Mechanistic Impact of KLF15 Polymorphism and its association with Liver Cancer

By

Humaira Bibi

(Registration No: 00000399991)

Department of Biomedicine

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

(2024)

Mechanistic Impact of KLF15 Polymorphism and its association with Liver Cancer

By

Humaira Bibi

(Registration No: 00000399991)

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

in partial fulfillment of the requirements for the degree of

Master of Science in
Molecular Medicine

Supervisor: Dr. Yasmin Badshah

Co Supervisor: Dr. Maria Shabbir

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan (2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS Thesis written by Mr / Ms_____Humaira Bibi_____

(Registration No. 00000399991), of _ASAB_ (School/College/Institute) 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: _____
Dr. Yasmin Badshah
Assistant Professor
Atta-ur-Rahman School of
Applied Biosciences (ASAB
NUST Islamabad

Name of Supervisor __Dr Yasmin Badshah__

Date:____8/8/24_____

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

Date: __08/08/2024_____

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

Date: ____8/8/24_____

National University of Sciences & Technology

MS THESIS WORK

FORM TH-4

We hereby recommend that the dissertation prepared under our supervision by:

(Student Name & Regn No.) _____ Humaira Bibi, Regn No. **00000399991**

Titled: "Mechanistic Impact of Klf-15 polymorphism and its association with liver cancer " be accepted in partial fulfillment of the requirements for the award of _____ MS Molecular Medicine _____ degree with (A grade).

Examination Committee Members

1. Name: Dr. Rumeza Hanif   Signature: _____

2. Name: Dr. Peter John   Signature: _____

3. Name: Dr. Hashaam Akhtar   Signature: _____
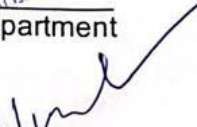
Supervisor's name: Dr. Yasmin Badshah   Signature: _____

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

Date: 24 - July - 2024

Dr. Hussain Mustahsan Shahab Waheed (HoD)
Head of Department Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

Head of Department

Date : 08/08/2024

COUNTERSINGED

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

Dean / Principal

AUTHOR'S DECLARATION

I .... Humaira Bibi .... hereby state that my MS thesis titled " .... Mechanistic impact of KLF15 polymorphism and its association with liver cancer .... " is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

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

Name of Student: __Humaira   bibi__

Date: _____8/08/24_____

This thesis has been dedicated to my beloved Father, who with love and effort has

accompanied me in this process, I would not be here without you.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor Dr Yasmin badshah, for her invaluable guidance, support, and encouragement throughout this research journey. I am immensely grateful to my co-supervisor, Dr Maria shabbir, for her insightful feedback and constructive critiques. Special thanks to my colleagues and friends, whose camaraderie and support were indispensable during the challenging moments. I also extend my heartfelt appreciation to my family, whose unwavering belief in me provided the foundation for my academic pursuits.

TABLE OF CONTENTS

LIST OF TABLES

Page No.

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS AND ACRONYMS

KLF15: Krüppel-Like Factor 15

HCC: Hepatocellular Carcinoma

CCA: Cholangiocarcinoma

ARMS-PCR: Amplification Refractory Mutation System - Polymerase Chain Reaction

SNP: Single Nucleotide Polymorphism

CADD: Combined Annotation Dependent Depletion

SIFT: Sorting Intolerant From Tolerant

REVEL: Rare Exome Variant Ensemble Learner

TFAP2A-AS1: Transcription Factor Activating Protein 2a - Antisense Rna 1

LUCAT1: Lung Cancer Associated Transcript 1

ABSTRACT

As liver cancer ranks among the top cancer-related fatalities worldwide, prognostic indicators and treatment targets for this disease need to be improved. This study attempts to ascertain whether the Krüppel-like factor 15 (KLF15) gene and its variants are linked to liver cancer through use of in-silico and lab-based analysis. Using in silico techniques including SIFT, PolyPhen, CADD, REVEL, Mutation Assessor, MetaLR, MutPred2, DynaMut, MUpro, MAESTRO, Project HOPE, and FATHMM, two pathogenic missense variants rs7557194 and rs7686768 were identified. Structural and functional analyses suggest that these variants most likely disrupt the protein's structure and function. Following statistical analysis and tetra ARMS-PCR investigation, it became apparent that in a population from Pakistan the rs7557194 variant was notably linked to liver cancer. The study underlines that a target for treatment and a marker for liver cancer diagnosis could be the R364P mutation of KLF-15. These findings show the potential for the use of the KLF15 R364P mutation as a liver cancer biomarker, therefore enabling more precise treatments and improved diagnostics. Future research with larger sample sizes and different populations should investigate the functional mechanisms of KLF15 variants in liver cancer progression utilizing in vitro and in vivo studies, therefore validating our results.

Keywords: Liver cancer, Krüppel-like factor 15 (KLF15), rs755719419, rs768676875, in silico analysis, SIFT, PolyPhen, CADD, REVEL, Mutation Assessor, MetaLR, AlphaFold,

MutPred2, DynaMut, MUpro, MAESTROweb, Project HOPE, FATHMM, tetra ARMS-PCR, genetic testing, diagnostic marker, therapeutic target.

1. INTRODUCTION

1.1 Cancer

The extracellular matrix represents a vital and integral element within all tissues and organs, playing a crucial role in the functioning of multicellular organisms. Its influence spans from the initial phases of organism development to the final stages of life, overseeing and refining various cellular functions within the body. In the context of cancer, alterations occur within the extracellular matrix at the biochemical, biomechanical, architectural, and topographical levels. Recent years have witnessed a notable surge in research and acknowledgment of the matrix's significance in solid tumor studies (Cox, 2021). The destructive agents of cancer are human cells that have been enlisted and somewhat altered into pathological entities or the foundational units of tumors. Cancers disrupt and capitalize on the mechanisms of multicellular structure, leading to complex philosophical dilemmas when seeking to comprehend them (Hausman, 2019).

1.2 Types of cancer

In a wider context, cancer encompasses over 277 various types of disease. Researchers have distinguished multiple stages of cancer, revealing that numerous gene mutations contribute to the development of cancer. These mutations result in atypical cell growth. Genetic disorders triggered by hereditary factors play a crucial part in the escalation of cell proliferation (Hassanpour & Dehghani, 2017). Among men, the most prevalent cancer types are found in the prostate, lung and bronchus, colon and rectum, and urinary bladder. Conversely, in women, the highest rates of cancer are observed in the breast, lung and bronchus, colon and rectum, uterine corpus, and thyroid. These statistics highlight the significant presence of prostate and breast cancer among men and women, respectively (Siegel et al., 2016). For children, the predominant types of cancer are those affecting the blood, as well as cancers linked to the brain and lymph nodes (Schottenfeld & Fraumeni Jr, 2016).

1.3 Incidence of cancer

The study by (Sung et al., 2021) provides a recent overview of the global cancer situation, based on the GLOBOCAN 2020 evaluations conducted by the International Agency for Research on Cancer. In the previous year, there were approximately 19.3 million new instances of cancer worldwide (excluding nonmelanoma skin cancer), resulting in nearly 10.0 million deaths (excluding nonmelanoma skin cancer). Female breast cancer has surpassed lung cancer as the most often diagnosed type, with over 2.3 million new cases (11.7%). Lung cancer follows closely behind with 11.4% of new cases, followed by colorectal cancer (10.0%), prostate cancer (7.3%), and stomach cancer (5.6%). Despite improvements in knowledge and medical care, lung cancer remained the leading cause of cancer-related deaths, accounting for around 1.8 million fatalities (18%). This was followed by colorectal cancer (9.4%), liver cancer (8.3%), stomach cancer (7.7%), and female breast cancer (6.9%). Significantly, the incidence rates were considerably higher in developed countries compared to underdeveloped countries, for both males and females. However, there was less variation in fatality rates between genders. By 2040, it is estimated that there will be a significant increase in the global cancer burden, with a predicted total of 28.4 million cases. This represents a 47% jump compared to the number of cases in 2020. The anticipated rise is projected to be more significant in nations undergoing transition (from 64% to 95%) compared to countries that have already transitioned (from 32% to 56%). This is primarily attributed to demographic changes and the widespread increase in risk factors linked to globalization and economic expansion (Sung et al., 2021).

1.4 Therapeutic resistance in cancer

Significantly, the genetic makeup of tumors often interlinks with non-genetic ways utilized to resist treatment (Marine et al., 2020). Notably, certain cancer-causing mutations enable the cells to swiftly adjust their transcriptional and metabolic schemes to withstand and survive the pressures of therapy. Conversely, other oncogenic drivers confer an inherent

2

adaptability to the cancer cell, facilitating lineage switching and/or evading the body's anticancer immune surveillance. The widespread occurrence and diverse spectrum of non-genetic resistance mechanisms present a new challenge for the field, demanding innovative approaches to monitor and counter these adaptive processes. In this Perspective, we investigate the fundamental concepts of cancer therapeutic resistance that is not hereditary. Tumors develop resistant to antiangiogenic treatment by altering several biological processes including invasiveness, metastases, stemness, autophagy, and metabolic reprogramming (Marine et al., 2020). Among these pathways are VEFGs, GM-CSF, G-CSF, and IL17. A hallmark of this resistance is increasing mobilization of bone marrow-derived cells including myeloid-derived suppressive cells, tumor-associated macrophages, and tumor-associated neutrophils. Among local stromal cells whose function is changed include pericytes, cancer-associated fibroblasts, and endothelial cells. Furthermore, interesting biomarkers for antiangiogenic treatment response prediction, among these biomarkers include imaging biomarkers, certain proteins, bone marrow-derived cells, stromal cell composition, non-coding RNAs in serum and plasma (Nisar et al., 2022).

1.5 Diagnosis of cancer

The application of cancer diagnosis techniques offers benefits in accurately identifying the exact location of the cancer site and distinguishing the boundary between the diseased and healthy tissue regions. This facilitates precise cancer diagnosis. Fluorescence imaging technology utilizes small-molecule fluorescent probes to enable visual diagnosis. The technology provides a non-invasive approach, with a high level of sensitivity and improved resolution in both space and time. Several fluorescent probes have been developed to identify biomarkers, differentiate between malignant and normal cells/tissues, and facilitate the imaging of solid tumors. Given this, it specifically examines the design strategies and uses of these small-molecule fluorescence probes in the diagnosis of cancer. In addition, assemblage and identification of biomarkers that are appropriate for cancer diagnosis, emphasizing their important functions in medical imaging (Wang et al., 2023). Only a few notable advancements have been made in the treatment of cancer, but much is

already known about the causes of, prevention from, and management of the disease. Early detection is clearly the only approach to raise life expectancy and quality of living. To detect circulating tumor cells (CTCs), biomarkers, or tumor-derived vesicles released rather early from the tumor into the blood which will significantly change the morbidity and mortality of the disease, Nano sensors and highly sensitive nanomaterial-based devices have been developed (Khazaei et al., 2023). According to the World Health Organization (WHO), liver cancer is the sixth most common cancer worldwide (WHO, 2024).

Absolute numbers, Incidence, Both sexes, in 2022
World
(Top 15 cancer sites)
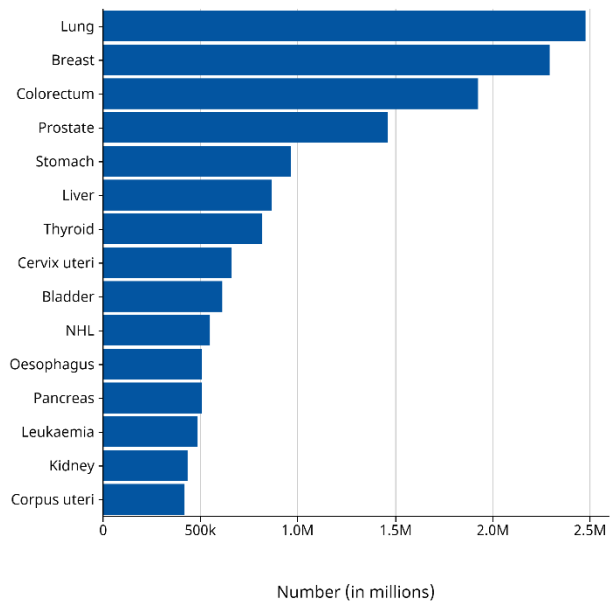
Cancer TODAY | IARC - https://gco.iarc.who.int/today
Data version : Globocan 2022 (version 1.1)
© All Rights Reserved 2024

International Agency
for Research on Cancer
World Health
Organization

Figure 1: Incidence of cancer

1.6 Liver cancer

Hepatic fibrogenesis is a result of chronic liver injury and is characterized by an abnormal accumulation of proteins in the extracellular matrix. Hepatocytes, which are the main functional cells of the liver, interact with other types of cells such as hepatic stellate cells,

sinusoidal endothelial cells, and immune cells (both resident and invading) during the complex process of fibrosis, the review focuses on the key cellular components responsible for liver fibrosis, the specific cytokines and chemokines that influence its progression, and the circumstances in which fibrosis might reverse, the regulation of the fibro genic process in hepatic stellate cells by mitochondria and metabolic changes. In addition, there is an impact of fibrosis on the development of hepatocellular carcinoma (Dhar et al., 2020).

In this study Rizvi et al. (2021) , it is stated that the fundamental processes influencing the progression of liver fibrosis, including its associated complications such as cirrhosis, portal hypertension, and liver failure, often necessitating liver transplantation. Additionally, it underscore the significant contribution of hepatic stellate cells (HSCs) in the development of liver fibrosis. Emphasizing the role of oxidative stress, mitochondrial dysfunction, and metabolic reconfiguration, it is highlighted that how these factors impact HSC activation, thereby indicating potential avenues for therapeutic intervention (Rizvi et al., 2021).

1.7 Biomarkers and genetic alterations in liver cancer

Biomarkers play a crucial role in the effective clinical management of cancer patients, significantly contributing to enhanced survival rates and the optimization of medical interventions (refer to Figure 2). Within the context of hepatocellular carcinoma (HCC), there is an urgent demand for biomarkers in the following clinical domains: (1) the categorization of risk and early detection of HCC, (2) the prediction of prognosis, and (3) the anticipation of systemic therapy response. Conversely, in cholangiocarcinoma (CCA), several unmet requirements persist, including: (1) improved diagnosis of CCA, particularly in individuals with primary sclerosing cholangitis; (2) enhanced prognostic capabilities post-curative treatment to guide adjuvant therapy; and (3) the identification of biomarkers that can predict responses and resistance to first and second-line systemic treatments (Nault & Villanueva, 2021).
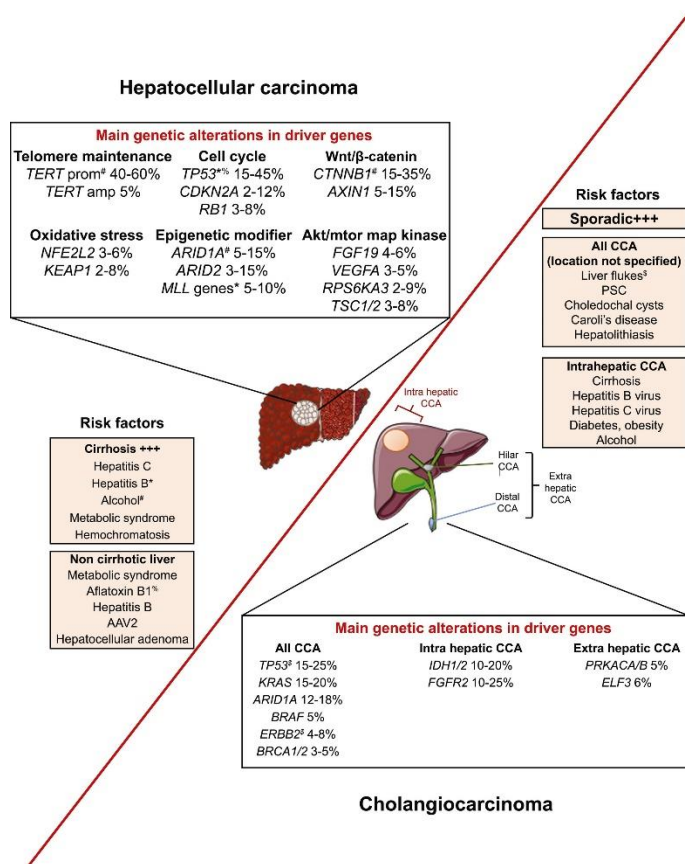
Figure 2: The primary somatic alterations in key driver genes within hepatobiliary cancers, specifically hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA), were summarized, establishing their association with risk factors implicated in tumor. Source :(Nault & Villanueva, 2021)

Instances of an enrichment of genetic alterations linked to specific risk factors were denoted by the symbols *, #, %, and $. Several important genes and risk factors were highlighted, including AAV2 (adeno-associated virus type 2), CDKN2A (cyclin-dependent kinase inhibitor 2A), ELF3 (E74-like ETS transcription factor 3), MLL (myeloid/lymphoid leukemia), NFE2L2 (nuclear factor, erythroid 2 like 2), PRKACA/B (protein kinase cAMP-activated catalytic subunit alpha/beta), PSC (primary sclerosing cholangitis), RB1 (RB transcriptional corepressor 1), RPS6KA3 (ribosomal protein S6 kinase A3), and TSC1/2 (TSC complex subunit 1/2).

1.8 Krüppel-like factors (KLFs) and KLF15

The Krüppel-like factors (KLFs) represent a group of 17 transcription factors characterized by conserved zinc finger domains. Several of these proteins are found in various tissues and exhibit specific activities and functions unique to each tissue (Y. Zhao et al., 2019). KLFs play a critical role in regulating numerous physiological processes, including growth, development, differentiation, proliferation, and embryogenesis. Among these, KLF15, identified as a kidney-enriched member of the KLF family, is also commonly known as KKLF (Uchida et al., 2019). Previous research Zhao et al. (2021) indicates that KLF15 is involved in a number of mammalian systems, including the transcription of KLF3 in adipocytes and the state of rhodopsin. Additionally, it has been linked to both promoting and inhibiting gastric (GC) cell growth and metastasis, as well as lung adenocarcinoma. The study Zhao et al. (2021) also demonstrates how KLF15 inhibits human cancers, as shown by the upregulation of TFAP2A-AS1.

1.9 KLF15 and cancer

Among the transcription factors that have endured over time are Krüppel-like factors (KLF) and specificity proteins (SP). Three zinc finger domains define these proteins from the C-terminus. Among all the members of the SP/KLF family, these domains are unique for being the only ones displaying significant homogeneity. The patterns these proteins show in their middle and N-terminal areas control protein interactions and determine the selectivity of DNA binding, hence there is great variety in them (Orzechowska-Licari et al., 2022). (Rane et al., 2019) report that eighteen KLFs have been found across several taxa. Member of the KLF family, KLF15 has been connected to several diseases. KLF15, belonging to the KLF family, has been implicated in a range of diseases, such as diabetes (Patel et al., 2017), cardiac hypertrophy (Y. Zhao et al., 2019), , muscle atrophy (Hirata et al., 2019), and specific types of cancer (Yoda et al., 2015). Still, there is a dearth of present studies on how KLF15 fuels the tumorigenicity of breast cancer (BrCa). Although the

precise anti-tumor mechanism is currently unknown, Yoda et al. (2015) discovered that KLF15 could help to lower the growth of tumors. (Liu et al., 2020) argued that managing breast cancer (BC) depends critically on the "LUCAT1/miR-181a-5p axis". In comparison, LUCAT1 stimulates tumor growth but miR-181a-5p reduces it. Direct links between LUCAT1 and miR-181a-5p have lately been discovered. Moreover, it has been discovered that downregulation of KLF6 and KLF15 is a main control of miR-181a-5p activity in breast cancer. The LUCAT1/miR-181a-5p axis (Liu et al., 2020) seems to be a good therapeutic target for BC, miR-376a-3p, (Y. Wang et al., 2020) found that TTN-AS1 favorably controls KLF15. In vivo testing validated these findings. TTN-AS1 increased in tissues and cell lines from colorectal cancer (CRC). Functional studies using TTN-AS1 inhibition found that cell mortality was increased while CRC cell invasion and proliferation were lowered. Computational analysis pinpointed miR-376a-3p as a target of TTN-AS1. The introduction of a miR-376a-3p inhibitor via transfection mitigated the effects triggered by TTN-AS1 knockdown (Y. Wang et al., 2020) . Although the aforementioned research has indicated the involvement of KLF15 in the advancement, growth, and varying levels of expression in cancer, the overall impact of KLF15 polymorphism on tumor development remains unclear.

The existence of a genetic connection between KLF15 SNPs and an elevated susceptibility to liver cancer remains unknown. The main objective of this research is to identify the potentially detrimental SNPs present in KLF15 and establish a connection between them and liver cancer. By conducting further research on KLF15, we can gain a deeper understanding of its characteristics. This knowledge may potentially result in the identification of a new therapeutic target and a prognostic marker for the early detection of liver cancer.

1.10 Problem statement

The challenging prognosis, chemotherapy resistance, and substantial morbidity and mortality linked with liver cancer have considerably hampered patient outcomes. Consequently, this study's principal aim is to identify unique variations in the KLF15 gene that could serve as a promising target for diagnostics.

1.11 Aims and objective

- To investigate the structure and function of the KLF15 genetic variant.
- To determine the association of KLF15 variant with liver cancer in Pakistani population.

1.12 Rationale and significance of study

Often referred to as liver cancer, hepatic carcinoma is one of the primary causes of cancer-related fatalities worldwide. The complexity of its pathophysiology demands a complete awareness of the genetic elements influencing its beginning and progression. Among the several biological mechanisms under recognized regulation of KLF15 (Krüppel-like factor 15) are cell proliferation, inflammation, and metabolism. Variations in the KLF15 gene might alter the way the protein is handled, thereby influencing the rate of liver cancer development and the probability of occurrence. Finding the function of KLF15 polymorphisms in the etiology of liver cancer might inspire the detection of unique therapeutic approaches and diagnostic instruments. The current study is significant as it may close a research gap on the genetic elements causing liver cancer. By defining the role of KLF15 polymorphisms, this work has the potential to promote cancer medicine; this will open the path for more focused methods of illness prevention and treatment considering genetic composition. Furthermore, by creating focused treatments based on the found pathways and processes by which KLF15 influences hepatic cancer, one can improve patient outcomes and reduce the global burden of this disease.

2. LITERATURE REVIEW

2.1 Introduction to cancer

The extracellular matrix is an essential component found in all tissues and organs, playing a critical role in the survival of multicellular organisms. The body's regulatory system governs and optimizes all cellular activities, spanning from the initial phases of organism growth to the end of life. Solid tumors in cancer cause changes to the extracellular matrix at various levels, including biochemical, biomechanical, architectural, and topographical. In recent years, there has been a significant increase in research and knowledge of the importance of the matrix (Cox, 2021). The characteristic of cancer conceptualization is the development of a framework that aims to simplify the immense complexity of cancer phenotypes and genotypes into a preliminary set of guiding principles. As our understanding of the mechanisms behind cancer has increased, other facets of the disease have emerged as potential areas for improvement. The paper by (Hanahan, 2022), discusses the potential distinction between disturbed differentiation and phenotypic flexibility as hallmark capacities. Non-mutational epigenetic reprogramming and polymorphism microbiomes are proposed as unique facilitating features that contribute to the development of hallmark capacities (Hanahan, 2022). The existing group of eight hallmarks, shown in Figure 1 (on the left), includes the acquired abilities related to maintaining cell division signaling, evading growth inhibitors, resisting programmed cell death, enabling unlimited replication, initiating and accessing blood vessel formation, activating invasion and spread to other parts of the body, reprogramming cellular metabolic processes, and avoiding destruction by the immune system. In a more recent update of this paradigm (2), two specific traits, namely the disruption of cellular metabolism and the ability to avoid immunological responses, were identified as "emerging hallmarks." Nevertheless, it is now clear that these two distinguishing characteristics, similar to the initial six, can be seen as

essential and inherent attributes of cancer. Consequently, they are included as such in the current depiction. In 2012, the International Agency for Research on Cancer (IARC) calculated that there were 8.2 million deaths caused by cancer and 14.1 million new cases of cancer worldwide (Sarwar & Saqib, 2017). In Pakistan, the prevalence of cancer has been steadily rising, according to the World Health Organization. The five most common cancers, according to the study (Tufail & Wu, 2023), were breast cancer (24.1%), oral cancer (9.6%), colorectal cancer (4.9%), esophageal cancer (4.2%), and liver cancer (3.9%). Oral cavity cancer (14.9%), colorectal cancer (6.8%), liver cancer (6.4%), prostate cancer (6.0%), and lung cancer (6.0%) were all more common in men than in women. Breast (6.9%), oral cavity (5.5%), cervix (4.7%), and uterus (4.1%) cancers were the most prevalent among women (41.6%). The highest risk group for cancer development was middle-aged people (43.0%), followed by seniors (30.0%) and adults (20.0%).
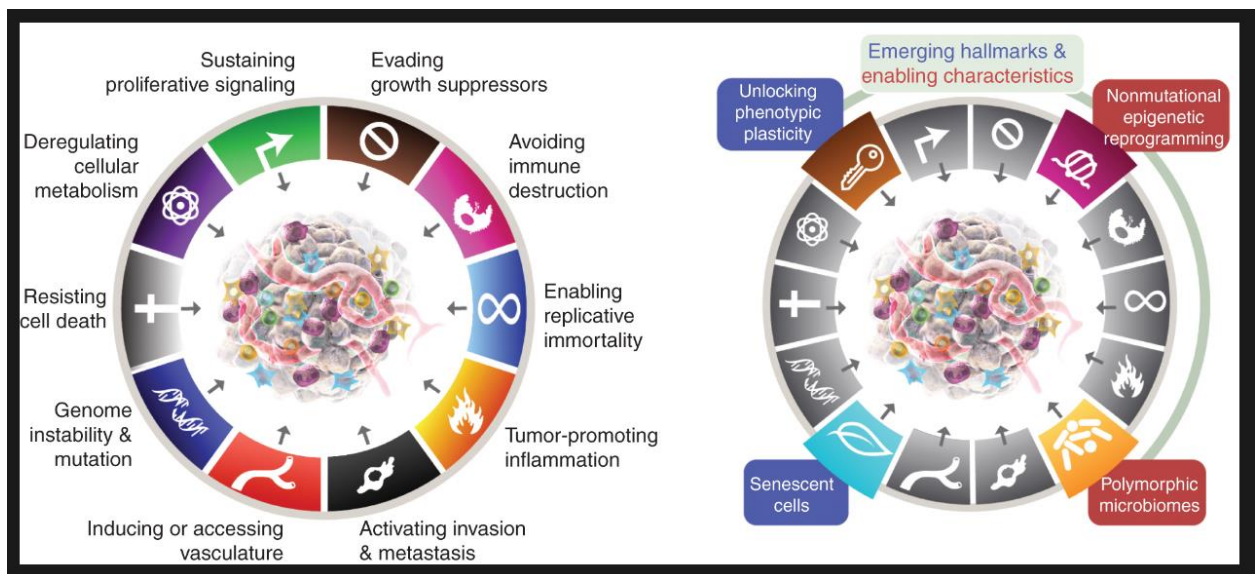
Figure 3: Eight hallmark skills and two enabling traits are currently embodied by the Hallmarks of Cancer (left).

The two initial "emerging hallmarks" introduced in 2011, namely cellular energetics (now commonly known as "reprogramming cellular metabolism") and "avoiding immune

destruction," have been thoroughly validated and are now considered integral components of the core set, along with the six acquired abilities proposed in 2000, known as the Hallmarks of Cancer. This attribute is also commonly described as the capacity to stimulate or otherwise gain entry to blood vessels that facilitate the expansion of tumors, as it is now recognized that tumors can acquire sufficient blood supply either by initiating the formation of new blood vessels or by taking over existing ones in healthy tissues (128). Furthermore, the 2011 update introduced "tumor-promoting inflammation" as an additional enabling characteristic, alongside the primary factors of "genome instability and mutation." These factors are crucial in activating the eight hallmark capabilities necessary for tumor growth and progression. This review also considers additional potential emergent characteristics and facilitating qualities such as "activating phenotypic flexibility," "nonmutational epigenetic reprogramming," "diverse microbiomes," and "senescent cells." The picture depicting the symptoms of cancer was modified from (Hanahan, 2022).

2.2 Introduction to liver cancer

The liver is the sixth most frequent site for primary cancer in people and often develops when there is cirrhosis and inflammation. Additionally, due to the liver's physical position and architecture, as well as its distinct metabolic milieu and immunosuppressive environment, metastases from malignancies of other organs (especially the colon) typically infiltrate it (Li et al., 2021). Liver cancer stands as one of the most prevalent lethal malignancies on a global scale; however, within the United States, it holds the fifth position in terms of incidence (Anwanwan et al., 2020). A frequent challenge in managing liver cancer lies in the late-stage diagnosis that often ensues, consequently contributing to its unfavorable prognosis (Siegel et al., 2018). Hepatocellular carcinoma (HCC) constitutes more than 90% of all liver cancer cases, with chemotherapy and immunotherapy serving as the foremost therapeutic options (El-Serag et al., 2019). Given the critical need for enhanced treatment modalities for liver cancer patients, there is a pressing requirement to explore novel avenues. The utilization of natural compounds in conjunction with nanotechnology presents a promising avenue that holds the potential to deliver improved

patient outcomes while mitigating systemic toxicity and minimizing adverse effects (Li et al., 2019). The pursuit of advanced treatments could herald a brighter outlook for liver cancer patients, ultimately resulting in more favorable prognoses. Chronic hepatitis B and C (HBV/HCV) infection is responsible for a sizable share of liver malignancies. If established measures like immunization can be applied on a wide scale, liver cancer may become the second cancer after cervical to be effectively controlled globally. The predicted global mortality rate for liver cancer in 2018 was 8.5 per 100 000 people (Shi et al., 2021).

The findings (Foda et al., 2023) were verified in a separate population. In liver cancer, chromatin and genomic alterations, particularly those resulting from transcription factor binding sites, were reflected in altered cfDNA fragmentation. These findings offer a biological explanation for modifications in cfDNA fragmentation in liver cancer patients and a practical method for noninvasive cancer detection.

(Wu et al., 2021) investigated in their research how closely TLSs are to cancer cells influences their distinctive composition. They so proposed a TLS-50 capability for precise TLS spatial detection. By clarifying the complex PLC tumour environment, (Wu et al., 2021)   discovered fresh data that might support cancer treatment. Since there are no medications that target important dependencies, treating liver cancer remains challenging. People with hepatocellular carcinoma gain little from broad-spectrum kinase inhibitors including sorafenib. Combining senolysis, a second medicine meant to target and destroy old cancer cells, with senescence induction may help cancer patients (Wang et al., 2019).

2.3 KLF Family Proteins

The liver serves as a focal point and regulates a number of crucial physiological functions, including metabolism and xenobiotic detoxification. These pleiotropic effects are made possible at the cellular level via hepatocyte transcriptional regulation. Hepatic disorders emerge as a result of defects in hepatocyte function and its transcriptional regulatory mechanisms, which have a negative impact on liver health (Yerra & Drosatos, 2023). This

study (Yerra & Drosatos, 2023) covers the role that Krüppel-like factors (KLF), members of the zinc finger family of transcription factors, play in physiological hepatocyte functions as well as their role in the beginning and progression of hepatic disorders (Yerra & Drosatos, 2023).
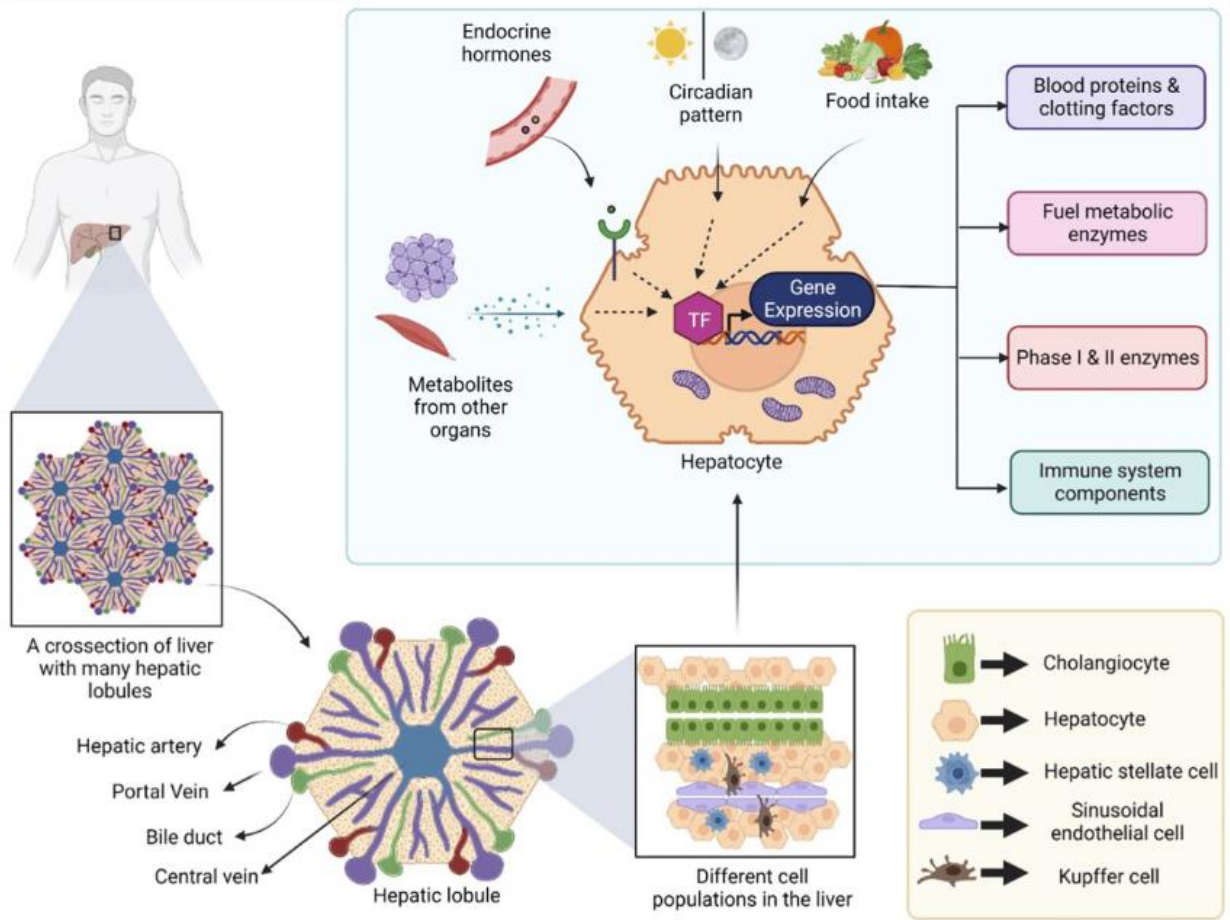
Figure 4: demonstrating the transcriptional control of hepatocytes and several cell types in the liver: Each of the four lobes of the human liver contains many hepatic lobules.

Different types of epithelial, endothelial, and immunological cells can be found in each hepatic lobule. A specialized parenchymal epithelial cell known as a hepatocyte is capable of producing various blood proteins, clotting factors, metabolic enzymes, etc. The

day/night cycle, dietary intake, endocrine hormones, metabolites from other metabolically active organs, and other environmental factors all have an impact on the gene transcription in hepatocytes. **Source: (Yerra & Drosatos, 2023)**

Investigating in mice gave a basic understanding of the function of Krüppel-like factors (KLFs) in embryonic liver development. Recently discovered is a major role of the Erythroid Krüppel-like factor (EKLF) transcription factor in controlling foetal liver RBC production. Researchers saw foetal anaemia in EKLF1-deficient mice to support this (Nuez et al., 2018). Later studies revealed the function of KLF6 in mouse blood cells and embryonic stem cells as well as in liver development. (Matsumoto et al., 2019) claimed that embryos lacking Klf6 had poor hematopoiesis and insufficient yolk sac vascularization, hence lacking clearly defined liver development. Laboratory findings confirmed those in the wild: stem cells extracted from these embryos showed slowed blood cell development (Matsumoto et al., 2019). Apart from its participation in early hematopoiesis, these studies revealed the particular function of KLF6 in hepatogenesis remained unknown.

More research on zebrafish revealed that klf6 (zebrafish: copeb) is necessary for appropriate liver development during the early phases of endodermia, in which case the liver is either entirely or partially absent. When Klf6 was absent from mouse embryonic stem cells, endoderm marker genes—including Hnf3β, Gata4, Sox17, and CxCr4—found their expression down. On the other hand, the expression levels of these genes were recovered upon Klf6's overexpression (Zhao et al., 2021). Moreover, at E14.5 researchers have discovered KLF13 expression in blood vessels of the developing liver; the precise role of this gene is currently unknown (Lavallée et al., 2019). Studies have indicated that KLF15 helps to regulate the differentiation of embryonic hepatoblasts produced from mouse liver into mature hepatocyte-like cells. The paper reports that hepatocyte marker genes Tat, Cps1, Cyp1a2, and Cyp2e1 were triggered when KLF15 was overexpressed in hepatoblasts produced from induced pluripotent stem cells (iPSCs) by retrovirus-mediated

15

techniques. This indicates that a lineage unique to hepatocytes cannot develop without KLF15 (Anzai et al., 2021).

Among the several physiological and biological events influenced by Krüppel-like factors (KLF), a family of conserved transcription factors containing zinc fingers, cell division, proliferation, development, and death. Among the bioinformatics approaches suggested to help to better grasp KLF proteins include phylogenetic reconstruction, gene synteny analysis, sequence similarity searches, multiple sequence alignment, and phylogenetic reconstruction (Le et al., 2021). Post-translational modifications can either change the metabolic activity or induce KLFs to alter their metabolism. The Krüppel protein regulates thorax and front abdomen segmentation in an embryo developing in a fruit fly (Drosophila melanogaster). This protein bears rather close resemblance to KLFs (Preiss et al., 2018). Like the KLF family, the transcription factor Sp1 binds to GC-rich areas of DNA using three C2H2-type zinc fingers (Brayer & Segal, 2018). The Sp1/KLF family incorporates these zinc fingers as they are discrete domains seen in KLF proteins (Kadonaga et al., 2019. Sp1/KLF family members have been shown to regulate the intricate interactions of multiple genes involved in the development and maintenance of different types of tissue. However, initially, this family was believed to be a general transcription factor responsible for controlling the basic expression of essential genes (McConnell & Yang, 2010).

2.4 KLF15 – a unique member of KLF family

The study (Suzuki et al., 2022) demonstrate that Krüppel-like factor 15 (Klf15) coregulates regeneration enhancers identified by the ATAC-seq and H3K27ac landscape analysis for transposase-accessible chromatin. Furthermore, it was demonstrated that the adrenergic receptor gene is a downstream target of Klf15, and that therapy with an agonist for this receptor promotes the regeneration of nephric tubules and restores organ size. According to these findings, the regeneration pathway heavily depends on Klf15-dependent adrenergic receptor signaling. Through altering O6-methylguanine-DNA methyltransferase expression, KLF15 improved glioma sensitivity to temozolomide

16

cytotoxicity. According to this study (Huang et al., 2022), KLF15 expression was higher in healthy tissues than it was in CRC tissues, and low KLF15 expression was associated with poor CRC patient outcomes. The degree of methylation of Cg sites in CRC was negatively correlated with the expression of KLF15. Therefore, it is simple to conclude that hypermethylation of its promoter region is a contributing factor in the reduction of KLF15 expression. The findings (Kanyomse et al., 2022) demonstrated that promoter methylation of KLF15 partially explains the substantial downregulation of KLF15 in breast cancer cell lines and tissues. When KLF15 was expressed exogenously, it caused TNBC cells to go into apoptosis and the G2/M phase of the cell cycle, which reduced cell growth, metastasis, and in vivo carcinogenesis. The CCL2 and CCL7 chemokine ligands were specifically targeted by KLF15, according to investigations on the mechanism.

Transcriptome and metabolomic analyses show that KLF15 is implicated in important metabolic and anti-tumor regulating pathways in TNBC, according to (Kanyomse et al., 2022). Reducing CCL2 and CCL7 is the primary method by which KLF15 slows the spread and development of TNBC cells. KLF15 is one possible prognostic biomarkener for TNBC. Recent studies ((X. Zhao et al., 2019) show that KLF15 modulates molecules like myocardin, GATA-binding protein 4 (GATA4), transforming growth factor (TGF), and myocyte enhancer factor 2 (MEF2), hence suppressing pathogenic cardiac hypertrophy and fibrosis. KLF15 is one likely therapeutic target for heart failure and other cardiovascular illnesses. SFe therapy reduced p-CREB and KLF15 expression in muscle tissue, according to (Choi et al., 2021). This implies that by cAMP-PKA/CREB-KLF15 signalling, SFe could control atrogin-1 and MuRF1, hence preventing protein breakdown. Hyperglycemia was seen to raise the expression of the KLF15 protein in the skeletal muscles of diabetic rats ((Hirata et al., 2019). We thus prevented the ubiquitin-dependent degradation of KLF15 and lowered the E3 ubiquitin ligase WWP1 to reach this. These data indicate that the WWP1/KLF15 pathway causes muscle atrophy in hyperglycemia, a major diabetes consequence. Preventing the loss of skeletal muscle mass brought on by diabetes could make this pathway a therapeutic target. Transfection of miR-181a-5p produced overexpression of miR-181a-5p, which in turn reduced LUCAT1's capacity to induce cell

17

proliferation. Furthermore shown by bioinformatics and a luciferase reporter test was miR-181a-5p's targeting of the 3′-UTR region of the tumours suppressor genes KLF6 and KLF15. The LUCAT1/miR-181a-5p axis controlled KLF6 and KLF15 expression in both living organism and laboratory environments (Liu et al., 2020). This study (Wang et al., 2020b) revealed that in tissues and cell lines connected to CRC-related disease TTN-AS1 was overexpressed. Downregulating TTN-AS1 reduced CRC cell growth and invasion and increased cell death, according studies of functional components. Target of TTN-AS1 is MiR-376a-3p according a bioinformatics analysis. Transfecting with an inhibitor of miR-375a-3p helped to offset the effects of TTN-AS1 knockdown. Furthermore positively regulated by TTN-AS1 was miR-376a-3p. Furthermore supported by in vivo studies were all these findings (Wang et al., 2020b). Ultimately, TTN-AS1's promotion of CRC invasion and spread depends critically on the miR-376a-3p/KLF15 axis. TTN-AS1 shown promise as a possible target for CRC treatment (Y. Wang et al., 2020).

2.5 KLF15 in Disease: Implications for Cardiomyocyte Hypertrophy

Many organs, including the kidney, show transcription factors of the Krüppel-like family. Acute kidney injury alters the expression of many KLFs in either an adaptive or maladaptive manner, therefore influencing multiple cellular pathways (Anwanwan et al., 2020). Ang II activated p38 kinase via TAK1 kinase, therefore inducing the TGF-β expression and blocking KLF15 expression. E2 and -LGND2 both dropped all rather dramatically. E2's anti-hypertrophic effects as well as -LGND2's reduced myocyte hypertrophy generated by KLF15 knockdown. Important elements were validated with an in vivo model of heart hypertrophy. More ER anti-hypertrophic properties discovered by (Anwanwan et al., 2020) could be investigated in humans to stop the spread of heart disease from development. The main goal was to find in cultured neonatal rat cardiomyocytes how oestrogenic compounds and AngII affected KLF15 expression. AngII substantially lowered the protein and messenger RNA of the anti-hypertrophic transcription factor. ER abolished the E2-suppressed AngII effects on KLF15 protein observed in presence of control siRNA. The effects of ER and TGF on the reduction in KLF15 expression were

then investigated. TGF synthesis and AngII lower KLF15 mRNA and protein in a p38 kinase-dependent manner, which causes cardiomyocyte hypertrophy (Leenders et al., 2012). The KLF15 protein shows a strong association with the anti-hypertrophic transcription factor. This TF physically interacts with and inhibits the myocardin protein function in the nucleus of cardiomyocytes, inhibiting the increased production of genes that promote hypertrophy (Leenders et al., 2012) . Here expanded to encompass AngII and TGF-induced TAK1 signaling. The TAB family of proteins (TAB 1-3) complexing with TAK1 closely regulates its activity, and all of its activities are affected by a variety of post-translational changes (Hirata et al., 2017). Through its effects on elements including myocyte enhancer factor 2 (MEF2), GATA-binding protein 4 (GATA4), transforming growth factor (TGF), and myocardin, recent investigations have also indicated a crucial role for KLF15 as an inhibitor of pathological cardiac hypertrophy and fibrosis. Therefore, KLF15 might be a useful therapeutic target for treating heart failure and other cardiovascular illnesses (Y. Zhao et al., 2019).

2.6 KLF15 in Liver Disease: Implications and Insights

Forkhead box protein O1 and insulin-targeting SREBP1c moreover inhibited TWIST2 expression, but dexamethasone-targeting Krüppel-like factor 15 increased it through direct interactions with the Twist2 promoter DNA. Together, our studies show that TWIST2 plays a critical role and regulates hepatic homeostasis by reducing steatosis, inflammation, and oxidative stress via the NF-B-FGF21 or SREBP1c-FGF21 pathway, which may offer a novel treatment option for nonalcoholic fatty liver disease (Zhou et al., 2019). Through transcriptional regulation of important metabolic processes, glucocorticoids can coordinate systemic metabolic balance by targeting the transcription factor KLF15 (Hsieh et al., 2019). KLF15 expression in adipose tissue was decreased in obese patients, and KLF15 is linked to an elevated body mass index in obesity (Hsieh et al., 2019). The glucocorticoid receptor downstream is well known to be mediated by KLF15. KLF15 affects the supply of

gluconeogenic carbon substrates from skeletal muscle as well as their usage in the liver. This includes both direct control of hepatic gluconeogenic enzymes and indirect control of these processes (Gray et al., 2017). The metabolism of the liver is significantly impacted by KLF15. Deletion of the KLF15 gene in mice fed a high-fat diet (HFD) has no effect on endoplasmic reticulum stress, insulin resistance, or hepatic inflammatory response. Animals on a high-fat diet show improvements in indices of endoplasmic reticulum stress (Tian et al., 2020), even if the deletion of the KLF15 gene in mice has no impact on endoplasmic reticulum stress or the inflammatory response in the liver connected with insulin resistance. The liver responds well from KLF15's anti-apoptotic properties. Treating ALI brought on by LPS or DGaIN using exogenous KLF15 Moreover, KLF15 can lower ALI by blocking the p38MAPK/ERK1/2 signalling pathway and so lowering of sepsis-induced inflammation (Tian et al., 2020). Teneligliptin, a mechanism mediated by phosphatidylinositol 3-kinase (PI3K)/AKT/Krüppel-like factor 15 (KLF15), clearly affected the expression of the BA synthesis inhibitory factor Fgf15. KLF15's inhibition neutralised this impact. Teneligliptin increases BA generation, hence improving the treatment of metabolic diseases (G. Wang et al., 2020).

2.7 KLF 15-Role in cancer

Stomach cancer tissues show significantly less KLF15 than surrounding normal tissues. Furthermore inversely linked to clinical stage, lymphatic metastases, and distant metastases was KLF15 expression. Moreover, KLF15 expression can be a prognostic indicator for GC patients. Furthermore, as (Sun et al., 2017) point out, an overabundance of KLF15 expression may possibly lower cell proliferation by changing CDKN1A/p21 and CDKN1C/p57. TTN-AS1 used overarchingly the miR-376a-3p/KLF15 axis to encourage CRC invasion and spread. Potential target for treating colorectal cancer, TTN-AS1 showed promise according to (Y. Wang et al., 2020). Reports state that KLF15 is linked to higher cancer risk. Think of the study of (Gao et al., 2017), who discovered that aberrant KLF15

expression was present in lung cancer tissue and increased cell proliferation and metastases in vitro. Researchers in this work identified KLF15 as a target of miR-376a-3p by bioinformatics means. Confirming this was done with a luciferase test. Most often used method of control by lncRNA is sponging target miRNAs upregulates the genes controlled by those miRNAs. LncRNA CRNDE modulated the miR-384/IRS1 axis to act as an oncogene in pancreatic cancer, hence illustrating the concept (Wang et al., 2017). Expression of KLF15 was shown to be elevated in CRC tissues and positively correlated with TTN-AS1 expression. Then, by positively controlling KLF15 through sponging miR-376a-3p, TTN-AS1 might have limited CRC advancement. TTN-AS1 was therefore found for the first time as a colorectal cancer biomarkers. Targeting the TTN-AS1/miR-376a-3p/KLF15 pathway, (Y. Wang et al., 2020)propose as advance method to treat colorectal cancer.

2.8 KLF-15- Signaling pathway

The negative regulatory elements of cardiac hypertrophy have gotten less attention, despite the fact that we are learning more about the signaling pathways that encourage it. By modulating the transcriptional pathways that control heart metabolism, Krüppel-like factor 15 (KLF15) appears to play a significant role in sustaining cardiac function. The actions of KLF15 on molecules including myocyte enhancer factor 2 (MEF2), GATA-binding protein 4 (GATA4), transforming growth factor (TGF)- (TGF-), and myocardin have been linked to a critical role for KLF15 as an inhibitor of pathological cardiac hypertrophy and fibrosis. For the treatment of heart failure and other cardiovascular illnesses, KLF15 may therefore be a useful therapeutic target (Y. Zhao et al., 2019).
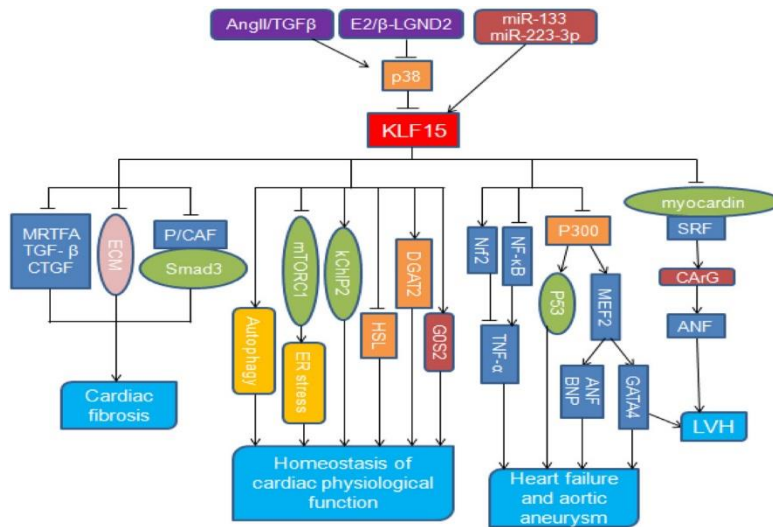
Figure 5: it depicts the diverse functions of KLF15 in the heart. Initially, KLF15 maintains cardiac equilibrium by regulating adipocyte lipid metabolism, systemic energy balance, managing ER stress, controlling autophagy signaling, and modulating cardiac function. Source : (Y. Zhao et al., 2019)

In terms of countering fibrosis, KLF15 inhibits key factors in cardiac fibroblasts, reducing myocardial fibrosis and improving cardiac function. KLF15 also serves as a transcriptional inhibitor of left ventricular hypertrophy (LVH) by strongly inhibiting myocardin. Furthermore, it is regulated reciprocally by AngII/TGFβ and E2/β-LGND2, and it is directly targeted by miR-133 and miR-223-3p. Lastly, the KLF15-p300-p53 pathway holds potential as a therapeutic target for congestive heart failure and aortic aneurysm.

When KLF15 levels are elevated, as can occur during fasting, they promote the recruitment of p300 to specific metabolic targets. Conversely, reduced KLF15 levels, which may happen during pressure overload, lead to a decrease in the expression of these metabolic genes. At the same time, this reduction in KLF15 levels frees up p300 molecules, allowing them to co-activate prohypertrophic transcription factors like GATA-binding protein 4 (GATA4) and myocyte enhancer factor 2 (MEF2) (Prosdocimo et al., 2014).
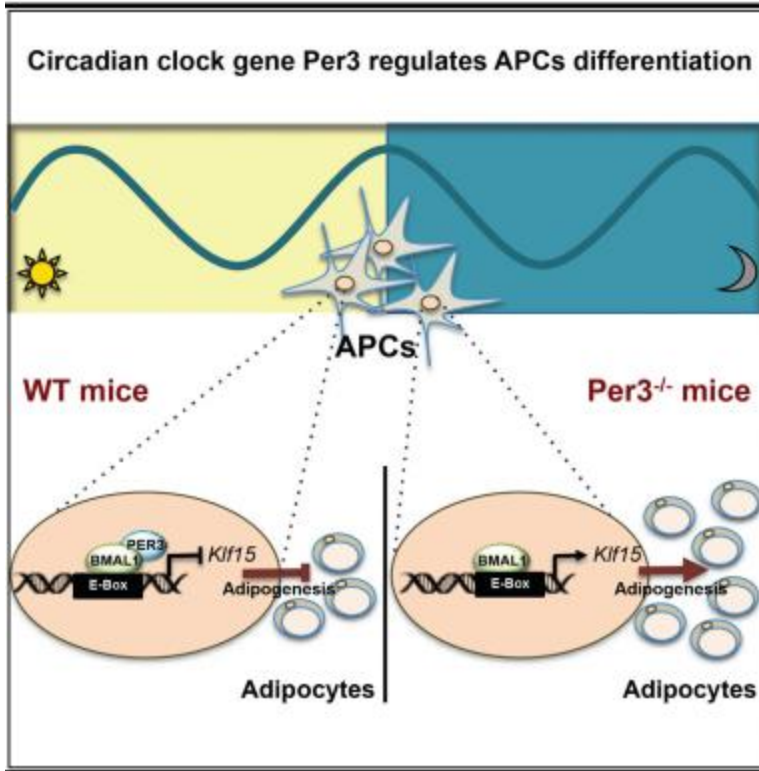
Figure 6: Uncertainty exists regarding the peripheral circadian clock's role in the biology of adipocytes. Source: (Aggarwal et al., 2017)

According to (Aggarwal et al., 2017) adipocyte precursor cells have a circadian clock that includes context-sensitive components. Period 3 (Per3), in particular, plays a significant role in the clock of adipocyte precursor cells and defines an output pathway to Klf15 that controls adipogenesis.
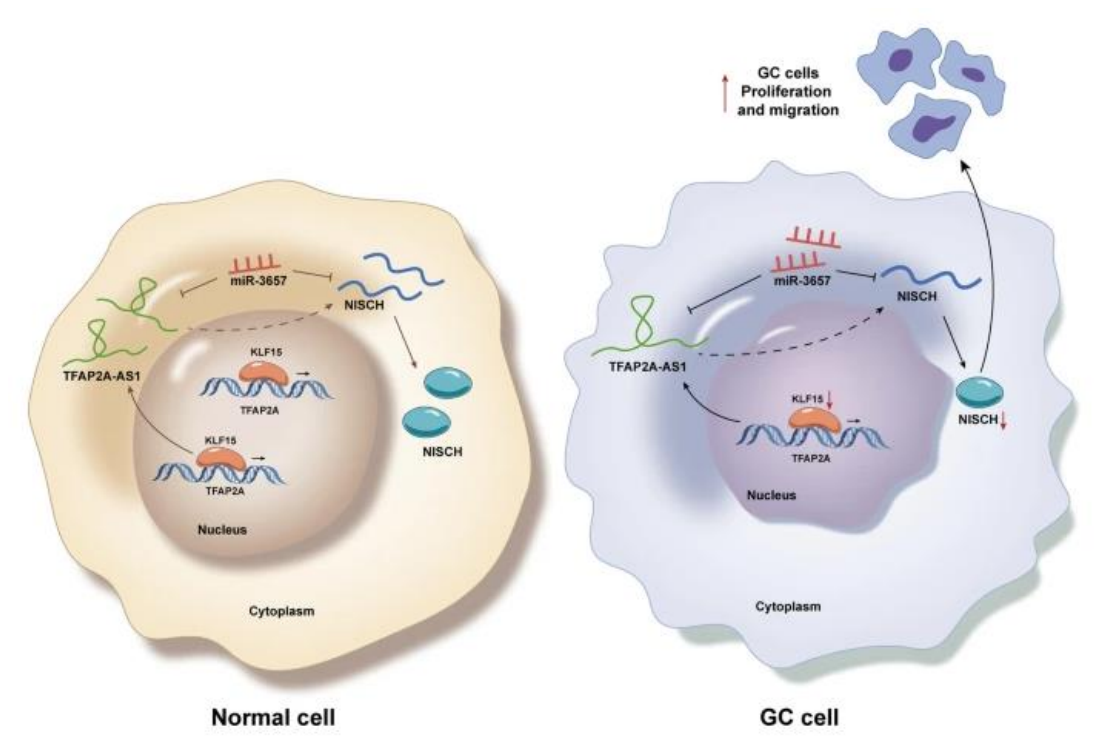
Figure 7: The proliferation and migration of GC cells are inhibited by TFAP2A-AS1. MiR-3657 was identified as TFAP2A-AS1's downstream gene and NISCH as its target in the downstream regulation pathway. Source: (Zhao et al., 2021)

Additionally, NISCH reduces GC cell migration and proliferation. Transcriptional factor KLF15 favorably mediates TFAP2A-AS1 to inhibit GC cell proliferation and migration in the upstream regulatory process.

3. MATERIAL AND METHODS

3.1. Structure Prediction of KLF15

KLF15 gene protein sequence was obtained from the ENSEMBL database. The format of the protein sequence obtained was in FASTA format with the transcript ID: ENSG00000163884. The obtained sequence consists of 416 amino acids. Other than the protein sequence the ENSEMBLE database provides information about the variants of the gene, linkage to the disease, and a prediction of protein function (Cunningham et al., 2019). The structure of the protein KLF was obtained from AlphaFold by submitting the amino acid sequence of the protein the FASTA format to the tool (Jumper et al., 2021). The 3-D protein structure provided by the AlphaFold was subsequently applied in the analysis. KLF-15's FASTA amino acid sequence was submitted in InterProScan for domain prediction of the protein (Blum et al., 2021).

3.2. Retrieval of Pathogenic SNP

ENSEMBLE yielded variants of KLF15; the variant table was obtained from which 4483 variant were found. Then pathogenicity prediction was based on filtered missense variants. Strong link with the disease guided the choice of these selected SNPs depending on the pathogenicity score of the in-silico tools.

3.3. Analysis of missense SNPs

These missense SNPs were further analyzed through different tools like REVEL (Ioannidis et al., 2016) Mutation Accessor (Bott et al., 2011) , metaLR (Wickramarachchi et al., 2020), SIFT (Sim et al., 2012), Polyphen2 (Adzhubei et al., 2013) and CADD (Rentzsch et al., 2019). These tools helped in screening the deleterious SNPs. These tools assisted to

screen the potentially disease-causing SNPs. Two SNPs in all were obtained for the study for the further analysis.

3.4. Protein stability analysis

Protein structural stability was investigated in response to KLF15 variants using MUpro, MAESTRO, MutPred-2, DynaMut, and HOPE. We investigated how selected SNPs will influence protein stability. We acquired delta delta G (DDG) values to project the effect on protein stability resulting from a single nucleotide modification. A lower DDG value suggests decreased stability of the protein, which changes its typical pathways. The DDG value between roughly 0.5 and 50.5 increases both protein flexibility and stiffness.

3.5. Structural and functional study of SNP

The structural and functional effect of the single nucleotide polymorphisms on the protein was determined through Project Hope. The Project Have Our Protein Explained (Project Hope) tool was used to study the change in protein structure because of any mutation. Project Hope considers physicochemical alterations based on the size, by hydrophobicity, and charge of amino acids, as well as the ensuing loss of salt-bridge or hydrogen-bonding-based intra- or inter-residual contacts. Project Hope detect the heritable disease because of change in the structure and function of the protein due to its mutation (Zheng et al., 2022).

3.6. Association of KLF15 with cancer

The computational technique FATHMM was applied (Rogers et al., 2018) to ascertain whether the KLF15 missense variants correlated with the disease. This method forecasts the protein's link with disease using knowledge about protein sequence and function. We investigated the noted variations for any possible cancer-related connections using this approach. Whether human cancer was linked to the KLF15 missense mutations was

ascertained. The tool would note either a passenger or cancer once all the mutations were verified.

3.7. KLF15 flexibility analysis

The tool named DynaMut was used to identify the fluctuation and distortion in protein because of the missense variants. It was also used to check the effect of mutation on the molecular motion of the protein. In this regard, this tool uses the Normal Mode Analysis approach which analyzes the vibrational entropy change in the protein because of the single nucleotide polymorphisms (Rodrigues et al., 2018).

3.8. In-Situ Mutagenesis

The mutation was induced in the original KLF15 structure by changing the amino acids in the wild-type structure with the variant amino acid. The structure of the protein was predicted and aligned through PyMOL. For this in-situ mutagenesis a tool named mutagenesis wizard tool in the PyMOL v4.0.4 (Ogun et al., 2023) was used.

3.9. Molecular Dynamic Simulations of wild type and variant of KLF15

The molecular dynamic simulation for the wild-type structure of KLF-15 and its variant of the proteins is performed. The conformational changes on the inside of the protein structure because of the missense mutations can be checked by performing the simulation of 20ns. The stability of the model for molecular dynamic simulation will be obtained by using GROMACS 2016 (Abraham et al., 2015) and the force-field of OPLS-AA (Kulig et al., 2015). The initial energy minimization of the MD simulation was carried out at the 50,000 steps along with the NVT and NPT equilibration of 100 ps. The built-in programs of Gromacs were used for constructing trajectories (gmx_triconv) and also to carry out structural analysis. The trajectory coordinates were saved every 10ps. The time of 20ns is enough to check different parameters such as Radius of gyration (Rg), Root means square fluctuations (RMSF), the total number of intra- molecular hydrogen bonds, Solvent Accessibility Surface Area (SASA), and Root mean square deviations (RMSD). Also, the

27

rearrangements in the side chains of the wild-type structure of a protein can be completely understood. The Gromacs commands such as gmx_rms, gmx_rmsf, gmx_gyrate, gmx_hbond, and gmx_sasa were used to calculate the above-mentioned parameters. For the backbone of the structure RMSD was calculated. The protein and sidechains analysis will be done through RMSF. The radius of gyration was calculated for the protein and backbone. The results were plotted on scatter smooth line graphs.

3.10. Experimental Analysis

3.10.1. Sample Collection

For the study of liver cancer, the samples were collected from the patients above 18 years of age and without any co-morbidity. The patients were asked to sign a consent form that they are giving the samples by their own choice and are being part of the study. Also, ethical approval from the institutional review board of Atta-ur Rehman School of Applied Biosciences NUST was obtained.

3.10.2. Primer Designing

The primers for tetra ARMS PCR are designed computationally by a tool named Primerl (Zhong et al., 2018). To get the ARMS PCR primers from Primer1 the input in the form upper layer of supernatant was discarded, and the lower pellet layer was kept safe for further processing. Then the pellet was resuspended in 500ul of solution A, shaken vigorously, and a second centrifugation was carried out for 1 minute at the speed of 13,000 pn. Again, the supernatant after the centrifugation was discarded, and in the nuclear pellet solution B of volume, 400μl was added. After that again 400ul of the solution was added to the pellet separated. The other compounds that were added in the above solution of nuclear pellet and solution B are Sul of proteinase K (20μl/ml stored at -20°C) and 12μl of 20% SDS. It was then kept overnight in an incubator at 37°C.

On the following day, 250µl of solution C and 250µl of the solution were added. After inward mixing of the tube several times, it was centrifuged at 13.000 rpm for 10 minutes. On completion of centrifugation, two layers were observed in the tube. The upper layer containing DNA was separated very carefully into another Eppendorf tube. In it, 55µl of Sodium acetate and 500ul of isopropanol were added for the precipitation of the DNA. These tubes were then centrifuged again for 10 minutes at 13,000 rpm. After the centrifugation, the supernatant was discarded, and the pellet collected at the bottom was suspended in 200µl of 100% ethanol (chilled at -20°C). Centrifugation was carried out for 8 minutes at 13,000 rpm. After centrifugation, the DNA was dried at room temperature. 75µl of Tris-EDTA buffer was added to the DNA pellet.

Forward inner primer (T allele):	215 CACCTCAAGGCCCACCGGT 233
Reverse inner primer (C allele):	250 CTCACCCGTGTGCCGTCG 233
Forward outer primer (5' - 3'):	35 AAGTTTGTGCGCATTGCCCC 54
Reverse outer primer (5' - 3'):	410 CCCCTCCCTCCCCTCCCT 393

3.10.4. Gel Electrophoresis

The quality of the extracted DNA was analyzed on agarose gel electrophoresis. For this purpose, 1% agarose gel was used. To prepare 50ml of agarose gel, Sml 10X TAE buffer was mixed with 45ml of distilled water. Then 0.5g of agarose was added to the above solution in a beaker. This mixture was microwaved for 1-2 minutes, and Sul of ethidium bromide was then added. The gel mixture was poured into the gel tank and after removing the bubbles the comb was placed in the tank. The gel was placed at room temperature for

30 minutes to solidify. For 100ml of the gel, 10ml of 10X TAE buffer was added to 90ml of distilled water and Ig of agarose was added to the solution and the amount of ethidium bromide added was Sul. The next step after the solidification of the gel was to load the DNA samples into the wells. 8µl of DNA was mixed with 0.5µl of the loading dye; this mixing was carried out on the parafilm. After mixing, 8.5ul of the sample was added to the wells in the gel that was placed in the electrophoresis tank containing the IX TAE buffer. The gel was run at 100 volts for 15 minutes. The gel was visualized under the UV-Transilluminator. After visualization of the bands, the DNA samples were stored at 4°C for further use. Amplified PCR products were analyzed on 2% agarose gel. The agarose gel was prepared as described above.

3.10.5. Polymorphic genotyping

To detect the point mutation in the DNA extracted from the blood samples, Tetra ARMS PCR (Amplification-refractory mutation system polymerase chain reaction) was used. For ARMS PCR two inner (inner forward and inner reverse,) and two outer (outer forward and outer reverse) sets of primers are used. The inner primers are specific for the allele, but the outer primers are to detect the SNP by the amplification of the gene. For PCR 20µl of reaction was prepared with 1µl of all four primers, 8 ul of Solis BioDyne master mix, 2ul of the sample, and 6ul of water. To mix the PCR samples properly they were given a short spin before going for PCR. The reaction was optimized at 60°C of the annealing temperature. To carry out the samples at multiple temperatures simultaneously, a Gradient PCR machine was used. This machine has a specialty of using different temperatures for one go as each row in it has a different temperature. It will take 1.5-2 hours to complete the 35 cycles of the reaction.

PCR Steps and Conditions

Step 1: In the first step of the reaction the hydrogen bonds between the strands of the DNA are broken for the denaturation of the DNA molecule. This denaturation was carried out at 95°C for 5 minutes which will completely melt the molecule of DNA. The next step will not repeat the above step of denaturation, but it will complete the 35 cycles of denaturation of DNA at 94°C for 30 seconds.

Step 2: Multiple temperatures were set at this step as we were using gradient PCR. This step will also take 30 seconds.

Step 3: In this step, just after annealing the DNA will be polymerized for 30 seconds at 72°C and after that, the final polymerization of DNA will occur at 72°C for 7 minutes.

3.10.7. Statistical Analysis

The genotype data that was obtained through the ARMS PCR was further analysis by applying some statistical tools. Here the Graphpad Prism software 9 (Mavrevski et al., 2018) was used to obtain the significance values for the genotypes to show their link with the breast cancer. The p-value of 0.005 or less was regarded as statistically significant. In this tool the chi square method was used for the analysis and to obtain graphs between the desired values.

4. RESULTS

4.1 KLF-15 Structure Prediction

The 3D modelling was done through Alpha Fold online. The structure was visualized in PyMOL. The structure shows all the main regions including the kinase domain, regulatory domain as well as the hinge domain in a 3D manner. The protein structure was verified using InterPro, a web-based technology that classifies proteins into families and predicts their individual domains. The four predicted domains of the KLF 15 protein are the color red for the PE/DAG-bd domain, orange for the C2 domain, magenta for the protein kinase domain, and blue for the AGC kinase C domain that can be seen in figure 8. This gave us information related to the amino acids and their spatial arrangement. It also provides us with the information for designing experiments related to in-situ mutagenesis. Transcriptional regulator that binds to the GA element of the CLCNKA promoter. Binds to the KCNIP2 promoter and regulates KCNIP2 circadian expression in the heart (By similarity). Is a repressor of CCN2 expression, involved in the control of cardiac fibrosis. It is also involved in the control of cardiac hypertrophy acting through the inhibition of MEF2A and GATA4 (By similarity). Involved in podocyte differentiation (By similarity). Inhibits MYOCD activity. Is a negative regulator of TP53 acetylation. Inhibits NF-kappa-B activation through repression of EP300-dependent RELA acetylation (NCBI).
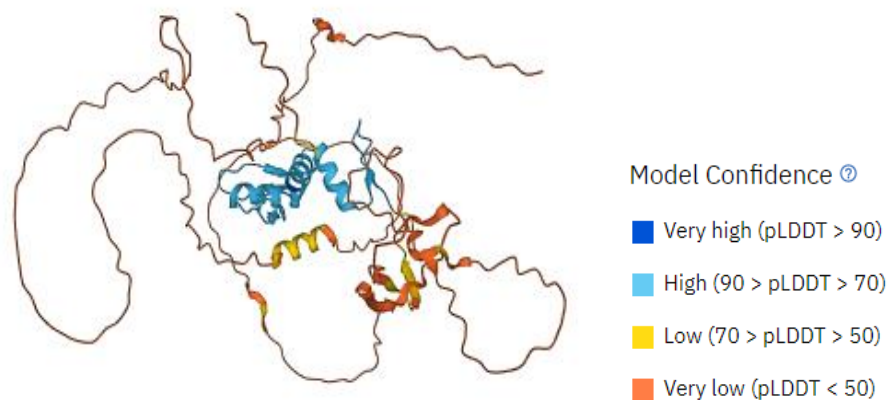
Figure 8: 3D- Structure Representation of KLF-15. The image represents all different regions of KLF-15. Each color represents a different region.

The optimized structure was validated in SAVES, and PROCHECK analysis revealed that 63.2 percent of amino acid residues had phi-psi angles in the most preferred regions, 25.8 percent in additionally allowed regions, 7.5 percent in generously allowed regions, and 3.5 percent residue had phi-psi angles in the disallowed regions of the Ramachandran plot. 83.607% is the quality factor of predicted structure analysis of KLF-15 (Figure 10).

Ramachandran Plot
saves
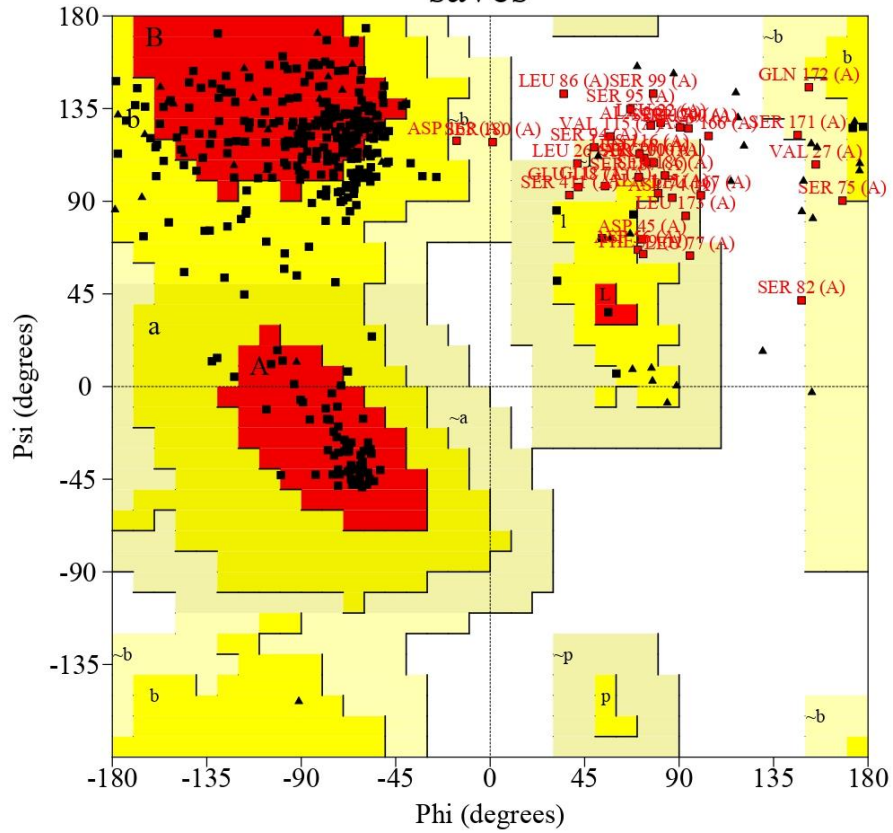
Phi (degrees)

Psi (degrees)

Plot statistics

Residues in most favoured regions [A,B,L]	201	63.2%
Residues in additional allowed regions [a,b,l,p]	82	25.8%
Residues in generously allowed regions [~a,~b,~l,~p]	24	7.5%
Residues in disallowed regions	11	3.5%
	----	------
Number of non-glycine and non-proline residues	318	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	46	
Number of proline residues	50	

Total number of residues	416	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms
and R-factor no greater than 20%, a good quality model would be expected
to have over 90% in the most favoured regions.

Figure 9 : Ramachandran Plot showing the measurements of angles in KLF15

Program: ERRAT2
File: AF-Q9UIH9-F1-model_v4.pdb
Chain#:A
Overall quality factor**: 83.607

*On the error axis, two lines are drawn to indicate the confidence with
which it is possible to reject regions that exceed that error value.

**Expressed as the percentage of the protein for which the calculated
error value falls below the 95% rejection limit. Good high resolution
structures generally produce values around 95% or higher. For lower
resolutions (2.5 to 3A) the average overall quality factor is around 91%.

Figure 10:Quality prediction analysis of predicted KLF-15 3D model

PyMOL was used for highlighting the various regions of KLF15 i.e homologous
superfamily and zinc finger domains, Figure 10 shows homologous superfamily was
highlighted via PyMol software, 335-404 (Red).
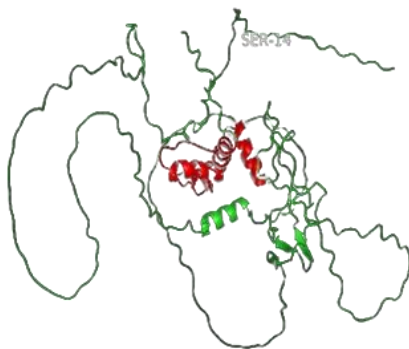
Homologous superfamily:

- 335-404 (Red)

Figure 11: Homologous superfamily was highlighted via PyMol software
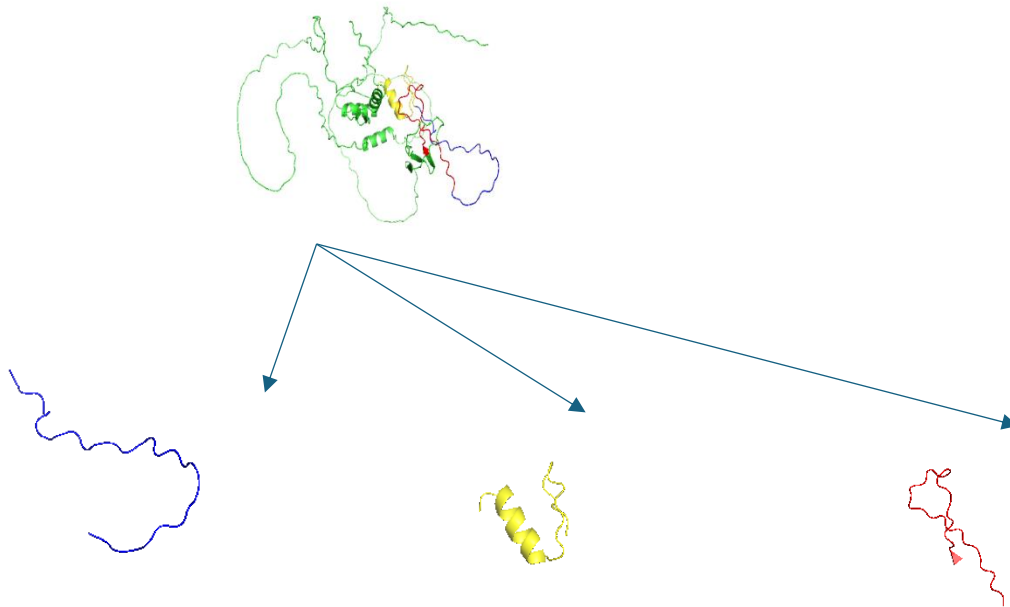
Figure 12: Zinc finger domains, 261-290 (red), 291-320(blue), 321-348(yellow)

4.2 Subcellular localization and phylogenetic tree

DEEPLOC 1.0 was able to anticipate where the KLF15 would be found subcellularly. Here we can see the KLF15 localization path in Figure 12. The red line represents the protein's pathway to its intracellular compartment. The score indicates the 40th percentile of the event's probability or likelihood. The score suggests that KLF15 is located in the nucleus. Figure 13 shows the evolutionary tree of KLFs, which shows that all KLFs have a common ancestor. The score quantifies the number of substitutions made at each site, providing insight into the relative evolutionary history of different proteins in the same family. The evolutionary link between KLFs families was found using Clustal Omega program.

Prediction: Nucleus, Soluble

Table 1: Subcellular localization

Localization	Likelihood
Nucleus	0.9997
Cytoplasm	0.0003
Cell membrane	0
Mitochondrion	0sg
Endoplasmic reticulum	0
Golgi apparatus	0
Peroxisome	0
Extracellular	0
Lysosome/Vacuole	0
Plastid	0
Type	**Likelihood**
Soluble	0.6948
Membrane	0.3052
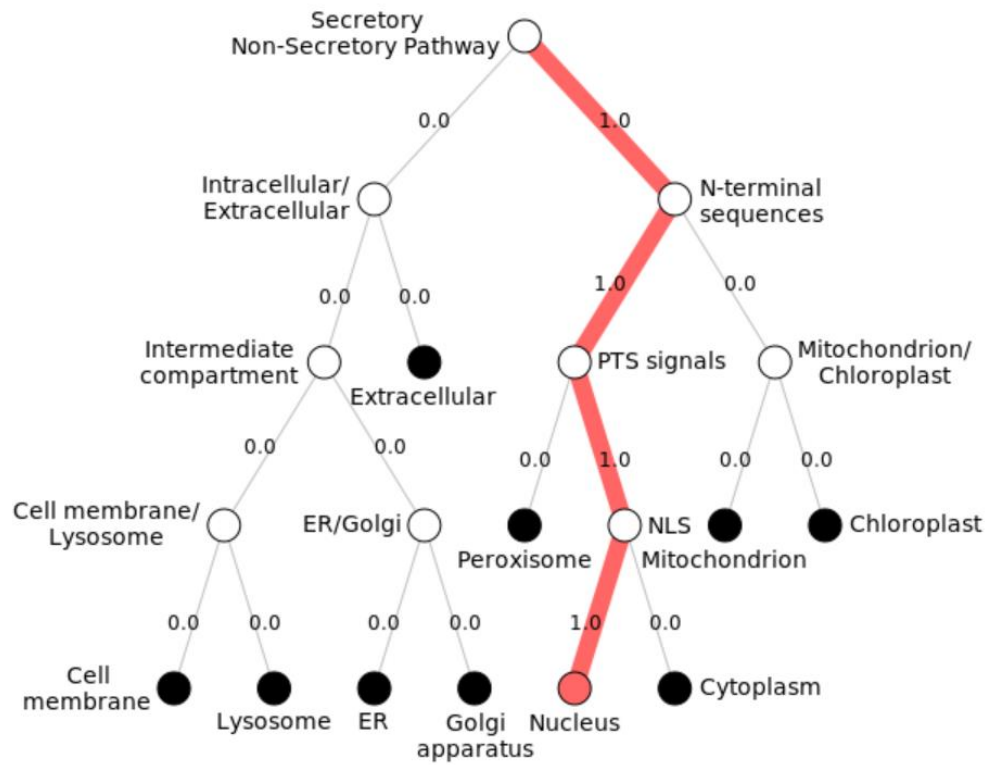
Figure 13: Shows the KLF15 protein's localization route as well as the likelihood score. The path of localization is shown in red.
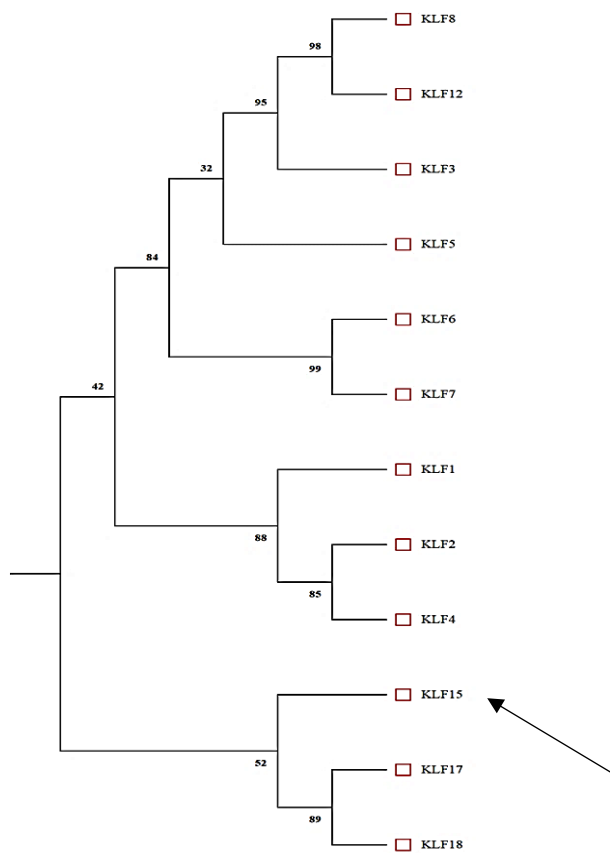
Figure 14: Represents the Phylogenetic Tree of the KLFs proteins. All the KLFs originated from a common root and then evolute into three different classes

Using Clustal Omega to align the sequences of the KLFs family's isoforms reveals that whereas some parts are shared by all members of the family, others differ. The Krüppel-like factor (KLF) family is characterized by highly conserved C-terminal regions containing three C2H2 zinc fingers, which are critical for DNA binding, while the N-terminal regions show significant variability and are involved in transcriptional regulation and protein interactions. A study (Ling et al., 2023) demonstrate their conserved nature and

show that the C2H2 zinc fingers are essential for nuclear localization and DNA-binding capabilities across several KLFs. Emphasizing this conserved role, KLF1, KLF4, and KLF8 use their zinc fingers for nuclear localization. On the other hand, the N-terminal areas show great variety, which helps to define different regulating roles and interactions with other cellular proteins. As Figure 15 shows, this diversity allows KLFs to engage in a broad spectrum of cellular activities including differentiation, cell cycle control, and responsiveness to extracellular signals.

```
KLF2-201    ELLEAKPKRGRRSWPRKRTATHTCSYAGCGKTYTKSSHLKAHLRTHTGEKPYHCNWDGCG    310
KLF4-201    MPEEPKPKRGRRSWPRKRTATHTCDYAGCGKTYTKSSHLKAHLRTHTGEKPYHCDWDGCG    434
KLF5-201    QNIQPVRYNRRSNPDLEKRRIHYCDYPGCTKVYTKSSHLKAHLRTHTGEKPYKCTWEGCD    411
KLF6-206    TSGKPGDKGNGDASPDGRRRVHRCHFNGCRKVYTKSSHLKAHQRTHTGEKPYRCSWEGCE    238
KLF7-201    QSDSDQGGLGAEACPENKKRVHRCQFNGCRKVYTKSSHLKAHQRTHTGEKPYKCSWEGCE    257
KLF3-201    --PGKRPLPVESPDTQRKRRIHRCDYDGCNKVYTKSSHLKAHRRTHTGEKPYKCTWEGCT    298
KLF8-204    ----MAQMQGEESLDLKRRRIHQCDFAGCSKVYTKSSHLKAHRRIHTGEKPYKCTWDGCS    312
KLF12-203   SIESTRRQRRSESPDSRKRRIHRCDFEGCNKVYTKSSHLKAHRRTHTGEKPYKCTWEGCT    355
KLF18-201   QLPKQKTQSCQFWKNPEVSRPYVCTYEDCKMSYSKACHLRTHMRKHTGEKPYVCDVEGCT    1002
KLF17-201   -EKNSRPQEGTGRRGSSEARPYCCNYENCGKAYTKRSHLVSHQRKHTGERPYSCNWESCS    321
KLF15-201   -----MGQKFPKNPAAELIKMHKCTFPGCSKMYTKSSHLKAHLRRHTGEKPFACTWPGCG    359
KLF11-201   ----TSSQNCVPQVDFSRRRNYVCSFPGCRKTYFKSSHLKAHLRTHTGEKPFNCSWDGCD    432
KLF10-201   -------AKVTPQIDSSRIRSHICSHPGCGKTYFKSSHLKAHTRTHTGEKPFSCSWKGCE    407
KLF9-201    -----LHPGVAAKGKHASEKRHKCPYSGCGKVYGKSSHLKAHYRVHTGERPFPCTWPDCL    181
KLF13-201   ----RVRRGRSRADLESPQRKHKCHYAGCEKVYGKSSHLKAHLRTHTGERPFACSWQDCN    205
KLF14-201   ----AADQAPRRRSVTPAAKRHQCPFPGCTKAYYKSSHLKSHQRTHTGERPFSCDWLDCD    233
KLF16-201   -------------APSAAAKSHRCPFPDCAKAYYKSSHLKSHLRTHTGERPFACDWQGCD    165
                         : *  . .*    * * .** :* * ****:*: *    .*


KLF1-201    WRFARSDELTRHYRKHTGQRPFRCQLCPRAFSRSDHLALHMKRHL---------------    362
KLF2-201    WKFARSDELTRHYRKHTGHRPFQCHLCDRAFSRSDHLALHMKRHM---------------    355
KLF4-201    WKFARSDELTRHYRKHTGHRPFQCQKCDRAFSRSDHLALHMKRHF---------------    479
KLF5-201    WRFARSDELTRHYRKHTGAKPFQCGVCNRSFSRSDHLALHMKRHQN--------------    457
KLF6-206    WRFARSDELTRHFRKHTGAKPFKCSHCDRCFSRSDHLALHMKRHL---------------    283
KLF7-201    WRFARSDELTRHYRKHTGAKPFKCNHCDRCFSRSDHLALHMKRHI---------------    302
KLF3-201    WKFARSDELTRHFRKHTGIKPFQCPDCDRSFSRSDHLALHRKRHMLV------------    345
KLF8-204    WKFARSDELTRHFRKHTGIKPFRCTDCNRSFSRSDHLSLHRRRHDTM------------    359
KLF12-203   WKFARSDELTRHYRKHTGVKPFKCADCDRSFSRSDHLALHRRRHMLV------------    402
KLF18-201   WKFARSDELNRHKKRHTGERPYLCSICSKNFARSDHLKQHAKVHNIRPGL----------    1052
KLF17-201   WSFFRSDELRRHMRVHTRYRPYKCDQCSREFMRSDHLKQHQKTHRPGPSDPQANNNNG--    379
KLF15-201   WRFSRSDELSRHRRSHSGVKPYQCPVCEKKFARSDHLSKHIKVHRFPRS----SRSVRSV    415
KLF11-201   KKFARSDELSRHRRTHTGEKKFVCPVCDRRFMRSDHLTKHARRHMTTKKIPGWQAEVGKL    492
KLF10-201   RRFARSDELSRHRRTHTGEKKFACPMCDRRFMRSDHLTKHARRHLSAKKLPNWQMEVS--    465
KLF9-201    KKFSRSDELTRHYRTHTGEKQFRCPLCEKRFMRSDHLTKHARRHTEFHPSMIKRSKKALA    241
KLF13-201   KKFARSDELARHYRTHTGEKKFSCPICEKRFMRSDHLTKHARRHANFHPGMLQRRGGGSR    265
KLF14-201   KKFTRSDELARHYRTHTGEKRFSCPLCPKQFSRSDHLTKHARRHPTYHPDMIEYRGRRRT    293
KLF16-201   KKFARSDELARHHRTHTGEKRFSCPLCSKRFTRSDHLAKHARRHPGFHPDLLRRPGARST    225
               * ***** ** : *:   : : *   * : * *****   * : *
```

Figure 15: Sequence allignment through clustal omega

```
1           11          21          31          41
MVDHLLPVDE  NFSSPKCPVG  YLGDRLVGRR  AYHMLPSPVS  EDDSDASSPC
51          61          71          81          91
SCSSPDSQAL  CSCYGGGLGT  ESQDSILDFL  LSQATLGSGG  GSGSSIGASS
101         111         121         131         141
GPVAWGPWRR  AAAPVKGEHF  CLPEFPLGDP  DDVPRPEQPT  LEEIEEFLEE
151         161         171         181         191
NMEPGVKEVP  EGNSKDLDAC  SQLSAGPHKS  HLHPGSSGRE  RCSPPPGGAS
201         211         221         231         241
AGGAQGPGGG  PTPDGPIPVL  LQIQPVPVKQ  ESGTGPASPG  QAPENVKVAQ
251         261         271         281         291
LLVNIQGQTF  ALVPQVVPSS  NLNLPSKFVR  IAPVEIAAKP  VGSGPLGPGP
301         311         321         331         341
AGLLMGQKEP  KNPAAELIKM  HKCTFPGCSK  MYTKSSHLKA  HLRRHTGEKP
351         361         371         381         391
FACTWPGCGW  RFSRSDELSR  HRRSHSGVKP  YQCPVCEKKF  ARSDHLSKHI
401         411
KVHRFPRSSR  SVRSVN
```

e – An exposed residue according to the neural network algorithm.

b – A buried residue according to the neural network algorithm.

f – A predicted functional residue (highly conserved and exposed).

s – A predicted structural residue (highly conserved and buried).

X – Insufficient data – the calculation for this site was
performed on less than 10% of the sequences.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |

Variable   Average   Conserved

Insufficient Data
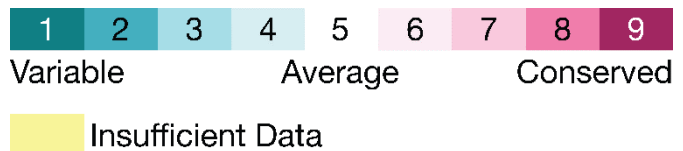
Figure 16: ConSurf investigation of KLF-15 . The diagram depicts the evolutionary conservation scale and showcases the PKC γ residues categorized as buried (b), exposed (e), functional (f), and structural (s).

4.3 In silico mutagenesis

Insilco mutagenesis induced via PyMol software, where the amino acid arginine at position 364 R is altered to P.

R 364 P

Figure 17: Insilco mutagenesis induced via PyMol software, where the amino acid arginine at position 364 R is altered to P.

R343C
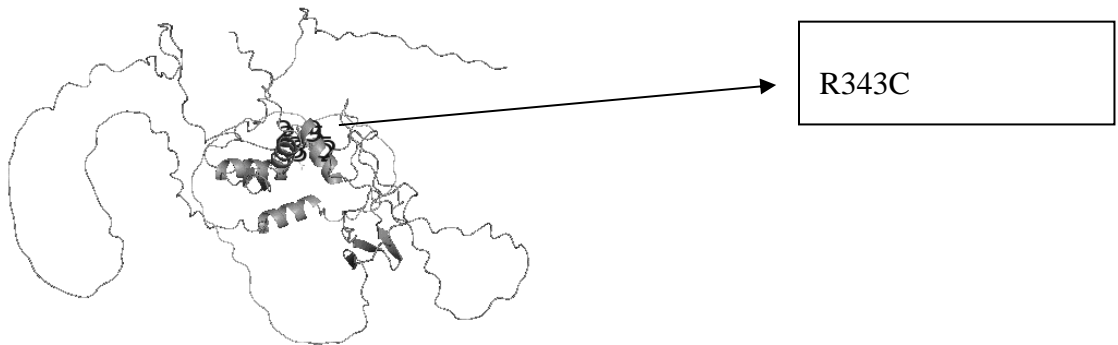
Figure 18: Insilco mutagenesis induced via PyMol software, where the amino acid arginine at position 343 R is altered to C.

4.4 Identification of KLF Variants

We used three tools for data retrieval: ENSEMBLE, GENOMAD, and COSMIC. ENSEMBLE revealed that the transcript has three exons, twenty-eight domains and features, forty-48 variant alleles, and maps to two hundred and eighty-two oligo probes and four hundred and sixteen amino acids. COSMIC also supplied a variant table. Ensemble has the most entries (4982 to be exact), according to the data gathered from different sources. The number of entries in the GenomAD database is substantially lower at 418. Out of the three databases, the cosmic database has the fewest entries 298 to be exact. When contrasted with the more niche collections included in GenomAD and cosmic, the Ensemble database stands out for its comprehensive breadth.
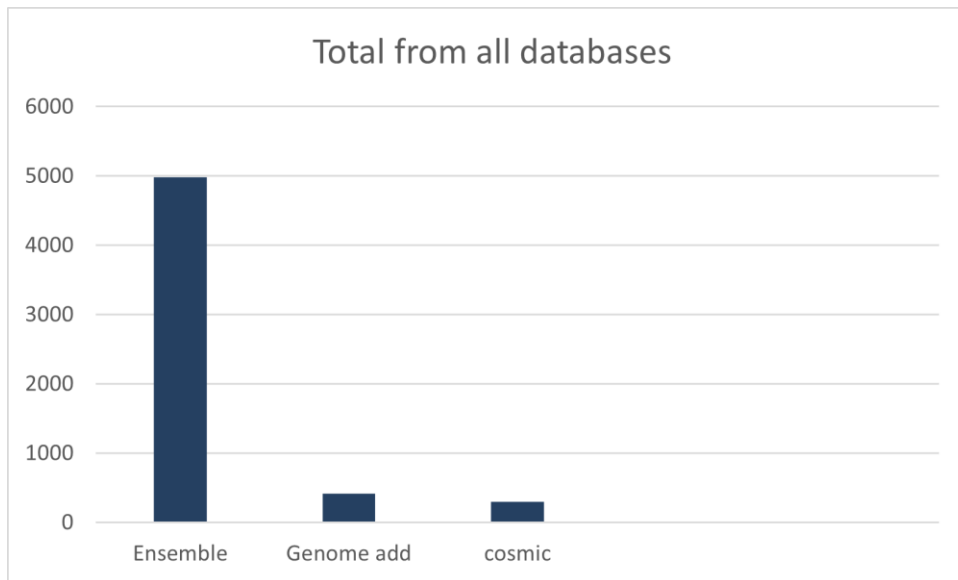
Figure 19: Total variants from all the selected data bases
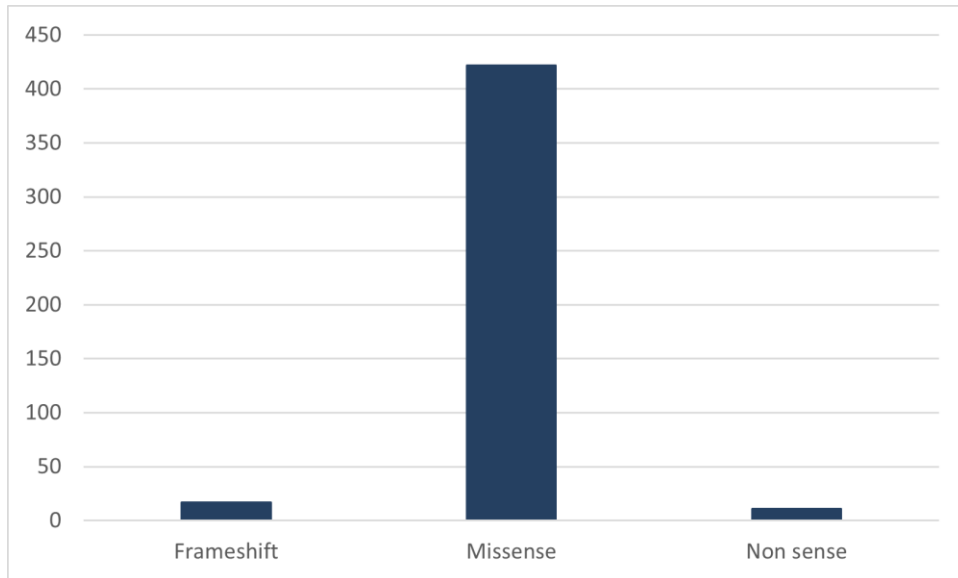
Figure 20: Non-synonymous variants from all the selected data bases

The data illustrates the number of missense variants reported across different databases. The missense category itself has a total of 422 entries. Among the specific databases, Ensemble has recorded 330 missense variants, while GenomAD has documented 212, and cosmic has reported 131. This distribution underscores the varying levels of missense variant documentation across these databases.
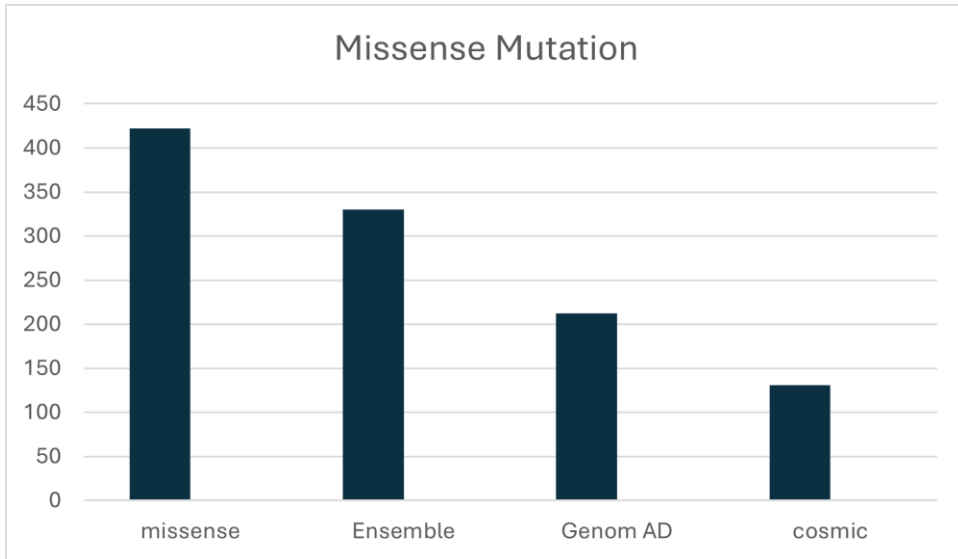
Figure 21: Unique missense variants from all the selected databases

The data from various databases highlight the total number of entries for different types of genetic variations. Missense variants are the most prevalent, with a total of 422 entries. Frameshift variants follow with 17 entries, while nonsense variants are the least common, with only 11 entries. This summary provides a clear comparison of the frequency of these genetic variations across all databases.
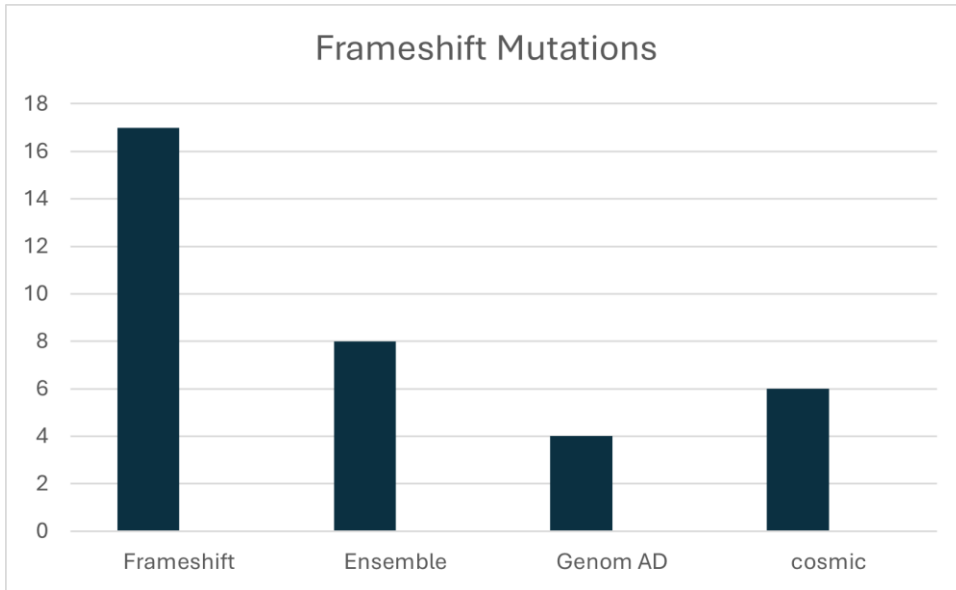
Figure 22: Total frameshift variants from the selected databases

The total number of frameshift variants across all databases is 17. Among these, Ensemble has recorded 8 frameshift variants, GenomAD has documented 4, and cosmic has reported 6. This data indicates the distribution of frameshift variants within these specific databases.

In analyzing the data provided, the term "nonsense" appears most frequently, with a count of 11, indicating its significant relevance or focus in the given context. "Cosmic" follows with a count of 7, suggesting it also plays a notable role but to a lesser extent. "Ensemble" is mentioned 5 times, showing a moderate level of importance. Lastly, "Genom AD" appears only once, implying it is the least emphasized term among the four. This distribution highlights a primary focus on "nonsense" and "cosmic," with secondary attention to "ensemble," and minimal mention of "Genom AD."

Non sense mutations
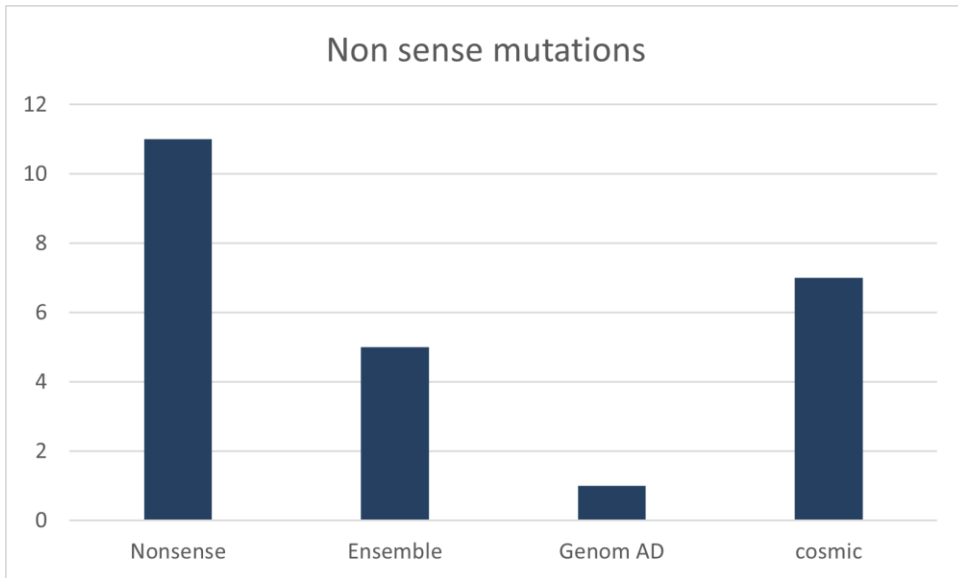
KLF15 has 4 exons, one exon has codes for 42% AA and 68% variants frequency, exon 2 codes for 38 % AA and 62% variant frequency, exon 3 codes for 40% AA and 60% variants frequency, exon 4 codes for 45 % AA and 55% variants frequency.
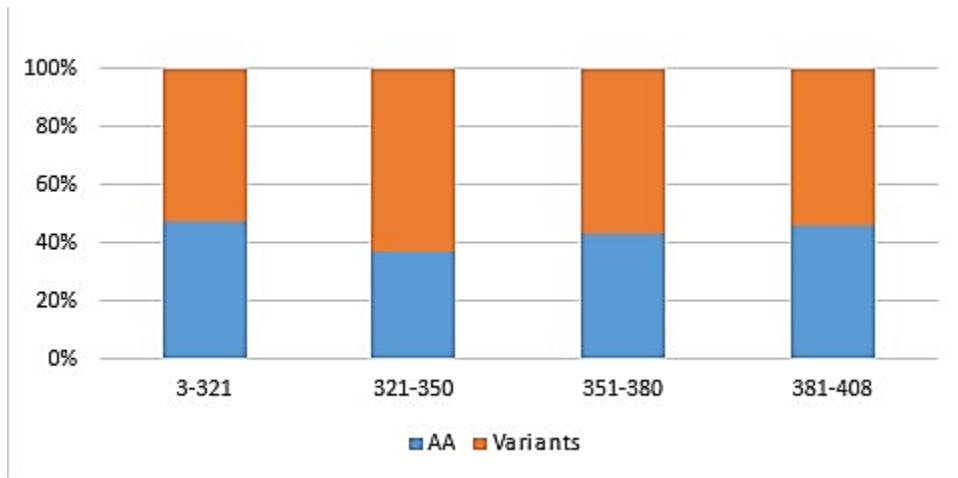
Figure 23: SNP frequency per exon

4.5 Pathogenicity analysis

In order to focus on one SNP for additional analysis, the missense data was filtered. Initially, the SNPs were sorted according to the percentage of pathogenicity. The pathogenicity percentages were calculated using SIFT, PolyPhen, CADD, REVEL, Mutation Assessor, and MetaLR. Following that, the pathogenicity percentage for each variant was calculated and displayed graphically.

Table 2: Filtered SNPs for analysis

Variant ID	Alleles	Conseq. Type	AA	AA coord	sift_class	Polyphen	Cadd	Revel	meta_lr	Mutation Accessor
rs755719419	C/G	missense variant	R/P	364	deleterious	probably damaging	likely deleterious	likely disease causing	Tolerated	Medium
rs768676875	G/A	missense variant	R/C	343	deleterious	probably damaging	likely deleterious	likely disease causing	Tolerated	Medium

49

The missense data was sifted through to isolate one SNP for further investigation. The SNPs were initially filtered based on their pathogenicity percentages. SIFT, PolyPhen, CADD, REVEL, Mutation Assessor, and MetaLR pathogenicity percentages were determined using the pathogenicity scores provided by the tools stated in the methodology. After that, the percentage pathogenicity for all variants was determined and plotted in a clustered column graph. The above-mentioned techniques were used to filter all data about KLF-15 missense SNPs. Missense SNPs above the 75% threshold were selected, and further filtration was done to sort out the most pathogenic SNPs. The SIFT; harmful SNPs (score: 0 0.01), PolyPhen; certainly, damaging and maybe damaging SNPs (score: 0.983-1), Revel; likely disease-causing SNPs (score: 0.804-0.978), MetaLR; damaging SNPs (0.77-1), CADD score cutoff value was maintained at 27, and Mutation Assessor (score: 0.5-0.9); medium and high scoring SNPs were maintained in the filtration process. After the filtration, twenty-three highly deleterious variants were sorted out. After all the screening the SNPs with variant IDs rs755719419 and rs768676875 were selected for further in silico analysis and wet lab validation. The missense variants per amino acid residue were determined to see the abundance of variants present in the coding region.

nsSNPs
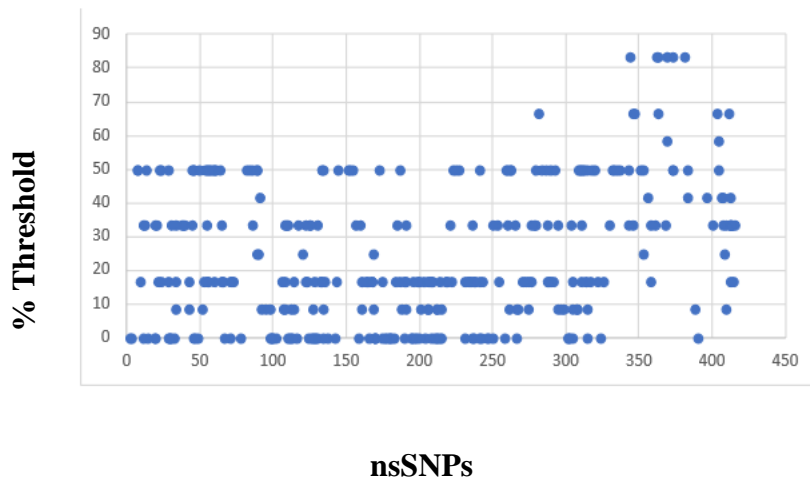
Figure 24 : Pathogenicity Percentage above Threshold Level

4.6 Stability analysis

MUpro was done for both variants and results are given below. MUpro analysis was conducted for two variants, rs755719419 and rs768676875, employing SVM and Neural Network models with sequence-only inputs. For rs755719419, SVM predicted a decrease in stability with a $\Delta\Delta G$ value of -1.5224059 kcal/mol, corroborated by Neural Network's prediction with a confidence score of -0.502399548128. Similarly, for rs768676875, SVM indicated decreased stability, supported by both the $\Delta\Delta G$ value of 0.8825882 kcal/mol and Neural Network's prediction with a confidence score of -0.945564764538754. These findings suggest potential destabilizing effects associated with the analyzed variants.

Table 3 : Variant ID: rs755719419 results from MUpro

Method	Model	Input	Prediction	Confidence Score
1	SVM	Sequence only	DECREASE stability -1.5224059 (Delta delta G Value)	-1.9224059
2	SVM	Sequence only	DECREASE stability	-0.32174172
3	Neural Network	Sequence only	DECREASE stability	-0.502399548128

Table 4: Variant ID: rs768676875 results from MUpro

Method	Model	Input	Prediction	Confidence Score
1	SVM	Sequence only	-Delta delta G =0.8825882 (DECREASE stability)	-1.5224059

| 2 | SVM | Sequence only | DECREASE stability | -0.90835183 |
| 3 | Neural Network | Sequence only | DECREASE stability | - 0.945564764538754 |

For variant rs755719419, characterized by the substitution R364P in the KLF15_Human gene, MutPred2 assigned a score of 0.881. Notably, the mutation affects multiple PROSITE and ELM motifs, including ELME000053, ELME000062, ELME000136, ELME000159, PS00006, and PS00028. Molecular analysis indicates probable alterations in DNA binding, supported by a probability of 0.35 and a significant p-value of 0.05. Additionally, a potential loss of helical structure is suggested, with a probability of 0.32 and a p-value of 0.05. Conversely, variant rs768676875, featuring the substitution R343C in KLF15_Human, received a MutPred2 score of 0.733. These variants influences various PROSITE and ELM motifs, including ELME000008, ELME000012, ELME000062, ELME000102, ELME000108, ELME000334, PS00004, and PS00028. With a p-value of 0.05 and a chance of 0.35, molecular analysis points to possible loss of inherent disorder. Furthermore, indicated alterations in DNA binding give a p-value of 0.05 and a likelihood of 0.15. These investigations help to clarify the possible functional results of the changes in protein structure and molecular mechanisms.

Table 5: Variant ID: rs755719419 results from MutPred2

ID	Substitution	MutPred2 Score	Remarks	Affected Prosite and ELM Motife

KLF15_Human	R364P	0.881	-	ELME000053, ELME000062, ELME000136, ELME000159, PS00006, PS00028
Molecular mechanism with P-values <= 0.05			Probability	P-value
Altered DNA binding			0.35	0.05
Loss of Helix			0.32	0.05

Table 6: Variant ID: rs768676875 results from MutPred2

ID	Substitution	MutPred2 score	Remarks	Affected PROSITE and ELM Motifs
KLF15_HUMAN	R343C	0.733		ELME000008, ELME000012, ELME000062, ELME000102, ELME000108,

				ELME000334, PS00004, PS00028
Molecular mechanisms with P-values <= 0.05	Probability	P value		
Loss of Intrinsic disorder	0.35	0.05		
Altered DNA binding	0.15	0.05		

4.7 MAESTRO web

We obtained prediction insights for the variant rs7557194 which is denoted by the substitution R364A using MAESTRO web. The expected ΔΔG value of 1.5160718 for this mutation suggests a possible destabilizing influence on the protein structure. Attached furthermore to this forecast is a confidence score of 0.76848. Conversely, S343A was evaluated as a substitute and displayed a ΔΔG value of 0.3760340 and a higher confidence score of 0.970638 (c_pred). These predictions provide vital information on how amino acid substitutions may influence protein stability, so helping us to learn more about the effects of variations on protein structure and function. In MAESTRO, if ΔΔG value is less than 0.0, it is stabilizing.
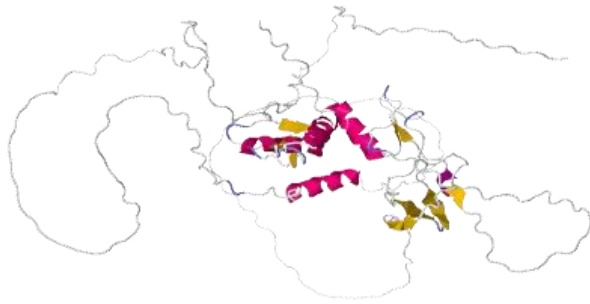
Table 7: Variant ID: rs755719419 results from MAESTRO web

Substitution	ddG_pred	c_pred
R364.A(P}	1.516071863	0.76848

Table 8: Variant ID: rs768676875 results from maestro

substitution;	ddG_pred;	c_pred
S343.A(A}	0.376034087	0.970638

4.8 DYNAMUT

From their ΔΔG values, both variation 1 and variant 2 exhibit destabilizing effects. The ΔΔG value of variation 1, -0.996 kcal/mol, shows a clear drop in stability when compared to the wild-type protein. With a ΔΔG value of -0.437 kcal/mol, variation 2 also shows a tendency towards destabilization. As our data show that these polymorphisms may influence protein stability, more investigation on the functional consequences of these variants and their putative relationships with disease phenotypes is required.
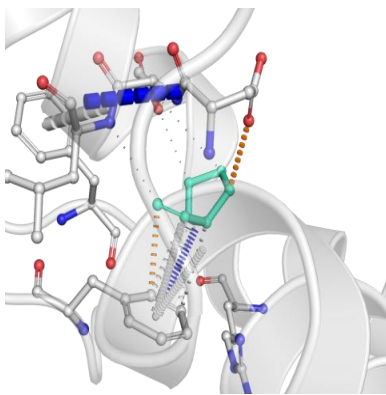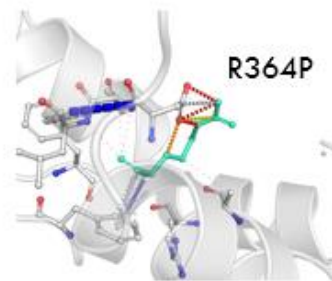
Wild type KLF-15   **Mutant 1**   **Mutant 2**
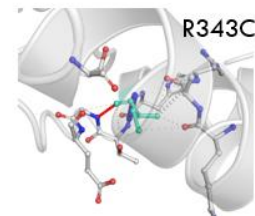
4.9 Hope

The effects on size, hydrophobicity, and charge mean that substituting cysteine for arginine at position 343 will interfere with the DNA-binding zinc-finger domain. The mutation will

affect the Zinc Finger C2H2-type and Superfamily domains since the residue is highly conserved and close to another conserved site. This suggests the mutation will destroy domains.

Table 9: Variant ID: rs768676875 results from ProjectHOPE

Category	Information
Amino Acid	Mutation: Arginine to Cysteine at position 343
Structures	Original (left) and mutant (right) amino acid structures provided. Mutant is smaller, has a neutral charge, and is more hydrophobic.
Structure	Mutation disturbs a Zinc-finger domain known to bind DNA.
Conservation	Residue is 100% conserved; mutation is likely damaging based on conservation. Mutant residue is near a highly conserved position.
Domains	Zinc Finger C2H2-Type and Zinc Finger C2H2 Superfamily domains affected.
Amino Acid Props	Charge loss in mutant, size difference (smaller mutant), and increased hydrophobicity.
Images	No structural information available, so no images generated.
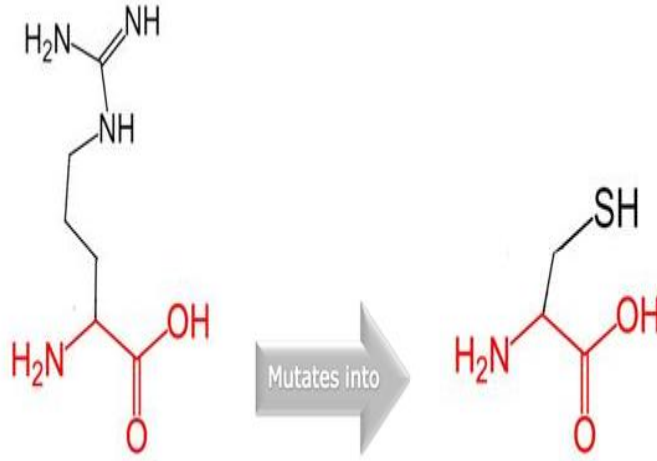
Figure 25: Mutation of Arginine to Cysteine

The size and charge of an amino acid changes from positively charged to neutral and hydrophobicity increases as glycine is converted to alanine. This mutation affects the Zinc Finger C2H2-Type and Superfamily domains, which may impair DNA binding. The mutation's closeness to a conserved location and high conservation suggests it's harmful. This may affect protein folding, size, hydrophobicity, charge, and interactions. Table shows results from HOPE rs755719419.

Table 10: Variant ID: rs755719419 results from MAESTROweb

Category	Information
Amino Acid properties	Mutation from Glycin to Alanin
Charge	Positive to neutral
Size	Larger to small
Hydrophobicity	Less hydrophobic to more hydrophobic
Zinc finger C2H2-type	Affected

Zinc finger C2H2 superfamily	Affected
Wild type consistency	highly conserved
Mutation impact	Likely damaging based on conservation scores
Mutant residue	Located near a highly conserved position
Impact on charge	Loss of charge may impact interactions
Impact on size	Smaller size may lead to loss of interactions
Impact on hydrophobicity	Increased hydrophobicity can disrupt hydrogen bonding and folding
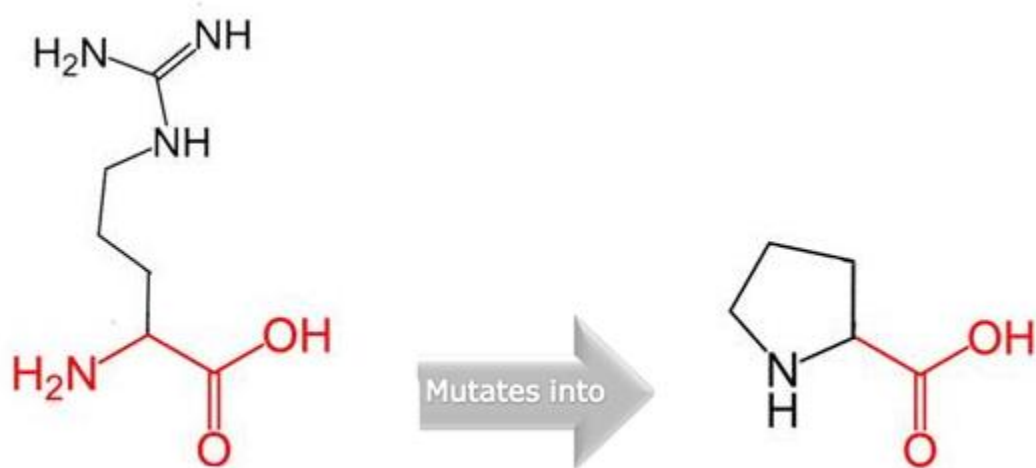
Figure 26: Mutation of Glycin to Alanin

60

4.10 FATHMM

Table 11: FATHMM analysis of variant ID (rs755719419)

Chromosome	Position	Variant	Coding score	Noncoding score	Messages
3	126343887	A/G		0.537127	oncogenic

Table 12: FATHMM analysis of variant ID (rs768676875)

Chromosome	Position	Variant	Coding score	Noncoding score	Messages
3	**126351896**	**G/A**		**0.442581**	**Benign**

The first table describes a variant located on chromosome 3 at position 126343887, with an A/G substitution. This variant has a noncoding score of 0.537127, indicating a moderate potential impact on gene regulation. The variant is classified as "oncogenic," suggesting it may contribute to cancer development. In contrast, the second table features a variant on chromosome 3 at position 126351896, with a G/A substitution. This variant has a noncoding score of 0.442581, indicating some potential impact, though less significant than the first variant. It is labeled as "benign," implying that it is not associated with cancer or harmful effects.

4.11 RNA fold

The variant exhibits a Free Energy of -15.23 kcal/mol, indicating the Gibbs free energy of the thermodynamic ensemble. This value suggests a certain level of stability; however, it's essential to note that lower values signify greater stability in the system at constant temperature and pressure. The MFE Structure Frequency stands at 26.09%, indicating the prevalence of the Minimum Free Energy (MFE) structure within the ensemble. Ensemble Diversity is measured at 4.58, indicating a relatively diverse ensemble of structures. Comparatively, the Centroid Structure Energy is -14.10 kcal/mol, representing the energy of the centroid secondary structure in dot-bracket notation, which is the structure with the minimum free energy within the ensemble. This variant's MFE and structural parameters differ from those of the wild type, suggesting potential effects on stability and structural dynamics. Further analysis is warranted to elucidate the functional implications of these differences in the variant.
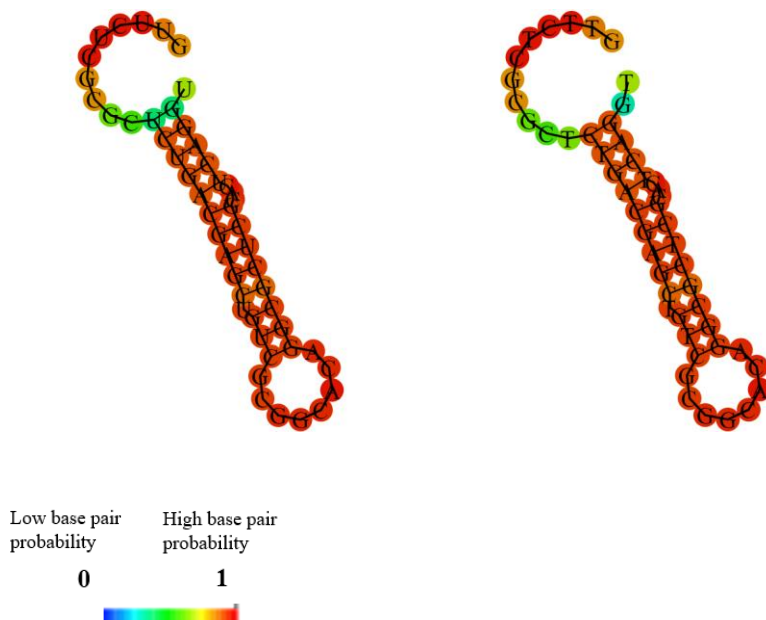
Figure 27: MFE of variant is -14.10 kcal/mol (left) and the wild type is of -15.23 kcal/mol (right)

The variant under scrutiny exhibits a Free Energy of -102.62 kcal/mol, reflecting the Gibbs free energy of the thermodynamic ensemble. Lower values suggest greater stability within the ensemble. The MFE Structure Frequency is notably low at 0.13%, indicating a sparse occurrence of the Minimum Free Energy (MFE) structure within the ensemble, implying a diverse set of structures. Ensemble Diversity is high at 37.05, signifying a broad spectrum of possible structures present within the ensemble. The Centroid Structure Energy is -95.80 kcal/mol, denoting the energy of the centroid secondary structure in dot-bracket notation, which represents the most thermodynamically stable structure within the ensemble. These parameters collectively highlight the variant's distinct RNA secondary structure characteristics compared to the wild type, suggesting potential alterations in stability and structural dynamics. Further investigation is warranted to comprehend the functional implications of these differences in the variant's RNA structure.
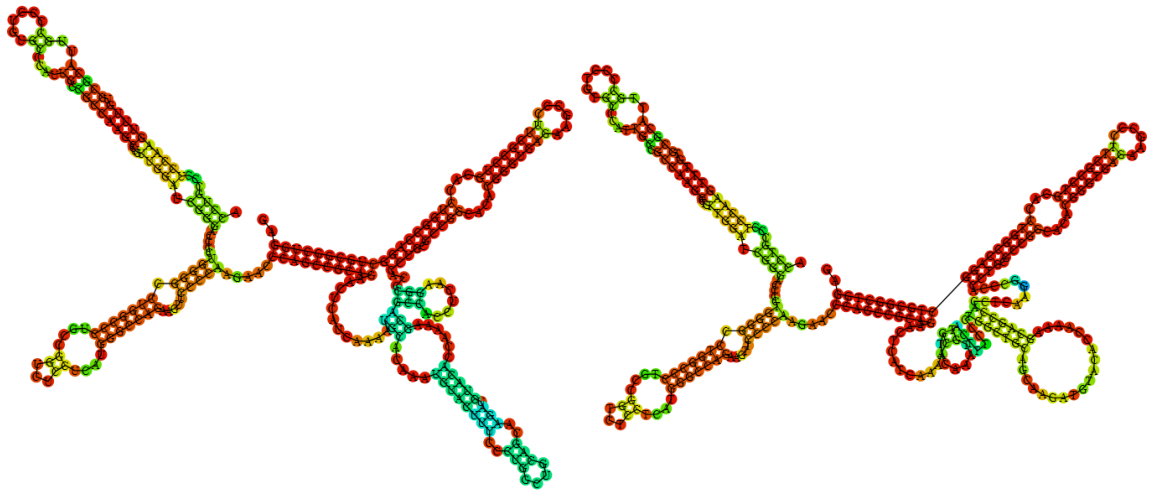
Figure 28: MFE is -95.80 kcal/mol for a variant (left) --102.62 kcal/mol for a wild type (right)

4.12 Molecular Dynamics

a)

RMSF Plot

b)

Radius of Gyration Over Time

c)

Hydrogen Bonds Over Time

d)

Solvent Accessible Surface Area Over Time
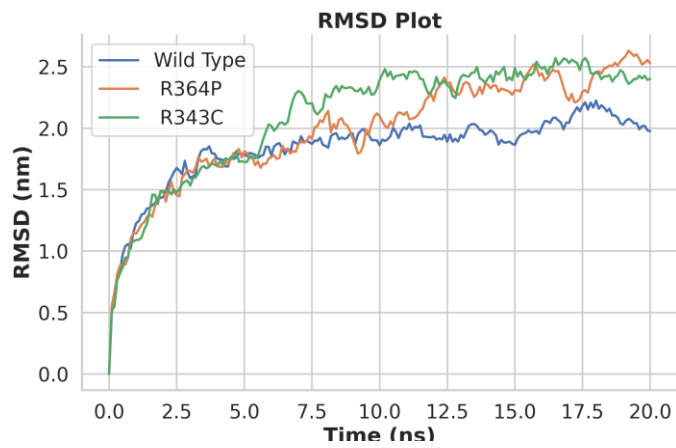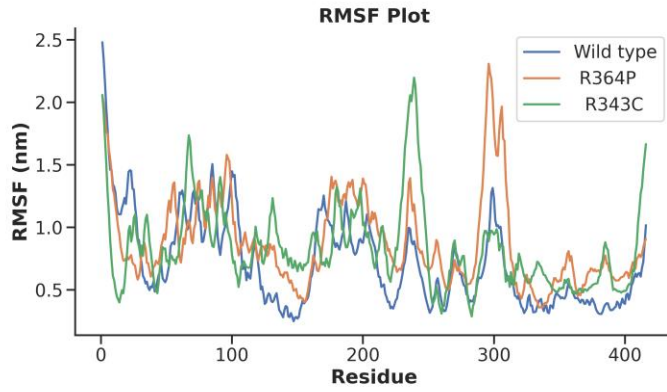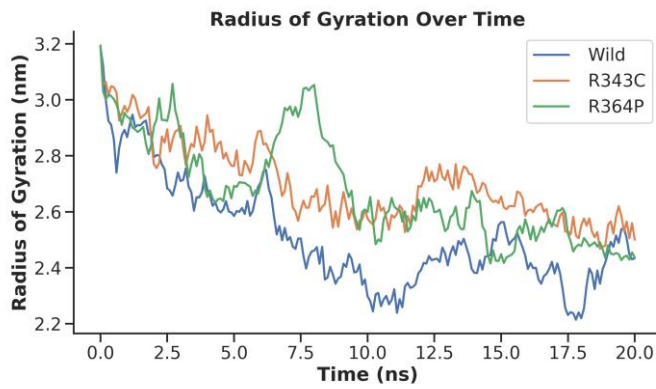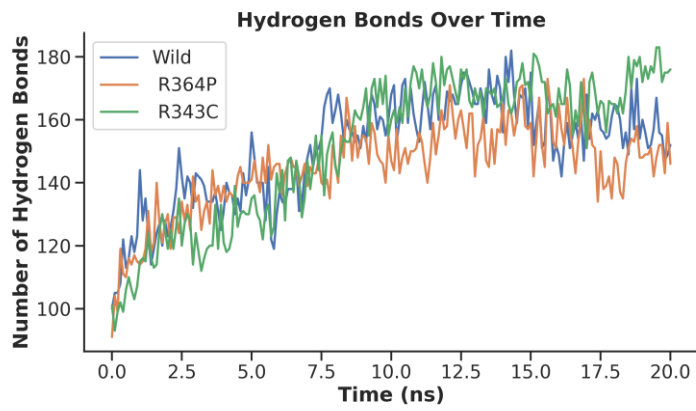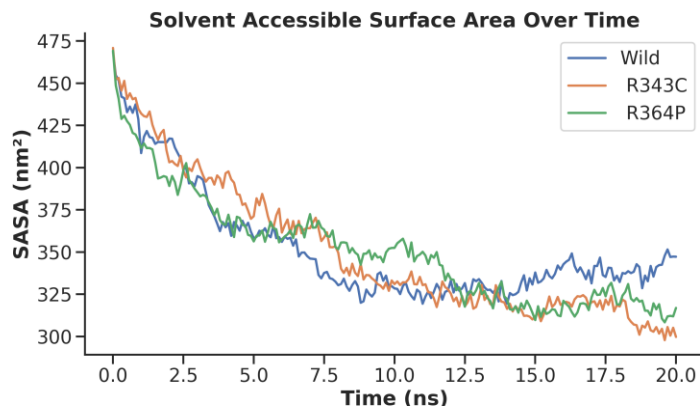
e)

Figure 29: Molecular dynamics of KLF-15 and its SNPs. a) Root mean square deviation (RMSD), b) Root mean square fluctuation (RMSF), c) Radius of gyration, d) Number of hydrogen bonds, and e) Solvent accessibility surface area (SASA).

The first graph (a) displays the RMSD (Root Mean Square Deviation) with time, therefore displaying the structural stability of the proteins. The Wild Type protein indicates more stability by having the lowest RMSD values over the simulation when compared to the R364P and R343C variants. Emphasizing how flexible each residue is, the second graph (b) displays the link between the number of residues and the root mean square fluctuation (RMSF). The R343C variety exhibits higher fluctuations than the Wild Type and R364P, which would indicate that it is perhaps more flexible and unstable in some places.

Plotting the Radius of Gyration with time in the third graph (c) helps one to demonstrate the simplicity of the protein structures. Unlike R343C and R364P, which show more variability, the Wild Type maintains mostly constant radius, implying a strong compact

66

structure. The Wild Type generates more stable hydrogen bonds, which keeps it together, as seen in the fourth graph (d), which displays the quantity of hydrogen bonds with time. Finally, graph (e) displays the Solvent Accessible Surface Area (SASA) with time; a lower SASA for the Wild Type denotes less solvent exposure and more robust stability. Based on the whole analysis, the Wild Type protein shows greater stability, compactness, and structural integrity than the variants in R364P and R343C.

4.13 Laboratory based Experimentation Results

Genotype Data of Hepatic Cancer in Patients and Healthy Control Samples

Table 13: Genotype Data of Patient and Control of R364P(rs755719419) Mutation

		Patients %	Controls %	Odds ratio	95% CI (odds ratio)	Relative risk	95% CI (relative risk)	P value
Genotypes	**CC**	18.18 %	61.46%	0.139	0.07492 to 0.2642	0.3572	0.2370 to 0.5171	<0.005
	GG	19.09 %	2.08%	2.933	2.772 to 48.66	1.651	1.460 to 2.253	
	CG	62.73 %	36.46%	11.09	1.658 to 5.150	1.877	1.266 to 2.190	
Alleles	**C**	49.09 %	80.21%	0.2379	0.1298 to 0.4431	0.5521	0.4310 to 0.7040	<0.005
	G	50.91 %	19.79%	4.203	2.257 to 7.705	1.811	1.421 to 2.320	

Table 13 shows for both patients and controls the genotyping data for the R364P (rs755719419) mutation. Looking at the relative risk (RR) and odds ratio (OR), the control group clearly differs from the patients for several genotypes and alleles. Given their statistically significant RR and OR values, the distribution of these genotypes and alleles between the two groups likewise follows. P-values below 0.05 also help to corroborate this. The CG genotype and G allele of the patient group raise their risk of the disease since their RR and OR values far exceed 1. These results suggest that, given its distribution across patients and controls, the R364P mutation might have a function in disease susceptibility. GG and CG genotype is significantly associated with disease due to high OR and RR. while genotype CC is significant and has a protective role.

Table 14: Association of R364P(rs755719419) SNP with Clinical Features of hepatic Cancer Patients.

					Positive family history		Negative family history	
		P value	Positive FH %	Negative FH %	Odds ratio (95% CI)	Relative risk (95% CI)	Odds ratio (95% CI)	Relative risk (95% CI)
Genotypes	CC		21.05%	2.63%	9.867 (1.604 to 108.3)	1.682 (1.175 to 2.141)	0.1014 (0.009230 to 0.6234)	0.5945 (0.4671 to 0.8514)
	GG	<0.005	21.05%	2.63%	9.867 (1.604 to 108.3)	1.682 (1.175 to 2.141)	0.1014 (0.009230 to 0.6234)	0.5945 (0.4671 to 0.8514)
	CG		94.74%	57.89%	13.09 (3.305 to 58.40)	1.93 (1.462 to 2.567)	0.07639 (0.01712 to 0.3026)	0.5181 (0.3896 to 0.6839)

Table 14 shows that, overall, family history significantly influences the connection between genotypes and disease risk. Higher OR and RR values indicate an increasing risk in cases of a positive family history; lower OR and RR values show a protective impact in cases of a negative family history. The results show their robustness since statistical significance across all genotypes and family histories supports them.

Table 15: Association of R364P(rs755719419) SNP with Clinical Features of hepatic Cancer Patients.

					Fatty liver		Non-Fatty liver	
		P value	Fatty liver %	Non-Fatty liver %	Odds ratio (95% CI)	Relative risk (95% CI)	Odds ratio (95% CI)	Relative risk (95% CI)
Genotypes	CC	0.7896	20.59%	17.11%	0.7959 (0.2835 to 2.067)	0.9286 (0.6076 to 1.236)	1.256 0.4838 to 3.527	1.077 (0.8093 to 1.646)
	GG	0.6008	21.05%	14.71%	0.884 (0.3993 to 1.970)	1.13 (0.7961 to 1.437)	1.131 (0.5076 to 2.505)	0.8848 (0.6961 to 1.256)
	CG	0.8334	64.71%	61.84%	1.547 (0.5254 to 4.135)	0.963 (0.7517 to 1.273)	0.6466 (0.2419 to 1.903)	1.038 (0.7856 to 1.330)

Table 15 shows that none of the genotypes (CC, GG, CG) statistically correlate with either non-fatty or fatty liver diseases. Since the P-values for all conditions and genotypes are non-significant (Ns), fatty liver disease is not much affected by these genotypes.

Table 16: Genotype Data of Patient and Control of R364P(rs755719419) Mutation

Genotype	Gender	Patient %	Control %	Odds ratio	95% CI Odds ratio	Relative risk	95% CI Relative risk	P value
CC		24.62%	60.00%	0.2177	0.09269 to 0.5197	0.556	0.3624 to 0.7895	<0.005
GG	Female	18.46%	5.71%	2.533	0.8172 to 17.44	1.375	0.9499 to 1.760	
CG		56.92%	34.29%	3.736	1.120 to 5.694	1.391	1.030 to 1.884	

CC		8.89%	62.30%	0.05905	0.02111 to 0.1918	0.1487	0.05792 to 0.3527	
GG	Male	20.00%	2.00%	4.067	3.981 to 5.964	2.283	1.810 to 6.822	
CG		71.11%	37.70%	*4.9725*	1.762 to 8.910	2.694	1.394 to 3.900	

Table 16 shows that gender influences the variation in the connection between genotypes and disease risk. In both sexes the CC genotype has a statistically significant protective effect. With the CG and GG genotype, both sexes show a rather higher risk, while the relationship is more evident in men. The statistical significance of the P-values gives strong proof for the conclusions.

5. DISCUSSION

According to the World Health Organization (WHO), liver cancer is the sixth most common cancer worldwide (WHO, 2024). The incidence of liver cancer is significantly higher in sub-Saharan African and Southeast Asian countries compared to the United States. It is the predominant form of cancer in many of these countries. Over 800,000 individuals worldwide receive a diagnosis of this illness annually. Liver cancer is a prominent contributor to global cancer mortality, resulting in over 700,000 fatalities annually (Society, 2024). The risk factors for this condition encompass hepatitis B virus, hepatitis C virus, fatty liver disease, cirrhosis caused by alcohol consumption, smoking, obesity, diabetes, iron overload, and other dietary exposures. The outlook for liver cancer is unfavorable. Only a small percentage, ranging from 5% to 15% of people meet the criteria for surgical removal. This procedure is only suited for patients in the early stages of the disease and who have a reduced ability for their liver to regenerate, often without cirrhosis. It is important to note that the right hepatectomy poses a greater risk for post-operative complications compared to left hepatectomy. For more advanced phases, the available treatment options are as follows: (a) Trans-arterial chemoembolization (TACE) results in a 23% increase in the 2-year survival rate compared to conservative therapy for patients with intermediate stage hepatocellular carcinoma (HCC). (b) Administering sorafenib orally is the preferred treatment for advanced cases, as it is a kinase inhibitor. However, less than 33% of patients experience positive outcomes from the treatment, and the development of drug resistance becomes apparent within six months after starting the regimen(Anjum et al., 2021) . Extended usage of chemotherapeutic medications, such as sorafenib, might lead to additional problems such as toxicity and/or drug ineffectiveness. Consequently, both current ablation treatments and chemotherapy have limited effectiveness in enhancing outcomes of this debilitating condition (Anwanwan et al., 2020). Hepatocellular carcinoma (HCC) was once regarded as an orphan illness. Nevertheless, due to significant advancements in comprehending its cancer-causing

mechanisms and innovative therapeutic choices, HCC has become a subject of extensive translational and clinical investigations, leading to substantial scientific advancements. Various subclasses of hepatocellular carcinoma (HCC) have been suggested according on molecular profiles. These subclasses have been associated with pathological characteristics, clinical symptoms, and the aggressiveness of the illness (Wege et al., 2021). The aim of our study is to evaluate the prognostic significance of the KLF 15 polymorphism in liver cancer patients. The two unique SNP variants in KLF-15 are identified and linked with liver cancer to analyze its effect on the Pakistani Population. More and more biomarkers are appearing that could be of benefit, it is highly probable that KLF15's function in increasing cell proliferation is responsible for its association with enhanced tumor growth in pancreatic, endometrial, and breast malignancies (Gao et al., 2017). On the other hand, researchers have paid little attention to the fundamental functions of KLF15. In light of this, we set out to determine what part KLF15 plays in liver cancer and how it works in the present study. A study by (Gao et al., 2017) showed that EMT-related markers revealed that KLF15 knockdown triggered EMT by decreasing N-cadherin and Vimentin levels and increasing E-cadherin expression, This show that KLF15 controlled EMT in LADC cancer cells, which resulted in metastasis (Gao et al., 2017). Found in eukaryotic life, the KLFs are a family of transcription factors. Among the several physiological and pathogenic mechanisms in which they are involved are cell differentiation, angiogenesis, organ and tissue development, protooncogene mutations, and early embryonic growth and development. Researchers have so discovered seventeen different KLF components in mammals, which makes them an essential regulator up to now. Based on their discovered sequence, they were allocated numerals KLF 1–17. Made by many different types of cells in many different sorts of animals, the KLF factors have many various purposes. These purposes include regeneration of tissues, promoting vascular regeneration, causing phenotypic alterations in cells, so preventing blood cell production, aiding in cancer growth, and so fostering stem cell differentiation. KLF15 brings about notable modifications in hepatic metabolism. Though it has no effect on endoplasmic reticulum stress or hepatic inflammatory response (Tian et al., 2020), deleting the KLF15

gene in mice increases insulin resistance in response to a high-fat diet (HFD). (Tian et al., 2020) claim that by either transcriptional or post-transcriptional pathways, KLF15 can regulate cell differentiation, proliferation, death, and fibrosis. KLF15-/-mice lower hepatic steatosis by blocking the motorC1 signaling pathway by raising fatty acid oxidation. The KLF15 gene is thereby absolutely vital. Controlling liver metabolism depends on the KLF15 gene; changes in its expression account for the hepatic pathologies induced by a high-fat diet (HFD) (Fan et al., 2018).

Research indicates that the expression profiles of the KLF isoenzymes differ depending on the type of cancer. Many studies have shown that some genetic variations can change protein expression, structure, and function (Robert & Pelletier, 2018). Found in the functional areas of genes, missense SNPs have demonstrated in studies to enhance oncogene function and cause structural changes (Choi et al., 2019). Missense SNPs in particular genes may increase cancer risk. Certain SNPs in the KLF family have been associated to some forms of cancer (Marrero-Rodríguez et al., 2014).

The results of DeepLoc 1.0 analysis of KLF15 imply that its nuclear localization is accurate with a probability of 0.9997. This corresponds with the function of KLF15 as a transcription factor in the nucleus. In addition, the protein is projected to be soluble with a likelihood of 0.6948. This aligns with its function in regulating gene expression, indicating that it is not expected to be bound to the cell membrane. The results depicted in Figure 2 align with the evolutionary lineage of KLF proteins, which can be traced back to a shared ancestor, as seen by the evolutionary tree generated using Clustal Omega. The evolutionary analysis measures the number of substitutions at each site, giving us information about the relative evolutionary divergence within the KLF family. This prediction highlights the fact that KLF15 is likely to be found in the nucleus and is soluble, which is consistent with its known biological roles. KLF15 evolution was analyzed by aligning the sequences of all KLF family proteins to indicate conserved areas among the proteins in this family. The phylogenetic tree that shows that all members of the KLF family descended from a common ancestor and have some conserved regions present in all KLF isoforms. KLF15

is located on chromosome 3q21.3 in humans and encodes a transcription factor protein that is composed of 416 amino acids. Previous studies have shown that KLF15 acts as a transcriptional activator for multiple significant target genes (Marrero-Rodríguez et al., 2014).

The KLF 15 protein structure was first searched in Protein Data Bank (PBD). No data was retrieved from the database that gives us the structure, so we have used AlphaFold to predict the structure of KLF15, structure retrieved from AlphaFold was viewed via PyMol to validate the structure obtained, It has previously been proven that the highly accurate protein structure prediction via AlphaFold has done in a variety of researches (Jumper et al., 2021). The latest AlphaFold model exhibits a substantial enhancement in accuracy compared to previous specialized tools. It surpasses the docking tools in accuracy for protein-ligand interactions, outperforms nucleic-acid-specific predictors in accuracy for protein-nucleic acid interactions, and demonstrates significantly improved accuracy in antibody-antigen prediction compared to AlphaFold-Multimer v2.37,8 (Abramson et al., 2024). These demonstrate that it is feasible to achieve excellent accuracy in modeling across the entire range of biomolecular structures using a single integrated deep learning framework. The study aimed to establish a connection between single nucleotide polymorphisms (SNPs) and Liver cancer. SNPs in the KLF15 gene were acquired via the Ensemble Genome Browser. Only missense single nucleotide polymorphisms (SNPs) were selected due to their direct influence on the structure of proteins and their potential to modify protein function to some degree. In order to forecast the impact of a single nucleotide polymorphism on a protein, the collected non-synonymous SNPs underwent several filtration techniques, relying on their scores in SIFT, PolyPhen, Revel, MetaLR, CADD, and Mutation Assessor. SNPs that were deemed extremely pathogenic were filtered. Following extensive filtration, a total of 25 missense single nucleotide polymorphisms (SNPs) were selected from a large pool of 429 missense SNPs that exhibit a detrimental impact. Afterwards, the list was narrowed down to two options, specifically rs755719419 and rs768676875 which were selected for further examination. KLF15 is composed of multiple functional domains, which include the C-terminal DNA-binding

75

domain, the N-terminal transactivation domain, and the zinc finger motifs that are crucial for DNA binding. KLF15 possesses a DNA-binding domain that enables it to specifically attach to particular DNA sequences located in the promoter regions of target genes. This interaction ultimately governs the expression of these genes.

The stability analysis of the selected variants rs755719419 (R364P) and rs768676875 (R343C) was performed using the tools named MutPred 2, DynaMut and HOPE, MUpro, MAESTRO, which predict that the mutation at this site may impair protein by decreasing its stability. SVM and Neural Network models with sequence-only inputs were used for MUpro analysis of rs755719419 and rs768676875. A decrease in stability was predicted for rs755719419 by SVM with a $\Delta\Delta G$ value of -1.5224059 kcal/mol and confirmed by Neural Network with a confidence score of -0.502399548128. SVM showed lower stability for rs768676875, verified by $\Delta\Delta G$ value of 0.8825882 kcal/mol and Neural Network's confidence score of -0.945564764538754. These results indicate that the variations may destabilize the protein. MutPred2 gave variation rs755719419, defined by KLF15_Human gene substitution R364P, molecular investigation suggests DNA binding changes with a probability of 0.35 and a significant p-value of 0.05. Helical structural loss is also possible with a chance of 0.32 and a p-value of 0.05. Variant rs768676875, which replaced KLF15_Human with R343C, scored 0.733 on MutPred2 and molecular analysis suggests intrinsic disorder loss with a likelihood of 0.35 and a significant p-value of 0.05. DNA binding changes are also suggested with a likelihood of 0.15 and p-value of 0.05. The results offer understanding of the possible effects of variations on protein structure and molecular dynamics. We obtained prediction insights for the variant rs7557194, which is denoted by the substitution R364P, by means of MAESTRO web. Given the expected $\Delta\Delta G$ value of 1.5160718, this mutation most certainly has a destabilising effect on the protein structure. The projection also features a 0.76848 confidence score. On the other hand, S343A was considered as a possible replacement having a $\Delta\Delta G$ value of 0.3760340 and a greater confidence score of 0.970638 (c_pred). This provides important information on how amino acid changes could affect protein stability, therefore helping one to better grasp the consequences of differences on protein structure and function. Their $\Delta\Delta G$ values help

to show the destabilizing consequences of both variants 1 and 2. Stability is considerably reduced in the ΔΔG value of -0.996 kcal/mol for rs7557194 when compared to the wild-type protein. Similar to rs768676875, a destabilizing trend is observed with a ΔΔG value of -0.437 kcal/mol. These variations may affect protein stability, indicating the need for further research into their functional effects and disease manifestations. Replacement of arginine with cysteine at position 343 should dramatically impair the DNA-binding zinc-finger domain. A bigger, positively charged, hydrophilic arginine is replaced with a smaller, neutral, hydrophobic cysteine. The modification is anticipated to be particularly harmful because the residue at this position is 100% conserved and next to another highly conserved site. The DNA-binding zinc-finger C2H2-Type and Superfamily domains are severely damaged. Loss of charge, size difference, and enhanced hydrophobicity of mutant cysteine destabilize the domain. Although no structural images are available, the data shows the mutation's profound impact on protein function. The mutation from glycine to alanine increases hydrophobicity, neutralizes charge, and reduces size. This change may decrease DNA binding in the Zinc Finger C2H2-Type and Superfamily domains. The mutation's proximity to and location at a highly conserved region suggest it's deleterious. Mutations may affect protein folding, size, hydrophobicity, charge, and interactions. Alanine's hydrophobicity and smaller size than glycine may disrupt hydrogen bonding and folding, causing essential interactions to be lost. Based on conservation scores, the mutation's loss of charge affects protein interactions, indicating its potential harm. The effect of mutant KLF15 SNPs (R364P and R343C) on protein structure and function is evaluated by observing the stability of mRNA secondary structure. The variant exhibits a Free Energy of -102.62 kcal/mol, suggesting greater stability compared to the wild type's Free Energy of -15.23 kcal/mol. Despite this, the variant shows a low MFE Structure Frequency of 0.13%, indicating a diverse ensemble of structures, while the wild type has an MFE Structure Frequency of 26.09%. The Ensemble Diversity is 37.05 for the variant, much higher than the wild type's 4.58, signifying a broader spectrum of possible structures. The Centroid Structure Energy for the variant is -95.80 kcal/mol, compared to the wild type's -14.10 kcal/mol. These differences imply that, although the variant appears more

stable thermodynamically, its increased structural diversity might affect stability and structural dynamics differently than the wild type. Further investigation is needed to understand the functional implications of these changes. Structural stability and dynamics vary greatly across Wild Type, R364P, and R343C protein variants according to molecular dynamics (MD) simulation study. Plotting RMSD and RMSF shows that the Wild Type protein has less flexibility and higher structural stability; nonetheless, the R343C variant shows larger fluctuations and probable instability in some regions. The Radius of Gyration shows more evidence that the Wild Type protein preserves a more compact structure over time than the R364P and R343C variants, whose compactness is more changeable. A higher number of stable hydrogen bonds which are necessary for maintaining the stability of the protein is linked with the enhanced structural integrity of the Wild Type protein. Given its reduced SASA values, the Wild Type protein may be more stable as it is less exposed to solvents. The findings show that among mutations in the R364P and R343C, the Wild Type protein is more stable and structurally complete. This result could influence the proteins' performance under physiological conditions as well as their general integrity. These results could clarify the molecular foundations of the mutations as well as their possible influence on protein stability and function.

The association between allele change and hepatic cancer was studied since the in-silico study indicated that a mutation in KLF-15 (R364P) could modify the structure and consequently the function of the protein. Two sets of primers "two outer and two inner" were planned for this objective utilizing Primer1 against both the variations followed by DNA extraction. Tetra ARMS-PCR followed the extraction. PCR using variant produces the same results. Distribution of this mutation, R364P, among patients and controls follows the genotypes and alleles indicated in table 13. The major P-values (<0.05) for all genotypes (CC, GG, CG) and alleles (C, G) demonstrate statistically significant variations between the two groups. Particularly, the greatly higher odds ratio (OR) and relative risk (RR) indicate a strong association between disease susceptibility and the CG, GG genotype and G allele. If the statistics are to be taken, people with the G allele or the CG, GG genotype are much more likely than those without to get the illness. This indicates that the

R364P mutation is a main cause of disease pathogenesis. Table 14 shows how genotype distribution varies depending on genetic background. The CG and GG genotypes are connected to a somewhat higher risk for persons with a positive family history based on odds ratios (OR) larger than 1 and statistically significant p-values (0.013). On the other hand, a negative family history for these genotypes (OR and RR values less than 1) points to a lower risk, which is validated by noteworthy P-values as well. The CG genotype shows a more complex link whereby a positive family history is linked with a higher risk. The results indicate how family risk interacts with hereditary elements, therefore highlighting the importance of genetic predisposition in view of a person's family tree. Table 15 examines the relationship between genotypes and fatty liver. The non-significant P-values (Ns) suggest that neither fatty liver nor non-fatty liver disorders have a statistically significant correlation between any of the genotypes (CC, GG, CG). This information suggests that the R364P mutation has no effect on the likelihood of fatty liver. The absence of statistical significance of the R364P mutation in this environment suggests that other environmental or genetic factors might be more crucial in the beginning of fatty liver. Table 16 shows genotype distributions by gender. The fact that the CC genotype is connected to a rather reduced risk of disease in females (P-value 0.001) suggests a preventive influence. Conversely, the CG, GG genotype in women carries a higher risk (P=0.0373). Whereas the CC genotype shows a clear protective impact (P-value <0.0001) in men, the GG and CG genotypes are substantially related with an enhanced risk of illness (P-values 0.0003 and 0.0008, respectively). These results support the theory that genetic counseling and risk assessments customized to each gender would be beneficial given the significant differences in genetic risk profiles between the sexes.

SUMMARY OF RESEARCH WORK

This study aims to determine, with a focus on the Pakistani population especially, whether or not Krüppel-like factor 15 (KLF15) polymorphisms have any predictive power for liver cancer outcomes. Though it is a major global issue, liver cancer is more common in Southeast Asia and sub-Saharan Africa. Examining the likely role of several KLF15 single nucleotide polymorphisms (SNPs) in the gene helps this study test the hypothesis that some variations influence cancer risk and development.

This work extensively applied in silico and lab-based analysis. We validated the estimated KLF15 protein structure obtained using AlphaFold, PyMOL and InterPro. Pathogenicity prediction was done utilising programmes like SIFT, PolyPhen, CADD, REVEL, and

Mutation Assessor after extracting missense SNPs from databases including ENSEMBLE, GENOMAD, and COSMIC. Two pathogenic variants rs755719419 (R364P) and rs768676875 (R343C) were investigated further. For stability studies, we applied MUpro, MAESTRO, and DynaMut; for structural and functional impact evaluations, Project HOPE and RNAfold. We investigated patient and control samples using Tetra ARMS-PCR, then investigated the association between these SNPs and liver cancer.

Preliminary in silico investigations suggested that changes in R364P and R343C might lower protein stability, therefore affecting the function of KLF15. Comparative secondary structures of the mutant and wild-type proteins in RNA exposed clear structural changes. Particularly in those with a positive family history and unique gender-based genotype distributions, genotyping data revealed a strong association between the R364P mutation with an increased risk of liver cancer. Although the CC genotype appeared to offer some protection, those with the GG and CG genotypes clearly had a much higher risk of liver cancer.

The results show that the R364P mutation of KLF15 is a major actor in the evolution of liver cancer. This study underlines the importance of inherited inclination, especially in populations where liver cancer is rather widespread. The thorough analysis of KLF15's anatomy and physiology clarify how these variations in DNA could affect cancer growth, therefore guiding the path towards new opportunities in focused treatment and tailored healthcare.

This dissertation presents a complete analysis of the function of KLF15 polymorphisms in liver cancer by including in silico predictions with validation in the lab. The intensity of the association between the R364P mutation and the incidence of liver cancer emphasizes the potential of KLF15 as a target for therapy and a prognostic marker. Future research on the molecular mechanisms underlining KLF15's participation in cancer could help to evaluate the practical relevance of these findings.

CONCLUSIONS AND FUTURE RECOMMENDATION

In Silico identification of KLF-15 disease related variants led to the selection of a most likely pathogenic SNP. Whereas the selected KLF-15 protein missense mutation Variant ID (*rs755719419)* variation was found to be detrimental and substantially linked with hepatic cancer, the statistics shows that persons with the G allele or the GG genotype are much more likely than those without to get the illness. This indicates that the R364P mutation is one of the causes of disease pathogenesis. Examining the statistical analysis closely demonstrates that, with other factors including gender and family history also influencing disease susceptibility, the several risks associated to distinct genotypes clearly highlight the effect of the mutation, the GG and CG genotypes in particular exhibit a significant correlation with a higher risk of disease. The protective influence of the CC genotype in both sexes provides the complexity of genetic connections. These findings enable more exact genetic testing and tailored medical treatments, therefore advancing our understanding of the hereditary foundation of illness. Small sample size can be the limiting factor; so, it is advisable to see the influence of the mutation sample size by raising it. Moreover, the sample requirements include the population from a specific place, diverse ethnicity throughout the globe displays distinct allele distribution which can be connected to liver cancer advancement. Predicting the structure of the R364P mutant protein, structural and functional studies of that mutation produced lower general stability. In general, the Wild Type protein is less flexible, more compact, and structurally stable than the R364P and R343C variants. Reduced RMSF, consistent RMSD, and more stable hydrogen bonds all point to Wild Type proteins as more stable and whole than mutants. These findings highlight the need of protein stability in performance and the possible functional consequences of the R364P and R343C mutations.

Additionally displaying relationship with hepatic cancer in wet lab experiments is the R364P mutation. Acting as a new possible therapeutic target and a prognostic marker to help in early Hepatic cancer identification, the found SNP could affect the expression of

KLF-15 after this mutation may open new directions in the field of cancer treatment. Moreover, more investigation should be done for better validation of the conclusions on the large sample size. The sample criteria could be expanded, and with any other advanced technique to check the significance of this mutation. In order to close existing gaps, more in vitro and in vivo research is required. The KLF-15 association with various coding and non-coding variants can be examined in other cancer types as well. Comprehensive studies are necessary to offer proof of how mutant KLF-15 influences the interaction path. This knowledge helps one to better grasp the precise mechanisms and potential effects of KLF-15 mutations in several biological processes. Using both animal and cell-based models helps one to grasp the whole extent of these mutations' consequences on cellular performance and organismal health. These kinds of research is essential for advancing our knowledge in this field and offer tailored treatment options.

REFERENCES

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., & Bambrick, J. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 1-3.

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, *76*(1), 7.20. 21-27.20. 41.

Aggarwal, A., Costa, M. J., Rivero-Gutiérrez, B., Ji, L., Morgan, S. L., & Feldman, B. J. (2017). The circadian clock regulates adipogenesis by a Per3 crosstalk pathway to Klf15. *Cell reports*, *21*(9), 2367-2375.

Anjum, S., Hashim, M., Malik, S. A., Khan, M., Lorenzo, J. M., Abbasi, B. H., & Hano, C. (2021). Recent advances in zinc oxide nanoparticles (ZnO NPs) for cancer diagnosis, target drug delivery, and treatment. *Cancers*, *13*(18), 4570.

Anwanwan, D., Singh, S. K., Singh, S., Saikam, V., & Singh, R. (2020). Challenges in liver cancer and possible treatment approaches. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, *1873*(1), 188314.

Anzai, K., Tsuruya, K., Ida, K., Kagawa, T., Inagaki, Y., & Kamiya, A. (2021). Kruppel-like factor 15 induces the development of mature hepatocyte-like cells from hepatoblasts. *Scientific Reports*, *11*(1), 18551.

Blum, M., Chang, H.-Y., Chuguransky, S., Grego, T., Kandasaamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., & Raj, S. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic acids research*, *49*(D1), D344-D354.

Bott, M., Brevet, M., Taylor, B. S., Shimizu, S., Ito, T., Wang, L., Creaney, J., Lake, R. A., Zakowski, M. F., & Reva, B. (2011). The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21. 1 losses in malignant pleural mesothelioma. *Nature genetics*, *43*(7), 668-672.

Brayer, K. J., & Segal, D. J. (2018). Keep your fingers off my DNA: protein–protein interactions mediated by C2H2 zinc finger domains. *Cell biochemistry and biophysics*, *50*, 111-131.

Choi, C.-M., Jang, S.-J., Park, S.-Y., Choi, Y.-B., Jeong, J.-H., Kim, D.-S., Kim, H.-K., Park, K.-S., Nam, B.-H., & Kim, H.-R. (2019). Transglutaminase 2 as an independent prognostic marker for survival of patients with non-adenocarcinoma subtype of non-small cell lung cancer. *Molecular cancer*, *10*, 1-10.

Choi, H.-J., Yeon, M.-H., & Jun, H.-S. (2021). Schisandrae chinensis Fructus Extract Ameliorates Muscle Atrophy in Streptozotocin-Induced Diabetic Mice by

Downregulation of the CREB-KLF15 and Autophagy–Lysosomal Pathways. *Cells*, *10*(9), 2283.

Cox, T. R. (2021). The matrix in cancer. *Nature Reviews Cancer*, *21*(4), 217-238.

Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., Bennett, R., Bhai, J., Billis, K., & Boddu, S. (2019). Ensembl 2019. *Nucleic acids research*, *47*(D1), D745-D751.

Dhar, D., Baglieri, J., Kisseleva, T., & Brenner, D. A. (2020). Mechanisms of liver fibrosis and its role in liver cancer. *Experimental Biology and Medicine*, *245*(2), 96-108.

El-Serag, H. B., Marrero, J. A., Rudolph, L., & Reddy, K. R. (2019). Diagnosis and treatment of hepatocellular carcinoma. *Gastroenterology*, *134*(6), 1752-1763.

Fan, L., Hsieh, P. N., Sweet, D. R., & Jain, M. K. (2018). Krüppel-like factor 15: Regulator of BCAA metabolism and circadian protein rhythmicity. *Pharmacological research*, *130*, 123-126.

Foda, Z. H., Annapragada, A. V., Boyapati, K., Bruhm, D. C., Vulpescu, N. A., Medina, J. E., Mathios, D., Cristiano, S., Niknafs, N., & Luu, H. T. (2023). Detecting liver cancer using cell-free DNA fragmentomes. *Cancer discovery*, *13*(3), 616-631.

Gao, L., Qiu, H., Liu, J., Ma, Y., Feng, J., Qian, L., Zhang, J., Liu, Y., & Bian, T. (2017). KLF15 promotes the proliferation and metastasis of lung adenocarcinoma cells and has potential as a cancer prognostic marker. *Oncotarget*, *8*(66), 109952.

Gray, S., Wang, B., Orihuela, Y., Hong, E.-G., Fisch, S., Haldar, S., Cline, G. W., Kim, J. K., Peroni, O. D., & Kahn, B. B. (2017). Regulation of gluconeogenesis by Krüppel-like factor 15. *Cell metabolism*, *5*(4), 305-312.

Hanahan, D. (2022). Hallmarks of cancer: new dimensions. *Cancer discovery*, *12*(1), 31-46.

Hassanpour, S. H., & Dehghani, M. (2017). Review of cancer from perspective of molecular. *Journal of cancer research and practice*, *4*(4), 127-129.

Hausman, D. M. (2019). What is cancer? *Perspectives in biology and medicine*, *62*(4), 778-784.

Hirata, Y., Nomura, K., Senga, Y., Okada, Y., Kobayashi, K., Okamoto, S., Minokoshi, Y., Imamura, M., Takeda, S. i., & Hosooka, T. (2019). Hyperglycemia induces skeletal muscle atrophy via a WWP1/KLF15 axis. *JCI insight*, *4*(4).

Hirata, Y., Takahashi, M., Morishita, T., Noguchi, T., & Matsuzawa, A. (2017). Post-translational modifications of the TAK1-TAB complex. *International Journal of Molecular Sciences*, *18*(1), 205.

Hsieh, P. N., Fan, L., Sweet, D. R., & Jain, M. K. (2019). The Krüppel-like factors and control of energy homeostasis. *Endocrine Reviews*, *40*(1), 137-152.

Huang, Z., He, H., Qiu, F., & Qian, H. (2022). Expression and prognosis value of the KLF family members in colorectal cancer. *Journal of Oncology*, *2022*.

Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., & Karyadi, D. (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, *99*(4), 877-885.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., & Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589.

Kanyomse, Q., Le, X., Tang, J., Dai, F., Mobet, Y., Chen, C., Cheng, Z., Deng, C., Ning, Y., & Yu, R. (2022). KLF15 suppresses tumor growth and metastasis in Triple-Negative Breast Cancer by downregulating CCL2 and CCL7. *Scientific Reports*, *12*(1), 19026.

Khazaei, M., Hosseini, M. S., Haghighi, A. M., & Misaghi, M. (2023). Nanosensors and their applications in early diagnosis of cancer. *Sensing and Bio-Sensing Research*, 100569.

Lavallée, G., Andelfinger, G., Nadeau, M., Lefebvre, C., Nemer, G., Horb, M. E., & Nemer, M. (2019). The Kruppel-like transcription factor KLF13 is a novel regulator of heart development. *The EMBO journal*, *25*(21), 5201-5213.

Le, N. Q. K., Do, D. T., & Le, Q. A. (2021). A sequence-based prediction of Kruppel-like factors proteins using XGBoost and optimized features. *Gene*, *787*, 145643.

Leenders, J. J., Wijnen, W. J., Van Der Made, I., Hiller, M., Swinnen, M., Vandendriessche, T., Chuah, M., Pinto, Y. M., & Creemers, E. E. (2012). Repression of cardiac hypertrophy by KLF15: underlying mechanisms and therapeutic implications. *PloS one*, *7*(5), e36754.

Li, X., Ramadori, P., Pfister, D., Seehawer, M., Zender, L., & Heikenwalder, M. (2021). The immunological and metabolic landscape in primary and metastatic liver cancer. *Nature Reviews Cancer*, *21*(9), 541-557.

Li, X., Sun, R., & Liu, R. (2019). Natural products in licorice for the therapy of liver diseases: Progress and future opportunities. *Pharmacological Research*, *144*, 210-226.

Ling, X., Wang, Q., Zhang, J., & Zhang, G. (2023). Genome-Wide Analysis of the KLF Gene Family in Chicken: Characterization and Expression Profile. *Animals*, *13*(9), 1429.

Liu, Y., Cheng, T., Du, Y., Hu, X., & Xia, W. (2020). LncRNA LUCAT1/miR-181a-5p axis promotes proliferation and invasion of breast cancer via targeting KLF6 and KLF15. *BMC molecular and cell biology*, *21*(1), 1-11.

Marine, J.-C., Dawson, S.-J., & Dawson, M. A. (2020). Non-genetic mechanisms of therapeutic resistance in cancer. *Nature Reviews Cancer*, *20*(12), 743-756.

Marrero-Rodríguez, D., Taniguchi-Ponciano, K., Jimenez-Vega, F., Romero-Morelos, P., Mendoza-Rodríguez, M., Mantilla, A., Rodriguez-Esquivel, M., Hernandez, D., Hernandez, A., & Gomez-Gutierrez, G. (2014). Krüppel-like factor 5 as potential molecular marker in cervical cancer and the KLF family profile expression. *Tumor Biology*, *35*, 11399-11407.

Matsumoto, N., Kubo, A., Liu, H., Akita, K., Laub, F., Ramirez, F., Keller, G., & Friedman, S. L. (2019). Developmental regulation of yolk sac hematopoiesis by Kruppel-like factor 6. *Blood*, *107*(4), 1357-1365.

McConnell, B. B., & Yang, V. W. (2010). Mammalian Krüppel-like factors in health and diseases. *Physiological reviews*, *90*(4), 1337-1381.

Nault, J. C., & Villanueva, A. (2021). Biomarkers for hepatobiliary cancers. *Hepatology*, *73*, 115-127.

Nisar, Q. A., Haider, S., Ali, F., Gill, S. S., & Waqas, A. (2022). The Role of Green HRM on Environmental Performance of Hotels: Mediating Effect of Green Self-Efficacy & Employee Green Behaviors. *Journal of Quality Assurance in Hospitality & Tourism*, 1-34.

Ogun, O. J., Soremekun, O. S., Thaller, G., & Becker, D. (2023). An In Silico Functional Analysis of Non-Synonymous Single-Nucleotide Polymorphisms of Bovine CMAH Gene and Potential Implication in Pathogenesis. *Pathogens*, *12*(4), 591.

Orzechowska-Licari, E. J., LaComb, J. F., Mojumdar, A., & Bialkowska, A. B. (2022). SP and KLF Transcription Factors in Cancer Metabolism. *International Journal of Molecular Sciences*, *23*(17), 9956.

Patel, S. K., Wai, B., Lang, C. C., Levin, D., Palmer, C. N., Parry, H. M., Velkoska, E., Harrap, S. B., Srivastava, P. M., & Burrell, L. M. (2017). Genetic variation in Kruppel like factor 15 is associated with left ventricular hypertrophy in patients with type 2 diabetes: discovery and replication cohorts. *EBioMedicine*, *18*, 171-178.

Preiss, A., Rosenberg, U. B., Kienlin, A., Seifert, E., & Jäckle, H. (2018). Molecular genetics of Krüppel, a gene required for segmentation of the Drosophila embryo. *Nature*, *313*(5997), 27-32.

Prosdocimo, D. A., Anand, P., Liao, X., Zhu, H., Shelkay, S., Artero-Calderon, P., Zhang, L., Kirsh, J., Rosca, M. G., & Vazquez, E. (2014). Kruppel-like factor 15 is a critical

regulator of cardiac lipid metabolism. *Journal of Biological Chemistry*, *289*(9), 5914-5924.

Rane, M. J., Zhao, Y., & Cai, L. (2019). Krüppel-like factors (KLFs) in renal physiology and disease. *EBioMedicine*, *40*, 743-750.

Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research*, *47*(D1), D886-D894.

Rizvi, S., Wang, J., & El-Khoueiry, A. B. (2021). Liver cancer immunity. *Hepatology*, *73*, 86-103.

Robert, F., & Pelletier, J. (2018). Exploring the impact of single-nucleotide polymorphisms on translation. *Frontiers in genetics*, *9*, 419121.

Sarwar, M. R., & Saqib, A. (2017). Cancer prevalence, incidence and mortality rates in Pakistan in 2012. *Cogent Medicine*, *4*(1), 1288773.

Schottenfeld, D., & Fraumeni Jr, J. F. (2016). *Cancer epidemiology and prevention*. Oxford University Press.

Shi, J. F., Cao, M., Wang, Y., Bai, F. Z., Lei, L., Peng, J., Feletto, E., Canfell, K., Qu, C., & Chen, W. (2021). Is it possible to halve the incidence of liver cancer in China by 2050? *International Journal of Cancer*, *148*(5), 1051-1065.

Siegel, R. L., Miller, K. D., & Jemal, A. (2016). Cancer statistics, 2016 [Article]. *CA Cancer Journal for Clinicians*, *66*(1), 7-30. https://doi.org/10.3322/caac.21332

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA: a cancer journal for clinicians*, *68*(1), 7-30.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research*, *40*(W1), W452-W457.

Society, A. C. (2024). https://www.cancer.org/cancer/types/liver-cancer/about/what-is-key-statistics.html#:~:text=More%20than%20800%2C000%20people%20are,Center%20for%20more%20key%20statistics.

Sun, C., Ma, P., Wang, Y., Liu, W., Chen, Q., Pan, Y., Zhao, C., Qian, Y., Liu, J., & Li, W. (2017). KLF15 inhibits cell proliferation in gastric cancer cells via up-regulating CDKN1A/p21 and CDKN1C/p57 expression. *Digestive Diseases and Sciences*, *62*, 1518-1526.

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, *71*(3), 209-249.

Suzuki, N., Kanai, A., Suzuki, Y., Ogino, H., & Ochi, H. (2022). Adrenergic receptor signaling induced by Klf15, a regulator of regeneration enhancer, promotes kidney reconstruction. *Proceedings of the National Academy of Sciences*, *119*(33), e2204338119.

Tian, L.-L., Zhang, J., Wang, Z.-Z., Chen, S.-C., Zou, X.-B., Yu, Z.-K., & Kang, C.-C. (2020). KLF15 reduces the level of apoptosis in mouse liver induced by sepsis by

inhibiting p38MAPK/ERK1/2 signaling pathway. *European Review for Medical & Pharmacological Sciences*, *24*(20).

Tufail, M., & Wu, C. (2023). Exploring the Burden of Cancer in Pakistan: An Analysis of 2019 Data. *Journal of Epidemiology and Global Health*, 1-11.

Uchida, S., Tanaka, Y., Ito, H., Saitoh-Ohara, F., Inazawa, J., Yokoyama, K. K., Sasaki, S., & Marumo, F. (2019). Transcriptional regulation of the CLC-K1 promoter by myc-associated zinc finger protein and kidney-enriched Krüppel-like factor, a novel zinc finger repressor. *Molecular and cellular biology*, *20*(19), 7319-7331.

Wang, C., Vegna, S., Jin, H., Benedict, B., Lieftink, C., Ramirez, C., de Oliveira, R. L., Morris, B., Gadiot, J., & Wang, W. (2019). Inducing and exploiting vulnerabilities for the treatment of liver cancer. *Nature*, *574*(7777), 268-272.

Wang, G., Pan, J., Zhang, L., Wei, Y., & Wang, C. (2017). Long non-coding RNA CRNDE sponges miR-384 to promote proliferation and metastasis of pancreatic cancer cells through upregulating IRS 1. *Cell proliferation*, *50*(6), e12389.

Wang, G., Wu, B., Cui, Y., Zhang, B., Jiang, C., & Wang, H. (2020). Teneligliptin promotes bile acid synthesis and attenuates lipid accumulation in obese mice by targeting the KLF15-Fgf15 pathway. *Chemical Research in Toxicology*, *33*(8), 2164-2171.

Wang, K., Liu, C., Zhu, H., Zhang, Y., Su, M., Wang, X., Liu, M., Rong, X., & Zhu, B. (2023). Recent advances in small-molecule fluorescent probes for diagnosis of cancer cells/tissues. *Coordination Chemistry Reviews*, *477*, 214946.

Wang, Y., Jiang, F., Xiong, Y., Cheng, X., Qiu, Z., & Song, R. (2020). LncRNA TTN-AS1 sponges miR-376a-3p to promote colorectal cancer progression via upregulating KLF15. *Life sciences*, *244*, 116936.

Wege, H., Schulze, K., von Felden, J., Calderaro, J., & Reig, M. (2021). Rare variants of primary liver cancer: Fibrolamellar, combined, and sarcomatoid hepatocellular carcinomas. *European Journal of Medical Genetics*, *64*(11), 104313.

WHO. (2024).

Wickramarachchi, A., Mallawaarachchi, V., Rajan, V., & Lin, Y. (2020). Metabcc-lr: meta genomics b inning by c overage and c omposition for l ong r eads. *Bioinformatics*, *36*(Supplement_1), i3-i11.

Wu, R., Guo, W., Qiu, X., Wang, S., Sui, C., Lian, Q., Wu, J., Shan, Y., Yang, Z., & Yang, S. (2021). Comprehensive analysis of spatial architecture in primary liver cancer. *Science Advances*, *7*(51), eabg3750.

Yerra, V. G., & Drosatos, K. (2023). Specificity Proteins (SP) and Krüppel-like Factors (KLF) in Liver Physiology and Pathology. *International Journal of Molecular Sciences*, *24*(5), 4682.

Yoda, T., McNamara, K. M., Miki, Y., Onodera, Y., Takagi, K., Nakamura, Y., Ishida, T., Suzuki, T., Ohuchi, N., & Sasano, H. (2015). KLF15 in breast cancer: a novel tumor suppressor? *Cellular Oncology*, *38*, 227-235.

Zhao, X., Chen, L., Wu, J., You, J., Hong, Q., & Ye, F. (2021). Transcription factor KLF15 inhibits the proliferation and migration of gastric cancer cells via regulating the TFAP2A-AS1/NISCH axis. *Biology Direct*, *16*(1), 1-11.

Zhao, X., Monson, C., Gao, C., Gouon-Evans, V., Matsumoto, N., Sadler, K. C., & Friedman, S. L. (2019). Klf6/copeb is required for hepatic outgrowth in zebrafish and for hepatocyte specification in mouse ES cells. *Developmental biology*, *344*(1), 79-93.

Zhao, Y., Song, W., Wang, L., Rane, M. J., Han, F., & Cai, L. (2019). Multiple roles of KLF15 in the heart: Underlying mechanisms and therapeutic implications. *Journal of Molecular and Cellular Cardiology*, *129*, 193-196.

Zheng, G., Dong, X., Wei, J., Liu, Z., Aslam, A., Cui, J., Li, H., Wang, Y., Tian, H., & Cao, X. (2022). Integrated methylome and transcriptome analysis unravel the cold tolerance mechanism in winter rapeseed (Brassica napus L.). *BMC plant biology*, *22*(1), 414.

Zhong, C., Sun, S., Li, Y., Duan, C., & Zhu, Z. (2018). Next-generation sequencing to identify candidate genes and develop diagnostic markers for a novel Phytophthora resistance gene, RpsHC18, in soybean. *Theoretical and applied genetics*, *131*, 525-538.

Zhou, L., Li, Q., Chen, A., Liu, N., Chen, N., Chen, X., Zhu, L., Xia, B., Gong, Y., & Chen, X. (2019). KLF15-activating Twist2 ameliorated hepatic steatosis by inhibiting inflammation and improving mitochondrial dysfunction via NF-κB-FGF21 or SREBP1c-FGF21 pathway. *The FASEB Journal*, *33*(12), 14254-14269.

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: Humaira Bibi
Assignment title: Quick Submit
Submission title: Mechanistic Impact of KLF15 Polymorphism in Liver Cancer
File name: Humaira_thesis_for_plagrism.docx
File size: 117.34K
Page count: 48
Word count: 14,372
Character count: 82,534
Submission date: 04-Jul-2024 11:35PM (UTC-0700)
Submission ID: 2412730153

Mechanistic Impact of KLF15 Polymorphism in Liver Cancer

By

Humaira Bibi

00000399991

Master of Science in Molecular Medicine

Supervisor

Dr. Yasmin Badshah

Co-Supervisor

Dr. Maria Shabbir

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences & Technology (NUST)

Islamabad, Pakistan

7   David Anwanwan, Santosh Kumar Singh, Shriti Singh, Varma Saikam, Rajesh Singh. "Challenges in liver cancer and possible treatment approaches", Biochimica et Biophysica Acta (BBA) - Reviews on Cancer, 2020
    Publication   <1%

8   Yanhua Wang, Fang Jiang, Yan Xiong, Xiaoliang Cheng, Zhimin Qiu, Rongfeng Song. "LncRNA TTN-AS1 sponges miR-376a-3p to promote colorectal cancer progression via upregulating KLF15", Life Sciences, 2020
    Publication   <1%

9   Lulu Zhou, Qinjin Li, Ao Chen, Na Liu, Ning Chen, Xiaojun Chen, Lin Zhu, Benzeng Xia, Yuqing Gong, Xiaodong Chen. " KLF15-activating Twist2 ameliorated hepatic steatosis by inhibiting inflammation and improving mitochondrial dysfunction NF-κB-FGF21 or SREBP1c-FGF21 pathway ", The FASEB Journal, 2019
    Publication   <1%

10 www.ncbi.nlm.nih.gov
    Internet Source   <1%

11 Douglas Hanahan. "Hallmarks of Cancer: New Dimensions", Cancer Discovery, 2022
    Publication   <1%

12 Masoud Khazaei, Marzieh Sadat Hosseini, Ali Moshfegh Haghighi, Majid Misaghi. "Nanosensors and their applications in early diagnosis of cancer", Sensing and Bio-Sensing Research, 2023
Publication

<1%

13 Nanoka Suzuki, Akinori Kanai, Yutaka Suzuki, Hajime Ogino, Haruki Ochi. "Adrenergic receptor signaling induced by Klf15, a regulator of regeneration enhancer, promotes kidney reconstruction", Proceedings of the National Academy of Sciences, 2022
Publication

<1%

14 www.ijnrd.org
Internet Source

<1%

15 Submitted to Al Quds University
Student Paper

<1%

16 Fizzah Abid, Talha Iqbal, Khushbukhat Khan, Yasmin Badshah et al. "Analyzing PKC Gamma (+ 19,506 A/G) polymorphism as a promising genetic marker for HCV-induced hepatocellular carcinoma", Biomarker Research, 2022
Publication

<1%

17 Geissler, Catherine, Powers, Hilary. "Human Nutrition", Human Nutrition, 2023
Publication

<1%

18 Guang Wang, Bing Wu, Yang Cui, Bo Zhang, Chunyan Jiang, Heyuan Wang. "Teneligliptin Promotes Bile Acid Synthesis and Attenuates Lipid Accumulation in Obese Mice by Targeting the KLF15-Fgf15 Pathway", Chemical Research in Toxicology, 2020
Publication

<1%

19 Neil Hoa, Lisheng Ge, Kenneth S. Korach, Ellis R. Levin. "Estrogen receptor beta maintains expression of KLF15 to prevent cardiac myocyte hypertrophy in female rodents", Molecular and Cellular Endocrinology, 2017
Publication

<1%

20 Submitted to Rwanda Defence Forces Command and Staff College
Student Paper

<1%

21 Josh Abramson, Jonas Adler, Jack Dunger, Richard Evans et al. "Accurate structure prediction of biomolecular interactions with AlphaFold 3", Nature, 2024
Publication

<1%

22 archive.org
Internet Source

<

23 Khushbukhat Khan, Hania Shah, Areeba Rehman, Yasmin Badshah, Naeem M Ashraf, Maria Shabbir. "Influence of PRKCE non-synonymous variants on protein dynamics