Influence of Deleterious KPCQ Variants on Human Breast Cancer

Andleeb Farooq

00000361674

Supervisor

Dr. Yasmin Badshah

Co-Supervisor

Dr. Maria Shabbir

Department of Healthcare Biotechnology

Atta-ur-Rehman School of Applied Biosciences, National University of Sciences and Technology Islamabad, Pakistan

2023

Influence of Deleterious KPCQ Variants on Human Breast Cancer

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

Andleeb Farooq

00000361674

Department of Healthcare Biotechnology

Atta-ur-Rehman School of Applied Biosciences, National University of Sciences and Technology Islamabad, Pakistan

2023

FORM TH-4

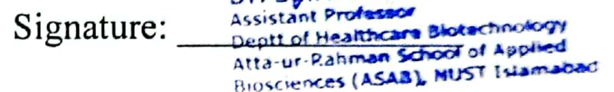
National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: **Andleeb Farooq** Reg No. **361674** Titled: **Influence of Deleterious KPCQ Variants on Human Breast Cancer** be accepted in partial fulfillment of the requirements for the award of **MS degree in Healthcare Biotechnology** degree with (___A___ grade).

Examination Committee Members

1. Name: **Peter John** Signature:

2. Name: **Dr. Saria Justin** Signature: 

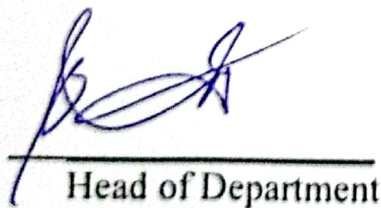
3. Name: **Dr. Erum Dilshad** Signature:

Co- Supervisor's name: **Dr. Maria Shabbir** Signature:

Date: 13/6.23

Supervisor's name: **Yasmin Badshah** Signature:

Date: 13/6/23



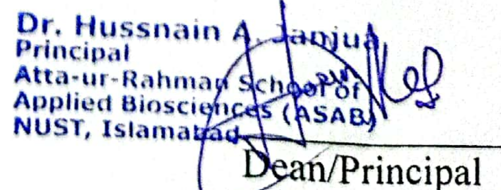
Head of Department

Dr. Sobia Manzoor

Date:: 22.06.2023

COUNTERSINGED

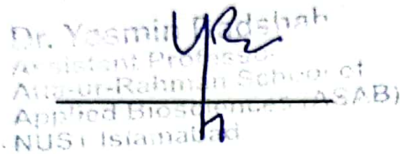
Date: 23.06.2023



Dean/Principal

THESIS ACCEPTANCE CERTIFICATE

It is certified that the final contents of the MS/MPhil thesis entitled **"Influence of Deleterious KPCQ Variants on Human Breast Cancer."** submitted by **Andleeb Farooq, (Registration number 00000361674)**, of ASAB, has been verified by the undersigned, observed total in all regards according to NUST Status/Regulations, is free of plagiarism, and is acknowledged as partial fulfillment for grant of MS/MPhil degree. It is additionally ensured that fundamental revisions as called attention to by GEC individuals from the researcher have likewise been joined in the said proposal.

Signature: 

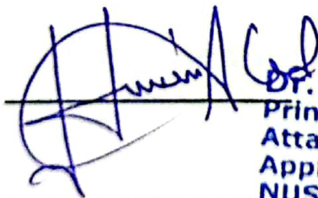
Supervisor: **Dr. Yasmin Badshah**

Date: 15/6/2023

Signature (HOD):

Dr. Sobia Manzoor

Date: 22.06.2023

Signature (Dean/Principal): 

Dr. Hussnain A. Janjua
Principal
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST, Islamabad

Date: 23.06.2023

CERTIFICATE FOR PLAGIARISM

It is certified that the MS thesis entitled "**Influence of Deleterious KPCQ Variants on Human Breast Cancer**" submitted by Andleeb Farooq has been examined.

I undertake the following:

- The thesis has significant new work/knowledge as compared to already published, or under consideration to be published work. No sentence, equation, diagram, table, paragraph, or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
- The work presented is original and the author's own work, i.e., there is no plagiarism. No ideas, results, or work of others have been presented as the author's own work.
- There is no fabrication of data or results, which have been compiled/ analyzed.
- There is no falsification by manipulating research materials, equipment/processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
- The thesis has been checked by using Turnitin (a copy of the originality report is attached) and found within the limits as per HEC plagiarism policy and instructions issued from time to time.

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

Supervisor

Dr. Yasmin Badshah

DECLARATION

I certify that my research work, "**Influence of Deleterious KPCQ Variants on Human Breast Cancer**" is my own work. The work has not been presented elsewhere for assessment. The materials that have been used from other sources have been properly acknowledged/referred.

Signature of Student

Andleeb Farooq

Dedicated to my Father (late), my Mother, Siblings, Family, Teachers, and Friends

Acknowledgments

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

I would like to express my sincere gratitude to my supervisor Dr. Yasmin Badshah and co-supervisor Dr. Maria Shabbir for their guidance and motivation that made this project possible. I would like to extend my regards to my senior Khushbukhat Khan for her support and expertise in the project.

I would also like to thank my seniors Hania Shah, and Fizzah Abid colleagues Tooba Khan, Umama Ayub, Mahnoor Basit, and Muhammad Armaghan, and friends Nazish Majeed, Nawal Naveed Abbasi, Amna Fayyaz, Somia Khan, Saman Noor, Tazeen and Roshaan Ijaz for their constant support, encouragement, and help throughout the project.

I would like to extend my appreciation to my late father, although he is no longer with us, his presence continues to be felt in every step I take. His wisdom, integrity, and unconditional love has been a source of inspiration for me. His life's lessons and the values he instilled in me are guiding lights that I carry in my heart, propelling me forward.

Lastly, a special thanks to my mother, my siblings, and all family members as this journey would not have been possible without their love and support.

Thank you.

Andleeb Farooq

TABLE OF CONTENTS

ABSTRACT..1

CHAPTER 1: INTRODUCTION..2

1.1 Cancer..2

1.1.1 Carcinoma..3

1.1.2 Sarcoma..3

1.1.3 Myeloma...3

1.1.4 Leukemia...3

1.1.5 Lymphoma...4

1.1.6 Mixed Types..4

1.2 Breast cancer...4

1.3 Diagnostic methods of breast cancer..5

1.4 Treatment Strategy for breast cancer...5

1.5 Problems in the diagnostic and treatment of breast cancer..............................6

1.6 Biomarkers..7

1.7 KPCQ as a potential molecular target...8

1.8 Single nucleotide polymorphism...8

1.9 Interaction between KPCQ and KPCδ..9

1.10 Computational and experimental approaches..10

1.11 Study objectives...10

CHAPTER 2: LITERATURE REVIEW...11

2.1 Overview of cancer...11

2.2 KPC family..13

2.3 KPCQ...14

2.3.1 Discovery of KPCQ...15

2.3.2 Structure of KPCQ..15

2.3.3 Localization of KPCQ......15
2.4 Role of KPCQ in cell......16
2.4.1 Role of KPCQ in cell cycle 17
2.5 Effect of overexpression of KPCQ......18
2.6 Effect of loss of KPCQ......19
2.7 KPCQ and cancer 19
2.7.1 Leukemia 19
2.7.2 Breast cancer 21
2.7.3 Role of KPCQ in different cancers......23
2.8 Molecular targets and interactions of KPC-θ 24
2.8.1 KPCQ role in the activation of AP1 24
2.8.2 Role of KPCQ in the activation of NAFT 24
2.9 Potential KPCQ substrates 25
CHAPTER 3: MATERIAL AND METHODS 27
3.1 In-silico analysis......27
3.1.1 Structural prediction of KPCQ......27
3.1.2 In silico SNP effect analysis......27
3.1.3 Evaluation of deleterious nsSNPs......28
3.1.4 Stability analysis......28
3.1.5 Structural and functional effect of SNP......28
3.1.6 Prediction of disease association......28
3.1.7 Analysis of filtered variant's flexibility......29
3.1.8 In-situ mutagenesis 29
3.1.9 Molecular dynamic simulations......29
3.1.10 Primer designing......30
3.2 Experimental work 30
3.2.1 Sample collection......30

3.2.2 DNA extraction.......31
3.2.3 TETRA ARMS PCR.......31
3.2.3.1 PCR steps and conditions.......32
3.2.4 Gel electrophoresis.......32
3.2.4.1 Agarose gel preparation.......32
3.2.5 Statistical analysis.......33
CHAPTER 4: RESULTS.......34
4.1 Structural prediction of KPCQ.......34
4.2 Prediction of deleterious nsSNPs.......34
4.3 KPCQ DOMAIN.......38
4.4 Mutation land mapping.......39
4.5 Effect of nsSNP on structure and function.......41
4.6 Association of pathogenic SNPs with disease.......42
4.7 Impact of SNP on KPCQ stability.......43
4.8 Flexibility and stability analysis.......44
4.9 In-situ Mutagenesis.......47
4.10 Molecular dynamic simulation.......48
4.10.1 Radius of gyration.......48
4.10.2 Number of hydrogen bonds.......48
4.10.3 Root mean square deviation.......50
4.10.3 Root mean square fluctuation.......50
4.10.4 SASA analysis.......51
4.11 Genotype frequency distribution and association of KPCQ with breast cancer.......52
CHAPTER 5: DISCUSSION.......57
CONCLUSION AND FUTURE PRESPECTIVE.......61
REFRENCES.......63

LIST OF TABLES

Table 1. The most common cases of cancer reported in 2020 are enlisted. 2

Table 2. In 2020, the most common causes of cancer deaths 3

Table 3. Primers for KPCQ 30

Table 4. Details of 9 Variants including IDs, genomic and amino acid coordinated. 35

Table 5. Scores of KPCQ gene nine deleterious variants were obtained through multiple consensus tools. 36

Table 6 KPCQ nsSNPs properties, location, and effect on its structure predicted through PROJECT HOPE...41

Table 7 Estimated association of KPCQ SNPs with cancer FATHMM. 42

Table 8 Influence of SNP stability analysis on variant pathogenicity through I-Variant. 43

Table 9. Frequency distribution and genetic association of KPCQ variant with breast cancer...53

Table 10. Genotype frequency distribution and association of KPCQ variant with untreated and treated state of breast cancer 54

Table 11. Frequency distribution and genetic association of KPCQ variant with non metastatic and metastatic breast cancer...54

Table 12. Genotype frequency distribution and association of KPCQ variant with a non-family history and family history that is linked with breast cancer. 55

LIST OF FIGURES

Figure1. Fears and barriers in breast cancer diagnosis and treatment 6
Figure 2. Incidence of major cancers in Pakistan 12
Figure 3. The Hallmarks of Cancer 13
Figure 4. Pathogenicity percentage graph of KPCQ missense variants. 38
Figure 5. Map showing the domains and regions of KPCQ. Six SNPs (C29Y, P27S, P27L, R6L, R6Q, D13Y) present in C2-like domain in regulatory region. Variants I550N, L495P, and V394F localized in the protein kinase domain. 39
Figure 6(a). Mutational landscape of missense variants spread across 706 amino acid residues of KPCQ...40
Figure 7. SNP stability analysis of KPCQ variants through I-Variant v2.0. Except for D13Y, all other variants were estimated as destabilizing mutations. Estimation was based on Delta G value where negative Delta G values were regarded as destabilizing and positive Delta G values were taken as stabilizing......................44
Figure 8. Flexibility analysis of wild and variant structure of KPCQ through DynaMut. 45
Figure 9. Comparison of interatomic interatomic interactions of wild and variant structure of KPCQ variants. 46
Figure 10. In-situ mutagenesis of KPCQ variants. 47
Figure 11. This graph shows the compactness of both variant and wild protein in a dynamic setting over time. 49
Figure 12. This graph represents the difference in the number of hydrogen bonds between variant and wild-type proteins structure. 49
Figure 13. This graph shows the significant difference and deviation in wild type and variant residues over the time. 50
Figure 14. Graphical representation of RMSF values of both wild type and variant protein. 51
Figure 15. Graphical representation of SASA values of both wild type and variant protein. 52
Figure 16. Genotyping analysis of KPCQ variants with breast cancer.................53
Figure 17. Genotyping analysis of KPCQ variants of both untreated and treated states in relation to breast cancer. 54

Figure 18. Genotyping analysis of KPCQ variants of both metastatic and non-metastatic breast cancer. 55
Figure 19. Genotype frequency distribution and association of KPCQ variant with a family history that is linked with breast cancer. 56

LIST OF ACRONYMS

ARMS	Amplification-Refractory Mutation System
MRI	Magnetic Resonance Image
SNP	Single Nucleotide Polymorphism
DAG	Diacylglycerol
EMT	Epithelial–Mesenchymal Transition
IS	Immunological Synapse
GIST	Gastrointestinal Stromal Tumor Cells
ALL	Acute Lymphocytic Leukemia
PCR	Polymerase Chain Reaction
FASTA	Fast-All
HOPE	Have Our Protein Explained
MD	Molecular Dynamics
CADD	Combined Annotation Dependent Depletion
CI	Confidence Interval
OR	Odd Ratio
RR	Relative Risk
TBK1	TANK-binding kinase 1
TBKBP1	TBK-binding protein 1
AP1	Activator Protein 1
NFAT	Nuclear Factor of Activated T- Cells

// ABSTRACT

The cancer that is commonly diagnosed all over the world is breast cancer, with an estimated 2.26 million cases recorded in 2020. It is the main cause of death in women. In past eras, the role of different genes including TP53, PTEN, STK11, CDH1, ATM, PALB2, CHEK2, and BRCA1 and BRCA2 in breast cases have been studied and identified, but the result was not effective for treatment and diagnosis at an early stage of breast cancer. Therefore, there is a need for another novel gene for early prognosis and better treatment for breast cancer. For this purpose, novel KPC family isoforms are preferable as a potential molecular target, which lower the incidence rate and recurrence of breast cancer. The KPCQ is a member of novel KPCs and belongs to serine /threonine kinase. KPCQ plays an important role in the pathophysiology of numerous cancers, particularly autoimmune diseases, and various cancers. The determination of the impact of missense SNPs on the structure and function of the KPCQ gene and the identification of a novel KPCQ missense variant's association with Breast cancer and its clinical features is the main purpose of the study. In this study, KPCQ variant C29Y was selected, and this variant was further analyzed to validate their effect on function, structure, and stability of protein through experimental and in vitro analysis and these predict that protein structure alters and then leads to the progression of breast cancer. Further, genotyping analysis was performed to confirm the association of this variant with breast cancer and this analysis confirmed that the C29Y variant of KPCQ was involved in the progression of breast cancer. The molecular dynamic simulation was performed which proved that this variant of KPCQ is linked with the progression of breast cancer. As a result, this study laid the foundations for considering the variant rs1248923790 as a predictive prognostic biomarker for breast cancer. It could be used at the clinical level after more validation.

Keywords: Kinase protein C theta (KPCQ); Breast cancer; auto-immune disorders; molecular dynamics simulation.

CHAPTER 1: INTRODUCTION

1.1 Cancer

Cancer can be defined as a group of diseases that are caused by uncontrolled and irregular growth of cells that are triggered by various changes in the expression of a gene which are responsible for cell proliferation and cell death and eventually lead to the population of cells that can infect tissue and spread to different locations, which further result in causing the massive illness and if it is not properly diagnosed and treated at an early stage then it causes the death of an individual (Ruddon, 2007).

Many factors are responsible for the cause of cancer. These factors are tobacco, radiation, chemicals, infectious organisms, hormones, random mutation, and inherited mutations. These factors that are responsible for the cause of cancers are very complex and diverse. Many things are known to enhance the risk of cancer, including infections, pathogens, dietary factors, lack of physical activity, and environmental pollutants (Anand et al., 2008a). Thus, all these factors may act collectively in the body and stimulate carcinogenesis which is directly responsible for the destruction of the DNA and thus leading to cause cancer which is the major cause of death.

Cancer is basically explained by the changes that occur in the expression of various genes which are responsible for causing abnormalities in the normal cellular process of the cell cycle and then enhancing cell death. The cells that are responsible for causing cancer easily attack nearby cells and then further spread into lymph nodes and then proliferate into various organs in the body (Ruddon, 2007).Cancer is a major cause of death globally, which reports about 10 million deaths in 2020 (Piñeros et al., 2021).

Table 1. The most common cases of cancer reported in 2020 are enlisted.

S.no.	Cancer	Cases
1	Breast	2.26 million
2	Lung	2.21 million
3	Colon and rectum	1.93 million
4	Skin	1.20 million
5	Stomach	1.09 million

Table 2. In 2020, the most common causes of cancer deaths.

S.no.	Cancer	Deaths
1	Lung	1.80 million
2	Colon and rectum	916 000
3	Liver	830 000
4	Stomach	769 000
5	Breast	685 000

From a histological point, there are multiple cancers, which are divided into six major classes (NIH).

1.1.1 Carcinoma

It is the cancer of epithelium or cancer of the internal or external lining of the body. It accounts for about 80 to 90 percent of all cancer cases.

1.1.2 Sarcoma

This is a class of cancer that develops from the tissue that protects and provide support to the other organs and tissues such as tendons, muscles, cartilage, and bones of the body. It generally occurs in adults, but the most common sarcoma often develops as a painful mass on the bone.

1.1.3 Myeloma

It basically develops in the plasma cells of bone marrow. These plasma cells produce some of the proteins which are present in the blood.

1.1.4 Leukaemia

Leukaemia's also known as liquid cancer or blood cancer. It is a cancer of the bone marrow (it is a site of blood cell production). The word leukaemia is derived from a Greek word that means "white blood". This cancer is often associated with the production of many immature leukocytes.

1.1.5 Lymphoma

Lymphomas originate in the glands or nodes of the lymphatic system which is a network of vessels, nodes, and organs (particularly the spleen, tonsils, and thymus) that clear bodily fluids and produce lymphocytes that fight against the disease.

1.1.6 Mixed Types

This class of cancer may be included within one category or from different categories.

1.2 Breast cancer

It is the second greatest cause of death and most common form of malignancy in women after lung cancer, with approximately one in eight women are at risk of suffering from this cancer in their lifetime. Almost 210,000 women were detected and confirmed with BC in 2010 (Downs-Holmes & Silverman, 2011). Due to high incidence and mortality rates in recent years, breast cancer become the most common cause of death in women globally. According to the IARC, namely the GLOBOCAN program, there were 2.1 million reported cases of breast cancer in 2018 as compared to 2008, when cases of breast cancer were 12.7 million ("International agency for research on cancer," 2019), which estimated that in women one in four cases of cancer were reported (Bray et al., 2015).

The incidence rates of breast cancer are increasing which may be explained based on the aging and growth of the world population as well as the acquisition of lifestyle that enhanced the cancer risk. The occurrence of molecular approaches like the technique of RNA sequencing, has made it possible to establish the profiles of gene expression, molecular classification of breast cancer and identify the heterogeneity of tumor. Based upon the expression of PR, ER, estrogen-related genes such as FOXA1, ESR1, and GATA3 , and genes that are associated with the start of the process of proliferation like HER2 and many other genes that are found in the region of the HER2 amplicon on the chromosome number 17 have been used for the identification of six subtypes of breast cancer (Bray et al., 2015).

It is a multifactorial disease and numerous factors have been identified that significantly contribute to the etiology and promote the risk of growth of breast cancer Some factors are endogenous, like inherited mutations, hormones, and immunity conditions and some are exogenous, such as tobacco, radiations, and chemicals (heavy metals, pesticides, and polycyclic aromatic hydrocarbons) (Mena et al., 2009)

1.3 Diagnostic methods of breast cancer

It is a complex, heterogeneous disease which responsible for causing a large number of death of women all over the world. Some medical diagnostic techniques are developed such as mammography, ultrasonography, positron emission tomography, and magnetic resonance image (MRI) (Thamaraiselvan et al., 2014), and all these techniques are used for the detection of small cancers, palpable masses where a cyst is in the differential diagnosis (Jackson, 1990) and then further to analyse the solid breast masses(Schor & Schor, 1983), breast carcinoma and ductal carcinoma (Boetes et al., 1994).

In addition to these above techniques, there are some additions that are available for diagnosis of breast cancer which includes, needle aspiration in which the biopsies of the mass are analysed for cyst, while in needle biopsy the cells are directly eliminated from the mass for evaluation and in a surgical biopsy, a part of the mass is removed for further examination (Warner et al., 2004).

1.4 Treatment Strategy for breast cancer

For the treatment of breast cancer, different strategies are used and are analysed by several factors and are mostly based on the stages of the tumor, type, and characteristics. Mostly used treatment methods for breast cancer are surgery, radiation therapy, hormonal therapy, biological therapy, chemotherapy, and chemoprevention (Thamaraiselvan et al., 2014).

1.5 Problems in the diagnostic and treatment of breast cancer

It is a multifactorial disease that causes major health issues at the national and international levels. To address the main problems of the detection of cancer at primary stage, several factors need to be observed during the treatment and diagnosis of breast cancer which includes the misconception about cancer, the growth rate of the tumor, the age at which cancer detect, different strategies about health breast, cancer understanding and awareness in women. Some methods which include chemotherapy, radiotherapy, and surgery are available options for the treatment of this cancer. For the detection of breast cancer, biopsy, and mammography are accessible methods in Pakistan. for detection. Problems can arise when there is a delay in the procedure of diagnosis. Treatment failure and delays in the treatment process of breast cancer co-occurrences of some reasons which include surgical, consultation, and biopsy (Ades et al., 2017).

There are some progress has been made in the drug development and treatment and this cancer (Zhuang et al., 2020), but some studies have proved that there is no better option for the treatment of breast cancer, and still, cases and deaths rates of this cancer are increased and not controlled effectively due to lack of cancer awareness, expensive treatment techniques , poor understanding about the growth and development of cancer progression (Wapnir et al., 2018).

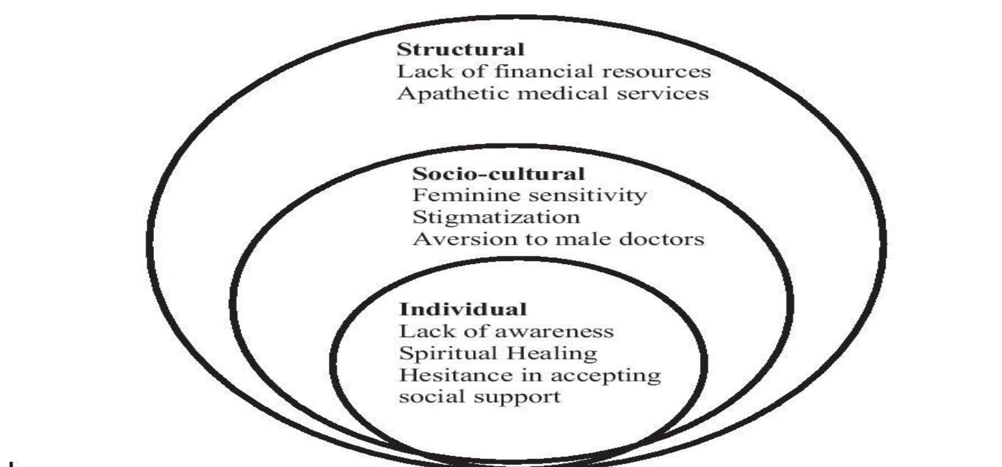


Figure1. Fears and barriers in breast cancer diagnosis and treatment (Saeed et al., 2021).

1.6 Biomarkers

Throughout the whole process of treatment, the most important biomarkers for effective management and early detection of cancer are those that aid in the prognosis and diagnosis of breast cancer (Hayes et al., 2001; Weigel & Dowsett, 2010). They can be used to evaluate the biological state of a disease, which helps in the identification of cancer type, its progression rate, and responses to medication which help in reducing the breast cancer rate (Hsing-Ju Wu & Pei-Yi Chu, 2021; Hsing-Ju Wu & Pei-Yi Chu, 2021). A combination of biomarkers is preferred for the early detection of cancer because of the heterogeneity of cancer cells and a single biomarker is insufficient for the timely detection of the growth and development rate of cancer. A large number of biomarkers have been discovered in blood and tissue because of prominent advancements in the molecular methods of signaling and genetic fingerprinting. These biomarkers may be used for the identification and detection of a chance of spread of cancer, prediction, therapy tolerance and re-emergence. The selectivity and sensitivity of some of the biomarkers, that have been used at clinical trials are zero (Nalejska et al., 2014; Hsing-Ju Wu & Pei-Yi Chu, 2021). For this purpose, there is a need for new and improved biomarkers.

The struggle to create, develop, and identify the optimized biomarkers that assist in improving the prognosis and therapeutic results of breast cancer was set up by disease's early detection. Breast cancer associated biomarkers consist of macromolecules which include protein, whole cell, and nucleic acids. The development in molecular methods has precisely recognized all categories of biomarkers which are specifically investigated for implication of prognostic, diagnostic, therapeutic, and drug resistance. Drug resistance issue, which is difficult and major issue in treatment of breast cancer may be resolved with the help of these biomarkers. The progression of breast cancer is developed by abnormal expression of non-coding RNAs including micro-RNA and circular RNA. As compared to individual miRNA, the circulating miRNA profiles provide improved diagnostic and prognostic sensitivity and performance. The circular RNA that are present in bodily fluid, are more important and potential novel diagnostic biomarkers (Afzal et al., 2022).

In the past era, role of different genes in breast cancer have been studied and identified, but the result was not effective for effective diagnosis and better treatment of breast cancer and cases of breast cancer increased every year. These genes were PTEN,ATM,CHEK2,CDH1,PALB2, BRCA1 and BRCA2 and TP53 (Turnbull & Rahman, 2008) .

So, there is a need for another novel gene for early prognosis and better treatment for breast cancer. For this purpose, KPC family isoforms are preferable as a potential molecular target and lowered the incidence rate and recurrence of breast cancer among patients and provide effective results and better treatment options (Khan et al., 2020).

1.7 KPCQ as a potential molecular target

The KPCQ is a member of novel KPCs and it belongs to serine /threonine kinase and is used as a potential molecular target for better treatment of breast cancer so that the incidence rate of this complex and multifactorial disease is controlled. Different cellular function which include survival of cell, differentiation, and progression are controlled by KPCQ (Nicolle et al., 2021). It generally maps to a region of chromosome 10p15, a region which is generally variant in T cell immunodeficiencies cells leukemia, and lymphomas (Zhang et al., 2013). KPCQ involves in the outbreak of pathophysiology of many disorders, particularly autoimmune diseases, and it also involve in several cancers such as Leukemia, Gastrointestinal Stromal Tumor, Breast Cancers, lungs cancer, renal cancer cells, and ovarian cancer are also reported (Nicolle et al., 2021). The level of KPCQ increased particularly in ER-negative human breast cancers (Belguise & Sonenshein, 2007). Furthermore, KPCQ’s 3-dimensional structure has not yet predicted. the three-dimensional structure of KPCQ is not predicted yet. In this study, identify novel KPCQ variants and predict the three-dimensional structure of KPCQ and analyse the influence of deleterious KPCQ variants on human breast cancer.

1.8 Single nucleotide polymorphism

In human genome, the most common variation of sequence is SNP, which result in cause by only one base substitution. SNP has a lower allelic frequency which is > 1%

in at least one population (Risch, 2000). The following factors make SNPs the most appropriate biomarkers for diagnosis and prognosis of the disease.

1) Common frequency
2) Ease of analysis
3) Economical genotyping costs.
4) Probability to carry out association studies based on statistical and bioinformatics tools.

Consequently, SNPs are considered significant as vital drivers in disease related studies (Srinivasan et al., 2016). If these genetic variants exist in protein-coding genetic sequences, they have an extreme propensity to have damaging effect on normal 3-dimensional structure of protein which can cause multiple fatal diseases mainly cancer (Deng et al., 2017; Gao et al., 2009). The most prevalent kind of genetic variation in human genome is single nucleotide polymorphism (SNP). Genes contain SNPs that control DNA mismatch repair, cell cycle adjustment, metabolism and immunity are connected with a genetic vulnerability to cancer (Deng et al., 2017).

1.9 Interaction between KPCQ and KPCδ

KPC – theta and KPC -delta both belongs to the novel KPC family and can be cleaved with the help of caspase-3 which results in the production of a constitutively activated kinase, which, when introduced into cells, can activate apoptosis (Datta et al., 1997; Schultz et al., 2003). Based on this observation these kinases are often classified as "pro-apoptotic". However, there is only a little information about how KPC-theta plays a significant role in apoptosis. Instead of this, KPC-theta appears to be critical for immune cell function and T cell survival, and severe problems for signalling and activation of T -cells are caused because of KPC-theta knockout mice. Although, the function of KPC-delta in apoptotic cells has been well established (Reyland, 2009).

KPCδ and KPCQ are two main novels KPC isotypes that involve in human autoimmunity and inflammation diseases (Mochly-Rosen et al., 2012) KPCδ and KPCQ both serve as epigenetic enzymes in T cells (Lim et al., 2015).

1.10 Computational and experimental approaches

Computational approaches are easy to use, less laborious, have no need for wet lab resources, can greatly facilitate and help in designing, implementing, evaluating, and analysing experiments, and generally help to lift research to a new and better level. This approach basically integrates the huge amount of data that is generated by experimental researchers. Advanced bioinformatics tools are used for screening potential biomarkers that can be used for breast cancer. Experimental approaches are used in research because it provides the researchers with a high level of control. Experimental research is a simple and basic form of research. This helps to improve the sustainability of a concept for products, ideas, and theories. This allows anyone to be able to check and confirm published results, which often allows for better results to be attained because the exact steps can produce the exact results. This research is mostly used in the medical and pharmaceutical industries to evaluate the impact of various treatments and drugs.

1.11 Study objectives

The determination of the impact of missense SNPs on the structure and function of the KPCQ gene and the identification of the novel KPCQ missense variant's association with breast cancer and its clinical features are the main aims and objectives of this present study.

CHAPTER 2: LITERATURE REVIEW

2.1 Overview of cancer

Cancer can be defined as a group of diseases that are caused by uncontrolled division, growth, and the spread of abnormal cells. The stage at which cancer cells are spread and not controlled is known as metastasis and it can result in death. Many external factors which are responsible for the cause of cancer are tobacco, radiation, chemicals, and infectious organisms, whereas there are also some internal factors which include hormones, random mutation, and inherited mutations that are responsible for the cause of cancer. These factors that cause cancers are very complex and diverse. There are many things that are known to enhance the risk of cancer which include certain infections, pathogens, dietary factors, lack of physical activity, and environmental pollutants (Anand et al., 2008b). Thus, all these factors may act collectively in the body and to stimulate carcinogenesis which is directly responsible for damaging the DNA and thus leading to cause cancer which is the main leading cause of death.

Cancer is basically explained by the changes that occur in the expression of various genes which are responsible for causing abnormalities in the normal cellular process of the cell cycle and then enhancing cell death. The cells that are responsible for causing cancer easily to invade nearby cells and then further spread into lymph nodes and then to metastasize into various organs in the body (Ruddon,2007).

The primary cause of morbidity and mortality all around the whole world is cancer. There is an estimation that 14.1 million new cases and approximately 8.2 million deaths occurred in the year 2012. The number of incidences that occurred because of cancer is increasing day by day and is expected to increase by about 70%, over the coming two decades. There is a huge range of variability across the entire world of cancer incidences. Cancer incidence rates in middle-income countries are very high, and those countries that are less developed have lower rates on average are revealed by some studies.

According to world statistics of 2012, the more common cancers in men were lung cancer, prostate cancer, colorectal cancer, stomach cancer, and liver cancer whereas the most common cancers that are in women were cancer of lung, breast cancer, colon cancer, stomach cancer, and cervical cancer. Those countries which have a high

income, and the cancers that were common in men in such countries were prostate, lung, colorectal, and bladder cancer whereas breast, colorectal, lung, and endometrium cancer were most common in women, and in low-income countries, most common cancers in men were prostate, liver, oesophagus, and lung cancers while breast, cervical, ovarian, and oesophageal cancer among women. Thus, different incidence rates and differential patterns of cancer occur in different countries are different which manifests that different populations are exposed to diverse factors distributed unevenly (Whiteman & Wilson, 2016).

Cancer classification is very complex because of a huge variety of cancer develop in different tissue and organs of the body and incidence rate of cancer also increased. Depend upon its primary site of origin or based on types of tissue, cancer are classified (Guruvayoorappan et al., 2015).

There are six major types of cancer, based on their ontology. These are sarcoma, carcinoma, lymphoma, myeloma, leukaemia, and mixed type (which include different carcinomas that are adenosquamous carcinoma, mixed mesodermatumor, carcinosarcoma, and teratocarcinoma) (Guruvayoorappan et al., 2015).

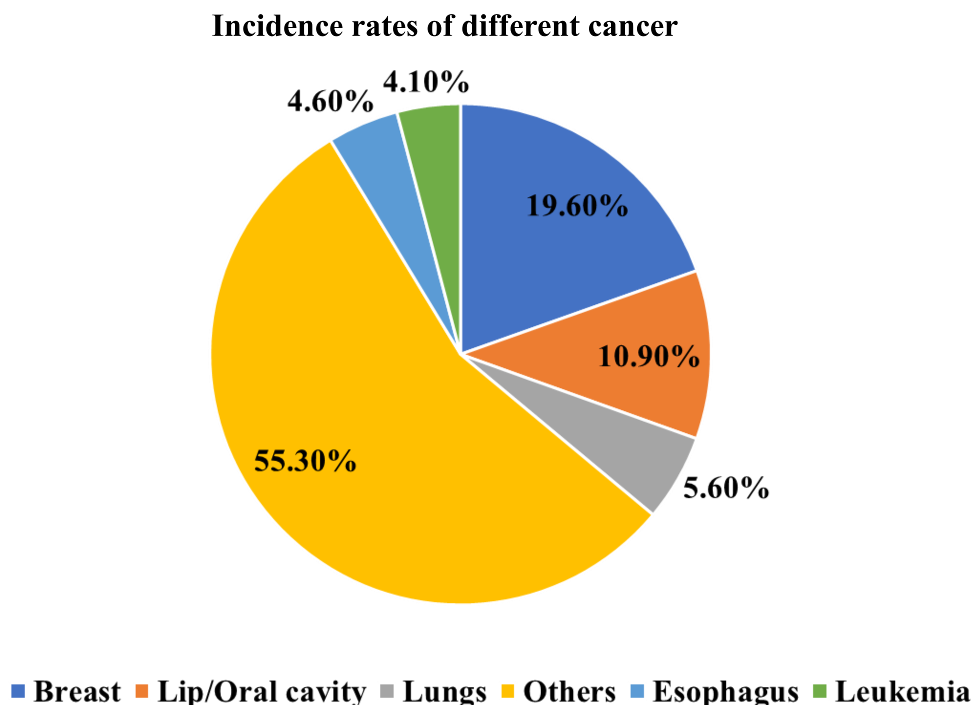


Figure 2. Incidence of major cancers in Pakistan (Khokhar et al., 2020).

The Hallmarks of Cancer

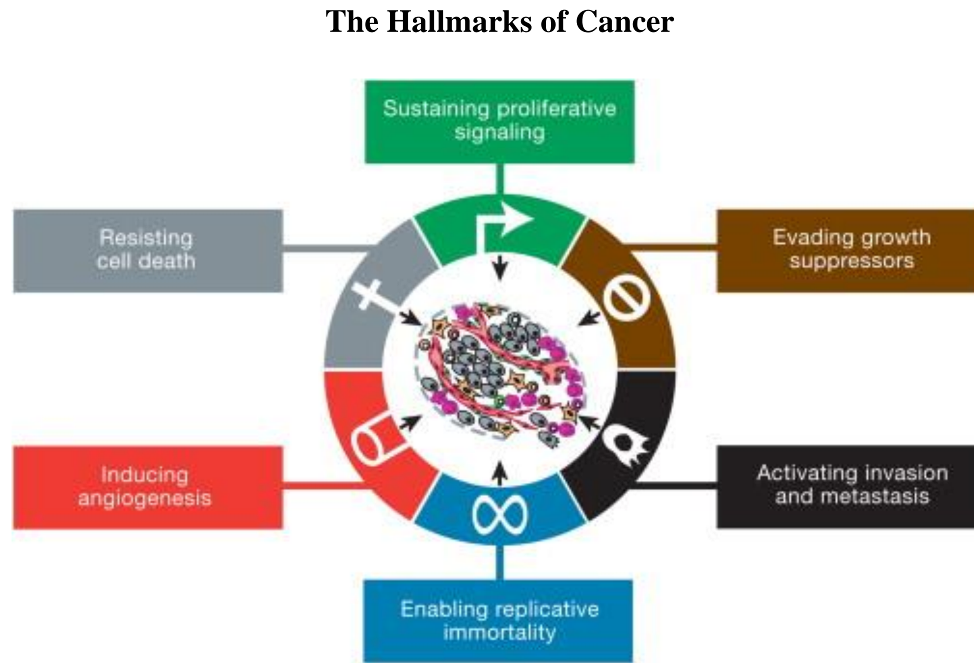


Figure 3. The Hallmarks of Cancer (Hanahan & Weinberg, 2011).

2.2 KPC family

The KPC represents the serine/threonine kinase family and it is basically part of the protein kinase AGC superfamily (Rosse et al., 2010; Ryu et al., 2010). There are various cellular processes of a cell type that are required for them. Depending upon their structure and their ability that how it reacts with DAG and calcium, KPC further classified into three subfamilies. These three subfamilies are classical, novel, and atypical (Steinberg, 2008).

The members of the KPC family basically take part in the signal transduction events of a few significant pathways that consist of cell progression, cell death and differentiation (Mackay & Twelves, 2007). The isotypes of the KPC family are prime elements of signalling pathways that accord a number of cellular phenomenon in both acquired and natural immunity (Baier, 2003).

All members of the KPC family for membrane interaction have their n phosphatidylserine (PS) binding domain. Catalytic and regulatory domains are organized at the C and N-terminus correspondingly in the structure of KPC molecules. These both domain structures are formed from a number of the conserved region that is scattered with regions of lower homology commonly named variable domains (Webb et al., 2000).

For physiological activation of classical KPC calcium and DAG are required. Whereas only DAG is enough for the activation of novel KPCs. Contrastingly, there is no need of calcium and DAG for activation of atypical PKPCs. Rather than the protein-protein interaction that is interceded by a PB1 domain together with carboxyl-terminal PDZ ligand motif help in the activation of their function. Cytokine and growth receptors involvement put forward a route for activation of PLCβ which further split (PI(4,5) P2) to produce soluble second messenger inositol triphosphate and lipid second messenger diacylglycerol. When lipid second messenger diacylglycerol is formed, then it recruited both novel and classical KPCs to plasma member, where they undertake an adaptational changes that result in complete activation (Black & Black, 2013).

2.3 KPCQ

The KPCQ is a member of novel KPCs which are Ca2+-independent and DAG-dependent KPC subfamily. It is primarily expressed in the T lymphocytes and muscles (Altman et al., 2000; Bauer et al., 2000). KPCQ dependent on DAG but they lack calcium. The location of this gene on chromosome 10p15, which is basically a part that is normally variant in cancer, leukemia T cells and immunodeficiencies (Zhang et al., 2013).

The level of KPCQ is high in lymphoid organs and skeletal muscle, primarily in lymph nodes and in thymus and to less extent in spleen but they show no expression in bone marrow (Meller et al., 1998). In T cells, among blood stem cells it is most abundant. At a lower level the enzyme is normally present in antibodies, leukocytes, macrophages/monocytes, and erythrocytes. From whole body mRNA expression analysis, this expression pattern was established ,which also manifest that the KPCQ mRNA is expressed in few part of the mouse brain and the peripheral nervous system (Bauer et al., 2000).

The KPCQ plays an important role in controlling the survival, T cells activation, and differentiation. Another important function is that they balance the growth of muscle cells, maintain stability, and alteration in skeletal muscle cells (Marrocco et al., 2014). They involved in the histopathology of numerous diseases, apart from its physiological functions. Low KPCQ activity in skeletal muscles tissue and immune system result in cause inflammation as well as autoimmune disease and insulin resistance and Type 2 diabetes (Zhang et al., 2013).

2.3.1 Discovery of KPCQ

Altman et al. demonstrated the novel KPC isotype's function in development, growth and in activation of T cells (Altman, 2000), at the same time they were also discovered and isolated KPCQ in early 1990s (Baier, 1993). Later on, other investigators cloned KPCQ cDNA from both humans and mice in various parts of tissue (HASHIMOTO, 1992; Osada, 1992).

The location of the human KPCQ gene is found with the help of chromosomal mapping within the short arm of chromosome 10 (10p15), and it is a specific position of a gene on chromosomes that have been previously identified as being linked with the variations that responsible for causing T cell lymphoma and immunodeficiency of T cell (Monaco, 1991; Verma, 1987).

2.3.2 Structure of KPCQ

It is a novel isotype of the KPC family which comprised of 706 amino acids. Molecular weight of KPCQ is 82kDA (G Baier et al., 1993). KPCQ is basically dependent on DAG but independent of Ca2+. It is a protein kinase that belongs to novel KPC subfamily and its structure is composed of several functional domains (Steinberg, 2008).

The C-terminal catalytic domain of KPCQ contains C3 and C4 domains whereas, N-terminal regulatory region accommodates the pseudo substrate region, C2-like domain as well as DAG-binding domain. The hinge region also called as V3 motif is well defined and specific to each KPC isoforms, the area which spilt both these catalytic and regulatory domains.

2.3.3 Localization of KPCQ

KPCQ shows few distinctive characteristics, which differ from other T cells expressed KPCs, significant because of its distinct and unique cellular localization in a clearly defined subregion of T cell. KPCQ is unique because its relocated to plasma membrane T cells , at the point of contact that joins the MHC and antigen specific T cells, (Monks et al., 1997), that are normally named as IS (immunological synapse) (Grakoui et al., 1999). For activation of T cells, IS act as the program that helps in carrying out sustained activation signals. The assembly of signalling complexes that include proteins, TCRs, and costimulatory receptors that are take part in signal transmission across the membrane of cell is necessary for functional IS (Dustin & Chan, 2000), the

reorganization of the actin cytoskeleton (Dustin & Cooper, 2000; Penninger & Crabtree, 1999) and the clustering of specified membrane microdomains that are named as lipid rafts (Bi & Altman, 2001; Miceli et al., 2001).

Compartmentalization of molecules in at least three distinguishable synaptic region is indicated by T cell antigen-presenting cells contact point allegedly named central SMAC, peripheral SMAC, and distal SMAC (Dustin, 2005; Monks, Freiberg, Kupfer, Sciaky, & Kupfer, 1998). KPCQ co-localizes along with TCR in SMAC (Monks et al., 1998). At high stoichiometry, the localization happened, that lasts for several hours. The dimensional association and duration of the IS influence the functional outcomes and promote the crucial phenomenon of differential T cell signalling. The cSMAC collecting of KPCQ, occurs simultaneously with its catalytic activation and it only happened after the T cells's energetic activation of T cell. Hence, KPCQ localization in the immunological synapse is closely related to the activation of T cells.

KPCQ translocate to membrane lipid raft in response to the regulation of Ag, and these lipid raft primarily localize to immunological synapse (Bi et al., 2001). As long as, KPCQ have not any clear targeted raft-signalling protein like palmitoylation sites then the structural beginning for the connection of KPCQ with lipid rafts remains unclear. The connection of KPCQ with another targeted raft-signalling protein including Lck tyrosine kinase is indicated by localization of lipid raft (Bi et al., 2001) which is known to be reside in lipid rafts through its double acylation and which we discovered to be linked with KPCQ regulated T cells (Liu et al., 2000). Relocation of KPCQ, the immunological synapse and lipid rafts is intensified by CD 28 stimulation. Another recent study demonstrates that the CD 28 stimulation has a specific role in selective concentration of KPCQ in the cSMAC.

The starting dislocation of KPCQ in the IS, where another process may work to distinctively enrolls it to the cSMAC, may be regulated by lipid raft of receptor-induced clustering that may act as an important driving force. For encouraging the proper cSMAC localization of KPCQ, lipid rafts at the IS may be particularly important (Hayashi & Altman, 2007).

2.4 Role of KPCQ in cell

KPCQ is particularly expressed in thymus, liver, nervous system, skeletal muscle cells as well as in natural killer cells, mastocytes, thrombocytes and hematopoietic system,

especially in T cells (Zhang et al., 2013). In case of immune system KPCQ has many known isotypes-specific functions and mice which lack of KPCQ expression exhibit problems in T cells activation because of signalling of abnormal Ca+ 2, NFkB, and NFAT activation pathways (Altman et al., 2004; Pfeifhofer et al., 2003; Sun et al., 2000).

By managing the pro- and anti-apoptotic Bcl2 family members expression, KPCQ modifies the T cells (Villalba et al., 2001). The role of KPCQ in immune responses is supported by several pieces of evidence, and there is no need for much immunity in resistance viral and bacterial infectious agents(Valenzuela et al., 2009). However, KPCQ appears to be needed for immunological responses that are linked with allograft rejection and inflammatory disorders, and this may be due to a specific require for KPCQ in the growth and development of Th17 cells, a subset of CD4+ T cells (Guéry & Hugues, 2015; Tan et al., 2006).

However, the role of KPCQ remains to be completely clear in non-hematopoietic tissues and cancers. It has greatly showed expression in a subset of GIST (gastrointestinal stromal tumor cells) cancers and as a result of cell division modulates increased number of cells (Ou et al., 2008).

2.4.1 Role of KPCQ in cell cycle

A variety of cell types which include gastrointestinal stromal cancerous cell and breast cancer cells, KPCQ act as a positive stimulator, where it mask the expression of p21C1p1 or p27Kip1 (Belguise & Sonenshein, 2007; Ou et al., 2008), and in case of capillary endothelial cells, KPCQ enhance the progression of G2/M (Tang et al., 1997).

KPCQ is basically entangled in T cells activation and helps to build-up proliferative signal in numerous ways which include those that are basically activated by CD28,TCR and TNF-α while it is not essential for CD4+ and CD8+ T cells differentiation (Altman et al., 2000; So & Croft, 2012). Early in T cells activation, KPCQ is enrolled to the IS, where it is essential for the development of the CBM complex, that plays crucial function in settling the downstream signalling during activation of T cells (Rawlings et al., 2006).

A variety of transcription factors that modulate the progression and activation of T cells, which include, NFAT, Ap1, and NF-κB pathways are triggered by KPCQ signalling

(Pfeifhofer et al., 2003). In regulatory T cells, KPCQ has an antagonist effect where it is segregated from IS and then stimulate the growth inhibition, but it plays a positive role in case progression of effector T cells.

2.5 Effect of overexpression of KPCQ

The upregulation of KPCQ improve the utilization of glucose into L6 cells which may be reflected in increases the activity in GS. The effects of insulin on 2-deoxyglucose intake were produced to ascertain whether this was the case. There was no notable difference in the intake of 2-deoxyglucose between L6u and L6 myotubes.

The effect of increased level of KPCQ was found to be linked with effects on basal and insulin-stimulated GS activation instead of enhanced glucose flux into cells. KPCQ has been entangled to play an important function in insulin resistance and insulin action and it is most abundant isotype of KPC in skeletal muscle (Monks et al., 1997; Yamada et al., 1995).

The KPCQ expression is upregulated in human skeletal muscles cell during differentiation, with increased KPCQ mRNA and immunoreactive protein being noticeable after the joining of myoblasts to multinucleated myotubes. The higher amount of expression of KPCQ in rat myoblasts did not impede with progression of cells before fusion to form differentiated myotubes. Although, overexpression in KPCQ may be act as a final marker for the differentiating of muscle cells but it is not essentially required for the process.

The high level of immunoreactive protein in the cell membrane of white muscle take place in fructose-induced insulin resistance rats but not in red muscle as reported by Reaven and his colleague (Donnelly et al., 1994) . They suggest that activation of KPCQ in fructose-fed rats was a feasible cause of insulin resistance in skeletal muscle (Donnelly et al., 1994). They also demonstrated that KPC act as an important inhibitor of insulin action, maybe by phosphorylation of the insulin receptor.

In differentiation of skeletal myoblast high expression of KPCQ was a normal event in and which myotubes from type 2 diabetic patients revealed the impaired expression of KPCQ. In response to insulin, skeletal myotubes from type 2 diabetic patients also display impaired activation of GS and transport of glucose, which would declare against high activation of KPCQ being engaged in insulin resistance (Donnelly et al., 1994).

2.6 Effect of loss of KPCQ

Loss of function mutation and in particular KPC isotypes, have increased the chances that some members of KPC family may function as tumor-suppressive (Newton, 2018). In case of immunological functions, a large number of the known isotypes-specific KPCQ functions, there are lacking of KPCQ expression in mice indicate the defects in T cells activation because of impaired signalling of Ca+ and activation of NFAT (Altman et al., 2004; Manicassamy et al., 2006). The T cells in mice that have insufficient KPCQ expose impaired differentiation toward T-helper subsets, and a complex loss of chemokines (Hayashi & Altman, 2007). In culture and triple-negative breast tumor xenografts, the novel isozyme of KPC, KPCQ, lower the expression of gene that stop the development of triple-negative breast cancer cells.

The loss of activity of KPCQ is responsible for both inflammatory, autoimmune disorders, and insulin resistance and Type 2 diabetes, in case of skeletal muscle tissue and immune system (Zhang etal.2013). A low levels of KPCQ expression remarkably reduced the number of apoptotic Schwann cells (Wang et al., 2021).

2.7 KPCQ and cancer

2.7.1 Leukaemia

KPCQ play an important role in regulation, cell progression, and survival of leukemic T cells and cytotoxic lymphocytes and leukemic T cells. Scientist name, Garaude et al. talked over the role of KPCQ in the development of T cell leukemia by using wild and the variant mice that have deficiency of KPCQ with Moloney-murine leukemia virus. Furthermore, they discovered that eventuality rate of leukemia was excessive in KPCQ deficient mice in comparison to wild type mice with identical leukemic cell and similar size of thymus and spleen. Hence it is proved that KPCO plays a key role in immunologic response to mice leukaemia (Garaude et al., 2008).

Notch signalling plays a crucial role by interconnecting with the pathways of pre T cell receptor, at the stage of T-cell differentiation and leukemogenesis, which converges on the regulation of well-defined pathway of NF-κB (Vacca et al., 2006). Felli et al. examined that a KPCQ was the downstream target of signalling of Notch3 and it could enhance the growth of the Notch3-induced T-cell leukemogenesis (Felli et al., 2005).

Later on, the graft-versus-leukaemia effect exhibits the most significant and influential immunotherapy to treat hematopoietic infections which is called leukaemia since the graft consist of donor T cells that can remove the residual cancerous cells (Appelbaum, 2001).

Furthermore, donor T cells in recipient identified the histocompatibility antigens of the healthy cells and produce an immune attack against the recipient cells, which result in destroying the organs of the recipient. They are potentially acts as an important outcome of BMT that attacks the body of recipient and called as graft-versus-host disease. Apart from it, factors that are used could restrict the graft-versus-host disease by inhibit the activation of T cells are broadly immunosuppressive and thus could result in cause contagious infections. So, thus search a target of therapeutic that maintain the GVL impacts as well as the immunological response to the antigen at the same time is very crucial and also inhibit GVDH (Isakov & Altman, 2012; Valenzuela et al., 2009).

By using mouse model, The scientist named Valenzuela et al show that T cells that deficit KPCQ were not able to undertake a powerful growth and thus cause destruction to the receiving organs, but these T cells are most significant that despite that maintain both graft-versus-leukemia and both virus contamination responses after bone marrow transplant (Valenzuela et al., 2009). Furthermore, these discoveries demonstrated after bone marrow transplant, KPCQ could be an important therapeutic target that is essential for deleterious but not beneficial functions of donor T cells.

By using genetic and pharmacologic techniques, The scientist ,Haarberg et al. proved that targeting KPCQ, along with KPC alpha, by using pharmacological inhibitors would be a good therapeutic strategy to stop the Graft versus host disease, while protecting functional GVL immune responses after BMT (Nangalia et al., 2013).

Both B lymphocyte and T lymphocytes are subtypes of Acute lymphocytic leukemia (ALL). Various therapeutic procedures are used to treat B cells and T cells. Hence it is necessary to recognize the techniques to differentiate B cells and T cells at once so that the possible treatment could be applied. Consequently, the scientist, Ma et al manifest that both KPCQ and CD3D (CD3 Delta Subunit Of T-Cell Receptor Complex) jointly can assist to distinguish the two types of ALL, but it is tough that a single gene is not able to distinguish between the two efficaciously (Ma et al., 2016).

2.7.2 Breast cancer

Breast cancer is the most common cancer that occurs in women and its incidence rate increased day by day. It is the leading cause of death among women globally (Harbeck et al., 2019; Liu et al., 2017)

It is considered a lethal and deadly malignancy, BC is hazardous and contagious and it becoming a serious threat and issue that tackle women all over the world (Ferlay et al., 2015; Livaudais-Toman et al., 2015)

In estrogen receptor-negative human breast tumours, KPCQ showed high expression, at transcript (Belguise & Sonenshein, 2007; Chin et al., 2006) and protein levels (Soyombo et al., 2013),but it somehow show weak expression in ER+ breast tumours but not expressed in normal breast epithelia. Epithelial breast cancerous cells lack progesterone receptor and Her2 protein and don't exhibit the estrogen receptor, these cancers basically belong to the triple-negative breast cancer. In comparison to ER+, the triple negative breast cancer is more invasive cancer. As compared to the ER+ tumor, KPCQ was highly expressed in triple negative breast cancer, in addition to tumor samples, and it is reflected by phosphor-538 KPCQ expression as it is present under its active form (Belguise et al., 2012) .

KPCQ was entangled in the mammary tumorigenesis that are induced by NF-κB family member c-Rel, and further investigation take place and showed them the cytoplasmic KPCQ function in cancer cells. Then, further, by KPCQ, discovered a novel NF-κB activation pathway. Expression of ERα stop through the activation of the Akt/FOXO3A pathway if there is a high amount of KPCQ present. Furthermore , the inhibition expression of ERα result in enhanced the c-Rel activity and led to the transcription of c-Rel target genes, therefore responsible for causing tumorigenesis (Belguise & Sonenshein, 2007).

Further investigations showed that KPCQ enhanced the RelB expression which is basically another member of the NF-κB family that control the expression of ERα (Nicolle et al., 2021) which further stimulate the mesenchymal phenotype of breast cancer cells (Wang et al., 2007). Later, it was reported another role of cytoplasmic KPCQ, that KPCQ induced the migration and invasion of breast cancer cells. Moreover, activation of KPCQ helps in enhancing the phosphorylation of AP-1 family member, FRA-1.

To increase the stability of protein and transcriptional activity of FRA-1, KPCQ-mediated phosphorylation, thus further induce the FRA-1 target genes expression (Belguise et al., 2017; Belguise et al., 2012). Whereas, KPCQ stabilize the FRA-1 protein through the ERK and SPAK pathways, instead of direct phosphorylation (Belguise et al., 2012). In addition to its localization in the cytoplasm, in its active form KPCQ was also found to distribute in the nucleus of triple negative breast cancer cells, in addition to be localized in cytoplasm (Zafar et al., 2014). However, according to the study from Zafar et al. showed that KPCQ is localized into the nucleus of triple negative breast cells (Zafar et al., 2014), localization of KPCQ in TNBC cells appeared to be particularly cytoplasmic with a smaller amount of it present in the nucleus. Due to subcellular fractionation methos, this variation may be occurred. A previous study indicated that signalling of phosphoinitide developed DAG, which is a co-activator of KPCQ , appeared to be present in nucleus, supporting the idea that KPCQ actively involved in nucleus (Visnjic & Banfic, 2007) .

Chromatin-bound-KPCQ was part of an active transcription complex that modulated the gene expression in breast cancerous cells which are primarily involved in cancer stem cell and EMT as reported by Zafar et al. (Zafar et al., 2015; Zafar et al., 2014). KPCQ enhanced epithelial to mesenchymal transition, which is an important process that are involved in the initiation of cancer by help of transcriptional control.

microRNAs transcription is negatively modulated by chromatin bound KPCQ and it is described by Sutcliffe et al., like, miR-9, and miR-200c, in T cells (Sutcliffe et al., 2011). These microRNAs play an important role in invasive breast cancer cells and in process of epithelial to mesenchymal transition with higher amount of KPCQ, and they are the most important and key factors in tumor metastasis (Gregory et al., 2008). Chromatin-bound-KPCQ phosphorylated Lysine-specific demethylase 1 (LSD1) was phosphorylated by Chromatin bound KPCQ and to regulate its demethylase activity and enhance the effect of LSD1 effect on epithelial to mesenchymal transition as reported by Boulding et al. (Boulding et al., 2018). Altogether, all these studies proved that KPCQ was involved in several steps of progression of breast cancer which include invasion, proliferation, and migration.

2.7.3 Role of KPCQ in different cancers

The adaptor protein that are known as, TBK-binding protein 1 (TBKBP1), has been showed to regulate direct phosphorylation of KPCQ of TANK-binding kinase 1 (TBK1) in results to stimulate the epidermal growth factor which further led to the activation of the oncogenic activity of TBK1, in a mouse model of lung cancer. The EGF of PD1 ligand 1 (PD-L1), a protein that is associated with immunosuppression, was induced because of TANK-binding kinase 1. Thus , these data explain a signalling pathway that is KPCQ-TBKBP1-TBK1which could help to control tumor growth and immunosuppressed (Zhu et al., 2019).

Englerin A and tonantzitlolone are two natural compounds, which activate the KPCQ and serve as a compound that act against tumor in KPCQ-dependent manner, in case of renal carcinoma cells. These two compounds induced an insulin-resistant phenotype in tumor cells through the activation of the heat shock factor, because of the activation of KPCQ (Sourbier et al., 2015; Sourbier et al., 2013). Thus , KPCQ inhibit the formation of actin stress fibre, Akt activation, and cell migration and it has also been used to stop and phosphorylate the Gα-interacting, vesicle-associated protein (GIV) activity in Hela cells, and then further inhibit Akt activation, formation of actin stress fibre and cell migration (López-Sánchez et al., 2013).

The higher level of expression of nuclear KPCQ could linked with poor survival and disease re-emergence, as reported by scientists, Chu et al. and they also found an expression of KPCQ that is not normal in oral squamous cancerous cells (Chu et al., 2012) .

Phosphorylated KPCQ (Ser-695) has been detected at protein level, but its role is not clear in case of ovarian cancer (Hu et al., 2020). There are two tumors, one is Ewing sarcoma/primitive neuroectodermal tumor (ES/PNETs) and the second one is malignant peripheral nerve sheath tumors (MPNSTS), in which Cytoplasmic KPCQ was show expression. For the diagnosis in ES/PNETs, a particular “dot-like” pattern of KPCQ staining could be used, whereas the function of KPCQ in both ES/PNETs and MPNSTS remain unclear (Kang et al., 2009).

Furthermore, the expression of KPCQ in other cancers might be interesting, and deciphering the functioning of KPCQ in all these cancers could be crucial to understand its pathological role.

2.8 Molecular targets and interactions of KPCQ

2.8.1 KPCQ role in the activation of AP1

API is activated in response to the stimulation of TCR and it is most important transcription factor. Requirement of AP1 is necessary for the T-cells activation and cytokines secretion. There are two most important components of API complexes one is c-Jun and the second is c-fos are upregulated during activation of T cells (Eferl & Wagner, 2003).

MAPK cascade results in causing the phosphorylation of c-Jun and c-fos and this cascade is triggered by T cells. Furthermore, an active heterodimer (AP1) complex is basically formed by c-Jun and c-for. There are some previous in-vitro studies that use Jurkat cells that have implicated JNK in connecting KPCQ to AP1 (Baier-Bitterlich et al., 1996). So, activation of JNK is normal in KPCQ-/- T cells and further suggests that an alternative process may take part in KPC-θ-regulated API. It was reported that JNK1/2 is not required for production of cytokines and primary T cells activation, these findings remained the same (Sabapathy et al., 2001).

At the transcription and translation level, API is upregulated. KPCQ also regulates AP1 at the transcription level. There is novel study that indicated that through the SPAK pathway, KPCQ regulates AP1 activation (Li et al., 2004).

KPCQ phosphorylates SPAK at serine 311 in vitro by direct interlinking with SPARK. KPC-θ-mediated activation of AP1 is decreased by an excess of mut form of SPAK or knockdown of SPAK. Serine 311 phosphorylation is influenced by T cells activation remains unclear. However, this study's observation strongly indicates that for the regulation of AP1 activation SPAK lies downstream of KPCQ.

2.8.2 Role of KPCQ in the activation of NAFT

There are at least five members are including in NAFT which is a member of a transcription factor. For activation of T cells, progression, and differentiation NAFT are most important for them. Furthermore, AICD and T cell homeostasis is controlled by NAFT signalling. The pathway that helps in the regulation and activation of NAFT is the Ca2+/calcineurin-dependent signalling pathway (Crabtree, 1999; Rao et al., 1997).

The phosphorylated form of NFAT is sequestered in the cytoplasm, under the resting phase. The TCR stimulation help to increase in intracellular calcium, which further play role in the activation of serine/ threonine phosphatase calcineurin. Then further Activated calcineurin dephosphorylates NFAT to reveal its nuclear localization sequences, that lead to translocation of NFAT into the nucleus. With the help of stimulating the influx of Ca2+ ions, activation of NFAT is increased by KPCQ which is proved by several studies (Altman et al., 2004; Manicassamy et al., 2006; Pfeifhofer et al., 2003).

The abnormal activation of PLCγ1, production of IP3, the influx of Ca2+, and translocation of NFAT to the nucleus are revealed by KPCQ-/- T cells. How KPCQ help in the activation of PLCγ1 is not yet clearly understood. Tec family of kinase member, Itk, is a connection between both KPCQ and PLCγ1. The target for Itk is PLCγ1 is the target for Itk and when Itk expression level is high ,it strongly activate NFAT and PLCγ1 without the help of crosslinking of TCR (Tomlinson et al., 2004).

T cells from mice deficient in the Tec family of kinases Itk and Rlk T cells deficient mice show impaired production of Ca2+ and IP3 influx due to low activity of PLCγ1, similar to KPCQ-/- T cells (Liu et al., 1998; Schaeffer et al., 2000; Schaeffer et al., 1999). Tec family of kinase help in the activation of PLCγ1 in restimulated KPCQ-/- T cells is reported by scientist named Altman et al. (Altman et al., 2004). In contrast to Itk, in naïve T cells , Tec is upregulated in restimulated T cells and show less expression in naïve T cells as compared to Itk (Tomlinson et al., 2004). KPCQ and Itk were present in the same complex. Furthermore, activation of Itk is decreased upon TCR stimulation in KPCQ-/- T cells. Effector T cells that are dependent on Tec to control KPCQ that help in regulation of NFAT and calcium influx and by the help of Itk, KPCQ control the influx of calcium in naïve T cells.

Thus, demonstrated that, KPCQ play role in the regulation of several signaling pathways. How KPCQ interlinked these pathways is one of the future directions in understanding the molecular mechanisms that underly KPCQ regulated T cell function.

2.9 Potential KPCQ substrates

By biochemical studies, many potential KPCQ substrates have been recognized. KPCQ has appeared to associate with them and then further phosphorylate adaptor protein that

is Cbl (Liu, 1999), a pro-apoptotic protein that is Bad (Villalba, 2001), and the cytoskeleton linker meiosis (Pietromonaco, 1998).

The latest study revealed that when KPCQ binds to and phosphorylates SPAK, this is important for the activation of AP1 (Li, 2004). Furthermore, KPCQ phosphorylates HePTP at Ser225 upon antigen stimulation of T cells and this KPCQ is also associate with hematopoietic protein tyrosine phosphatase (HePTP) (Nika, 2006). The phosphorylation of HePTP at Ser225 is important and thought that it is necessary to modulate the translocation of HePTP to the lipid rafts. When there is a mutation of S225A in HePTP occurred this result in increased KPC-θ-mediated activation of NFAT/AP1 pathways.

Therefore, observation from this study signifies that phosphorylation of HePTP is controlled by KPCQ and it is important for regulating TCR-mediated signals that are needed for activation of NFAT and AP1 processes.

CHAPTER 3: MATERIAL AND METHODS

3.1 In-silico analysis

3.1.1 Structural prediction of KPCQ

SNP data were obtained from databases such as ENSEMBL(McLaren et al., 2016). This database contained information on base and residue alterations, genome coordinates, amino acid coordinates (based on the GRCh38/hg38 genome assembly GRCh38/hg38), and variant ID. KPCQ transcript ENST00000263125.10 was employed for mapping the SNPs. The protein sequence of KPCQ, was retrieved from ENSEMBLE database in FASTA format with transcript ID KPCQ- **ENSG00000065675**. Information on the KPCQ protein was obtained from Uniport and ENSEMBL (Uniport ID Q04759).).

Data retrieved from the database consists of SNPs and variants related to genes, sequence retrieval, and disease association. Alpha-fold is the most authentic software that is used to predict the three-dimensional structure of KPCQ and InterPro(Hunter et al., 2009) is used to further validate the structure of KPCQ. By using, amino acid sequence as input in FASTA format and used to visualize and analyse structure using the PyMol(DeLano, 2002).

3.1.2 In silico SNP effect analysis

SNPs data of KPCQ was retrieved from ENSEMBL on 2nd August 2022 (Hubbard et al., 2002). The data that was retrieved provided us with information about amino acid coordinates, amino acid residues, genomic coordinates, variants IDs, and variant bases. Only missense variants were further considered which was1166 according to the data retrieved.

Synonymous, intronic, and in-frame SNPs all were eliminated after the processing of all types of variants and only missense or non-synonymous SNPs solely led to an altered protein product and were subjected for further analysis.

3.1.3 Evaluation of deleterious nsSNPs

Six in silico tools including SIFT, PolyPhen, CADD, REVEL, MetaLR, and Mutation Assessor were used to predict missense SNPs obtained from databases. These tool assign scores to the variants and categorize them into various classes of pathogenicity. The average percentage of pathogenicity obtained from these tools for the non-synonymous variants was determined and plotted on a graph. Pathogenicity percentage >70% was considered as a threshold. The highly pathogenic variants will be chosen based on their criteria, and a total of nine non-synonymous SNPs out of 1166 nsSNPs were predicted and selected by these techniques based on average and percentage that crossed the 70%threshold level. Only one of the most pathogenic non-synonymous variant C29Y were selected for further in vivo analysis that can change the structure and function of KPCQ's gene.

3.1.4 Stability analysis

I-Variant (Wang et al., 2013) which is a bioinformatics tool is used to evaluate the impact of non-synonymous SNPs on the structural stability of KPCQ. Protein sequence in FASTA format, and variant amino acids along with their coordinates were submitted to the tool I-Variant2.0 that presented the results in the form of DDG values (Kcal/mole). SNPs with delta G values below zero were considered to be unstable, whereas a DDG value greater than 0 indicates growing stability.

3.1.5 Structural and functional analysis

Variation in amino acid is induced by the presence of nsSNPs in a particular position of a protein. Nine variants of KPCQ were analyzed for their impacts on structure and function, and then Project HOPE (Mustafa et al., 2020) was used to predict the effects of altered amino acids. Project Hope considers physicochemical alterations caused by changes in the size of amino acids, hydrophobicity, and charge of amino acids, as well as the resulting loss of salt-bridge or hydrogen-bonding-based interresidual or interresidual interactions.

3.1.6 Prediction of disease association

Bioinformatic tool, FATHMM (Rogers et al., 2018), is used to predict the co-relation of KPCQ variants with cancer association of these variants with cancer. Variants are classified as passenger or carcinogenic according to scores obtained through this tool.

3.1.7 Analysis of filtered variant's flexibility

The effect of the mutation on a protein's dynamics was computationally assessed using DynaMut. It is an online program that uses normal mode analysis to predict modifications in protein levels. The difference in vibrational entropy (DDS) between variant and wild-type proteins is also predicted using this method.

3.1.8 In-situ mutagenesis

By changing the amino acids, the mutation was introduced in the wild type of structure of KPCQ and convert it into the variant structure. PyMol was used to predict and align the KPCQ's structure. For this in-situ mutagenesis, a bioinformatic software named mutagenesis wizard in the PyMol v4.0.4 (DeLano, 2002) was used. In .pdb format, variant structures were saved that were subsequently used for MD simulations.

3.1.9 Molecular dynamic simulations

Wild and variant structure of KPCQ were performed through Molecular dynamic simulations. These simulations were performed on supercomputers that use the open-source software package named as GROMACS 2016 (Van Der Spoel et al., 2005)that used OPLS-AA force field (Kulig et al., 2015). However, the commands for running the simulations were given using software PuTTY (Yanuar et al., 2018)(Yanuar, Chavarina, & Syahdi, 2018), and WinSCP (Boze & Libraries, 2002) was used that acted as a bridge between the supercomputer and the personal computer by assisting in transferring the files using ssh file transfer protocol(SFTP). The initial energy minimization was carried out at 50,000 steps, followed by 100ps volume and pressure equilibrium for molecular dynamic simulation.

GROMACS 2016 has an inbuilt tendency to construct the dynamic trajectories and it can also perform structural analysis of MD complexes using appropriate input commands. At every 10ps, trajectory coordinates were saved. The time of 20ns is enough to analyze the side chains, number of hydrogen bonds, protein backbone throughout course of simulation with the help of different parameters including Radius of gyration, Number of hydrogen bonds, Solvent accessible surface, Root mean square fluctuation, and Root mean square deviation.

3.1.10 Primer designing

The PCR primers were designed using Primer 1 (Collins & Ke, 2012), a bioinformatic tool for tetra ARMS PCR. The genome mapped from chromosome assembly 10p15 was used as an input in primer1. For the amplification of particular region of gene containing the targeted variants, four primers which contains two outer, and two inner primers were designed. Primers were also validated by UCSC in silico PCR (Hinrichs et al., 2006). Length of primer, temperature of primer, allele difference and position of SNP were chosen, but the remaining options were left as default.

3.2 Experimental work

3.2.1 Sample collection

First taking the approval from the Institutional Review Board of ASAB and CHM (Combined Military Hospital) and then collecting the 100 samples of patients with breast cancer and 100 blood samples of healthy controls. All procedures were carried out by the recommendations made by the ethical review board. Patients who volunteered to give their blood signed a consent form.

Patients with breast cancer as primary cancer were incorporated whereas patients with co-morbidity, pregnant and lactating females were excluded from this study. After gathering all the necessary data, including the patient's age, tumor grade, cancer type, treatment status, and receptor subtypes, history forms for the patients were completed. In addition to pathological characteristics, data on family history, breastfeeding, post- or premenopausal age, and alcohol and smoking use were also collected from patients.

Table 3. Primers for KPCQ

Variant ID	Sequences	Denaturation	Annealing	Extension
rs1248923790	**Forward Inner** CGAGGCTGTTAACCCTTACTG	**95 °C**	**63.5 °C**	**72 °C**
	Reverse Inner CATACTCTTTGACGAGCACAGCAA			
	Forward Outer GCACTGGTTAAGTAGACATGGG			
	Reverse Outer AATGTGTGTCCATGAGGAAGTC			

3.2.2 DNA extraction

The phenol-chloroform method was used to isolate genomic DNA from both patient and control samples. The blood samples were collected in EDTA tubes. After sample collection, take 500mico litters of blood in an Eppendorf tube and add the same amount of solution A, after that mixed it well, and then placed it at room temperature for 10 minutes. Then placed it in centrifuged at 13000 rpm for 1 minute, after that discard the supernatant and again add solution A, mixed it well, and again centrifuge it for 1 minute. Then discard the supernatant and add solution B (400micro litters), mixed it well until the pellet disappeared, centrifuged it for 1 minute, then discard the supernatant and add the solution B,12 microliters of SDS, and proteinase K in it, mixed it well, and placed in an incubator at 37°C overnight.

Then add solution C and solution D OF 250 microliters in the sample incubated overnight, inverted slightly, and centrifuged for 10 minutes. Then, the DNA-containing layer was discarded into separated layers, adding 55 microliters of sodium acetate and 500 chilled isopropanol, and inverted the tube was several times so the DNA gets precipitated, then centrifuged for 10 minutes. Discard the supernatant and add 200micro litters of ethanol, centrifuge it for 8 minutes, and let it dry so the ethanol gets evaporated after that add 40 microliter PCR water to it.

3.2.3 TETRA ARMS PCR

Detection of point mutations in DNA extracted from blood samples was found using ARMS PCR. There are two outer primers which are outer forward, outer reverse and two inner primers which are inner forward and inner reverse were used in TETRA ARMS. The purpose of outer primers was to amplify the entire gene and to detect the SNP inner primers that were utilized which were allele specific (table 3).

The reaction of PCR was carried out in a single tube which contain about twenty microliters of reaction volume which is composed of master mix (8 microliters), Forward outer/inner and Reverse outer/inner (1 microliter), template DNA (2 microliters), and nuclease-free water (6 microliters). Every PCR reaction consisted of 35 cycles and three major stages. Add the required reagents (DNA, PRIMERS, NUCLEASE FREE WATER, AND MASTER MIX) in a PCR tube and then gently tap them. The PCR samples were then microcentrifuge for four seconds quickly to mix all entities properly. The reaction was optimized at 63.5°C of the annealing temperature.

After that set the 96-well thermocycler machine and carry out samples at multiple temperatures simultaneously, as each row has different temperatures. A total of 35 cycles was run which takes approximately 1.5hours.

3.2.3.1 PCR steps and conditions

The first step, denaturation, or separation, of the two strands of the DNA molecule, is the first stage. The starting material is heated for five minutes at a temperature of about 95 °C to completely melt the DNA. This step was not repeated in stage two, instead DNA denaturation took place for 35 cycles at 94°C for 30 seconds.

In the second step, the annealing process was done at 63.5°C for variant **rs1248923790**, and 30 seconds were allotted for it. The annealing of DNA occurs in this step as gradient PCR was used so multiple temperatures were specified and this step required 30 seconds as well.

The third step involves DNA polymerase and begins adding nucleotides onto the ends of the annealed primers. The DNA was first polymerized at 72°C for 30 seconds following the annealing stage, and the DNA was polymerized there for a further 7 minutes.

3.2.4 Gel electrophoresis

It is a method that is used to separate DNA fragments based upon their size. DNA samples are loaded into wells at one end of a gel, and an electric current is applied to pull them through the gel. Due to the negative charge, DNA fragments move toward the positive electrode. Smaller DNA fragments move through the gel more quickly than larger ones because they all have the same amount of charge per mass. Due to the differences in each fragment's electrophoretic mobility, DNA fragments of various sizes will be separated. The DNA fragments can be seen as bands in a gel visualized in a UV Illuminator.

3.2.4.1 Agarose gel preparation

Make the 2% agarose gel, for this, measure the 2g of agarose and mixed it with 100ml of 1XTAE buffer. Then mix it in the flask and boil it in the oven until the solution is clear. After that, add 3 microliters of ethidium bromide to this solution. After placing the comb in gel casting tray, pouring the agarose into the gel tray and let it solidify at room temperature for 20 minutes. After the gel solidifies, take the comb out, remove

the scotch tape, and after that run the 5-microliter ladder in the first well after that run 15 microliter PCR samples in the next well. Electrophoresis was carried out at 80V and 400 A for 45-60 minutes to visualize the gel in a UV Illuminator. Similar procedures were used to visualize the DNA isolated from the samples. 8-10 µl of DNA samples were mixed with 1 µl of loading dye and placed into agarose gel wells.

3.2.5 Statistical analysis

To perform the statistical analysis of genotyping data GraphPad Prism (Mavrevski et al., 2018) were used. Two types of tests were performed, one is Fisher Exact and Chi-square test on the samples of both control and patients. The relative risk and odd ratio analysis were performed and established the appropriate confidence intervals. A P-value less than or equal to 0.05 P-value was statistically significant. Statistically significant P-Value shows the association of variants genotypes with breast cancer.

CHAPTER 4: RESULTS

4.1 Structural prediction of KPCQ

Through ENSEMBLE database, amino acid sequences of KPCQ were retrieved. This database provides detailed information about base and alteration of residue, amino acid coordinates, variants ID, and genomic coordinates. KPCQ transcript ENST00000263125.10 and the retrieved sequences, which have the transcript ID KPCQ- ENSG00000065675 were saved in FASTA format. The amino acid sequence of KPCQ was further used to predict the 3-dimensional structure of KPCQ, for this purpose, protein structure prediction, and alpha-fold were used which is the most reliable and authentic tool. By using, amino acid sequence as input in FASTA format, visualization of the 3-D structure of KPCQ was performed through PyMol v4.0.4. The information related to domains and amino acid residues of KPCQ was got from literature, which was predicted through software InterPro.

4.2 Prediction of deleterious nsSNPs

Total 1166 SNPs of KPCQ were obtained from ENSEMBLE database, which are used for predicting deleterious nsSNPs. This database basically consists of information about functional annotations of proteins, variants, sequence data, and the association of disease. Across 706 amino acids residues of KPCQ, coding SNPs were present. Only 429 deleterious non-synonymous SNPs were chosen and evaluate further their association with disease progression and impact on structure and function of protein were also checked. The Pathogenicity of KPCQ variants was predicted through various bioinformatics tools. These programs work on certain programming languages and algorithms based on which they are allocated specific scores to each variant. Based on these scores, all the variants were further categorized into pathogenicity group which includes, Deleterious variants, moderately damaging variants, and Tolerated variants. To sum up the effect of non-synonymous variants, average Percentage pathogenicity for all the selected variants was also determined (figure 4). Based on that, there were 19 variants with pathogenicity percentage greater than the threshold value of 70% and they were regarded as completely deleterious. 19 variants were chosen for additional evaluation and further stringent filtration criteria were applied to obtain the most deleterious variants in KPCQ. The selection criteria include that these variants must have POLYPHEN (>0.9), MetaLR (>0.5), CADD (≥30), SIFT (≤0.05), REVEL (>0.5),

and Mutation Accessor (≥0.8=Medium, >1.9=high). Based on this selection, nine most pathogenic variants were yielded in KPCQ and one of the topmost pathogenic variant C29Y(rs1248923790) was selected for further in silico and in vivo analysis. rsIDs, chromosomal and amino acid coordinates nine nsSNPs are listed in (Table 4). SNP effect scores from consensus tools for these nsSNPs are given (Table 5).

Table 4. Details of 9 Variants including IDs, genomic and amino acid coordinated.

SR	VARIANT ID	LOCATION	ALLELES	AA	AA CORD
1	rs769348033	10:6442080	A/T	I/N	550
2	rs1403981107	10:6462327	A/G	L/P	495
3	rs1380861138	10:6479165	C/A	V/F	394
4	rs1248923790	10:6515050	C/T	C/Y	29
5	rs145984477	10:6515056	G/A	P/L	27
6	rs780998510	10:6515057	G/A	P/S	27
7	rs765069464	10:6515119	C/A/T	R/L	6
8	rs765069464	10:66515119	C/A/T	R/Q	6

9	rs777420994	10:6515099	C/A/T	D/Y	13

Table 5. Scores of KPCQ gene nine deleterious variants were obtained through multiple consensus tools.

VARIANT ID	SIFT	POLYPHEN	CADD	REVEL	META_LR	MUTATION_ASSESSOR
rs769348033	Deleterious (0)	probably damaging (0.973)	likely deleterious (32)	likely benign (0)	Damaging (0.61)	Medium (0.763)
rs1403981107	Deleterious (0)	probably damaging (1)	likely deleterious (32)	likely benign (0)	Damaging (0.72)	High (0.974)
rs1380861138	Deleterious (0.01)	probably damaging (0.97)	likely deleterious (33)	likely benign (0)	Damaging (0.852)	High (0.997)
rs1248923790	Deleterious (0)	probably damaging (0.996)	likely benign (26)	likely disease-causing (0.708)	Damaging (0.524)	Medium (0.675)
rs145984477	Deleterious (0.03)	probably damaging (0.997)	likely benign (27)	likely disease-causing (0.654)	Damaging (0.537)	Medium (0.683)
rs780998510	Deleterious (0.03)	probably damaging (0.996)	likely benign (26)	likely disease-causing (0.634)	Damaging (0.523)	Medium (0.683)

rs765069464	Deleterious (0)	possibly damaging (0.995)	likely benign (28)	likely disease-causing (0.676)	Damaging (0.536)	Medium (0.706)
rs765069464	Deleterious (0)	possibly damaging (0.992)	likely benign (29)	likely disease-causing (0.57)	Damaging (0.536)	Medium (0.706)
rs777420994	Deleterious (0)	probably damaging (0.967)	likely benign (27)	likely disease-causing (0.538)	Damaging (0.52)	Medium (0.535)

Based on pathogenicity percentage, first filtered the SNPs. Pathogenicity percentages were determined using the pathogenicity scores provided by SIFT, PolyPhen, CADD, REVEL, Mutation Assessor, and MetaLR tools. After that, the percentage pathogenicity for all variants was determined, and plotted the graph for all nsSNPs KPCQ variants (figure 4).

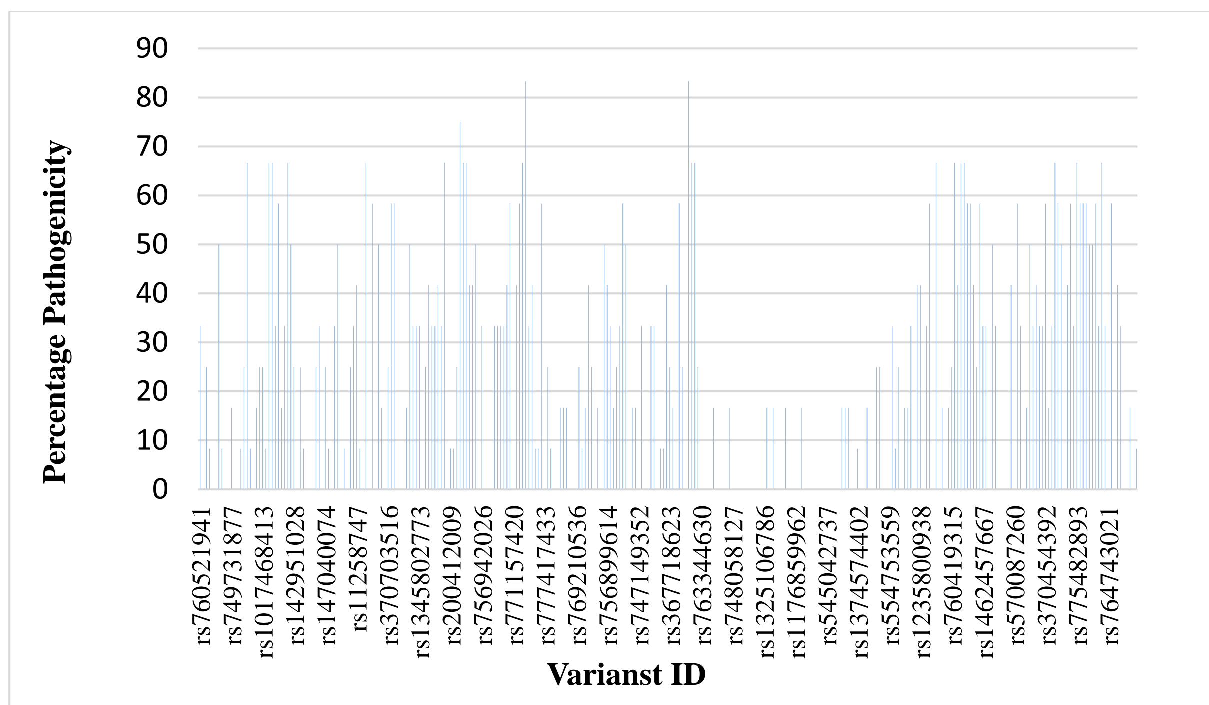


Figure 4. Pathogenicity percentage graph of KPCQ missense variants.

4.3 KPCQ DOMAIN

Alpha-fold is the most authentic and new software that is used to predict the three-dimensional structure of KPCQ. INTERPRO (Blum et al., 2021) is used to further validate the structure of KPCQ and identify four domains of protein. KPCQ consists of 706 amino acids, and it comprised of four domains including C1domain/DAG/PE binding domain, C2-like domain, C3 domain (binding site for ATP kinase domain) and C4 domain (AGC kinase domain). Six SNPs (C29Y, P27S, P27L, R6L, R6Q, D13Y) residing in the C2 like domain and three variants I550N, L495P, and V394F present in the protein kinase domain (Figure 5).

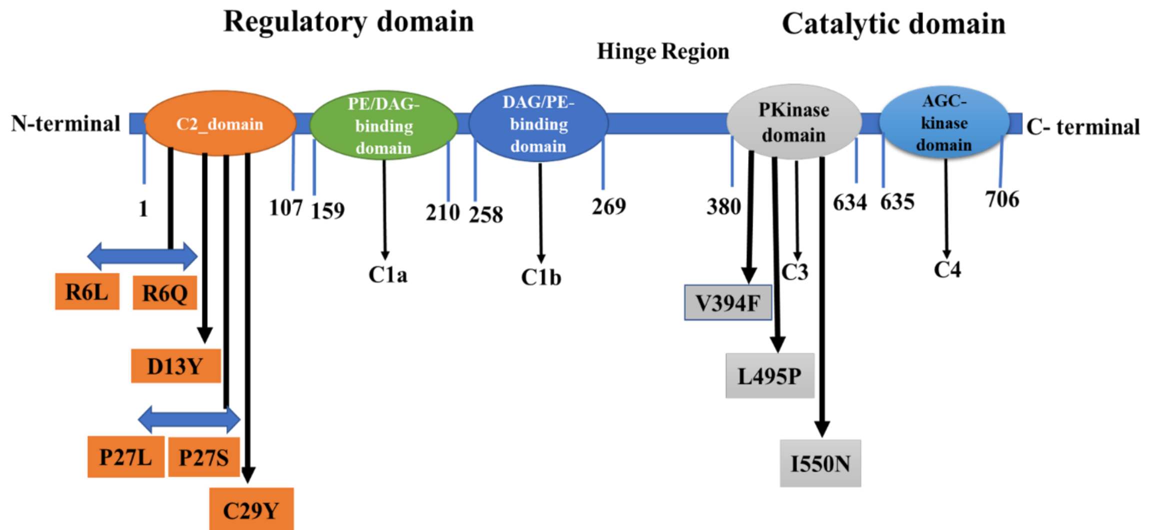


Figure 5. Map showing the domains and regions of KPCQ. Six SNPs (C29Y, P27S, P27L, R6L, R6Q, D13Y) present in C2-like domain in regulatory region. Variants I550N, L495P, and V394F localized in the protein kinase domain.

4.4 Mutation land mapping

The ENSEMBLE genome browser contains a total of 116 SNPs variants of KPCQ. Out of these 1166,429 missenses SNPs/ non-synonymous were selected. The missense variants of 706 amino acids were determined to see the abundance of variants present in the coding region which show that residues 11,44,131, and 592 had the highest frequency of missense variants (Figure.6a).

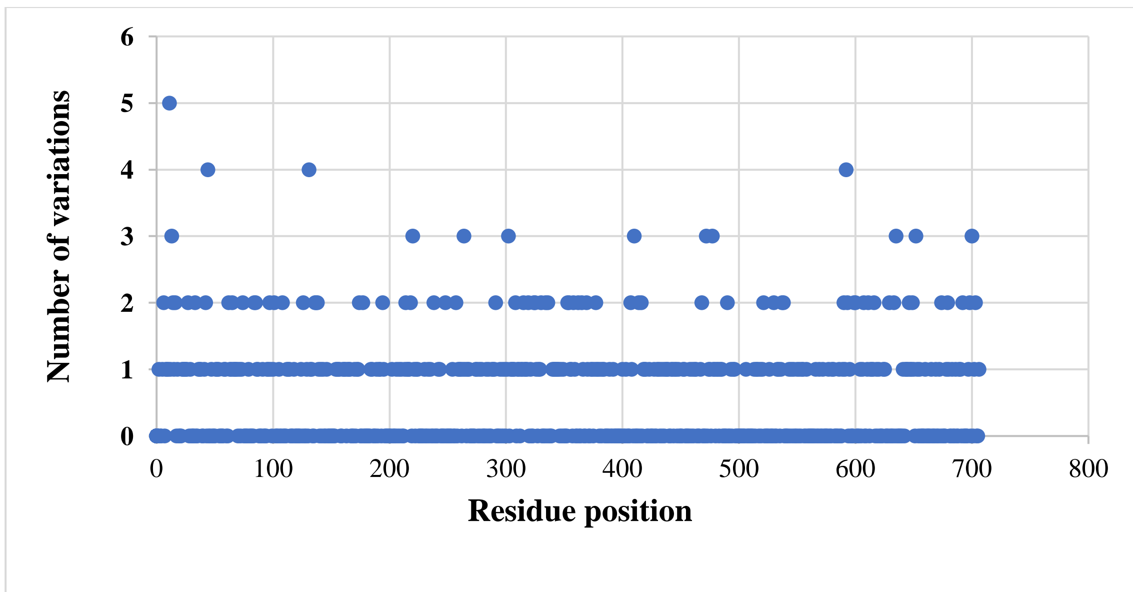


Figure 6(a). Mutational landscape of missense variants spread across 706 amino acid residues of KPCQ.

The relative abundance of missense variants across the 15 exons of KPCQ was determined. Exon 1 has the highest number of variation whereas the lowest number of variations were present at exon 4 and 12(Figure.6b).

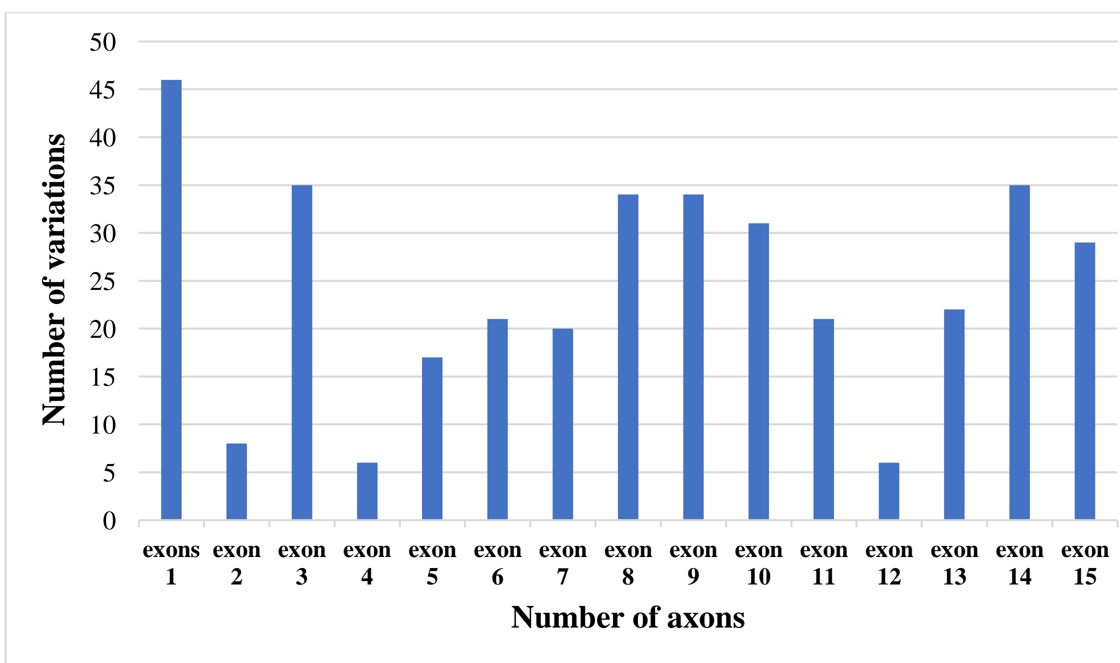


Figure 6(b). Distribution of SNPs across the fifteen exons of KPCQ.

4.5 Effect of nsSNP on structure and function

At a specific position in a protein, nsSNP variants cause an amino acid change. Nine KPCQ nsSNPs were evaluated through Project Hope which provide the information related to the properties of altered amino acids and their predicted impact. In comparison to wild-type, five (I550N, V394F, C29Y, P27L, and D13Y) out of nine variants that brought amino acid substitution were bigger. Nine SNPs were selected, and five (I550N, C29Y, P27S, R6L, and D13Y) of them exhibited greater hydrophobicity. Smaller size amino acids residues variation causes the protein folding disturbance whereas variant residues that were bigger in size and were responsible for the creation of repulsion and bumps. The charge of residue is change from negative to neutral in case of D13Y variant and change from positive to neutral in case of R6L and R6Q(Table.6). The loss of inter amino acids molecular interaction which include hydrogen bonds, was induced by variants that caused the substitution of large and neutrally charge amino acids. Flexibility was reduced due to non-polar amino acids substitution.

Table 6. KPCQ nsSNPs properties, location, and effect on its structure predicted through PROJECT HOPE.

Variants Variant amino acids Mutation location					
Size charge				**Hydrophobicity**	
Wild type Variant					**Domain**
I550N	Bigger	.	.	more	Kinase Domain
L495P	Smaller	.	.	.	Kinase Domain
V394F	Bigger	.	.	.	Kinase domain
C29Y	Bigger	.	.	more	C2
P27L	Bigger	.	.	.	C2
P27S	smaller			more	C2
R6L	Smaller	+	Neutral	more	C2

R6Q	Smaller	+	Neutral	.	C2
D13Y	Bigger	_	Neutral	more	C2

4.6 Association of pathogenic SNPs with disease

The association of KPCQ variants with cancer was predicted through the FATHMM tool. Individual mutations result from FATHMM are in form of functional scores. Out of nine variants, three variants (I550N, L495P, V394F) were passengers and six variations (C29Y, P27L, P27S, R6L, R6Q, and D13Y) were carcinogenic (Table.7). This suggests that six SNPs specifically C29Y, P27L, P27S, R6L, R6Q, and D13Y have association with cancer and they responsible for causing dysregulation of protein.

Table 7. Estimated association of KPCQ SNPs with cancer FATHMM.

VARIANTS	PREDICTION	SCORE
I550N	PASSENGER	0.69
L495P	PASSENGER	0.13
V394F	PASSENGER	-0.58
C29Y	CANCER	-1.18
P27L	CANCER	-1.87
P27S	CANCER	-1.92
R6L	CANCER	-1.11
R6Q	CANCER	-1.16
D13Y	CANCER	-1.07

4.7 Impact of SNP on KPCQ stability

Through I-Variant, bioinformatic too, all nine KPCQ variants were evaluated for their potential influence on protein stability. All variants are responsible to bring about the destabilization in structure of protein except D13Y variant. Based on the value of DDG, score less than zero cause the structural destabilization of protein. Out of nine variants, four variants L495, V394F, P27S, and R6Q had the lowest DDG score which indicate that these variants responsible for causing the destabilization of protein (Table.8) (figure 7).

Table 8. Influence of SNP stability analysis on variant pathogenicity through I-Variant.

Position	**Original residue name**	**Variant residue name**	**Stability**	**DDG**
550	I	N	Decrease	-1.42
495	L	P	Decrease	-1.49
394	V	F	Decrease	-1.88
29	C	Y	Decrease	-0.01
27	P	L	Decrease	-0.80
27	P	S	Decrease	-2.57
6	R	L	Decrease	-1.18
6	R	Q	Decrease	-1.46
13	D	Y	Increase	0.22

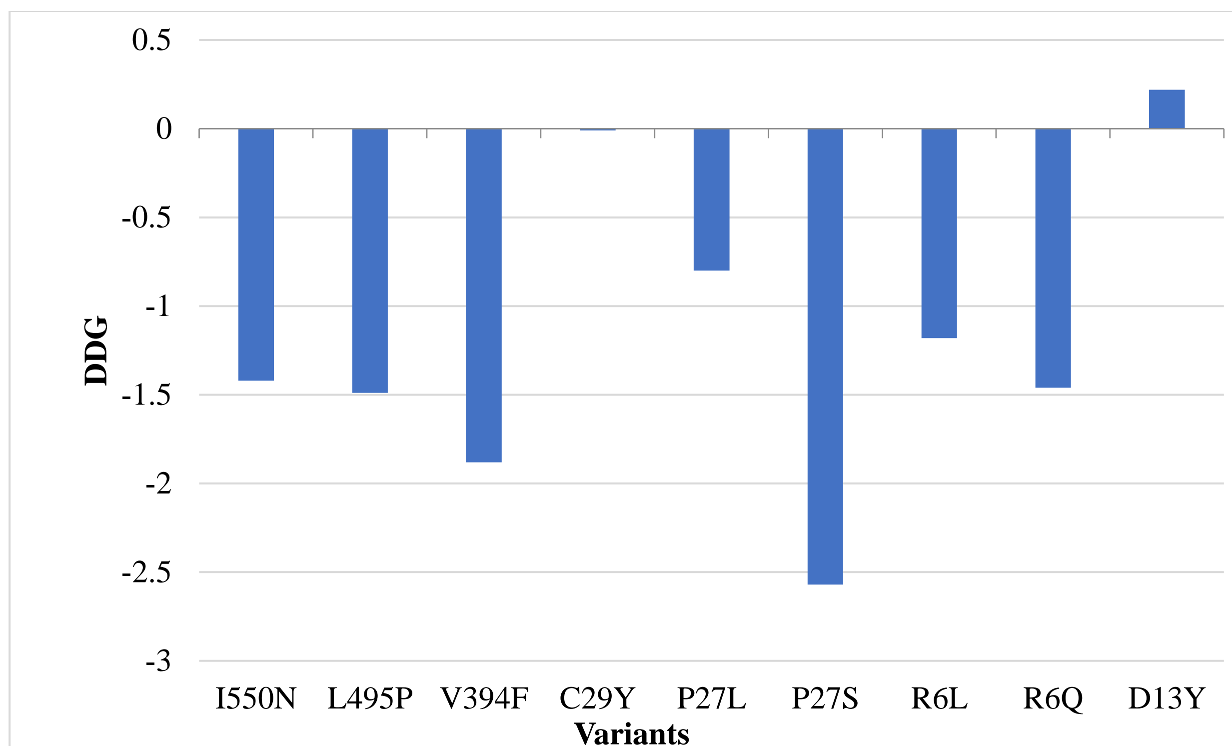


Figure 7. SNP stability analysis of KPCQ variants through I-Variant v2.0. Except for D13Y, all other variants were estimated as destabilizing mutations. Estimation was based on Delta G value where negative Delta G values were regarded as destabilizing and positive Delta G values were taken as stabilizing.

4.8 Flexibility and stability analysis

DynaMut is used to evaluate the impact of nine SNPs on protein's flexibility. Through ENCom values, stability effect was analysed. These variants have impact on protein structure and function, according to ENCom values. Increase vibrational entropy resulted in increase the molecular flexibility in case of four variations (I550N, L495P, R6L, and R6Q) out of nine variants of KPCQ whereas, Five variations (V394F, C29Y, P27L, P27S, and D13Y) lower molecular flexibility. Interatomic interaction of protein may be influence by the change of molecular flexibility of protein. A comparison of intramolecular interactions of wild and variant structures is done in figure 8.

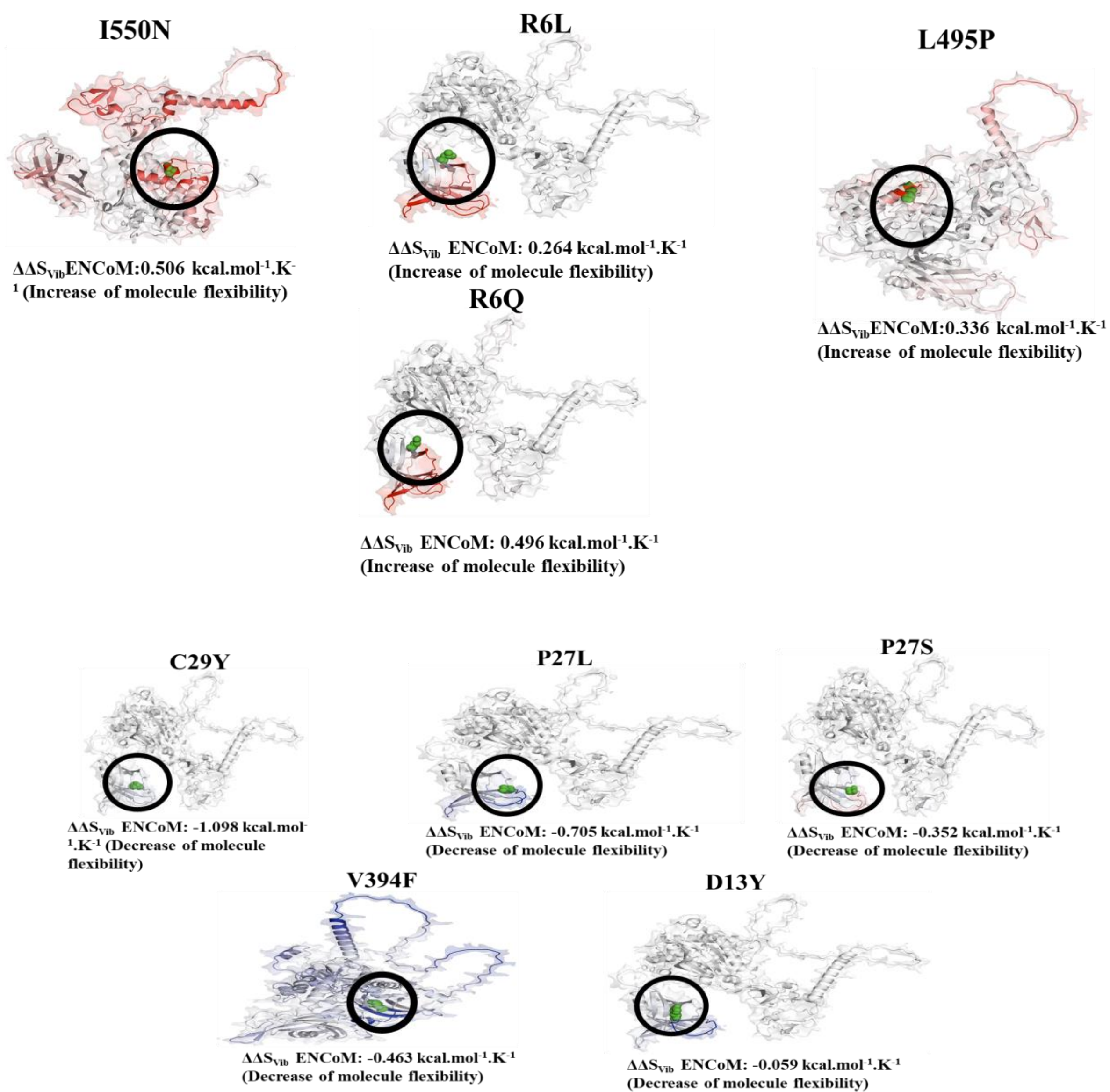


Figure 8. Flexibility analysis of wild and variant structure of KPCQ through DynaMut.

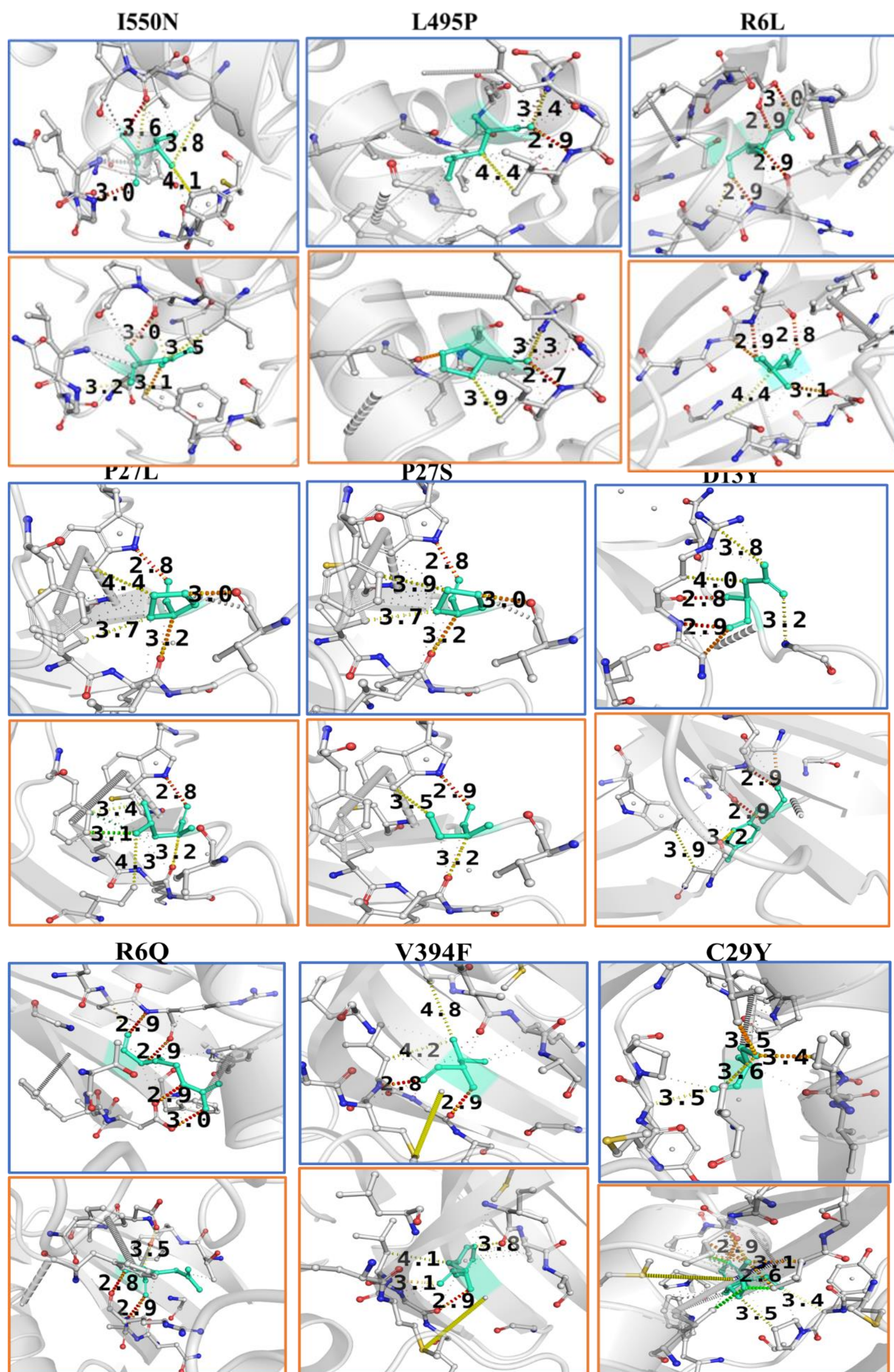


Figure 9. Comparison of interatomic interatomic interactions of wild and variant structure of KPCQ variants.

4.9 In-situ Mutagenesis

Alpha-fold is used to predict the three-dimensional structure of protein. of KPCQ predicted by alpha-fold and converted into the variant structure. Furthermore, the structure of KPCQ was visualised and aligned through PyMol. For this in-situ mutagenesis, a bioinformatic software named mutagenesis wizard in the PyMol v4.0.4 was used figure 9.

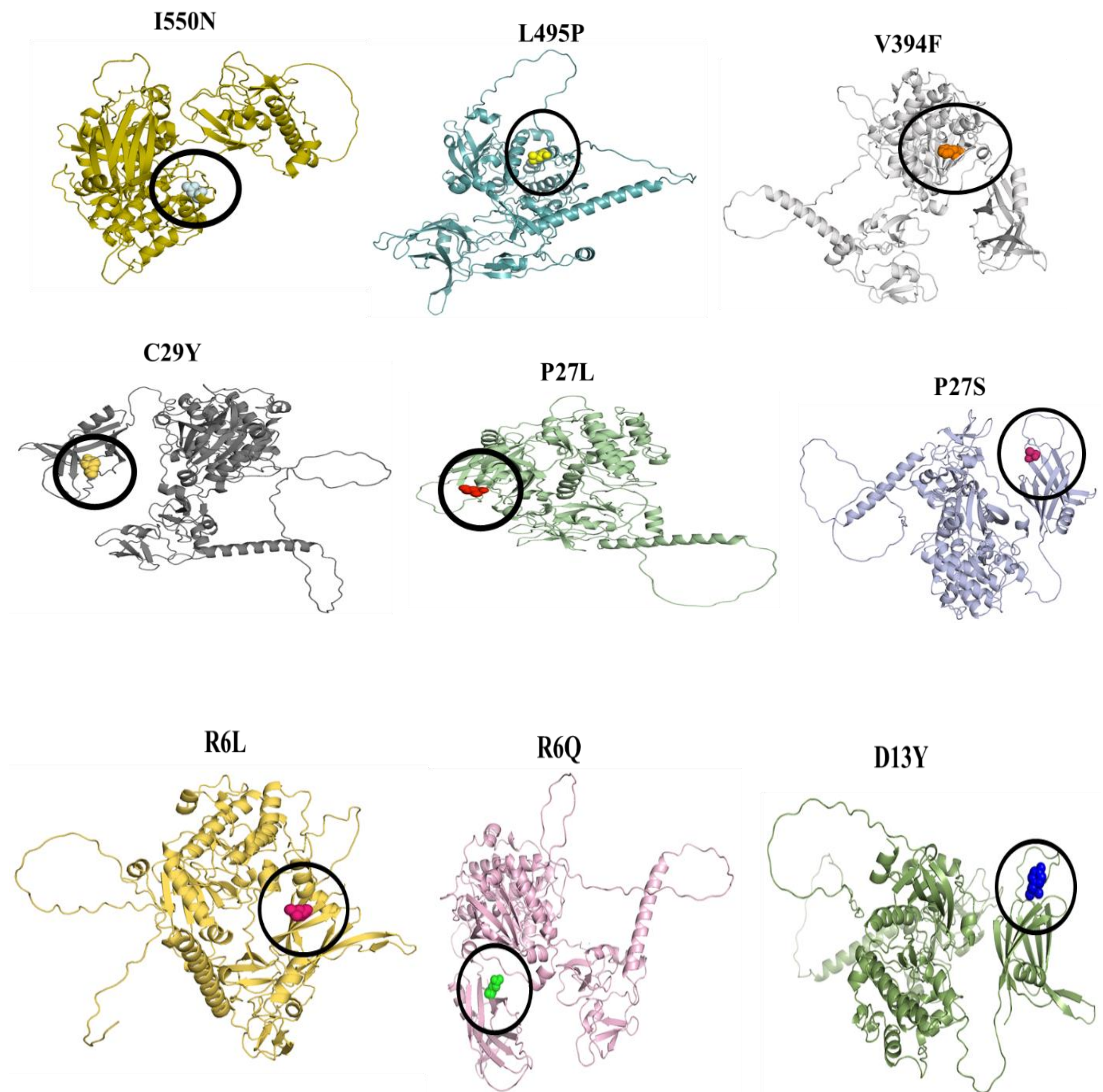


Figure 10. In-situ mutagenesis of KPCQ variants.

4.10 Molecular dynamic simulation

The stability of the predicted model of KPCQ structure was evaluated with the help of bioinformatic software GROMACS. Mutagenesis wizard tool of PyMol was used to induce the point mutation. GROMACS version 5.1 tool is used to determine the effect of obtained variant structures on wild type protein structure. 20NS MD simulation was run for both variant and wild type KPCQ C29Y variants. Different parameters including Root mean square fluctuation, Radius of gyration (Rg), Root mean square deviation (RMSD), Solvent accessible surface (SAS2) and number of hydrogen bonds were used to analyze the trajectory files.

4.10.1 Radius of gyration

It is defined as the radial distance of all the atoms in a protein from their common axis. In a dynamic situation, Rg is basically the radius of a protein structure, which indicates the compactness of protein and change in folding over the time. The overall radius of gyration of the variant structure of C29Y(rs1248923790) is greater than the wild type of structure. The radius of gyration of the variant structure of C29Y variants is highest at 3.49nm during 4.6ns. This indicates that the compactness of the variant structure of protein was decreased as the stimulation proceeded more significantly compared to the wild type as shown in figure 10.

4.10.2 Number of hydrogen bonds

Number of hydrogen bonds of both wild type and variant C29Y(rs1248923790) variant of KPCQ show no significant difference as there is substitution of only one amino acid. The line of both variant and wild type proteins superimposed on each other indicating that there is no significant difference as shown in figure 11.

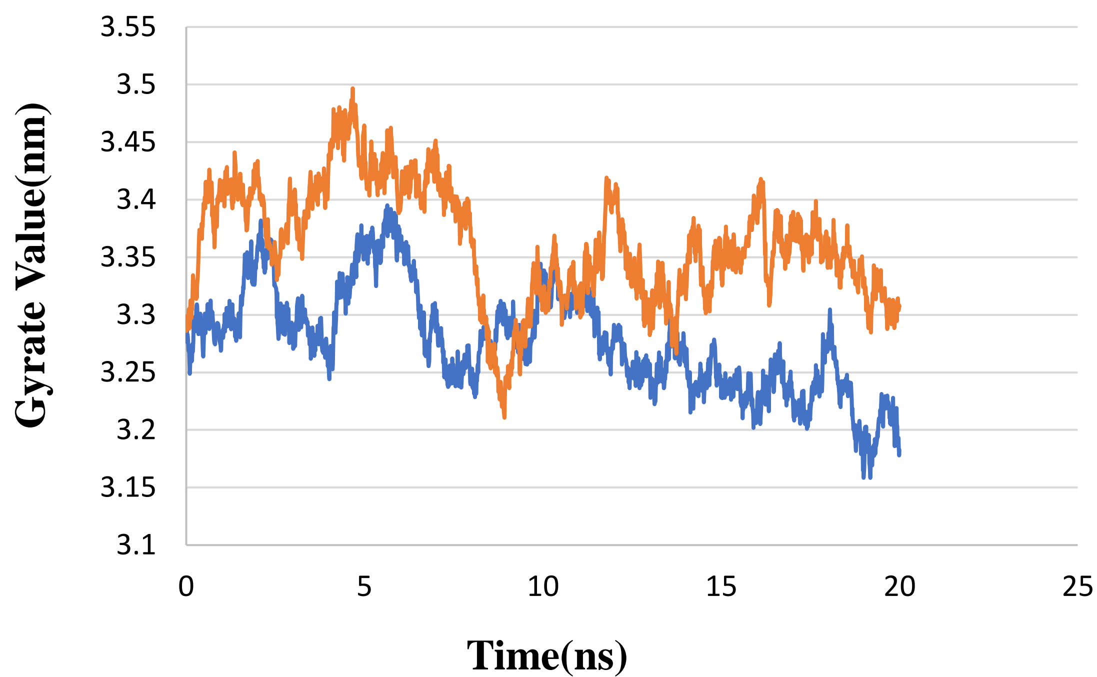


Figure 11. This graph shows the compactness of both variant and wild protein in a dynamic setting over time.

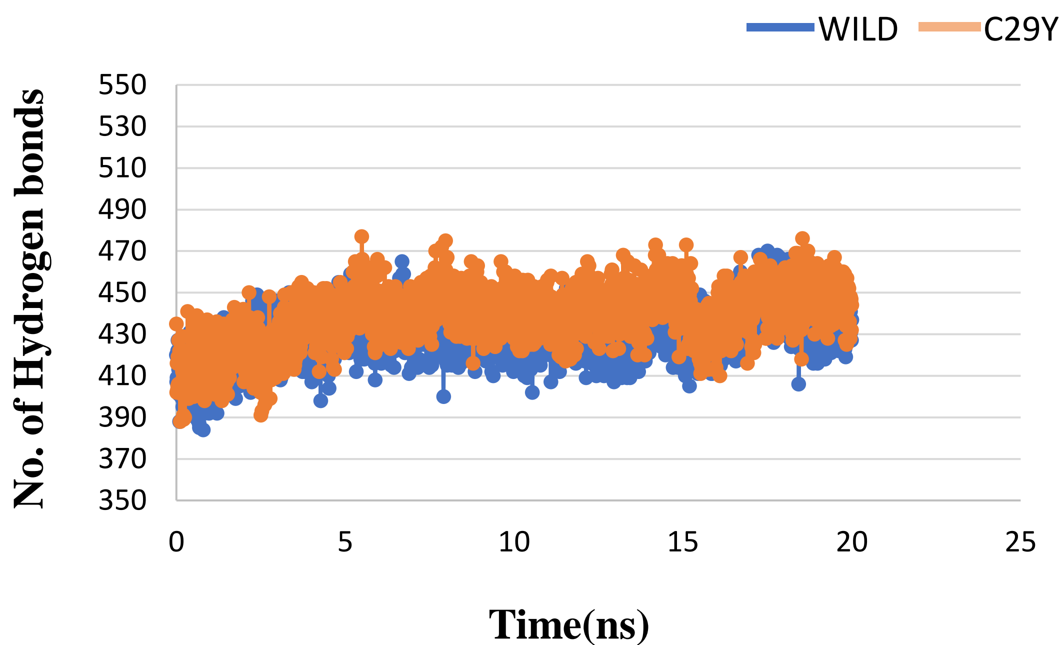


Figure 12. This graph represents the difference in the number of hydrogen bonds between variant and wild-type proteins structure.

4.10.3 Root mean square deviation.

The deviation of different atoms in protein from its mean point is defined as RMSD (Root mean square deviation). As compared to wild type, variant structure of KPCQ C29Y variant deviates significantly from its mean position. Within 0.37 nanosecond at 0.36nm the fluctuation is noticed in the variant protein and this pattern is followed till the end with the highest deviation 1.60nm at 19 nanosecond. The RMSD analysis of variant and wild structure shown in figure 12.

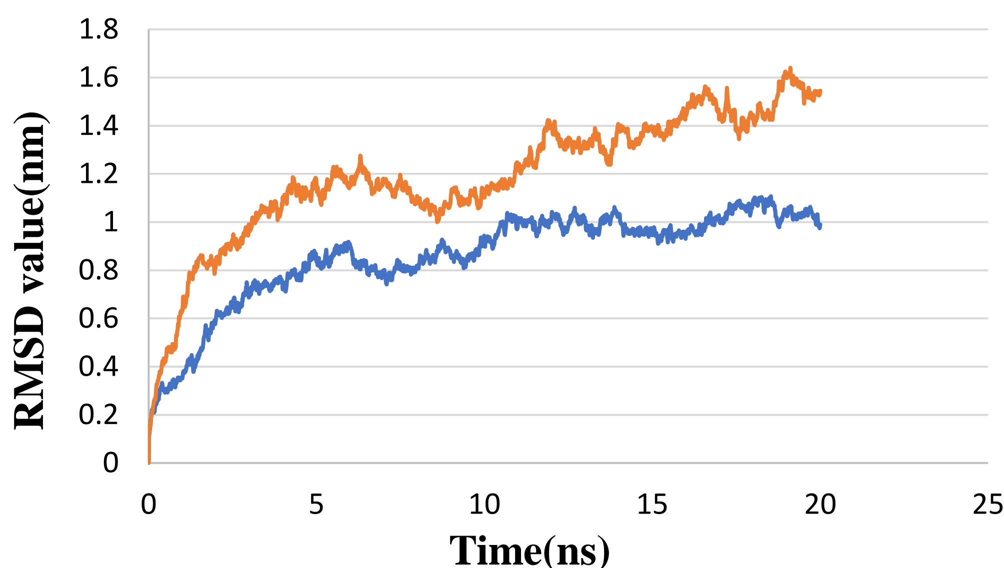


Figure 13. This graph shows the significant difference and deviation in wild type and variant residues over the time.

4.10.3 Root mean square fluctuation.

It represents that the individual residue deviates from its mean point. RMSF analysis showed the difference in the fluctuation of both variant and wild type protein residues. The region from 16-135,213-271,315-351 and 430-646 residues of the variant protein has shown significance variation from its mean position, which shows that the structure of expands as the simulation proceeds. Graphical representation of RMSF values of both wild type and variant protein are shown in figure 13.

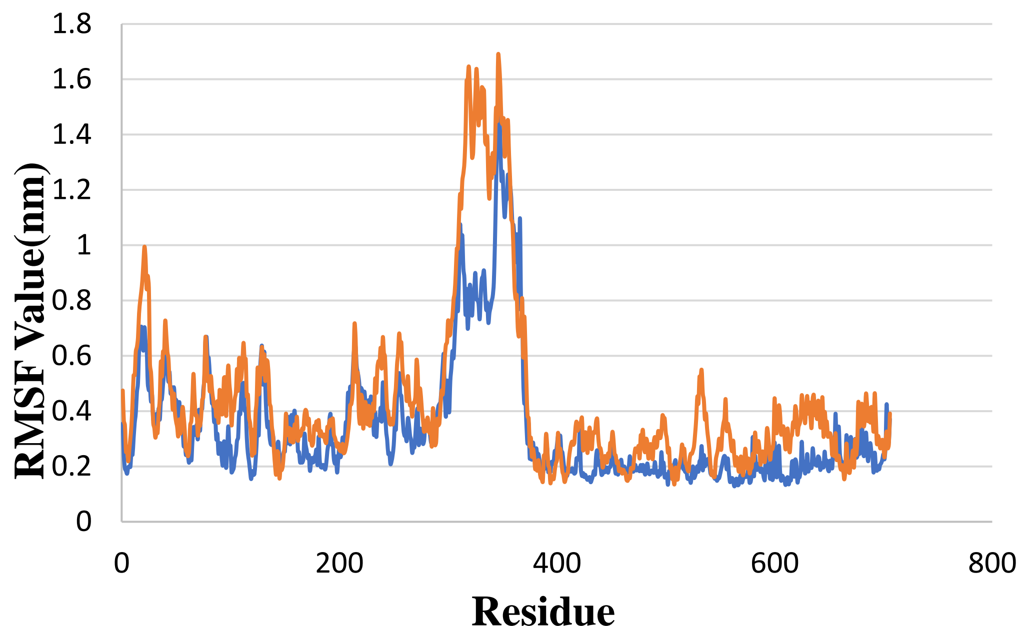


Figure 14. Graphical representation of RMSF values of both wild type and variant protein.

4.10.4 SASA analysis

Calculating the solvent-accessible surface area (SASA) is a simple method that address the surface features of protein, which include whether the surface is polar or non-polar, or to differentiate between exposed and buried sections or amino acids. The SASA plot are shown in figure 14. The SASA value of variant protein structures is larger than that of the wild type of protein structure, which represent that the variant structures are to some extent unstable. The variant structure of the C29Y variant fluctuated from 2.4ns -4.9ns from wild type structure but after 5ns the wild type and both variant type proteins show almost similar values of solvent-accessible surface area. This result indicates that the surface accessibility of C29Y is high compared to wildtype structure.

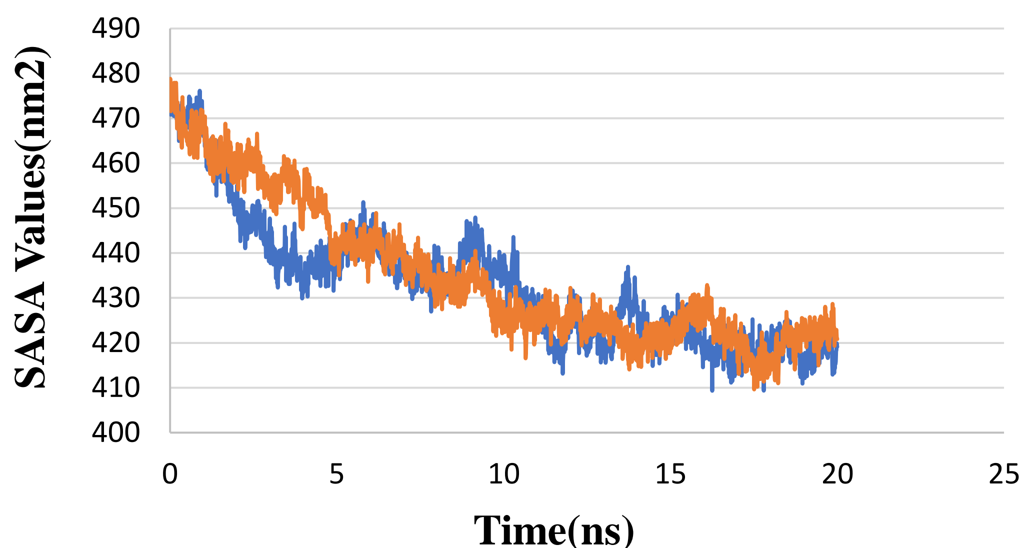


Figure 15. Graphical representation of SASA values of both wild type and variant protein.

4.11 Genotype frequency distribution and association of KPCQ with breast cancer

Phenol-chloroform DNA extraction procedure is used for the extraction of DNA from blood samples. Genotyping analysis was performed on DNA samples to predict the KPCQ non-synonymous genetic variant C29Y with variant ID rs1248923790. Tetra ARMS-PCR is used to perform genotyping analysis, and this PCR used to amplify the sequence of targeted gene and polymorphic variants that present in that genomic sequence were genotyped.

Genotype frequency distribution of both control and patient samples for KPCQ genetic variant C29Y are shown in the table. It has been observed that the GG genotype of variant C29Y was significantly correlated with breast cancer with a distribution frequency of 60% which was the highest among all genotypes. Genotype TT of this variant was also statistically correlated with the incidence of breast cancer. The Heterozygous genotype GT was statistically insignificant and could not be linked with the incidence of breast cancer. Statical analysis for the distribution of allele frequency represent that both GG and TT alleles were significantly showing association with breast cancer. Frequency distribution and genetic association of KPCQ variant with breast cancer were given in table 9 and show in figure 15.

Table 9. Frequency distribution and genetic association of KPCQ variant with breast cancer.

	Frequency Distribution		**Odd Ratio**		**Relative Risk**		**P-Value**
Genotypes	**Control**	**Patient**	**Value**	**95%CI**	**Value**	**95%CI**	
GG	60%	40%	2.25	1.271 to 3.919	1.5	1.131 to 2.017	0.0071
GT	8%	10%	0.7826	0.2975 to 2.165	0.8792	0.4787 to 1.358	0.8056
TT	32%	50%	0.4706	0.2647 to 0.8479	0.6772	0.4896 to 0.9129	0.0143

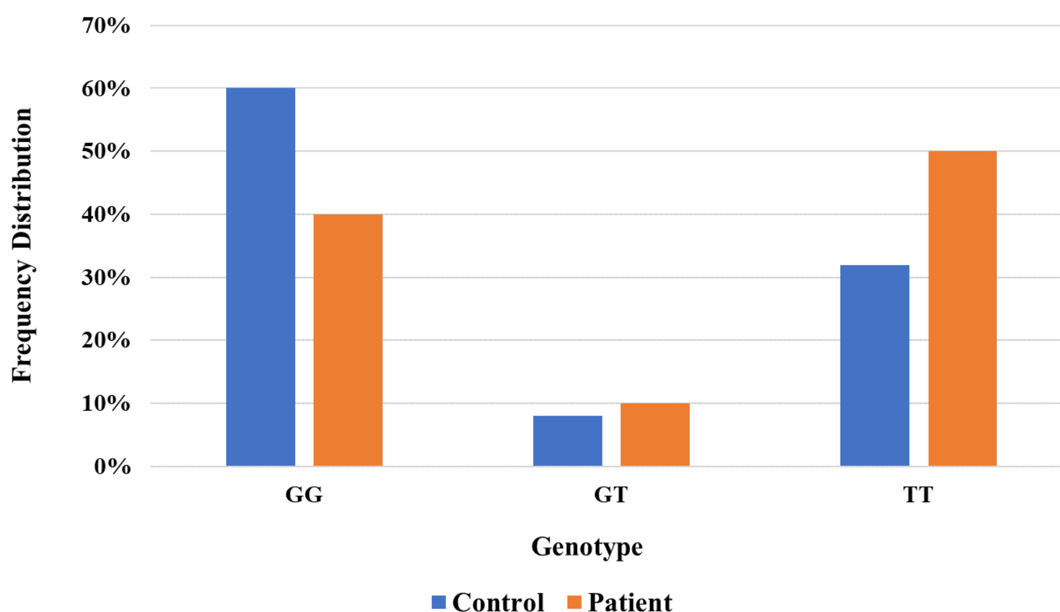


Figure 16. Genotyping analysis of KPCQ variants with breast cancer.

For the treated state of breast cancer patients genotype distribution frequency was determined. It has been revealed that in variant C29Y all genotypes were significantly related to the incidence of breast cancer and the influence of treatment is 37% in the case of the TT genotype, and there is a highly significant value of the TT genotype which shows that there is a high risk of breast cancer. Distribution frequencies of KPCQ variant C29Y genotypes related to treated states is given in the table 10 and show in figure 10.

Table 10. Genotype frequency distribution and association of KPCQ variant with treated state of breast cancer.

	Frequency Distribution		**Odd Ratio**		**Relative Risk**		**P-Value**
Genotypes	**Untreated**	**Treated**	**Value**	**95%CI**	**Value**	**95%CI**	
GG	9%	31%	4.543	2.019 to 10.43	1.797	1.376 to 2.271	0.0002
GT	1%	9%	9.791	1.545 to 108.5	1.879	1.222 to 2.290	0.0185
TT	13%	37%	3.93	1.914 to 7.723	1.762	1.353 to 2.250	0.0001

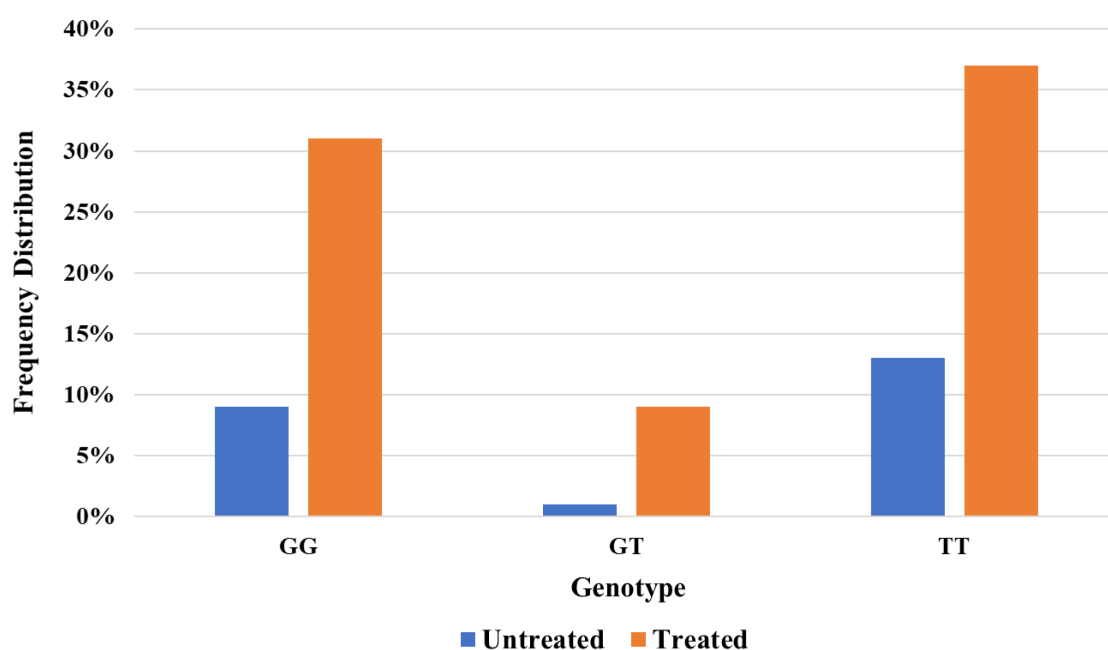


Figure 17. Genotyping analysis of KPCQ variants of both untreated and treated states in relation to breast cancer.

For metastatic state of cancer patients genotype distribution frequency was also determined. It has been revealed that in KPCQ variant C29Y, none of the genotypes showed any significant correlation to metastatic states in breast cancer, and the GG, GT, and TT genotypes were shown a protective role from metastasis. Distribution frequencies of KPCQ variants genotypes related to metastatic states is given in the table 11.

Table 11. Frequency distribution and genetic association of KPCQ variant with metastatic breast cancer.

	Frequency Distribution		Odd Ratio		Relative Risk		P-Value
Genotypes	**Non metastatic**	**Metastatic**	**Value**	**95%CI**	**Value**	**95%CI**	
GG	19%	21%	1.133	0.5652 to 2.313	1.063	0.7369 to 1.438	0.8599
GT	5%	5%	1	0.3140 to 3.184	1	0.4679 to 1.583	>0.9999
TT	26%	24%	0.8988	0.4667 to 1.713	0.9474	0.6651 to 1.284	0.8704

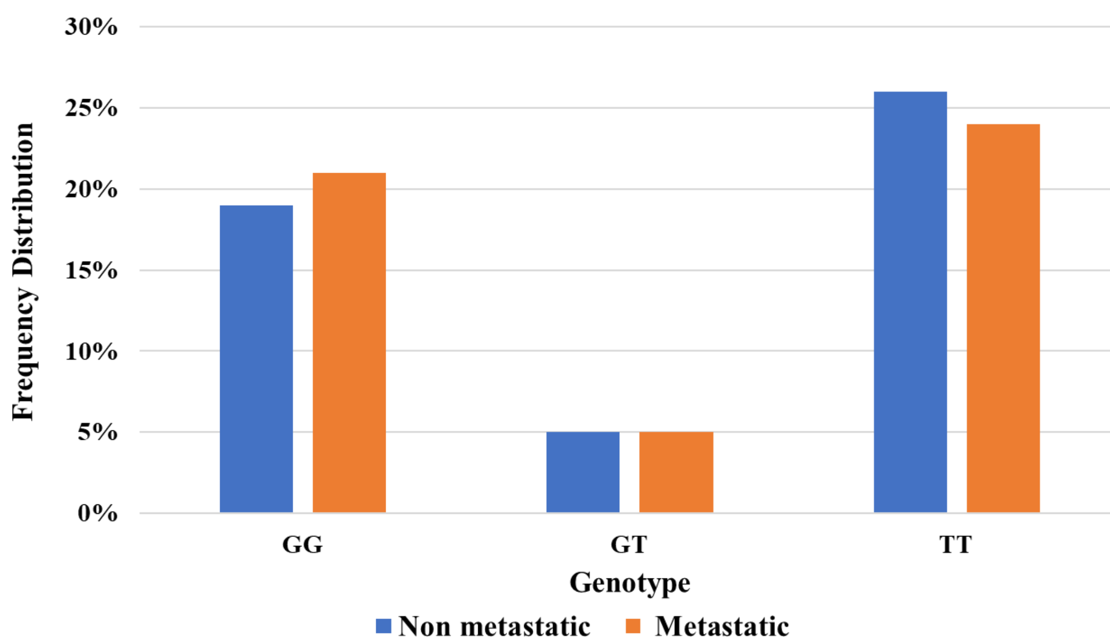


Figure 18. Genotyping analysis of KPCQ variants of both metastatic and non-metastatic breast cancer.

Based on cancer family history, genotype frequency distribution of C29Y variant was also determined. Genotype analysis revealed that all genotypes GG, GT, and TT in variant C29Y were statistically insignificant and could not correlated with breast cancer. Genotyping data of KPCQ variant with family history that is linked with breast cancer is given in table12.

Table 12. Genotype frequency distribution and association of KPCQ variant with a family history that is linked with breast cancer.

	Frequency Distribution		**Odd Ratio**		**Relative Risk**		**P-Value**
Genotypes	**No family history**	**Family history**	**Value**	**95%CI**	**Value**	**95%CI**	
GG	17%	23%	1.458	0.7308 to 2.848	1.195	0.8477 to 1.587	0.3769
GT	7%	3%	0.4109	0.1133 to 1.545	0.5876	0.2095 to 1.207	0.3311
TT	20%	30%	1.714	0.8990 to 3.239	1.286	0.9475 to 1.681	0.1412

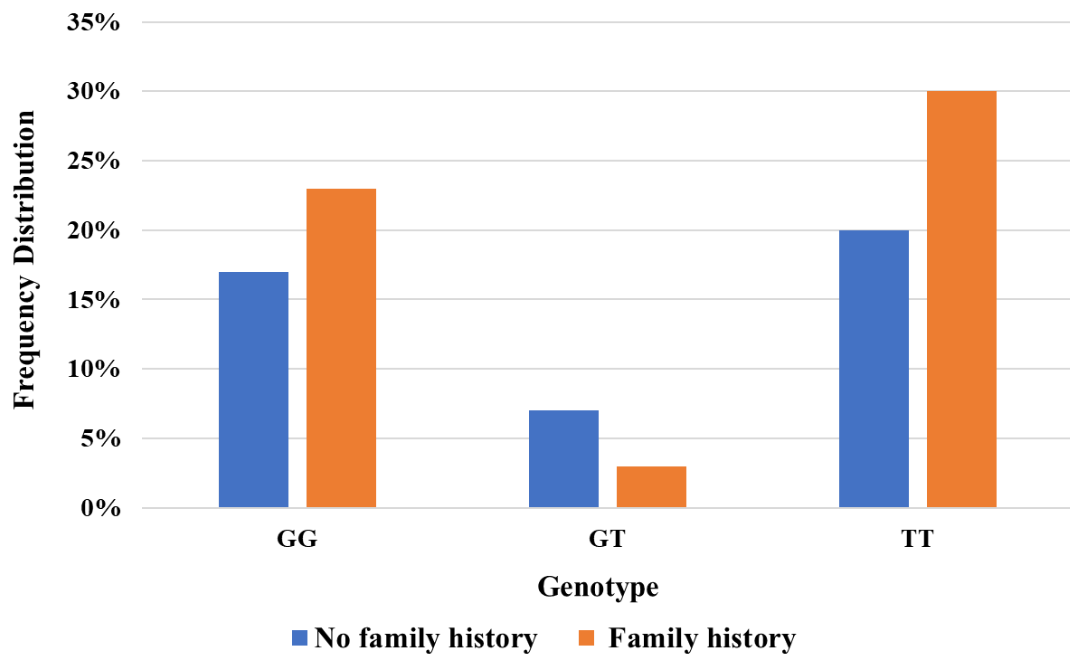


Figure 19. Genotype frequency distribution and association of KPCQ variant with a family history that is linked with breast cancer.

CHAPTER 5: DISCUSSION

The most lethal and multifactorial disease that causes the death of many women globally is breast cancer. This cancer is the second largest disease that is responsible for women's death (Sun et al., 2017). Because of its high incidence and death rate, it causes major health problems in women. According to GLOBOCAN 2018 data which is produced by the International Agency for Research on Cancer According to IARC, namely the GLOBOCAN program, in 2018 from 185 countries investigated that the breast cancer death rate is 6.9% and the morbidity rate is 11.7% (Kashyap et al., 2022). In developed countries, there are more incidence rates of breast cancer as compared to developing countries, indicating their connection with the globalization. Breast cancer is commonly referred to as a group of diseases due to the presence of numerous molecular biological subtypes that reflect the signs and symptoms of a disease and distinct molecular profile(Ghoncheh et al., 2016; Kashyap et al., 2022). Radiation therapy, surgical, and chemotherapy, endotherapy and immunotherapy are available options for treatment of breast cancer (Nounou et al., 2015; Sharma et al., 2010). Mortality and morbidity rate of breast cancer are not controlled easily, and they remain high, although there are some available options for treatment and this result in cause major problems throughout the world (Brand et al., 2000; Gawde et al., 2018). In past eras, the role of different genes including PTEN. TP53, STK11,ATM CDH1,PALB2, CHEK2, PALB2 and BRCA1 and BARC2 (Turnbull & Rahman, 2008) in breast cases had been studied and identified, but the result was not effective for early detection and better breast cancer treatment and incidence and death rates of breast cancer increased every year. For early prognosis and better treatment for breast cancer, there is a need for another novel gene. Novel KPC family isoforms are preferable as a potential biomarker, which basically reduce the incidence rates and recurrence of breast cancer among patients and providing effective results and better treatment options. Therefore, the main objective of this study is to determine the impact of missense SNPs on the structure and function of the KPCQ gene and identification of novel KPCQ missense variant's association with Breast cancer and its clinical features.

The KPC (protein kinase C) belongs to the serine/threonine kinases, and it is classified into three subfamilies (classical, novel, and atypical) based on their structure and ability to respond to calcium and DAG(Steinberg, 2008). The novel KPC isoform is KPCQ which is a protein kinase and it consists of 706 amino acids with a molecular weight of approximately 82 kDa (Gâ Baier et al., 1993). KPCQ is a Ca2+-independent but DAG-dependent, protein kinase whose structure contains various functional domains that are conserved among the novel KPC subfamily(Steinberg, 2008). KPCQ differs from other KPC isoforms because its physiological function is restricted to some types of cells such as T cells, thrombocytes, and skeletal muscle cells. KPCQ controls numerous important cellular processes such as survival, proliferation, and differentiation. KPCQ involves in the outbreak of pathology of various diseases, especially autoimmune disorders, and various cancer such as Leukaemia, Gastrointestinal Stromal Tumours, lung cancer, renal cancer cells, and ovarian cancer are also reported. In these types of cancers, the high expression of KPCQ leads to abnormal cell expansion, migration, and invasion which result in malignant phenotype. Therefore, KPCQ is considered a prognostic marker and a druggable target for the treatment of cancer(Nicolle et al., 2021).

The sequence of KPCQ was retrieved from ENSEMBLE and saved in FASTA format. Alpha-fold is used to predict the structure of KPCQ, further InterPro is used to validate the structure and domains of KPCQ, and the 3-D structure of this protein is visualized by PyMol(DeLano, 2002). This protein consists of 706 amino acids with four domains.

Non-synonymous single nucleotide polymorphisms(nsSNPs) are a type of genetic variant that are considered the cause of change in the structure, function, and expression of the protein (Vallejos-Vidal et al., 2020). The presence of these SNPs in those genes that help in the regulation of fundamental cellular pathways, can change their normal cellular functioning which can further enhance the risk of transformation and development of cancer(Deng et al., 2017). In addition, these nsSNPs can be of great impotence as they are regarded as important prognostic and diagnostic markers for disease because of their higher rate of prevalence. In this study, KPCQ SNPs data was collected from three genomic databases including Ensemble, COSMIC, and genomAD, and resulted in a total of 1166 of SNPs that consist of intronic, splice, UTRs, frameshift, nonsense, and non-synonymous variants. Out of these, there were 429 non-synonymous SNPs in that data that were selected, and then their percentage pathogenicity was

analysed with the help of six different bioinformatic tools which include SIFT, PolyPhen2, REVEL, MetaLR, Mutation Assessor, and CADD(Arshad et al., 2021), which predicted 19 nsSNPs that were present beyond the pathogenicity threshold of 70% and hence considered as completely deleterious. These non-synonymous SNPs were then further passed from stringent selection criteria as shown in table 5 to get the top nine pathogenic SNPs of KPCQ and one of them variant which is C29Y(rs1248923790) was selected for further in vivo and in silico analysis.

Structural and functional analysis of selected variant C29Y was performed through I-variant and PROJECT HOPE which means that mutation at this site results in a change the function of protein function and causes protein destabilization. DDG value for C29Y -0.01kcal/mol, which shows that the mutation results in a decrease the protein stability and destabilizes the protein. Project hope calculated that substitution of amino acid in C29Y was bigger in size than wild type and hydrophobicity is enhanced. This show that mutation at this site result in in bumps in protein structure and disrupt the protein activity.

FATHMM tool is used for the association of selected pathogenic SNPs C29Y with the cancer. The score for C29Y obtain through this tool is -1.18 which shows that it involves in protein deregulation and is responsible for causing cancer. Flexibility and stability analysis of the C29Y variant was performed through DynaMut. Vibrational entropy energy of both wild and variant protein structure was determined by the help of this tool which showed that alteration in protein flexibility caused by mutation. Value for C29Y variant is **$\Delta\Delta S_{Vib}$ ENCom: -1.098 kcal.mol^{-1}. K^{-1}** which results in a decrease the molecular flexibility and causes protein destabilization. The results obtained through all these tools indicate the protein structure effect which results in promoting cancer development.

Molecular dynamic simulations have enabled to get in-depth information on different molecular processes of bio-macromolecules which include their dynamics, folding, stability, and confirmational changes(Kumar & Purohit, 2012). MD-simulations for both wildtype and variant C29Y variant of KPCQ was performed. These simulations were performed on supercomputers that use the open-source software package named as GROMACS 2016(Van Der Spoel et al., 2005). Molecular Dynamic simulations evaluated that the mutation of KPCQ at position 29 from Cysteine (C) to Tyrosine (Y) may change the structure of the protein. The radius of gyration for wildtype and variant

structure was plotted, demonstrating the number of fluctuations in both structures. The graph of RG indicated as compared to wild type structure, the variant structure of C29Y variant show more fluctuation which indicates that the compactness of variant structure was decreased. The impact of C29Y variant on the number of hydrogen bonds between wild-type and variant protein structure was determined and plotted a graph which indicates that the line of both variant and wild type protein superimposed over each other which shows that there is no significant difference in both wildtype and variant structure. RMSD for variant structure was found to be higher as compared to that of the wildtype protein which result in decreased protein stability, whereas the RMSF graph was plotted for the carbon-backbone for each residue and compared to the wildtype structure. This analysis indicates that the variant protein significantly fluctuates from its mean position, which shows that the structure of protein expands over time. Lastly, SAS2 value of variant protein structures to some extent larger than that of the wild type of structure of protein. The variant structure of the C29Y variant is fluctuating from 2.4ns -4.9ns from wild type structure but after 5ns the wild type and both variant type proteins show an almost similar value of SAS2. This overall result indicates that the surface accessibility area is not much exposed and there is minimal effect on protein structure. All these results show a significant difference in RMSD, RMSF, Radius of gyration, and SASA analysis in both wildtype and variant proteins structure, which can be linked with the progression of breast cancer.

In-vivo analysis was performed to further validate the results of C29Y variant of KPCQ. For this purpose, first for DNA extraction, phenol-chloroform DNA extraction method used (Ghatak et al., 2013). For genotype analysis, TETRA ARMS-PCR was performed to confirm the role and association of the C29Y variant with Breast cancer. Genotype data analysis showed that genotype GG shows a more significant P-value and has a strong correlation with breast cancer and the distribution frequency of the GG genotype is 60% which is the highest. Genotype TT also shows significant value and association with breast cancer, but the heterozygous GT genotype was not showing statistically significant value and could correlate with breast cancer. Statical analysis for the distribution of allele frequency showed that both GG and TT genotypes were significantly correlated with the incidence of breast cancer.

There are various studies that demonstrated the role of KPCQ in different cancers. KPCQ is important for survival, growth, and modulation of leukaemia T cells and

cytotoxic lymphocytes. KPCQ was a downstream target, and it may enhance the Notch3-induced T-cell leukemogenesis progression of signalling of Notch3 and it may enhance the Notch3-induced T-cell leukemogenesis progression as reported by Felli et al.(Felli et al., 2005). In renal cancerous cells, englerin A and tonantzitlolone which are natural products, act as an anti-cancer agent in KPCQ dependent manner and help in KPCQ activation. Furthermore, these natural products induce an insulin resistant phenotype in cancers cell by activating KPCQ, through heat shock factor 1 activation (Sourbier et al., 2015; Sourbier et al., 2013). KPCQ was shown to highly expressed only in GIST but not in any other types of epithelial or mesenchymal cancer tissue which include non-GIST KIT-positive tumours as reported by Blay et al. and they evaluated the transcriptional profiling data of sarcoma(Blay et al., 2004). It has been proposed that KPCQ inhibit and phosphorylate the activity of GIV in Hella cells and Gα-interacting , as a result of this it also inhibit Akt activation, development of actin stress fibre and cell migration(López-Sánchez et al., 2013). Chu et al. reported an abnormal KPCQ expression in oral squamous cell carcinoma and their data demonstrated that increase KPCQ nuclear expression may be associated with a reoccurrence of the disease (Chu et al., 2012). The result from this study also proposed that SNPs in KPCQ might help in upregulating the expression of KPCQ in breast tissues which may lead to an increase in the chance of breast cancer.

CONCLUSION AND FUTURE PRESPECTIVE

SNPs act as an important prognostic marker for the occurrence and development of cancer which was proved by different studies. Therefore, there is a need to identify the new pathogenic variant that enhances the risk of cancer so that cancer at an early stage is diagnosed and treated by new diagnostic approaches. In this study, KPCQ variant C29Y was selected, and this variant was further analysed to validate their effect on function, structure, and stability of protein through in vitro analysis and these predict that protein structure alters and then leads to the progression of breast cancer. Further, genotyping analysis was performed to confirm the association of this variant with breast cancer and this analysis confirmed that the C29Y variant of KPCQ is involved in the progression of breast cancer. However, there is a need for more functional studies should be performed on this variant of KPCQ, so that this could be used as a potential

prognostic biomarker for better prognosis and early diagnosis as well as molecular therapeutic target for the treatment of breast cancer

REFRENCES

Ades, F., Tryfonidis, K., & Zardavas, D. (2017). The past and future of breast cancer treatment—from the papyrus to individualised treatment approaches. *ecancermedicalscience*, *11*.

Afzal, S., Hassan, M., Ullah, S., Abbas, H., Tawakkal, F., & Khan, M. A. (2022). Breast Cancer; Discovery of Novel Diagnostic Biomarkers, Drug Resistance, and Therapeutic Implications. *Frontiers in Molecular Biosciences*, *9*.

Altman, A., Isakov, N., & Baier, G. (2000). Protein kinase Cθ: a new essential superstar on the T-cell stage. *Immunology today*, *21*(11), 567-573.

Altman, A., Kaminski, S., Busuttil, V., Droin, N., Hu, J., Tadevosyan, Y., . . . Villalba, M. (2004). Positive feedback regulation of PLCγ1/Ca2+ signaling by PKCθ in restimulated T cells via a Tec kinase-dependent pathway. *European journal of immunology*, *34*(7), 2001-2011.

Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., . . . Aggarwal, B. B. (2008a). Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical research*, *25*, 2097-2116.

Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., . . . Aggarwal, B. B. (2008b). Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical research*, *25*(9), 2097-2116.

Appelbaum, F. R. (2001). Haematopoietic cell transplantation as immunotherapy.*Nature genetics*, *411*(6835), 385-389.

Arshad, S., Ishaque, I., Mumtaz, S., Rashid, M. U., & Malkani, N. (2021). In-silico analyses of nonsynonymous variants in the BRCA1 gene. *Biochemical Genetics*, *59*(6), 1506-1526.

Baier-Bitterlich, G., Uberall, F., Bauer, B., Fresser, F., Wachter, H., Grunicke, H., . . . Baier, G. (1996). Protein kinase C-theta isoenzyme selective stimulation of the transcription factor complex AP-1 in T lymphocytes. *Molecular cellular biology 16*(4), 1842-1850.

Baier, G. (2003). The PKC gene module: molecular biosystematics to resolve its T cell functions.*Immunological reviews*, *192*(1), 64-79.

Baier, G., Telford, D., Giampa, L., Coggeshall, K., Baier-Bitterlich, G., Isakov, N., & Altman, A. (1993). Molecular cloning and characterization of PKC theta, a novel member of the protein kinase C (PKC) gene family expressed predominantly in hematopoietic cells. *Journal of Biological Chemistry*, *268*(7), 4997-5004.

Baier, G., Telford, D., Giampa, L., Coggeshall, K., Baier-Bitterlich, G., Isakov, N., & Altman, A. (1993). Molecular cloning and characterization of PKC theta, a novel member of

the protein kinase C (PKC) gene family expressed predominantly in hematopoietic cells. *Journal of Biological Chemistry*, *268*(7), 4997-5004.

Bauer, B., Krumböck, N., Ghaffari-Tabrizi, N., Kampfer, S., Villunger, A., Wilda, M., . . . Überall, F. (2000a). T cell expressed PKCθ demonstrates cell-type selective function. *European Journal of Immunology*, *30*(12), 3645-3654.

Bauer, B., Krumböck, N., Ghaffari-Tabrizi, N., Kampfer, S., Villunger, A., Wilda, M., . . . Überall, F. (2000b). T cell expressed PKCθ demonstrates cell-type selective function. *European journal of immunology*, *30*(12), 3645-3654.

Belguise, K., Cherradi, S., Sarr, A., Boissière, F., Boulle, N., Simony-Lafontaine, J., . . . Chalbos, D. (2017). PKCθ-induced phosphorylations control the ability of Fra-1 to stimulate gene expression and cancer cell migration. *Cancer Letters*, *385*, 97-107.

Belguise, K., Milord, S., Galtier, F., Moquet-Torcy, G., Piechaczyk, M., & Chalbos, D. (2012a). The PKCθ pathway participates in the aberrant accumulation of Fra-1 protein in invasive ER-negative breast cancer cells. *Oncogene*, *31*(47), 4889-4897.

Belguise, K., Milord, S., Galtier, F., Moquet-Torcy, G., Piechaczyk, M., & Chalbos, D. (2012). The PKCθ pathway participates in the aberrant accumulation of Fra-1 protein in invasive ER-negative breast cancer cells. *Oncogene*, *31*(47), 4889-4897.

Belguise, K., & Sonenshein, G. E. (2007). PKCθ promotes c-Rel–driven mammary tumorigenesis in mice and humans by repressing estrogen receptor α synthesis. *The Journal of clinical investigation*, *117*(12), 4009-4021.

Bi, K., & Altman, A. (2001). Membrane lipid microdomains and the role of PKCθ in T cell activation. Seminars in Immunology,

Black, A. R., & Black, J. D. (2013). Protein kinase C signaling and cell cycle regulation. *Frontiers in immunology*, *3*, 423.

Blay, P., Astudillo, A., Buesa, J. M., Campo, E., Abad, M., García-García, J., . . . Losa, R. (2004). Protein kinase C θ is highly expressed in gastrointestinal stromal tumors but not in other mesenchymal neoplasias. *Clinical cancer research*, *10*(12), 4089-4095.

Boetes, C., Barentsz, J. O., Mus, R. D., Van Der Sluis, R., Van Erning, L., Hendriks, J., . . . Ruys, S. (1994). MR characterization of suspicious breast lesions with a gadolinium-enhanced TurboFLASH subtraction technique. *Radiology*, *193*(3), 777-781.

Boulding, T., McCuaig, R. D., Tan, A., Hardy, K., Wu, F., Dunn, J., . . . Bert, A. G. (2018). LSD1 activation promotes inducible EMT programs and modulates the tumour microenvironment in breast cancer. *Scientific reports*,*8*(1), 1-18.

Boze, A. (2002). WinSCP 2.0 Beta. *J Information Technology Libraries 21*(4), 190-192.

Brand, T. C., Sawyer, M. M., King, T. A., Bolton, J. S., & Fuhrman, G. M. (2000). Understanding patterns of failure in breast cancer treatment argues for a more thorough

investigation of axillary lymph nodes in node negative patients. *The American journal of surgery*, *180*(6), 424-427.

Bray, F., Ferlay, J., Laversanne, M., Brewster, D., Gombe Mbalawa, C., Kohler, B., . . . Antoni, S. (2015). Cancer I ncidence in F ive C ontinents: inclusion criteria, highlights from Volume X and the global status of cancer registration. *International Journal of Cancer*, *137*(9), 2060-2071.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P. T., Roydasgupta, R., Kuo, W.-L., . . . Ryder, T. (2006).Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell*, *10*(6), 529-541.

Chu, P.-Y., Hsu, N. C.-H., Tai, H.-C., Yeh, C.-M., Lin, S.-H., Hou, M.-F., & Yeh, K.-T. (2012). High nuclear protein kinase Cθ expression may correlate with disease recurrence and poor survival in oral squamous cell carcinoma. *Human pathology*, *43*(2), 276-281.

Collins, A., & Ke, X. (2012). Primer1: primer design web service for tetra-primer ARMS-PCR. *The Open Bioinformatics Journal*, *6*, 55-58.

Crabtree, G. R. (1999). Generic signals and specific outcomes: signaling through Ca2+, calcineurin, and NF-AT. *Cell*,*96*(5), 611-614.

Datta, R., Kojima, H., Yoshida, K., & Kufe, D. (1997). Caspase-3-mediated cleavage of protein kinase C θ in induction of apoptosis. *Journal of Biological Chemistry*, *272*(33), 20317-20320.

DeLano, W. L. (2002). Pymol: An open-source molecular graphics tool. *J CCP4 Newsl. Protein Crystallogr*, *40*(1), 82-92.

Deng, N., Zhou, H., Fan, H., & Yuan, Y. (2017). Single nucleotide polymorphisms and cancer susceptibility. *Oncotarget*, *8*(66), 110635.

Donnelly, R., Reed, M. J., Azhar, S., & Reaven, G. M. (1994). Expression of the major isoenzyme of protein kinase-C in skeletal muscle, nPKC theta, varies with muscle type and in response to fructose-induced insulin resistance. *Endocrinology*, *135*(6), 2369-2374.

Downs-Holmes, C., & Silverman, P. (2011). Breast cancer: overview & updates.*The Nurse Practitioner*, *36*(12), 20-26.

Eferl, R., & Wagner, E. F. (2003). AP-1: a double-edged sword in tumorigenesis. *Nature Reviews Cancer*,*3*(11), 859-868.

Felli, M. P., Vacca, A., Calce, A., Bellavia, D., Campese, A. F., Grillo, R., . . . Palermo, R. (2005). PKCθ mediates pre-TCR signaling and contributes to Notch3-induced T-cell leukemia. *Oncogene*, *24*(6), 992-1000.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., . . . Bray, F. (2015).Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, *136*(5), E359-E386.

Gao, L.-B., Rao, L., Wang, Y.-Y., Liang, W.-B., Li, C., Xue, H., . . . Lv, M.-L. (2009). The association of interleukin-16 polymorphisms with IL-16 serum levels and risk of colorectal and gastric cancer. *Carcinogenesis*, *30*(2), 295-299.

Garaude, J., Kaminski, S., Charni, S., Aguilò, J. I., Jacquet, C., Plays, M., . . . Anel, A. (2008). Impaired anti-leukemic immune response in PKCθ-deficient mice.*Molecular Immunology*, *45*(12), 3463-3469.

Gawde, K. A., Sau, S., Tatiparti, K., Kashaw, S. K., Mehrmohammadi, M., Azmi, A. S., & Iyer, A. K. (2018). Paclitaxel and di-fluorinated curcumin loaded in albumin nanoparticles for targeted synergistic combination therapy of ovarian and cervical cancers.*Colloids Surfaces B: Biointerfaces 167*, 8-19.

Ghatak, S., Muthukumaran, R. B., & Nachimuthu, S. K. (2013). A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. *Journal of biomolecular techniques: JBT*, *24*(4), 224.

Ghoncheh, M., Pournamdar, Z., & Salehiniya, H. (2016). Incidence and mortality and epidemiology of breast cancer in the world. *Asian Pacific journal of cancer prevention*, *17*(sup3), 43-46.

Grakoui, A., Bromley, S. K., Sumen, C., Davis, M. M., Shaw, A. S., Allen, P. M., & Dustin, M. L. (1999). The immunological synapse: a molecular machine controlling T cell activation. *Science*, *285*(5425), 221-227.

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., . . . Goodall, G. J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biology*, *10*(5), 593-601.

Guéry, L., & Hugues, S. (2015). Th17 cell plasticity and functions in cancer immunity. *BioMed research international*, *2015*.

Guruvayoorappan, C., Sakthivel, K. M., Padmavathi, G., Bakliwal, V., Monisha, J., & Kunnumakkara, A. B. (2015). Cancer preventive and therapeutic properties of fruits and vegetables: an overview. *Anticancer properties of fruits vegetables: A scientific review* 1-52.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646-674.

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., . . . Cardoso, F. J. R. D. P. (2019). Breast cancer (Primer) Nat. *66*.

Hayashi, K., & Altman, A. (2007). Protein kinase C theta (PKCθ): a key player in T cell life and death. *Pharmacological research*, *55*(6), 537-544.

Hayes, D. F., Isaacs, C., & Stearns, V. (2001). Prognostic factors in breast cancer: current and new predictors of metastasis. *Journal of mammary gland biology neoplasia 6*(4), 375-392.

Hinrichs, A. S., Karolchik, D., Baertsch, R., Barber, G. P., Bejerano, G., Clawson, H., . . . Hsu, F. (2006). The UCSC genome browser database: update 2006. *Nucleic acids research*, *34*(suppl_1), D590-D598.

Hu, Y., Sun, L., Zhang, Y., Lang, J., & Rao, J. (2020). Phosphoproteomics reveals key regulatory kinases and modulated pathways associated with ovarian cancer tumors. *OncoTargets and therapy*, *13*, 3595.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., . . . Down, T. (2002). The Ensembl genome database project. *Nucleic acids research*, *30*(1), 38-41.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., . . . Duquenne, L. (2009). InterPro: the integrative protein signature database.*Nucleic acids research*, *37*(suppl_1), D211-D215.

International agency for research on cancer. (2019). *World Health Organization*.

Isakov, N., & Altman, A. (2012). PKC-theta-mediated signal delivery from the TCR/CD28 surface receptors. *Frontiers in immunology*, *3*, 273.

Jackson, V. P. (1990). The role of US in breast imaging. *Radiology*, *177*(2), 305-311.

Kang, G. H., Kim, K.-M., Park, C. K., & Kang, D. Y. (2009). PKC-theta expression in Ewing sarcoma/primitive neuroectodermal tumour and malignant peripheral nerve sheath tumour. *Histopathology*, *55*(3), 368-369.

Kashyap, D., Pal, D., Sharma, R., Garg, V. K., Goel, N., Koundal, D., . . . Belay, A. (2022). Global increase in breast cancer incidence: risk factors and preventive measures. *BioMed research international*, *2022*.

Khan, K., Safi, S., Abbas, A., Badshah, Y., Dilshad, E., Rafiq, M., . . . Shabbir, M. (2020). Unravelling structure, localization, and genetic crosstalk of KLF3 in human breast cancer.*BioMed research international*, *2020*.

Khokhar, M. A., Ali, M. M., Liaqat, S., Moin, A., Sarwar, H. A., & Sarwar, M. Z. (2020). A review of access to cancer facilities in Punjab, Pakistan. *Cancer Reports*, *3*(3), e1245.

Kulig, W., Pasenkiewicz-Gierula, M., & Róg, T. (2015). Topologies, structures and parameter files for lipid simulations in GROMACS with the OPLS-aa force field: DPPC, POPC, DOPC, PEPC, and cholesterol. *Data in brief*, *5*, 333-336.

Kumar, A., & Purohit, R. (2012). Computational screening and molecular dynamics simulation of disease associated nsSNPs in CENP-E. *Mutation Research/Fundamental Molecular Mechanisms of Mutagenesis 738*, 28-37.

Li, Y., Hu, J., Vita, R., Sun, B., Tabata, H., & Altman, A. (2004). SPAK kinase is a substrate and target of PKCθ in T-cell receptor-induced AP-1 activation pathway. *The EMBO journal*,*23*(5), 1112-1122.

Lim, P. S., Sutton, C. R., & Rao, S. (2015). Protein kinase C in the immune system: from signalling to chromatin regulation. *Immunology today*, *146*(4), 508-522.

Liu, F.-C., Lin, H.-T., Kuo, C.-F., See, L.-C., Chiou, M.-J., & Yu, H.-P. (2017). Epidemiology and survival outcome of breast cancer in a nationwide study. *Oncotarget*, *8*(10), 16939.

Liu, K.-Q., Bunnell, S. C., Gurniak, C. B., & Berg, L. J. (1998). T cell receptor–initiated calcium release is uncoupled from capacitative calcium entry in Itk-deficient T cells. *The Journal of experimental medicine*, *187*(10), 1721-1727.

Livaudais-Toman, J., Karliner, L. S., Tice, J. A., Kerlikowske, K., Gregorich, S., Pérez-Stable, E. J., . . . Kaplan, C. P. (2015). Impact of a primary care based intervention on breast cancer knowledge, risk perception and concern: a randomized, controlled trial. *The Breast*, *24*(6), 758-766.

López-Sánchez, I., Garcia-Marcos, M., Mittal, Y., Aznar, N., Farquhar, M. G., & Ghosh, P. (2013). Protein kinase C-theta (PKCθ) phosphorylates and inhibits the guanine exchange factor, GIV/Girdin. *Proceedings of the National Academy of Sciences 110*(14), 5510-5515.

Ma, D., Zhong, S., Liu, X., Mai, H., Mai, G., Xu, C., & Zhou, F. (2016). CD3D and KPCQ work together to discriminate between B-cell and T-cell acute lymphoblastic leukemia. *Computers in Biology Medicine 77*, 16-22.

Mackay, H. J., & Twelves, C. J. (2007). Targeting the protein kinase C family: are we there yet? *Nature Reviews Cancer*, *7*(7), 554-562.

Manicassamy, S., Sadim, M., Richard, D. Y., & Sun, Z. (2006). Differential roles of PKC-θ in the regulation of intracellular calcium concentration in primary T cells. *Journal of Molecular Biology*, *355*(3), 347-359.

Marrocco, V., Fiore, P., Madaro, L., Crupi, A., Lozanoska-Ochser, B., & Bouché, M. (2014). Targeting PKCθ in skeletal muscle and muscle diseases: good or bad? *Biochemical Society transactions*, *42*(6), 1550-1555.

Mavrevski, R., Traykov, M., Trenchev, I., & Trencheva, M. (2018). Approaches to modeling of biological experimental data with GraphPad Prism software. *WSEAS Trans Syst Control*, *13*(1), 242-247.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., . . . Cunningham, F. (2016). The ensembl variant effect predictor. *Genome biology*, *17*(1), 1-14.

Meller, N., Altman, A., & Isakov, N. (1998). New perspectives on PKCθ, a member of the novel subfamily of protein kinase C. *Stem cells*, *16*(3), 178-192.

Mena, S., Ortega, A., & Estrela, J. M. (2009). Oxidative stress in environmental-induced carcinogenesis. *Mutation Research/Genetic Toxicology Environmental Mutagenesis 674*(1-2), 36-44.

Miceli, M. C., Moran, M., Chung, C. D., Patel, V. P., Low, T., & Zinnanti, W. (2001). Co-stimulation and counter-stimulation: lipid raft clustering controls TCR signaling and functional outcomes. Seminars in Immunology,

Mochly-Rosen, D., Das, K., & Grimes, K. V. (2012). Protein kinase C, an elusive therapeutic target? *Nature reviews Drug discovery*, *11*(12), 937-957.

Monks, C. R., Kupfer, H., Tamir, I., Barlow, A., & Kupfer, A. (1997). Selective modulation of protein kinase C-θ during T-cell activation. *Nature genetics*, *385*(6611), 83-86.

Mustafa, H. A., Albkrye, A. M. S., AbdAlla, B. M., Khair, M. A. M., Abdelwahid, N., & Elnasri, H. A. (2020). Computational determination of human PPARG gene: SNPs and prediction of their effect on protein functions of diabetic patients. *Clinical Translational Medicine 9*, 1-10.

Nalejska, E., Mączyńska, E., & Lewandowska, M. A. (2014). Prognostic and predictive biomarkers: tools in personalized oncology. *Molecular diagnosis therapy 18*(3), 273-284.

Nangalia, J., Massie, C., Baxter, E. J., Nice, F. L., Gundem, G., Wedge, D., . . . Kent, D. G. (2013). The genomic landscape of myeloproliferative neoplasms: somatic calr mutations in the majority of JAK2-wildtype patients. In: American Society of Hematology Washington, DC.

Newton, A. C. (2018). Protein kinase C: perfectly balanced. *Critical reviews in biochemistry molecular biology 53*(2), 208-230.

Nicolle, A., Zhang, Y., & Belguise, K. (2021). The emerging function of pkctheta in cancer. *Biomolecules*, *11*(2), 221.

NIH.CANCERCLASSIFICATION. https://training.seer.cancer.gov/disease/categories/classification.html

Nounou, M. I., ElAmrawy, F., Ahmed, N., Abdelraouf, K., Goda, S., & Syed-Sha-Qhattal, H. (2015). Breast cancer: conventional diagnosis and treatment modalities and recent patents and technologies. *Breast cancer: basic clinical research 9*, BCBCR. S29420.

Ou, W., Zhu, M., Demetri, G. D., Fletcher, C. D., & Fletcher, J. A. (2008). Protein kinase C-θ regulates KIT expression and proliferation in gastrointestinal stromal tumors. *Oncogene*, *27*(42), 5624-5634.

Pfeifhofer, C., Kofler, K., Gruber, T., Ghaffari Tabrizi, N., Lutz, C., Maly, K., . . . Baier, G. (2003). Protein kinase C θ affects Ca2+ mobilization and NFAT activation in primary mouse T cells. *The Journal of experimental medicine*, *197*(11), 1525-1535.

Piñeros, M., Mery, L., Soerjomataram, I., Bray, F., & Steliarova-Foucher, E. (2021). Scaling up the surveillance of childhood cancer: a global roadmap. *JNCI: Journal of the National Cancer Institute 113*(1), 9-15.

Rao, A., Luo, C., & Hogan, P. G. (1997). Transcription factors of the NFAT family: regulation and function. *Annual Review of Immunology*, *15*(1), 707-747.

Rawlings, D. J., Sommer, K., & Moreno-García, M. E. (2006). The CARMA1 signalosome links the signalling machinery of adaptive and innate immunity in lymphocytes. *Nature Reviews Immunology*, *6*(11), 799-812.

Reyland, M. E. (2009). Protein kinase C isoforms: multi-functional regulators of cell life and death. *Frontiers in bioscience*, *14*, 2386.

Risch, N. J. (2000). Searching for genetic determinants in the new millennium. *Nature genetics*, *405*(6788), 847-856.

Rogers, M. F., Shihab, H. A., Mort, M., Cooper, D. N., Gaunt, T. R., & Campbell, C. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics*, *34*(3), 511-513.

Rosse, C., Linch, M., Kermorgant, S., Cameron, A. J., Boeckeler, K., & Parker, P. J. (2010). PKC and the control of localized signal dynamics. *Nature reviews Molecular cell biology*, *11*(2), 103-112.

Ruddon, R. W. (2007). *Cancer biology*. Oxford University Press.

Ryu, J. M., Lee, M. Y., Yun, S. P., & Han, H. J. (2010). High glucose regulates cyclin D1/E of human mesenchymal stem cells through TGF-β1 expression via Ca2+/PKC/MAPKs and PI3K/Akt/mTOR signal pathways. *Journal of cellular physiology*, *224*(1), 59-70.

Sabapathy, K., Kallunki, T., David, J.-P., Graef, I., Karin, M., & Wagner, E. F. (2001). c-Jun NH2-terminal kinase (JNK) 1 and JNK2 have similar and stage-dependent roles in regulating T cell apoptosis and proliferation. *The Journal of experimental medicine*, *193*(3), 317-328.

Saeed, S., Asim, M., & Sohail, M. M. (2021). Fears and barriers: problems in breast cancer diagnosis and treatment in Pakistan. *BMC Women's Health*, *21*(1), 1-10.

Schaeffer, E. M., Broussard, C., Debnath, J., Anderson, S., McVicar, D. W., & Schwartzberg, P. L. (2000). Tec family kinases modulate thresholds for thymocyte development and selection. *The Journal of experimental medicine*,*192*(7), 987-1000.

Schaeffer, E. M., Debnath, J., Yap, G., McVicar, D., Liao, X. C., Littman, D. R., . . . Schwartzberg, P. L. (1999). Requirement for Tec kinases Rlk and Itk in T cell receptor signaling and immunity. *Science*, *284*(5414), 638-641.

Schor, A. M., & Schor, S. L. (1983). Tumour angiogenesis. *The Journal of pathology*, *141*(3), 385-413.

Schultz, A., Jönsson, J.-I., & Larsson, C. (2003). The regulatory domain of protein kinase Cθ localises to the Golgi complex and induces apoptosis in neuroblastoma and Jurkat cells. *Cell Death Differentiation 10*(6), 662-675.

Sharma, G. N., Dave, R., Sanadya, J., Sharma, P., & Sharma, K. (2010). Various types and management of breast cancer: an overview. *Journal of advanced pharmaceutical technology research 1*(2), 109.

So, T., & Croft, M. (2012). Regulation of the PKCθ-NF-κB axis in T lymphocytes by the tumor necrosis factor receptor family member OX40. *Frontiers in immunology*, *3*, 133.

Sourbier, C., Scroggins, B. T., Mannes, P. Z., Liao, P.-J., Siems, K., Wolf, D., . . . Neckers, L. (2015). Tonantzitlolone cytotoxicity toward renal cancer cells is PKCθ-and HSF1-dependent. *Oncotarget*, *6*(30), 29963.

Sourbier, C., Scroggins, B. T., Ratnayake, R., Prince, T. L., Lee, S., Lee, M.-J., . . . Beutler, J. A. (2013). Englerin A stimulates PKCθ to inhibit insulin signaling and to simultaneously activate HSF1: pharmacologically induced synthetic lethality. *Cancer Cell*, *23*(2), 228-237.

Soyombo, A. A., Wu, Y., Kolski, L., Rios, J. J., Rakheja, D., Chen, A., . . . Ross, T. S. (2013). Analysis of induced pluripotent stem cells from a BRCA1 variant family. *Stem Cell Reports*,*1*(4), 336-349.

Srinivasan, S., Clements, J. A., & Batra, J. (2016). Single nucleotide polymorphisms in clinics: fantasy or reality for cancer? *Critical reviews in clinical laboratory sciences*, *53*(1), 29-39.

Steinberg, S. F. (2008). Structural basis of protein kinase C isoform function. *Physiological Reviews*, *88*(4), 1341-1378.

Sun, Y.-S., Zhao, Z., Yang, Z.-N., Xu, F., Lu, H.-J., Zhu, Z.-Y., . . . Zhu, H.-P. (2017). Risk factors and preventions of breast cancer. *International journal of biological sciences*, *13*(11), 1387.

Sun, Z., Arendt, C. W., Ellmeier, W., Schaeffer, E. M., Sunshine, M. J., Gandhi, L., . . . Schwartzberg, P. L. (2000). PKC-θ is required for TCR-induced NF-κB activation in mature but not immature T lymphocytes. *Nature genetics*, *404*(6776), 402-407.

Sutcliffe, E. L., Bunting, K. L., He, Y. Q., Li, J., Phetsouphanh, C., Seddiki, N., . . . Kelleher, A. D. (2011). Chromatin-associated protein kinase C-θ regulates an inducible gene

expression program and microRNAs in human T lymphocytes. *Molecular Cell*, *41*(6), 704-719.

Tan, S.-L., Zhao, J., Bi, C., Chen, X. C., Hepburn, D. L., Wang, J., . . . Na, S. (2006). Resistance to experimental autoimmune encephalomyelitis and impaired IL-17 production in protein kinase Cθ-deficient mice. *The Journal of Immunology*, *176*(5), 2872-2879.

Tang, S., Morgan, K. G., Parker, C., & Ware, J. A. (1997). Requirement for protein kinase C θ for cell cycle progression and formation of actin stress fibers and filopodia in vascular endothelial cells. *Journal of Biological Chemistry*, *272*(45), 28704-28711.

Thamaraiselvan, R., Peramaiyan, R., Natarajan, N., Balasubramanian, M., & Nishigaki, I. (2014). The epidemiology of human breast cancer. *International Journal of Medical Health Sciences 3*(1), 62-71.

Tomlinson, M. G., Kane, L. P., Su, J., Kadlecek, T. A., Mollenauer, M. N., & Weiss, A. (2004). Expression and function of Tec, Itk, and Btk in lymphocytes: evidence for a unique role for Tec. *Molecular and Cellular Biology*,*24*(6), 2455-2466.

Turnbull, C., & Rahman, N. (2008). Genetic predisposition to breast cancer: past, present, and future. *Annu. Rev. Genomics Hum. Genet.*, *9*, 321-345.

Vacca, A., Felli, M. P., Palermo, R., Di Mario, G., Calce, A., Di Giovine, M., . . . Screpanti, I. (2006). Notch3 and pre-TCR interaction unveils distinct NF-κB pathways in T-cell development and leukemia. *The EMBO journal*, *25*(5), 1000-1008.

Valenzuela, J. O., Iclozan, C., Hossain, M. S., Prlic, M., Hopewell, E., Bronk, C. C., . . . Blazar, B. R. (2009). PKCθ is required for alloreactivity and GVHD but not for immune responses toward leukemia and infection in mice. *The Journal of clinical investigation*, *119*(12), 3774-3786.

Vallejos-Vidal, E., Reyes-Cerpa, S., Rivas-Pardo, J. A., Maisey, K., Yáñez, J. M., Valenzuela, H., . . . Sandino, A. M. (2020). Single-nucleotide polymorphisms (SNP) mining and their effect on the tridimensional protein structure prediction in a set of immunity-related expressed sequence tags (EST) in Atlantic salmon (Salmo salar). *Frontiers in genetics*, *10*, 1406.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: fast, flexible, and free. *Journal of computational chemistry*, *26*(16), 1701-1718.

Villalba, M., Bushway, P., & Altman, A. (2001). Protein kinase C-θ mediates a selective T cell survival signal via phosphorylation of BAD. *The Journal of Immunology*, *166*(10), 5955-5963.

Visnjic, D., & Banfic, H. (2007). Nuclear phospholipid signaling: phosphatidylinositol-specific phospholipase C and phosphoinositide 3-kinase. *Pflügers Archiv-European Journal of Physiology*, *455*(1), 19-30.

Wang, L., Zheng, J., Luo, Y., Xu, T., Zhang, Q., Zhang, L., . . . Zhang, C. (2013). Construction of a genomewide RNA i variant library in rice. *Plant Biotechnology Journal*, *11*(8), 997-1005.

Wang, X., Belguise, K., Kersual, N., Kirsch, K. H., Mineva, N. D., Galtier, F., . . . Sonenshein, G. E. (2007). Oestrogen signalling inhibits invasive phenotype by repressing RelB and its target BCL2. *Nature Cell Biology*, *9*(4), 470-478.

Wang, Y., Gao, N., Feng, Y., Cai, M., Li, Y., Xu, X., . . . Yao, D. (2021). Protein kinase C theta (KPCQ) affects nerve degeneration and regeneration through the c-fos and c-jun pathways in injured rat sciatic nerves. *Experimental Neurology*, *346*, 113843.

Wapnir, I. L., Price, K. N., Anderson, S. J., Robidoux, A., Martín, M., Nortier, J. W., . . . Baena-Cañada, J. M. (2018). Efficacy of chemotherapy for ER-negative and ER-positive isolated locoregional recurrence of breast cancer: final analysis of the CALOR trial. *Journal of clinical oncology*, *36*(11), 1073.

Warner, E., Plewes, D. B., Hill, K. A., Causer, P. A., Zubovits, J. T., Jong, R. A., . . . Messner, S. J. (2004). Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA: Journal of the American Medical Association*, *292*(11), 1317-1325.

Webb, B. L., Hirst, S. J., & Giembycz, M. A. (2000). Protein kinase C isoenzymes: a review of their structure, regulation and role in regulating airways smooth muscle tone and mitogenesis. *British Journal of Pharmacology*, *130*(7), 1433.

Weigel, M. T., & Dowsett, M. (2010). Current and emerging biomarkers in breast cancer: prognosis and prediction. *Endocrine-related cancer*, *17*(4), R245-R262.

Whiteman, D. C., & Wilson, L. F. (2016). The fractions of cancer attributable to modifiable factors: A global review. *Cancer Epidemiology, Biomarkers & Prevention*, *44*, 203-221.

Wu, H.-J., & Chu, P.-Y. (2021). Recent discoveries of macromolecule-and cell-based biomarkers and therapeutic implications in breast cancer. *International Journal of Molecular Sciences*, *22*(2), 636.

Wu, H.-J., & Chu, P.-Y. (2021). Recent discoveries of macromolecule-and cell-based biomarkers and therapeutic implications in breast cancer. *International Journal of Molecular Sciences*, *22*(2), 636.

Yamada, K., Avignon, A., Standaert, M. L., Cooper, D. R., Spencer, B., & Farese, R. V. (1995). Effects of insulin on the translocation of protein kinase C-θ and other protein kinase C isoforms in rat skeletal muscles. *Biochemical Journal*, *308*(1), 177-180.

Yanuar, A., Chavarina, K. K., & Syahdi, R. R. (2018). Molecular Dynamic Simulation Analysis on Marine Fungi Compounds Against EGFR and VEGFR-2 Inhibitory Activity in Non-Small Cell Lung Cancer. *Journal of Young Pharmacists*, *10*(2s), S25.

Zafar, A., Hardy, K., Wu, F., Li, J., & Rao, S. (2015).The role of protein kinase-C theta in control of epithelial to mesenchymal transition and cancer stem cell formation. *Genomics Data*,*3*, 28-32.

Zafar, A., Wu, F., Hardy, K., Li, J., Tu, W. J., McCuaig, R., . . . Gregory, P. A. (2014). Chromatinized protein kinase C-θ directly regulates inducible genes in epithelial to mesenchymal transition and breast cancer stem cells. *Molecular and Cellular Biology*, *34*(16), 2961-2980.

Zhang, E. Y., Kong, K.-F., & Altman, A. (2013). The yin and yang of protein kinase C-theta (PKCθ): a novel drug target for selective immunosuppression. *Advances in pharmacology*, *66*, 267-312.

Zhu, L., Li, Y., Xie, X., Zhou, X., Gu, M., Jie, Z., . . . Cheng, X. (2019). TBKBP1 and TBK1 form a growth factor signalling axis mediating immunosuppression and tumourigenesis. *Nature Cell Biology*, *21*(12), 1604-1614.

Zhuang, J., Dai, S., Zhang, L., Gao, P., Han, Y., Tian, G., . . . Kui, L. (2020). Identifying breast cancer-induced gene perturbations and its application in guiding drug repurposing. *Current Bioinformatics*, *15*(9), 1075-1089.

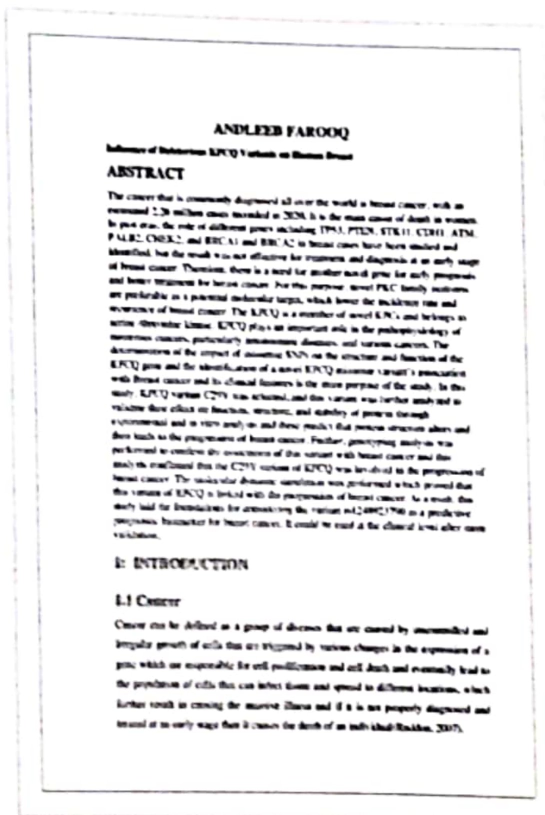
Dr. Yasmin B... Shah
Assistant Pr...ssor
Atta-ur-Rahman School of
Applied Biosciences (ASAB
NUST Isla...bad

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:	Andleeb Farooq
Assignment title:	Influence of Deleterious KPCQ Variants on Human Breast
Submission title:	Influence of Deleterious KPCQ Variants on Human Breast
File name:	Andleeb_Farooq.docx
File size:	154.45K
Page count:	41
Word count:	14,452
Character count:	78,318
Submission date:	02-Jun-2023 03:29AM (UTC-0700)
Submission ID:	2107388257



ANDLEEB FAROOQ

ABSTRACT

1. INTRODUCTION

1.1 Cancer

Copyright 2023 Turnitin. All rights reserved.

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

Influence of Deleterious KPCQ Variants on Human Breast

ORIGINALITY REPORT

9%	%	9%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Amandine Nicolle, Ye Zhang, Karine Belguise. "The Emerging Function of PKCtheta in Cancer", Biomolecules, 2021 Publication	4%
2	Charles E. Chalfant, Theodore P. Ciaraldi, James E. Watson, Svetlana Nikoulina, Robert R. Henry, Denise R. Cooper. " Protein Kinase Cθ Expression Is Increased upon Differentiation of Human Skeletal Muscle Cells: Dysregulation in Type 2 Diabetic Patients and a Possible Role for Protein Kinase Cθ in Insulin-Stimulated Glycogen Synthase Activity ", Endocrinology, 2000 Publication	1%
3	Hania Shah, Khushbukhat Khan, Naila Khan, Yasmin Badshah, Naeem Mahmood Ashraf, Maria Shabbir. "Impact of deleterious missense PRKCI variants on structural and functional dynamics of protein", Scientific Reports, 2022 Publication	1%

4 "Molecular Pathology of Breast Cancer", Springer Science and Business Media LLC, 2016
Publication
<1%

5 Khushbukhat Khan, Hania Shah, Areeba Rehman, Yasmin Badshah, Naeem M Ashraf, Maria Shabbir. "Influence of PRKCE non-synonymous variants on protein dynamics and functionality", Human Molecular Genetics, 2022
Publication
<1%

6 Black, Adrian R., and Jennifer D. Black. "Protein kinase C signaling and cell cycle regulation", Frontiers in Immunology, 2013.
Publication
<1%

7 "Next Generation Point-of-care Biomedical Sensors Technologies for Cancer Diagnosis", Springer Science and Business Media LLC, 2017
Publication
<1%

8 Santhakumar Manicassamy, Maureen Sadim, Richard D. Ye, Zuoming Sun. "Differential Roles of PKC-θ in the Regulation of Intracellular Calcium Concentration in Primary T Cells", Journal of Molecular Biology, 2006
Publication
<1%

9	Hussein N Rubaiy, Melanie J Ludlow, Karsten Siems, Katherine Norman et al. "Tonantzitlolone is a Nanomolar Potency Activator of TRPC1/4/5 Channels", British Journal of Pharmacology, 2018 Publication	<1 %
10	Patrick, Graham L.. "An Introduction to Medicinal Chemistry", An Introduction to Medicinal Chemistry, 2023 Publication	<1 %
11	Ndumiso N. Mhlongo, Mahmoud E. S. Soliman. "Single H5N1 influenza A neuraminidase mutation develops resistance to oseltamivir due to distorted conformational and drug binding landscape: multiple molecular dynamics analyses", RSC Advances, 2015 Publication	<1 %
12	Amnon Altman, Noah Isakov, Gottfried Baier. "Protein kinase Cθ: a new essential superstar on the T-cell stage", Immunology Today, 2000 Publication	<1 %
13	"Abstracts of the IDF Congress in Paris 2003", Diabetologia, 2003 Publication	<1 %
14	"Physicians Poster Sessions", Bone Marrow Transplantation, 2013. Publication	<1 %

15	Garaude, J.. "Impaired anti-leukemic immune response in PKC@q-deficient mice", Molecular Immunology, 200807 Publication	<1 %
16	Ramya Viswanathan, Safoura Sameni, Gavin Yong-Quan Ng, Wilner Martinez-Lopez, Prakash Hande. "Telomerase inhibition by MST-312 sensitises breast cancer cells to anti-cancer properties of Plumbagin", Cold Spring Harbor Laboratory, 2023 Publication	<1 %
17	K.N. ArulJothi, B. Suruthi Abirami, Sivaraj Irusappan, Amarnath Gautami, C. Swathine, A. Devi. "L55M and Q192R polymorphism of Paraoxonase gene and the risk of myocardial infarction in South Indian Tamil population", Meta Gene, 2018 Publication	<1 %
18	Prabhudutta Mamidi, Mahendra Gaur, Baijayantimala Mishra, Bharat Bhusan Subudhi, Sutapa Rath, Monalisa Mohanty. "The uniqueness of the Brazilian Yellow Fever Virus. Are the vaccines less effective against it?", Cold Spring Harbor Laboratory, 2023 Publication	<1 %
19	Lubna Al-Khalili, Thais de Castro Barbosa, Jörgen Östling, Julie Massart et al. "Proteasome inhibition in skeletal muscle cells	<1 %

unmasks metabolic derangements in type 2 diabetes", American Journal of Physiology-Cell Physiology, 2014
Publication

20	de Oliveira, T.C.. "Interfacial activation of snake venom phospholipases A"2 (svPLA"2) probed by molecular dynamics simulations", Journal of Molecular Structure: THEOCHEM, 20070930 Publication	<1%
21	E C Rhodes. "Effects of one year of resistance training on the relation between muscular strength and bone density in elderly women", British Journal of Sports Medicine, 2/1/2000 Publication	<1%
22	Apple, K.A.. "Differential Effects of Protein Kinase C Isoform Activation in Endothelin-Mediated Myocyte Contractile Dysfunction With Cardioplegic Arrest and Reperfusion", The Annals of Thoracic Surgery, 200608 Publication	<1%
23	Audrey E. Brown, Jane Palsgaard, Rehannah Borup, Peter Avery, David A. Gunn, Pierre De Meyts, Stephen J. Yeaman, Mark Walker. "p38 MAPK activation upregulates proinflammatory pathways in skeletal muscle cells from insulin-resistant type 2 diabetic	<1%

patients", American Journal of Physiology-Endocrinology and Metabolism, 2015
Publication

---

24 Avijeet Kulshrestha, Sudeep N Punnathanam, Rahul Roy, K Ganapathy Ayappa. "Cholesterol catalyzes unfolding in membrane inserted motifs of the pore forming protein cytolysin A", Cold Spring Harbor Laboratory, 2023
Publication
<1%

---

25 Kumar, Ambuj, and Rituraj Purohit. "Computational screening and molecular dynamics simulation of disease associated nsSNPs in CENP-E", Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2012.
Publication
<1%

---

26 Le Shu, Ao Chen, Linrui Li, Lun Yao, Yiduo He, Jianbo Xu, Wei Gu, Qiang Li, Kun Wang, Tongcun Zhang, Guoquan Liu. "NRG1 regulates Fra-1 transcription and metastasis of triple-negative breast cancer cells via the c-Myc ubiquitination as manipulated by ERK1/2-mediated Fbxw7 phosphorylation", Oncogene, 2022
Publication
<1%

---

27 Wisam Hoidy. "Development of a T-ARMS-PCR Assay for Detecting Genetic Polymorphism in the Catalase (rs7943316) Gene in the Iraqi
<1%

Population with Breast Cancer", Research Square Platform LLC, 2023
Publication

28	I. Lopez-Sanchez, M. Garcia-Marcos, Y. Mittal, N. Aznar, M. G. Farquhar, P. Ghosh. "Protein kinase C-theta (PKC) phosphorylates and inhibits the guanine exchange factor, GIV/Girdin", Proceedings of the National Academy of Sciences, 2013 Publication	<1%

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		