Evaluating the role of KLF17 variant as a key regulator in Pathogenesis of Breast Cancer

By

Asma Areesh

Registration No: 00000401316

Department of Biomedicine

Atta-ur-Rahman School of Applied Biosciences

National University of Sciences and Technology

Islamabad, Pakistan

2024

Evaluating the Role of KLF17 variant as a key regulator in pathogenesis of Breast Cancer

By

Asma Areesh

Registration No: 00000401316

A thesis submitted to the NUST in partial fulfillment of the requirements for the degree of MS in Healthcare Biotechnology

Supervisor: **Dr. Maria Shabbir**

Co Supervisor: **Dr. Yasmin Badshah**

Atta-ur-Rahman School of Applied Biosciences

National University of Sciences and Technology

Islamabad, Pakistan

2024

FORM TH-4

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: **Asma Areesh**
Reg No. **00000401316** Titled: **Evaluating the Role of KLF17 Variant as a Key Regulator in Breast Cancer Pathogenesis** be accepted in partial fulfillment of the requirements for the award of MS Degree in Healthcare biotechnology degree with (__A__ grade).
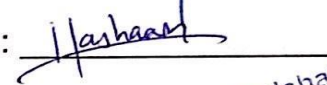
Examination Committee Members

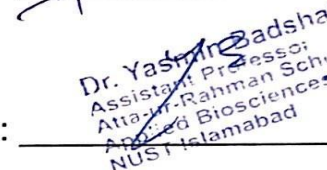1.   Name: <u>Dr. Peter John</u>   Signature: _____

2.   Name: <u>Dr. Rumeza Hanif</u>   Signature: _____

3.   Name: <u>Dr. Hashaam Akhtar</u>   Signature: _____

Co-Supervisor's name: Dr. Yasmin Badshah   Signature: _____

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

Supervisor's name: Dr. Maria
Shabbir

Dr. Maria Shabbir
Tenured Associate Professor
Atta-ur-Rahman School
Biosciences (ASAB), NUST Isla.

Signature: _____

Date: __09 - 08 -24__

Date: __13/08/2024__

Dr. Hussain Mustatab Wahedi
Head of Department (HoD)
Dept of Biotechnology
Deptt of ...

Head of Department

COUNTERSINGED

A/Principal & Dean
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST, Islamabad
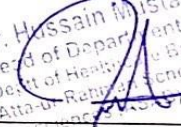
Date: __15/08/24__

Dean/Principal

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by Ms. **Asma Areesh** Registration No. 00000401316), 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 point out by GEC members and evaluators of the scholar have also been incorporated in the said thesis.
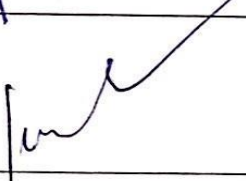
Signature: _____

Name of Supervisor Dr. Maria Shabbir.

Date: 09-08-24

Signature (HOD): _____

Date: 13/08/2024

Signature (Dean/ Principal) _____

Date: 15/08/24

CERTIFICATE FOR PLAGIARISM

It is to confirm that MS thesis entitled titled "**Evaluating the Role of KLF17 Variant as a Key Regulator in Pathogenesis of Breast Cancer**" of **Asma Areesh** Reg No. 00000401316 has been examined by me. I undertake that,

1. Thesis has significant new work/knowledge as compared to already elsewhere. No sentence, table, equation, diagram, paragraph, or section has been copied verbatim from previous work except when placed under quotation marks and duly referenced.

2. The work presented is original and own work of the author i.e., there is no plagiarism. No idea, results or work of others have been presented as author's own work.

3. There is no fabrication of data or results such that the research is not accurately represented in the records. The thesis has been checked using Turnitin, a copy of the original report attached and focused within the limits as per HEC plagiarism policy and instruction based on time to time.

(Supervisor)

Dr. Maria Shabbir

Biomedicine

ASAB, NUST

COPYRIGHT STATEMENT

DECLARATION

I **"ASMA AREESH"** hereby state that my MS thesis titled "**Evaluating the Role of KLF17 variant as a key regulator in Breast Cancer Pathogenesis**" 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:_____ASMA AREESH_____

Date: _____9-Aug-2024_____

DEDICATION

I dedicate this thesis to my beloved Parents and Siblings for their

unwavering support throughout this journey.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page No.

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

ARMS PCR   **Amplification- Refractory Mutation System**

CADD   **Combined Annotation Dependent Depletion**

DNA   **Deoxyribonucleic Acid**

EDTA   **Ethylenediaminetetraacetic Acid**

EMT   **Epithelial to Mesenchymal transition**

HOPE   **Have Our Protein Explained**

ID-1   **inhibitor of differentiation**

KLF   **Krüppel Like Factor**

MD   **Molecular Dynamics**

OR   **Odd Ratio**

PTC   **Papillary thyroid carcinoma (PTC**

RMSD   **Root Mean Square Deviation**

RMSF   **Root Mean Square Fluctuation**

SASA   **Solvent accessible surface area**

SDS   **Sodium Dodecyl Sulphate**

SNP   **Single Nucleotide Polymorphism**

TAE   **Tris Acetate EDTA**

TE   **Tris EDTA**

UV   **Ultraviolet**

ABSTRACT

Breast cancer is one of the most prevalent cancer among women globally with higher metastatic potential and recurrence rate leading to greater mortality rates in females of Pakistan which demands for a promising prognostic biomarker of breast cancer to make early detection possible. KLF17 is a transcription factor and has been observed to play important role as tumor suppressor specifically inhibiting EMT but protein has received little consideration in perspective of Single Nucleotide Polymorphism. Therefore, the aim of this study was to analyze the pathogenic variants of KLF17 and its association with breast. The SNPs in KLF17 effect both the structure and function of KLF17 was investigated through in-silico analysis and the results observed were validated through the blood-based genotypic analysis via ARMS PCR. A positive association of C345R SNP was observed with pathogenesis of breast cancer, suggesting that rs183242786 could be used as a prognostic biomarker of Breast Cancer.

Keywords: Breast cancer, KLF17, EMT, SNPs, Biomarker, ARMS PCR.

CHAPTER 1: INTRODUCTION

A SNP is a variation in DNA sequence at single position that results due to a change in nucleotide base at one given position. If SNP occurs at gene, then gene is said to have multiple alleles. In the human genome, SNPs are the most prevalent kinds of genetic changes.

SNPs act as a label for specific regions of DNA and so they can be examined to check for any variations in those particular areas. Although a SNP might not be called as a lethal mutation and may not lead to a disorder yet some of the SNP may be deleterious and show association with disease. Such SNP can last a drastic impact on the genetic makeup as well as structure and stability of Protein. ("SNP," 2014). Genetic liability to cancer is allied to SNPs in genes that regulates DNA mismatch repair, cell cycle regulation, metabolism, and immunology. It is necessary to Grasp the molecular etiology of cancers necessitates an expertise of the processes that underlie the impact of SNPs contributing to cancer susceptibility. SNPs have the potential to be both therapeutic and diagnostic biomarkers for a variety of cancer forms from a clinical standpoint (Deng et al., 2017). in addition to this, SNPs could expediate the development of personalized medicine by identifying the variations responsible for altered response to certain drugs (Britannica., 2024, May 27).

1.1 CANCER

Cancer is a broad category of disorders that can instigate in nearly any organ or tissue in the body when aberrant cells proliferate out of control, cross normal boundaries to infect other body parts, or spread to the other organs. The metastasizing phase is a primary contributor to cancer-related deaths. Other frequent terms for cancer include neoplasm and malignant tumor ("CANCER,").

Benign tumors usually do not penetrate or spread to neighboring tissues neither do they exhibit recurrency. On the other hand, malignant tumors occasionally do. However, in some cases benign tumors can grow to be rather enormous. Some of the benign brain tumors prove to be potentially fatal or result in severe symptoms.

Approximately 9.6 million mortalities, or 1 in 6 deaths, were attributed to cancer in 2018, making it the second most common cause of death worldwide. Men are most likely to develop lung, prostate, colorectal, stomach, and liver cancers, whereas women are more likely to develop breast, colorectal, lung, cervical, and thyroid cancers. ("CANCER,")

1.2 BREAST CANCER

The term "breast cancer" describes tumors that originate in the breast tissue, usually from the lobules that supply the milk ducts or the inner lining of the ducts. As most cancers are named after the part of the body they first appeared in, breast cancer describes the irregular growth and division of cells that start in the breast tissue. The glandular tissues and the stromal (supporting) tissues are the two primary tissue types that constitute the breast. The lobules and the ducts are contained in the glandular tissues, whereas stromal tissues

constitutes the fatty and fibrous connective tissues of the breast ("what is breast cancer ", 2010).

Tumors are classified into different types depending upon their development in different areas of breast. Most of these tumors are the result of changes in non-cancerous cells. A number of breast cancers initiate from duct lining cells while some arise in lobular lining cells.

Breast cancer is further classified into different types on the basis of their ability of invade and the region of origin. Non-invasive breast cancer as the name indicates do not invade and metastasize to the other tissues such as fatty and connective tissues of the breast.

Lobular carcinoma in situ and Ductal carcinoma in situ, in situ in name of the cancer types indicates its inability to invade and spread to the other organs of the body where in lobular carcinoma in situ there is increase in abnormal cells in milk gland of breast while DCIS originates in region of breast ductal comedocarcinoma and stay restricted to it.("Types of breast cancer,")

The invasive Cancer type include Invasive Lobular Carcinoma originating from milk glands and spreading to other areas of the body while Infiltrating Ductal Carcinoma another invasive cancer instigates from milk ducts of the breast metastasizing to the other parts of the body.(Sharma et al., 2010).

Breast cancer is a highly heterogeneous disease both at molecular and clinical levels with a diverse range of its subtypes that are barely studied in detail and a very less data is available about them creating hurdle in development of biomarkers, moreover, exact cause of breast cancer is still un clear while some of the risk factors associated with breast cancer have been identified such as Genetics, hormonal influence and standard of living, till date (Koren & Bentires-Alj, 2015). Another limitation of breast cancer is its early detection and diagnosis, hence leading to metastasis, that decreases the effectivity of treatment as breast cancer has highest metastatic potential. Late diagnosis and higher metastatic potential of breast cancer keeps its treatment a challenge. KLF17 has been reported as a tumor suppressor in various cancer including breast cancer where it indirectly inhibits EMT and so metastasis. The genetic polymorphism involved in KLF17 reported effect on the structure of Protein and may influence its function as well. Missense SNPs that occurred in KLF17 may increase the likelihood of Breast cancer development and progression. KLF17 is a member of KLFs, a family of transcription factor comprising of 18 members. KLF 17 located at 1p34.1 on chromosome (Wang et al., 2020). KLF17 has mostly been researched in the perspective of cancer, where it has been identified as a tumor suppressor through its interaction with TGF/SMAD signaling and p53, as well as its prevention of epithelial-to-mesenchymal transition (Ali, Zhang, et al., 2015). Further research revealed low expression of KLF17 levels promoting tumor cell invasion, migration, and EMT shift. KLF17 was reported to be involved in Lung Adenocarcinoma hepatocellular (HCC), Gastric Cancer (GC), Papillary thyroid carcinoma (PTC) (Ye et al., 2014), and Non-Small Cell Lung Cancer (NSCL). However, no clear studies on direct tumor suppression via

KLF17 have been demonstrated yet. Additionally, Genetic link between KLF17 SNP and breast cancer status is yet to be discovered. Hence the main objective of this research is

— To identify possible pathogenic SNPs of KLF17 and analyze its effect on stability of protein and validate the result through blood based genotypic analysis leading to development of KLF17 as a therapeutic target or and a prognostic Biomarker of breast cancer as well.

CHAPTER 2: LITERATURE REVIEW

2.1   What is Cancer

Cancer is an uncontrolled abnormal growth of normal cells developing into tumor masses. Not all abnormal cells are cancerous cells, some cells develop into benign that do not penetrate to the neighboring tissues neither do they revert back after removal; however, the malignant tumors metastasize to other parts of the body and show a high rate of recurrency after removal. Nevertheless, in some cases benign cells may also grow to be quite enormous becoming lethal *("CANCER," 2022).*

In 2020, Almost 18,094,716 million cases of cancer were reported across the Globe. The number of reported cases in men were higher (206.9 per 100,000) than in the women (178.1 per 100,000)("Global cancer data by country ", 2020). In 2020, Nearly 10 million people died from cancer worldwide, making it the top cause of death ("CANCER," 2020).

2.2   Breast Cancer

Breast cancer is the most frequent cancer in the world, overtaking lung cancer for the first time in 2020 with 2.3 million female diagnoses and 685,000 fatalities worldwide ("Breast Cancer," 2023). Every 14 seconds, a woman is diagnosed with breast cancer diagnosis across the globe.("Breast Cancer Statistics Worldwide," 2023).

Prevalence of Breast Cancer in Pakistan

Among Asian countries Pakistan is at the top of breast cancer cases reported, with one out of every nine women at risk of being diagnosed with breast cancer. (Zaheer et al., 2019).

In 2020, 178,388 new cases of breast cancer were reported and the rate of incidence of breast cancer is expected to rise by approximately 23.1 percent in 2020 to 60.7 percent in 2025, with the increase in cases of about 2300 in 5 years. (Zaheer et al., 2019)
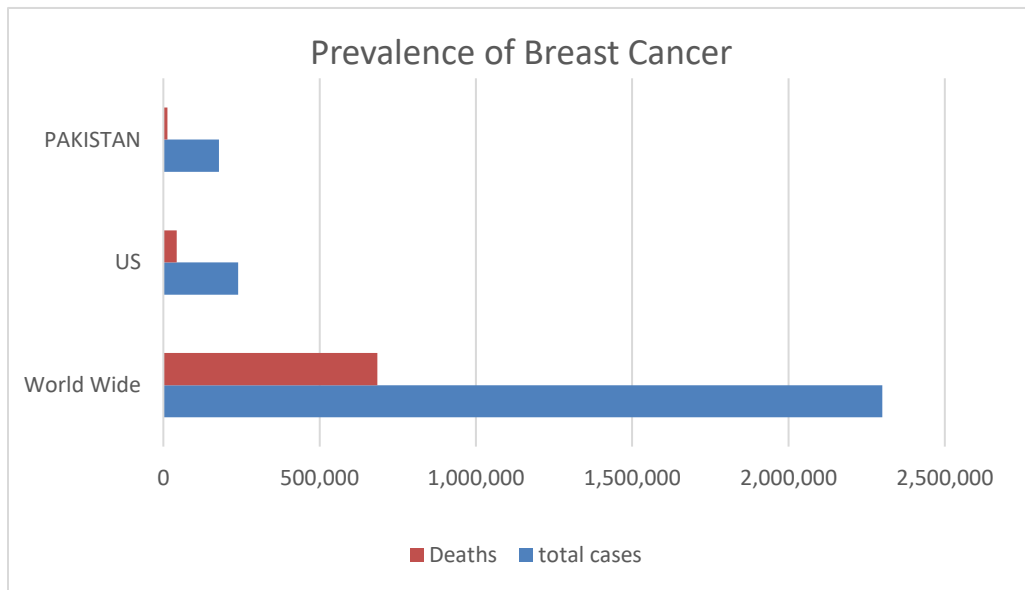
Figure 2.1 Prevalence of Breast Cancer in Pakistan and US In Comparison to Worldwide

2.2.1 Introduction to Breast Cancer

The term "breast cancer" describes tumors that originate in the breast tissue, usually from the lobules that supply the milk ducts or the inner lining of the ducts. As most cancers are named after the part of the body they first appeared in, breast cancer describes the irregular growth and division of cells that start in the breast tissue. The glandular tissues and the

stromal (supporting) tissues are the two primary tissue types that constitute the breast. The lobules and the ducts are contained in the glandular tissues, whereas stromal tissues constitutes the fatty and fibrous connective tissues of the breast ("what is breast cancer ", 2010) (Sharma et al., 2010).

2.2.2. Types of Breast Cancers

Tumors are classified into different types depending upon their development in different areas of breast. Most of these tumors are the result of changes in non-cancerous cells. A number of breast cancers initiate from duct lining cells while some arise in lobular lining cells.("What is Breast Cancer?," 2009).

Breast cancer is further classified into different types on the basis of their ability to invade and the region of origin. Non-invasive breast cancer as the name indicates do not invade and metastasize to the other tissues such as fatty and connective tissues of the breast.

Lobular carcinoma in situ and Ductal carcinoma in situ, in situ in name of the cancer types indicates its inability to invade and spread to the other organs of the body where, in lobular carcinoma in situ, there is increase in abnormal cells in milk gland of breast while DCIS originates in region of breast ductal comedocarcinoma and stay restricted to it.("Types of breast cancer,")

The invasive Cancer type include **Invasive Lobular Carcinoma or infiltrating lobular carcinoma,** accounting for 80 percent of all the cases originates from milk glands and

spreading to other areas of the body ("INVASIVE LOBULAR CANCER (ILC)," 2023) while **Infiltrating Ductal Carcinoma** another invasive cancer which is the most recurrent kind of breast cancer accounts for almost 80 percent of the cases. It instigates from milk ducts of the breast metastasizing to the other parts of the body by intruding the breast fatty tissue.(Sharma et al., 2010). ("INVASIVE DUCATL CANCER ").

The most frequent type of non-invasive breast cancer is ductal carcinoma in situ (DCIS), which accounts for 90 percent of all cases. Lobular carcinoma in situ (LCIS) is less prevalent and is thought to be a risk factor for breast cancer. Breast cancer cells that infect the surrounding fatty and connective tissues after breaking through the duct and lobular walls. cancer can be invasive without spreading to the lymph nodes or other organs (Lydia Choi Sep 2022).

Triple Negative Breast Cancer

In Triple-negative breast cancer there is lack all of the receptors witnessed in other types of breast cancer.

The three receptors are

— feminine hormone estrogen.

— progesterone, another female hormone and

— Human epidermal growth factor (HER2) ("Triple-Negative Breast Cancer," 2023)

2.3. KLF Family

KLFs are the family of transcriptional factors involved in tumor initiation, progression and elevation through transcriptional activation and repression. KLFs holds great importance in certain vital mechanisms such as cell proliferation, differentiation, cell death and metastatic property of cell (Jianping et al., 2020).

KLF family constitutes of 18 members and their expression patterns vary widely amongst different tissues. KLFs have been categorized into 3 main classes based on their structural characteristics and transcriptional activities. KLF3, KLF8, and KLF12, which interact with transcriptional repressors, are members of Group 1. Group 2 consists of the transcriptional activators KLF1, KLF2, KLF4, KLF5, KLF6, and KLF7, however they may also show interactions as transcriptional repressors have also been noted. KLF9, KLF10, KLF11, KLF13, KLF14, and KLF16 are the members of group 3, which have mostly been classified as transcriptional repressors. Because little is known about their protein interaction motifs, KLF15 and KLF17 have not been categorized in any of these groupings. (Kotlyarov & Kotlyarova, 2023). KLF isoforms are also involved in controlling energy homeostasis and metabolic pathways in a variety of organs, including the liver, adipose tissue, heart, skeletal muscle, lungs, and myeloid cells (Pollak et al., 2018).

KLF family is comprised of characteristic KLF like zinc fingers that have specific binding site CACCC elements and GC-rich regions of DNA at C terminal that play important role in activation and repression of transcription. Although the C-terminal domain of the KLFs

contains primarily nuclear localization signals and the DNA-binding region, the N-terminal domain bears regions that interact with other proteins. KLFs have also been linked to the development and upholding of pluripotency (Tetreault et al., 2013).

(Kin-Sang et al., 2015)
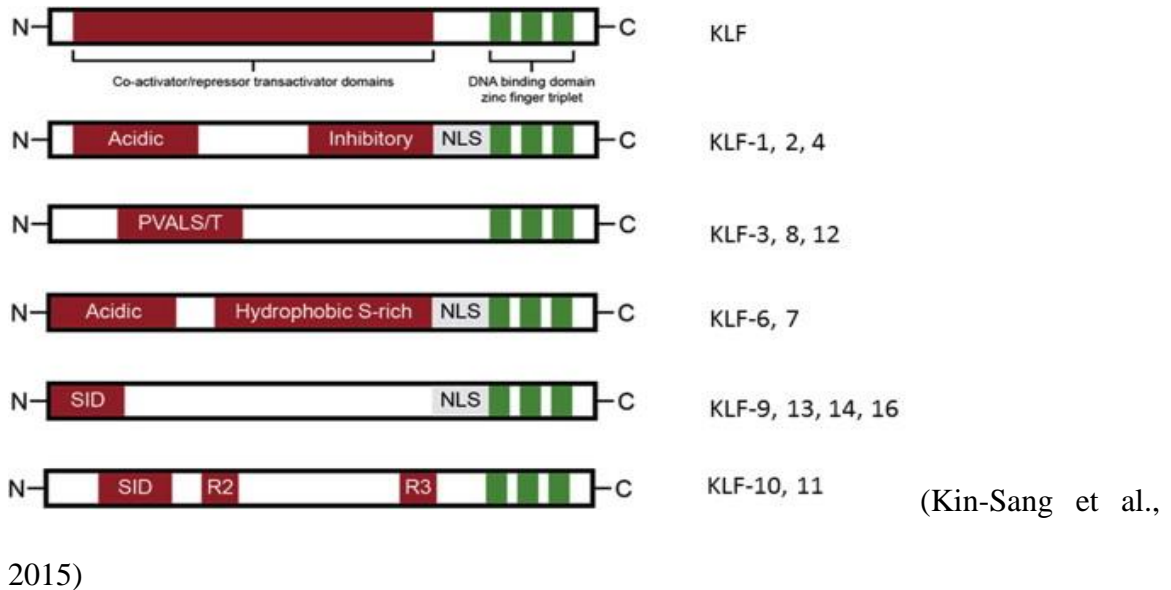
Figure 2.2: Structural domains of KLF family members

Most of the KLFs are ubiquitously expressed (such as KLFs 6, 10, and 11), but others are determined to be developmentally or temporally expressed in tissue- and cell-type. The tissue expression of the KLFs varies. KLF1 is expressed in the bone marrow in a particular way, KLF2 is very substantially expressed in adipose tissue and KLFs 4, 5 are plentiful in the digestive system. Apart from the aforementioned DNA binding domains, the structure also contains transcriptional regulation domains and nuclear localization signal (NLS) domains, though NLS is only found in a small number of KLFs (KLF 1, 4, 8, and 11),

which are found near or inside Zinc-finger motifs and serve as nuclear localization signals for KLF (Jianping et al., 2020).

2.4 KLF Maturation and Activation

The N-terminal regions of KLFs are primarily responsible for the selectivity of their transcriptional activation (Bialkowska et al., 2017).

KLFs and the proteins they interact with undergo post-translational changes that change their transcriptional activity or subcellular localization. Functional commonality exists across KLFs with structural homology, notably in the N-terminal domains (Li et al., 2022). Several acetyl-transferases, including as p300, p300/CBP-associated factor 7, 8, 9, 10, and 11, bind to a variety of KLF members from group 2, including KLF1, KLF2, KLF4, KLF5, and KLF6, as well as KLF13 from group 3. As a result of this interaction, KLFs are acetylated, which promotes their transcriptional activity, or histones are acetylated, which causes chromatin remodeling and activates transcription in KLF-targeted areas. Likewise, KLFs undergo deacetylation via the interaction with histone deacetylases (HDACs), which decreases their transcriptional activity. Interaction with the transcriptional repressors Sin3A , C-terminal binding protein (CtBP)1, and CtBP2 may lead transcriptional suppressions of KLFs through recruitment of HDACs , methyl transferases , and other silencing complexes like polycomb proteins and Ikaros. (Hsieh et al., 2019).
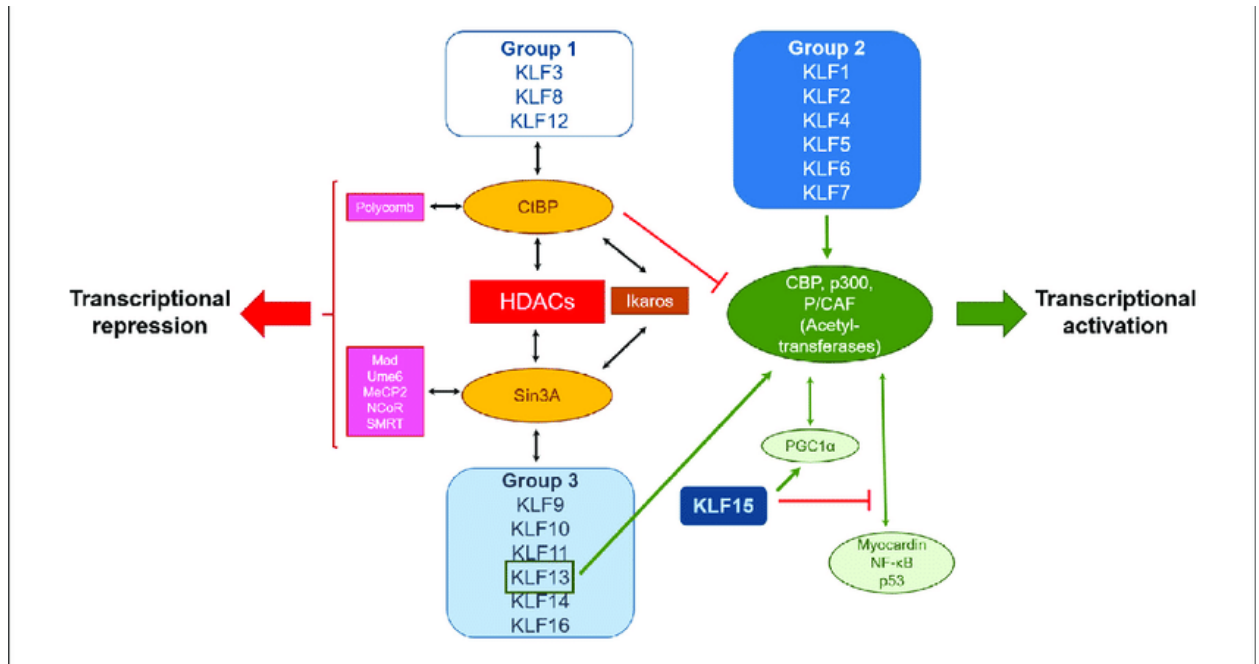
Figure 2.3 Activation and deactivation of KLF family

The post-translational modifications that KLFs exhibit (phosphorylation, methylation, SUMOlyation, and ubiquitination), as well as their interactions with co-activators or co-repressors, which add additional layers of specificity and governance over these factors, are also important differences (Bialkowska et al., 2017).

2.5 Structure of KLF Family

The existence of three highly conventional Cys_2/His_2 zinc fingers contribute to the unique characteristics of the KLF family. There are 23 residues in zinc finger 1 and 2 whereas the 3rd finger domain comprises only 21 residues the location of these fingers is the carboxyl terminus of the protein that aids in the binding of KLF to their specific CACCC and GC

boxes of DNA. Furthermore, these fingers are connected the specific and unique Krüppel-link, which is a highly conserved spacer region comprising of seven amino acids TGEKP(Y/F) X. each of the zinc finger combines with a single zinc ion that are further attached to four residues, two cysteine and two histidine. The regions responsible for determining sequence specificity are substantially conserved leading to sequence specificity (Richard et al., 2008) (Zhang et al., 2023) (McConnell & Yang, 2010).

2.6. KLF17: A unique member of KLF family

2.6.1. Discovery and Structure

KLF17 is a transcription factor and is also known as Zinc Finger 393. KLF 17 is human homologue of murine ZFP393 and is located on chromosome 1 at 1p34.1. The short arm of human chromosome is the most studied chromosome in human genome. Allelic deletions in the regions of 1p36 and 1p32 show strong association with poor survival (Wang et al., 2020). KLF17 is an affiliate of Sp/KLF family of transcription factors and in various studied it was observed to have closer association with KLFs belonging to the Sp1 C2H2-type zinc-finger protein family. KLF17 consists of three zinc finger domains (Shan Zhou et al., 2016). And its transcripts comprise of 3160 bps with 4 exons, out of which 3 exons encode amino acids with 27 domains. KLF 17 is comprised of 1170nt open reading frame encoding a total of 389 amino acids along with 3 c-terminal zinc fingers being the distinctive feature of KLF family (Jane et al., 2006).

The zinc fingers of KLF17 are hardly 60-70 percent identical to the other KLFs with quite a low level of similarity.
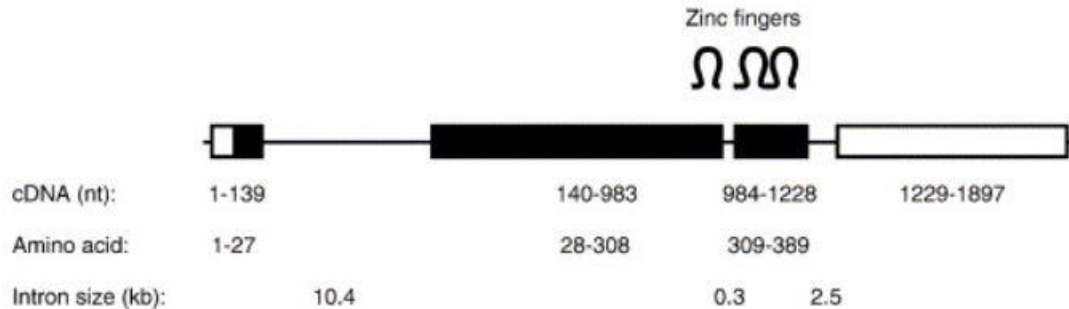
(Jane et al., 2006)

Figure 2.4: structural domains of KLF17

Figure depicts the human gene encoding KLF17. (Jane et al., 2006), (Abe et al., 2022).

2.6.2. Localization

KLF17 is predicted to be located at nucleus predominately with other organelles including Golgi apparatus, Cytosol, Mitochondria, Cytoskeleton and Extracellular Space. (*KLF17 Gene*, Aug 2, 2023).

2.7. Role Of KLF17

2.7.1. KLF17 in Cancer

KLF17 has mostly been researched in the perspective of cancer, where several studieas suggest its role as a tumor suppressor through various signaling pathways interactions such

as, TGF/SMAD signaling and p53, as well as in prevention of epithelial-to-mesenchymal transition.(Ali, Zhang, et al., 2015) Further research revealed that KLF17 expression levels were low and KLF17 expression reduction promoted tumor cell invasion, migration, and EMT shift.

Down-regulation of KLF17 has been discovered in a variety of human malignancies, including breast, lung, and liver cancer. Both human and chicken embryos have transactivation activity for KLF17. The majority of human malignancies, include BREAST CANCER (Ali et al., 2014), LUNG ADENOCARCINOMA , attenuation of KLF17 in patients with lung adenocarcinoma correlates with a momentary survival time, (Cai et al., 2012), HEPATOCELLULAR CARCINOMA (HCC) (Liu et al., 2013), GASTRIC CANCER, The abridged KLF17 protein expression in gastric cancer is associated with tumor bulk, lymph vascular invasion serves as a fair predictor of poor persistence in patients exhibiting gastric cancer surgery. (Peng et al., 2014), PAPILLARY THYROID CARCINOMA (PTC) (Ye et al., 2014), and NON-SMALL CELL LUNG CANCER (NSCLC), KLF17 and p53 have a tangible interaction that fosters KLF17's ability to prevent the proliferation of non-small cell lung cancer. (Ali, Bhatti, et al., 2015), and these have been shown to often have KLF17 down regulation that indicate the tumor suppressing activity of KLF17. In lung adenocarcinoma cells and primary tumor tissues KLF 17 was expressed at a lower level. Patients with HCC who have high KLF17 expression were observed to have a greater survival rate than those with low KLF17 expression, which is linked to a poor prognosis for HCC. KLF17 expression was observed to be considerably

downregulated. A lowered KLF17 expression was also noticed in PTC tissues on comparison to the nearby normal tissues. Revealing association of clinical-pathological characteristics and the prognosis of PTC patients with KLF17 expression.(S. Zhou et al., 2016).The majority of human tumors have diminished expression of KLF17, which indicates the association of downregulated KLF17 with cancer development and role of KLF17 as a tumor suppressor.

KLF17 was initially discovered as a new tumor suppressor via a direct genetic screen in a mouse model (K. Gumireddy et al., 2009). It exerts its protective effect against cancer by interacting with the promoters of EMT-related genes. A rising number of research in recent years have displayed that metastasis is aided by KLF17's suppressed expression. KLF17 has not always been demonstrated to inhibit metastasis, suggesting that it might be a context-dependent suppressive mechanism (Dong et al., 2014).

2.7.2. KLF 17 In Breast Cancer

Several studies have shown significant role of KLF17 in breast cancer as a tumor suppressor. KLF17 was observed to inhibit cancer cell invasion and EMT in breast cancer by binding to the consensus sequence ((5′-CACCC-3′) of DNA. Studies reveal the role of ID-1 as an oncogenic factor that facilitates the progression of breast cancer by inhibiting PTEN acting as a negative regulator of PTEN, a tumor suppressor protein. Overexpression of ID-1 was observed in highly invasive cancer cells such as prostate, cervical and breast

cancer. Linking the expression pattern of KLF17 and ID-1 may also pave way for a possible biomarker for lymph node metastatic breast cancer.(Ismail et al., 2014b).

In a study, A unique functional and molecular association between KLF17 and ER-dependent signaling in breast cancer cells was observed. It revealed that correlation of KLF17 and ER may also inhibit ER-dependent transcription. KLF17 blocks the production of ER-dependent genes by inhibiting the recruitment of ER to chromatin leading to downregulation of the expression of ER and the target genes. As a result of downregulated estrogen, the proliferation and survival of breast cancer cells was inhibited. (Jianping et al., 2020) (S. Zhou et al., 2016)

CHAPTER 3: METHODOLOGY

3.1 In-Silico Analysis

In Silico tools were used to predict the wildtype structure of KLF17. The pathogenic Missense variants were identified and their association with Breast cancer was also investigated.

3.1.1 Data Retrieval and Processing

Ensemble database was used to obtain the protein sequence of KLF17 gene with transcript ID ENST00000372299.4 and reference assembly of **GRCh38.CM000663.2** in FASTA format.

The gene sequence along with protein sequence was saved in FASTA format. Ensembl, Genome AD and Cosmic databases (Van Der Spoel et al., 2005) were used to retrieve the data of genetic variants.

Variants of KLF 17 were identified and processed while excluding the in-frame variants. The retrieved Variants were categorized into two groups i.e., Coding variants (missense variants synonymous variants, non-synonymous variants, nonsense variants and frameshift variants) and Regulatory variants (splice site variants and untranslated region variants (UTRs). Out of all types of variants, missense variants resulting in distorted proteins products were narrowed down and were further analyzed.

3.1.2 Unique Klf17 Variants Identification from Different Databases

The total number of variants obtained from the three databases were reported and sum of all the variants for the particular databases was calculated and presented through graphs. Variants from each of the databases were compared and redundancy of the common variants among all databases was eliminated.

3.1.3 Pathogenicity percentage OF SNPs

Six different in silico tools were used to determine the pathogenicity percentage of missense variants such as SIFT, PolyPhen, CADD, Mutation Accessor, Revel and Meta Lr. THESE tools assign particular score to the variants depending upon the level of their pathogenicity. SIFT assigns score ranging from 0-1 with 0 indicating variant to be deleterious and 1 or closer to 1 indicating that the variant is benign. PolyPhen allots each variant a score between 0 and 1, with 0.4 being considered benign, 0.4–0.8 as perhaps damaging, and >0.9 as probably damaging. Mutation accessor divides all the variants into four categories i.e., high (>0.9), medium (0.5-0.9), low (0.2-0.4), and neutral (0.0-0.1). MetaLR divides the variations into two categories: harmful (0.0-0.4) and tolerable (0.4 to 0.9). Same as MetaLR, REVEL also divides the variants into two categories where variants with score ranging from 0.0-0.4 are considered to be disease-causing if their scores while variants with score between 0.5 to 0.9 are regarded as benign. The Last tool CADD, assigns score by diving variants into two categories: likely benign and likely deleterious where score ranges from 0 to 35. The average proportion of obtained missense variants

pathogenicity was represented through graph. A threshold of 70 percent pathogenicity was applied to the variants. Variants with pathogenicity percentage greater than 70 percent were considered pathogenic. Rest of the variants either considered benign or tolerated. A total of 5 pathogenic variants of KLF17 were yielded and of these rs183242786 was chosen for further analysis in wet lab.

3.1.4 Prediction of 3D structure of Protein

To predict the 3D structure of KLF17 protein, The In-Silico tool ALPHAFOLD, was used to. The protein sequence of KLF17 retrieved from Ensembl data base in FASTA Format was submitted to ALPHAFOLD tool which in turn predicted the structure of protein which was saved in pdb format and was further visualized by using another tool i.e., PyMOL

Interpro was used to identify the domains of protein and amino acid residues found in those domains which were highlighted in Protein structure using PyMOL.

The predicted structure was validated through bioinformatics tools of saves 6.0. saves 6.0 is a validation server. ERRAT in saves6.0 (Colovos & Yeates, 1993) was used to assess the quality of the predicted structure and PROCHECK to validate the predicted structure by generating a Ramachandran plot.

3.1.5 Subcellular Localization & Phylogenetic Analysis

The in-silico tool Deeploc 1.0 (Almagro Armenteros et al., 2017) was used to determine the localization of KLF17. This tool generates a hierarchal tree to illustrate the expected subcellular localization of protein along with the probability score.

For phylogenetic analysis clustal omega was used to perform multiple sequence alignment among all members of KLF family which uses a progressive alignment tool to align the sequences exposing conserved and variable regions among proteins.

Another tool MEGA7 was also used for the phylogenetic analysis of KLF17 which generates a phylogenetic tree to trace down the evolutionary development of protein to determine its ancestor protein.

3.1.6 Stability Analysis of Protein

To determine the influence of variation on the structure and stability of Proteins various bioinformatics tools were used including MUpro, DynaMut2, MUTPRED and Maestro web.

MUpro predicts the change in stability of protein due to SNPs by revealing a change in energy score (DDG) that ranges from -1 to 1. A greater than 0 DDG score indicates an increase in stability while decrease in stability of protein is indicated by the DDG score less than 0.

Maestro Web uses the link between pH and $\Delta\Delta G$ as pH of solvent influences the protein stability largely. Pdb file of predicted structure of protein was uploaded on Maestro Web

and mutational change was specified resulting in predicting the change in stability of protein where stabilizing mutation is indicated by $\Delta\Delta G < 0.0$.

Another tool DynaMut2 was also used to assess how Single amino acid variations affect the stability of proteins and the intra molecular attractions. It also predicts vibration entropy (DDS) variation between wild type and mutant proteins.

3.1.7 Structural and Functional Effects of Variants

For structural and functional analysis, a tool Mutpred was used that explains the changes in the structure and function of protein due to SNP. A general score (g), or the likelihood that the amino acid alteration is harmful, is the result of MutPred2 uses a general score to indicate whether the amino acid alteration is pathogenic or not. A score end point of 0.50 implies pathogenicity. MutPred2 requires submission of protein FASTA sequence and substitution information in FASTA format.

DynaMut2 is also used to predict the effect of SNP on structure and function of protein. Dynamut2 also determines the rigidity or flexibility of the protein upon variation, making prediction of altered protein structure and function simpler.

Another tool HOPE, describes the changes in activity of protein and variation in charge and size of protein as well as its molecular interactions due to single amino acid variation.

3.1.8 Association of pathogenic SNPs with disease

A tool FATHMM was used to evaluate the association of pathogenic SNPs with Breast Cancer. Results of the Fathmm are in form of functional score which was applied to the both SNPs. On the basis of those score Fathmm assigns the SNPs as passenger or carcinogenic.

3.1.9. Effects of SNPs on mRNA Secondary structure

In-silico tool RNAFold was used to predict the secondary structure of mRNA of wild type gene and of both the variants as well. The differences in the structure and MFE energy as a result of variation were analyzed.

3.1.10 In Silico Mutagenesis

PyMOL was used to induced in Silico mutagenesis in wildtype structure of KLF17 that replaced the wild type amino acid of the original KLF17 protein with the one observed in the variant. To induce mutation wizard option was selected. Then, mutagenesis was selected from object panel and the wild type residue to be mutated was selected and replaced with the required amino acid. This mutagenesis resulted in slight change in structure protein and to make the change obvious the replaced amino acid was colored differently than the original one. The file was saved in. pdb format. MD simulations were run on both the wildtype and mutated structures.

3.1.11 MD Simulation

To further analyze the molecular dynamics of wild type and the two variants, MD Simulations were run using GROMACS software (Van Der Spoel et al., 2005) on a super computer. PuTTY, an open-source emulator, was used to provide data to the supercomputer and WinSCP was used to transfer files between local PC and supercomputer. Various commands were used to perform energy minimization, pressure equilibration and volume equilibration to run the final command of 20ns MD After completion of MD steps, different metrices such as Root Mean Square Deviation RMSD, Root Mean Square Fluctuation RMSF, solvent accessible surface analysis SASA, Radius of Gyration and total hydrogen bonds were observed and were presented through graphical representation.

3.1.12 Primer Designing

Primers of variant rs183242786 for genotypic analysis via ARMS PCR were designed via Primer 1. Primers for Arms PCR consisted of set of four Primers. Inner Reverse Primer, Outer Reverse Primer, Outer Forward Primer and Inner Reverse Primer. IF primer and IR primers are allele specific and show SNP whereas OF and OR primers and non-allele specific and bind to the strand irrespective of presence of any SNP.

Primer was designed using primer 1. Sequence containing the SNP position was selected and copy pasted onto primer 1 sequence box. The details of SNP such as position of mutant allele and non-mutant allele were provided. Upon submission the tool generated a set of 10 primers of which one with the best optimum temperature was selected and validated.

Table 3.1: Sequences of Primers and PCR conditions used for genotyping of KLF17 variants rs183242786

Variant Rs ID	PRIMERS	Denaturation	Annealing	Extension
rs183242786 **T/C**	**Forward inner primer (C allele):** GGTACACACCAGATATCGACCATATACAC	94º	57º	72º
	Reverse inner primer (T allele): ATGAACTCCCGGCTGCACTGATCCCA			
	Forward outer primer (5' - 3'): CTGGACTGCTCTGCCTCACAGAGTTAGA			
	Reverse outer primer (5' - 3'): AACCCAGAATGAGGAGAGGAGACTGACC			

3.2. Laboratory Based Analysis

3.2.1 IRB Approval and Sample Collection

Before starting the study, Institutional review board approval was obtained from parent department, ASAB. Blood samples were collected from patients with their consent at BBH, RWP and DHQ, RWP. A tourniquet was tied around 3 cm above a visible vein on hand to collect the blood. About 3-5 ml blood was drawn with sterile syringes and was transferred to EDTA tubes immediately. The selection criteria for sample collection included:

Inclusion Criteria: Only females who have been clinically diagnosed with localized or metastatic Breast cancer.

Exclusion Criteria: Female patient whose tumors have been cleared and had visited just for their follow up.

3.2.1.1 Sample Collection

Samples were collected from the female patients who were diagnosed with the localized or metastatic breast cancer that were currently undergoing chemotherapy or radiotherapy along with a proportion of sample from such patients that had not undergone any treatment yet. Each patient's confidentiality was made sure by assigning a unique code to each of them. History forms of all the participating patients were filled properly. Patients participated belonged to different ethnicities.

3.2.2 Genomic DNA Extraction

Genomic DNA extraction was carried out by phenol-chloroform method for both patients' samples and controls. Buffer solution A, B, C and D were prepared for this purpose. Solution A comprised of 0.32M sucrose, 10Mm tris base (Ph=7.5) and 5mM MgCl2 which was dissolved in autoclaved distilled water to make a solution of 500ml and was autoclaved later. After the solution had cooled down completely, 5ml of 1percent v/v triton was added as it is thermos labile. Solution A was then stored at 40 degrees.

For solution B, 10mM Tris-base pH=7.5 (0.6057g), 400mM sodium chloride (11.685g), and 2mM ethylene diamine tetra acetic acid (0.58g) were dissolved in distilled autoclaved

water. pH of each of these was adjusted between 7-7.5. the pH was maintained by adding HCL to lower the pH value, the solution was raised up to 500ml.

Solution C is composed of only Phenol. Upon exposure to light phenol oxidizes as it is a photo reactive molecule and so it was kept in dark region covered properly in aluminum foil.

To make 25ml of solution D, 24ml of chloroform and 1ml of isoamyl alcohol were dissolved and was diluted up to 500 ml with distilled water.

DAY 1

1. On first day of DNA extraction 750μl of blood was aliquoted in an Eppendorf with the equal volume of solution A. After settling for 10 minutes, it was centrifuged at 13000 rpm for 10 minutes.

2. After centrifugation supernatant was discarded, the pellet obtained was dissolved in 400μl of solution A. Sample was centrifuged again at 13000 rpm for 10 minutes.

3. After centrifugation supernatant was discarded again. Solution A was added again to break the cell membranes, escaping the DNA out of the cells.

4. After the second washing, 400μl of solution B was added and the pellet was dissolved in solution which was then subjected to centrifugation at 13000 rpm for 10 minutes.

5. After centrifugation, supernatant was again discarded and pellet was dissolved in 20μl of SDS and 5μl of proteinase K. Sample was left for overnight incubation at 37°

DAY2

1. On Day 2 Tubes were taken out of incubator. 250μL of solution C and 250μl was added in an empty Eppendorf's which was then transferred to sample Eppendorf's. Solution C separates organic and aqueous phase in solution whereas solution D stabilized the coagulation of protein.

2. Upon addition of solution C and D and centrifugation two layers were formed. Upper layer was carefully removed using pipettes and transferred to the fresh Eppendorf. Where 500μl of isopropanol and 55μl of sodium acetate was added and inverted several times until a thread like DNA appeared and was again subjected to centrifugation for 10 minutes.

3. After centrifugation, the pellet was washed with 75 percent ethanol and centrifuged for 8 minutes. After centrifugation Eppendorf's were air dried so that ethanol may evaporate. Upon drying, 200μl PCR water was added and mixed thoroughly.

Bands of the extracted DNA were then observed through Gel Electrophoresis under UV light.

3.2.3 Tetra ARMS PCR

Tetra ARMS-PCR is the revised version of the Amplification Refractory Mutation System-Polymerase Chain Reaction (ARMS-PCR). It is viewed as the "Gold Standard" for genotypic analysis in genomic DNA through the discovery of SNPs in the genome. Tetra ARMS-PCR uses two pairs of primers, two internal primers and two outer primers for the variant identified after performing the Insilico analysis. A 20 µl reaction mixture was used for each PCR cycle. 1 µl of each primer (primer 1-4), 2 µl of DNA taken from both control and positive blood samples, 2µl of Nuclease free water and 12 µl of Solis master mix were included in each of the 20 µl reaction mixture. The PCR buffer, magnesium, dNTPs, Taq polymerase, and loading dye were all included in master mix. PCR tubes holding the reaction mixture were subjected to spin for a few seconds to integrate everything and were placed in the thermocycler then. During primer optimization Multiple PCRs with varied chemical concentrations and Tm were carried out. The optimum temperature ranges for PCR cycles were determined using the gradient PCR technique was used to determine optimum temperature for PCR cycle. Each PCR reaction was set for 30 cycles, with three key phases.

In the first stage of the ARMS-PCR, the template DNA undergoes denaturation for around 5 minutes at 95 degrees C. The double-stranded template DNA was denatured to single-stranded DNA in this phase. The daughter DNA strands were denatured in stage 2 of the PCR cycle. This stage consisted of 30 cycles, each lasting for 00:30 seconds at 94 degrees Celsius. The annealing was done at 57 degrees C for variant rs183242786, and this stage took 30 seconds. Extension phase following The Annealing phase, lasted 30 seconds at 72

degrees C. The final stage of PCR was the final extension, which took 7 minutes at 72 degrees C. After the last extension, the PCR cycle was put on pause at 4 degrees C to allow the PCR product to cool down and the PCR tubes were stored at -40 degrees C till further visualization through Gel Electrophoresis.

3.2.4 Gel Electrophoresis

The gel electrophoresis technique was used to visualize the genomic DNA as well as the PCR products. An agarose gel was utilized for this. For this purpose, A 2 percent (W/V) agarose gel was constructed to visualize the PCR products obtained in the foregoing step. In 100ml of 1X TAE buffer, 2 grams of agarose powder was dissolved. The agarose was completely dissolved into the buffer solution after being heated in the microwave until a clear solution appeared. When the gel mixture had cooled down slightly, 5-7 µl of ethidium bromide was added to the gel solution and carefully stirred to prevent the formation of bubbles. Meanwhile, the gel solution was transferred into the gel casting tray that had been prepared earlier. The gel was allowed to solidify at room temperature. Then the gel was placed in the electrophoresis tank, filled with 1X TAE buffer. The wells formed in the gel were filled with 8-l0 µl of PCR products and 5 µl of the ladder. As the master mix already contained the loading dye, there was no need to mix the PCR products with loading dye. Finally, electrophoresis was performed for 45-60 minutes at 80V and 500 A, or until the bands were properly separated and fully observable. To examine the PCR results on the gel, the gel doc or UV transilluminator were used. Similar procedures were used to see the DNA isolated from the samples. For this purpose, 1 percent agarose gel was made and 8-

10 µl of DNA samples were combined with 1 µl of loading dye and placed into agarose gel wells.

3.2.5 Statistical Analysis

To statistically analyze the genotyping data, Graph Pad Prism was used. Both the control samples and the patients with breast cancer samples were subjected to the Fisher exact test and statistical significance of genotypes were determined in relation to breast cancer. Moreover, odds ratios and relative risks associated with the genotypes and alleles were also evaluated. Any P-value below 0.005 was considered to be statistically significant

CHAPTER NO 4: RESULTS

4.1 KLF17 Variant Identification

Around 874 variants of KLF were identified and retrieved through different databases Ensembl, Genome AD and Cosmic comprising of 416,300 and 158 variants respectively. The 874 variants included Missense variants, nonsense variants, splice region variants and frameshift variants. Where Ensembl contained 333 missense variants,15 nonsense variants, 43 splice region variants and 25 frameshift variants. On the other hand, Genome AD consisted of 257 missense variants,12 nonsense variants, 15 splice region variants and 16 frameshift variants and Cosmic consisted of 150 Missense variants, 7 nonsense variants and 1 frameshift variant. Of 874 variants 526 were unique variants while rest of 348 variants were redundant
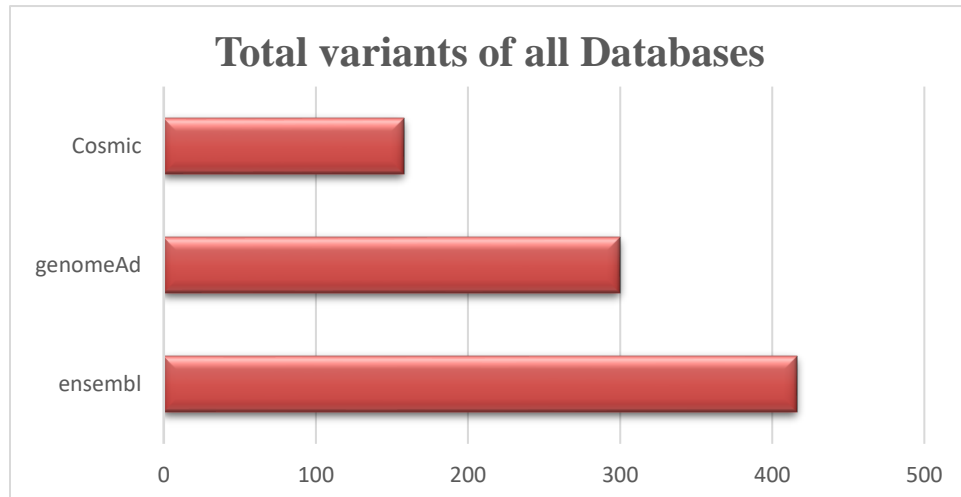
Figure 4.1 Total variant in each database.
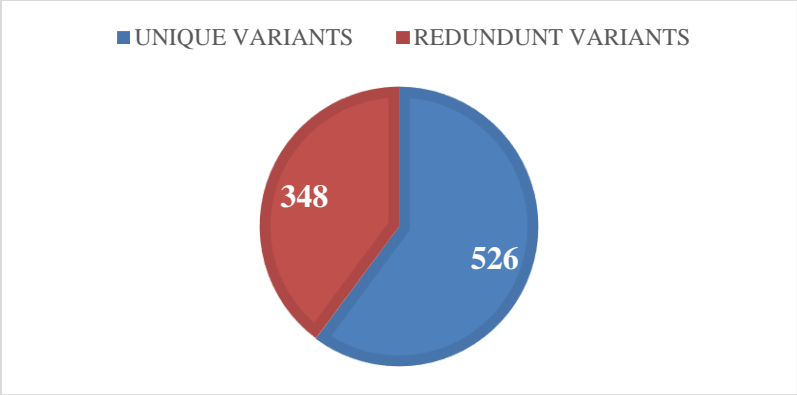
Figure 4.2: Filtered unique and redundant variants of all databases.
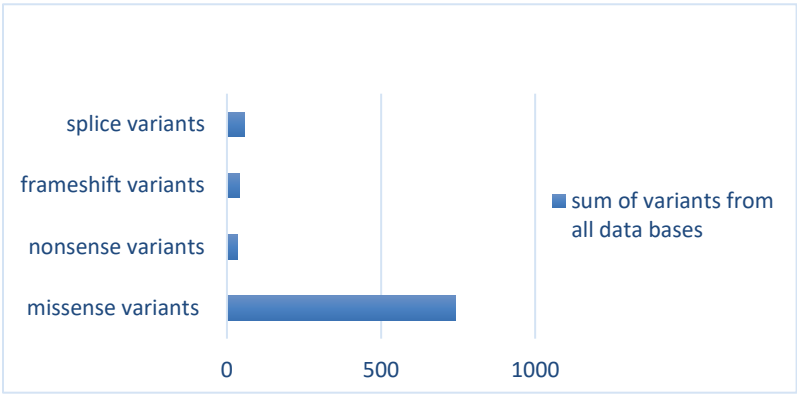
Figure 4.3: Types of all variants count in all databases.
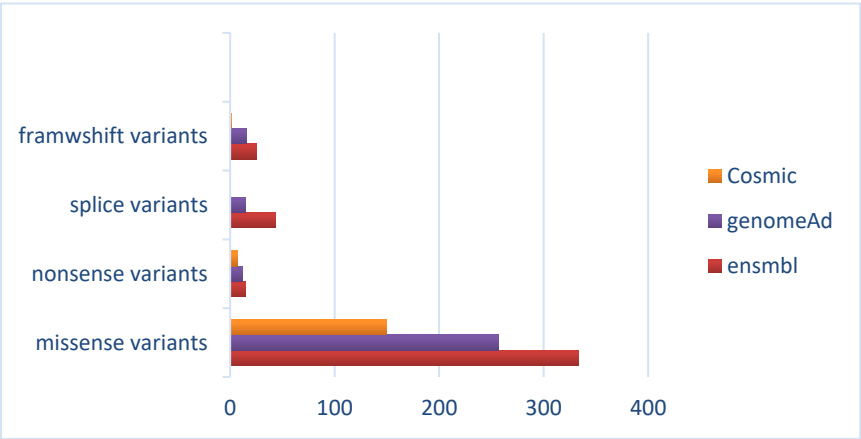
Figure 4.4: Numbers of each type of variant in every database.

Among the filtered variants missense variants were found to be abundant so were filtered out for further analysis excluding nonsense variants, frameshift variants and splice regions variants. To calculate the relative missense variant per SNP per Exon further analysis was carried out. A total of 4 exons were identified in KLF17 encoding around 389 amino acids. Exon 1 covering N-terminal region of transcript encodes 27 amino acids and showed SNP frequency of 19 whereas exon 2 encodes 260 amino acids in N-terminal region and 22 amino acids in C-terminal with a total of 248 missense SNP frequency. Exon 3 encodes 80 amino acids in C-Terminal region with a SNP frequency of 66. While Exon 4 does not encode any amino acid. (Figure 4.5)
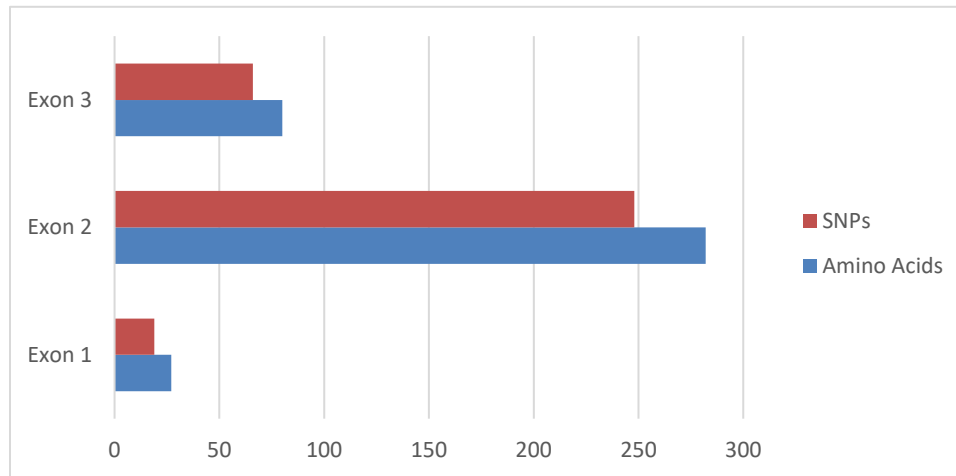
Figure 4.5: Frequency of missense variants per exons.

The ensemble database contained total 5163 variants/ exonic SNPs of KLF17 out of which 333 missense SNPs were selected and pathogenicity of which were calculated using in

silico tools like SIFT, PolyPhen, REVEL, MetaLR and Mutation Accessor. These tools can help indicate the pathogenic association of SNP and effect on its structure and function. These in silico tools have their own way of scoring results. Pathogenicity of all selected variants were calculated as shown in figure 4.6
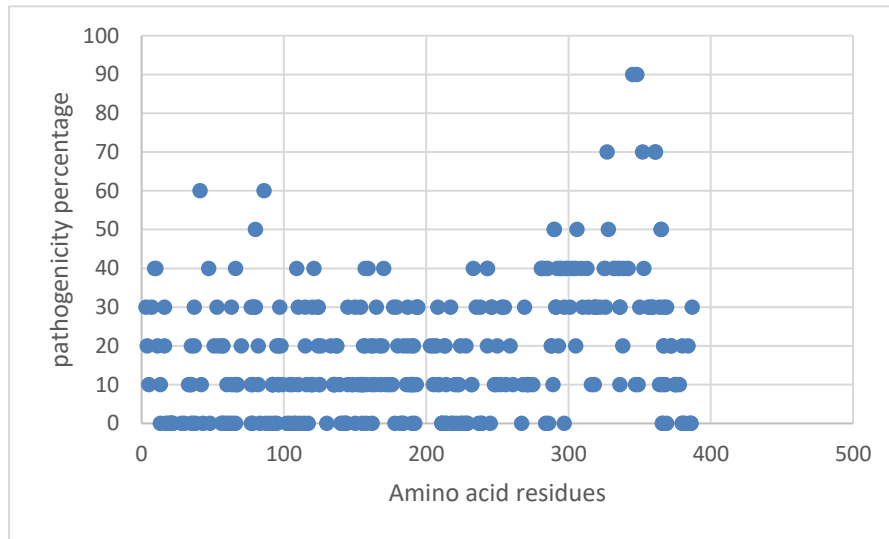
Figure 4.6: Percentage pathogenicity of selected missense variants from Ensembl database.

Table 4.1: Pathogenicity table of selected missense variants.

SR.NO	Variant ID	AA Coord	PATHOGENEICITY %AGE
1	rs775862566	3	30
2	rs747380426	4	20
3	rs768956475	5	10
4	rs1466868607	7	30
5	rs777044721	9	40

6	rs761955311	10	40
7	rs1191730059	11	20
8	rs895680201	13	10
9	rs895680201	13	0
10	rs1292351628	16	30
11	rs1292351628	16	20
12	rs1392486329	17	0
13	rs1459845884	18	0
14	rs201635350	20	0
15	rs1196683397	20	0
16	rs1384406493	21	0
17	rs201093755	21	0
18	rs201093755	21	0
19	rs994439989	22	0
20	rs141824161	28	0
21	rs758567338	30	0
22	rs755730666	33	10
23	rs1313040254	34	10
24	rs11210969	35	20
25	rs11210969	35	10
26	rs756737268	35	0
27	rs1173205062	36	0
28	rs780679632	37	30

29	rs1389288063	37	20
30	rs752182180	38	0
31	rs755386139	41	60
32	rs781645367	42	10
33	rs181031157	43	0
34	rs770019075	47	40
35	rs1284180052	48	0
36	rs1453801096	51	20
37	rs771038219	53	30
38	rs774265767	54	20
39	rs374100654	56	0
40	rs772010943	57	20
41	rs2429051	57	20
42	rs2429051	57	0
43	rs1460167383	58	0
44	rs1371277195	60	10
45	rs763751748	60	0
46	rs1412280550	62	0
47	rs1368024086	63	30
48	rs1407574759	63	10
49	rs1215628498	64	0
50	rs761384044	65	10
51	rs749970205	66	40

52	rs749970205	66	0
53	rs755476285	67	10
54	rs755476285	67	10
55	rs1249941365	70	20
56	rs753058278	77	10
57	rs144146386	77	0
58	rs144146386	77	30
59	rs770985004	78	30
60	rs770985004	78	0
61	rs377749405	78	10
62	rs755453227	79	30
63	rs771953734	80	50
64	rs6656945	80	30
65	rs768404279	82	10
66	rs1225283717	82	20
67	rs966528895	83	0
68	rs776336787	86	60
69	rs373989636	87	0
70	rs1447593450	90	0
71	rs866116560	92	10
72	rs866116560	92	10
73	rs866116560	92	10
74	rs139530696	92	10

75	rs200013903	93	0
76	rs781778682	95	20
77	rs200239795	95	0
78	rs756494319	96	10
79	rs756494319	96	20
80	rs147552179	96	10
81	rs1162275301	97	30
82	rs1446964935	98	20
83	rs757403692	99	10
84	rs1317003541	102	0
85	rs758518853	104	0
86	rs780007411	104	10
87	rs768655136	106	10
88	rs776227278	107	0
89	rs747987683	108	0
90	rs1257985524	109	40
91	rs373734125	110	30
92	rs370566517	110	10
93	rs1185526523	111	0
94	rs1470294732	114	0
95	rs776050611	115	20
96	rs149226855	115	30
97	rs764667525	116	10

98	rs754157499	117	0
99	rs1431009749	119	10
100	rs1301073074	120	10
101	rs1336623249	120	10
102	rs1336623249	120	30
103	rs757637368	121	40
104	rs1277879040	124	30
105	rs145700906	124	20
106	rs750681943	124	30
107	rs758461714	125	10
108	rs575683674	125	10
109	rs1421653568	126	20
110	rs1438077316	130	0
111	rs144529908	133	20
112	rs769752999	135	10
113	rs769752999	135	10
114	rs1571992756	135	10
115	rs749005256	136	10
116	rs200960950	137	20
117	rs1397602310	137	20
118	rs773666549	139	10
119	rs759117084	140	0
120	rs764455518	141	0

121	rs1230922636	143	0
122	rs1230922636	143	0
123	rs1339758687	144	0
124	rs762131869	145	30
125	rs765655179	145	10
126	rs145084649	148	10
127	rs1300864486	148	10
128	rs139636116	150	0
129	rs751544291	150	30
130	rs112819362	152	10
131	rs752750606	154	10
132	rs755957342	154	30
133	rs777651690	155	0
134	rs777651690	155	10
135	rs777651690	155	10
136	rs2485652	156	10
137	rs1571992893	156	20
138	rs1423361318	157	20
139	rs1345272185	157	40
140	rs1325338066	158	10
141	rs1325338066	158	0
142	rs1204421906	159	40
143	rs1571992924	161	20

144	rs1326946271	162	0
145	rs1326946271	162	10
146	rs745344925	163	20
147	rs1276037880	164	10
148	rs1022192032	165	30
149	rs1311056571	167	20
150	rs771648624	168	10
151	rs1272007488	169	20
152	rs140666029	170	40
153	rs138303007	172	10
154	rs1571993007	174	10
155	rs1446909205	176	10
156	rs1571993019	177	30
157	rs773530149	178	0
158	rs1034871775	179	30
159	rs763402328	180	20
160	rs143224105	182	0
161	rs751848711	183	0
162	rs1299203098	184	0
163	rs767765022	184	20
164	rs752836277	186	10
165	rs1415025608	187	20
166	rs1415025608	187	30

167	rs551004452	188	10
168	rs144255993	189	10
169	rs141307073	190	20
170	rs141307073	190	0
171	rs1317620438	190	10
172	rs1215024029	191	20
173	rs778395622	191	10
174	rs778395622	191	10
175	rs531602229	192	0
176	rs267598619	193	10
177	rs267598619	193	30
178	rs138244178	194	30
179	rs138244178	194	30
180	rs1363253252	203	20
181	rs1051166117	205	10
182	rs1377192491	205	20
183	rs368038713	207	20
184	rs759744470	208	10
185	rs767855234	208	30
186	rs775671290	209	10
187	rs760859256	211	0
188	rs764096822	211	0
189	rs764096822	211	0

190	rs373793338	213	20
191	rs373793338	213	20
192	rs750191918	214	0
193	rs750191918	214	10
194	rs758007346	214	0
195	rs1249920439	217	30
196	rs1437980822	218	0
197	rs746500531	218	0
198	rs1396867376	220	10
199	rs1454636692	221	0
200	rs1175129616	222	10
201	rs1357163166	223	10
202	rs149698028	224	20
203	rs149698028	224	0
204	rs556051294	225	0
205	rs199651256	227	0
206	rs1234143689	228	20
207	rs1234143689	228	0
208	rs1299664261	229	0
209	rs1309567263	232	10
210	rs1309567263	232	10
211	rs370557101	233	40
212	rs772368262	235	30

213	rs202045166	237	0
214	rs760805966	238	30
215	rs746959776	239	0
216	rs746959776	239	0
217	rs773548701	243	20
218	rs1461907202	243	40
219	rs1461907202	243	40
220	rs1327763250	245	0
221	rs974077986	246	30
222	rs776418350	246	30
223	rs867352052	248	10
224	rs761749908	249	10
225	rs765119923	250	20
226	rs1394505426	252	10
227	rs750185270	253	30
228	rs148405031	255	30
229	rs200292026	256	10
230	rs1279901807	259	20
231	rs1295269737	261	10
232	rs144379257	267	0
233	rs757739139	268	10
234	rs779305958	269	30
235	rs1186927146	271	10

236	rs746209131	271	10
237	rs1423979872	272	10
238	rs1418122062	275	10
239	rs371659233	281	40
240	rs371659233	281	40
241	rs776735636	282	40
242	rs1165338985	284	0
243	rs761823672	285	40
244	rs761823672	285	40
245	rs368227907	286	0
246	rs766222742	288	20
247	rs766222742	288	20
248	rs12760706	289	10
249	rs1334424511	290	50
250	rs1334424511	290	50
251	rs767103784	291	30
252	rs767103784	291	30
253	rs755624924	292	40
254	rs902190699	293	20
255	rs1487226761	294	40
256	rs1257850488	295	40
257	rs779440513	297	30
258	rs1179640821	297	0

259	rs751037720	298	40
260	rs138188954	299	40
261	rs142735791	301	30
262	rs200657181	302	40
263	rs768875857	304	40
264	rs781319287	305	20
265	rs200840562	305	40
266	rs1571993674	306	50
267	rs867800845	309	40
268	rs1384288617	310	30
269	rs1235780436	313	40
270	rs1311219776	314	30
271	rs1320655865	316	10
272	rs1194743535	318	10
273	rs1385183693	318	30
274	rs777766657	319	30
275	rs768144719	319	30
276	rs774318789	322	30
277	rs1329421990	325	40
278	rs745641456	326	30
279	rs771825513	326	40
280	rs1324200171	327	70
281	rs1186210887	328	50

282	rs763841467	332	40
283	rs763841467	332	40
284	rs776240055	335	40
285	rs1240178860	335	40
286	rs761460881	336	10
287	rs761460881	336	30
288	rs1255619127	336	30
289	rs752180233	338	20
290	rs752180233	338	40
291	rs141181356	341	40
292	rs777990070	342	40
293	rs183242786	345	90
294	rs1420285745	347	10
295	rs778813952	348	90
296	rs371898795	349	10
297	rs775079569	350	30
298	rs139118447	352	70
299	rs1401007670	353	40
300	rs761403973	356	30
301	rs1356953637	357	30
302	rs772796419	359	30
303	rs780277452	359	30
304	rs1264312672	361	70

305	rs1354891411	361	70
306	rs374824606	364	30
307	rs374824606	364	10
308	rs1287199192	365	50
309	rs768103063	365	50
310	rs369271264	366	0
311	rs756547380	366	0
312	rs756547380	366	10
313	rs754122457	367	20
314	rs757312344	367	30
315	rs757312344	367	20
316	rs1187275423	368	10
317	rs144526061	369	0
318	rs1160157792	369	30
319	rs746750521	372	20
320	rs746750521	372	20
321	rs1376937554	375	10
322	rs1376937554	375	10
323	rs1571994290	376	10
324	rs199995724	378	10
325	rs769578320	380	0
326	rs113334663	380	0
327	rs113334663	380	20

328	rs540269615	382	0
329	rs772432491	384	20
330	rs1324704005	385	0
331	rs1483713949	386	0
332	rs1483713949	386	0
333	rs572748957	387	30

To filter out the potentially deleterious variants out of the 333 missense variants, 70% threshold was adjusted resulting in 6 Missense variants with more than 70% pathogenicity where two of the amino acids (rs183242786) and (rs778813952) at position of 345 and 348 respectively had greatest pathogenicity of 90% and were selected for further study.

Table 4.2: Missense variants after threshold pathogenicity sorting.

Variant ID	AA coord	Alleles	sift	polyphen_class	revel_class	Meta lr	Mutation_assessor	pathogenicity %age
rs1324200171	327	C/T	1	1	0.5	1	0	70
rs183242786	345	T/C	1	1	0.5	1	1	90
rs778813952	348	T/C	1	1	0.5	1	1	90
rs139118447	352	C/A/T	1	1	0.5	1	0	70
rs1264312672	361	A/G	1	1	0.5	1	0	70
rs1354891411	361	C/G	1	1	0.5	1	0	70

4.2 Subcellular Localization of KLF17

Subcellular localization of KLF17 was predicted using the tool Deeploc1.0.and it predicted KLF17 to be present in nucleus with a likelihood of 99.99 percent along with a very low likelihood of 0.01 percent in cytoplasm as well. KLF17 was predicted to be absent in rest of the organelles.

Deeploc1.0 also predicted the type of KLF17. The result showed the protein to be almost 71% soluble and 29% membranous protein.
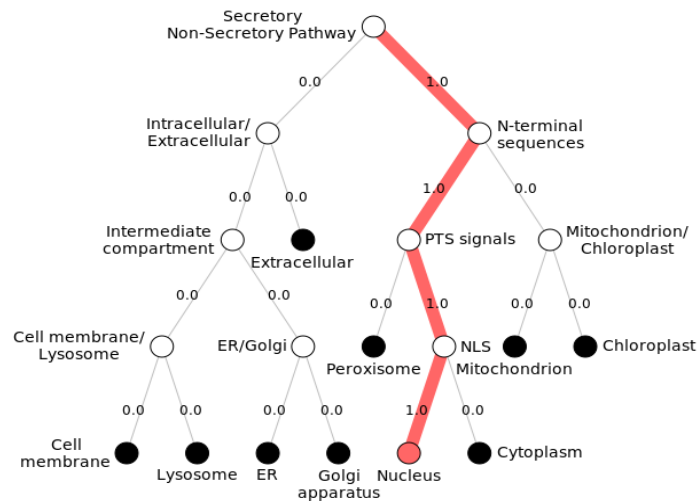
Figure 4.7: subcellular localization of ZFP393 using Deeploc1.0 depicting the strong evidence of presence of KLF17 in nucleus. It is directed to Nucleus via PTS as well as NLS signals as depicted by score 1.0 in red.

Table 4.3 : predicted subcellular localization of KLF17.

Localization	Likelihood
Nucleus	0.9999

Cytoplasm	0.0001
Cell membrane	0
Endoplasmic reticulum	0
Golgi apparatus	0
Peroxisome	0
Extracellular	0
Mitochondrion	0
Lysosome/Vacuole	0
Plastid	0

Table 4.4: Predicted type of KLF17 Protein.

Type	Soluble	Membrane
Likelihood	0.7096	0.2904

4.3 Structure Prediction of KLF17 Protein

KLF17 protein's structure was predicted by Alpha fold, which used per residue model confidence score (pDDLT) ranging from 0 to 100. Based on this model a 3D structure of KLF17 protein was obtained. whereas, 3 C-terminal conserve domains showed a high confidence score (70-90 pDDLT score), the few residues within C-terminal domain also showed low confidence score (50-70 pDDLT score). Other than C-terminal region the residues showed a lower confidence score (<50 pDDLT score) Hence, only C-terminal domains consisting of 3 alphas helical conformation was predicted shown in figure 4.8

While Interpro analysis provided labelled domains of KLF17 protein consisting of variable N-terminal domain covering 9-287 amino acids, while C-terminal Domain consisted of 3 Zinc Finger motifs covering the 283-312, 313-342, 343-365 amino acids respectively as shown in figure 4.8.
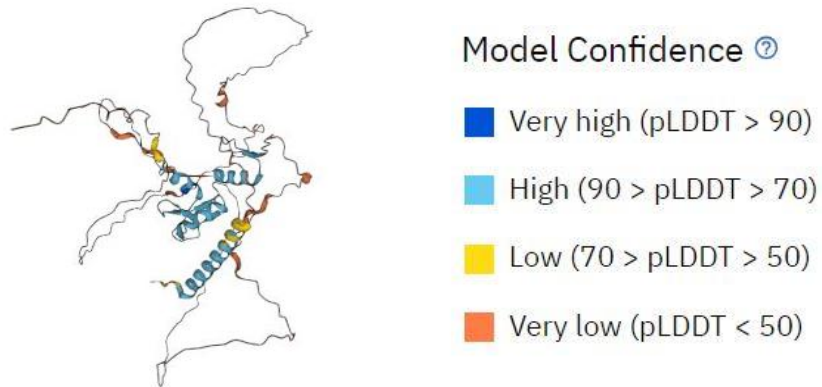
Figure 4.8: KLF12 protein's 3D predicted structure, C-terminal region domains showed high confidence score (70-90) of predicted structure.

PyMOL was used to highlight and visualize the various regions of gene KLF17 (Fig: 4.9 b).

a.
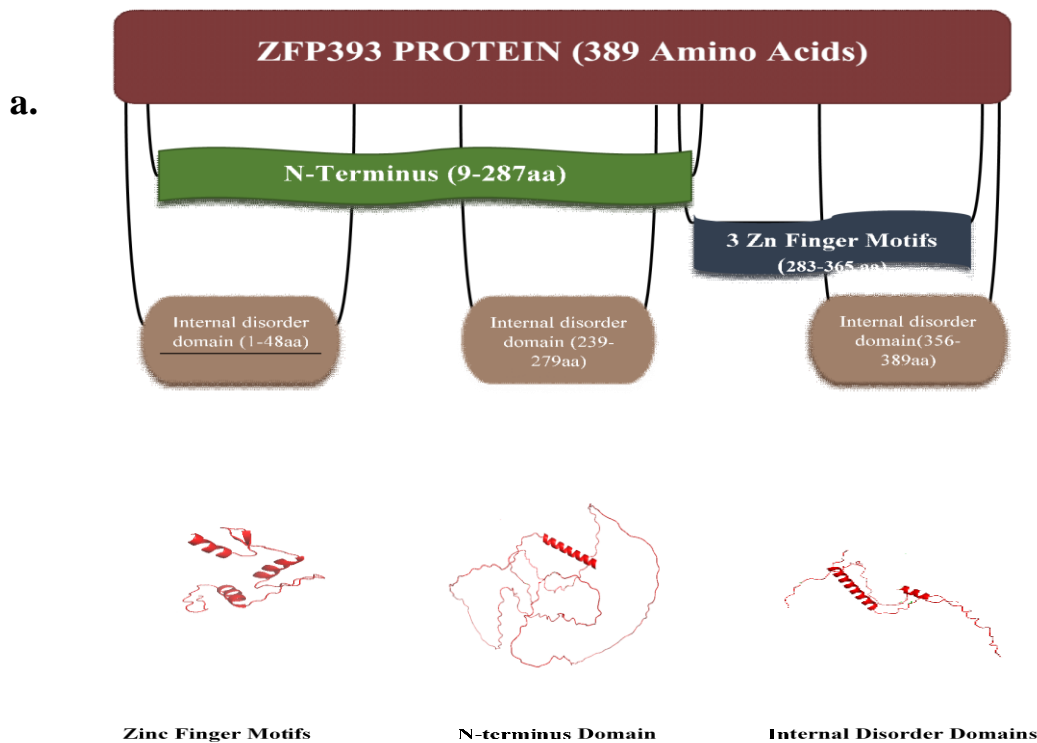
Figure 4.9: a) Highlighted domains of KLF12 protein, consist of N-terminal, internal disorder, and C-terminal domains, b) 3d structure of highlighted domains

The predicted structure of KLF17 was validated through SAVES 6.0 structure validation tools. On the basis of non-bonding interactions within different atoms ERRAT (tool of saves6.0) evaluated the quality factor of the structure. The quality factor was 94.068 indicating a large section of protein structure to be in allowed region, suggesting a higher quality protein structure. Another tool of Saves6.0, PROCHECK, was used to attain Ramachandran Plot to further validate the structure of KLF17 protein. Ramachandran Plot showed 62.2% residues in most favored regions,24.8% residues in additional allowed regions and 4.9% residues in generously allowed region while in disallowed regions only

8.1% regions were found that yields a total of 91.9% residues present in allowed region showing the higher quality of the structure.
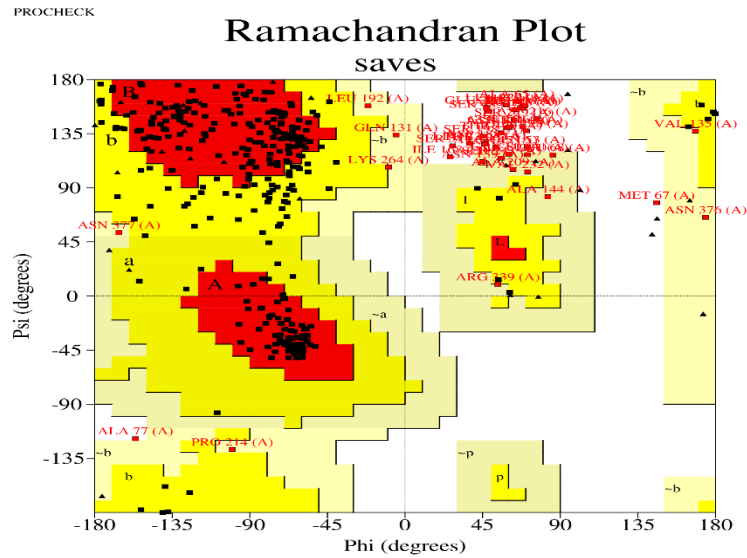
Figure 4.10 Ramachandran plot of the KLF17 predicted structure validation of via Saves6.0. The evaluation indicated that 91.9% residues of KLF17 lie in allowed regions whereas only 8.1% lies in disallowed regions.

4.4 Phylogenetic Analysis

The phylogenetic tree of KLF family via MEGA7 is shown below (fig 4.10) along with the bootstrap values cited on the branches. The tree show how KLF17 is related to the other members of KLF Family whereas the bootstrap value indicates the confidence of members belonging to common ancestor. A value of 100 shows the maximum confidence The confidence value for KLF17 and KLF18 belonging to same ancestor is 89.Phylogenetic

analysis was also performed via clustalW tool that resulted in almost same pattern as that of MEGA7 depicting the close association of all proteins belonging to the same ancestor.

a.   b.
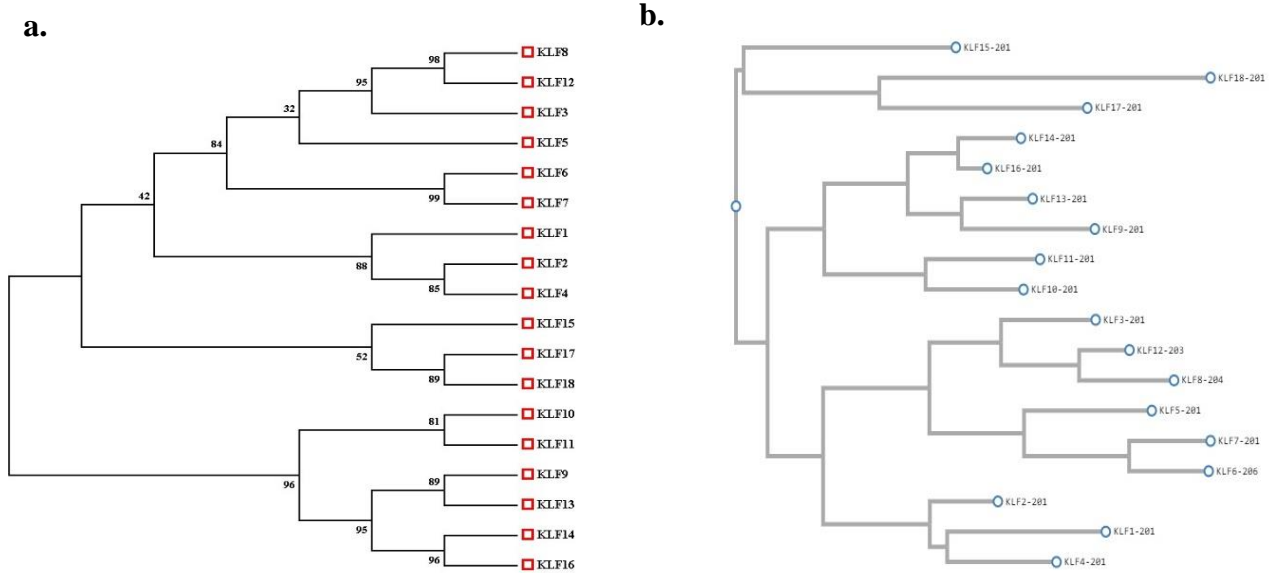
Figure 4.11 Molecular Phylogenetic analysis by (a) Maximum Likelihood method. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (-5112.96) is shown.

4.5 Conservation Analysis of KLF17

Conservational analysis of KLF17 was done through in-silico tool Consurf which used the surface mapping of phylogenetic information to identify the functional regions in proteins in the form of highlighted amino acid residues. Consurf uses different codes to highlight the amino acid in numerous colors based on the conservation of amino acid residues. The

conservation on scale 1-4 depict those amino acids are variable, scale 4-6 show moderately conserved amino acids where scale 7-9 show highly conserved amino acid residues. The protein kinase domain of KLF17 is evolutionary conserved.
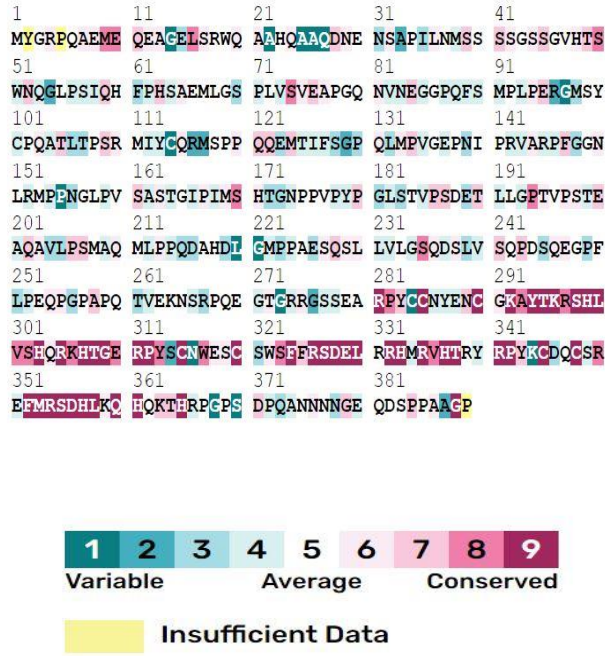
Figure 4.12 Protein Conservation KLF17 showing *Consurf* analysis results for 389 amino acid residues of KLF17

4.6 Protein Stability Analysis

To analyze the structural and functional variations in selected two missense variants, In-silico tools such as MUpro, Mutpred, Maestro web and dynamut2 were used. As a result of MUpro a Delta-Delta G (Gibbs Free Energy) were generated that indicated the stability

of the selected SNPs. Hence the DDG value of both the variants i.e., -1.1221967 (C345R) and -0.9518535 (C348R) indicated decrease in stability of the protein due to mutation.

Dynamut2 proposed DDG value of -0.247 kcal/mol for C348R and for the variant C345R the DDG value was -0.637 kcal/mol indicating a destabilizing effect of mutation in both SNPs due to altered multiple interaction and bonding between the amino acids' residues.

The DDG values provided by Maestro Web for selected variants **rs183242786**, and **rs778813952**, are 1.61937832 and 1.044 kcal/mol with confidence score estimation (C_{pred}) 0.797303 and 0.752 kcal/mol respectively. The obtained result implicates the DDG>0 suggested destabilizing variation with higher reliability as C_{pred} score 0.0 (not reliable) and 1.0 (highly reliable). Hence both variants lead to destabilization in protein.

Table 4.5: Protein stability analysis.

Tools	rs 183242786		rs778813952	
	DDG Value	**Consequence**	**DDG Value**	**Consequence**
MUpro	-1.1221967kcal/mol.	Decrease Stability	-0.9518535	Decrease Stability
Dynamut 2	-0.637 kcal/mol	Decrease Stability	-0.247 kcal/mol	Decrease Stability
Maestro Web	1.61493783 kcal/mol	Decrease Stability	1.044 kcal/mol	Decrease Stability

4.7. Variant's Effect on Protein's Structure and Function

To evaluate the structural and functional changes of amino acid variation, Mutpred, HOPE and dynamut2 tools were used. Results from MutPred showed that both the SNPs C345R and C348R resulted in altered disordered interface with probability of 0.40 and 0.35 being pathogenic respectively and p-value of 3.5e-03 and 5.5e-03 respectively.

Table 4.6: List of possible alteration in KFL17 protein when amino acid number 345 and 348 altered. Estimated p-Values were also provided by MutPred2.

Residual Change	Mechanisms	p-Value	Probability
C345R	Altered Disordered Interface	3.5e-03	0.40
C348R	Altered Disordered interface	4.4e-03	0.35

Dynamut tool was used to evaluate change in proten structure flexibility caused by variation by measuring the change in vibrational entropy energies (ΔSVib ENCoM) of wild type and variants protein structures The ΔSVib ENCoM value for C345R variant was 0.044 kcal.mol^{-1}. K^{-1} that depicted the Increase of molecule flexibility whereas for variant C348R ΔSVib ENCoM was -0.273 kcal.mol^{-1}. K^{-1} that depicted the decrease of molecule flexibility (fig 4.13)

Table 4.7. protein stability of C345R and C348R

variants	$\Delta\Delta S_{vib}$ ENCoM:	$\Delta\Delta G$
C345R	-0.273 kcal.mol^{-1}.K^{-1} (Decrease of molecule flexibility)	-0.637 kcal/mol (Destabilizing)
C348R	0.044 kcal.mol^{-1}.K^{-1} (Increase of molecule flexibility)	-0.247 kcal/mol (Destabilizing)

a.
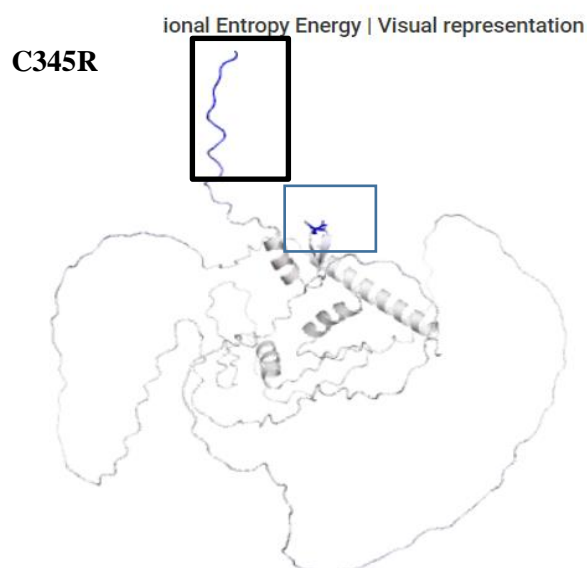
b.
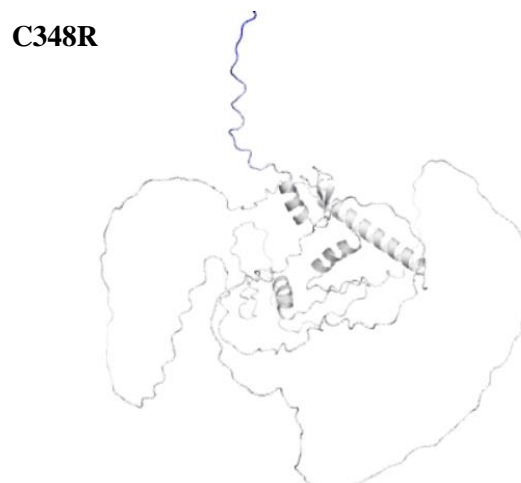
Figure 4.13 Vibrational entropy energy: Amino acid residues are colored according to the vibrational entropy changes upon variation of amino acid residue. Blue region represents a decrease in flexibility (rigidification) and red suggested the gain in flexibility of structure.

HOPE project was used to evaluate the structural and functional changes in protein KLF17 due to variations , C345R and C348R And it was analyzed that Both the mutants C345R

and C348R were bigger than the wildtype in their size which might lead to the bumps, as well it was also observed that the wildtype molecule was neutral while the mutant C345R gained negative charge and SNP C348R gained a positive as a result of the variation, due to the charges, loss in hydrophobicity was also noted in SNPs and hence hydrophobic interactions of protein.

Table 4.8: HOPE functional and structural analysis of variant 1(C345R) and variant 2 (C348R)

Variants	Size	Charge	Hydrophobicity	Location	Conservation	MetaRNN score	Pathogeneicity
Wild type	Normal	Neutral	More hydrophobic	-	More	-	-
C345R	bigger	Positive	Less hydrophobic	Zinc Finger Domain	Less	0.9512116	Probably Damaging
C348R	Bigger	Positive	Less hydrophobic	Zinc Finger Domain	Less	0.9327874	Probably Damaging
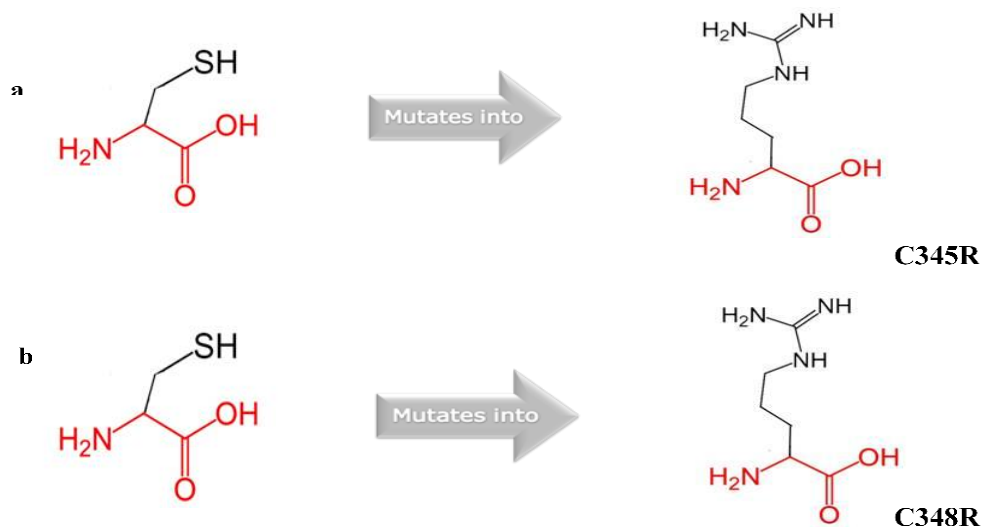
Figure 4.14 The figure above shows the schematic structures of the original (left) and the mutant (right) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black. a) Structure of C345R b) Structure of C348R

4.7.1 Protein Stability prediction

Structural analysis from hope project also that infered that the mutaions C345R AND C348R in KLF17 lead to disturbance in Zinc finger domain.The MetaRNN score of hope project for C345R and C348R was 0.9512116 and 0.9327874 respectively that declare both of the variant as highly pathogenic MetaRNN score can range from 0.0 to 1.0. The higher, the more likely it is to be pathogenic. Conservation in both the SNPs against wild type was observed to be very low declaring both the SNPs to be highly damaging.

4.8.Association of pathogenic SNPs with disease

A tool FATHMM was used to evaluate the association of pathogenic SNPs with Breast Cancer. Results of the Fathmm are in form of functional score which was applied to the

both pathogenic SNPs C345R and C348R and Fathmm declared both of the SNPs to have association with cancer and are responsible of the dysregulation in the structure and function of the protein.

Table 4.9. Estimated association of KLF17 SNPs with cancer using FATHMM

VARIANTS	PREDICTION	SCORE
C345R	CANCER	-1.16
C348R	CANCER	-5.50

4.9. Effects of SNPs on mRNA Secondary structure

mRNA Secondary structure of wild type KLF17 and SNPs were predicted using in-silico tool RNAFold. RNAFold calculated Minimum Free Energy (MFE) for both wild and SNPs. On comparing the results of RNAFold, secondary structure of the wild type mRNA Showed Slight changes for both the variants demonstrating the overall effect of variants alleles on structure of mRNA. MFE of wild type was calculated and compared with that of variant 1(C345R) MFE for wild type T C345 was -6.40 where for R345 it was -6.90 indicating a slight increase in MFE energy of variant 1 leading the mRNA Towards stability (fig 4.15b). Where MFE energy of R348 was calculated to be -7.70 indicating significant increase in MFE energy of variant as compared to that of Wild type resulting in a more stable mRNA Secondary Structure (fig 4.15c)
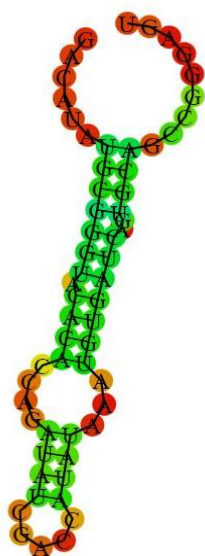
a. Wild Type   b.C345R   c.C348R
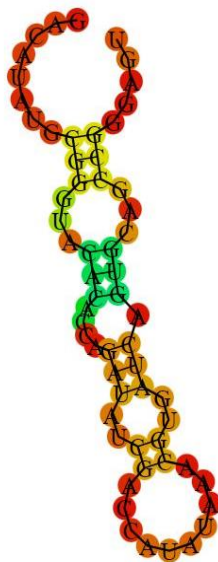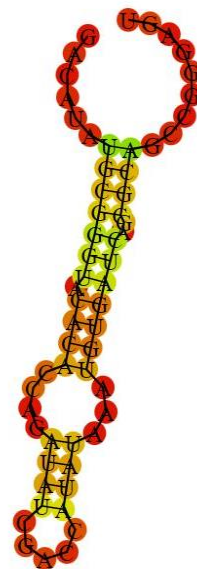
MFE = -6.40   **MFE= -6.90**   **MFE= -7.70**

Figure 4.15: Effect of SNPs on secondary structure of mRNA of KLF17 and change in the values of MFE as a result of variation

4.10. MD Analysis

Running MD simulations for both variants and wildtype generated various files. To analyze the result graphs were plotted for each file. Five characteristics: Root Mean Square Deviation, Root mean square fluctuations, radius of gyration, No of Hydrogen bonds and solvent accessible surface analysis were particularly used.

RMSD (Root mean square deviation) shows the variation of distinct atoms in a protein from its typical position. When compared with Wildtype, RMSD analysis of variants

showed that altered proteins varied greatly from the reference positions. The RMSD analysis was compared for over a period of 20ns plotted in graphs fig 4.16. Both the variants show variation within 2.0 nanoseconds, variant 2 follows the pattern at the end whereas variant 1 aligns with wildtype at 17.5nanoseconds till end. Largest deviation shown by variant 1 is 2.7nm at 14ns whereas for variant 2 is 2.4 nm at 20ns as shown in fig 4.16. Overall, the graph of RMSD for variant 1 is greater than wildtype indicating decreased stability whereas for variant2 graph is smaller indicating it to be more stable. The atypical peaks in graphs of both mutants and wildtype indicate presence of unknown loop structures deviating it significantly from established a-helixes and b-sheets.
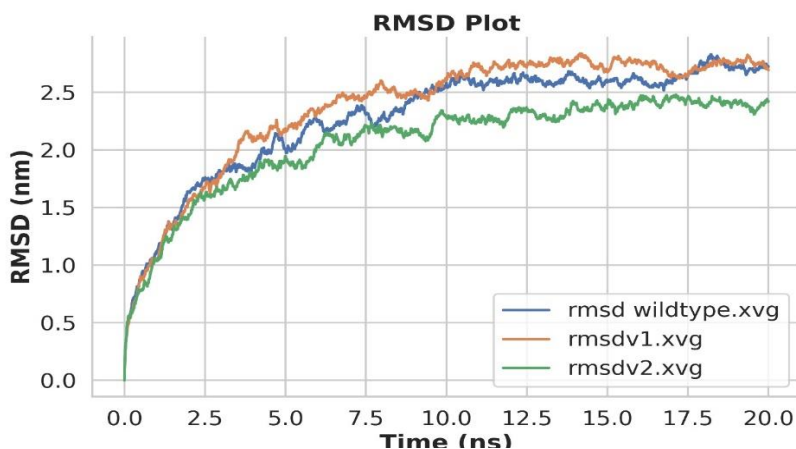

Figure 4.16 The RMSD graph shows the deviation in structures of wildtype protein and variants over a period of 20ns

RMSF (root mean square fluctuation) depicts the deviation in individual residues from their mean position. A higher graph of RMSF indicates a more flexible movements during simulation whereas a lower graph means restricted movements.to analyze the fluctuation

difference in variants and wildtype protein a graph was plotted as shown in fig 4.17. The peaks in graph indicate increased Kinetic Energy. The fluctuation in variants and wildtype range from 0.3-2.4. The variant 1 C345R shows the maximum fluctuation of 2.4 at region of 45-55 amino acids that lie in N-Terminal Domain as well as fluctuation of 2.3 around 140-150 amino acids while in last Zinc finger domain the peak gets lower than the wildtype. RMSF graph of variant 2(C348R) show lower peaks as compared to that of variant 1 and wildtype but at some residues greater fluctuations than the wildtype were observed i.e., 1.5-1.8 at 180-220 residues and at 390-400 residues it exceeds both the wildtype and variant 1 indicating flexible regions.

Consequently, it can be predicted that variant 1 C345R and variant 2 C348R are responsible for highest number of residual fluctuations in N terminal Domain which might contribute to effect on protein stability. The fluctuation of amino acid residing in the Zinc finger binding domain may affect the binding ability of proteins.
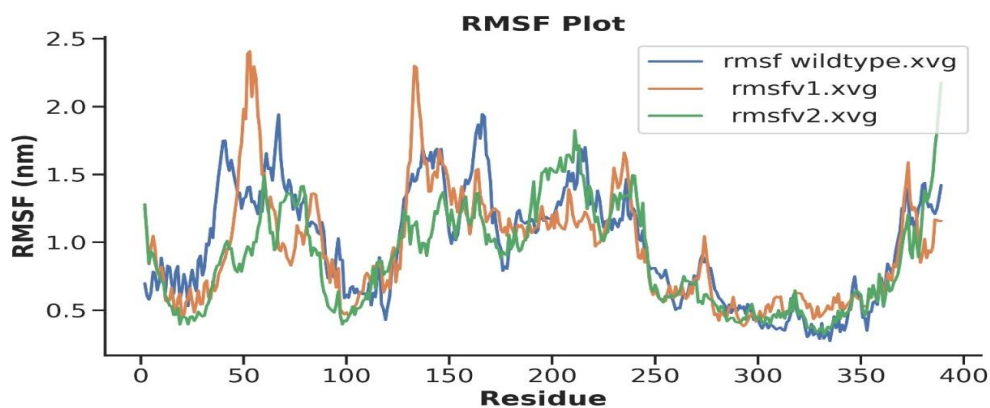
Figure 4.17 RMSF graph shows the total fluctuation difference in wildtype and variant residues

The radius of gyration (Rg) is the radial distance of all atoms in their proteins from a common axis. Rg is the radius of protein in a dynamic situation that indicates the proteins compression and the modifications in fold over the time. The evaluated radius of gyration of wildtype and mutant are depicted in the figure below. The radius of gyration for C345R was decreased at the beginning of simulation i.e., till 2.6nm at 5.0 ns and eventually increased up to 3.0 nm at 5.3 nm and then gradually decreased. At 20ns variant 1 showed radius of 2.5nm which was greater than both of the wildtype and variant 2. On the other hand, variant 2 started from 3.1 nm radius gradually decreased to 1.0 nm radius by the end of simulation gaining maximum compactness.

From this analysis it can be conferred that both the variants had lost their compactness and gradually gained it during simulation, with variant 2 having greater compactness than variant 1.
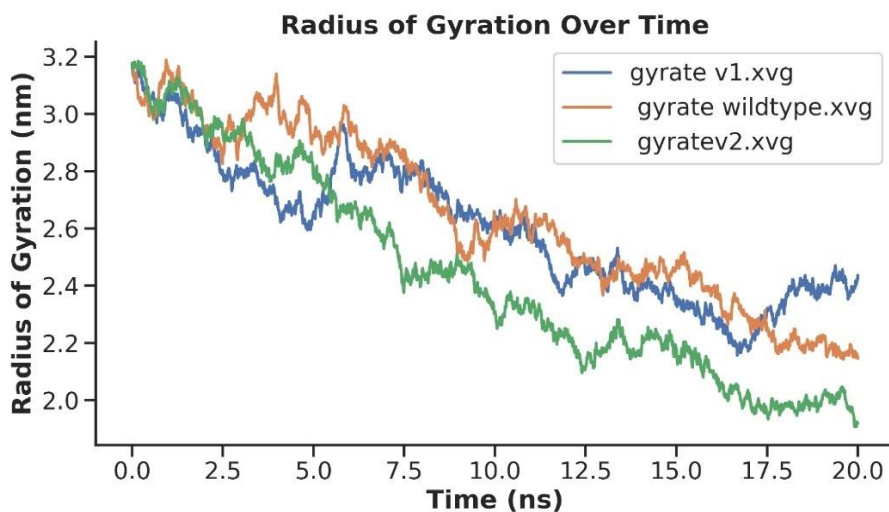
Figure 4.18 Radius of gyration of both the wildtype and variants

Hydrogen bonds develop between polar side chains of amino acid residues Graph fig4.19 depicts the amount of all hydrogen bonds formed in wildtype and both variant KLFs. The total number of H-bonds formed in Wildtype and variants are gradually increased as the simulation is run starting from 140 bonds, the variant 2 and wildtype gained up to 300 hydrogen bonds by the end of simulation whereas variant 1 had less hydrogen bonds of almost 270.
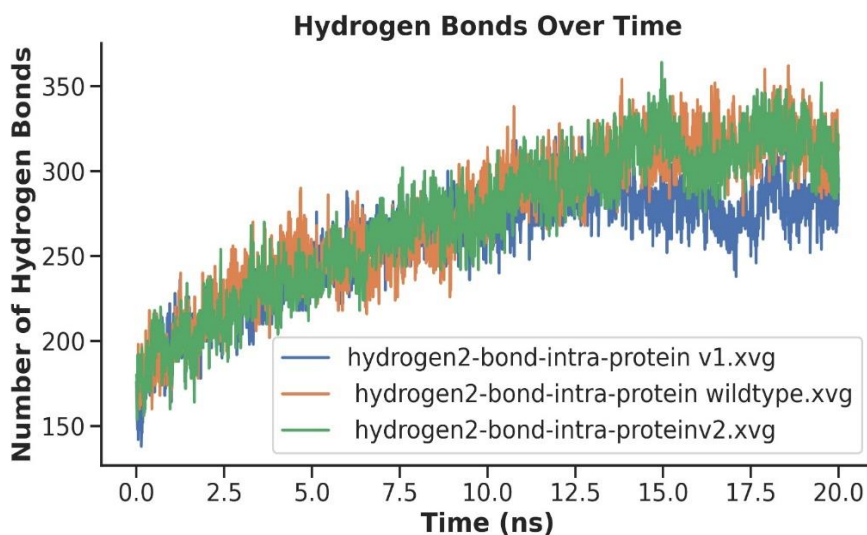
Figure 4.19 The difference in number of hydrogen bonds between wildtype and variants

To address the surface features of peptide or protein a simple technique of calculating the solvent accessible surface analysis is used which is basically investigating whether the protein is polar or non-polar or whether you want to differentiate between exposed and buried sections. SASA plot is shown in fig 4.20. According to the graph variant 1 has equal SASA 370nm2 at 10ns but it gets lower till 290 towards the end of simulation indicating a decrease in solvent accessible surface which makes protein stable . Hence variant 1

becomes more stable Where for variant 2 SASA is little higher as compared to wild type ranging from 480-340 nm2 indicating an expansion in solvent accessible surface hence making protein unstable.
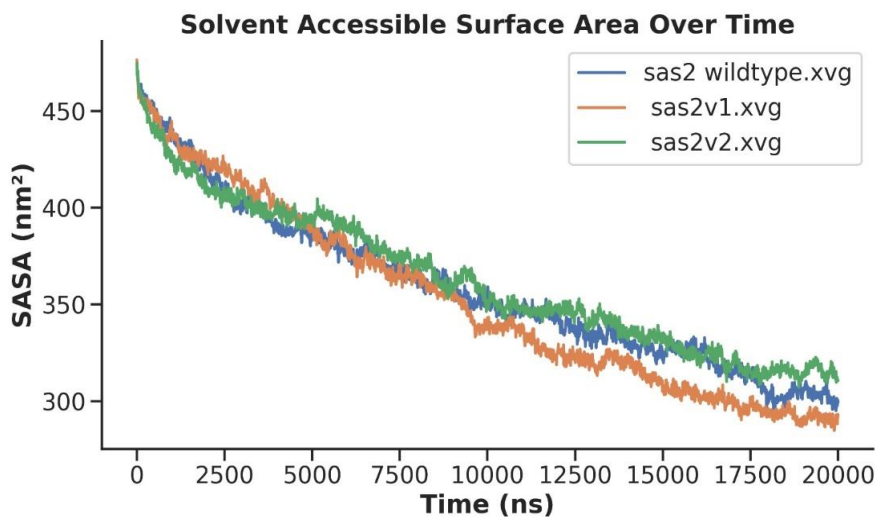
Figure 4.20 SASA analysis of wildtype and variant proteins

4.11. Laboratory Based Experimentation Analysis

4.11.1 Genotype Data of Cancer Patients and Healthy Control Samples

The result for breast cancer genotyping data analysis revealed that R345C SNP in the heterozygous wild form CT is statically significant for breast cancer with a P- value of 0.0010 and relative risk and odds ratio of 1.606 and 2.731 respectively and could be related to cancer progression and development when compared to homozygous mutant and homozygous wild form. The homozygous wild form shows association with breast cancer

but the odds ratio and relative risk being less than 1 indicate its role against the Cancer, where homozygous mutant CC with a P-value of 0.8146 i.e., >0.005 signifies no association with breast cancer. Table 4.10 shows the genotypic data of patients and controls.

Table 4.10: Genotype Data of Patient and Control of R345C (r) Mutation

Genotype	Patient %	Control %	Odds Ratio	95% CI –odds ratio	Relative Risk	95% CI-relative risk	p-Value
TT	38.83	64.08%	0.3559	0.2041 to 0.6269	0.5990	0.4461 to 0.7927	<0.05
CT	50.49%	27.18%	2.731	1.506 to 4.962	1.606	1.229 to 2.100	
CC	10.68	8.74	1.249	0.6788 to 1.570	1.112	0.5024 to 2.985	
C	35.58	22.33	1.921	1.034 to 3.559	1.353	1.022 to 1.750	0.0461
T	64.08	77.67	0.5128	0.2764 to 0.9541	0.7331	0.5658 to 0.9719	

4.11.2 Association of R345K (rs1331232028) SNP with Clinical Features of Breast Cancer

To further investigate the association R345C with breast cancer, the genotypic analysis of R345C with the clinical feature "TREATMENT" was performed. The analysis revealed that none of the genotypes in R345C showed any association with treated or untreated breast cancer as the p-value for all of them was P-value > 0.05.

Table 4.11: Genotypic dissemination of KLF17 SNP C345R correlated with the treatment level of breast cancer patients

Genotype	Treated	Untreated	Odds Ratio	95% CI –odds ratio	Relative Risk	95% CI- relative risk	p-Value
TT	41.56	33.33	1.422	0.5584 to 3.335	1.093	0.8541 to 1.371	0.4999
CT	49.35	51.85	0.9048	0.3974 to 2.270	0.9744	0.7669 to 1.234	0.5000
CC	9.09	14.81	0.5750	0.1577 to 1.893	0.8455	0.4657 to 1.171	0.4053

CHAPTER NO 5: DISCUSSION

Breast cancer is one of the most prevalent cancers to affect women across the globe that accounted for about 570,000 deaths in 2015. Every year over 1.5 million women are diagnosed with breast cancer; breast cancer is a highly metastatic cancer that can proliferate to distant organs explaining its deadliness. There are many risk factors associated with breast cancer that can upsurge the progression of breast cancer, including sex, ageing, estrogen's, family history, genetic mutations, and an unhealthy lifestyle. Early detection of the breast cancer may result in early treatment, good prognosis and an elevated survival rate. (Sun et al., 2017).

KLFs belong to family of serine/threonine kinases comprising 18 eukaryotic Zinc Finger transcription factors that may function as both activators and repressors in various biological activities (Pestal et al., 2024). KLFS are involved in tumor initiation, progression and elevation holding great importance in certain vital mechanisms such as cell proliferation, differentiation, cell death and metastatic property (Jianping et al., 2020). One such KLF Family member i.e. KLF17 which is a transcriptional regulatory protein also known as zinc finger 393 i.e. zfp393 comprising of 389 amino acids (Jane et al., 2006) located at 1p34.1 on chromosome (Wang et al., 2020), has been described as a tumor suppressor in various cancer as well as breast cancer through its interaction with TGFβ/SMAD3 as well as indirectly inhibiting EMT (Ali, Zhang, et al., 2015). And so, in this study, role of KLF17 in pathogenicity of breast cancer was evaluated through blood-

based genotyping that may pave wave towards a novel breast cancer biomarker hence, leading to early diagnosis and better treatment options. The main objective of this study is to identify the influence of Missense SNPs on the structure and function of KLF17 and identification of its Unique missense SNP's association with breast cancer pathogenicity and its clinical features.

To study the correlation of KLF17 and breast cancer it is crucial to find the missense SNPs of KLF17 that may be pathogenic. To analyze the effect of missense variants on KLF17, in silico tool, Alpha Fold was used to predict the structure of KLF17 through FASTA sequence of protein retrieved from Ensembl. The structure predicted was validated through Ramachandran Plot (via SAVES6.0) that yielded a total 91.9 percent residues to be present in allowed region whereas ERRAT (another tool of SAVES6.0) resulted in 94.068 percent quality factor for the predicted structure. To get more insights on the structure of the protein, a bioinformatics tool Interpro was used that resulted in detection of different zinc finger domains and sequence of the KLF17 protein. To get a valuable understanding of the location of KLF17, an in-silico tool Deeploc 1.0 was used and it is predominantly found in nucleus with a probability of 99.99 percent with a very low likelihood of 0.01 percent to be present in cytoplasm. Deeploc 1.0 also predicted KLF to be 71 percent soluble and 29 percent insoluble.

Phylogenetic analysis of all members of KLF family via MEGA 7 and ClustalW showed a close association among protein belonging to a common ancestor where KLF17 being

closest to KLF18 with a confidence value of 89 percent. While another in silico tool Consurf showed the kinase domain of KLF17 to be evolutionary conserved.

Because the goal of the study was to link single nucleotide polymorphisms (SNPs) to breast cancer A total of 5163 variants of KLF17 were retrieved through ensembl,503 Genome AD and 326 from cosmic. Missense variants, nonsense, frameshift and splice region variants were filtered out of all these databases and the numbers from the three databases were compared. The pathogenicity of the variants was studied through six computational tools including SIFT, PolyPhen, CADD, REVEL Mutation Accessor and MetaLR. A threshold of 70 percent was applied resulting in 5 missense variants. Of the 5 variants two variants R345C R348C with 90% pathogenicity, indicating higher probability of association with cancer, were selected for in silico analysis. Whereas wet lab analysis was performed on only one of these SNPs i.e., R345C.

To investigate the effect of SNPs on the stability of protein tools such as MUpro, maestro web and Dynamut were used which predicted a decrease in stability due to both the SNPs.to evaluate the structural and functional changes in protein due to variants tools like Mutpred, HOPE and DynaMut were used. Mutpred predicted altered disordered interface with a P value of 3.5e05 for variant R345C whereas for R348C predicted altered disordered interface with a P value of 4.4e03. DynaMut assessed the vibrational entropy energy between the wild-type and mutants to emphasize the change in protein flexibility caused by mutation, indicating a decrease in flexibility of molecule due to R345C hence destabilizing the protein whereas R348C SNP led to increase in flexibility of molecule

eventually resulting in unstable protein. Because the two residues arginine and cysteine differ in size, HOPE calculated that the mutation in that location would results in bulges in protein structure, disrupting the activity of the protein's zinc finer domain, causing the loss of activity and interactions with the substrate. HOPE project observed that the wildtype molecule was neutral while the mutant C345R gained negative charge and SNP C348R gained a positive as a result of the variation, due to the charges, loss in hydrophobicity was also noted in SNPs and hence hydrophobic interactions of proteins. From all of these predictions from MutPred2, DynaMut and HOPE it can be confered that impact of a Single amino acid variation is not restricted to just that residue but can also drastically effect the the structure and function of additional residues reversing the role of protein and promoting cancer development.

Molecular Dynamics Simulations were run at 20ns using GROMACS to observe the behavior of wildtype and mutants. The results of MD were analyzed through five characteristic graphs including RMSD, RMSF Radius of Gyration, SASA and No of hydrogen bonds that indicated Variant C345R to be more unstable than the Variant C348R and wildtype protein.

RMSD graph for variant C345R was higher than the wildtype one indicating unstable behavior of the variant protein. RMSF for variant one also depicted more fluctuations in the residues of variant C345R. Radius of gyration indicated variant C345R to be more flexible whereas variant C348R gained more compactness by the end of the simulations. No of Hydrogens for variant C345R were lower that the variantC348R and Wildtype by

the end of the simulations. According to SASA analysis, by the end of simulations, variant C345R had less surface exposed as compared to variant C348R, surface of which was more available for the interactions.

To validate the results of C345R of KLF17 genotypic analysis was performed in wet lab. For which, first Phenol-chloroform extraction of DNA from blood samples was performed. For genotypic analysis, Tetra ARMS PCR method was used to confirm the role and association of C345R variant of KLF17 with breast cancer. Genotypic data analysis depicted that TT and CT showed significant association with p-value of 0.0005 and 0.0010 respectively where CT have risk association with breast cancer (odds value 2.731 and relative risk 1.606) while TT the wild type genotype, has rather protective role against breast cancer (odds ratio 0.3559) and relative risk 0.5990). whereas the homozygous mutant genotype CC showed no correlation with breast cancer.

Various studies indicate the role of KLF17 in breast cancer. KLF17 also plays a critical role in regulation of numerous cellular processes such as erythropoiesis, cell survival, and the transformation of differentiated cells into stem cells. (Kiranmai Gumireddy et al., 2009). KLF17 is a negative regulator of epithelial to mesenchymal transition (EMT) and metastasis. Studies suggest that KLF17 plays significant role in inhibition of cancer cell invasion, migration, cell cycle progression, and development (Amjad et al., 2016).

This study suggests the association of KLF 17 with breast cancer and heterozygous C345R SNP to be involved in breast cancer development and progression and can be employed as a prognostic biomarker leading to the early detection of breast cancer. Furthermore, expression analysis of C345R SNP needs to be done that may lead to new paths in the field of Cancer therapy.

CONCLUSION

Breast cancer is one of the most prevalent cancers to affect women worldwide, which accounted for almost 570,000 deaths in 2015. Every year over 1.5 million women (or 25% of all women diagnosed with cancer) are diagnosed with breast cancer. Breast cancer is a heterogenous disease with higher potential of metastasis and recurrency.

SNPs are one of the most prevalent kinds of genetic changes in human genome that control DNA mismatch repair, cell cycle regulation, metabolism, and immunity are also linked to the cancer. Thus, to find novel pathogenic genetic variations associated with the incidence of breast cancer is vital, that may boost the progression of innovative diagnostic and therapeutic approaches, such as personalized medicine.

In this study, the most pathogenic variant of KLF17 i.e., rs183242786 was identified. Structure of KLF17 was predicted along with its structural domains and subcellular localization was predicted. Moreover, effect of missense variant on protein's structure and function was investigated. The result revealed that SNP led to disruption in stability and activity of KLF17.

To analyze the association between SNP and breast cancer as well as its influence on treatment genotyping of KLF17 was carried out. The results exposed a close association of heterozygous CT with progression of breast cancer and suggested that it can serve as a biomarker for early detection and diagnosis of the disease. Moreover, expression analysis of C345R SNP needs to be done that may lead to new paths in the field of Cancer therapy.

REFERENCES

1) Abe, M., Saeki, N., Ikeda, Y., & Ohba, S. (2022). Kruppel-like Factors in Skeletal Physiology and Pathologies. *International Journal of Molecular Sciences*, *23*(23), 15174. https://www.mdpi.com/1422-0067/23/23/15174

2) Ali, A., Bhatti, M. Z., Shah, A. S., Duong, H. Q., Alkreathy, H. M., Mohammad, S. F., Khan, R. A., & Ahmad, A. (2015). Tumor-suppressive p53 Signaling Empowers Metastatic Inhibitor KLF17-dependent Transcription to Overcome Tumorigenesis in Non-small Cell Lung Cancer. *J Biol Chem*, *290*(35), 21336-21351. https://doi.org/10.1074/jbc.M114.635730

3) Ali, A., Shah, A. S., & Ahmad, A. (2014). Gain-of-function of mutant p53: mutant p53 enhances cancer progression by inhibiting KLF17 expression in invasive breast carcinoma cells. *Cancer Lett*, *354*(1), 87-96. https://doi.org/10.1016/j.canlet.2014.07.045

4) Ali, A., Zhang, P., Liangfang, Y., Wenshe, S., Wang, H., Lin, X., Dai, Y., Feng, X. H., Moses, R., Wang, D., Li, X., & Xiao, J. (2015). KLF17 empowers TGF-β/Smad signaling by targeting Smad3-dependent pathway to suppress tumor growth and metastasis during cancer progression. *Cell Death Dis*, *6*(3), e1681. https://doi.org/10.1038/cddis.2015.48

5) Almagro Armenteros, J. J., Sønderby, C. K., Sønderby, S. K., Nielsen, H., & Winther, O. (2017). DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics*, *33*(21), 3387-3395. https://doi.org/10.1093/bioinformatics/btx431

6) Amjad, A., Irina, I., Huda Mohammad, A., & Abid Ali, K. (2016). KLF17 attenuates estrogen receptor α-mediated signaling by impeding ERα function on chromatin and determines response to endocrine therapy. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1859*(7), 883-895. https://doi.org/https://doi.org/10.1016/j.bbagrm.2016.04.009

7) 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

8) Breast Cancer. (2023). *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

9) Breast Cancer Statistics Worldwide. (2023). *Breast Cancer Statistics And Resources*. https://www.bcrf.org/breast-cancer-statistics-and-resources/

10) Britannica., T. E. o. E. (2024, May 27). single nucleotide polymorphism. In *Britannica*.

11) Cai, X. D., Che, L., Lin, J. X., Huang, S., Li, J., Liu, X. Y., Pan, X. F., Wang, Q. Q., Chen, L., Lin, M. J., Huang, Z. H., Ma, H. M., Wu, Y., Liu, S. M., & Zhou, Y. B. (2017). Krüppel-like factor 17 inhibits urokinase plasminogen activator gene expression to suppress cell invasion through the Src/p38/ MAPK signaling pathway in human lung adenocarcionma. *Oncotarget*, *8*(24), 38743-38754. https://doi.org/10.18632/oncotarget.17020

12) Cai, X. D., Zhou, Y. B., Huang, L. X., Zeng, Q. L., Zhang, L. J., Wang, Q. Q., Li, S. L., Feng, J. Q., & Han, A. J. (2012). Reduced expression of Krüppel-like factor 17 is related to tumor growth and poor prognosis in lung adenocarcinoma. *Biochem Biophys Res Commun*, *418*(1), 67-73. https://doi.org/10.1016/j.bbrc.2011.12.129

13) CANCER. *WHO* https://www.who.int/health-topics/cancer#tab=tab_1

14) CANCER. (2020). *World Health Organization*. https://www.who.int/news-room/fact-sheets/detail/cancer

15) CANCER. (2022). *cleveland clinic*. https://my.clevelandclinic.org/health/diseases/12194-cancer

16) cancer statistics at a glance. (2020). *Centers for disease control and prevention*. https://gis.cdc.gov/Cancer/USCS/#/AtAGlance/

17) Cancer Stats: Female breast cancer. (2023). *National Cancer Institue*. https://seer.cancer.gov/statfacts/html/breast.html

18) *CIB3 GENE*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CIB3

19) Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*, *2*(9), 1511-1519. https://doi.org/10.1002/pro.5560020916

20) 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

21) Dong, P., Kaneuchi, M., Xiong, Y., Cao, L., Cai, M., Liu, X., Guo, S. W., Ju, J., Jia, N., Konno, Y., Watari, H., Hosaka, M., Sudo, S., & Sakuragi, N. (2014). Identification of KLF17 as a novel epithelial to mesenchymal transition inducer via direct activation of TWIST1 in endometrioid endometrial cancer. *Carcinogenesis*, *35*(4), 760-768. https://doi.org/10.1093/carcin/bgt369

22) DUCTAL CARCINOMA IN SITU. (2022). *NATIONAL BREAST CANCER FOUNDATION,INC*. https://www.nationalbreastcancer.org/dcis/

23) Gao, S. P., Sun, H. F., Li, L. D., Fu, W. Y., & Jin, W. (2017). UHRF1 promotes breast cancer progression by suppressing KLF17 expression by hypermethylating its promoter. *Am J Cancer Res*, *7*(7), 1554-1565.

24) Global cancer data by country (2020). *WCRF International*. https://www.wcrf.org/cancer-trends/global-cancer-data-by-country/#:~:text=Globally%2C%2018%2C094%2C716%20million%20cases%20of,women%20(178.1%20per%20100%2C000)

25) Gumireddy, K., Li, A., Gimotty, P. A., Klein-Szanto, A. J., Showe, L. C., Katsaros, D., Coukos, G., Zhang, L., & Huang, Q. (2009). KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. *Nat Cell Biol*, *11*(11), 1297-1304. https://doi.org/10.1038/ncb1974

26) Gumireddy, K., Li, A., Gimotty, P. A., Klein-Szanto, A. J., Showe, L. C., Katsaros, D., Coukos, G., Zhang, L., & Huang, Q. (2009). KLF17 is a negative regulator of epithelial–mesenchymal transition and metastasis in breast cancer. *Nature Cell Biology*, *11*(11), 1297-1304. https://doi.org/10.1038/ncb1974

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

28) Hu, W., Yan, F., Ru, Y., Xia, M., Yan, G., Zhang, M., Wang, H., Wu, G., Yao, L., Shen, L., Li, X., & Wang, Q. (2020). MIIP inhibits EMT and cell invasion in prostate cancer through miR-181a/b-5p-KLF17 axis. *Am J Cancer Res*, *10*(2), 630-647.

29) INVASIVE DUCATL CANCER *NATIONAL BREAST CANCER FOUNDATION,INC*. https://www.nationalbreastcancer.org/invasive-ductal-carcinoma/

30) INVASIVE LOBULAR CANCER (ILC). (2023). *NATIONAL BREAST CANCER FOUNDATION,INC*. https://www.nationalbreastcancer.org/invasive-lobular-cancer/

31) Ismail, I. A., Kang, H. S., Lee, H. J., Kim, J. K., & Hong, S. H. (2014a). DJ-1 upregulates breast cancer cell invasion by repressing KLF17 expression. *British Journal of Cancer*, *110*(5), 1298-1306. https://doi.org/10.1038/bjc.2014.40

32) Ismail, I. A., Kang, H. S., Lee, H. J., Kim, J. K., & Hong, S. H. (2014b). DJ-1 upregulates breast cancer cell invasion by repressing KLF17 expression. *Br J Cancer*, *110*(5), 1298-1306. https://doi.org/10.1038/bjc.2014.40

33) Jane, Linda, A. C., Kate, G. R. Q., Robert, C., Andrew, C. P., & Merlin, C. (2006). Human KLF17 is a new member of the Sp/KLF family of transcription factors. *Genomics*, *87*(4), 474-482. https://doi.org/https://doi.org/10.1016/j.ygeno.2005.12.011

34) Jianping, Z., Guangliang, L., Lifeng, F., Haiqi, L., & Xian, W. (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

35) Kin-Sang, C., Chenying, G., Justin, C., Juliet, C. Y., Ruilin, Z., Zhigang, H., & Dong Feng, C. (2015). Chapter 12 - The Intrinsic Determinants of Axon

Regeneration in the Central Nervous System. In S. Kwok-Fai & X. Xiao-Ming (Eds.), *Neural Regeneration* (pp. 197-207). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-801732-6.00012-4

36) *KLF17 GENE*. https://www.genecards.org/cgi-bin/carddisp.pl?gene=KLF17#pathways_interactions

37) *KLF17 Gene*. (Aug 2, 2023). https://www.genecards.org/cgi-bin/carddisp.pl?gene=KLF17#localization

38) *KLF17 interactions*. https://string-db.org/network/9606.ENSP00000361373

39) *KLF 17 INTERACTIONS*.

40) Koren, S., & Bentires-Alj, M. (2015). Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. *Mol Cell*, *60*(4), 537-546. https://doi.org/10.1016/j.molcel.2015.10.031

41) Kotlyarov, S., & Kotlyarova, A. (2023). Participation of Krüppel-like Factors in Atherogenesis. *Metabolites*, *13*(3). https://doi.org/10.3390/metabo13030448

42) Li, L., Zhang, C., Li, Y., Zhang, Y., & Lei, Y. (2020). DJ-1 promotes epithelial-to-mesenchymal transition via enhancing FGF9 expression in colorectal cancer. *Biol Open*, *9*(5). https://doi.org/10.1242/bio.051680

43) Li, Y., Zhao, X., Xu, M., & Chen, M. (2022). Krüppel-like factors in glycolipid metabolic diseases. *Molecular Biology Reports*, *49*(8), 8145-8152. https://doi.org/10.1007/s11033-022-07565-0

44) Lin, Y., Chen, Q., Liu, Q. X., Zhou, D., Lu, X., Deng, X. F., Yang, H., Zheng, H., & Qiu, Y. (2018). High expression of DJ-1 promotes growth and invasion via the PTEN-AKT pathway and predicts a poor prognosis in colorectal cancer. *Cancer Med*, *7*(3), 809-819. https://doi.org/10.1002/cam4.1325

45) Liu, F. Y., Deng, Y. L., Li, Y., Zeng, D., Zhou, Z. Z., Tian, D. A., & Liu, M. (2013). Down-regulated KLF17 expression is associated with tumor invasion and poor prognosis in hepatocellular carcinoma. *Med Oncol*, *30*(1), 425. https://doi.org/10.1007/s12032-012-0425-3

46) Lydia Choi (Sep 2022). Breast Cancer. *MERCK MANUAL.* https://www.merckmanuals.com/home/women-s-health-issues/breast-disorders/breast-cancer

47) 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

48) Muller, P. A., & Vousden, K. H. (2014). Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell*, *25*(3), 304-317. https://doi.org/10.1016/j.ccr.2014.01.021

49) Peng, J. J., Wu, B., Xiao, X. B., Shao, Y. S., Feng, Y., & Yin, M. X. (2014). Reduced Krüppel-like factor 17 (KLF17) expression correlates with poor survival in patients with gastric cancer. *Arch Med Res*, *45*(5), 394-399. https://doi.org/10.1016/j.arcmed.2014.06.005

50) Pestal, K., Slayden, L. C., & Barton, G. M. (2024). Krüppel-like Factor (KLF) family members control expression of genes required for serous cavity and alveolar macrophage identities. *bioRxiv*. https://doi.org/10.1101/2024.02.28.582578

51) Pollak, N. M., Hoffman, M., Goldberg, I. J., & Drosatos, K. (2018). Krüppel-like factors: Crippling and un-crippling metabolic pathways. *JACC Basic Transl Sci*, *3*(1), 132-156. https://doi.org/10.1016/j.jacbts.2017.09.001

52) Richard, P., Jacqueline, F., Sally, E., Merlin, C., & Shisan, B. (2008). Krüppel-like transcription factors: A functional family. *The International Journal of Biochemistry & Cell Biology*, *40*(10), 1996-2001. https://doi.org/https://doi.org/10.1016/j.biocel.2007.07.018

53) Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., & Sharma, K. K. (2010). Various types and management of breast cancer: an overview. *J Adv Pharm Technol Res*, *1*(2), 109-126.

54) SNP. (2014). *scitable by Nature Education*. https://www.nature.com/scitable/definition/snp-295/

55) Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., Shi, W., Jiang, J., Yao, P. P., & Zhu, H. P. (2017). Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci*, *13*(11), 1387-1397. https://doi.org/10.7150/ijbs.21635

56) team, T. A. C. S. m. a. e. c. (2022). what is cancer *American Cancer Society*. https://www.cancer.org/cancer/understanding-cancer/what-is-cancer.html

57) Tetreault, M. P., Yang, Y., & Katz, J. P. (2013). Krüppel-like factors in cancer. *Nat Rev Cancer*, *13*(10), 701-713. https://doi.org/10.1038/nrc3582

58) Triple-Negative Breast Cancer. (2023). *Centers for disease control and prevention*. https://www.cdc.gov/cancer/breast/triple-negative.htm

59) Types of breast cancer. (2010). *Rethink Breast Cancer*. http://www.rethinkbreastcancer.com/types_of_breast_cancer.html .

60) Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *J Comput Chem*, *26*(16), 1701-1718. https://doi.org/10.1002/jcc.20291

61) Wang, C., Wang, Z., He, M., Zhou, T., Niu, Y., Sun, S., Li, H., Zhang, C., Zhang, S., Liu, M., Xu, Y., Dong, N., & Wu, Q. (2020). Krüppel-like factor 17 upregulates uterine corin expression and promotes spiral artery remodeling in pregnancy. *Proc Natl Acad Sci U S A*, *117*(32), 19425-19434. https://doi.org/10.1073/pnas.2003913117

62) what is breast cancer (2010, 18 March 2010). *American Cancer Society* http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_is_breast_cancer_5.asp .

63) what is breast cancer. (2010). *imaginis*. http://www.imaginis.com/breast-health/what-is-breast-cancer-2

64) What is Breast Cancer? (2009). *American Cancer Society*. https://www.cancer.org/cancer/types/breast-cancer/about/what-is-breast-cancer.html

65) Ye, W. C., Gao, L., Huang, J., Fang, X. M., & Xie, G. (2014). Suppressed Krüppel-like factor 17 expression induces tumor proliferation, metastasis and a

poor prognosis in papillary thyroid carcinoma. *Mol Med Rep*, *10*(4), 2087-2092. https://doi.org/10.3892/mmr.2014.2429

66) Zaheer, S., Shah, N., Maqbool, S. A., & Soomro, N. M. (2019). Estimates of past and future time trends in age-specific breast cancer incidence among women in Karachi, Pakistan: 2004–2025. *BMC Public Health*, *19*(1), 1001. https://doi.org/10.1186/s12889-019-7330-z

67) Zhang, Y., Yao, C., Ju, Z., Jiao, D., Hu, D., Qi, L., Liu, S., Wu, X., & Zhao, C. (2023). Krüppel-like factors in tumors: Key regulators and therapeutic avenues. *Front Oncol*, *13*, 1080720. https://doi.org/10.3389/fonc.2023.1080720

68) Zhou, J., Liu, H., Zhang, L., Liu, X., Zhang, C., Wang, Y., He, Q., Zhang, Y., Li, Y., Chen, Q., Zhang, L., Wang, K., Bu, Y., & Lei, Y. (2018). DJ-1 promotes colorectal cancer progression through activating PLAGL2/Wnt/BMP4 axis. *Cell Death & Disease*, *9*(9), 865. https://doi.org/10.1038/s41419-018-0883-4

69) Zhou, S., Tang, X., & Tang, F. (2016). Krüppel-like factor 17, a novel tumor suppressor: its low expression is involved in cancer metastasis. *Tumor Biology*, *37*(2), 1505-1513. https://doi.org/10.1007/s13277-015-4588-3

70) Zhou, S., Tang, X., & Tang, F. (2016). Krüppel-like factor 17, a novel tumor suppressor: its low expression is involved in cancer metastasis. *Tumour Biol*, *37*(2), 1505-1513. https://doi.org/10.1007/s13277-015-4588-3 .

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:	Asma Areesh
Assignment title:	Quick Submit
Submission title:	Evaluating the Role of KLF17 in pathogenisis of Breast Cancer
File name:	ASMA_AREESH_THESIS_Draft.docx
File size:	122.63K
Page count:	38
Word count:	10,216
Character count:	54,432
Submission date:	04-Jul-2024 10:35PM (UTC-0700)
Submission ID:	2412718379

Evaluating the Role of KLF17 in pathogenisis of Breast Cancer

By

ASMA AREESH

00000-801316

Masters of science in Healthcare Biotechnology

Supervisor

Dr. Maria Shabbir

Co Supervisor

Dr. Yasmin Badshah

Department of Healthcare Biotechnology
ATTA-UR-Rehman School of Applied Biosciences
National University of Sciences and Technology
Islamabad, Pakistan
2024

Evaluating the Role of KLF17 in pathogenisis of Breast Cancer

mammalian stem cells and development",
Development, 2017
Publication

20 **Submitted to Santa Monica College**
Student Paper
<1 %

21 Syed Umair Ahmad, Yasir Ali, Zainab Jan,
Salman Rasheed et al. " Computational
screening and analysis of deleterious nsSNPs
in human 14ARF (gene) protein using
molecular dynamic simulation approach ",
Journal of Biomolecular Structure and
Dynamics, 2022
Publication
<1 %

22 elib.uni-stuttgart.de
Internet Source
<1 %

23 mccann.us.to
Internet Source
<1 %

24 **Submitted to London Metropolitan University**
Student Paper
<1 %

25 hdl.handle.net
Internet Source
<1 %

26 journal.waocp.org
Internet Source
<1 %

27 **Submitted to Texas A & M University, Kingville**
Student Paper
<1 %

www.science.gov

28
Internet Source

<1%

29 krishikosh.egranth.ac.in
Internet Source

<1%

30 link.springer.com
Internet Source

<1%

31 Submitted to Jawaharlal Nehru University (JNU)
Student Paper

<1%

32 Submitted to Callaghan Campus
Student Paper

<1%

33 Submitted to Curtin University of Technology
Student Paper

<1%

34 Submitted to Taylor's Education Group
Student Paper

<1%

35 bmccancer.biomedcentral.com
Internet Source

<1%

36 www.thermofisher.com
Internet Source

<1%

37 vdoc.pub
Internet Source

<1%

38 www.cancer.net
Internet Source

<1%

39 www.dailymail.co.uk
Internet Source

<1%

40 Kemal Yuce, Ahmet Ismail Ozkan. "The Kruppel-like Factor (KLF) Family, Diseases, and Physiological Events", Gene, 2023
Publication
<1%

41 Sayed El Azhar, R. Verhe, M. Proot, P. Sandra, W. Verstraete. "Binding of nitrite-N on polyphenols during nitrification", Plant and Soil, 1986
Publication
<1%

42 eprints.usm.my
Internet Source
<1%

43 www.researchandmarkets.com
Internet Source
<1%

44 studentsrepo.um.edu.my
Internet Source
<1%

45 www.amhsr.org
Internet Source
<1%

46 www.frontiersin.org
Internet Source
<1%

47 Submitted to Association of Educators
Student Paper
<1%

48 Diaz-Herrera, Diana Mabel. "Methods to Integrate Genetic and Clinical Data for
<1%

Disease Subtyping.", Wayne State University, 2020
Publication

49   Jaclyn A. Biegel, Tracy M. Busse, Bernard E. Weissman. "SWI/SNF chromatin remodeling complexes and cancer", American Journal of Medical Genetics Part C: Seminars in Medical Genetics, 2014
Publication

<1%

50   diglib.tugraz.at
Internet Source

<1%

51   healthdocbox.com
Internet Source

<1%

52   ugspace.ug.edu.gh
Internet Source

<1%

53   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
Publication

<1%