1 | 1 | 1 | 0.5 | 90% |
| 187 | rs1569194785 | 333 | 1 | 1 | 0 | 0 | 0 | 40% |
| 188 | rs1251320193 | 337 | 0 | 1 | 0 | 0 | 0.5 | 30% |
| 189 | rs756805881 | 337 | 1 | 1 | 0 | 0 | 0 | 40% |
| 190 | rs1196572444 | 339 | 1 | 1 | 1 | 1 | 1 | 100% |
| 191 | rs780645349 | 340 | 1 | 1 | 0 | 0 | 0.5 | 50% |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 192 | rs1016774145 | 341 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 193 | rs1470090266 | 341 | 1 | 0 | 0 | 0 | 0 | 20% |
| 194 | rs1169656171 | 344 | 1 | 1 | 0 | 0 | 0 | 40% |
| 195 | rs1397734593 | 345 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 196 | rs745530968 | 345 | 1 | 1 | 0 | 0 | 0.5 | 50% |
| 197 | rs1298321992 | 348 | 1 | 0 | 0 | 0 | 0 | 20% |
| 198 | rs1555939511 | 355 | 1 | 1 | 1 | 0 | 0 | 60% |
| 199 | rs1281874686 | 357 | 1 | unknown | 0 | 0 | 0.5 | 30% |
| 200 | rs201887872 | 358 | 1 | unknown | 0 | 0 | 0 | 20% |

After that, the percentage pathogenicity for all variants was determined and plotted as shown in the following graph.

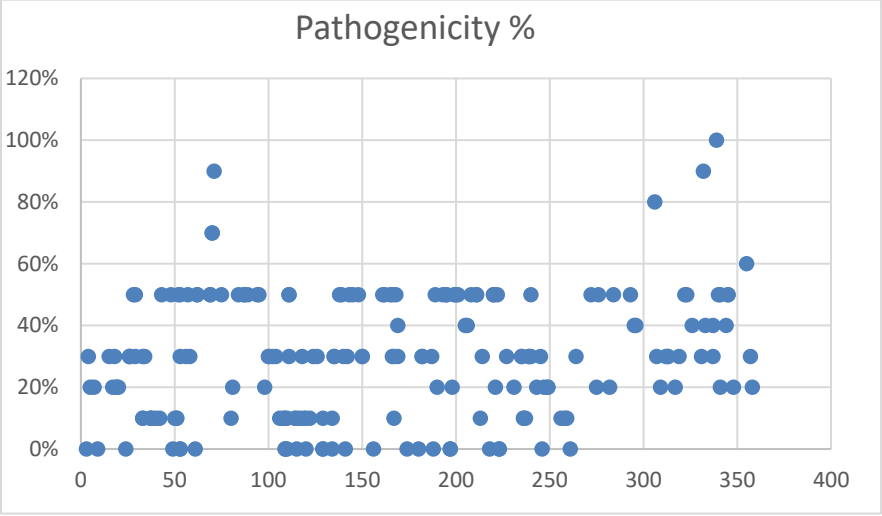



**Figure 4.11** Pathogenicity percentages of the missense variants

Out of 200 missense variants, threshold 85% pathogenicity was adjusted to filter out deleterious variants that could affect resultant protein structure and functions. After

applying threshold pathogenicity filter, 3 pathogenic variants were obtained with pathogenicity 85% or above shown in table 4.2. and two variants (rs868781835, and rs1196572444) at amino acid number 71 and 339 with highest pathogenicity 90% and 100% respectively were shortlisted for further protein analysis.

**Table 4.3** missense variants after threshold pathogenicity sorting

| Variant ID | Alleles | AA | AA coord | sift_class | polyphen_class | revel_class | meta_lr_class | mutation_assessor_class |
|---|---|---|---|---|---|---|---|---|
| rs868781835 | C/T | P/L | 71 | D | PD | LDC | DA | M |
| rs1456582819 | A/G | K/E | 332 | D | PD | LDC | DA | M |
| rs1196572444 | T/G | C/G | 339 | D | PD | LDC | DA | H |

(D= Deleterious, PD= Probably damaging, LDC= Likely disease causing, D= damaging, H= High, M= Medium)

**Table 4.4:** Selected Missense Variants according to Pathogenicity

| Variant ID | AA coord | Conseq. Type | sift_class | polyphen_class | revel_class | meta_lr_class | mutation_assessor_class | Pathogenicity % |
|---|---|---|---|---|---|---|---|---|
| rs868781835 | 71 | missense variant | 1 | 1 | 1 | 1 | 0.5 | 90% |
| rs1456582819 | 332 | missense variant | 1 | 1 | 1 | 1 | 0.5 | 90% |
| rs1196572444 | 339 | missense variant | 1 | 1 | 1 | 1 | 1 | 100% |

The SNP frequency per Exon was analyzed as shown below.

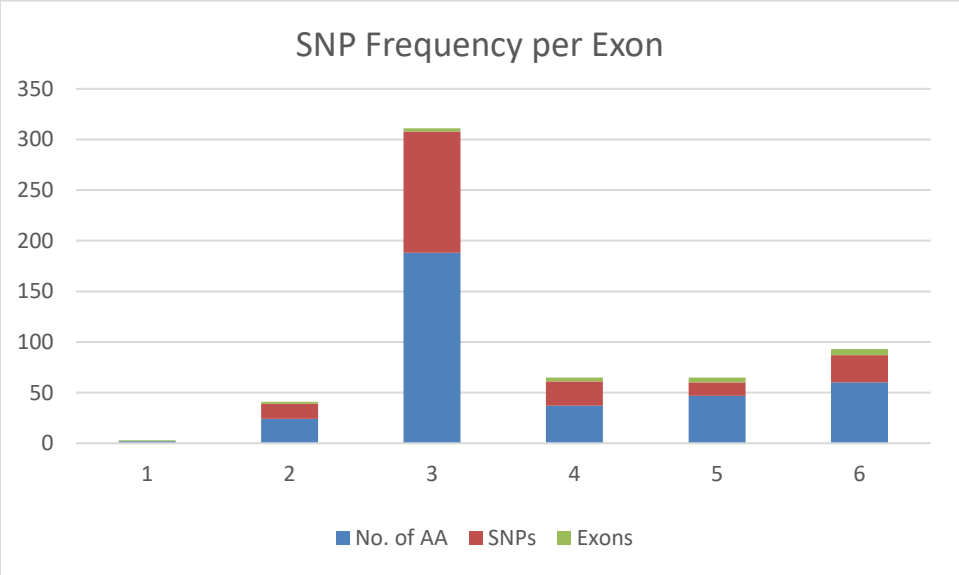



**Figure 4.12** SNP Frequency per Exon

*4.1.5 Protein Stability Analysis*

MuPro, DynaMut and Maestroweb were used to do the structural and functional investigation of the SNPs rs1196572444 and rs868781835. MUpro generated a DDG value of -1.7939459 kcal/mol for SNP1, indicating decreasing stability of the protein structure and a value of 0.75721874 kcal/mol, indicating an increasing stability. DynaMut DDG values indicated decreasing stability of both the protein structures associated with the SNPs under study. The DDG value for both the SNPs is 0.383 and 0.856 Kcal/mol, according to MaestroWeb results, indicating an increase in stability.

**Table 4.5:** Protein Stability Analysis

| Tools | rs1196572444 | | rs868781835 | |
|---|---|---|---|---|
| | **DDG Value** | **Consequence** | **DDG Value** | **Consequence** |
| **MUpro** | -1.7939459 kcal/mol | Decreasing Stability | 0.75721874 kcal/mol | Increasing Stability |
| **DynaMut** | 0.056 kcal/mol | Decreasing Stability | -0.292 kcal/mol | Decreasing Stability |
| **MAESTRO WEB** | 0.383 kcal/mol | Increasing Stability | 0.856 kcal/mol | Increasing Stability |

*4.1.6 Structural and Functional Analysis of Variants*

MutPred2, DynaMut, and HOPE were used to check the effect of amino acid variants on the structure and function of total proteins. The SNPs rs1196572444 and rs868781835 had a general probability score of 0.882 and 0.762 according to MutPred2.

According to Mutpred results, inducing the mutation C339G can result in an altered disordered interface and gain of N-linked glycosylation at N340. Also, the induction of mutation P71L can cause loss of intrinsic disorder, loss of B-factor and loss of ubiquitylation at K67.

**Table 4.6:** Mutpred2 Scoring

| ID | Substitution | MutPred2 score |
|---|---|---|
| **KLF8_HUMAN** | **P71L** | **0.762** |
| **KLF8_HUMAN** | **C339G** | **0.882** |

**Table 4.7** MutPred2 Results

| Residual Change | Mechanism | P-value |
|---|---|---|
| **C339G** | Altered Disordered interface | 0.02 |
| | Gain of N-linked glycosylation at N340 | 0.03 |
| **P71L** | Loss of Intrinsic disorder | 9.3e-03 |
| | Loss of B-factor | 0.03 |
| | Loss of Ubiquitylation at K67 | 0.04 |

DynaMut2 calculated the vibrational entropy energy between the wild-type and mutant proteins to demonstrate a difference in protein flexibility because of mutation. The results are depicted in the following table.

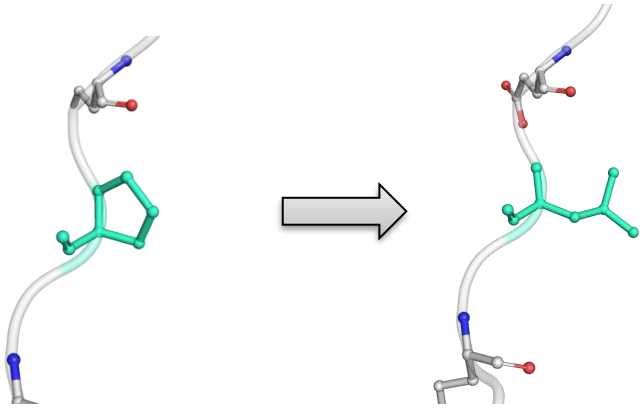
**Table 4.8:** DynaMut2 Results

| Variant | $\Delta\Delta S_{Vib}$ ENCoM | ΔΔG |
|---|---|---|
| **C339G** | -0.070 kcal.mol$^{-1}$.K$^{-1}$ (Decrease of molecule flexibility) | 0.056 kcal/mol (Destabilizing) |
| **P71L** | 0.365 kcal.mol$^{-1}$.K$^{-1}$ (Increase of molecule flexibility) | -0.292 kcal/mol (Destabilizing) |

C339G P71L

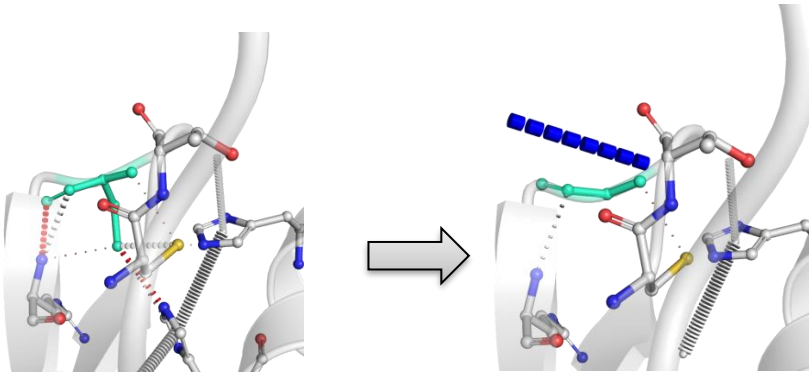
**Figure 4.13:** Variant Structures. Colored Amino acids show change in vibrational entropy due to mutation. BLUE represents a rigidification of the structure and RED a gain in flexibility

Intra-atomic interactions were also predicted by DynaMut. The results indicate different amino acid interactions when compared to wild type. The wild type contained cysteine which was converted to glycine at 339th position in one of the SNPs. In another selected SNP, proline was converted to leucine at position 71.



Wild type C339G

**Figure 4.14:** Intra-atomic Interactions of wildtype and C339G variant



Wild type P71L

**Figure 4.15:** Intra-atomic Interactions of wildtype and P71L variant

In addition to these techniques, the HOPE report shown below in Table revealed that both the variant residue is substantially larger than the wild-type residue, possibly resulting in bumps in the protein structure.

**Table 4.9:** HOPE Report

| SNP ID | Variant Amino Acid | | | | Variant Location | | Functional | Damage |
|---|---|---|---|---|---|---|---|---|
| rs1196572444 | size | charge | | hydrophobicity | Domain | Conservation | Effects on DNA binding | damaging |
| | smaller | Wild | Variant | More hydrophobic | Zinc-finger domain | Highly conserved | | |
| | | Neutral | Neutral | | | | | |
| rs868781835 | bigger | Neutral | Neutral | - | - | Highly conserved | | Probably damaging |

The variant and the wild type residue were neutral even after the mutation, C339G and P71L as shown in the figures below.


**Figure 4.16:** Cysteine is shown on the right side which is converted to Glycine located on the left The size of Cysteine is larger than Glycine as the -SH group is removed

At this position, mutation converts cysteine to glycine which reduces the rigidity of structure and this disturbs the zinc-finger domain. The resulting residue is less hydrophobic as compared to the wild-type. This will lead to a loss of hydrophobic interactions at the core or surface of the protein.


**Figure 4.17** Proline is shown on the right side which mutates to Leucine shown on the left side. The size of Leucine is larger than Proline as the ring structure is opened

This mutation leads to a reduction in rigidity of the protein structure. The mutant residue is larger. This can disturb the structure through the formation of bumps. The HOPE report does not mention the effect of this mutation on hydrophobicity and domains of the protein structure.

*4.1.7 Results of Molecular Dynamics (MD) Simulation*

The files containing data on MD simulations were run for the wild-type and variant proteins. To analyze the results, graphs were generated through Gromancer. To examine the differences between wild type and mutant proteins, four characteristics were used: root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration, number of hydrogen bonds, and solvent accessible surface area (SASA) analysis. The variation of distinct atoms in a protein from its typical position is shown by the root mean square deviation (RMSD). In comparison to wild type, RMSD analysis of the wild and mutant proteins found that the altered protein deviates greatly from its reference position. By comparing the RMSD values, the structural deviations of both wild type and variant proteins over the period of 20ns were determined and plotted in a graph. Both the variant proteins showed deviations when compared to the wildtype. The maximum deviation could be seen around 20 ns starting from around 0.3 ns. This shows that the variants have low stability.

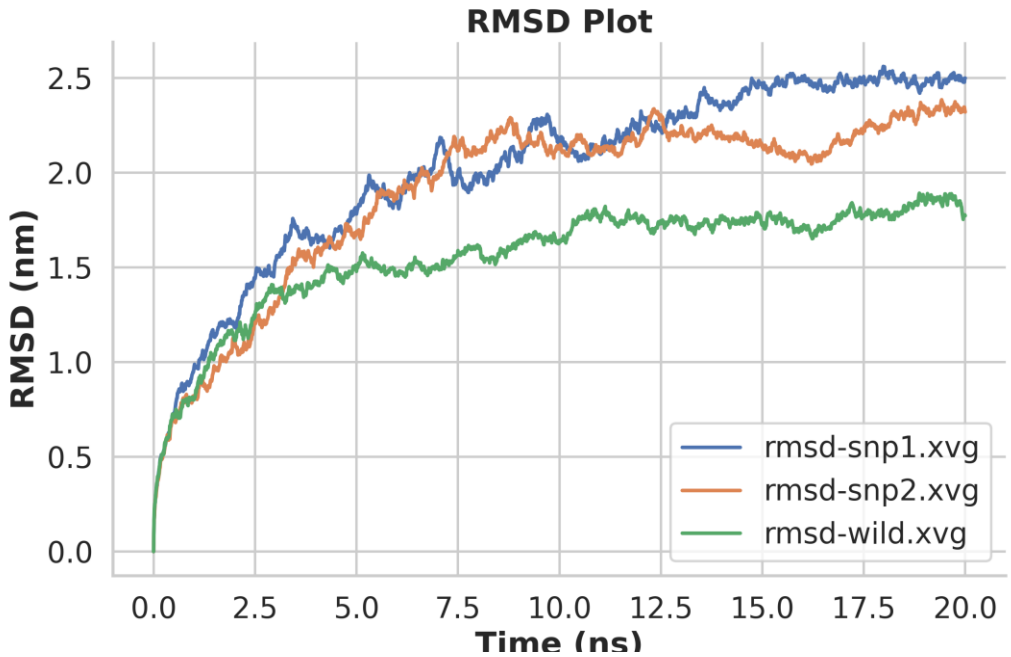



**Figure 4.18:** RMSD plot for wild-type and variant proteins

Individual residues' deviation from their mean position is depicted by root mean square fluctuation (RMSF). A higher RMSF number indicates more flexible movements during simulation, whereas a lower RMSF value indicates limited movements in comparison to its typical position. The difference in the fluctuation of wild and mutant protein residues can be seen using RMSF analysis in figure 25. To see the difference in the dynamic behavior of residues caused by the mutations, RMSFs of wild and both mutant KPCG structures were computed. The peaks represent the residues' highest kinetic energies.

The fluctuation in residues of wild and both variant KLF8 structures ranges from 0.1nm to >2nm. The C339G variant shows higher peaks than the wild and other variant, P71L. The region around 100-150 residues and 200-300 show a higher peaks for SNP1 (C339G) and the region 50-100 and 200-250 show highest peaks for the variant P71L. The wild-type shows highest peaks around 50-150 only. The overall graph shows nearly similar fluctuations for SNP2 while SNP1 shows much difference where the fluctuations are decreasing considerably after 300 residues.

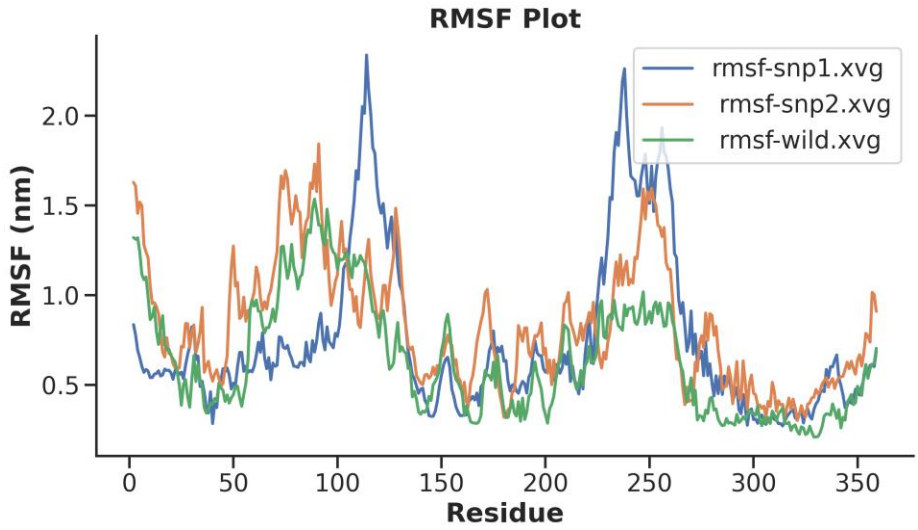



**Figure 4.19:** RMSF plot of wild-type and variant proteins

The radial distance of all the atoms in a protein from their common axis is known as the radius of gyration (Rg). Rg is the radius of a protein structure in a dynamic situation, indicating the protein's compactness and changes in folding over time. The computed radius of gyration for wild and modified proteins is shown in figure 4.20. it can be seen that the Rg value of wild-type protein increases significantly at 2.5 ns and then fluctuates throughout the course of simulation. However, for SNP1 the values drop in the beginning, then increase and then again decrease around 20ns showing an increase in compactness. For SNP2, the graph fluctuates but the values concluded around 20ns are similar to the wild-type.

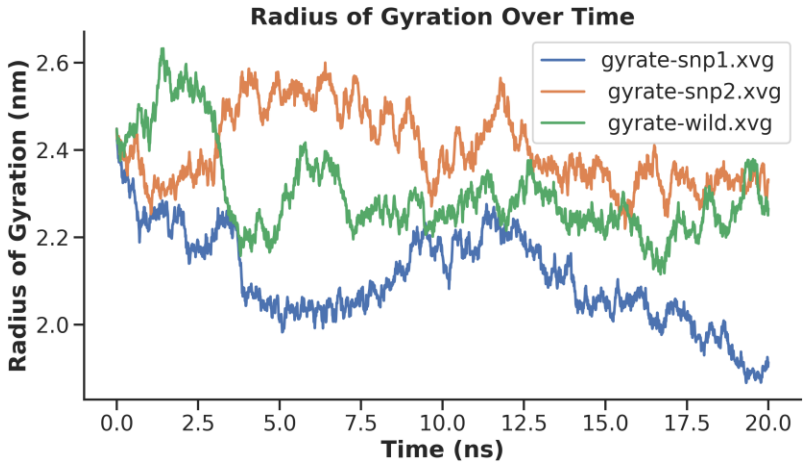



**Figure 4.20:** Radius of gyration of wild-type and variant proteins

Between the polar side chains of amino acid residues, hydrogen bonds develop. The amount of hydrogen bonds formed in all KLF8 structures throughout a 20ns simulation period is depicted in figure 4.21. The number of hydrogen bonds in variant P71L can be seen to be decreasing around 7.5 to 15ns when compared to the wild-type. The wild-type KLF8 hydrogen bonds are overall increasing during the 20ns simulation. However, variation is observed for C339G where the hydrogen bonds are even more in number than the wild KLF8.

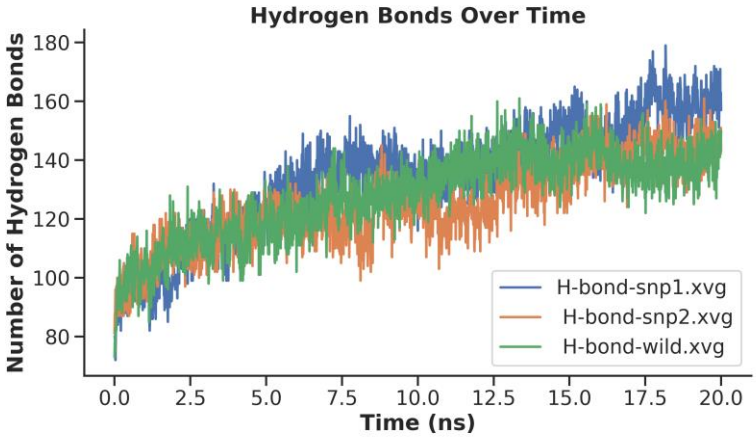



**Figure 4.21:** Difference in number of hydrogen bonds in wild-type and variant proteins

Calculating the solvent accessible surface area (SASA) is a simple technique to address

the surface features of peptides or protein, such as whether the surface is polar or non-polar, or to distinguish between exposed and buried sections or amino acids. the figure 4.22, shows the SASA plot.

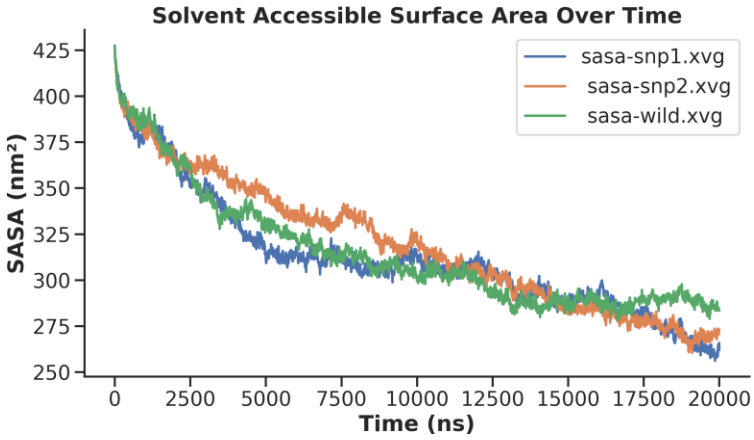



**Figure 4.22:** SASA Analysis of wild-type and variant proteins

The SASA value of both variant protein structures is overall smaller than that of the wild type protein structure, indicating that variant structures are more stable, particularly the C339G variant. The wild and mutant protein structures SASA range is between 250 and 425 nm². The SASA value begins to decrease for the wild-type as well as the variant proteins throughout the simulation period. Both variations resulted in a decrease in the surface area of protein, indicating a decrease in the solvent accessible surface area, which leads to a increase in protein stability. Figure 4.22 shows the graphical representation of surface area over the stimulation period.

*4.1.8 Effects of SNPs on mRNA Secondary Structure*

The mRNA secondary structure of KLF8 wild-type and variant was predicted applying in-silico method. For both, the wild-type and variant, minimum free energy was calculated. For the variant rs1196572444 T/G, the minimum free energy (MFE) of RNA secondary structure changed from -10 kcal/mol of wild-type to -12.6 kcal/mol of the variant. As a result, a significant change in structure was also observed. The decrease in minimum free energy signify an increase in stability of the mRNA structure. In case of the variant rs868781835 C/T, no change in MFE was observed. This showed a value of -7.5 kcal/mol.

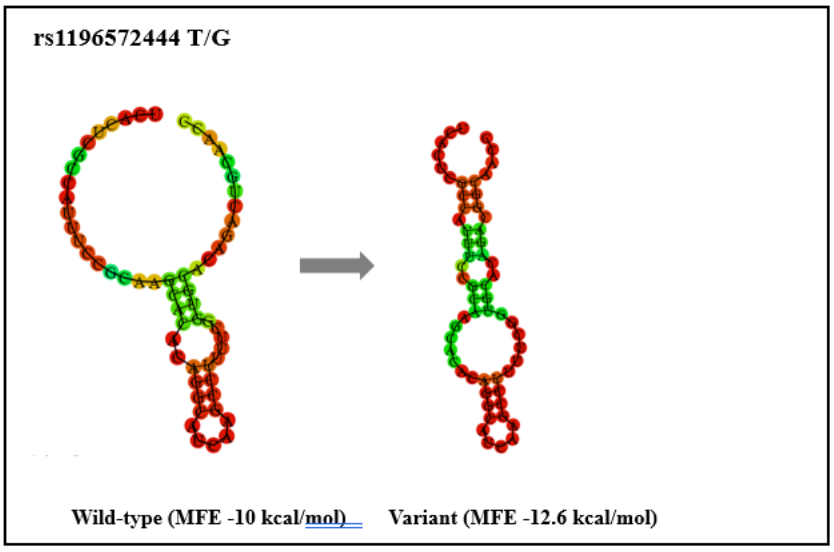



**Figure 4.23:** Minimum free energy of wild-type and variant C339G

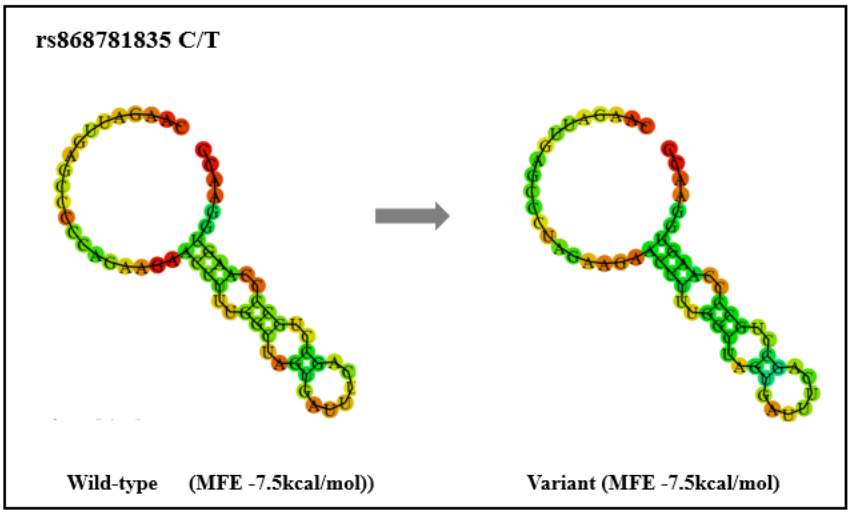



**Figure 4.24:** Minimum free energy of wild-type and variant P71L

## 4.2 Wet-lab Analysis

### *4.2.1 Genotype Analysis of Breast Cancer Patients and Control Samples*

The genotype analysis showed that the genotype TT has an odds ratio and relative risk of 40.16 and 6.143. These values indicate that these homozygous alleles could be related to breast cancer development and association when compared to homozygous GG or heterozygous GT. The homozygous genotype GG shows an odds ratio of 0.1856 and relative risk of 0.3293 which means that this might not be involved in the disease.

**Table 4.10** Genotyping data of C339G (rs1196572444)

| Genotype/ Alleles | Cancer % | Control % | Odds Ratio | Relative risk | P-value |
|---|---|---|---|---|---|
| **GG** | 3% | 14.29% | 0.1856 | 0.3293 | <0.005 |
| **GT** | 11% | 72.45% | 0.047 | 0.1748 | |
| **TT** | 86% | 13.27% | 40.16 | 6.143 | |
| **G** | 8% | 23.15% | 0.2887 | 0.4611 | <0.005 |
| **T** | 92% | 76.85% | 3.464 | 2.169 | |

### *4.2.2 Association of C339G SNP (rs1196572444) with Metastasis in Breast Cancer*

The genotype GG and GT have higher values than the TT genotype. However, the genotype GT has the highest value. Therefore, we can say that the patients possessing this genotype might have a chance of metastasizing.

**Table 4.11:** Statistical Analysis of Genotype relation with metastasis

| Genotype | Disease State | | | | | |
|---|---|---|---|---|---|---|
| | Metastatic | | | Non-metastatic | | |
| | P-value | Relative Risk | Odds Ratio | P-value | Relative Risk | Odds Ratio |
| GG | >0.9999 | 1.176 | 1.527 | >0.9999 | 0.7698 | 0.6548 |
| TT | 0.3832 | 0.7651 | 0.4821 | 0.3832 | 1.587 | 2.074 |
| GT | 0.3428 | 1.321 | 2.177 | 0.3428 | 0.6068 | 0.4594 |

# CHAPTER 5: DISCUSSION

The accumulation of mutations in genes that regulate cell cycle, apoptosis, invasiveness and metastasis leads to cell transformation in the form of malignancy. This happens because such genes act as caretaker genes for maintaining genomic stability (Listgarten et al., 2004). Breast cancer development occurs either due to modifiable or non-modifiable factors. Out of these, a non-modifiable factor is single-nucleotide polymorphism (SNP) (Łukasiewicz et al., 2021b). These SNPs are accountable for increasing cancer susceptibility (Hubner & Houlston, 2017).

The KLF family of proteins is the transcription factor family. These are either transcriptional activators or repressors and are zinc-finger proteins. There are a total of eighteen members of this family. KLF8 is one of them. This protein is involved in cell migration, proliferation, epithelial to mesenchymal transition, DNA repair, as well as resistance to drugs in cancer progression (Yuce & Ozkan, 2024). The aim of this study was to identify the most pathogenic variant of KLF8 and check its impact on overall protein structure and function. It further aimed at checking its link with breast cancer development and prognosis.

To check the localization of this protein, Deeploc 1.0 was utilized. Deeploc 1.0 uses deep learning to predict the subcellular localization of proteins found in eukaryotes. It is able to differentiate among ten different localizations. These include nucleus, cytoplasm, mitochondrion, cell membrane, endoplasmic reticulum, chloroplast, Golgi apparatus, lysosome or vacuole, peroxisome and extracellular. The predicted likelihood score turned out to be 0.998 in nucleus for KLF8. This indicates that this protein is found in the nucleus. The Clustal Omega software was used to find the evolutionary relationship between Kruppel-like Factor family members. KLF8 evolution was analyzed by aligning the sequences of all KPC family proteins to indicate conserved areas among the proteins in this family. The phylogenetic tree showed that all members of this family descended from a common ancestor and have some conserved regions present in all KLF isoforms. The conserved regions were visualized through ConSurf.

We searched the structure of KLF8 in Protein Data Bank (PDB). No structure was found so we used AlphaFold to predict its three-dimensional structure. This software applies scoring system to predict model confidence. While the 3D structure was predicted by AlphaFold, another software, Interpro was used to see the domains in KLF8. One N-terminal domain and three C2H2 domains were identified. according to AlphaFold, the three C2H2 zinc-finger domains showed high confidence score. The structure was validated using SAVES and Procheck analysis showed that majority of the angles lie in the most preferred regions pointing towards structure stability. This was depicted through a Ramachandran plot.

As this research study was aimed at checking the link between KLF8 SNPs and breast cancer, the SNP data was acquired from Ensembl Genome Browser. Then, the missense SNPs were filtered out and considered for further analysis as these have an impact on protein structure and sometimes, on its function as well. Further filtration was conducted to identify highly pathogenic variants. For this, the scores of various softwares were considered. These include SIFT, PolyPhen, Revel, Mutation Assessor, and MetaLR. Pathogenicity percentage of the variants was calculated by self-scoring the variants keeping benign and pathogenic on two extremes as 0 and 1 and a midpoint was set at 0.5. Then, a threshold value was selected as 85%. All these variants were shown in a graph pf pathogenicity percentage. Based on the threshold value, two highly pathogenic variants further analyzed. These had rs IDs, rs868781835(P71L) and rs1196572444(C339G).

Stability analysis of the selected variants was carried out using software tools such as MUpro, DynaMut and MaestroWeb. For the variant rs868781835(P71L), MUpro and MaestroWeb showed increasing stability while DynaMut showed decreasing stability. For the variant rs1196572444(C339G), MUpro and DynaMut showed decreasing stability while MaestroWeb showed increasing stability. The decreasing stability of protein indicates that the SNP might be associated with breast cancer.

For structural and functional analysis of selected SNPs, MutPred2, DynaMut and HOPE were used. These tools predict whether a change in protein structure may affect its function. MutPred2 gives a score for predicting the pathogenicity and has set 0.5 as a threshold for

pathogenicity. On the other hand, the value for our selected variants came out as 0.762 for P71L and 0.882 for C339G which renders them as pathogenic. Moreover for C339G, this showed an altered disordered interface and gain of N-linked glycosylation at N340 and for P71L, it showed loss of intrinsic disorder, loss of B-factor and loss of Ubiquitylation at K67. These mean that there is a decrease in stability of the mutant protein. DynaMut also predicted that the overall protein stability is decreasing due to the mutation. Moreover, the intra-atomic interactions were also altered as compared to the wild-type. Hope analysis revealed that the size of the mutant proteins differs from the wild-type. In case of SNP P71L the size was bigger with probably damaging effect, whereas for C339G, the size was smaller with damaging effect. These results suggest that a change in even a single amino acid can alter the protein structure and it may affect its function, leading to cancer development and disease progression.

The effect of such change on protein structure and function was also evaluated by observing the stability of mRNA secondary structure. in case of P71L, the change does not affect mRNA stability. Contrary to this, in SNP C339G, there is a change in stability. The minimum free energy for this structure changes from -10 kcal/mol to -12.6 kcal/mol. This decrease in MFE indicates an increase in stability of the structure.

As the in-silico analysis predicted that the polymorphism of KLF8 in SNP C339G might have a damaging role, leading to breast cancer development. For practically checking this effect, ARMS-PCR was conducted. Two groups, disease and control were made. Blood samples were collected from 100 subjects each and DNA was extracted. Then, two sets of primers, two inner and two outer were designed through Primer1. The obtained results were analyzed through statistical evaluation. According to these, the genotype GG might have a protective role in breast cancer development whereas the genotype TT might be involved in disease development as the odds ratio and relative risk turned out to be 40.16 and 6.143 for the genotype TT. The statistical analysis suggests that the genotype GG and GT might be involved in metastasis in breast cancer patients.

## *5.1 Conclusion*

Breast cancer is the most prevalent type of cancer in females worldwide and its incidence rate is also higher. There are a lot of ways in which the disease develops. The disease proves to be fatal as it is usually diagnosed at later stages. Early detection can lead to an increased chance of survival. However, biomarkers for detection and therapies might help in this aspect. Single nucleotide polymorphism is an important factor in disease development. In case of KLF8, little is known about its variants and their role in breast cancer.

In our study, the KLF8 SNPs were identified through in-silico analysis. Two of them were selected for structural and functional analysis through computational methods. One of these (rs868781835) turned out to be probably damaging while the other (rs1196572444) turned out to be damaging. This SNP was further chosen for genotype analysis. The results suggested that this was not associated with breast cancer development, but it was involved in disease progression through metastasis. The results might be different for different ethnicities.

This SNP can be used as a prognostic marker and a therapeutic target in breast cancer. The expression profile of this SNP calls is needed for further investigation as it might provide some cutting-edge findings in cancer therapy. Further research is required to validate the acquired results on a large sample size and different ethnicities. The results can also be applied to other cancer types to check the role of this variant of KLF8.

# REFERENCES


Aiqin Sun, & Jie Hao. (n.d.). *Regulation of Krüppel-like factor 8 by the NEDD4 E3 ubiquitin ligase*.

Barzaman, K., Karami, J., Zarei, Z., Hosseinzadeh, A., Kazemi, M. H., Moradi-Kalbolandi, S., … Farahmand, L. (2020). Breast cancer: Biology, biomarkers, and treatments. *International Immunopharmacology*, *84*, 106535. https://doi.org/10.1016/J.INTIMP.2020.106535

Bessenyei, B., M□rka, M., Urb□n, L., Zeher, M., & Semsei, I. (2004). Single nucleotide polymorphisms: aging and diseases. *Biogerontology*, *5*(5), 291–304. https://doi.org/10.1007/s10522-004-2567-y

Bialkowska, A. B., Yang, V. W., & Mallipattu, S. K. (2017). Krüppel-like factors in mammalian stem cells and development. *Development*, *144*(5), 737–754. https://doi.org/10.1242/dev.145441

Boland, M. R., Ryan, É. J., Dunne, E., Aherne, T. M., Bhatt, N. R., & Lowery, A. J. (2020). Meta-analysis of the impact of progesterone receptor status on oncological outcomes in oestrogen receptor-positive breast cancer. *British Journal of Surgery*, *107*(1), 33–43. https://doi.org/10.1002/bjs.11347

Brufsky, A. M., & Dickler, M. N. (2018). Estrogen Receptor-Positive Breast Cancer: Exploiting Signaling Pathways Implicated in Endocrine Resistance. *The Oncologist*, *23*(5), 528–539. https://doi.org/10.1634/theoncologist.2017-0423

Cokol, M., Nair, R., & Rost, B. (2000). Finding nuclear localization signals. *EMBO Reports*, *1*(5), 411–415. https://doi.org/10.1093/embo-reports/kvd092

Cotton, C. M., & Murray, J. C. (2018). The Human Genome and Neonatal Care. In *Avery's Diseases of the Newborn* (pp. 180-189.e2). Elsevier. https://doi.org/10.1016/B978-0-323-40139-5.00017-6

Deng, N., Zhou, H., Fan, H., & Yuan, Y. (2017). Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget*, *8*(66), 110635–110649. https://doi.org/10.18632/oncotarget.22372

Funnell, A. P. W., Norton, L. J., Mak, K. S., Burdach, J., Artuz, C. M., Twine, N. A., … Crossley, M. (2012). The CACCC-Binding Protein KLF3/BKLF Represses a Subset of KLF1/EKLF Target Genes and Is Required for Proper Erythroid Maturation *In Vivo*. *Molecular and Cellular Biology*, *32*(16), 3281–3292. https://doi.org/10.1128/MCB.00173-12

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, *144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

He, L., Wick, N., Germans, S. K., & Peng, Y. (2021). The Role of Breast Cancer Stem Cells in Chemoresistance and Metastasis in Triple-Negative Breast Cancer. *Cancers*, *13*(24), 6209. https://doi.org/10.3390/cancers13246209

He, Y., Liu, H., Chen, Q., Shao, Y., & Luo, S. (2019). Relationships between SNPs and prognosis of breast cancer and pathogenic mechanism. *Molecular Genetics & Genomic Medicine*, *7*(9). https://doi.org/10.1002/mgg3.871

Hsieh, P. N., Fan, L., Sweet, D. R., & Jain, M. K. (2019). The Krüppel-Like Factors and Control of Energy Homeostasis. *Endocrine Reviews*, *40*(1), 137–152. https://doi.org/10.1210/er.2018-00151

Hubner, R. A., & Houlston, R. S. (2017). Single Nucleotide Polymorphisms and Cancer Susceptibility. In *The Molecular Basis of Human Cancer* (pp. 231–239). New York, NY: Springer New York. https://doi.org/10.1007/978-1-59745-458-2_14

Jessie K. Cable, & Michael H. Grider. (n.d.). *physiology, progesterone*.

Jie Hao, & Heng Lu. (n.d.). *Role of krüppel-like factor 8 for therapeutic drug-resistant multi-organ metastasis of breast cancer*.

Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, *119*(6), 1420–1428. https://doi.org/10.1172/JCI39104

Karki, R., Pandya, D., Elston, R. C., & Ferlini, C. (2015). Defining "mutation" and "polymorphism" in the era of personal genomics. *BMC Medical Genomics*, *8*(1), 37. https://doi.org/10.1186/s12920-015-0115-z

Kemp W.L. (n.d.). *Neoplasia*.

Khan, N. H., Duan, S.-F., Wu, D.-D., & Ji, X.-Y. (2021). Better Reporting and Awareness Campaigns Needed for Breast Cancer in Pakistani Women. *Cancer Management and Research*, *Volume 13*, 2125–2129. https://doi.org/10.2147/CMAR.S270671

Krishnamurti, U., & Silverman, J. F. (2014). HER2 in Breast Cancer. *Advances in Anatomic Pathology*, *21*(2), 100–107. https://doi.org/10.1097/PAP.0000000000000015

Kumar, S., Behera, A., Saha, P., & Kumar Srivastava, A. (2021). The role of Krüppel-like factor 8 in cancer biology: Current research and its clinical relevance. *Biochemical Pharmacology*, *183*, 114351. https://doi.org/10.1016/j.bcp.2020.114351

Le Minh, G., Esquea, E. M., Dhameliya, T. T., Merzy, J., Lee, M.-H., Ball, L. E., & Reginato, M. J. (2023). Kruppel-like factor 8 regulates triple negative breast cancer stem cell-like activity. *Frontiers in Oncology*, *13*. https://doi.org/10.3389/fonc.2023.1141834

Li, T, Lu, H., Shen, C., Lahiri, S. K., Wason, M. S., Mukherjee, D., … Zhao, J. (2014). Identification of epithelial stromal interaction 1 as a novel effector downstream of Krüppel-like factor 8 in breast cancer invasion and metastasis. *Oncogene*, *33*(39), 4746–4755. https://doi.org/10.1038/onc.2013.415

Li, Tianshu, Lu, H., Mukherjee, D., Lahiri, S. K., Shen, C., Yu, L., & Zhao, J. (2015a). Identification of epidermal growth factor receptor and its inhibitory microRNA141 as novel targets of Krüppel-like factor 8 in breast cancer. *Oncotarget*, *6*(25), 21428–21442. https://doi.org/10.18632/oncotarget.4077

Li, Tianshu, Lu, H., Mukherjee, D., Lahiri, S. K., Shen, C., Yu, L., & Zhao, J. (2015b). Identification of epidermal growth factor receptor and its inhibitory microRNA141 as novel targets of Krüppel-like factor 8 in breast cancer. *Oncotarget*, *6*(25), 21428–21442. https://doi.org/10.18632/oncotarget.4077

Listgarten, J., Damaraju, S., Poulin, B., Cook, L., Dufour, J., Driga, A., … Zanke, B. (2004). Predictive Models for Breast Cancer Susceptibility from Multiple Single Nucleotide Polymorphisms. *Clinical Cancer Research*, *10*(8), 2725–2737. https://doi.org/10.1158/1078-0432.CCR-1115-03

Lomberk, G., & Urrutia, R. (2005). The family feud: turning off Sp1 by Sp1-like KLF proteins. *Biochemical Journal*, *392*(1), 1–11. https://doi.org/10.1042/BJ20051234

Lu, H, Hu, L., Yu, L., Wang, X., Urvalek, A. M., Li, T., … Zhao, J. (2014). KLF8 and FAK cooperatively enrich the active MMP14 on the cell surface required for the metastatic progression of breast cancer. *Oncogene*, *33*(22), 2909–2917. https://doi.org/10.1038/onc.2013.247

Lu, Heng, Hu, L., Li, T., Lahiri, S., Shen, C., Wason, M. S., … Zhao, J. (2012). A Novel Role of Krüppel-like Factor 8 in DNA Repair in Breast Cancer Cells. *Journal of Biological Chemistry*, *287*(52), 43720–43729. https://doi.org/10.1074/jbc.M112.418053

Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021a). Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers*, *13*(17), 4287. https://doi.org/10.3390/cancers13174287

Łukasiewicz, S., Czeczelewski, M., Forma, A., Baj, J., Sitarz, R., & Stanisławek, A. (2021b). Breast Cancer—Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies—An Updated Review. *Cancers*, *13*(17), 4287. https://doi.org/10.3390/cancers13174287

Ma, D., Liu, H., Qin, Y., Tian, Z., Li, S., & Liang, N. (2019). KLF8 overexpression promotes the growth of human lung cancer cells by promoting the expression of JMJD2A. *Cancer Cell International*, *19*(1), 258. https://doi.org/10.1186/s12935-019-0970-3

McConnell, B. B., & Yang, V. W. (2010). Mammalian Krüppel-Like Factors in Health and Diseases. *Physiological Reviews*, *90*(4), 1337–1381. https://doi.org/10.1152/physrev.00058.2009

Mehta, T. S., Lu, H., Wang, X., Urvalek, A. M., Nguyen, K.-H. H., Monzur, F., … Zhao, J. (2009). A unique sequence in the N-terminal regulatory region controls the nuclear localization of KLF8 by cooperating with the C-terminal zinc-fingers. *Cell Research*, *19*(9), 1098–1109. https://doi.org/10.1038/cr.2009.64

Menhas, R. , & U. S. (2015). (n.d.). *Breast Cancer among Pakistani Women*.

Mo, Y.-Y. (2012). Roles of microRNAs in cancer stem cells. *Frontiers in Bioscience*, *S4*(3), 301. https://doi.org/10.2741/s301

Mukherjee, D., Lu, H., Yu, L., He, C., Lahiri, S. K., Li, T., & Zhao, J. (2016). Krüppel-like factor 8 activates the transcription of C-X-C cytokine receptor type 4 to promote breast cancer cell invasion, transendothelial migration and metastasis. *Oncotarget*, *7*(17), 23552–23568. https://doi.org/10.18632/oncotarget.8083

Nature Education. (n.d.). *Gene Expression*.

Nicholson, G., Smith, A. V., Jónsson, F., Gústafsson, Ó., Stefánsson, K., & Donnelly, P. (2002). Assessing Population Differentiation and Isolation from Single-Nucleotide Polymorphism Data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *64*(4), 695–715. https://doi.org/10.1111/1467-9868.00357

Pollak, N. M., Hoffman, M., Goldberg, I. J., & Drosatos, K. (2018). Krüppel-Like Factors. *JACC: Basic to Translational Science*, *3*(1), 132–156. https://doi.org/10.1016/j.jacbts.2017.09.001

Robert, F., & Pelletier, J. (2018). Exploring the Impact of Single-Nucleotide Polymorphisms on Translation. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/fgene.2018.00507

Satadru K Lahiri, & Jihe Zhao. (n.d.). *Krüppel-like factor 8 emerges as an important regulator of cancer*.

*Stanford Medicine*. (n.d.).

Sudhakar, A. (2009). History of Cancer, Ancient and Modern Treatment Methods. *Journal of Cancer Science & Therapy*, *01*(02), i–iv. https://doi.org/10.4172/1948-5956.100000e2

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Suzuki, T., Aizawa, K., Matsumura, T., & Nagai, R. (2005). Vascular Implications of the Krüppel-Like Family of Transcription Factors. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *25*(6), 1135–1141. https://doi.org/10.1161/01.ATV.0000165656.65359.23

Upadhyay, A. (2021). Cancer: An unknown territory; rethinking before going ahead. *Genes & Diseases*, *8*(5), 655–661. https://doi.org/10.1016/j.gendis.2020.09.002

Vliet, J. v., Turner, J., & Crossley, M. (2000). Human Kruppel-like Factor 8: a CACCC-box binding protein that associates with CtBP and represses transcription. *Nucleic Acids Research*, *28*(9), 1955–1962. https://doi.org/10.1093/nar/28.9.1955

Wakim, S., & Mandeep Grewal. (n.d.). *Importance of Gene Expression*.

Wang, M.-D., Xing, H., Li, C., Liang, L., Wu, H., Xu, X.-F., … Yang, T. (2020). A novel role of Krüppel-like factor 8 as an apoptosis repressor in hepatocellular carcinoma. *Cancer Cell International*, *20*(1), 422. https://doi.org/10.1186/s12935-020-01513-3

Wang, X, Lu, H., Urvalek, A. M., Li, T., Yu, L., Lamar, J., … Zhao, J. (2011). KLF8 promotes human breast cancer cell invasion and metastasis by transcriptional activation of MMP9. *Oncogene*, *30*(16), 1901–1911. https://doi.org/10.1038/onc.2010.563

Wang, X, & Zhao, J. (2007). KLF8 transcription factor participates in oncogenic transformation. *Oncogene*, *26*(3), 456–461. https://doi.org/10.1038/sj.onc.1209796

Wang, Xianhui, Zheng, M., Liu, G., Xia, W., McKeown-Longo, P. J., Hung, M. C., & Zhao, J. (2007). Krüppel-Like Factor 8 Induces Epithelial to Mesenchymal Transition and Epithelial Cell Invasion. *Cancer Research*, *67*(15), 7184–7193. https://doi.org/10.1158/0008-5472.CAN-06-4729

WHO, 2020. (n.d.). *Global Cancer Observatory*.

WHO Globocan 2020. (n.d.). *Cancer Prevalence in Pakistan*.

Xianhui Wang, & Heng Lu. (n.d.-a). *Krüppel-like factor 8 promotes tumorigenic mammary stem cell induction by targeting miR-146a*.

Xianhui Wang, & Heng Lu. (n.d.-b). *Krüppel-like factor 8 promotes tumorigenic mammary stem cell induction by targeting miR-146a*.

Yi, R., Chen, B., Zhao, J., Zhan, X., Zhang, L., Liu, X., & Dong, Q. (2014). Krüppel-like Factor 8 Ameliorates Alzheimer's Disease by Activating β-Catenin. *Journal of Molecular Neuroscience*, *52*(2), 231–241. https://doi.org/10.1007/s12031-013-0131-4

Yuce, K., & Ozkan, A. I. (2024). The kruppel-like factor (KLF) family, diseases, and physiological events. *Gene*, *895*, 148027. https://doi.org/10.1016/j.gene.2023.148027

Zhang, J., Li, G., Feng, L., Lu, H., & Wang, X. (2020). Krüppel-like factors in breast cancer: Function, regulation and clinical relevance. *Biomedicine & Pharmacotherapy*, *123*, 109778. https://doi.org/https://doi.org/10.1016/j.biopha.2019.109778

ZHANG, M. H., MAN, H. T., ZHAO, X. D., DONG, N., & MA, S. L. (2014). Estrogen receptor-positive breast cancer molecular signatures and therapeutic potentials (Review). *Biomedical Reports*, *2*(1), 41–52. https://doi.org/10.3892/br.2013.187

turnitin

# Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

| | |
|---|---|
| Submission author: | Amna Ali |
| Assignment title: | Quick Submit |
| Submission title: | Oncogenic Role of KLF8 Variants in Breast Cancer |
| File name: | Plag_check_file-_Syeda_Amna_Ali_1.docx |
| File size: | 382.72K |
| Page count: | 26 |
| Word count: | 8,709 |
| Character count: | 46,969 |
| Submission date: | 04-Jul-2024 10:51PM (UTC-0700) |
| Submission ID: | 2412721526 |

Oncogenic Role of KLF8 Variants in Breast Cancer

By

Syeda Amna Ali

(Registration No: [illegible])

A thesis submitted to the National University of Sciences and Technology, Islamabad,

in partial fulfillment of the requirements for the degree of

Master of Science in

Healthcare Biotechnology

Supervisor: Dr. Maria Shabbir

Co Supervisor: Dr. Yasmin Bakhsh

Atta Ur Rahman School of Applied Biosciences

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

Dr. Maria Shabbir
Tenured Associate Professor
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Dr. Maria Shabbir
Tenured Associate Professor
Deptt of Health Care Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

# Oncogenic Role of KLF8 Variants in Breast Cancer

ORIGINALITY REPORT

| 14% | 8% | 6% | 6% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | Submitted to Higher Education Commission Pakistan<br>Student Paper | 3% |
| 2 | hdl.handle.net<br>Internet Source | 1% |
| 3 | Li, T, H Lu, C Shen, S K Lahiri, M S Wason, D Mukherjee, L Yu, and J Zhao. "Identification of epithelial stromal interaction 1 as a novel effector downstream of Krüppel-like factor 8 in breast cancer invasion and metastasis", Oncogene, 2013.<br>Publication | <1% |
| 4 | Kanupriya Jha, Amit Kumar, Kartik Bhatnagar, Anupam Patra, Neel Sarovar Bhavesh, Bipin Singh, Sarika Chaudhary. "Modulation of Krüppel-like factors (KLFs) interaction with their binding partners in cancers through acetylation and phosphorylation", Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2024<br>Publication | <1% |

| | | |
|---|---|---|
| 5 | Lu, H, L Hu, L Yu, X Wang, A M Urvalek, T Li, C Shen, D Mukherjee, S K Lahiri, M S Wason, and J Zhao. "KLF8 and FAK cooperatively enrich the active MMP14 on the cell surface required for the metastatic progression of breast cancer", Oncogene, 2013.<br>Publication | <1 % |
| 6 | www.science.gov<br>Internet Source | <1 % |
| 7 | nust.edu.pk<br>Internet Source | <1 % |
| 8 | M. Yazdi, M. Kolahi, E. Mohajel Kazemi, A. Goldson Barnaby. "Study of the contamination rate and change in growth features of lettuce (Lactuca sativa Linn.) in response to cadmium and a survey of its phytochelatin synthase gene", Ecotoxicology and Environmental Safety, 2019<br>Publication | <1 % |
| 9 | life.futuregenerali.in<br>Internet Source | <1 % |
| 10 | www.biorxiv.org<br>Internet Source | <1 % |
| 11 | www.ncbi.nlm.nih.gov<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 12 | Submitted to Imperial College of Science, Technology and Medicine<br>Student Paper | <1 % |
| 13 | network.bepress.com<br>Internet Source | <1 % |
| 14 | Submitted to University of Nebraska, Lincoln<br>Student Paper | <1 % |
| 15 | www.frontiersin.org<br>Internet Source | <1 % |
| 16 | www.ir.juit.ac.in:8080<br>Internet Source | <1 % |
| 17 | Agnieszka B. Bialkowska, Vincent W. Yang, Sandeep K. Mallipattu. "Krüppel-like factors in mammalian stem cells and development", Development, 2017<br>Publication | <1 % |
| 18 | Submitted to Grand Canyon University<br>Student Paper | <1 % |
| 19 | physrev.physiology.org<br>Internet Source | <1 % |
| 20 | Submitted to Murray State University<br>Student Paper | <1 % |
| 21 | Ping He, Vincent W. Yang. "The role of Krüppel-like factors in generating induced pluripotent stem cells", Elsevier BV, 2022<br>Publication | <1 % |

| | | |
|---|---|---|
| 22 | www.hindawi.com<br>Internet Source | <1 % |
| 23 | Submitted to Roosevelt University<br>Student Paper | <1 % |
| 24 | prr.hec.gov.pk<br>Internet Source | <1 % |
| 25 | Göransson, Sara. "Cell-Matrix Interactions: Master Regulators of Cancer Cell Fate?", Karolinska Institutet (Sweden), 2023<br>Publication | <1 % |
| 26 | "Abstracts from the 56th European Society of Human Genetics (ESHG) Conference: Hybrid Posters", European Journal of Human Genetics, 2024<br>Publication | <1 % |
| 27 | Siti Maulida, Kartini Eriani, Nur Fadli, Mohd N. Siti-Azizah et al. "Effect of type and concentration of antioxidant on sperm motility, viability, and DNA integrity of climbing perch Anabas testudineus Bloch, 1792 (Pisces: Anabantidae) post-cryopreservation", Cryobiology, 2024<br>Publication | <1 % |
| 28 | academic.oup.com<br>Internet Source | <1 % |

| | | |
|---|---|---|
| 29 | Min Zhang, Jian Huang, Xing Xie, C. D'Arcy J. Holman. "Dietary intakes of mushrooms and green tea combine to reduce the risk of breast cancer in Chinese women", International Journal of Cancer, 2008<br>Publication | <1 % |
| 30 | Submitted to Bennett University<br>Student Paper | <1 % |
| 31 | Cigdem Sevim Bayrak, Aayushee Jain, David Stein, Kumardeep Chaudhary et al. "Identification of Discriminative Gene-level and Protein-level Features Associated with Gain-of-Function and Loss-of-Function Mutations", Cold Spring Harbor Laboratory, 2021<br>Publication | <1 % |
| 32 | Submitted to De La Salle University<br>Student Paper | <1 % |
| 33 | Submitted to University of Westminster<br>Student Paper | <1 % |
| 34 | Submitted to Washington State University System<br>Student Paper | <1 % |
| 35 | era.ed.ac.uk<br>Internet Source | <1 % |
| 36 | mafiadoc.com<br>Internet Source | |

| | | |
|---|---|---|
| | | <1 % |
| 37 | s3.amazonaws.com<br>Internet Source | <1 % |
| 38 | www.nature.com<br>Internet Source | <1 % |
| 39 | core.ac.uk<br>Internet Source | <1 % |
| 40 | doaj.org<br>Internet Source | <1 % |
| 41 | edoc.hu-berlin.de<br>Internet Source | <1 % |
| 42 | ethesis.nitrkl.ac.in<br>Internet Source | <1 % |
| 43 | www.spandidos-publications.com<br>Internet Source | <1 % |
| 44 | Hussain, Tabinda, Sobia Manzoor, Yasir Waheed, Huma Tariq, and Khushbakht Hanif. "Phylogenetic analysis of torque teno virus genome from Pakistani isolate and incidence of co-infection among HBV/HCV infected patients", Virology Journal, 2012.<br>Publication | <1 % |
| 45 | Submitted to University of Central Florida<br>Student Paper | <1 % |

| | | |
|---|---|---|
| 46 | Yanxing Niu, Anna Rogiewicz, Lan Shi, Rob Patterson, Bogdan A. Slominski. "The effect of enzymatically-modified canola meal on growth performance, nutrient utilization, and gut health and function of broiler chickens", Animal Feed Science and Technology, 2023<br>Publication | <1 % |
| 47 | dr.ntu.edu.sg<br>Internet Source | <1 % |
| 48 | www.mb.com.ph<br>Internet Source | <1 % |
| 49 | www.scielo.br<br>Internet Source | <1 % |
| 50 | Nina M. Pollak, Matthew Hoffman, Ira J. Goldberg, Konstantinos Drosatos. "Krüppel-Like Factors", JACC: Basic to Translational Science, 2018<br>Publication | <1 % |
| 51 | Vinicius Matheus Knihs, Fabíola Branco Filippin-Monteiro. "GLP1R (glucagon-like-peptide-1 incretin receptor), diabetes and obesity phenotypes: An in silico approach revealed new pathogenic variants", Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2024<br>Publication | <1 % |
| | cris.brighton.ac.uk | |

| | | |
|---|---|---|
| 52 | Internet Source | <1 % |
| 53 | docplayer.net<br>Internet Source | <1 % |
| 54 | pdffox.com<br>Internet Source | <1 % |
| 55 | www.mdpi.com<br>Internet Source | <1 % |
| 56 | Alessandra Picollo. "Vesicular CLC chloride/proton exchangers in health and diseases", Frontiers in Pharmacology, 2023<br>Publication | <1 % |
| 57 | Xianhui Wang, Heng Lu, Tianshu Li, Lin Yu, Gang Liu, Xu Peng, Jihe Zhao. "Abstract A100: KLF8 induces breast stemness and chemoresistance via miRNAs associated with EMT.", Molecular Cancer Therapeutics, 2013<br>Publication | <1 % |
| 58 | Dean A. Jackson. "Cell Nucleus Biogenesis, Structure and Function", Wiley, 2006<br>Publication | <1 % |
| 59 | Nguyen Quoc Khanh Le, Duyen Thi Do, Trinh-Trung-Duong Nguyen, Quynh Anh Le. "A sequence-based prediction of Kruppel-like factors proteins using XGBoost and optimized features", Gene, 2021<br>Publication | <1 % |

| | | |
|---|---|---|
| 60 | www.zora.uzh.ch<br>Internet Source | <1 % |