CAPITAL UNIVERSITY OF SCIENCE AND TECHNOLOGY, ISLAMABAD

Insilico Analysis of Moonlight Proteins Associated with Breast Cancer

by

Zakia Batool

A thesis submitted in partial fulfillment for the degree of Master of Science

in the

Faculty of Health and Life Sciences

Department of Bioinformatics and Biosciences

2022

I am dedicating this thesis to my Mother and Father Whose affection, Love encouragement and prays of day and night makes me able to get such honor, and to my teachers whose encouragement has always been my source of motivation.

CERTIFICATE OF APPROVAL

Insilico Analysis of Moonlight Proteins Associated with Breast Cancer

by

Zakia Batool

(MBS203007)

THESIS EXAMINING COMMITTEE

S. No.	Examiner	Name	Organization
(a)	External Examiner	Dr. Mazhar Qayyum	PMAS Arid, Rawalpindi
(b)	Internal Examiner	Dr. Erum Dilshad	CUST, Islamabad
(c)	Supervisor	Dr. Sahar Fazal	CUST, Islamabad

Dr. Sahar Fazal
Thesis Supervisor
November, 2022

Dr. Syeda Marriam Bakhtiar
Head
Dept. of Bioinfo. and Biosciences
November, 2022

Dr. Sahar Fazal
Dean
Faculty of Health and Life Sciences
November, 2022

Author's Declaration

I, **Zakia Batool** hereby state that my MS thesis titled "**Insilico Analysis of Moonlight Proteins Associated with Breast Cancer**" is my own work and has not been submitted previously by me for taking any degree from Capital University of Science and Technology, Islamabad or anywhere else in the country/abroad.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw my MS Degree.

(Zakia Batool)
Registration No: MBS203007

Plagiarism Undertaking

I solemnly declare that research work presented in this thesis titled "**Insilico Analysis of Moonlight Proteins Associated with Breast Cancer**" is solely my research work with no significant contribution from any other person. Small contribution/help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Capital University of Science and Technology towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS Degree, the University reserves the right to withdraw/revoke my MS degree and that HEC and the University have the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized work.

(Zakia Batool)
Registration No: MBS203007

Acknowledgement

All praise should be directed toward the Almighty Allah and the Prophet Muhammad (SAW). I want to extend my sincere gratitude to my family for their kind support throughout my pursuit of the MS degree. I sincerely appreciate having Dr. Sahar Fazal as my supervisor and like to thank Dr. Sahar Fazal for her, thoughtful collaboration, encouragement advice and guidance with computational methods. Lastly, my family deserves endless gratitude. Thank to everybody.

(Zakia Batool)

Abstract

A "Moonlighting protein" is a single protein that can carry out various tasks without the use of gene fusions, multiple RNA splicing, or promiscuous enzyme proteolytic activity. Similar to people who work many jobs while, moonlighting proteins are multitasking polypeptide chains. These proteins have so far been found in humans, yeast, worms, bacteria, plants, viruses, archaea, and other novel creatures. Moonlight proteins related to Breast cancer were retrieved by using COREMINE , PubMed ,OMIM, gene bank SwissProt , and multiple sources of information generated .Three different types of textual data were extracted for each protein from UniProt KB. Retrieved list of proteins from Corermine and UniProt were manually compared to cross verify the protein among this list. The list retrieved from UniProt KB was refined by removing the accession no, gene name and only the UniProtKB Ids and proteins name were saved. DextMP was used to generate the moonlight proteins by using the list prepared from UniProtKB. Manual verification of predicted proteins using UniProtKB's functional description and quick searches of publication titles was performed. Cross validiation from MoonDB was also performed. DAVID was used to carry out the functional annotation and enrichment analysis of predicted moonlight proteins. DextMP predicted 84 proteins as moonlight proteins out of the list of 2246 proteins prepared and verified from coremine, UniProt and literature. Out of 84 moonlight proteins, only 58 were present in MoonDB. The proteins present in Moon DB were verified and remaining 27 proteins were catergorized as predicted proteins. Functional annotation generated 5 clusters of 55 proteins .Cross talk was performed using PathwaxII tool. The proteins were mapped on five classes of pathways with significant crosstalk. Pathways were further divided into cellular processes, environmental information processing, human diseases, metabolism and organismal system. Out of the 84 proteins, 58 were verified from MoonDB, and the other 27 proteins were predicted to be moonlight breast cancer proteins that needed to be verified in laboratory.

Contents

List of Figures

List of Tables

Abbreviations

AO	Alcohol oxidase
ATM	Ataxia telangiectasia mutated
Apaf-1	Protease-activatin 1
BLAST	Basic Local Alignment
BRCA1	Breast Cancer gene 1
CKI	Cyclin dependent kin ase inhibitor
DAVID	The Database for Annotation, Visualization and Integrated Discovery
DextMP	Deep dive into text for predicting moonlighting proteins
DLD	Dihydrolipoamide dehydrogenase
ER	Eestrogen receptor
FASTA	FAST Alignment
HDI	Human Development Index
HSP	Heat shock proteins
TNBC	Triple-negative breast cancer
MP	Moonlight Protein
NFB	Nuclear factor kappa B (CKI)
OMIM	Online Mendelian Inheritance in Man
PGK1	Phosphoglycerate kinase 1
PKM2	Pyruvate kinase M2
PR	Progesterone Receptor
Psi-BLAST	Position-Specific Iterated Basic Local Alignment Search Tool
SOPMA	Self Optimized Prediction Method with Alignment
STAT3	Signal transducer and activator of transcription 3
Tri	Tri carboxylic acid

UNIPROT Universal Protein Resource

Chapter 1

Introduction

A single protein that performs several functions without the use of, multiple RNA splicing, gene fusions or promiscuous enzyme proteolytic activity is known as a "moonlighting protein." It has been shown that many ribosomal protein components carry out essential extra-ribosomal tasks. Similar to people who work many jobs refers as moonlighting, also moonlighting proteins are multitasking polypeptide chains. There are several ribosomal protein components that carry out significant extra-ribosomal function, as mentioned in many studies. They typically perform a range of biological processes that are unique, physiologically significant, or unrelated [1]. These proteins are present in a variety of eukaryotes and prokaryotes, including yeast, bacteria, and humans [2].

The first moonlight model was published in 1980 by Piatigorsky and Wistow.They understood that crystallin, a structural protein found in the lens of the eye, also has an enzymatic role. These proteins have so far been found in humans, yeast, worms, bacteria, plants, viruses, archaea, and a number of other novel creatures and here been performing multiple functioning [3].

In order to maintain track of the information relevant to these proteins, several online databases have been created. MoonProt [4], MultitaskProtDB-II [5], and MoonDB [6] each reported 400, 694, and 238 proteins respectively in their most recent updates. There are several MP types, including: Various sites in the same

domain for various purposes .Multiple sites in different domains for various domains using a residue for several uses utilising different residues from the same site for varied roles, with varying structural makeup or foldings methods [7],

The majority of recognized moonlighting proteins are immensely expressed enzymes that are conserved. Owing to prior proof of these proteins' contribution in the growth of numerous infections, comprising contagious disorders and cancer, study of these proteins is now gaining attention. 80S ribosomes in eukaryote are huge intracellular composite structures made up of four ribosomal RNAs i[rRNAs] and approximately 80 ribosomal proteins [RPs]. Protein biosynthesis is carried out by these vastly invariant and conserved organelles Many RPs have been demonstrated to intricate in ribosome biogenesis, including RNA folding, pre ribosome transport, ribosomal subunit assemblage, and rRNA maintenance, in addition to being components of ribosomes [8]

New research suggests that ribosomal stress is brought on by extracellular or intracellular stressors, which prevent ribosomal biogenesis and result in the buildup of free RPs. DNA damage repair, drug resistance, apoptosis, cell propagation and differentiation, and cell migration and attack are examples of extra ribosomal activities. The involvement of ribosomal proteins in both tumor-suppressing and -promoting actions has been studied. Ribosomal proteins detach from the ribosome compound in response to stressors. To cause a physical outcome in the cell, a special interaction with some RNA or protein that is not a part of the ribosome is required [9].

Moonlighting proteins affect genomic sequence analysis and annotation since arrangement of homologs might share all, none, one, or a few functions. Moonlighting in systems biology adds another depth to our understanding of the complex yet controlled cellular protein network. For instance, a moonlighting protein may be in charge of a system for organising and coordinating the numerous intracellular routes, or it may be in charge of a switch that allows the cell to switch between pathways in response to environmental changes. Various cell types within an organism can communicate with one another and organise themselves via a moonlighting protein [10].

Our understanding of the complicated but tightly regulated cellular protein network has been improved by side work in systems biology. Numerous crucial genes for the development of tumours are connected to the moonlight Protein. One of the basic essential regulators for the stimulation of genes connected to cell production and existence are nuclear factor kappa B [NFB]i DNA-binding protein complexes. The five Rel subunits Rel [cRel], RelA [p65], RelB, NFB1 [p50/p105], and NFB2 [p52/p100] are collectively referred to as NFB in the homo- and heterodimer complexes. It has been confirmed that RPS3 act as a non-Reli component in the NFB complex. Body may rapidly and effectively reuse your material since this protein directly fixes to the Rel homology domain of the p65 homodimer through the K homology [KH] domain in the N terminals region of the cytoplasm and the nucleus [11].

Cell cycle disorders and excessive cellular proliferation are the main foundations of cancer. By speeding the G1/S transition and decreasing the production of p27 mRNA, overexpression of RPS13 has been initiated to stimulate the growth and development of gastric cancer cells through the cell cycle. The unique tumour suppressor p27 is a cyclin dependent kinase inhibitor (CKI), which manages CDK activity and hence controls cell cycle and act as moonlight Protein [12].

Signal transducers and activators of transcription 5 (STAT5) are the means through which RPL11 promotes cell proliferation in erythroid cells. A protein called RPS9 that binds to flavonoids causes cell cycle inhibition by activating the CDK1 enzyme. RPL19 has been shown to enhance the growth of cyclin D1, D3, and lung cancer cell lines when expressed in colon cancer cells. By increasing cycline expression, RPL6 overexpression also quickens the GES gastric cancer cell line's passage from G1 to S phase and determine its fate as Moonlight Protein [13].

Carcinogenesis, is characterized by six key features, resulting in the degenerative changes that cause a large percentage of malignancies. Resistance of apoptosis is one of the main mechanisms that promotes its advancement, along with a propensity for endless division, improved angiogenes is, immunity to anti-growth signals, and the ability to metastasis as well as the induction of own growth signals. Carcinogenesis is a complex process that is largely influenced by both genetic factors

and environmental factors as well. Uncomfortably, the number of fatalities from cancer is rising [14].

The number of deaths from cancer is alarmingly rising every year, making it one of the foremost reasons of death globally. However a major portion of cancers do not always end in death, they significantly reduce quality of life and result in higher overall costs. According to data from GLOBOCAN 2020, breast malignancy at present is one of the utmost repeatedly detected tumor and the fifth leading reason of melanoma-related deaths, with an estimated 2.3 million new cases worldwide. When compared to transitioned nations , expiries from breast malignancy are reported more frequently in the transitioning regions of the world with an incidence rate that is roughly 88% higher [15].

Over the past three decades, equally the incidence and life expectancy tolls of breast cancer have grown-up. Breast cancer incidence increased by more than double between 1990 and 2016 in 60 of 102 countries (including Afghanistan, the Philippines, Brazil, and Argentina), whereas fatalities increased by twofold in 43 of 102 countries (including Yemen, Paraguay, Libya, and Saudi Arabia) [16]. According to current estimates, there will be 2.7 million new cases diagnosed yearly over the world by 2030, while there will be 0.87 million fatalities [17]. Breast cancer incidence in low- and middle-income countries is predicted to rise further as a result of westernizing lifestyles (e.g., postponed pregnancies, less breastfeeding, early menarche, inactivity, and poor nutrition), improved cancer registration, and cancer diagnosis [18].

In women, breast cancer accounts for more than 20% of all cases of the disease, making it the most prevalent kind. Worldwide, more than a million women receive a breast cancer diagnosis each year, and 500,000 lose their lives to the condition. The cause of 5–10% of breast cancer cases in these women is hereditary predisposition. Numerous heritable diseases have also been linked to an increased risk of breast cancer. Over 50% of the genetic risk for breast cancer in families is still unknown, though. Due to a number of molecular alterations that promote cell proliferation and genetic instability, breast cancer is a challenging illness that can become more invading and resistant. This complexity leads to the emergence of

numerous molecular groups, which have a variety of clinical outcomes and therapeutic responses [19].

The understanding of the molecular processes that lead to breast cancer genesis has been increased by recent advancements in fundamental research. 10% of instances of familial breast cancer are caused by mutations in the p53, BRCA1, and PTEN genes. With one million new cases identified each year, breast cancer affects women more frequently than any other type of cancer. Additionally, it is the second important reason of death among women [20].

The WHO estimates that 107.8 million Disability-Adjusted Life Years (DALYs) are associated with malignant neoplasms. 2.26 million [95% UI, 2.24-2.79 million] new cases of breast cancer will be detected in women worldwide in 2020 [21]. In the US, breast cancer will likely represent 29% of all new cases of cancer in females. Age-standardized incidence rates (ASIRi) of breast cancer are directly correlated to the Human Development Index (HDIi) in many parts of the world. Data from 2020 show that the ASIR was highest in countries with very high HDI (75.6 per 100,000), whereas it was more than 200% lower in countries with medium and low HDI (27.8 per 100,000 and 36.1 per 100,000, respectively) [22].

In addition to being the most prevalent, breast cancer also kills more women from cancer than any other type. Breast cancer caused 684,996 deaths worldwide at a rate of 13.6/100,000 when adjusted for age. Although industrialized nations had the greatest incidence rates, 63% of all fatalities worldwide occurred in Asia and Africa in 2020. In high-income nations, the majority of breast cancer patients survive; however, this is not the case for many women in low- and middle-income nations [23].

As a realistic measure of 5-year survival rates, the mortality-to-incidence ratio (MIR) for breast cancer in 2020 was 0.30 worldwide [24]. In countries with modern healthcare standard, the 5-year existence rate of domestic cases and 75.4% for regional cancer cases on investigating depicted the wide experimental range of breast cancer of 89.6% and 75.4%, respectively. The survival rates for localised and

regional breast cancer in less developed nations (Costa Rica, India, Philippines, Saudi Arabia, Thailand) were 76.3% and 47.4%, respectively [25].

In Pakistan, one in nine patients is diagnosed with breast cancer, making it the most common type of cancer in women. It is 2.5 times more common in Pakistan than in neighboring countries such as Iran. Common risk elements for breast cancer includes age, family history, and menopausal hormone exposure , include estrogen and progestin, alcohol use, physical inactivity, poor socioeconomic status, and ignorance of the disease. New studies have revealed the extra ribosomal involvement of moonlighting ribosomal proteins in the growth of human cancers. Accurate measurement of gene expression levels is made possible by the discovery of genes whose expression is unaffected by cancer characteristics and patient characteristics [26].

1.1 Problem Statement

The term "moonlighting proteins" has been developed to describe well-known proteins that have recently been discovered to utilise new behaviours that are reportedly unrelated to their original roles. Subcellular localization change can result an assignment additional function to particular protein. A comprehensive evaluation and characterization of moonlighting protein networks can be pivotal to determine cancer prognosis and can also help in the development of more effective cancer therapeutic strategies.

1.2 Aims and Objectives

The aim of this study is to predict the proteins that act as moonlight in breast cancer progression for therapeutic purpose.

- To identify Proteins that act as moonlight proteins in breast cancer using text mining

- To functionally annotate the identified moonlight proteins involved in breast cancer
- To construct and analyze Protein - Protein interaction network to prioritize the key moonlight Protein
- To investigate the role of moonlight proteins associated with breast cancer in pathways

Chapter 2

Literature Review

2.1 Evolution of Moonlighting Proteins

Moonlighting proteins consist of proteins with two distinct activities that are combined into a single polypeptide. They exclude multifunctional proteins resulting from gene fusions, homologous protein families, splice variations, and promiscuous enzyme activities. They contain a variety of distinct protein varieties and functional combinations. Although the presence of a protein at an unexpected site can imply that the protein serves a second purpose, further evidence is needed to indicate that the protein actually fulfills two distinct biochemical functions in the two locations [27].

2.2 Characteristics of Moonlight Proteins

Most of the presently recognized moonlight proteins are extremely preserved enzymes, also referred to as ancient enzymes. Particularly sugar-metabolizing enzymes seem to work extra hours. The moonlighting role of seven out of ten glycolytic pathway proteins and seven out of eight tri carboxylic acid [TCA] cycle enzymes have been recommended to have a moonlighting role. Why moonlighting roles are so repeatedly seen in vastly preserved proteins is still a mystery. Perhaps

because highly conserved proteins are present in an extensive diversity of animals, there is a higher likelihood that one of them will be found to fulfill a secondary function than a protein that is not highly conserved. Moonlighting functions appear to be more common in proteins that are indicated at comparatively elevated levels constitutively [28].

2.3 Significance of Moonlight Proteins

Moonlight proteins have been discovered in prokaryotes, yeast, mammals, plants and yeast [29]. Although there are more samples of moonlight proteins in yeast, this is perhaps because these species have been the center of much investigation. The currently recognized moonlighting roles are highly diversified and implicated in a wide variety of biological activities [30].

2.3.1 *Escherichia Coli* and Thioredoxin

A prokaryotic moonlighting protein is thioredoxin, an anti-oxidant protein found in *E. coli* [31]. When *E. coli* is infected with the bacteriophage T7i, thioredoxin forms a complex with T7 DNA polymerase, resulting in increased T7 DNA replication a critical step in T7 infection success. Thioredoxin attached to a protein called thioredoxin. Thioredoxin's anti-oxidant action is completely separate and distinct from its role in T7 DNA replication, where most likely the protein shows a fundamental part [32].

2.3.2 In Methylotrophic Yeast

A well-studied enzyme called pyruvate carboxylase catalyses the initiation of the tricarboxylic acid cycle by carboxylating pyruvate to oxaloacetate .Surprisingly, effective pointing and assemblage of the protein alcohol oxidase [AO] peroxisomalare are also dependent on pyruvate carboxylase in methylotrophic yeast species as *Hansenula polymorpha* and *Pichia pastoris*. The homo-octameric flavoenzyme

alcohol oxidase is the first enzyme in the metabolism of methanol [33]. In wild-type cells, enzyme is introduced as active octamers in the peroxisomal matrix. Pyruvate carboxylase has a second, entirely independent role in the assemblage and significance of a peroxisomal matrix protein, as shown by the clustering of FAD-deficient AO monomers in the cytoplasm in cells missing the enzyme [34].

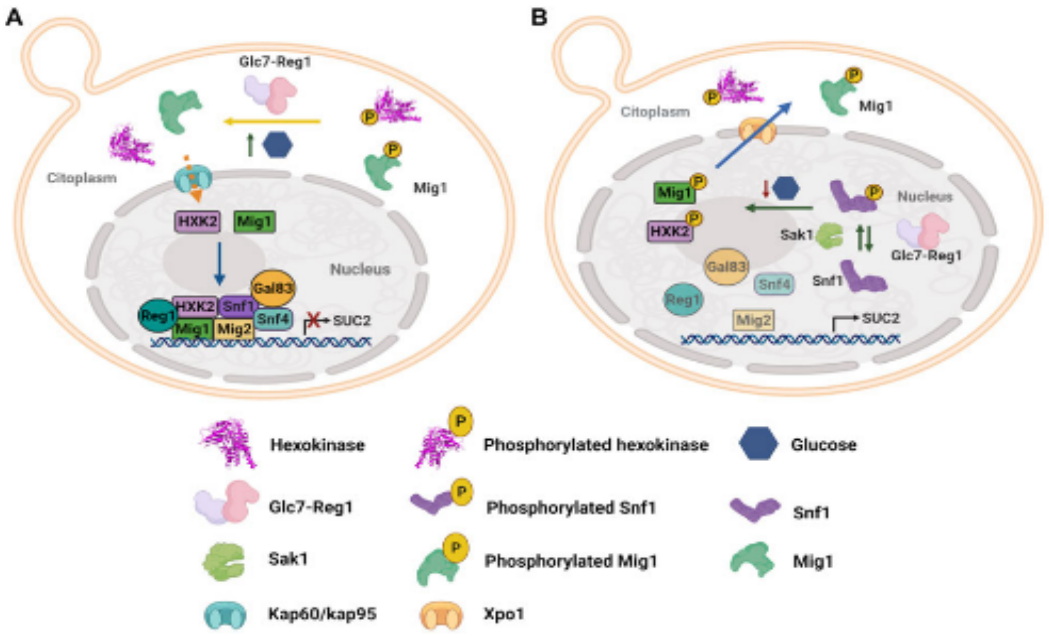


FIGURE 2.1: Yeast hexokinase moonlighting function and regulation [35]

2.3.3 Pyruvate Carboxylase

Methylotrophic yeast contains the enzyme pyruvate carboxylase. Pyruvate carboxylase serves this function, but how exactly it performs this function is uncertain. As it is also essential for a true moonlighting protein, the function in AO import/assembly is completely free for the activity of enzyme of pyruvate carboxylase, it has been suggested that amino acid replacements completely deactivate the activity of enzyme in pyruvate carboxylase without disturbing its role in assembly and import. On the other hand, mutations have been found that have no effects on the protein's enzymatic activity but entirely affect pyruvate carboxylase's role in import and assembly [36].

2.3.4 *Physcomitrella Patens* Presenilin

Presenilin is an enzyme that catalyses secretase enzyme complex a multiprotein, which splits chief proteins including Notch and amyloid precursor protein (APP)i, both of which have been associated with neurologic disorder Mammalian presenilin, postulated for secondary functions, though it is challenging to investigate these activities in mammals. The moss *P. patens* is used to streamline the inquiry since it has -secretase but neither Notch nor APP. When the gene that makes presenilin was removed, P. patens phenotypic deficiencies were seen, proving that presenilin is involved in the moss cytoskeletal network. This special function is unconnected to presenilin's enzymatic activity because presenilin mutants that lack enzymatic activity were nevertheless able to repair the abnormal shape. Surprisingly, an enzymatically inactive version of human presenilin was able to restore the phenotype when injected into P. patens. This indicates that presenilin may have a secondary function that has maintained throughout evolution given that it is found in both plants and mammals [37].

2.3.5 Cytochrome c

Cytochrome c , a protein located in the mitochondrial intermembrane space, is a component of the electron transport chain. The protein does, however, play a significant role in apoptosis when released into the cytosol. When cytosolic cytochrome and apoptotic protease-activating factor 1 (Apaf-1) combine, a signalling cascade that results in apoptotic cell death is stimulated. It is likely to produce an altered form of cytochrome c correctly in respiration but does not bind to Apaf-1. Cytochrome c's redox and pro-apoptotic actions are completely distinct from one another. Cytochrome c therefore shows all the features of a true moonlight protein [38].

2.3.6 STAT3

According to new findings, STAT3 in mammals appears to be a real moonlighting protein. Proteins can either be signal transducers or transcription activators (STATs) [30]. Phosphorylated STATs go from the cytoplasm to nucleus, where a

number of genes are present to control their expression. Leptin activates the protein STAT3, which controls energy intake and metabolism throughout the body Furthermore, a STAT3 mutant has been well-known that is transcriptionally active but not capable to renovate function of mitochondria. Therefore, decreasing STAT3's ability to control transcription has no impact on how it contributes to mitochondrial respiration, and vice versa . According to Wegrzyn et al. a share of the cellular protein STAT3 is localised to mitochondria where it takes part in oxidative phosphorylation [39].

2.3.7 Gene Duplication and Role of Moonlight Protein

Moonlighting proteins may encounter stress similar to that faced by numerous individuals who hold down two jobs. It's possible that the expression pattern needed for one function simply does not work for another. Additionally, a mutation may rise the effectiveness of one task while decreasing the effectiveness of another. The fourth enzyme of the urea cycle, arginine succinate lyase, serves as an example. This enzyme is a dual-purpose protein in ducks and ostriches because it also functions as a moonlight protein . One of them serves as structural Crystallin in the eye lens and is the enzyme's inactive form. The second one is the urea cycle's enzymatically active enzyme [40].

2.4 Moonlight Proteins in Cell Cycle

2.4.1 Cell Cycle Proteins

In addition to stress conserved proteins and metabolic enzymes and a novel class of proteins with a longer history is that which is involved in cell division. When environmental conditions are favorable, cell division serves as a mode of reproduction in unicellular organisms, and these proteins are produced in a way that is consistent with how they regulate cell division in these species. . Yet, in organism that consists of more than one cell, cells specialized into different cells, and their capacity for reproduction is frequently lost or diminished. This specialisation causes a variety of proteins involved in regulating and checking activities throughout the

cell series of event to lose function or become extinct. These protein may change into new secondary roles if ,produced in non-dividing cells and help other cells functionally. Otherwise, as cell division mechanisms proceeds , proteins required for metabolic phase may take additional function during cell division to aid in the smooth functioning of the process [41].

2.4.2 As a Chromatin Modifiers

Chromatin remodelers act as chromatin modifiers, that utilize the energy of ATP. They were all well studied for both their in vivo and in vitro actions, as well as their effects on transcription. Two ATPase, have recently been found to influence spindle motion during cell cycle. Microtubules and associated proteins, is a molecular framework that separates during cell division. Gene expression is regulated by these proteins in the nucleus during interphase, when they are segregated from interphase microtubules. They shift sites during mitosis [42].

As the nuclear envelope disintegrates during mitosis, they migrate, augmented on the microtubule. Spindle stability is regulated by CHD4, and spindle assembly is encouraged by ISWI. Furthermore, neither of these proteins' ATPase activities is required for either their capacity to bind microtubules or for their mitotic function, illustrating functions in cell division are different from those during transcription mechanism [43].

2.4.3 Association of Spindle Assembly Checkpoint Proteins with Insulin Signaling

Mad2 is a checkpoint response driver for spindle assemblies [44]. This protein is a participant of the family of proteins recognized as HORMADs, or HORMAi domain-containing proteins. Members of this family are distinguished from one another by their capacity to alter the conformation of required peptide bonds found in other proteins. HORMAD protein function is frequently influenced by its shape. For instance, when Mad2 is free, it does not participate in checkpoint

signaling. In an associated protein, after binding a peptide it takes on a locked structure. Chromosomes are monitored by spindle assembly checkpoint proteins to see if they are connected to the microtubules. They send out a signal that triggers a cell cycle if this attachment is defective or lacking [45].

The fact that several of these proteins have more recently been given interphase roles, however, suggests that in addition to their primary role during cell division, they can also work more extensively depending on the kind of cell. Accepting extra, moonlight , activities may also provide understanding into tumor diagnosis and cure assumed that genes associated with spindle check point frequently relates with cancer growth and diagnosis protein and can block the cell cycle. As a result, any procedure employing a protein that has one of these peptides available for binding qualifies and identifies it as a moonlight protein [46].

2.4.4 Regulation of Kinetochore Microtubule Interactions with Membrane Trafficking Proteins

Membrane trafficking proteins are a different class of proteins involved in mitosis. Membrane trafficking proteins control kinetochore microtubule connections. A more thorough description of these proteins is available, but for the purposes of this article, Clathrin and TRAMM will be highlighted. Clathrin and TRAMM control vesicle trafficking during interphase and encourage stable chromosome-microtubule interactions during mitosis. These two proteins work together to regulate vesicles. Trafficking is encouraged during interphase, whereas mitosis encourages permanent links between chromosomes and microtubules [47].

2.5 Moonlight Protein and Human Health

The involvement of moonlighting proteins may be connected to the complex phenotypes of different illnesses. There are few well-studied specimens of moonlight proteins that may contribute to illness, despite the fact that the majority of these statements lack supporting evidence. Moonlighting proteins are frequently investigated from the viewpoints of cellular and molecular level, but they are very important to human well-being as the connection among human illness and moonlight

proteins. Particularly, certain proteins with well-known moonlighting capabilities have been connected to human illness [48].

One disease type in particular that has been connected to moonlighting Protein is cancer. Either reveal a newly unknown moonlighting function or the moon lighting function. This third strategy could be particularly pertinent to take into consideration given that metabolic and housekeeping proteins commonly evolve side-functions, and cancer cells' metabolic profiles are intricately described. Examining all of these potentials, even if they don't happen under normal physical circumstances, can aid in diagnosis and therapy because protein excess expression has been associated with patient outcomes and cancer development [49].

2.5.1 Role of GAPDH and Cancer

Human, breast, pancreatic, colorectal ,skin, carcinomas ,colorectal, lung, and cervical kidney have all been shown to overexpress GAPDH. The ways by which tumour cells interfere with GAPDH's activities are almost as varied as GAPDH's functions themselves [50].

Tumor cells damage GAPDH function in three separate ways to aid in their own existence .As initially shown phosphorylated serine kinase Aktp phosphorylates GAPDH in heart muscle cells, creates a GAPDHpi -Aktp complex, unable to translocate into the nucleus and carry out its apoptotic function because it is entrapped in this complex, leading to cell survival [51].

2.5.2 Role of Protein Kinases and Metabolic Kinases in Cancer

Studies have shown that, in addition to GAPDH, other metabolic kinases, including protein kinases, have been associated with cancer. The last stage of the method is completed by PKM2. Glycolysis is the process by which phosphoenolpyruvate is converted to pyruvate to create ATP [53].

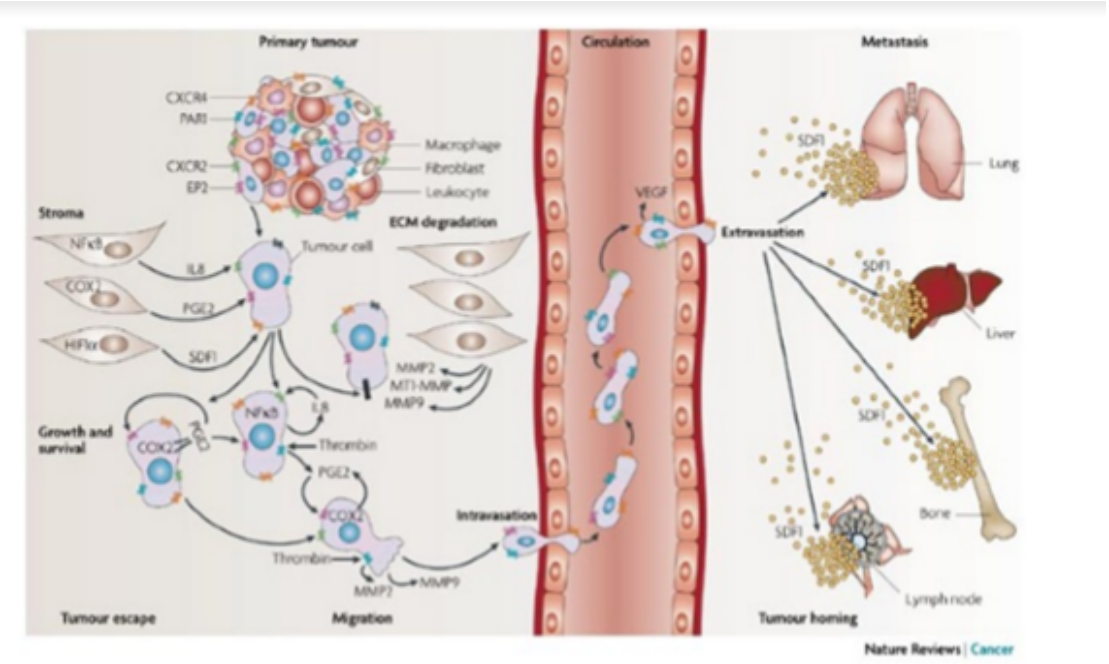


FIGURE 2.2: G-coupled proteins and cancerous tumour cell proliferatio [52]

Similar to PKM2, PGK1 also possesses protein kinase activity that has been connected to cancer and creates ATP during glycolysis. By phosphorylating the, PGK1 suppresses mitochondrial pyruvate metabolism and encourages cancer. Additionally, PGK1 phosphorylates, which is necessary for the beginning of autophagy, an important process that is typically accelerated in cancer cells. Through regulating glycolysis, mitochondrial metabolism, and other processes, PGK1 appears to show a vital role in the development and spread of cancer [54].

2.5.3 (DLD) Dihydrolipoamide Dehydrogenase

Dihydrolipoamide dehydrogenase, or DLD, is a mitochondrial enzyme that is located in at least five multienzyme complexes [55]. DLD is a component of these enzyme complexes, which makes it crucial for redox balance and energy metabolism. Insufficiencies in DLD action in children are associated with significant issues such metabolic disorders, hypotonia, and failure to thrive. On the other hand, the severity of the symptoms varies greatly and is based on the gene mutation. The protein is mostly monomeric under particular circumstances, such like mitochondrial matrix acidification, which results in the loss of DLD enzyme function.

According to Babady et al.[56] homodimer-destabilizing mutations have an extra effect that improves DLD's capacity to act as a protease by increasing the disclosure of a catalytic dyad at the dimer interface. The enzymatic action of the enzyme is unrelated to its proteolytic activity. DLD's secondary function might be harmful to metabolic health [57].

2.6 Moonlight Proteins in Microbes

2.6.1 Pathogenesis

The virulence of bacterial and fungal diseases has been linked to a surprising quantity of moonlighting proteins [58].

Unexpectedly many moonlight proteins have been connected to the pathogenicity of bacterial and fungal infections. The ability of these housekeeping proteins to moonlight during illness depends on their release outside of the cell. Generally, they are engaged in chaperone function, stress response, or metabolism. Unexpectedly, their secretion occurs without the aid of well-researched sorting mechanisms that seek for extracellular localisation. Their relationship to the cell surface is also unknown. Outside pathogens, these proteins promote signalling or adherence and can even act as toxins [59].

2.6.2 Enolase

Another protein having a special purpose in metabolism is enolase. Enolase is an enzyme that catalyses glycolysis, just like GAPDH does. Pathogenic bacteria and other species of enolase diffuse on the surface of cells as enolase moonlighters. Enolase's ability to moonlight has been seen in a number of Streptococcus species. Enolase has the ability to bind plasminogen, cytokeratin 8 (*Streptococcus gallolyticus*), *S. pneumoniae*, *Streptococcus canis*, *Streptococcus gordonii*, and salivary mucin (*S. mutans*) [60]. Enolase has several functions in *S. pneumoniae*. The most widespread multifunctional protein among disease causing microbes researched so

far appears to be plasminogen binding, which is seen in *Aeromonas hydrophila* [61].

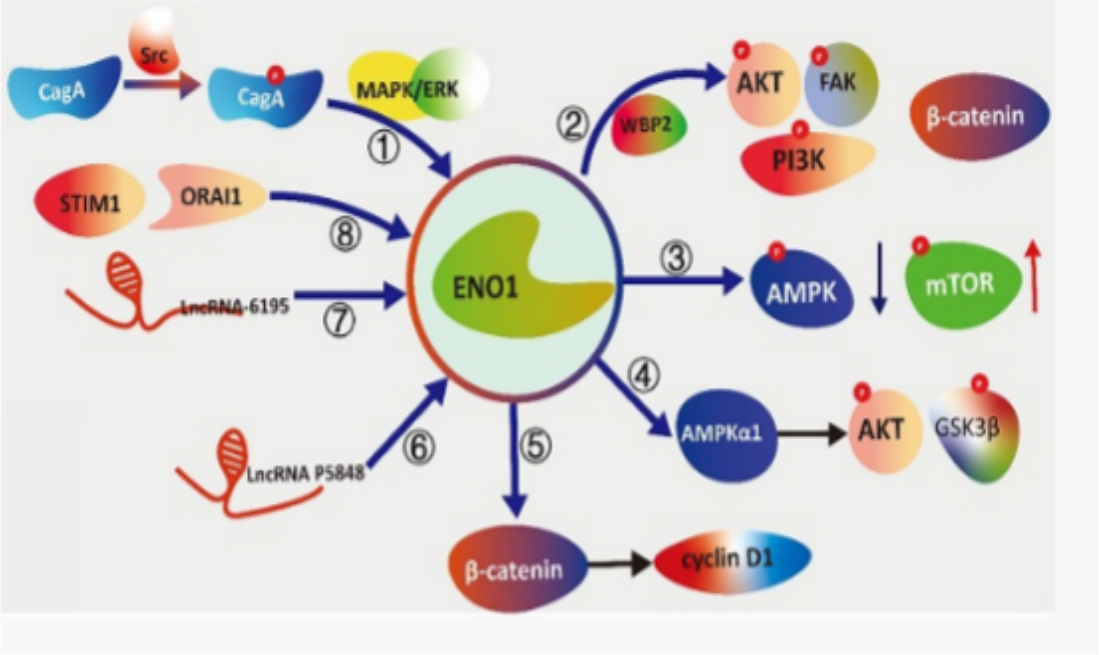


FIGURE 2.3: Enolase 1 (ENO1) function as a glycolysis enzyme and DNA binding protein is a moonlighting protein [62]

2.6.3 *Mycobacterium tuberculosis* Glutamate Racemase

The disease causing bacteria Mycobacterium tuberculosis is what causes human TB . Highly infectious and potentially lethal if neglected, this illness. *M. tuberculosis* can be treated with the antibiotic ciprofloxacin, which has a wide range of action . It facilitates the production of DNA interruptions when attached to a DNA gyrase [63].

The *M. tuberculosis MurI* protein's covert activity inhibits the effects of ciprofloxacin. MurI is necessary for *M. tuberculosis* cell wall (peptidoglycan) production. It facilitates the transformation of l-glutamate , a component of peptidoglycans, into d-glutamate. In many species of bacteria, including *M. tuberculosis*, MurI can also operate as a DNA gyrase inhibitor by decreasing gyrase binding . Whether or not the enzyme is active, MurI overproduction defends *M. tuberculosis* in contrast to the side effects of ciprofloxacin [64].

2.6.4 GAPDH in Pathogenesis

It is generally recognised that (GAPDH) glyceraldehyde-3-phosphate dehydrogenase participates in glycolysis, the process by which glucose is broken down to create energy in the cell. Most scientists believe it to be a housekeeping gene since it is produced at extraordinary level in the majority of tissues. Furthermore it is linked to other processes such as apoptosis , transcriptional control ,membrane fusion, , iron transport vesicle ,transport from the Golgi apparatus to the endoplasmic reticulum, and cellular response to environmental stress like hypoxia and oxygen deficiency. [65].

These processes are distributed throughout the cell, including the cytosol ,cell membrane, and nucleus GAPDH performs many more biological tasks during cell-cell interactions. For instance, it functions as a double operator in four distinct species of Streptococcus. GAPDHi functions, as an adhesin and invasin in the cell surface of Streptococcus pyogenes. It also functions as a neutrophil protein [66].

Moonlights GAPDH in Mycoplasma genitaliumi, as cell surface protein involved in binding mucin. More moonlighting roles of GAPDH will be discovered. In four distinct Streptococcus species , each of the moonlight performs differently. GAPDH stimulates B cells in Streptococcus agalactiae by acting as an immunomodulator. Nonpathogenic E.coli bacteria do not express GAPDH on their cell surfaces, while enterohemorrhagic and enterohaemorrhagic strains do. It is possible that when additional bacterial species are investigated, more GAPDH side functions will be identified [67].

2.6.5 Chaperones

Like moonlighting proteins, which were originally identified, chaperones are a protein family that are extremely well-preserved that bacterial pathogens have employed to contaminate hosts. Similar to metabolic enzymes, four bacterial chaperones or stress response proteins, chaperonin (Hsp) 10, peptidylprolyl isomerase, chaperonin (Hsp) 60, and DNAK/Hsp70, have been found to mediate either signalling of host immune cells or adherence to host tissue during colonisation [68].

Chaperones on the cell surfaces of *Mycobacterium TB* and *Helicobacter pylori* trigger monocytes to create pro-inflammatory cytokines. It's noteworthy to note that the same chaperone, chaperonin (Hsp) 60, may function as both a signalling protein and a chaperon during M. tuberculosis infection [69].

2.7 Genes Involved in Breast Cancer

2.7.1 BRCA1 and BRCA2

The BRCA1 gene, which is situated on chromosome 17, was the first significant gene connected to hereditary breast cancer [70].

Breast and other cancers are more likely to develop when one of the BRCA1 or BRCA2 genes is mutated. Large deletions and rearrangements in BRCA1 or BRCA2 can also affect how the genes operate, causing an analogous clinical condition to that found in carriers of these gene abnormalities. Mutations in BRCA2 and BRCA1 are autosomal dominantly pass on to generations, although they function as tumor suppressor genes on the cellular level that are engaged in DNA disruption in a recessive manner[71].

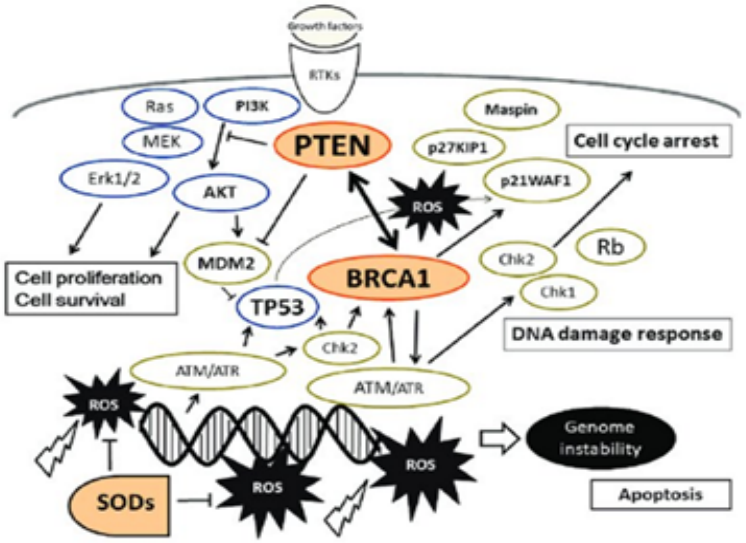


FIGURE 2.4: PTEN and BRCA1 signaling pathways [72]

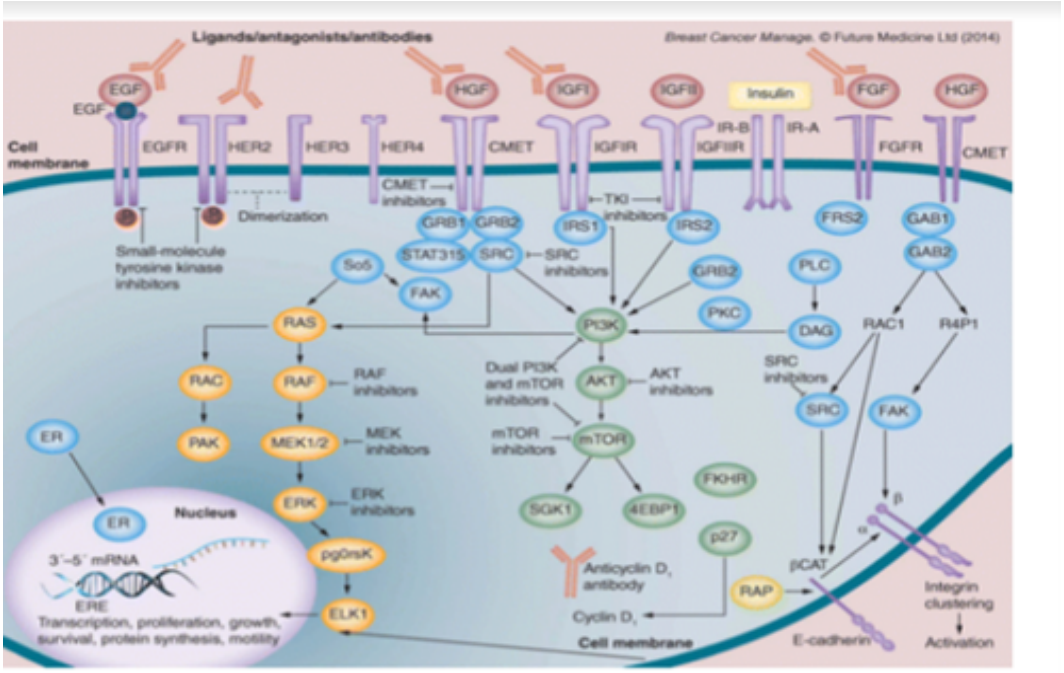


FIGURE 2.5: Multiple pathways in breast cancer [75]

In comparison to BRCA2 carriers, who are thought to have a 5%–10% lifetime risk, male BRCA1 carriers needs an elevated threat of breast melanoma, but toward a smaller extent. The disorders' further characteristics are listed. Most significantly, there is a higher threat of ovarian cancer, with lifetime risks for BRCA1 carriers estimated to be between 10% and 40% and for BRCA2 carriers to be between 10% and 20%. [73]. The disorders' further characteristics are listed in table 2.1. Most significantly, there is a greater possibility of ovarian cancer, with lifetime risks for BRCA1 carriers estimated to be between 10% and 40% and for BRCA2 carriers to be between 10% and 20%. Biallelic BRCA2 mutations considerably enhance the risk of juvenile malignancies and present with the Fanconi anaemia type D1 clinical presentation. Rarely documented biallelic BRCA1 mutations are probably embryonally fatal in the majority of instances [74].

2.7.2 BRIP1

A protein called BRIP1 (BACH1) is encoded by the BRCA1 C-Terminus (BRCT) domain. About 1% of tumors of breast are caused by changes in BRIP1. In women with a significant family history of breast melanoma, a genetic defect in BRIP1 is

TABLE 2.1: Breast cancer moderate-penetrance genes

Gene	**Function**	**Breast Cancer risk**	**Biallelic Phenotype i**
CHEK2	Involved in cell cycle regulation at G2.Activated CHEK2 stabilizes p53 and interacts with BRCA1	Female: RR 1.70, 95% CI 1.3–2.2 Male: RR 10.3, 95% CI 3.5–30.0	None known presumed to be embryonic lethal
BRIP1 (BACH1)	Cooperates with the BRCA1 C-Terminus (BRCT) domain of BRCA1	Women: RR 2.0, 95% CI 1.2–3.2 <50 ages: RR 3.5, 95% CI 1.9–5.7	Fanconi anemia, type J no major growth in infantile cancers
ATM	Protein kinase intricate in observing and repair of ds-DNA and regulation of BRCA1 and CHEK2	RR 2.37, 95% CI 1.5–3.8	Ataxia-telangiectasia i autosomal recessive Inheritance
PALB2	Connections with BRCA2. Intricate in nuclear localization and stability	All women: RR 2.3, 95% CI 1.4–3.9 <50 years: RR 3.0, 95% CI 1.4–5.5	Fanconi anemia type N-advanced occurrence of juvenile cancers

linked, with an increased risk for early-onset breast cancer. The bulk of BRIP1 mutations that have been reported so far truncate proteins. Without an apparent rise in paediatric malignancies, biallelic BRIP1 is linked to Fanconi anaemia type J [76].

2.7.3 ATM

A protein kinase called ATM controls the activity of BRCA1 and CHEK2 as well as the monitoring and repair of dsDNA. Ataxia-telangiectasia is an autosomal recessive condition brought on by a biallelic ATM mutation. It is expected that 1% of ATM mutations are monoallelic. A current meta-analysis found that the RR of A greater possibility of breast malignancy was seen in women under the age of fifty, and the chances of breast malignancy linked with an ATM mutation were 2.3%. Further genes implicated in DNA loss repair, such as RAD51C and genes in the

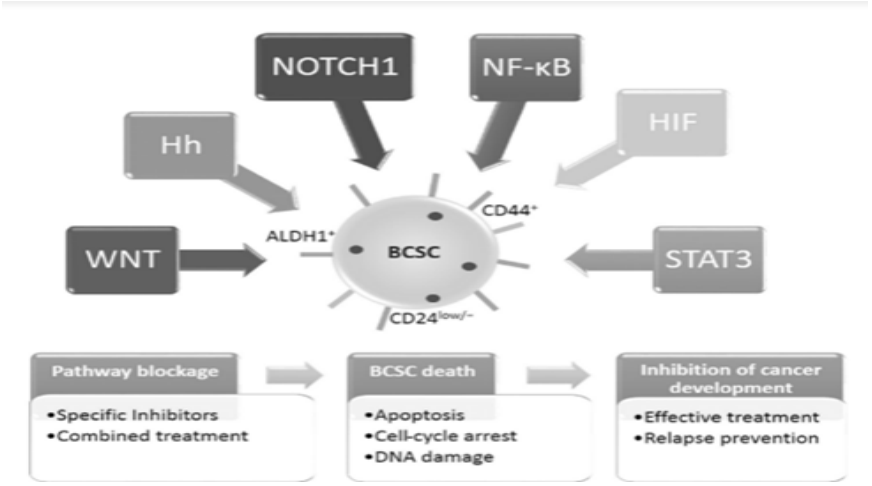


FIGURE 2.6: Breast cancer stem cells' cellular signaling pathways [77]

TABLE 2.2: Genes related to breast cancer associated Syndrome

Gene	**Syndrome**	**Breast cancer incidence**	**Other associated Cancers**	**Non Malignant syndrome feature**
BRCA1 BRCA2	Hereditary Breast-/Ovarian Cancer Syndrome	82% lifetime risk	Ovarian and fallopian tube cancer	Pathognomonic skin lesions Macrocephaly, benign breast and thyroid disease
PTEN	PTEN Hamartoma Tumor Syndrome Cowden Syndrome	85% lifetime risk	Non medullary thyroid cancer Endometrial cancer GU tumors, especially renal cell carcinomai	intestinal hamartomas,
TP53	Li-Fraumeni Syndrome	25% at age 74	Braintumor Sarcoma. Adrenocortical carcinoma. Leukemiai Lung-bronchoalveolar cancer	mental retardation
CDH1	Hereditary Diffuse	39% lobular breast cancer	Gastric cancer diffuse subtype Colorectal cancer	
STK11	Gastric Cancer Peutz-Jeghers Syndrome	GI cancers (esophagus, stomach, small bowel,colon)		

MRN DNA Investigations into the repair route have also been made. Nevertheless, families with high-risk were tested, no genetic change were demonstrably linked to an elevated risk of cancer or a effects in particular populations exist and aid in the emergence and spread of cancer [72].

2.8 Prognostic Marker

2.8.1 Estrogen Receptor

Since around 70–75 percent of invasive breast cancers have considerably increased ER expression, the estrogen receptor (ER) is a crucial diagnostic factor [78]. In accordance with current guidelines, both initial invasive tumours and recurring lesions must have their ER expression measured. Although the investigation of changed ER form is a highly significant step for choosing the right medication, the expression of ER may also be a projecting aspect since people with more ER expression typically have much healthier experimental consequences [79]. The effectiveness of ER expression as a breast cancer diagnostic indicator, particularly in situations of genetic peril, is further facilitated by the association that has been found between it and the family history of breast cancer [80].

2.8.2 Receptor for Progesterone

PR is rarely present (10%) among people suffering from breast tumour, with ER-negative as compared to people with ER-positive breast cancer [81]. Because ER controls PR expression, the physiological characteristics of PR provide information regarding the functionality of the ER i pathway. On the other hand, both are highly expressed in breast malignant tumor known as diagnostic breast cancer biomarkers (particularly for ER-positive tumours). Greater receptor for progesterone expression is crucial in determining time for treatment failure or progress, whereas lower receptor for progesterone thresholds remain most of the times linked with other hostile disease progression, worse prognosis, and longer times to recurrence and progression [82].

2.8.3 Receptor for Human Epidermal Growth Factor 2

(HER2) over expression during breast carcinogenesis, breast tumour accounts for 15–25% of breast cancers. As a result, HER2 status is largely significant for

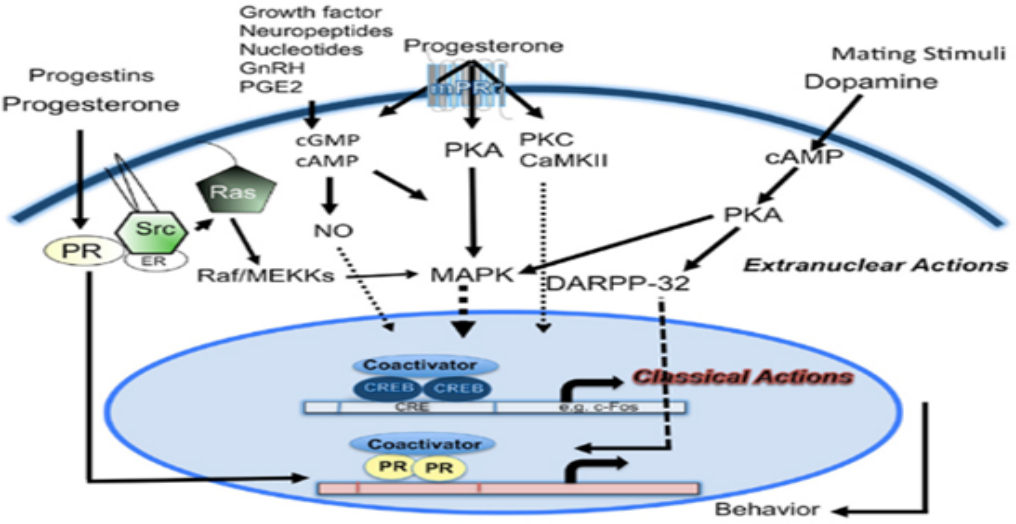


FIGURE 2.7: Interacellular And Extracellular Progesterone Signaling Pathway Cross- [83]

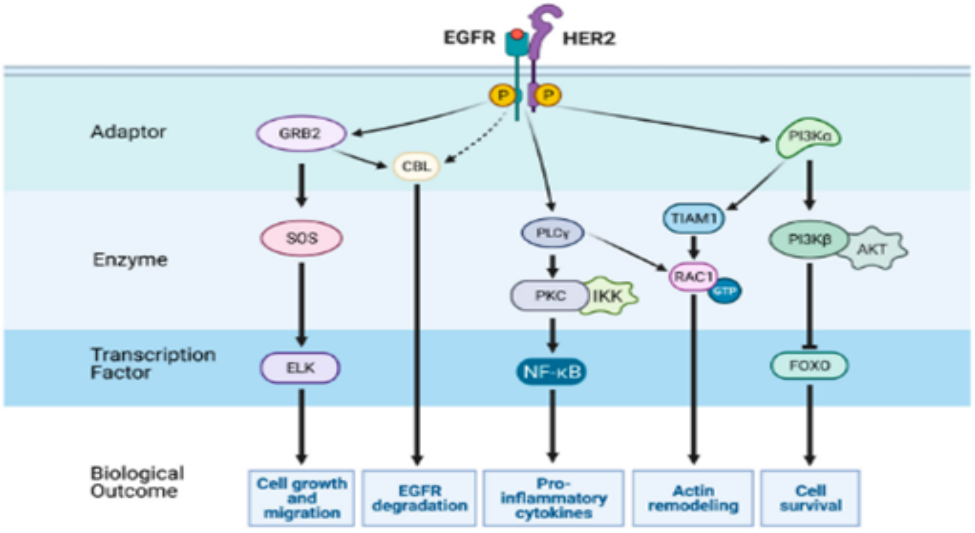


FIGURE 2.8: EGFR2 Pathway in Cancer [87]

choosing the best care for breast cancer patients [84]. The discovery level of metastatic cells or recurring breast tumours is also increased by HER2 from 50% to even more than 80% [85]. A possible immediate marker of tumour existence or repetition is the amount of serum HER2. In the situation of tumours like HER2-positive, HER2 amplification results in additional over stimulation of the oncogenic signalling paths, unchecked cancer cell proliferation, and worse clinical outcomes [86].

2.8.4 Antigen Ki-67

An effective indicator to offer evidences on the spread of malignant tissues, particularly in breast tumor, is the protein Ki-67, which is a biological indicator of production. The Ki-67i production index is based on the Ki-67 protein. The Ki-67i proliferation index is based on the Ki-67 protein. The Ki-67 proliferative activities measure the cancer's aggressiveness, treatment response, and duration between recurrences [88]. Therefore, Ki-67 is important for deciding on the best course of treatment and any necessary follow-ups in case of recurrence. Nevertheless, given the numerous restrictions on the diagnostic acceptability of Ki-67 expression levels have to be favourable while making treatment decisions. The overexpression of Ki-67 has remained connected to patients having poor clinical outcomes, according to a systematic review of 68,cases including 12,155,patients therefore it might also be thought of as a possible prognostic indicator. Poorer patient life expectancies in breast cancer patients are also associated with high expression of Ki-67. Ki-67 has been the subject of some concern as a potential prognostic marker, but the available evidence is currently few and inconsistent [89].

2.8.5 Mib1

Similar to Ki-67, the Mib1 proliferation index (beside anti-Ki-67) is still a valid diagnostic indicator for breast malignancy. A positive response of patients to effective therapy is related with a reduction in equally Ki-67i and Mib1i expressions [83]. Patients who also have concurrent p53 mutations have considerably higher levels of Mib1 .For biopsy specimens that are too small for mitotic index or S-phase fraction analysis, Mib1 assessment may be very helpful [90].

2.8.6 E-Cadherin

The epithelial-mesenchymal transition (EMT) requires the protein E-cadherin, and its absence causes a progressive change into the mesenchymal composition, which added to an increased threat of metastasis. Although the usefulness of E-cadherin study has suggested that its appearance may be related to a number of

breast cancer characteristics, including tumors size, lymph node status or TNMi stage, .The identification of the histologic subtype of breast cancer may be helped by slight or even complete loss of E-cadherin, expression. In terms of assessing patients' survival rates, E-cadherin, level do not appear to be favorable [87].

2.8.7 Circulating Circular RNA

Circular RNAs (circRNAs), which are non-coding RNAs, have lately been established to be important for a number of breast cancer indicators, such as apoptosis, increased production, or enlarged metastatic, prospective [91]. The hsa circ 0072309, which is highly stated in tumour patients and typically related through lesser existence , are two of the most thoroughly termed circRNAsi, mostly detailed to breast cancercircFBXW7, which remained suggested as a likely diagnostic biomarker also as beneficial instrument for victims with triple-negativei breast cancer (TNBCi). Has circ 0001785 is regarded as a promising breast cancer diagnostic biomarker [92].

2.8.8 P53

Numerous forms of cancer, including osteosarcomas, leukaemia, brain tumours, adrenocortical carcinomas, and breast cancer, have been associated to cause damage to the TP53 (P53) gene causing mutations [93]. The P53 protein mediates cellular stress reactions, and is crucial for healthy cellular homeostasis and sustaining the genome. The P53 gene silencing mutations are visible in the initial phases of cancer development. In terms of breast cancer prevalence, 10%i of patients with Luminal Aiidisease and 80% of TNBC patients both have the TP53 mutation [94].

2.8.9 MicroRNA

MicroRNAs (miRNAs) are a main class of noncoding RNAs, molecules (19–25 nucleotides) that have roles in various path. Micro RNAs linked to tumor growth, progress and reaction to therapy [96]. MiRNAs that exhibit aberrant expression have been examined as biomarkers in a number of studies. Two miRNAsi (miRNA-21i and miRNA-210i) were constantly elevated, whereas six miRNAs were regularly suppressed [97].

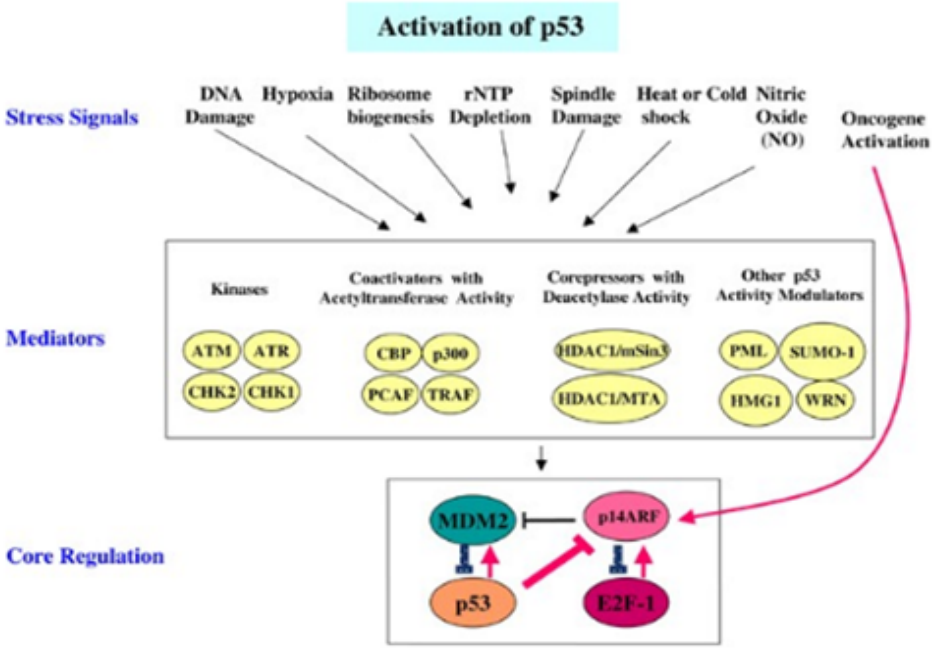


FIGURE 2.9: P53 Signaling Pathway: Positive and negative feedback loop [95]

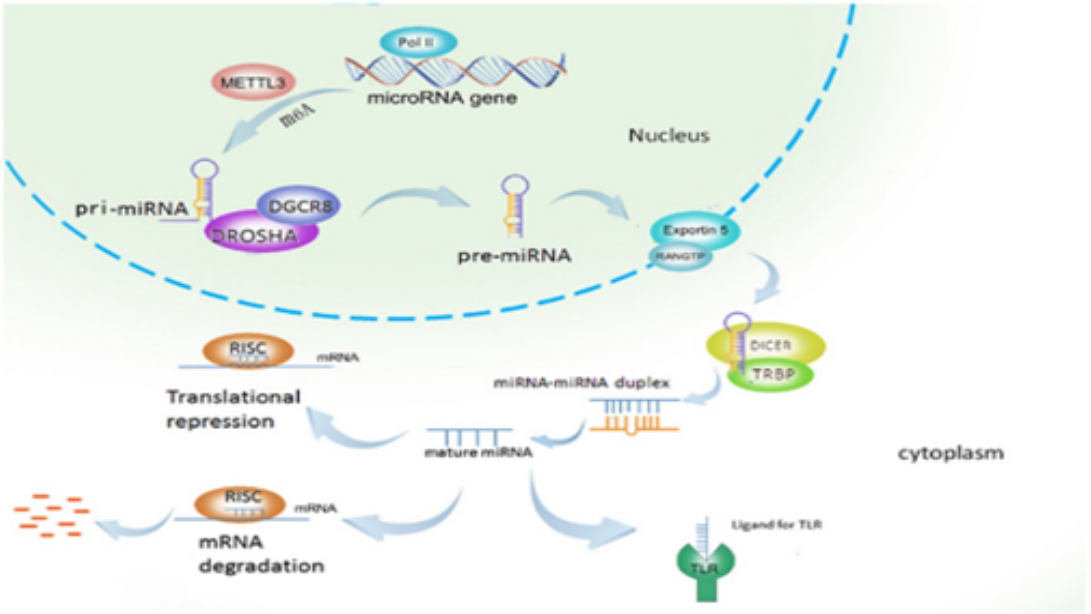


FIGURE 2.10: MicroRNA biogenesis in cancer [98]

2.8.10 Tumor-Associated Macrophages

Macrophages can be classified as M1- or M2-like states based on their morphologies and are well recognised for their immunomodulatory activities [99]. IL-12i

and tumour necrosis feature, which have antibacterial and anticancer actions, are secreted by M1 macrophages. Cytokines such IL-10, IL-1i type II receptor antagonist, and IL-1 idecoy receptor are produced by M2 macrophages. Therefore, M1-like i macrophages have remained associated to a favourable disorder outcome, whereas M2-like macrophages have been linked to a poor result, possibly due to immunosuppression, stimulation [100].M2 that support tumour development and metastasis are referred to as tumor-associated macrophages (TAMs). TAM density in breast cancer is correlated with position of hormone receptor, phase and histological status vascular attack, and lymph node metastasis, according to studies [101].

2.8.11 Models Based on Inflammation

Inflammation and host immunity in cancer cell and its surrounding are serious machineries in cancer development and progression. Tumor influence the inflammation leading to white blood cell causes changes in the peripheral blood cells [95]. White blood cells are modified due to tumor attack [98]. Consequently, there may be a relationship between peripheral blood inflammatory cells it serves as a nearby and initial process of forecasting a patient's diagnosis. New studies stated the prognostic function of cell causing inflammatory proportions: neutrophil-to-lymphocyte ratio, in various cancers [102].

2.8.12 The ratio of neutrophils to lymphocytes (NLR)

In patients with wide study on 27,031 malignant cells, in numerous cancers including breast tumor it was examined that predictive value of NLR establish a major association between NLR and breast tumour [103]. Lymphocytes plays a vital role in breast cancer immuno surveillance. Oppositely neutrophils destroy the cytolytic action of lymphocytes, leading to improved cancer development and proliferation [104]. Azab et al. stated that NLR before chemotherapy was a liberated cause for long-lasting mortality and related this one to stage and lump magnitude in breast cancer [10].

2.8.13 Ratio of Lymphocytes to Monocytes

The ratio of Lymphocytes to Monocytes (LMR) relation among patient diagnosis has been informed in numerous malignances [105]. Lymphocytes influence cell damage and limiting malignant tumor production, monocytes contribute in tumor formation. Lymphocytes influence cell damage and limiting malignant tumor production, monocytes contribute in tumor formation. In the lump region, free radicals and cytokines which are secreted by monocytes and macrophages, are related with tumor cell invasion and metastatic growth [106].

2.8.14 Platelet-to-Lymphocyte Ratio (PLR)

In various cancer forms, an excessive platelet calculation has been linked to a negative prognosis [107]. A meta-analytic analysis that included 5542 breast cancer patients focused at the predictive importance of PLR. Although a high PLR level was linked to a poor diagnosis (both general and uninfected survival), its therapeutic usefulness for molecular subtypes of breast malignancy was not proven. However, a correlation between PLR and clinic pathological characteristics of the tumour, such as stage lymph node metastasis, and distant metastasis, was discovered. Although a high PLR level was linked to a poor diagnosis (both general and uninfected survival), its therapeutic usefulness for molecular subtypes of breast malignancy was not proven. However, a correlation between PLR and clinic pathological characteristics of the tumour, such as stage lymph node metastasis, and distant metastasis, was discovered. While earlier research discovered a distinction between ER and PR hormonal states, the meta-analysis noted an alteration in the prevalence of high PLR level among HER2 status [108].

2.9 Research Gap

The concept of moonlight protein is not new but its implications in cancer is new. Due to novel implications in cancer, moonlight protein has been reported and confirmed in few cancer such as pancreatic and lung cancer but breast cancer has yet not been explored with reference to moonlight cancer.

2.10 Research Questions

- Which Moonlight proteins are associated with Breast cancer?
- What is the role of Moonlight in prognosis of Breast cancer?

Chapter 3

Methodology

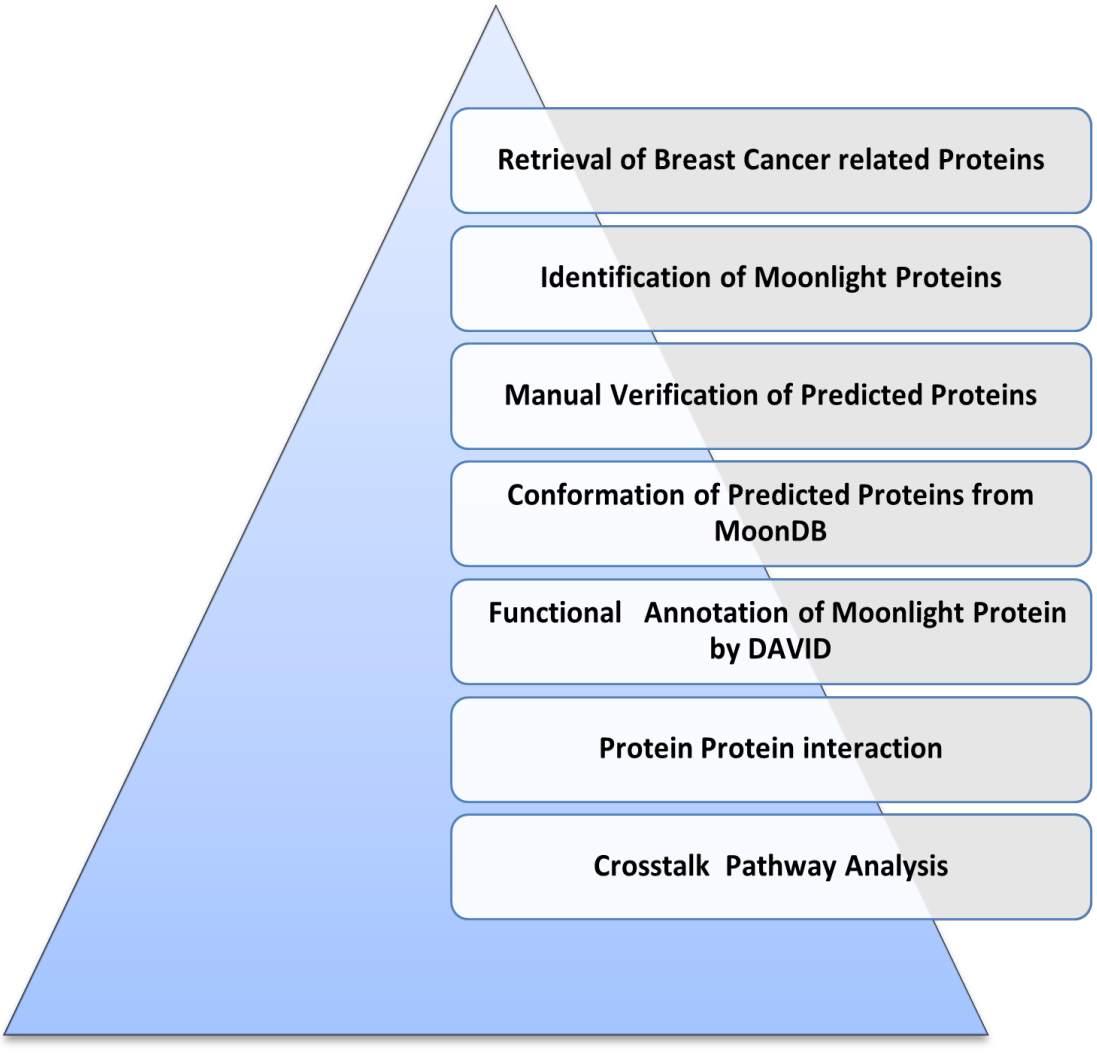


FIGURE 3.1: Flow Chart Methodology Conducted for the research

3.1 Retrieval of Breast Cancer Related Proteins

3.1.1 Retrieval of Proteins Related to Breast Cancer from UniProt and COREMINE

Candidate Proteins related to Breast cancer were retrieved by using COREMINE that use PubMed, OMIM, gene bank, SwissProt and multiple sources of information generated to answer the query. Mesh Term of "Breast Cancer" was used and query typed was disease protein association and the query key words were breast cancer.

3.1.2 Retrieval of Proteins from Literature

Literature related to breast cancer and associated protein was also searched from google scholar by typing the query breast cancer and proteins. The selection criteria also includes the year of publication of articles from 2000 and onward. In addition,to this literature related to breast cancer associated protein was also downloaded and save in folder with respective protein name.

3.1.3 Proteins Text Information from UniProt

For each protein, three distinct types of textual information were recovered. Each protein publication's titles were first listed. These titles were obtained directly from the UniProtKB entry for the protein's list of "PUBLICATIONS." The PubMed ID of each article was used as the database search key to extract information. Third, the protein's functional description text, which is found in the UniProtKB entry for the protein's function subsection in the "FUNCTION" section.

3.1.4 Comparison of Lists

Retrieved list of Protein from Coremine and UniProt were manually compared to cross verify the proteins among this list. One comprehensive list of Proteins was prepared by comparing these two list.

3.2 Identification of Moonlight Proteins

DextMP is a tool that is used to extract the moonlight proteins from text data. It is based on data mining . It uses textual information about the target proteins, these have broad application. UniProt KB was refined by removing the accession no, gene name and only the UniProt ID and protein name was kept. The prepared proteins list was uploaded to DextMP and run. A comprehensive list as mentioned in appendix I was uploaded in Dext MP and was run to find moonlight proteins.

3.3 Manual Verification of Predicted Proteins

Manual checking of the predicted moonlight protein was performed in two steps. First by using UniProtKB's functional description and quick searches of publication titles. If the protein had two different functions, it can be inferred from both textual information. Manual Checking-2, which involved a through analysis of the protein's literature. The two functions of the proteins were verified as distinct from one another in this final step by reviewing the literature.

3.4 Conformation of Predicted Proteins from Moon DB

After manual verification of predicted proteins from DextMP. Proteins from this list were searched in MoonDB. The purpose of this step was to find the status of already reported moonlight proteins.

3.5 Functional Annotation of Moonlight Protein by DAVID

DAVID was used to carry out the functional annotation of predicted moonlight. A well-known web server and web service for functional annotation and enrichment analyses of protein lists are included in the DAVID resource system for bioinformatics. It includes an extensive knowledgebase and several tools for performing

functional analyses [1]. A thorough description of these instruments has been developed in the Supplemental Information. List of predicted moonlight was uploaded and David was run by selecting protein UniProt ID and the gene list. Gene conversion tool was used to determine identifier type as protein. Since these are human gene. Homosapiens species were selected. There are certain parameters set to run the David Gene Ontology. Highest classification stringency was selected to screen the function of moonlight proteins.The higher stringency setting generates less functional group with more strongly related genes in each group so that more gene will be unclustered. Highest classification stringency was selected.

3.6 Protein Protein Interaction

Protein-protein interaction plays key role in predicting the protein function of target protein.

3.6.1 FunCoup

FunCoup is an acronym for functional coupling. A framework called Funcoup is used to identify functional couplings all over the genomes of 21 model species. functional coupling, also known as functional association, is a common word for association that includes both direct physical contact and more abstract forms of direct or indirect contact, such as regulatory contact or involvement in the same pathway or process. List of moonlight breast cancer proteins was uploaded and fun coup was run by selecting Homo sapiens as species. UniProt was selected as gene identifier and breast cancer tissue was selected in filter by tissue tab.

3.7 Crosstalk Pathway Analysis

Pathway Analysis with crosstalk, or PathwAX, is a web service for pathway annotation based on crosstalk. A framework for genome-extensive functional association networks.Select the specie as Homo sapiens. Submit the query breast cancer moonlight protein ID list. The IDs should be separated by commas or spaces when doing a search involving several genes. Select the pathway (KEGG or Reactome).

Chapter 4

Result and Discussion

4.1 Retrieval of Proteins from UniProt + COREMINE

A list of 2246 proteins involved in breast cancer was retrieved from the UniProt KB and Coremine against the query Breast Cancer Proteins. List was downloaded in excel that contains the information accession number ,gene name and protein name. List is available in appendex (appendix 1).

4.2 Identification of Moonlight Protein DeXTMP and Manual Verification

The refined list of 2246oproteins with UniProt Id and protein name when ran through the DextMP, predicted 84 proteins of breast cancer as moonlight (Table 4). Predicted moonlight proteins were manually scrutinized by checking the publication titles and the functional description in UnProt KB. The text information ,revealed the data whether two diverse functions are associated with one protein .These proteins were again manually verified through the literature evaluation of the proteins .This step was done to perform the different function of proteins that

are independent from each other. The function of 84 predicted moonlight proteins is available in table 4.1.

TABLE 4.1: Predicted Moonlight proteins

Sr. No.	**Breast Cancer Moonlight Protein**	**Protein Name**	**Gene Name**
1	Q9HCU9	Desmocollin 3	DSC3
2	P23381	Tryptophan–tRNA ligase	WARS1
3	P19525	Interferon-induced, double-stranded RNA-activated protein kinase	EF2AK2
4	P11511	Aromatase	CYP19A1
5	O95177	NADH dehydrogenase[ubiquinone] 1 alpha subcomplex subunit 3	GAS8-AS1 C16orf3
6	Q9H4B4	Serologically defined breast cancer antigen NY-BR-73	PLK3 CNK FNK PRK
7	P49639	Homeobox protein Hox-1F	HOX1F
8	Q9H093	Mutant early onset breast cancer susceptibility protein 2	NUAK2
9	Q15911	Zinc finger homeobox protein 3	ZFHX3
10	A6NNA2	Odontogenic ameloblast-associated protein (Apin)	SRRM3
11	P05109	calcium-binding proteinA8, Calgranulin	S100A8, CAGA CFAG ,MRP8
12	Q16678	Cytochrome P450 1 (Hydroperoxy icosatetraenoate dehydratase	CYP1B1
13	Q96EZ4	DAZ-associated protein 1	MYEOV OCM
14	O43175	D-3-phosphoglycerate dehydrogenase	PHGDH PGDH3
15	Q17RR3	MICOS complex subunit MIC60	PNLPRP3
16	Q8TC94	Actin-like protein 9	ACTL9 HSD21
17	Q96ND0	Janus kinase and microtubule-interacting protein 1	FAM210A
18	Q9HCU9	Desmocollin 3	BRMS1
19	Q9Y4K0	Latrophilin-2	LOXL2
20	O75443	Alpha-tectorin	TECTA
21	Q9UM73	Latrophilin-2	ALK
22	P21397	Amine oxidase	MAOA
23	P08183	ATP-dependent translocase ABCB1	MAOA
24	Q9NR30	Latrophilin-2	DDX21

Sr. No.	**Breast Cancer Moonlight Protein**	**Protein Name**	**Gene Name**
25	P10809		HSPD1 HSP60
26	P15172	Myoblasts (Myogenic factor 3, Myf-3	HSPD1 HSP60
27	Q5HYI8	Rab-like protein 3	RABL3
28	Q9H093	Breast cancer susceptibility protein 2	NUAK2 ,SNARK
29	O75582	ribosomal protein S6 kinase A5I	RPS6KA5 MSK1
30	Q9P032	Latrophilin-2	NDUFAF4
31	Q02809	Procollagen-lysine,2-oxoglutarate 5-dioxygenase1,	PLOD1 LLH PLOD
32	P05109	Protein S100-A8 (Calgranulin-A)	S100A8 MRP8
33	A6NNA2	Odontogenic ameloblast-associated protein (Apin)	SRRM3
34	P80192	Mitogen-activated protein kinase	MAP3K9 MLK1 PRKE1
35	P11310	MCAD Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADM
36	Q6PJQ5	Neutral cholesterol ester hydrolase 1, NCEH,- (Acetylalkylglycerol acetylhydrolase, 2-acetyl MAGE hydrolase,	FOXR2 FOXN6
37	O43240	Kallikrein-10, EC 3.4.21.- (Normal epithelial cell-specific 1) (Protease serine-like 1)	KLK10
38	Q9UM73	Latrophilin-2	ALK
39	Q9NR30	Latrophilin-2	DDX21
40	P09429	High mobility group protein B1 (High mobility group protein 1, HMG-1)	HMGB1 HMG1
41	P16444	Dipeptidase 1 (Beta-lactamase	DPEP1
42	P46527	Cyclin-dependent kinase inhibitor 1B	CDKN1B
43	Q96ND0	Janus kinase and microtubule-interacting protein 1	FAM210A
44	P50747	Biotin–protein ligase	HLCS
45	P08183	ATP-dependent translocase ABCB1	ABCB1 MDR1 PGY1
46	Q17RS7	Flap endonuclease GEN homolog 1	GEN1
47	Q15139	D1, Protein kinase	PRKD1

Sr. No.	**Breast Cancer Moonlight Protein**	**Protein Name**	**Gene Name**
48	Q15139	Serine/threonine-protein kinase D1	PRKD1 PKD PKD1 PRKCM
49	P58166	Gasdermin-D	NHBE
50	P11926	Ornithine decarboxylase, ODC	ODC1
51	P15514	AR Amphiregulin (Colorectum cell-derived growth factor, CRDGF)	AREG AREGB SDGF
52	Q9Y4K0	Latrophilin-2	LOXL2
53	O00273	DFF-45(DNAfragmentation factor subunit alpha, a 45 kDa subunit of DFF) (Inhibitor of CAD, ICAD)	DFFA DFF1 DFF45 H13
54	Q6PJQ5	Neutral cholesterol ester hydrolase 1	FOXR2 FOXN6
55	P61158	Actin-like protein 3	ACTR3 ARP3
56	P47712	Cytosolic phospholipase A2, cPLA2	PLA2G4A CPLA2 PLA2G4
57	O95177	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit	GAS8-AS1 C16orf3
58	P58166	Gasdermin-D (Gasdermin domain-containing protein 1)	NHBE
59	P01563	IFN-alpha-2, or interferon	FNA2 FNA2A FNA2B FNA2C
60	Q5HYI8	Rab-like protein 3	RABL3
61	Q9P032	Latrophilin-2	NDUFAF4 C6orf66 HRPAP20 HSPC125 My013
62	P02489	Alpha-crystallin A chain (Heat shock protein beta-4, HspB4)	CRYAA CRYA1 HSPB4
63	Q16678	Cytochrome P450 1B1, EC 1.14.14.1 (CYPIB1)	CYP1B1
64	P45452	Collagenase3,	MMP13
65	Q17RR3	MICOS complex subunit MIC60 (Mitochondrial inner membrane protein) (Mitofilin)	PNLPRP3
66	O00273i	(DNA fragmentation factor 45 kDa subunit, DFF-45i)	DFF1 DFF45

Sr. No.	Breast Cancer Moonlight Protein	Protein Name	Gene Name
67	Q17RS7	Flap endonuclease GEN homolog 1	GEN1
68	P09429	High mobility group protein B1 (High mobility group protein 1, HMG-1)	HMGB1 HMG1
69	P80192	Mitogen-activated protein kinase kinase kinase 9, (Mixed lineage kinase 1)	MAP3K9 MLK1 PRKE1
70	Q9H4B4	Serologically defined breast cancer antigen NY-BR-73	PLK3 CNK FNK PRK
71	P01574	Interferon beta, IFN-beta (Fibroblast interferon)	FNB1 FB FNB
72	Q8IZY5	Tensin-4 (C-terminal tensin-like protein)	BLD BRCC2
73	O75582	Ribosomal protein S6 kinase alpha-5, S6K-alpha-5	RPS6KA5 MSK1
74	P17612	cAMP-dependent protein kinase	PKACA
75	Q6R6M4	Ubiquitin carboxyl-terminal hydrolase	USP17L2 USP17M
76	Q9BTC8	Uncharacterized protein C5orf34	MTA3 KAA1266
77	P47712	Cytosolic phospholipase A2, cPLA2	PLA2G4
78	Q02809	Procollagen-lysine	PLOD
79	O43240	Kallikrein-10	KLK10 NES1 PRSSL1
80	O75417	DNA polymerase theta (also known as DNA polymerase eta)	POLQ POLH
81	O43175	3-PGDH, 2-oxoglutarate reductase	PHGDH PGDH3
82	P12273	Prolactin-inducible protein	PP GCDFP15 GPP4
83	P23381	Tryptophan–tRNA ligase, cytoplasmic	WARS1 F53 WARS WRS
84	P46527	Cyclin-dependent kinase inhibitor 1B	CDKN1B KP1

4.3 Cross Validation from MoonDB

Out of 84 moonlight proteins, only 58 were present in MoonDB .The proteins present in MoonDB were verified and remaining 27 proteins were catergorized as predicted proteins.The list of 27 predicted proteins are given in table A.1.

TABLE 4.2: Predicted Moonlight Protein list

MoonDB ID	**UniprotKB AC**	**Protein Full Name (MDB)**	**Moonlight Protein**
3	Q6UWE0	IE3 ubiquitin-protein ligase LR-SAM1	Eukaryotic translation initiation factor 2-
4	O00499	IMyc box-dependent-interacting protein 1	Aromatase, EC (Estrogen synthase)
7	P62256i	IUbiquitin-conjugating enzyme E2 Hi	Homeobox protein Hox-A1 (Homeobox protein Hox-1F)
9	P04792i	IHeat shock protein beta-1i	Zinc finger homeobox protein
13	P02511	IAlpha-crystallin B chain	DAZ-associated protein 1
16	P60520	IGamma-aminobutyric acid receptor-associated protein-like 2	Actin-like protein 9
18	Q9UMS4i	IPre-mRNA-processing factor 19i	Desmocollin 3
20	O15287	I Fanconi anemia group G protein	Alpha-tectorin
22	Q00613	Heat shock factor protein 1	Amine oxidase
25	O75674	ITOM1-like protein 1	60 kDa heat shock protein, ((Heat shock protein 60, HSP-60, Hsp60) (Mitochondrial matrix protein P1) (P60 lymphocyte protein)
26	Q13492	IPhosphatidylinositol-binding clathrin assembly protein	myoblast determination protein 1 (bHLHc1) (Myogenic factor 3, Myf-3)
35	Q15038	IDAZ-associated protein 2	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial, MCAD,
41	P27361	IMitogen-activatedprotein kinase 3	Dipeptidase 1, EC 3.4.13.19 ((Microsomal dipeptidase) (Renal dipeptidase,I hRDP)
44	P0DP23	Calmodulin-1	(Biotin apo-protein ligase)

MoonDB ID	UniprotKB AC	Protein Full Name (MDB)	Moonlight Protein
50	P27540	Aryl hydrocarbon receptor nuclear translocator	Ornithine decarboxylase, ODC,
51	Q16659	Mitogen-activated protein kinase 6	Amphiregulin, AR (Colorectum cell-derived growth factor, CRDGF)
55	Q13064	Probable E3 ubiquitin-protein ligase makorin-3	Actin-related protein 3 (Actin-like protein 3)i
59	O00204	Sulfotransferase family cytosolic 2B member 1	Interferon alpha-2, IFN-alpha-2 (Interferon alpha-A, LeIF A)
62	P28702	Retinoic acid receptor RXR-beta	Alpha-crystallin A chain (Heat shock protein beta-4, HspB4)
64	P68036	Ubiquitin-conjugating enzyme E2 L3	Collagenase 3, EC 3.4.24.- (Matrix metalloproteinase-13, MMP-13)
71	Q92997	Segment polarity protein dishevelled homolog DVL-3	Interferon beta, IFN-beta (Fibroblast interferon)
72	P00734	Prothrombin	Tensin-4 (C-terminal tensin-like protein)
74	Q05086	Ubiquitin-protein ligase E3A	cAMP-dependent protein kinase .
75	Q9BWF3	RNA-bindinG protein 4	Ubiquitin carboxyl-terminal hydrolase 17 (Deubiquitinating enzyme 17-like protein 2) (DUB-3, deubiquitinating protein)
76	P19971	Thymidine phosphorylase	Uncharacterized protein C5orf34
80	O00308	NEDD4-like E3 ubiquitin-protein ligase WWP2	DNA polymerase theta
82	Q9Y6X0	SET-binding protein	Prolactin-inducible protein (Gross cystic disease fluid protein 15, GCDFP-15)

4.4 Functional Annotation by David Tool

David tool was used to execute functional annotation. The results acquired after functional annotation were in the form of clusters. Functional categories based on a coexistence with a group of protein helped to unravel new bio- pathway processes. If proteins share related set of those terms, they are most likely involved in similar biological mechanisms.

The result of David tool gave us different clusters, it generated 5 cluster for Breast cancer moonlight proteins with the enrichment score ≥ 1 and p value as ≤ 0.01 a threshold. Out of 84 protein list 5 cluster of 55 proteins were generated .Mainly these 5 clusters were annotated as Protein Kinase, Catalytic domain binding site, Serine threonine protein kinase, immunity, infections and trans membrane proteins. Among the 5 cluster 2 clusters were highest priority due to enrichment score greater than 1.

4.4.1 Functional Annotation Clustering

Cluster 1 was categorize as Protein Kinase.

Annotation Cluster 1

The functions of Protein in this cluster were protein Kinase ATP binding site domain. Protein kinase, catalytic domain and Protein kinase-like domain. The enrichment score of this cluster was 3.58.

Annotation Cluster 1	Enrichment Score: 3.58	G		Count	P_Value	Benjamini
INTERPRO	Protein kinase, ATP binding site	RT		8	7.4E-5	1.3E-2
UP_SEQ_FEATURE	DOMAIN:Protein kinase	RT		8	3.2E-4	4.5E-2
INTERPRO	Protein kinase, catalytic domain	RT		8	3.5E-4	2.0E-2
INTERPRO	Protein kinase-like domain	RT		8	5.9E-4	2.5E-2

FIGURE 4.2: Functional annotation of moonlight breast cancer associated protein

5 Cluster(s) Download File

Annotation Cluster 1	**Enrichment Score: 3.58**			**Count**	**P_Value**	**Benjamini**
INTERPRO	Protein kinase, ATP binding site	RT		8	7.4E-5	1.3E-2
UP_SEQ_FEATURE	DOMAIN:Protein kinase	RT		8	3.2E-4	4.5E-2
INTERPRO	Protein kinase, catalytic domain	RT		8	3.5E-4	2.0E-2
INTERPRO	Protein kinase-like domain	RT		8	5.9E-4	2.5E-2
Annotation Cluster 2	**Enrichment Score: 3.09**			**Count**	**P_Value**	**Benjamini**
INTERPRO	Serine/threonine-protein kinase, active site	RT		7	1.9E-4	1.6E-2
GOTERM_MF_DIRECT	protein serine/threonine kinase activity	RT		7	6.4E-4	6.6E-2
SMART	S_TKc	RT		7	9.2E-4	4.3E-2
UP_KW_MOLECULAR_FUNCTION	Serine/threonine-protein kinase	RT		7	3.8E-3	1.5E-1
Annotation Cluster 3	**Enrichment Score: 0.94**			**Count**	**P_Value**	**Benjamini**
UP_KW_BIOLOGICAL_PROCESS	Innate immunity	RT		5	5.2E-2	6.0E-1
GOTERM_BP_DIRECT	innate immune response	RT		5	6.9E-2	1.0E0
UP_KW_BIOLOGICAL_PROCESS	Immunity	RT		5	4.2E-1	1.0E0
Annotation Cluster 4	**Enrichment Score: 0.65**			**Count**	**P_Value**	**Benjamini**
KEGG_PATHWAY	Hepatitis C	RT		3	1.3E-1	1.0E0
KEGG_PATHWAY	Influenza A	RT		3	1.5E-1	1.0E0
KEGG_PATHWAY	Kaposi sarcoma-associated herpesvirus infection	RT		3	1.8E-1	1.0E0
KEGG_PATHWAY	Coronavirus disease - COVID-19	RT		3	2.4E-1	1.0E0
KEGG_PATHWAY	Herpes simplex virus 1 infection	RT		3	6.0E-1	1.0E0
Annotation Cluster 5	**Enrichment Score: 0.02**			**Count**	**P_Value**	**Benjamini**
GOTERM_CC_DIRECT	integral component of membrane	RT		10	9.6E-1	1.0E0
UP_KW_DOMAIN	Transmembrane helix	RT		10	9.7E-1	1.0E0
UP_KW_DOMAIN	Transmembrane	RT		10	9.7E-1	1.0E0

FIGURE 4.1: Functional association clusters of moonlight proteins associated with breast cancer

Annotation Cluster 2

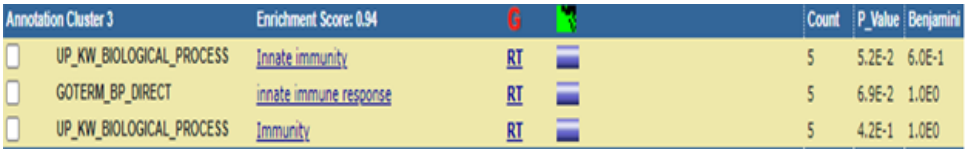
The Cluster was categorized as Serine/threonine-protein kinase group. The enrichment score was 3.09 (fig 4.3). The functions of these Proteins were group as Serine/threonine-protein kinase, activity, Serine/Threonine protein kinases, catalytic domain and Serine/threonine-protein kinase.

Annotation Cluster 2	**Enrichment Score: 3.09**			**Count**	**P_Value**	**Benjamini**
INTERPRO	Serine/threonine-protein kinase, active site	RT		7	1.9E-4	1.6E-2
GOTERM_MF_DIRECT	protein serine/threonine kinase activity	RT		7	6.4E-4	6.6E-2
SMART	S_TKc	RT		7	9.2E-4	4.3E-2
UP_KW_MOLECULAR_FUNCTION	Serine/threonine-protein kinase	RT		7	3.8E-3	1.5E-1

FIGURE 4.3: Functional annotation of moonlight breast cancer associated proteins

Annotation Cluster 3

The results in figure 4.4 indicates annotation of cluster 3 plays role in immunity. The enrichment score was 0.97 and non significant. The functions of these genes are involved in innate immunity, innate immunity response and immunity.

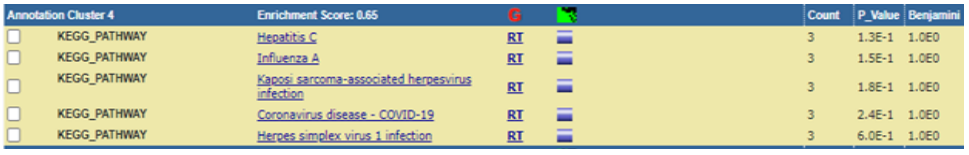


Annotation Cluster 3	Enrichment Score: 0.94			Count	P_Value	Benjamini
UP_KW_BIOLOGICAL_PROCESS	Innate immunity	RT		5	5.2E-2	6.0E-1
GOTERM_BP_DIRECT	innate immune response	RT		5	6.9E-2	1.0E0
UP_KW_BIOLOGICAL_PROCESS	Immunity	RT		5	4.2E-1	1.0E0

FIGURE 4.4: Functional annotation of moonlight breast cancer associated proteins

Annotation Clustering 4

The functions of moon light proteins are associated with multiple diseases. Cluster 4 was associated with multiple diseases. The functions of moon light proteins are associated with multiple diseases as Hepatitis C, Influenza A, Kaposi Sarcoma-Associated herpesvirus infection, Coronavirus disease COVID-19 and Herpes simplex virus 1 infection. The enrichment score was 0.650

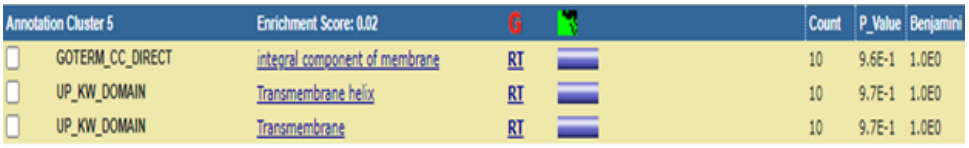


Annotation Cluster 4	Enrichment Score: 0.65			Count	P_Value	Benjamini
KEGG_PATHWAY	Hepatitis C	RT		3	1.3E-1	1.0E0
KEGG_PATHWAY	Influenza A	RT		3	1.5E-1	1.0E0
KEGG_PATHWAY	Kaposi sarcoma-associated herpesvirus infection	RT		3	1.8E-1	1.0E0
KEGG_PATHWAY	Coronavirus disease - COVID-19	RT		3	2.4E-1	1.0E0
KEGG_PATHWAY	Herpes simplex virus 1 infection	RT		3	6.0E-1	1.0E0

FIGURE 4.5: Functional annotation of moonlight breast cancer associated proteins

Annotation Clustering 5

This Cluster was associated with activities related to membrane. The enrichment score was 0.02 (non-significant). The functions of these proteins are involved in integral component of membrane, trans membrane helix and trans membrane.



Annotation Cluster 5	Enrichment Score: 0.02			Count	P_Value	Benjamini
GOTERM_CC_DIRECT	integral component of membrane	RT		10	9.6E-1	1.0E0
UP_KW_DOMAIN	Transmembrane helix	RT		10	9.7E-1	1.0E0
UP_KW_DOMAIN	Transmembrane	RT		10	9.7E-1	1.0E0

FIGURE 4.6: Functional annotation of moonlight breast cancer associated proteins

4.5 Protein Protein Interaction

4.5.1 FunCoup

Protein network revealed that Actin beta protein with 22 degree and PFC: 0.958 shows maximum interaction with other proteins. Aspartate transcarbamylase has a degree 3 interaction with PFC: 0.972. Serum response carbamoyl-phosphate synthetase 2 has degree 1 interaction (PFC: 0.994). Alpha kinase 2 of the eukaryotic translation initiation factor has a degree of 1 and a PFC of 1.00. The results of protein protein interactions for the query of 84 Breast cancer moonlight proteins cut-off of value of pfc > 0.25 showed 42 proteins interacted and with 169 links as mentioned in fiqure 4.7.

FIGURE 4.7: Protein Network

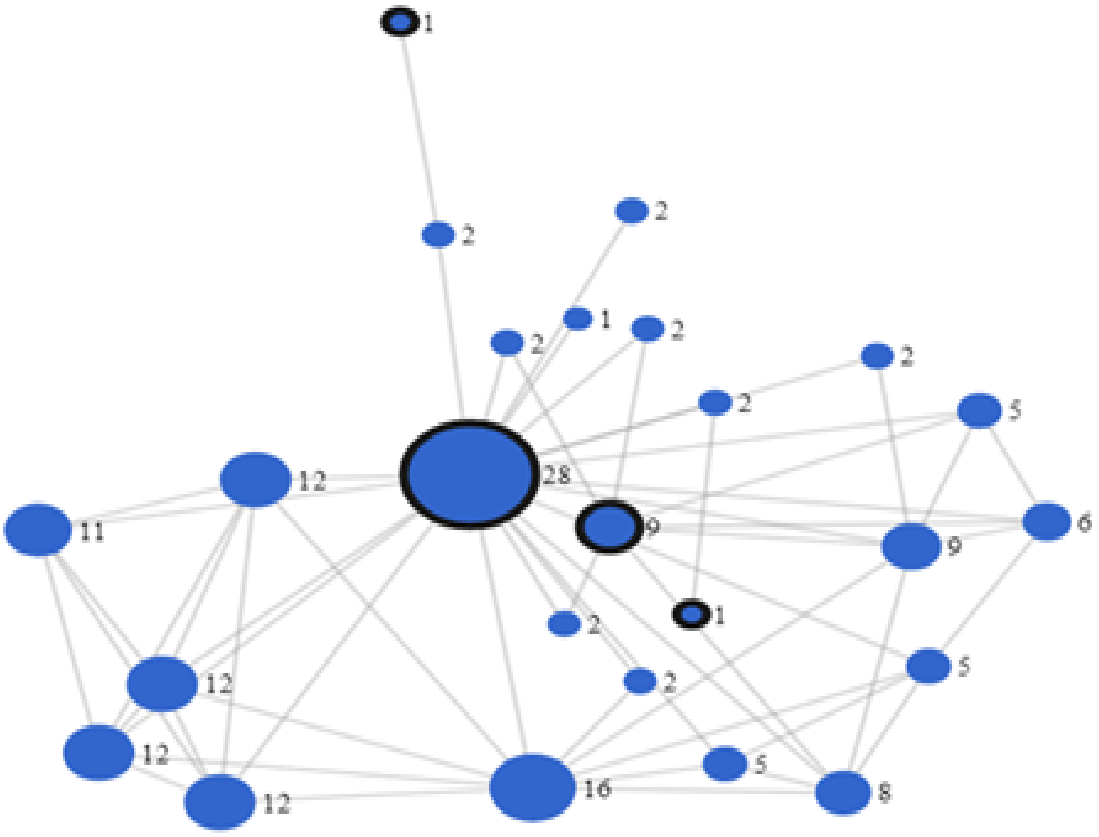


FIGURE 4.8: Fun Coup Protein Interactions

Cell cytoskeleton is mostly made up of the specific proteins actin.

It includes the six isoforms that are unique to certain cell types [109]. Since, we have proposed that it could operate as a biomarker intial stage cancer.Through integration into the normal cellular Factin network and altered actin binding protein interactions, aberrant actin subunit expression can provide cell withenhanced capacity for proliferation,migratory ability,and chemoresistance [110].

Since abnormal actin isoform expression has been observed in a variety of tumors, it can be proposed that it could operate as a biomarker for initial-phase cancer. Through integration the normal cellular F-actini network and altered actin binding proteini interactions, aberrant actin subunit expression can provide cells with enhanced capacity for proliferation, migratory ability, and chemoresistance. The function of each actin isoform has been clarified by a number of knockout studies [111].

The Circle size indicate the Protein node degree (links in the network) the nodes (Blue colour circle represent the protein while the grey edges (lines) indicate their function association. Black Border around the circle indicate the proteins.Highest PFC value indicate stronger probability of interaction .

While ACTG1, ACTG2 and ACTA2 knockouts resulted in a living organism with several muscular and cardiac abnormalities, Actin knockouts were embryonic, perinatal. Increased tumour metastatic potential is linked to the presence of fibroblasts that express ACTA1 in the stroma of prostate cancer [112].

High expression of ACTA1 is linked to a lower survival time in oral squamous cell carcinoma . A biomarker called ACTA1 has also been linked to chemoresistance in basal-like breast cancer [113].

The ACTC1 protein promotes oncogenesis via a variety of potential pathways. One route might be via annexins, which are phospholipidbinding, Ca2+-dependent proteins involved in cell proliferation, death, and vesicle trafficking. Smaller existence, carcinogenesis, and the development of malignant ovarian cancer are all associated with annexin expression. Notably, annexins and ACTA1 participate in a variety of physical interactions [114].

RhoA kinase (ROCK) is another enzyme downstream from RhoA that is active, and it inhibits the ability of the protein cofilin to cleave F-actin [115].

This in turn will encourage the production of stress fibres. Stress fibres and F-actin, which are involved in cytoskeletal stability, cell survival, migration, and adhesion, are more easily formed when ACTA1 subunits are abnormally produced in the cytoplasm.The expression of the ACTA1 gene is modified in several malignancies, according to bioinformatics studies [116].

For instance, in head and neck squamous cell carcinomas, it is downregulated, which is linked to carcinogenesis. In patients with head and neck squamous cell carcinoma, this decline in ACTA1 expression might be used as a predictive indicator for a poor clinical outcome .Through DNA hypermethylation, it is also downregulated in aggressive carcinogenesis-related diseases such as colorectal cancer, prostate cancer, and pancreatic adenocarcinoma [117].

The actin present in smooth-muscle in the blood vessels is encoded by the gene ACTA2. It serves as a contractile component of smooth muscle cells and is largely found in their microfilament bundles (i.e., during homoeostasis of body temperature).

It's interesting to note that elevated ACTA2 expression is linked to more distant metastases and a worse prognosis for lung adenocarcinoma, , early-onset colorectal cancer,HER2+ breast cancer , and non-small cell lung cancer . Additionally, pancreatic cancer, colorectal cancer, and head and neck cancer "stimulated" myofibroblastic cancer-associated fibroblasts have been shown to acquire ACTA2 [118].

Myofibroblasts' ability to contract mechanically depends on ACTA2, and higher levels of this protein are a sign that these cancer-associated fibroblasts have undergone an oncogenic transformation [119].

In a diverse range of biological processes, including cell growth, cell proliferation, cellular growth, immune response, gene expression, maintenance of cell stability, and cytoskeletal formation, ACTB, a widely expressed cytoskeletal protein, participates. The dysregulation of ACTB may contribute to the pathology of cancer, according to these roles. Intriguingly, elevated ACTB levels have been seen in sarcoma, colon adenocarcinoma, and hepatoma cell lines, three highly metastatic cancer types [120].

Normal cells cannot migrate without this isoform, as evidenced by the increased expression of genes that control myosin activity, increased creation of focal adhesions, and reduced membrane projections at the top edge of transferring cells when this isoform is knocked down . Actin's dynamic polymerization has been demonstrated to support tumour malignancy in a few malignancies. G-actin levels were reported to have dropped and F-actin levels to have increased in three cancer cell lines with strong ACTB expression [121].

According to the theory, amount of actin polymerization is required in cancer cell invasion to the surrounding tissues, these cells exhibit strong metastatic potential and invasion. Maintaining cell growth potential may require the assistance of ACTB. The work of *Kwiatkowski et al.* [122]. demonstrated that SET Domain

Containing 3 (an actin histidine methyltransferase) loss of ACTB This increased F-actin breakdown that leads to the instability of the actin cytoskeleton drains the cell of a significant amount of energy (up to 50% of total ATP consumption) and may force the cell to switch to anaerobic metabolism, which boosts lactate generation [123].

The increased ATP need and conversion of cellular metabolism to glycolysis are caused by the expedited breakdown of the hypomethylated F-actin fibres. Cancer cell with increase metabolic demand shows the interaction between F-actin and stability in ATP consumption prove to be crucial process with increase metabolic demand [124].

The cardiac sarcomere thin filaments, which are in charge of contracting the heart muscle, are mostly made up of a protein that is encoded by the ACTC1 gene . There is proof that the cardiac actin isoform is expressed at the earliest stages of mammalian neurodevelopment, despite the fact that it is primarily produced in the heart and less so in skeletal muscle. It's interesting to note that numerous cancer types, including brain, head and neck, bladder, urothelial, prostate, lung, and breast cancers, have recurrent ACTC1 expression [124].

Additionally, it has been previously shown that ACTC1 is a hub gene that imparts chemoresistance in a variety of tumours and that multi-drug resistant breast cancer cells express it at higher levels . The cytoskeleton protein -actin, which is abundant in the auditory hair cells of the cochlea and operates in non-muscle cells, is encoded for by the gene ACTG1. The internal cell motility of hair cells is influenced by this actin isoform, which is also necessary for the shape and functionality of the stereocilia .The motility of SH-EP neuroblastoma cells is hindered when this isoform is suppressed [125].

Interesting research by *Dong et al* [126] revealed that skin cancer tissue expresses ACTG1 at significantly greater levels. The rate of filament turnover affects cancer cell migration and invasion as well as the mitotic stress response Additionally, it has been shown that breast cancer cells and polyploidal large tumour cells both have more stress fibres with increased thickness and length. Additionally,

the actin cytoskeletal components in these cells are upregulated, which leads in stronger gross-tumor rheological characteristics and improved migratory capacity [127].

The pleiotropic proteins known as eukaryotic translation elongation factors 1 alpha, or eEF1A1 and eEF1A2, are widely expressed in human tumors such as breast cancer, ovarian cancer, and lung cancer. In addition to regulating the cytoskeleton and acting as a chaperone, eEF1A1 also regulates cell division and death. As evidenced by the fact that overexpressing eEF1A2 causes cellular transformation and the development of tumours in nude mice, eEF1A2 protein, on the other hand, encourages oncogenesis. The eEF1A2 protein promotes cancer by stimulating phospholipid signalling, activating Akt-dependent cell migration, and modifying actin. However,60 inactivation of eEF1A proteins promotes apoptosis and causes immunodeficiency as well as neurological and muscular abnormalities. Finally, the interaction of eEF1A proteins with a number of viral proteins enhances viral replication while reducing apoptosis and promoting cellular transformation. In this review, the most recent research on eEF1A proteins is reviewed, showing that these proteins are crucial for the development of cancer, prevent apoptosis, and promote viral pathogenesis, among other human disorders. [128].

In BT549 human breast cancer cells and non-transformed Rat2 cells, the expression of eEF1A2 is sufficient to promote the development of filopodia. Furthermore, while the siRNA-mediated down-regulation of eEF1A2 decreases Akt activity, its expression is sufficient to activate Akt in a PI3K-dependent manner. eEF1A2 expression increases cell migration and invasion in the breast cancer cell line BT549 [128].

This suggests that eEF1A2 controls oncogenesis through Akt and PI3K-dependent cytoskeleton remodelling .In actuality, eEF1A2 takes involved in the control of the signalling pathway for phospholipids. Phosphatidylinositols are membrane-bound, negatively charged phospholipids that play a role in the signaling cascades that control cell growth, survival, cytoskeleton structure, vesicular trafficking, andoncogenesis [129].

Phosphoinositols are made up of an inositol ring that has one or more OH groups at the 3, 4, and 5 positions that can be esterified with a phosphate group in any number of ways. These sites are phosphorylated by members of the PI3K, PI4K, and PI5K kinase families.Phosphatidylinositol 4-phosphate (PI4P) production in human cells is increased by overexpressing the eEF1A2 protein, which also increases total PI4Kactivity. Additionally, phosphatidylinositol-4 kinase III (a subfamily of PI4K), an enzyme that transforms phosphatidylinositol into PI4P, is directly interacted with and activated by eEF1A2.Phosphatidylinositol-4-kinase activity is decreased when eEF1A2 is knocked down using eEF1A2 [130].

Additionally, the production of 5-bisphosphate phosphatidylinositol-4, in the cytoplasm and at the plasma membrane is up-regulated by eEF1A2 expression. By binding to and activating PI4KIII, the ensuing rise in PI(4,5)P2 at the plasma membrane promotes the development of eEF1A2-induced filopodia [131].

As a result, eEF1A2 is implicated in actin remodelling and phosphatidylinositol signaling. Additionally, the high expression of eEF1A2 in plasmacytomas (PCT), which leads to the development of plasma cell neoplasms in both mice and human, was revealed by the gene expression profiling of primary mouse B cell lineage. Lastly, eEF1A2 expression is knocked down, which slows or prevents the IL-6-induced activation of the STAT3 and Akt signalling pathways [132].

This suggests that eEF1A2 is involved in the activation of STAT3 and Akt, which promotes cell proliferation, cell cycle progression, and the inhibition of apoptosis .Together, the PIK-Akt-STAT3 pathways, which have been well demonstrated to promote cellular transformation and oncogenesis, are activated by the eEF1A2 protein [133].

Utilizing comparative genomic hybridization and fluorescence in situ hybridization, it has been demonstrated that the genes at 20q13 are often increased in breast cancer .When metastatic and non-metastatic cell lines from the same parental rat mammary adenocarcinoma were screened separately, the metastatic cells had a 1.5-fold higher level of eEF1A expression than the non-metastatic cells [134].

TABLE 4.3: Network information: Fun coup

Protein	**Degree**	**PFC value**
Actin beta protein	28	0.958
Aspartate transcarbamylase	10	0.972
Serum response carbamoyl-phosphate synthetase 2	1	0.994
Alpha kinase 2 of the eukaryotic translation initiation factor	1	1.00
Ribosomal protein S3A	11	1
Ribosomal protein L23a	12	0.958
Ribosomal protein S4 X-linked	12	0.946
ribosomal protein L8	12	0.997
ribosomal protein S5	12	0.991
eukaryotic translation elongation factor 1 alpha 1	16	0.861
heat shock protein family D (Hsp60) member 1	5	1
heat shock protein family A (Hsp70) member 9	6	1
heat shock protein 90 alpha family class B member 1	9	1
valosin containing protein	2	0.950
Glucose-6-phosphate isomerase	2	0.949
Inorganic pyrophosphatase 1	2	0.907
Heat shock protein 90 alpha family class B member 1	9	0.907
Inorganic pyrophosphatase 1	0	0.99
ATP synthase F1 subunit beta	0	
Inorganic pyrophosphatase 1	0	0.907
Enolase 1	5	1
Tubulin beta class I	8	1

4.6 Cross Talk Pathway Analysis

PathwaxII was used to perform pathway crosstalk for the selected proteins. A threshold of q- value 0.01and KEGG v94.1 was selected for retrieving the results. Thirty-eight out of 84 input proteins were mapped on 38 pathways with significant

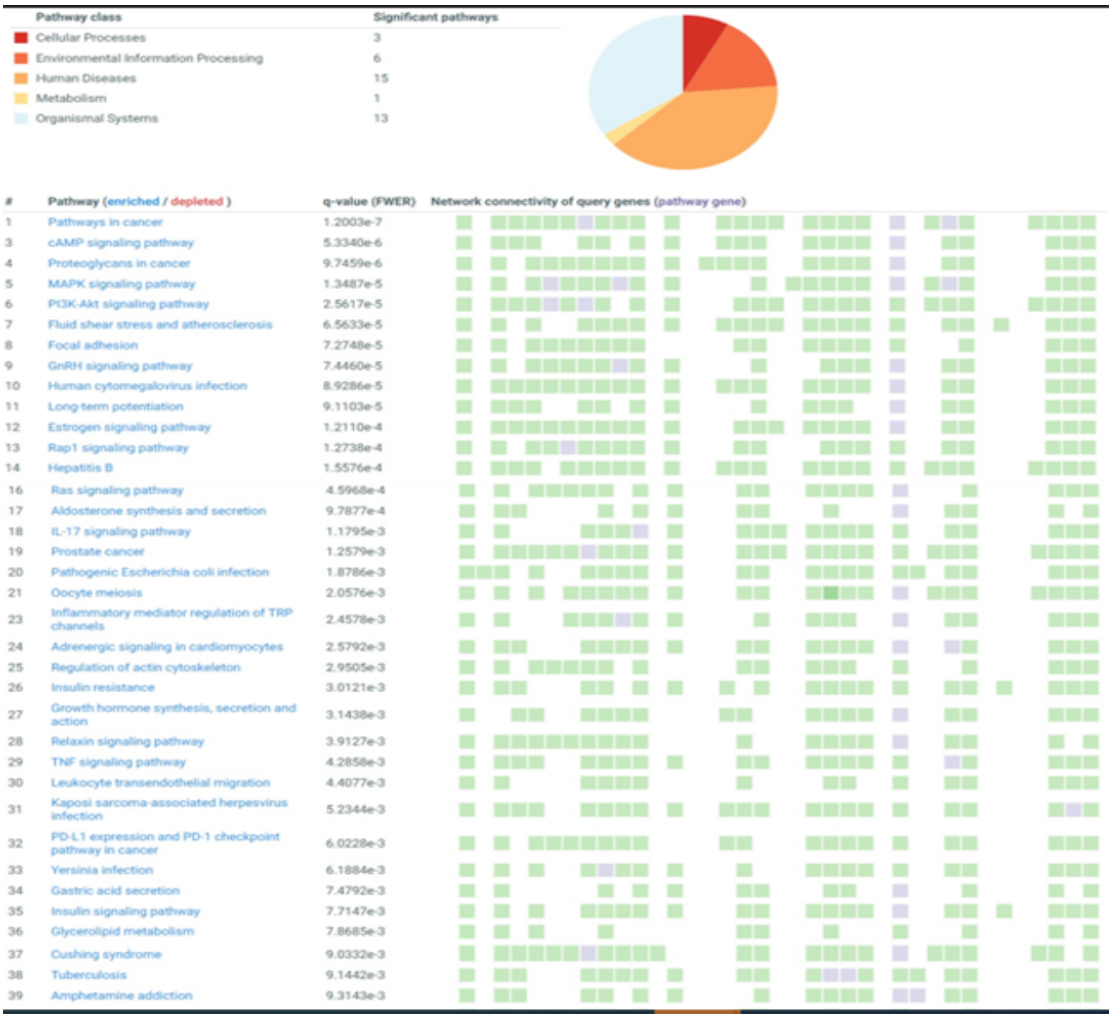


FIGURE 4.9: Cross talk pathways of moonlight proteins associated with breast cancer

crosstalk, the remaining 36 were not found in any pathway.The various colors of blocks in the fig 4.9 represents different status of presences and absences of that particular protein in that specific pathway, green (query protein having crosstalk links with other query and pathway proteins), white (no crosstalk links), purple (proteins shared by both query and pathway along with crosstalk links). To simplify the results the pathways were further divided into five classes of pathways i) Cellular processes (3), ii) Environmental information processing (6), iii) Human diseases (15), iv) Metabolism (1) and v) Organismal System (13).

Among cellular processes pathways (Focal adhesion, Oocyte meiosis and Regulation of actin cytoskeleton) only P17612 was shared between query and oocyte meiosis O43175, Q9NR30, P11926, Q15139, P46527, P61158, P05109, Q02809,

Q9P032, Q9HCU9, P10809, Q9H4B4, P17612, P09429, P23381, P19525 and P11310 showed significant crosstalk with almost all pathways.

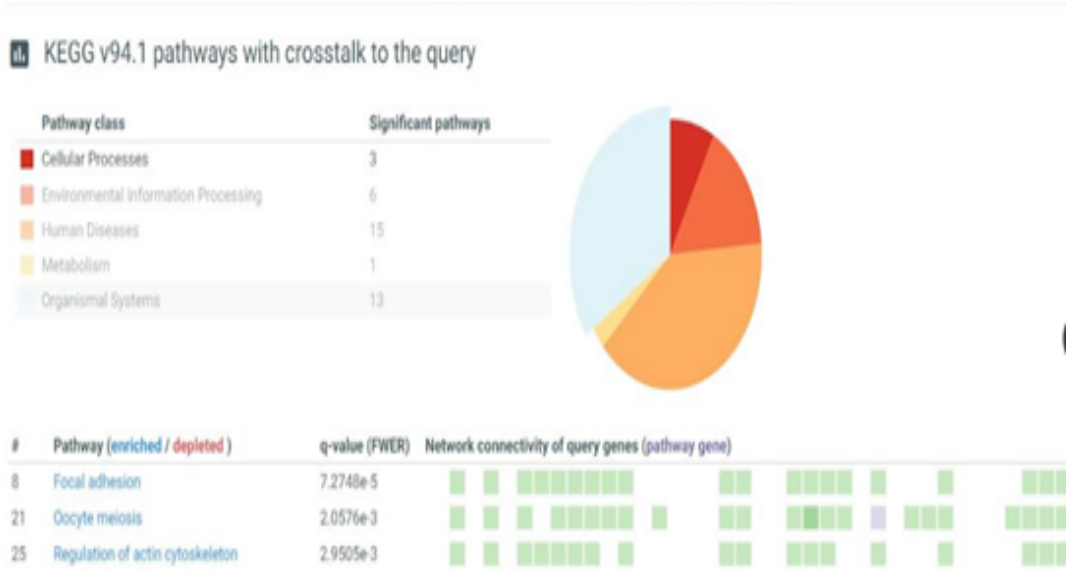


FIGURE 4.10: Cross talk pathways of moonlight proteins associated with breast cancer

Among the Human Diseases pathways (Pathways in cancer , Fluid shear stress and atherosclerosis, Proteoglycans in cancer, Focal, Human cytomegalovirus infection, Hepatitis B, Prostate cancer, Pathogenic Escherichia coli infection, Oocyte meiosis, Adrenergic signaling in cardiomyocytes, Regulation of actin cytoskeleton, Insulin resistance, Kaposi sarcoma-associated Relaxin signaling pathway, TNF signaling pathway, Leukocyte transendothelial migration, PD-L1 expression and PD-1 checkpoint pathway in cancer, Yersinia infection, cushing syndrome , Amphetamine addiction, Tuberculosis, and Amoebiasis). Only (P46527) Cyclin -dependent kinase inhibitor 1B cAMP-dependent protein kinase catalytic subunit alpha, and Ribosomal protein S6 kinase alpha-5, S6K-alpha-5, (Nuclear mitogen- and stress-activated protein kinase 1) was shared between query and pathways in cancer.

Only (P46527) Cyclin -dependent kinase inhibitor 1B cAMP-dependent protein kinase catalytic subunit alpha, and Ribosomal protein S6 kinase alpha-5, S6K-alpha-5, (Nuclear mitogen- and stress-activated protein kinase 1) was shared between query and pathways in cancer. O43175, Q9NR30, P11926, Q15139, P46527,

P61158, P05109, Q02809, Q9P032, Q9HCU9, P10809, Q9H4B4, P17612, P09429, P23381, P19525 and P11310 showed significant crosstalk.

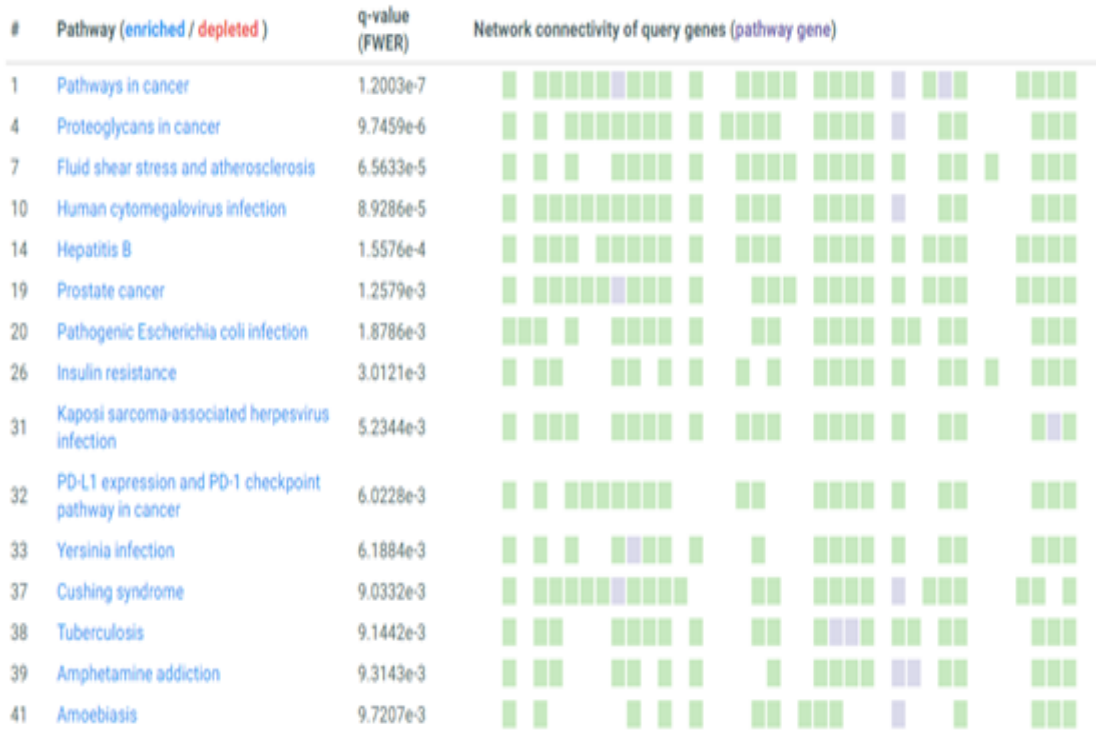


FIGURE 4.11: Cross talk pathways of moonlight proteins associated with breast cancer

In the Environmental information processing pathways, of cross talk. P17612 cAMP-dependent protein kinase catalytic subunit alpha, PKA C-alpha, was shared between query and camo signaling pathway, MAPK signaling pathway ,Rap1 signaling pathway, Ras signaling pathway, and TNF signaling pathway.

In the Environmental information processing pathways, of cross talk. O75582 ribosomal protein S6 kinase alpha-5, S6K was shared between query and TNF signaling pathway. O43175, Q9NR30, P11926, Q15139, P46527, P61158, P05109, Q02809, Q9P032,Q9HCU9, P10809, Q9H4B4, P17612, P09429, P23381, P19525 and P11310 showed significant crosstalk. Among the organismal systems P17612 cAMP-dependent protein kinase catalytic subunit alpha, PKA C-alpha, was shared between query protein and GnRH signaling pathway, Long-term potentiation estrogen signaling pathway, Aldosterone synthesis and secretion, IL-17 signaling pathway, Inflammatory mediator regulation of TRP channels, Adrenergic signaling in cardiomyocytes, growth hormone synthesis, O43175, Q9NR30, P11926, Q15139,

P46527, P61158, P05109, Q02809, Q9P032,Q9HCU9, P10809, Q9H4B4, P17612, P09429, P23381, P19525 and P11310 showed significant crosstalk. Among the organismal systems P17612 cAMP-dependent protein kinase catalytic subunit alpha, PKA C-alpha, was shared between query protein and GnRH signaling pathway.

Long-term potentiation estrogen signaling pathway, Aldosterone synthesis and secretion, IL-17 signaling pathway, Inflammatory mediator regulation of TRP channels, Adrenergic signaling in cardiomyocytes, growth hormone synthesis, secretion and action Relaxin signaling pathway, Leukocyte transendothelial migration, Gastric acid secretion and Insulin signaling pathway. O43175, Q9NR30, P11926, Q15139, P46527, P61158, P05109, Q02809, Q9P032, Q9HCU9, P10809, Q9H4B4, P17612, P09429, P23381, P19525 and P11310 showed significant crosstalk with almost all pathways.

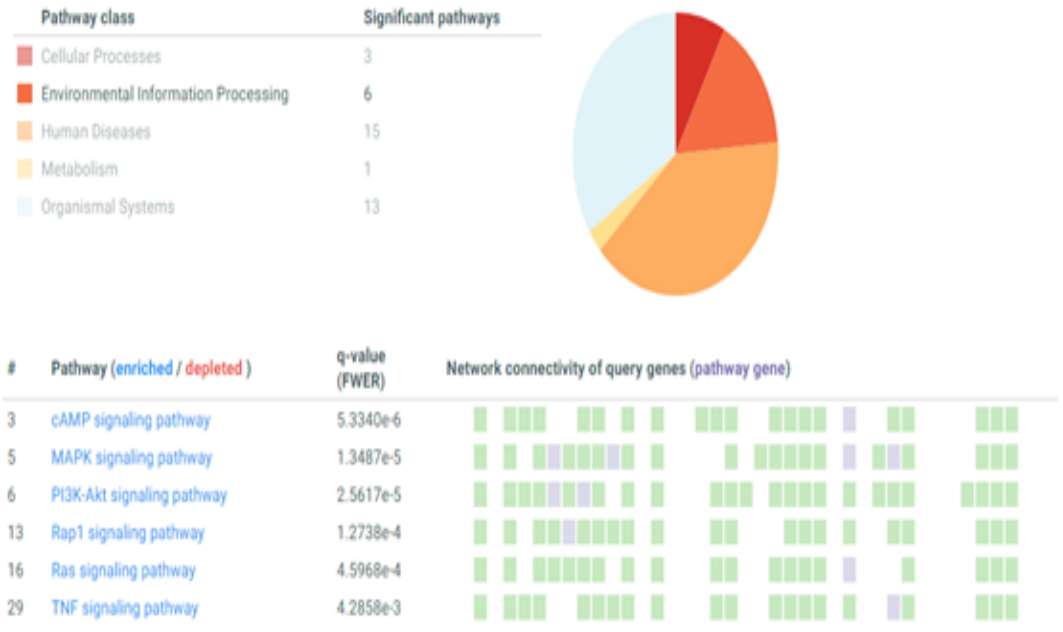


FIGURE 4.12: Cross talk pathways of moonlight proteins associated with breast cancer

Among the organismal systems P17612 cAMP-dependent protein kinase catalytic subunit alpha, PKA C-alpha, was shared between query protein and GnRH signaling pathway, Long-term potentiation estrogen signaling pathway, Aldosterone synthesis and secretion, IL-17 signaling pathway, Inflammatory mediator regulation

of TRP channels, Adrenergic signaling in cardiomyocytes, growth hormone synthesis, secretion and action Relaxin signaling pathway, Leukocyte transendothelial migration, Gastric acid secretion and Insulin signaling pathway. O43175, Q9NR30, P11926, Q15139, P46527, P61158, P05109, Q02809, Q9P032, Q9HCU9, P10809, Q9H4B4, P17612, P09429, P23381, P19525 and P11310 showed significant crosstalk with almost all pathways.

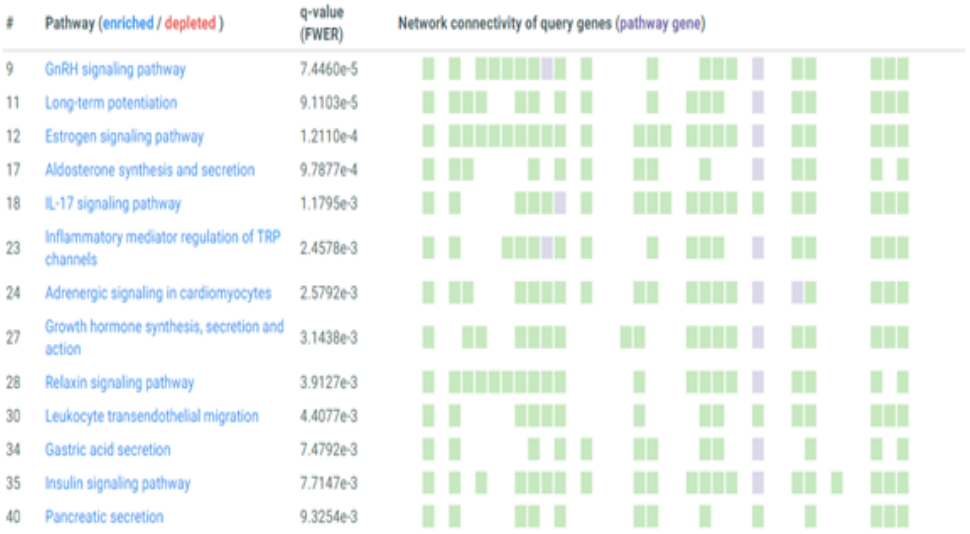


FIGURE 4.13: Cross talk pathways of moonlight proteins associated with breast cancer

In Metabolism Pathway there is no significantly shared pathway between query and Glycerolipid metabolism. O43175, Q9NR30, P11926, Q15139, P46527, P61158, P05109, Q02809, Q9P032, Q9HCU9, P10809, Q9H4B4, P17612, P09429, P23381, P19525 and P11310 showed significant crosstalk.

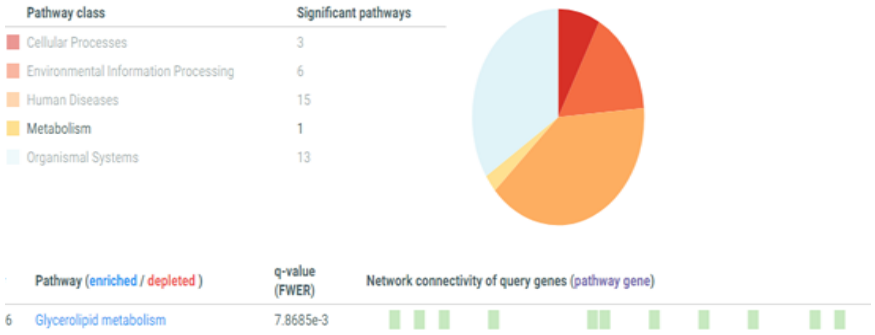


FIGURE 4.14: Cross talk pathways of moonlight proteins associated with breast cancer

Chapter 5

Conclusion and Future Recommendations

Moonlight Proteins involved in the development of numerous diseases, including infectious disorders and cancer. The understanding of the molecular processes that lead to breast cancer genesis has been increased by recent advancements in fundamental research. 10% of instances of familial breast cancer are caused by mutations in the p53, BRCA1, and PTEN genes. Moonlight Protein plays a key role in breast cancer tumours.

The first objective was to identify the proteins that act as moonlight proteins by using DextMP .For this purpose, a list of 2246 proteins involved in breast cancer was retrieved from the UniProt KB and Coremine against the inquiry Breast Cancer proteins. Three different types of textual data were extracted for each protein from UniProt KB. Retrieved list of proteins from Corermine and Uniprot were manually related to cross verify the protein among this list. The protein list was refer to DextMP .Manual Verification of Predicted Proteins using UniprotKB's functional description and quick searches of publication titles. Cross validiation from MoonDB was also conducted. The second objective was to functionally annotate the identified moonlight proteins involved in breast cancer. DAVID was used to carry out the functional annotation of predicted moonlight proteins and enrichment analyses. DextMP predicted 84 proteins as moonlight proteins. The

list of 2246 proteins was prepared and verified from Coremine, UniProt and literature. Out of 84 Moonlight proteins, only 58 were present in MoonDB. The proteins existing in Moon DB were verified, and remaining 27 proteins were categorised as predicted proteins. Functional annotation generated 5 clusters of 55 proteins. The third objective is to perform network analysis for proteins, interacting of moonlight to prioritise the significant MP proteins. Fourth objective was to perform cross talk . PathwaxII was used to perform pathway crosstalk for the selected moolight proteins. Thirty-eight out of 84 input proteins were mapped on 38 pathways with significant crosstalk. Pathways was mapped on five categories cellular processes, environmental information processing ,human diseases ,metabolism and organismal system .Out of the 84 proteins, 58 were verified from MoonDB, and the other 27 proteins were predicted to be moonlight breast cancer proteins that needed to be verified in vitro lab.

For Future recommendation Out of 84 proteins,27 proteins that was categorized as predicted needs to be valiadated in vitro study to explore their association with moonlight proteins in breast tumors. Moonlight proteins as a diagnostic target in tumor must also be explored to target multiple pathways. Moonlight proteins for diagnostic purpose in breast cancer can also be investigated to address the timely cancer therapy

Bibliography

[1] H. R. A. Douglas, "The hallmarks of cancer," *cell*, vol. 100, no. 1, pp. 57–70, 2000.

[2] D. Hanahan and R. A. Weinberg, "The hallmarks of cancer," *cell*, vol. 100, no. 1, pp. 57–70, 2000.

[3] J. Piatigorsky and G. J. Wistow, "Enzyme/crystallins: gene sharing as an evolutionary strategy," *Cell*, vol. 57, no. 2, pp. 197–199, 1989.

[4] M. Mani, C. Chen, V. Amblee, H. Liu, T. Mathur, G. Zwicke, S. Zabad, B. Patel, J. Thakkar, and C. J. Jeffery, "Moonprot: a database for proteins that are known to moonlight," *Nucleic acids research*, vol. 43, no. D1, pp. D277–D282, 2015.

[5] S. Hernandez, G. Ferragut, I. Amela, J. Perez-Pons, J. Pinol, A. Mozo-Villarias, J. Cedano, and E. Querol, "Multitaskprotdb: a database of multitasking proteins," *Nucleic acids research*, vol. 42, no. D1, pp. D517–D520, 2014.

[6] D. M. Ribeiro, G. Briere, B. Bely, L. Spinelli, and C. Brun, "Moondb 2.0: an updated database of extreme multifunctional and moonlighting proteins," *Nucleic acids research*, vol. 47, no. D1, pp. D398–D402, 2019.

[7] D. H. Huberts and I. J. van der Klei, "Moonlighting proteins: an intriguing mode of multitasking," *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1803, no. 4, pp. 520–525, 2010.

[8] N. Rahman, M. Shamsuzzaman, and L. Lindahl, "Interaction between the assembly of the ribosomal subunits: Disruption of 40s ribosomal assembly

causes accumulation of extra-ribosomal 60s ribosomal protein ul18/l5," *PloS one*, vol. 15, no. 1, p. e0222479, 2020.

[9] J. M. Dolezal, A. P. Dash, and E. V. Prochownik, "Diagnostic and prognostic implications of ribosomal protein transcript expression patterns in human cancers," *BMC cancer*, vol. 18, no. 1, pp. 1–14, 2018.

[10] B. Azab, N. Shah, J. Radbel, P. Tan, V. Bhatt, S. Vonfrolio, A. Habeshy, A. Picon, and S. Bloom, "Pretreatment neutrophil/lymphocyte ratio is superior to platelet/lymphocyte ratio as a predictor of long-term mortality in breast cancer patients," *Medical oncology*, vol. 30, no. 1, pp. 1–11, 2013.

[11] G. Molavi, N. Samadi, and E. Z. Hosseingholi, "The roles of moonlight ribosomal proteins in the development of human cancers," *Journal of cellular physiology*, vol. 234, no. 6, pp. 8327–8341, 2019.

[12] X. Guo, Y. Shi, Y. Gou, J. Li, S. Han, Y. Zhang, J. Huo, X. Ning, L. Sun, Y. Chen *et al.*, "Human ribosomal protein s13 promotes gastric cancer growth through down-regulating p27kip1," *Journal of cellular and molecular medicine*, vol. 15, no. 2, pp. 296–306, 2011.

[13] K. Miyoshi, J. M. Shillingford, G. H. Smith, S. L. Grimm, K.-U. Wagner, T. Oka, J. M. Rosen, G. W. Robinson, and L. Hennighausen, "Signal transducer and activator of transcription (stat) 5 controls the proliferation and differentiation of mammary alveolar epithelium," *The Journal of cell biology*, vol. 155, no. 4, pp. 531–542, 2001.

[14] L. Migliore and F. Coppedè, "Genetic and environmental factors in cancer and neurodegenerative diseases," *Mutation Research/Reviews in Mutation Research*, vol. 512, no. 2-3, pp. 135–153, 2002.

[15] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

[16] R. Sharma, "Breast cancer incidence, mortality and mortality-to-incidence ratio (mir) are associated with human development, 1990–2016: evidence from global burden of disease study 2016," *Breast Cancer*, vol. 26, no. 4, pp. 428–445, 2019.

[17] C. J. Cabasag, P. J. Fagan, J. Ferlay, J. Vignat, M. Laversanne, L. Liu, M. A. van der Aa, F. Bray, and I. Soerjomataram, "Ovarian cancer today and tomorrow: A global assessment by world region and human development index using globocan 2020," *International Journal of Cancer*, 2022.

[18] P. Porter, ""westernizing" women's risks? breast cancer in lower-income countries," *New England Journal of Medicine*, vol. 358, no. 3, pp. 213–216, 2008.

[19] E. Hormones and B. C. C. Group, "Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies," *Journal of the National Cancer Institute*, vol. 94, no. 8, pp. 606–616, 2002.

[20] T. Walsh, S. Casadei, K. H. Coats, E. Swisher, S. M. Stray, J. Higgins, K. C. Roach, J. Mandell, M. K. Lee, S. Ciernikova *et al.*, "Spectrum of mutations in brca1, brca2, chek2, and tp53 in families at high risk of breast cancer," *Jama*, vol. 295, no. 12, pp. 1379–1388, 2006.

[21] C. C. Benz, "Impact of aging on the biology of breast cancer," *Critical reviews in oncology/hematology*, vol. 66, no. 1, pp. 65–74, 2008.

[22] S. Łukasiewicz, M. Czeczelewski, A. Forma, J. Baj, R. Sitarz, and A. Stanisławek, "Breast cancer—epidemiology, risk factors, classification, prognostic markers, and current treatment strategies—an updated review," *Cancers*, vol. 13, no. 17, p. 4287, 2021.

[23] C. G. on Hormonal Factors in Breast Cancer *et al.*, "Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease," *The Lancet*, vol. 358, no. 9291, pp. 1389–1399, 2001.

[24] M. L. Baglia, M.-T. C. Tang, K. E. Malone, P. Porter, and C. I. Li, "Family history and risk of second primary breast cancer after in situ breast carcinomafamily history and second primary breast cancer," *Cancer Epidemiology, Biomarkers & Prevention*, vol. 27, no. 3, pp. 315–320, 2018.

[25] H. R. Brewer, M. E. Jones, M. J. Schoemaker, A. Ashworth, and A. J. Swerdlow, "Family history and risk of breast cancer: an analysis accounting for family structure," *Breast cancer research and treatment*, vol. 165, no. 1, pp. 193–200, 2017.

[26] N. H. Khan, S.-F. Duan, D.-D. Wu, and X.-Y. Ji, "Better reporting and awareness campaigns needed for breast cancer in pakistani women," *Cancer Management and Research*, vol. 13, p. 2125, 2021.

[27] S. D. Copley, "An evolutionary perspective on protein moonlighting," *Biochemical Society Transactions*, vol. 42, no. 6, pp. 1684–1691, 2014.

[28] C. J. Jeffery, "Moonlighting proteins," *Trends in biochemical sciences*, vol. 24, no. 1, pp. 8–11, 1999.

[29] G. Wang, Y. Xia, J. Cui, Z. Gu, Y. Song, Y. Q. Chen, H. Chen, H. Zhang, and W. Chen, "The roles of moonlighting proteins in bacteria," *Current issues in molecular biology*, vol. 16, no. 1, pp. 15–22, 2014.

[30] C. Gancedo, C.-L. Flores, and J. M. Gancedo, "The expanding landscape of moonlighting proteins in yeasts," *Microbiology and Molecular Biology Reviews*, vol. 80, no. 3, pp. 765–777, 2016.

[31] D. F. Mark and C. C. Richardson, "Escherichia coli thioredoxin: a subunit of bacteriophage t7 dna polymerase." *Proceedings of the National Academy of Sciences*, vol. 73, no. 3, pp. 780–784, 1976.

[32] S. Tabor, H. Huber, and C. Richardson, "Escherichia coli thioredoxin confers processivity on the dna polymerase activity of the gene 5 protein of bacteriophage t7." *Journal of Biological Chemistry*, vol. 262, no. 33, pp. 16 212–16 223, 1987.

[33] R. Boteva, A. J. Visser, B. Filippi, G. Vriend, M. Veenhuis, and I. J. van der Klei, "Conformational transitions accompanying oligomerization of yeast alcohol oxidase, a peroxisomal flavoenzyme," *Biochemistry*, vol. 38, no. 16, pp. 5034–5044, 1999.

[34] M. C. Sugden and M. J. Holness, "The pyruvate carboxylase-pyruvate dehydrogenase axis in islet pyruvate metabolism: Going round in circles?" *Islets*, vol. 3, no. 6, pp. 302–319, 2011.

[35] C. Rodríguez-Saavedra, L. E. Morgado-Martínez, A. Burgos-Palacios, B. King-Díaz, M. López-Coria, and S. Sánchez-Nieto, "Moonlighting proteins: the case of the hexokinases," *Frontiers in Molecular Biosciences*, p. 531, 2021.

[36] S. Jitrapakdee, M. St Maurice, I. Rayment, W. W. Cleland, J. C. Wallace, and P. V. Attwood, "Structure, mechanism and regulation of pyruvate carboxylase," *Biochemical journal*, vol. 413, no. 3, pp. 369–387, 2008.

[37] G. D'Onofrio, F. Panza, V. Frisardi, V. Solfrizzi, B. P. Imbimbo, G. Paroni, L. Cascavilla, D. Seripa, and A. Pilotto, "Advances in the identification of γ-secretase inhibitors for the treatment of alzheimer's disease," *Expert opinion on drug discovery*, vol. 7, no. 1, pp. 19–37, 2012.

[38] R. M. Kluck, L. M. Ellerby, H. M. Ellerby, S. Naiem, M. P. Yaffe, E. Margoliash, D. Bredesen, A. G. Mauk, F. Sherman, and D. D. Newmeyer, "Determinants of cytochrome c pro-apoptotic activity: The role of lysine 72 trimethylation," *Journal of Biological Chemistry*, vol. 275, no. 21, pp. 16 127–16 133, 2000.

[39] C. Schindler, D. E. Levy, and T. Decker, "Jak-stat signaling: from interferons to cytokines," *Journal of Biological Chemistry*, vol. 282, no. 28, pp. 20 059–20 063, 2007.

[40] H. M. Asif, S. Sultana, N. Akhtar, J. U. Rehman, and R. U. Rehman, "Prevalence, risk factors and disease knowledge of breast cancer in pakistan," *Asian Pacific journal of cancer prevention*, vol. 15, no. 11, pp. 4411–4416, 2014.

[41] H. Yokoyama, S. Rybina, R. Santarella-Mellwig, I. W. Mattaj, and E. Karsenti, "Iswi is a rangtp-dependent map required for chromosome segregation," *Journal of Cell Biology*, vol. 187, no. 6, pp. 813–829, 2009.

[42] D. J. Forbes, A. Travesa, M. S. Nord, and C. Bernis, "Nuclear transport factors: global regulation of mitosis," *Current opinion in cell biology*, vol. 35, pp. 78–90, 2015.

[43] C. R. Clapier and B. R. Cairns, "The biology of chromatin remodeling complexes," *Annual review of biochemistry*, vol. 78, no. 1, pp. 273–304, 2009.

[44] S. H. Kim, D. P. Lin, S. Matsumoto, A. Kitazono, and T. Matsumoto, "Fission yeast slp1: an effector of the mad2-dependent spindle checkpoint," *Science*, vol. 279, no. 5353, pp. 1045–1047, 1998.

[45] A. DeAntoni, V. Sala, and A. Musacchio, "Explaining the oligomerization properties of the spindle assembly checkpoint protein mad2," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 360, no. 1455, pp. 637–648, 2005.

[46] L. Aravind and E. V. Koonin, "The horma domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and dna repair," *Trends in biochemical sciences*, vol. 23, no. 8, pp. 284–286, 1998.

[47] S. J. Royle, "Mitotic moonlighting functions for membrane trafficking proteins," *Traffic*, vol. 12, no. 7, pp. 791–798, 2011.

[48] J. W. Kim, S. J. Kim, S. M. Han, S. Y. Paik, S. Y. Hur, Y. W. Kim, J. M. Lee, and S. E. Namkoong, "Increased glyceraldehyde-3-phosphate dehydrogenase gene expression in human cervical cancers," *Gynecologic oncology*, vol. 71, no. 2, pp. 266–269, 1998.

[49] Y. Higashimura, Y. Nakajima, R. Yamaji, N. Harada, F. Shibasaki, Y. Nakano, and H. Inui, "Up-regulation of glyceraldehyde-3-phosphate dehydrogenase gene expression by hif-1 activity depending on sp1 in hypoxic breast cancer cells," *Archives of biochemistry and biophysics*, vol. 509, no. 1, pp. 1–8, 2011.

[50] Q. Huang, F. Lan, Z. Zheng, F. Xie, J. Han, L. Dong, Y. Xie, and F. Zheng, "Akt2 kinase suppresses glyceraldehyde-3-phosphate dehydrogenase (gapdh)-mediated apoptosis in ovarian cancer cells via phosphorylating gapdh at threonine 237 and decreasing its nuclear translocation," *Journal of Biological Chemistry*, vol. 286, no. 49, pp. 42 211–42 220, 2011.

[51] M. Jacquin, J. Chiche, B. Zunino, M. Beneteau, O. Meynet, L. Pradelli, S. Marchetti, A. Cornille, M. Carles, and J. Ricci, "Gapdh binds to active akt, leading to bcl-xl increase and escape from caspase-independent cell death," *Cell Death & Differentiation*, vol. 20, no. 8, pp. 1043–1054, 2013.

[52] R. T. Dorsam and J. S. Gutkind, "G-protein-coupled receptors and cancer," *Nature reviews cancer*, vol. 7, no. 2, pp. 79–94, 2007.

[53] L. Qi, Y. Zhang, F. Song, Y. Han, and Y. Ding, "A newly identified small molecular compound acts as a protein kinase inhibitor to suppress metastasis of colorectal cancer," *Bioorganic Chemistry*, vol. 107, p. 104625, 2021.

[54] Q. Fu and Z. Yu, "Phosphoglycerate kinase 1 (pgk1) in cancer: a promising target for diagnosis and therapy," *Life Sciences*, vol. 256, p. 117863, 2020.

[55] M. T. Johnson, H.-S. Yang, T. Magnuson, and M. S. Patel, "Targeted disruption of the murine dihydrolipoamide dehydrogenase gene (dld) results in perigastrulation lethality," *Proceedings of the National Academy of Sciences*, vol. 94, no. 26, pp. 14 512–14 517, 1997.

[56] N. E. Babady, Y.-P. Pang, O. Elpeleg, and G. Isaya, "Cryptic proteolytic activity of dihydrolipoamide dehydrogenase," *Proceedings of the National Academy of Sciences*, vol. 104, no. 15, pp. 6158–6163, 2007.

[57] R. A. Vaubel, P. Rustin, and G. Isaya, "Mutations in the dimer interface of dihydrolipoamide dehydrogenase promote site-specific oxidative damages in yeast and human cells," *Journal of Biological Chemistry*, vol. 286, no. 46, pp. 40 232–40 245, 2011.

[58] B. Henderson and A. Martin, "Bacterial virulence in the moonlight: multitasking bacterial moonlighting proteins are virulence determinants in infectious disease," *Infection and immunity*, vol. 79, no. 9, pp. 3476–3491, 2011.

[59] J. C. Ranford, A. R. Coates, and B. Henderson, "Chaperonins are cell-signalling proteins: the unfolding biology of molecular chaperones," *Expert reviews in molecular medicine*, vol. 2, no. 8, pp. 1–17, 2000.

[60] M. E. Marquart, "Pathogenicity and virulence of streptococcus pneumoniae: Cutting to the chase on proteases," *Virulence*, vol. 12, no. 1, pp. 766–787, 2021.

[61] N. Salazar, M. C. L. de Souza, A. G. Biasioli, L. B. da Silva, and A. S. Barbosa, "The multifaceted roles of leptospira enolase," *Research in microbiology*, vol. 168, no. 2, pp. 157–164, 2017.

[62] G. Qiao, A. Wu, X. Chen, Y. Tian, and X. Lin, "Enolase 1, a moonlighting protein, as a potential target for cancer treatment," *International journal of biological sciences*, vol. 17, no. 14, p. 3981, 2021.

[63] I. Smith, "Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence," *Clinical microbiology reviews*, vol. 16, no. 3, pp. 463–496, 2003.

[64] S. Sengupta, M. Shah, and V. Nagaraja, "Glutamate racemase from mycobacterium tuberculosis inhibits dna gyrase by affecting its dna-binding," *Nucleic acids research*, vol. 34, no. 19, pp. 5567–5576, 2006.

[65] C. Guo, S. Liu, and M.-Z. Sun, "Novel insight into the role of gapdh playing in tumor," *Clinical and Translational Oncology*, vol. 15, no. 3, pp. 167–172, 2013.

[66] B. Henderson, "An overview of protein moonlighting in bacterial infection," *Biochemical Society transactions*, vol. 42, no. 6, pp. 1720–1727, 2014.

[67] B. Henderson and A. Martin, "Bacterial moonlighting proteins and bacterial virulence," *Between pathogenicity and commensalism*, pp. 155–213, 2011.

[68] W. Wang and C. J. Jeffery, "An analysis of surface proteomics results reveals novel candidates for intracellular/surface moonlighting proteins in bacteria," *Molecular bioSystems*, vol. 12, no. 5, pp. 1420–1431, 2016.

[69] B. Henderson, M. A. Fares, and P. A. Lund, "Chaperonin 60: a paradoxical, evolutionarily conserved protein family with multiple moonlighting functions," *Biological Reviews*, vol. 88, no. 4, pp. 955–987, 2013.

[70] R. Wooster, S. L. Neuhausen, J. Mangion, Y. Quirk, D. Ford, N. Collins, K. Nguyen, S. Seal, T. Tran, D. Averill *et al.*, "Localization of a breast cancer susceptibility gene, brca2, to chromosome 13q12-13," *Science*, vol. 265, no. 5181, pp. 2088–2090, 1994.

[71] E. M Rosen and M. J Pishvaian, "Targeting the brca1/2 tumor suppressors," *Current drug targets*, vol. 15, no. 1, pp. 17–31, 2014.

[72] L. Pizzamiglio, E. Focchi, and F. Antonucci, "Atm protein kinase: old and new implications in neuronal pathways and brain circuitry," *Cells*, vol. 9, no. 9, p. 1969, 2020.

[73] D. Saslow, C. Boetes, W. Burke, S. Harms, M. O. Leach, C. D. Lehman, E. Morris, E. Pisano, M. Schnall, S. Sener *et al.*, "American cancer society guidelines for breast screening with mri as an adjunct to mammography," *CA: a cancer journal for clinicians*, vol. 57, no. 2, pp. 75–89, 2007.

[74] B. Friedenson, "Brca1 and brca2 pathways and the risk of cancers other than breast or ovarian," *Medscape General Medicine*, vol. 7, no. 2, p. 60, 2005.

[75] D. de Melo Gagliato and A. M. Gonzalez-Angulo, "Targeting multiple pathways in breast cancer," *Breast Cancer Management*, vol. 3, no. 1, pp. 87–101, 2014.

[76] B. Frank, K. Hemminki, A. Meindl, B. Wappenschmidt, C. Sutter, M. Kiechle, P. Bugert, R. K. Schmutzler, C. R. Bartram, and B. Burwinkel, "Brip1 (bach1) variants and familial breast cancer risk: a case-control study," *BMC cancer*, vol. 7, no. 1, pp. 1–4, 2007.

[77] B. R. B. Pires, Í. S. S. DE AMORIM, L. D. E. Souza, J. A. Rodrigues, and A. L. Mencalha, "Targeting cellular signaling pathways in breast cancer stem cells and its implication for cancer treatment," *Anticancer research*, vol. 36, no. 11, pp. 5681–5691, 2016.

[78] Y. Li, D. Yang, X. Yin, X. Zhang, J. Huang, Y. Wu, M. Wang, Z. Yi, H. Li, H. Li *et al.*, "Clinicopathological characteristics and breast cancer–specific survival of patients with single hormone receptor–positive breast cancer," *JAMA Network Open*, vol. 3, no. 1, pp. e1 918 160–e1 918 160, 2020.

[79] A. Nasrazadani, R. A. Thomas, S. Oesterreich, and A. V. Lee, "Precision medicine in hormone receptor-positive breast cancer," *Frontiers in oncology*, vol. 8, p. 144, 2018.

[80] H.-P. Konan, L. Kassem, S. Omarjee, A. Surmieliova-Garnès, J. Jacquemetton, E. Cascales, A. Rezza, O. Trédan, I. Treilleux, C. Poulard *et al.*, "Erα-36 regulates progesterone receptor activity in breast cancer," *Breast Cancer Research*, vol. 22, no. 1, pp. 1–16, 2020.

[81] A. E. Obr and D. P. Edwards, "The biology of progesterone receptor in the normal mammary gland and in breast cancer," *Molecular and cellular endocrinology*, vol. 357, no. 1-2, pp. 4–17, 2012.

[82] N. Patani, L.-A. Martin, and M. Dowsett, "Biomarkers for the clinical management of breast cancer: international perspective," *International journal of cancer*, vol. 133, no. 1, pp. 1–13, 2013.

[83] E. T. N. W. T. F. L. S. A.-M. Ann-Marie Billgren, Lars Erik Rutqvist, "Proliferating fraction during neoadjuvant chemotherapy of primary breast cancer in relation to objective local response and relapse-free survival," *Acta Oncologica*, vol. 38, no. 5, pp. 597–601, 1999.

[84] B. A. Kohler, R. L. Sherman, N. Howlader, A. Jemal, A. B. Ryerson, K. A. Henry, F. P. Boscoe, K. A. Cronin, A. Lake, A.-M. Noone *et al.*, "Annual report to the nation on the status of cancer, 1975-2011, featuring incidence of breast cancer subtypes by race/ethnicity, poverty, and state," *Journal of the National Cancer Institute*, vol. 107, no. 6, p. djv048, 2015.

[85] K. Kontani, N. Kuroda, S.-i. Hashimoto, C. Murazawa, S. Norimura, H. Tanaka, M. Ohtani, N. Fujiwara-Honjo, Y. Kushida, M. Date *et al.*, "Clinical usefulness of human epidermal growth factor receptor-2 extracellular domain as a biomarker for monitoring cancer status and predicting the therapeutic efficacy in breast cancer," *Cancer Biology & Therapy*, vol. 14, no. 1, pp. 20–28, 2013.

[86] D. Furrer, C. Paquet, S. Jacob, and C. Diorio, "The human epidermal growth factor receptor 2 (her2) as a prognostic and predictive biomarker: Molecular insights into her2 activation and diagnostic implications," *Cancer Progn*, 2018.

[87] H. S. Qureshi, M. D. Linden, G. Divine, and U. B. Raju, "E-cadherin status in breast cancer correlates with histologic type but does not correlate with established prognostic parameters," *American journal of clinical pathology*, vol. 125, no. 3, pp. 377–385, 2006.

[88] R. Nishimura, T. Osako, Y. Okumura, M. Hayashi, Y. Toyozumi, and N. Arima, "Ki-67 as a prognostic marker according to breast cancer subtype and a predictor of recurrence time in primary breast cancer," *Experimental and therapeutic medicine*, vol. 1, no. 5, pp. 747–754, 2010.

[89] N. Pathmanathan, R. L. Balleine, U. W. Jayasinghe, K. L. Bilinski, P. J. Provan, K. Byth, A. M. Bilous, E. L. Salisbury, and J. Boyages, "The prognostic value of ki67 in systemically untreated patients with node-negative breast cancer," *Journal of clinical pathology*, vol. 67, no. 3, pp. 222–228, 2014.

[90] H. N. Horne, H. Oh, M. E. Sherman, M. Palakal, S. M. Hewitt, M. K. Schmidt, R. L. Milne, D. Hardisson, J. Benitez, C. Blomqvist *et al.*, "E-cadherin breast tumor expression, risk factors and survival: Pooled analysis of 5,933 cases from 12 studies in the breast cancer association consortium," *Scientific reports*, vol. 8, no. 1, pp. 1–11, 2018.

[91] H. S. Qureshi, M. D. Linden, G. Divine, and U. B. Raju, "E-cadherin status in breast cancer correlates with histologic type but does not correlate with

established prognostic parameters," *American journal of clinical pathology*, vol. 125, no. 3, pp. 377–385, 2006.

[92] W.-B. Yin, M.-G. Yan, X. Fang, J.-J. Guo, W. Xiong, and R.-P. Zhang, "Circulating circular rna hsa_circ_0001785 acts as a diagnostic biomarker for breast cancer detection," *Clinica chimica acta*, vol. 487, pp. 363–368, 2018.

[93] M. V. Iorio and C. M. Croce, "Microrna dysregulation in cancer: diagnostics, monitoring and therapeutics. a comprehensive review," *EMBO molecular medicine*, vol. 4, no. 3, pp. 143–159, 2012.

[94] A. Dumay, J.-P. Feugeas, E. Wittmer, J. Lehmann-Che, P. Bertheau, M. Espié, L.-F. Plassa, P. Cottu, M. Marty, F. André *et al.*, "Distinct tumor protein p53 mutants in breast cancer subgroups," *International journal of cancer*, vol. 132, no. 5, pp. 1227–1231, 2013.

[95] S. I. Grivennikov, F. R. Greten, and M. Karin, "Immunity, inflammation, and cancer," *Cell*, vol. 140, no. 6, pp. 883–899, 2010.

[96] E. A. Wiemer, "The role of micrornas in cancer: no small matter," *European journal of cancer*, vol. 43, no. 10, pp. 1529–1544, 2007.

[97] M. Adhami, A. A. Haghdoost, B. Sadeghi, and R. Malekpour Afshar, "Candidate mirnas in human breast cancer biomarkers: a systematic review," *Breast Cancer*, vol. 25, no. 2, pp. 198–205, 2018.

[98] G. J. Guthrie, K. A. Charles, C. S. Roxburgh, P. G. Horgan, D. C. McMillan, and S. J. Clarke, "The systemic inflammation-based neutrophil–lymphocyte ratio: experience in patients with cancer," *Critical reviews in oncology/hematology*, vol. 88, no. 1, pp. 218–230, 2013.

[99] A. Mantovani, S. Sozzani, M. Locati, P. Allavena, and A. Sica, "Macrophage polarization: tumor-associated macrophages as a paradigm for polarized m2 mononuclear phagocytes," *Trends in immunology*, vol. 23, no. 11, pp. 549–555, 2002.

[100] C. B. Williams, E. S. Yeh, and A. C. Soloff, "Tumor-associated macrophages: unwitting accomplices in breast cancer malignancy," *NPJ breast cancer*, vol. 2, no. 1, pp. 1–12, 2016.

[101] Z.-Y. Yuan, R.-Z. Luo, R.-J. Peng, S.-S. Wang, and C. Xue, "High infiltration of tumor-associated macrophages in triple-negative breast cancer is associated with a higher risk of distant metastasis," *OncoTargets and therapy*, vol. 7, p. 1475, 2014.

[102] H. Wang, Y. Ding, N. Li, L. Wu, Y. Gao, C. Xiao, H. Jiang, Y. Zheng, C. Mao, J. Deng *et al.*, "Prognostic value of neutrophil–lymphocyte ratio, platelet–lymphocyte ratio, and combined neutrophil–lymphocyte ratio and platelet–lymphocyte ratio in stage iv advanced gastric cancer," *Frontiers in oncology*, vol. 10, p. 841, 2020.

[103] A. Kinoshita, H. Onoda, N. Imai, A. Iwaku, M. Oishi, N. Fushiya, K. Koike, H. Nishino, and H. Tajiri, "Comparison of the prognostic value of inflammation-based prognostic scores in patients with hepatocellular carcinoma," *British journal of cancer*, vol. 107, no. 6, pp. 988–993, 2012.

[104] K. A. Mouchemore, R. L. Anderson, and J. A. Hamilton, "Neutrophils, g-csf and their contribution to breast cancer metastasis," *The FEBS journal*, vol. 285, no. 4, pp. 665–679, 2018.

[105] D. Tan, Y. Fu, W. Tong, and F. Li, "Prognostic significance of lymphocyte to monocyte ratio in colorectal cancer: a meta-analysis," *International journal of surgery*, vol. 55, pp. 128–138, 2018.

[106] C. E. Olingy, H. Q. Dinh, and C. C. Hedrick, "Monocyte heterogeneity and functions in cancer," *Journal of leukocyte biology*, vol. 106, no. 2, pp. 309–322, 2019.

[107] B. Li, P. Zhou, Y. Liu, H. Wei, X. Yang, T. Chen, and J. Xiao, "Platelet-to-lymphocyte ratio in advanced cancer: review and meta-analysis," *Clinica Chimica Acta*, vol. 483, pp. 48–56, 2018.

[108] S. I. Kubota, K. Takahashi, T. Mano, K. Matsumoto, T. Katsumata, S. Shi, K. Tainaka, H. R. Ueda, S. Ehata, and K. Miyazono, "Whole-organ analysis of tgf-β-mediated remodelling of the tumour microenvironment by tissue clearing," *Communications biology*, vol. 4, no. 1, pp. 1–15, 2021.

[109] R. Dominguez and K. C. Holmes, "Actin structure and function," *Annual review of biophysics*, vol. 40, p. 169, 2011.

[110] ——, "Actin structure and function," *Annual review of biophysics*, vol. 40, p. 169, 2011.

[111] K. Crawford, R. Flick, L. Close, D. Shelly, R. Paul, K. Bove, A. Kumar, and J. Lessard, "Mice lacking skeletal muscle actin show reduced muscle strength and growth deficits and die during the neonatal period," *Molecular and cellular biology*, vol. 22, no. 16, pp. 5887–5896, 2002.

[112] W.-D. Ni, Z.-T. Yang, C.-A. Cui, Y. Cui, L.-Y. Fang, and Y.-H. Xuan, "Tenascin-c is a potential cancer-associated fibroblasts marker and predicts poor prognosis in prostate cancer," *Biochemical and biophysical research communications*, vol. 486, no. 3, pp. 607–612, 2017.

[113] T. Wu, X. Wang, J. Li, X. Song, Y. Wang, Y. Wang, L. Zhang, Z. Li, and J. Tian, "Identification of personalized chemoresistance genes in subtypes of basal-like breast cancer based on functional differences using pathway analysis," *PLoS One*, vol. 10, no. 6, p. e0131183, 2015.

[114] I. Evsyukova, C. Plestant, and E. Anton, "Integrative mechanisms of oriented neuronal migration in the developing brain," *Annual review of cell and developmental biology*, vol. 29, p. 299, 2013.

[115] J. A. Carson and L. Wei, "Integrin signaling's potential for mediating gene expression in hypertrophying skeletal muscle," *Journal of applied physiology*, vol. 88, no. 1, pp. 337–343, 2000.

[116] K. Yang, S. Zhang, D. Zhang, Q. Tao, T. Zhang, G. Liu, X. Liu, and T. Zhao, "Identification of serpine1, plau and acta1 as biomarkers of head and neck

squamous cell carcinoma based on integrated bioinformatics analysis," *International journal of clinical oncology*, vol. 24, no. 9, pp. 1030–1041, 2019.

[117] N. M. White-Al Habeeb, L. T. Ho, E. Olkhov-Mitsel, K. Kron, V. Pethe, M. Lehman, L. Jovanovic, N. Fleshner, T. van der Kwast, C. C. Nelson *et al.*, "Integrated analysis of epigenomic and genomic changes by dna methylation dependent mechanisms provides potential novel biomarkers for prostate cancer," *Oncotarget*, vol. 5, no. 17, p. 7858, 2014.

[118] G. M. Son, M.-S. Kwon, D.-H. Shin, N. Shin, D. Ryu, and C.-D. Kang, "Comparisons of cancer-associated fibroblasts in the intratumoral stroma and invasive front in colorectal cancer," *Medicine*, vol. 98, no. 18, p. e15164, 2019.

[119] C. J. Hanley, M. Mellone, K. Ford, S. M. Thirdborough, T. Mellows, S. J. Frampton, D. M. Smith, E. Harden, C. Szyndralewiez, M. Bullock *et al.*, "Targeting the myofibroblastic cancer-associated fibroblast phenotype through inhibition of nox4," *JNCI: Journal of the National Cancer Institute*, vol. 110, no. 1, pp. 109–120, 2018.

[120] D. Nowak, A. Skwarek-Maruszewska, M. Zemanek-Zboch, and M. Malicka-Błaszkiewicz, "Beta-actin in human colon adenocarcinoma cell lines with different metastatic potential." *Acta Biochimica Polonica*, vol. 52, no. 2, pp. 461–468, 2005.

[121] C. Blanquicett, M. R. Johnson, M. Heslin, and R. B. Diasio, "Housekeeping gene variability in normal and carcinomatous colorectal and liver tissues: applications in pharmacogenomic gene expression studies," *Analytical biochemistry*, vol. 303, no. 2, pp. 209–214, 2002.

[122] S. Kwiatkowski, A. K. Seliga, D. Vertommen, M. Terreri, T. Ishikawa, I. Grabowska, M. Tiebe, A. A. Teleman, A. K. Jagielski, M. Veiga-da Cunha *et al.*, "Setd3 protein is the actin-specific histidine n-methyltransferase," *Elife*, vol. 7, p. e37921, 2018.

[123] L. Yu, X. Chen, X. Sun, L. Wang, and S. Chen, "The glycolytic switch in tumors: how many players are involved?" *Journal of Cancer*, vol. 8, no. 17, p. 3430, 2017.

[124] L. Pourcel, F. Buron, G. Arib, V. Le Fourn, A. Regamey, I. Bodenmann, P.-A. Girod, and N. Mermod, "Influence of cytoskeleton organization on recombinant protein expression by cho cells," *Biotechnology and bioengineering*, vol. 117, no. 4, pp. 1117–1126, 2020.

[125] M. S. Shum, E. Pasquier, S. T. Po'uha, G. M. O'Neill, C. Chaponnier, P. W. Gunning, and M. Kavallaris, "γ-actin regulates cell migration and modulates the rock signaling pathway," *The FASEB journal*, vol. 25, no. 12, pp. 4423–4433, 2011.

[126] X. Dong, Y. Han, Z. Sun, and J. Xu, "Actin gamma 1, a new skin cancer pathogenic gene, identified by the biological feature-based classification," *Journal of cellular biochemistry*, vol. 119, no. 2, pp. 1406–1419, 2018.

[127] S. Durinck, E. W. Stawiski, A. Pavía-Jiménez, Z. Modrusan, P. Kapur, B. S. Jaiswal, N. Zhang, V. Toffessi-Tcheuyap, T. T. Nguyen, K. B. Pahuja *et al.*, "Spectrum of diverse genomic alterations define non–clear cell renal carcinoma subtypes," *Nature genetics*, vol. 47, no. 1, pp. 13–21, 2015.

[128] A. N. Sasikumar, W. B. Perez, and T. G. Kinzy, "The many roles of the eukaryotic elongation factor 1 complex," *Wiley Interdisciplinary Reviews: RNA*, vol. 3, no. 4, pp. 543–555, 2012.

[129] L. Pecorari, O. Marin, C. Silvestri, O. Candini, E. Rossi, C. Guerzoni, S. Cattelani, S. A. Mariani, F. Corradini, G. Ferrari-Amorotti *et al.*, "Elongation factor 1 alpha interacts with phospho-akt in breast cancer cells and regulates their proliferation, survival and motility," *Molecular cancer*, vol. 8, no. 1, pp. 1–11, 2009.

[130] D. A. Fruman, R. E. Meyers, and L. C. Cantley, "Phosphoinositide kinases," *Annual review of biochemistry*, vol. 67, p. 481, 1998.

[131] T. D. Bunney and M. Katan, "Phosphoinositide signalling in cancer: beyond pi3k and pten," *Nature Reviews Cancer*, vol. 10, no. 5, pp. 342–352, 2010.

[132] S. Jeganathan and J. M. Lee, "Binding of elongation factor eef1a2 to phosphatidylinositol 4-kinase β stimulates lipid kinase activity and phosphatidylinositol 4-phosphate generation," *Journal of Biological Chemistry*, vol. 282, no. 1, pp. 372–380, 2007.

[133] G. Kulkarni, D. A. Turbin, A. Amiri, S. Jeganathan, M. A. Andrade-Navarro, T. D. Wu, D. G. Huntsman, and J. M. Lee, "Expression of protein elongation factor eef1a2 predicts favorable outcome in breast cancer," *Breast cancer research and treatment*, vol. 102, no. 1, pp. 31–41, 2007.

[134] A. P. Ayyappan and K. S. Jhaveri, "Ct and mri of hepatocellular carcinoma: an update," *Expert review of anticancer therapy*, vol. 10, no. 4, pp. 507–519, 2010.

Appendix A

An Appendix

Table A.1: Predicted Moonlight Protein list

MoonDB ID	UniprotKB AC	Protein Full Name (MDB)	Moonlight Protein
3	Q6UWE0	IE3 ubiquitin-protein ligase LR-SAM1	Eukaryotic translation initiation factor 2-
4	O00499	IMyc box-dependent-interacting protein 1	Aromatase, EC (Estrogen synthase)
7	P62256i	IUbiquitin-conjugating enzyme E2 Hi	Homeobox protein Hox-A1 (Homeobox protein Hox-1F)
9	P04792i	IHeat shock protein beta-1i	Zinc finger homeobox protein i
13	P02511	IAlpha-crystallin B chain	DAZ-associated protein 1
16	P60520	IGamma-aminobutyric acid receptor-associated protein-like 2	Actin-like protein 9
18	Q9UMS4i	IPre-mRNA-processing factor 19i	Desmocollin 3
20	O15287	I Fanconi anemia group G protein	Alpha-tectorin
22	Q00613	Heat shock factor protein 1	Amine oxidase

MoonDB ID	**UniprotKB AC**	**Protein Full Name (MDB)**	**Moonlight Protein**
25	O75674	ITOM1-like protein 1	60 kDa heat shock protein, ((Heat shock protein 60, HSP-60, Hsp60) (Mitochondrial matrix protein P1) (P60 lymphocyte protein)
26	Q13492	IPhosphatidylinositol-binding clathrin assembly protein I	myoblast determination protein 1 (bHLHc1) (Myogenic factor 3, Myf-3)
35	Q15038	IDAZ-associated protein 2 I	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial, MCAD,
41	P27361	IMitogen-activated I protein kinase 3	Dipeptidase 1, EC 3.4.13.19 ((Microsomal dipeptidase) (Renal dipeptidase,I hRDP)
44	P0DP23	Calmodulin-1 I	(Biotin apo-protein ligase)
50	P27540	Aryl hydrocarbon receptor nuclear translocator I	Ornithine decarboxylase, ODC,
51	Q16659	Mitogen-activated protein kinase 6 I	Amphiregulin, AR (Colorectum cell-derived growth factor, CRDGF)
55	Q13064	Probable E3 ubiquitin-protein ligase makorin-3 I	Actin-related protein 3 (Actin-like protein 3)i
59	O00204	Sulfotransferase family cytosolic 2B member 1 I	Interferon alpha-2, IFN-alpha-2 (Interferon alpha-A, LeIF A)
62	P28702	Retinoic acid receptor RXR-beta I	Alpha-crystallin A chain (Heat shock protein beta-4, HspB4)
64	P68036	Ubiquitin-conjugating enzyme E2 L3 I	Collagenase 3, EC 3.4.24.- (Matrix metalloproteinase-13, MMP-13)
71	Q92997	Segment polarity protein dishevelled homolog DVL-3 I	Interferon beta, IFN-beta (Fibroblast interferon)
72	P00734	Prothrombin I	Tensin-4 (C-terminal tensin-like protein)

MoonDB ID	**UniprotKB AC**	**Protein Full Name (MDB)**	**Moonlight Protein**
74	Q05086	Ubiquitin-protein ligase E3A I	cAMP-dependent protein kinase .
75	Q9BWF3	RNA-bindinG protein 4 I	Ubiquitin carboxyl-terminal hydrolase 17 (Deubiquitinating enzyme 17-like protein 2) (DUB-3, deubiquitinating protein)
76	P19971	Thymidine phosphorylase I	Uncharacterized protein C5orf34
80	O00308	NEDD4-like E3 ubiquitin-protein ligase WWP2 I	DNA polymerase theta, EC 2.7.7.7 (DNA polymerase eta)
82	Q9Y6X0	SET-binding protein I	Prolactin-inducible protein (Gross cystic disease fluid protein 15, GCDFP-15)