Proteomic Analysis for Identification of Genetic Markers for Brain Cancer

Done by:

Ayesha Ejaz (00000279485)

Iman Asad (00000280685)

Arwa Aatique (00000280670)

Saad Ahmed (00000279085)

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

National University of Sciences and Technology (NUST)

2022

Proteomic Analysis for Identification of Genetic Markers for Brain Cancer

Done by:

Ayesha Ejaz (00000279485)

Iman Asad (00000280685)

Arwa Aatique (00000280670)

Saad Ahmed (00000279085)

Supervised by: Dr. Maria Shabbir

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

National University of Sciences and Technology (NUST)

2022

Proteomic Analysis for Identification of Genetic Markers for Brain Cancer

Done by:

Ayesha Ejaz (00000279485)

Iman Asad (00000280685)

Arwa Aatique (00000280670)

Saad Ahmed (00000279085)

A thesis submitted to the National University of Science and Technology, Islamabad, in partial fulfillment of the requirements of the degree of

Bachelor of Science in

Applied Biosciences

Thesis Supervisor: Dr. Maria Shabbir

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

National University of Sciences and Technology (NUST)

2022

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of BS FYP Thesis written by Mr/Ms Ayesha Ejaz, Iman Asad, Arwa Aatique and Saad Ahmed (Registration No. 00000279485, 00000280685, 00000280670, 00000279085), of (School/College/Institute) "Atta-ur-rahman School of Applied Biosciences" has been vetted by undersigned, found complete in all respects as per NUST Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of BS degree. It is further certified that necessary amendments as pointed out, during final presentation of the scholar, have also been incorporated in the said thesis.

Signature: _____

Name of Supervisor _____

Date:_____

Signature (HOD): _____

Date: _____

Signature (Dean/Principal) _____

Date: _____

AUTHOR'S DECLARATION

I, Ayesha Ejaz, Iman Asad, Arwa Aatique and Saad Ahmed, hereby state that my BS FYP thesis titled "Proteomic Analysis for identification of Genetic markers for Brain cancer" is my own work and has not been submitted previously by me for taking any degree from this university National University of Science and Technology or anywhere else in the country/world.

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

Name of student:
Ayesha Ejaz: _____
Arwa Aatique: _____
Iman Asad: _____
Saad Ahmed: _____

Date:

Certificate for Plagiarism

It is certified that BS FYP Thesis Titled "Proteomic Analysis for identification of Genetic Markers for Barain cancer" by Ayesha Ejaz, Iman Asad, Arwa Aatique and Saad Ahmed has been examined by me. I undertake the following:-

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

Date:_____ **Student Signature:**

Ayesha Ejaz: _____

Iman Asad: _____

Arwa Aatique: _____

Saad Ahmed: _____

Date:_____ **Signature & Stamp of Supervisor**

This thesis is dedicated to our beloved parents, friends, and family whose encouragement and guidance has been unmatched.

Acknowledgments

Above all, we would like to first of all thank our research supervisors, Dr Maria Shabbir and Dr Yasmeen Badshah whose guidance and dedication helped this project come to its fruition. Their constant support has been much appreciated throughout the duration of our degree.
Secondly, with utmost gratitude, we thank Dr Hussnain A. Janjua and the entirety of the ASAB department for providing us with all of the required facilities that aided in the progress and eventual completion of our project.

We would also like to show particular gratitude to Ms Khushbakhat Khan who has aided us tremendously in the Insilico work as well as our thesis write up and its compilation for which we would like to thank Ms Hania Shah as well. Special thanks to Ms Amna Hafeez Khan, Ms Sameen Fatima and Ms Kunza who also aided us in the lab. Their guidance is sincerely valued.

Finally, and most importantly, the kindness and support of our parents and family, friends and fellow peers has been immensely appreciated. Without their encouragement, we would not be where we are today and we thank them from the bottom of our hearts.

Contents

Table of Figures

List of Tables

List of Abbreviations

GC-MS: Gas chromatography Mass spectroscopy

CNS: Central Nervous System

PNS: Peripheral Nervous System

GBM: Glioblastoma

MSTFA: N-Methyl-N-(trimethylsilyl)trifluoroacetamide

TMCS: Chlorotrimethylsilane

SNP: Single Nucleotide Polymorphism

VDR: Vitamin D Receptor

EGFR: Epidermal Growth Factor Receptor

TP53/P53: Tumor protein 53

IDH 1 or 2: Isocitrate dehydrogenase 1 or 2

Rb: Retinoblastoma protein

RTK: Receptor Tyrosine Kinase

MGMT: O^6 -alkylguanine DNA alkyltransferase

ERBB2: Erb B-2 Receptor Tyrosine kinase 2

MET: Mesenchymal Epithelial transition Receptor Tyrosine Kinase

PDGFRA: Platelet derived growth factor receptor

SIRT2: Sirtuin 2 gene

VGEF: Vascular endothelial growth factor

PKCβ: Protein kinase Cβ

Ras/MAPK: Ras-mitogen activated protein kinase

PARP-1: Poly (ADP-ribose) polymerase 1

SSRP1: Human FACT Complex SSRP1

RL7: 60S ribosomal protein L7

RS13: 40S ribosomal protein S13

RL4: 60S ribosomal protein L4

RAB5C: Ras related protein Rab-5C

RL8: 60S ribosomal protein L8

RS8: 40S ribosomal protein S8

RL13: 60S ribosomal protein L13

RAB5B: Ras-related protein Rab-5B

RL18: 60S ribosomal protein L18

SF3B3: Splicing factor 3B subunit 3

MET7A: Methyltransferase like 7A, isoform CRA_a

ACADV: Very long-chain specific Acyl-CoA dehydrogenase

DHE3: Glutamate dehydrogenase 1

RL18A: 60S ribosomal protein L18a

PSA7: Proteasome subunit alpha type-7

SEPT2: Septin-2

JAK1: Janus Kinase 1

LAMC1: Laminin subunit gamma-1

LAMB1: Laminin Subunit Beta 1

LAMA5: Laminin subunit alpha 5

CFB: Complement factor B

SMD3: Small nuclear ribonucleoprotein Sm D3

SEPT7: Septin-7

EIF3L: Eukaryotic translation initiation factor 3 subunit L

AT2A2: Sarcoplasmic/endoplasmic reticulum calcium ATPase 2

AT2A1: Sarcoplasmic/endoplasmic reticulum calcium ATPase 1

RL3: 60S ribosomal protein L3

ARPC4: Actin-related protein 2/3 complex subunit 4

hsa-mir: Human microRNA

NFkb: Nuclear factor kappa B

MTDH: Metadherin

Abstract

Brain tumors are amongst the leading source of cancer-related morbidity and mortality worldwide. However, the diagnostic techniques used for this are time consuming and costly. The aim of this study was to identify the potential genetic markers for brain cancer in the Pakistani population for diagnostic and prognostic purposes through the evaluation of serum samples using GC-MS.

A total of six samples (three patients and three controls) were collected for GC-MS analysis. The compounds that were identified were analyzed using scaffold software and based on t-test and foldchange calculation, 25 proteins were sorted. Afterwards, their location was determined and their metabolic pathways were mapped. PPP1R163 gene was selected for further processing based on its various roles in cancer processes and SNP data was retrieved and filtered that was then subjected to the processing that followed. Evolutionary conservation, structure analysis, stability analysis and investigation into its link with cancer was carried out.

Most proteins were localized in the cytoplasm while some were in the nucleus, mitochondria and endoplasmic reticulum. RAB5C and ACADV were anti-apoptotic leading to enhanced proliferation while RPL7 lead to cancer angiogenesis and metastasis. Additionally, SSRP1 was both pro-proliferative and tumor suppressive that indirectly activated PPP1R163 gene. Among the 7 deleterious variants of this gene, 4 of them; L387P, R274C, K246M and E92A had a significant impact on protein function while destabilization was caused by most of the SNPs. S405F was the most associated with cancer.

Due to limitations and lack of existing data on cancer in Pakistan, it is difficult to study cancer and its effects on the population. This study identifies the different prognostic genetic markers for brain cancer and brings focus on PPP1R163 as an important gene with its link to brain cancer and as a potential future therapeutic target.

Chapter 1

Introduction

1.1. Cancer

Cancer in a broader term that refers to a collection of more than 277 different types of cancer disease that are divided into different categories based on their origins such as brain cancer, breast cancer, liver cancer, skin cancer, prostate cancer and many others. Cancer arises when a cell breaks free from the constraints of normal cell division and proliferation that can lead to uncontrolled division, multiplication and spread of affected cells at a very fast pace, this can lead to the development of a lump of cells referred to as a tumor. Therefore, it can be defined as uncontrolled and unchecked division and multiplication of body cells (Weinberg, How Cancer Arises, 1996). Tumors can be malignant or benign. Cancerous or malignant tumors spread and invade nearby tissues and can travel to different organs of our bodies to form new tumors, a process referred to as metastasis. These tumors are referred to as malignant tumors. Benign tumors, however, do not spread into or invade nearby tissues. When removed, they usually don't grow back, unlike malignant tumors. Benign tumors are usually not as threatening as malignant tumors but sometimes can cause serious symptoms and be life threatening depending upon their position in the body such as a benign tumor located in the brain. Tumors composed of malignant cells can become more aggressive over time and can be lethal and life threatening when they disrupt the tissues and organs needed for the survival of the organism as a whole. This process of cancer and tumor development is not new but over the past few decades, scientists have uncovered a set of basic principles that govern the development of cancer. The malignant transformation of a normal cell into a cancer cell usually comes through the accumulation of mutations in specific classes of genes within that cell. These classes of genes can provide the key of understanding the processes at the root that can lead to development of a cancer (Chaurasia, Saurabh, & Tiwari, 2018).

1.2. Brain Cancer

The human brain is one the most complex organ structures with approximately 100 billion neurons and even more glial cells. It makes up 2% of the total body weight where 20% of the body's energy and oxygen is supplied (Maldonado & Alsayouri.,

2021). Unfortunately, some of the highly alarming tumors to exist in the world today are brain tumors (Aldape, 2019). Brain tumors can be benign or malignant but can become metastatic tumors, which are frequent in adults (Milica M. Badža, 2021). It is predicted that neuronal stem cells transform and create a transformed and malignant brain tumor hierarchy that leads to tumor formation. The perivascular niches, which are the microenvironment around the vessels, are said to have the nonmalignant and malignant stem cells. Glioblastomas, from glial cells, are the most frequent types of malignant brain tumors (Jigisha P Thakkar). Glioblastoma cells trigger tumor-associated macrophages for creation of immunosuppressive, tumor-promoting cytokines and amplified apoptosis of T cells (Wojciech Jelski, 2021). The leading causes of the disease are unknown but genetic susceptibility is seen to be significantly involved (Fabelo H, 2018). Some tumors are difficult to recognize due to the location in which they are present such as those located behind the blood–brain barrier (Aldape, 2019). This makes them more threatening and difficult to treat early.

They are recognized after the display of symptoms like headache, dizziness, nausea, mood swings, or loss of communication (Mroczko, 2021). In spite of this, most of the tumors are identified at a later stage leading to limited treatment options. This is because identification of tumor is done through MRI most of the time and the detection is usually made when it is the final stage of tumor (Fabelo H, 2018). With the progress of a tumor, there are a variety of biochemical and molecular changes that occur. There can be early identification through tests with the help of molecular biomarkers. These markers can be classified into distinct groups: molecular biomarkers, circulating free DNA, circulating tumor cells, circulating extracellular vesicles, and circulating cell-free microRNAs. With glioblastoma, EGFR could be a possible biomarker. Normally it is a part of growth factor signaling, but when there are cancer-related changes it gains ligand-independent oncogenic activity. There is more EGFR present in glioblastoma patients, which is approximately 40%, and can be related to high grade cancers. Furthermore, cells of the central nervous system like astrocytes produce glial fibrillary acidic protein, an intermediate fiber protein. It is seen to be present significantly more in tumor tissue compared to normal tissue. There are a variety of other biomarkers that could indicate the presence of brain tumor progression (Wojciech Jelski, 2021).

1.3. Types of brain cancer

Overall, there are two types of brain cancer;

1. Primary brain tumors: These originate in the brain and remain localized there.
2. Secondary brain tumors: These originate elsewhere and reach the brain. The most common type of cancers that metastasize to the brain are; lung, breast cancer and melanoma (Nayak L, 2012).

Brain tumors can either be benign, which are non-cancerous, or they can be malignant, which are cancerous. In order to further analyze the different brain tumors on the basis of how they look or how fast they spread, World Health Organization (WHO) has further graded them from 1-4; as the grade increases, so does the aggressiveness and abnormality of the cancer.

In a recent epidemiological study in Pakistan done in clinics, it was determined what type of cancer was the most common amongst children and adults and whether it was more prevalent in one gender compared to the other. It was found that it was more prevalent in males with a male to female ratio of 13.63:1 (Nauman Idris Butt, 2020). It was also found to be more common in children aged 6-8 years with the mean age of occurrence found to be 6.73 (Nauman Idris Butt, 2020). The incidence of the different types has been detailed in Table 1.

Table 1: Incidence of different types of brain tumors in both children and adults in Pakistan **(Nauman Idris Butt, 2020)**

Diagnosis (Children)	**Percentage %**
Benign	30.76%
Small Round Blue Cell Tumor	17.30%
Medulloblastoma	21.15%
Diagnosis (Adults)	**Percentage %**
Glioblastoma	21.95%
Meningioma	20.32%
Benign	20.32%
Low grade Glioma	18.69%

Herein would be discussed the two most commonly occurring tumors worldwide and in Pakistan; Meningioma and Gliomas.

Meningioma

These types of tumors originate in the outer layers of tissue that coat the brain just beneath the skull, the meninges. In the meninges, they grow out of the middle layer called arachnoid and do so rather slowly, so much so that it takes years to be detectable.

Glioma

These types of tumors originate in glial cells that surround the brain and provide support to its neurons. These cells may include astrocytes, ependymal cells or oligodendrocytes. The different types of brain tumors under this category thus include astrocytomas, brain stem gliomas, ependymomas, mixed gliomas, oligodendrogliomas and optic pathway gliomas.

1.4. Genetic pathways involved in brain cancer

It has been proven that there are certain types of genes that are responsible for the onset of primary cancers. These include:

1. Tumor suppressors
2. Genes necessary for DNA repair
3. Proto-oncogenes

Mutations in these genes often determine the type of brain cancer an individual is afflicted with. The degree of mutation results in different kinds of tumors (Gopal S Tandel, 2019).

These genes may include tumor protein 53 (TP53) which has a role in initiating apoptosis and DNA repair. It is also known to be involved in high-grade gliomas where its levels are elevated. Mutations in this gene are indicated in 80% of the tumors or ATRX (A-Thalassemia-mental retardation syndrome X-linked) which is involved in mutations of TP53 and IDH1. It has shown valuable use as a prognostic indicator during cases of tumors with IDH1 mutations. These along with IDH2 mutations, have also been seen in high-grade glioblastomas and oligodendroglial tumors. Abnormalities in repair genes such as MGMT are also seen as a cause of glioblastomas. Another gene involved is the RB1 gene or retinoblastoma which is a

tumor suppression gene whose mutation has been implicated in 75% of brain tumors especially in glioblastomas (Gopal S Tandel, 2019).

P53 protein is a transcriptional regulator that has an integral role in maintaining cell proliferation, genomic integrity and cell survival, promoting cell cycle arrest and apoptosis as well as regulating cellular homeostasis. It is also activated in instances of DNA damage. The p53 pathway is seen to deregulate in patients with glioblastomas (Ying Zhang, 2018).

Many signaling pathways have also been implicated in glioblastomas that have contributed significantly to its aggressiveness in patients. These have been reported to be the RTK (receptor tyrosine kinase) pathway in which amplification and mutation of EGFR, ERBB2, PDGFRA, and MET has been studied as well as phosphatidylinositol-3 kinase (PI3K)/Akt/mammalian target of rapamycin (mTOR), RB (retinoblastoma)/p53, signal transducer and activator of transcription 3, Notch, Wnt, and nuclear factor-kappaB (NF-κB), Ras/mitogen-activated protein kinase (MAPK) (Ae Kyung Park, 2019). Furthermore, studies have also proven the role of SIRT2 down regulation in patients afflicted with glioblastomas. In addition, vascular endothelial growth factor (VGEF) has also been implicated in the progression of cancer. Protein kinase Cβ (PKCβ), on the other hand, has been reported to activate angiogenesis and growth of cancer in intracranial carcinomas. (Mohamed Alshabi, 2019)

A recent study on the Pakistani population illustrated the role of Poly (ADP-ribose) polymerase 1 (PARP-1) which codes for a nuclear protein that plays an integral role in processes such as cell death and mitosis as well as detection of damage in DNA such as strand breakage and oxidative stress and its subsequent repair. Several single nucleotide polymorphisms have been detected in this gene. It plays an important part in the BER (base excision repair) pathway and hence, any malfunctions in this gene can negatively affect this pathway that can lead to tumorigenesis in the brain (Asad ullah Khan, 2019).

1.5. Problem Statement

Brain cancer often has late diagnosis and poor prognosis; therefore, there is a need of a proteome level analysis underlying the pathogenesis of brain cancer and its progression at a global level for quick and efficient diagnosis of brain cancer.

1.6. Objective

The primary aim of the study was

- The identification of the potential prognostic metabolic markers for brain cancer in the Pakistani population at proteome level through the evaluation of serum samples collected from patients using GC-MS.
- Further In silico analysis of a potential prognostic marker.

Chapter 2

Literature review

2.1. Cancer

Cancer in most cases is caused by a genetic mutation, that is, it is caused by changes in the genes that control the way the cells function, especially how they divide and grow (Weinberg, How Cancer Arises, 1996).

Genetic changes that can lead to cancer can be (Weinberg, How Cancer Arises, 1996):

1. Errors that might occur during the division process
2. Damage caused to the DNA because of external environmental factors such as chemicals in tobacco smoke, UV rays etc.
3. Inherited mutations in the genetic material (Weinberg, How Cancer Arises, 1996).

The body usually has the ability to get rid of the cells that have mutations before they turn cancerous, but this ability of body goes down as we age, which is why chances of cancer increase as one ages (Kapałczyńska, Kolenda, Przybyła, & Zajączkowska, 2018).

There are more than 100 different types of cancer. Types of cancer are usually named based on the organs or tissues where the cancer originates. For example, lung cancer starts in the lung, and brain cancer starts in the brain. Cancer can also be described by the type of cell that formed them, such as squamous cell or epithelial cell (Kapałczyńska, Kolenda, Przybyła, & Zajączkowska, 2018).

Carcinoma

These are one of the most common types of cancer. They are formed by epithelial cells; these cells cover the inside and the outside surfaces of the body. Carcinomas are further differentiated on the basis of different epithelial cells, for example, adenocarcinoma, Basal cell carcinoma, and transitional cell carcinoma (Garcia & Poletti, 2009).

Sarcoma

Sarcomas are the cancer that forms in the soft tissues and bones, of the body including muscle, tendons, fat, blood vessels, lymph vessels, nerves. Osteosarcoma is the most common cancer of the bone (Anderson ME, 2020).

Leukemia

Cancer that begins in the blood forming tissue of the bone marrow is called leukemia. These cancers don't form solid tumors, instead huge number of abnormal white blood cells (Campana & Behm, 2000).

Lymphoma

Lymphoma is the cancer that begins in the lymphocytes, T and B cells, these are the disease fighting white blood cells that are part of the immune system. In lymphoma, abnormal lymphocytes build up in the lymph nodes and lymph vessels (Mugnani, 2016).

2.2. Brain Cancer

The cancers of the brain and the overall nervous system are considered the 10^{th} major cause of deaths worldwide. In 2020, primary brain and CNS tumors caused the deaths of an estimated 251,329 people across the world and an estimated 308,102 were in turn, diagnosed with brain tumors as well. It is also intimated that an estimated amount of 25,050 adults will be diagnosed with brain cancer in 2022 (Board, 2022).

The average annual age-adjusted incidence rate of CNS tumors as well as all malignant and non-malignant brain tumors was seen to be 23.79 in a conducted study (Ostrom QT, 2020).

In Pakistan, 10,114 cases of brain cancer have been reported in the last 5 years (WHO, Pakistan, 2021). According to a study conducted in Agha Khan University, the incidence of brain cancer in males has reportedly been increasing in Pakistan at a rapid pace in the last 30 years with a percentage increase of 10.90% (Shamsi, 2020).

The incidence of brain tumors has been increasing in all ages in the last 20 years; however, it has increased more than 40 percent in adults. Global findings indicate a wide variation in the incidence of brain and CNS tumors with the standardization of age in different countries is between 0.01 and 12.7 in males and 0.01 and 10.7 in females per 100,000 individuals. The lowest incidence is in Africa while the highest

incidence is in northern Europe. The increase in incidence of neuronal cancer in Western countries may be due to improvements in diagnostic techniques such as CT scan and MRI compared to other countries. In addition, the difference in incidence between different regions is due to genetic backgrounds and ethnic differences among different populations. The highest incidence rates were in males of Armenia, Albania, Macedonia, Serbia, Georgia was 12, 12.7, 10.3, 10.8, and 11.9 per 100,000 individuals, respectively. In Sweden, the incidence was higher in females (Salehiniya & Farmanfarma, 2020) (Miller, Ostrom, Kruchko, & Patil, 2021).

According to a report in 2020, the overall death rate for cancers relating to the Brain and Central Nervous System was 3.4% (WHO, Pakistan fact sheet, 2020). It was also known to be the 8th most frequent cause of mortality in Pakistan in 2018 (Farheen Danish, 2021). Worldwide, the mortality rate due to the disease has been charted down to be 2.5% (Hyuna Sung PhD, 2021) and the 10th leading cause of death (Cancer.Net, 2022).

The recurrence rate for brain tumor varies widely amongst different types of the tumor. For example, after the complete removal of Meningioma, the chances of recurrence are low. However, in the next 15 years, the chances are still between 24-32% that the tumor would recur (Chetan Bettegowda, 2022).

The usual type of treatment for Brain cancer is surgery followed by chemotherapy and radiotherapy. After a surgery, bits of tumor tissue are accidently left behind or a high amount of brain tissue already been removed. Moreover, extra removal of brain tissue can lead to neurological deficits in the patient. Therefore, the brain tumor margins are too perplexing to clearly understand to perform accurate surgery (Fabelo H, 2018). There are new innovative therapies being studied to treat brain cancer such as nanomedicine technology (Adip G. Bhargav, 2020) and combinatory treatment strategies where combinations of certain chemotherapy drugs along with radiotherapy is given for targeted treatment (Hennequin C, 2019).

2.3. Types of brain cancer

There is a variety of brain tumor types ranging from benign to malignant and this will determine whether the tumor is cancerous or not. These are divided into two categories; Primary brain tumors, where the tumor emerges from cells within the

Central Nervous system, which can be benign or malignant (Sarah Lapointe, 2018), and Secondary brain tumors, a metastatic tumor that emerges from other parts of the body and these are malignant (Benjamin D. Fox, 2011).

Furthermore, WHO has created a grading system for the evaluation of brain tumors, which includes 4 grades ranging from low grade to high grade. High-grade tumors generally indicate immense malignancy and high aggressiveness of cancer. Meanwhile, low-grade tumors indicate the opposite in comparison to high grade. The following are the common types of brain tumors.

Benign

Benign tumors develop at a slower rate compared to malignant tumors and remain within the brain. Fortunately, they are non-cancerous tumors. Some of these types of tumors include:

- Meningiomas: this is a common intracranial and spinal neoplasm tumor, occurring in the meninges, a membrane between the brain and the skull (Martín-Alonso, 2020).
- Pituitary Adenoma: a common type of pituitary tumor that originates from the pituitary gland and is mostly benign (Huse, 2021).
- Schwannoma: these emerge from differentiated Schwann cells leading to nerve sheath tumors (Jahangir Moini, 2021).
- Craniopharyngioma: it is a rare intracranial tumor that emerges from areas close to the brain that can press on the nerves or parts of the brain (Greg James, 2018).

Malignant Meningiomas

Meningiomas could be benign and malignant depending on their grading scale provided by WHO. Following this, there are 15 meningioma subtypes classified over 3 grades based on their characteristics. Most of the benign meningiomas, which are the most common type, fall under grades I and II. Meanwhile, malignant meningiomas are rare and fall under grade III classification. As a result, malignant meningiomas in comparison to benign show poor prognosis indicating their severity (Martín-Alonso, 2020).

Malignant Gliomas

Gliomas are a type of primary brain tumor that emerges from glial cells in the brain and they are more commonly malignant (Paulina Zofia Gorynska, 2022). The glial cells are the non-neuronal cells within the CNS and PNS. They are involved in providing metabolic and physical support to the neurons (Chung, 2018). These cells are subdivided into three types of cells and these are:

- Astrocytes are involved in the control of blood flow, the formation of synapses, and offer synaptic cleft support (Chung, 2018).
- Oligodendrocytes are important for the formation of the myelin sheath around the axons and this provides insulation of neurons along with enhancing the electrical signals (Sarah Kuhn, 2019).
- Microglia are versatile cells involved in restoring injury and the development of the brain (Marco Prinz, 2019).

There are a variety of different tumors under gliomas and they are characterized by their location and type of cell. Glioblastomas are the highly occurring gliomas followed by astrocytoma, both emerging from astrocytes. According to the WHO grading system, glioblastomas fall under grade IV, high grade, showing its high malignancy due to undifferentiated cells' rapid division. Other tumors like astrocytoma and oligodendroglioma fall under grade I and II, low grade, showing its lower malignancy in comparison to high-grade tumors (Davis, 2018). Moreover, it shows quite a poor prognosis indicating its soaring recurrence rate.

Malignant pineal tumors

These are rare tumors originating in the pineal gland that is the neuroendocrine gland involved in the creation and delivery of melatonin into the bloodstream. These neoplasms are further classified as germinomas, pineal parenchymal tumors, and pineal metastasis. Germinomas are the most common type of pineal neoplasm (Gaia Favero, 2021).

2.4. Global Metabolomics

Metabolomics is the process of analyzing and studying small molecules, commonly known as metabolites present within cells, tissues or bio fluids of organisms. The molecular interaction of these metabolites within a biological system is known as the metabolome. Hence, metabolomics, in simpler words represents the molecular phenotype. Metabolomics of an organism or group of organisms is influenced by both

environmental and genetic factors. The concentration of metabolites can directly reflect the underlying biochemical activity and state of cells and tissues present within the organisms (Idle & Gonzalez, 2007).

Metabolomics can help us understand the biochemical and biological processes that are altered in case of a disease state, progression of a disease, as well as recovery from a disease. Cancer is one of the leading causes of death and high mortality rate from cancer is thought to be from lack of systemic therapies and the high rates of metastasis at the time when cancer is diagnosed. Hence, there is a need to recognize and study more exact molecular biomarkers for use in the clinical practice. Any measurable specific alteration of a cancer cell on the molecular level on DNA, RNA, protein, or metabolite level can be referred to as a cancer biomolecular marker. Using metabolomics for cancer diagnosis can be a breakthrough because, firstly, it can provide metabolic fingerprints, which can play a significant role in early cancer screening, diagnosis, prognosis and therapy. Secondly, it can provide biomarkers produced in response to an anticancer agent being used for cancer treatment and their response, as is shown in the field of toxicology (Liu & Locasale, 2017).

Cancer affects a large proportion of world population every year and hence remains a huge societal problem. The number of cancer patients is expected to increase in the next 50 years (Jemal, Bray, & Center, 2011). The difficulty of understanding cancer not only arises from the unknown factors regarding what triggers its onset and progression, but also due to the difficulty in understanding and detecting the biological events and changes that may lead to the onset of cancer. Even though there are only a small number of biomarkers that are associated with cancer but these can still be used for detection as the metabolic pathways in which these molecular biomarkers are used are associated with different cancers (Qian Bu, 2012).

Altered cell metabolism is one of the main factors associated with any cancer and plays a major role in promoting tumorigenesis, tumor growth and increased resistance of tumor cells to various chemotherapeutic processes. Metabolic alteration of glucose, fatty acids and anaplerotic amino acid metabolism has been associated with malignancy and improved resistance of tumors to various anticancer drugs (Qian Bu, 2012).

In recent metabolomics studies, it was found that 'Metabolic phenotypes' possess great potential for developing anticancer strategies for cancer treatment as well as for monitoring of the treatment process. A number of studies focus on the metabolic characterization of immortalized or cultured cells which included neurons, astrocytes, glial cells and neural system cells, their benign and malignant counterparts showed some interesting findings, for example the elevation or increase in hexamine pathway metabolic intermediates in brain cell lines is linked to cancer cell death following cisplatin treatment. NMR analysis of glioblastoma stem cells identified a high level of Alpha aminoadipate in GBM as a marker of tumor aggressiveness (Renu Pandey, 2017).

2.5. Gas Chromatography-Mass spectrometry

In order to analyze metabolomics data, several techniques are utilized that may be both qualitative and quantitative in their results. The type of technique that is chosen depends on two parameters; experimental objective and sample type. In order to achieve a detailed picture of the metabolomics data, usually two or more independent techniques are combined, sometimes even called "hyphenated techniques". One of those techniques is Gas Chromatography hyphenated to Mass Spectrometry. GC-MS has been described as one of the most methodical and well utilized tool in analytics to date (Beale, 2018). GC-MS is robust, has good separation capability, reproducibility, ease of usage, compound identification, sensitivity and selectivity (Beale, 2018). However, it is only able to separate and identify molecules of low molecular weight (50-600 Da) as well as compounds that are volatile (Beale, 2018). In order to identify non-volatile compounds, the process of chemical derivatization becomes essential as it improves volatility, detector response as well as sensitivity (Poojary, 2016).

In order to obtain reliable metabolomics data, parameters like sample preparation, sample extraction, derivatization, analysis and chromatography settings such as columns and detectors are important. Afterwards, bioinformatics methods are usually employed in order to decipher the data obtained reliant on the sample type/matrix (Beale, 2018). The process has been summarized in Figure 1.

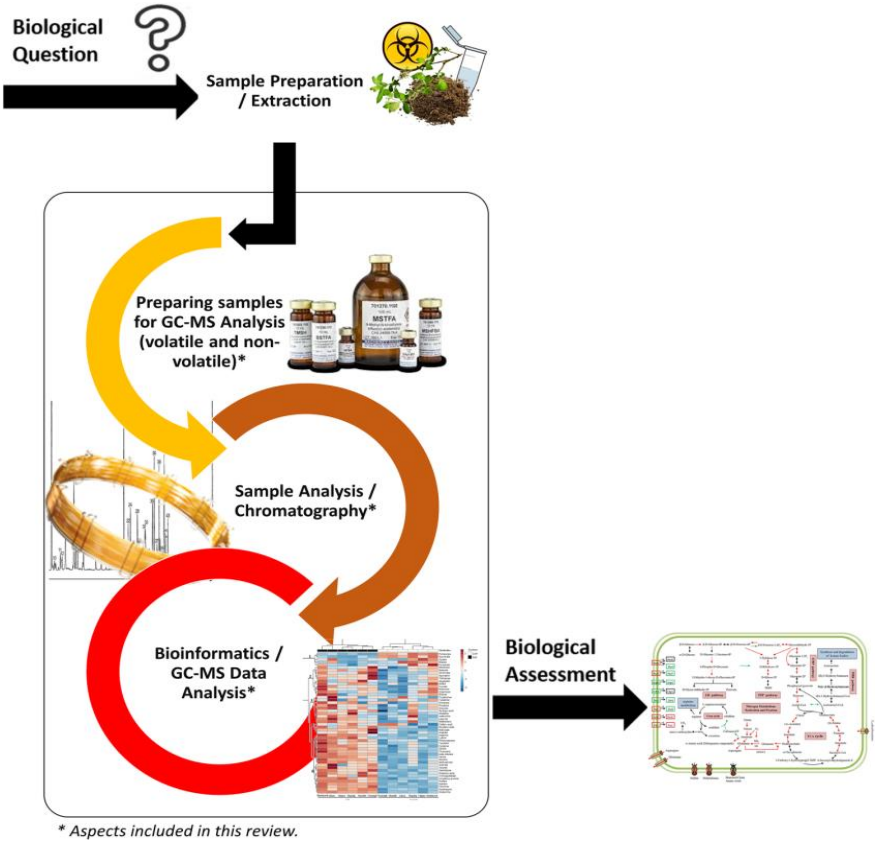

Figure 1: Schematics of GC-MS analysis taken from **(Beale, 2018)**

Chapter 3

Methodology

3.1. Sample collection

The project was first approved by the NUST Ethical review committee and the review board (IRB No. 10-2021-01/02). The samples were then collected from Combined Military Hospital (CMH), Rawalpindi. A total of 6 samples were collected. Three of them consisted of control samples from Healthy individuals (labeled C1, C2 and C3) and the rest from patients who had Brain cancer (labeled B1, B2 and B3). The following criterion (Table 2) was employed in eliciting patient information.

Table 2: Information taken from patients for sample collection

Cancer localization	
Cancer type	Primary Secondary
For primary cancer	
Type	Astrocytoma Meningiomas Oligodendrogliomas Glioblastoma
Molecular features	IDH1 IDH2 MGMT 1p/19q co-deletion Other
Cancer metastasis	Yes No
Cancer grade	I II III IV
Treatment status	Treated Not treated
If yes, duration of treatment	
Any other details (optional)	
Family History	Yes No
If Yes, please state the name of cancer: _____	
Relationship with the patient: _____	
Smoking	Yes No
Alcohol/drug consumption	Yes No

Co-Morbidity	Yes No
If Yes, please state the name of the disease.	

3.2. Sample handling and transportation

The samples were collected in tubes containing a clotting activator and then the tubes were shaken thoroughly and gently. The tubes were stored at room temperature to allow the blood to coagulate. 6 x sample collection vials were then transported in an ice box within one hour to the lab and centrifuged to separate the serum from the blood clot at 2,500 x g (4ºC) for 15 minutes. These were then stored in the freezer at -20ºC.

3.3. Preparation of samples

The samples were prepared using the following protocol (Beale, 2018);

The serum in the upper layer was separated out into 3 separate tubes and labeled B1*, B2* and B3*. This corresponded to the 3 patients from which they were collected from. Afterwards, 300 µl of extraction solvent was added to the serum samples and vortexed for 2 minutes. The samples were then stored on ice for an hour during which they were vortexed every 15 minutes. The cell insoluble matrices were centrifuged at 13,000 rpm for 10 minutes at 4ºC. Finally, the supernatants were collected and transferred to GC vials for drying using EZ-2 Plus at 37(+/-)ºC.

Derivatization

The dry samples were dissolved in 25µl of 20 mg/ml Methoxyamine Hydrochloride in Pyridine. They were then vortexed for 2 minutes and stored for at least 6 hours at 25º C prior to Silylation.

Silylation

25µl of MSTFA + 1% TMCS were added and then dissolved in 100µl of Pyridine. They were then vortexed for 2 minutes. The samples were incubated at 50ºC for 30 minutes. Lastly, they were transferred to 200µl micro-inserts and analyzed by GC-MS.

3.4. GC-MS Analysis

Instrument

The compounds in a mixture are first separated in a gas chromatography chamber and those compounds are converted into ions by using any ionization method in the Mass spectrometry chamber. The molecules can then be detected based on their mass/charge ratio. The mobile phase consists of mobile gas carrier such as N_2, H_2 or He, controls for the carrier gas, an injector for the gas that is heat-controlled, a heater for the column, a data recording device and a detector. The chamber consists of a capillary column that may be coated with the stationary phase consisting of either a solid-phase gas-solid chromatography or a polymeric liquid-phase gas-liquid chromatography. The analytes are carried by an inert carrier gas which are then vaporized and passed over a stationary phase. The analytes have different physical and chemical properties which mean that they interact differently with the stationary phase, some taking longer than others in passing through. This causes them to elute from it at different retention times. This interaction between the analyte and stationary phase is based on the partitioning of the analyte amidst the mobile gas phase and stationary phase.

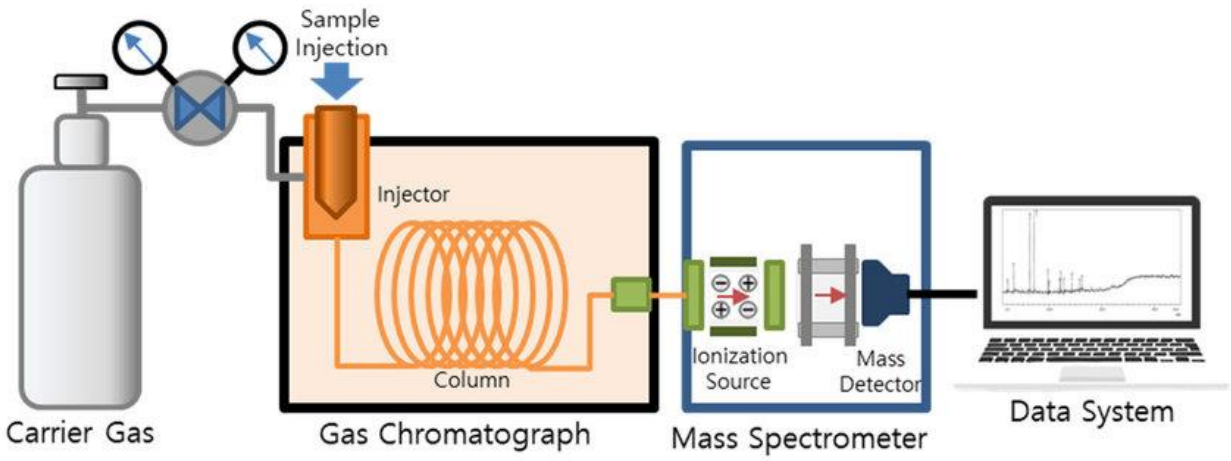

Figure 2: Main components of the GCMS instrument

It mainly consists of 3 components (shown in Figure 2);

1) **Injector:** This aids in injecting the samples into the GC chamber. It would utilize heat in order to vaporize the sample so it could be introduced into the column as a mobile gas phase. These could be three types of inlets;

split/splitless, programmed temperature vaporizers (PTV) and cool-on-column (COC) inlet. For the split/splitless there are two modes that are activated. Split mode vents most of the vapors whereas splitless mode allows most of the transfer when employed. The latter is utilized when analyte concentration is high and the former for trace analysis. PTV is operated on a mixed mode where the samples are concentrated after the solvent has been vented out and then vaporized. In COC, the sample is deposited on the column directly and is employed for those samples that are either sensitive or thermo-labile and is used for trace analysis.

2) **A separation column:** Here, the sample is separated into its individual components based on its interaction with both the mobile and stationary phases.
3) **Mass Spectrometer detector:** Both the concentration and composition of the analytes are detected here. Different analyzers may be used for these purposes which include: Single Quadrupole analyzer, Time of Flight analyzer and Orbitrap analyzer. In the Single Quadrupole Analyzer, two types of voltages (AC & DC) are employed, and based on the stability of the ions in the presence of these two voltages; they can be detected and analyzed differentially from the rest. In Time of Flight analyzer, the ions that are generated are accelerated and since their velocities would depend on their mass/charge ratio, they can be separated and then detected. The Orbitrap Analyzer consists of two electrodes; an inner and an outer one. The ions get trapped in the inner electrode and are then analyzed using Fourier transform spectrometer.

GC-MS equipment specifications

The model of the instrument was 7890 GC containing a 7683 autosampler and 5975C quadruple MSD. The column specifications were; DB-5MS, 30 m, 0.25 mm ID, and 0.25 μm stationary-phase film thickness.

Identification of proteins

The settings and parameters for GCMS were adapted from Abdul-Hamid M. Emwas et al, 2015 (Emwas, 2015). The chromatograms obtained for each of the control and diseased samples were superimposed to generate a final chromatogram with peak

differences for further proteome analysis using scaffold software (shown in Figure 3 ,Figure 4 and Figure 5). The peaks were then identified via NIST library.

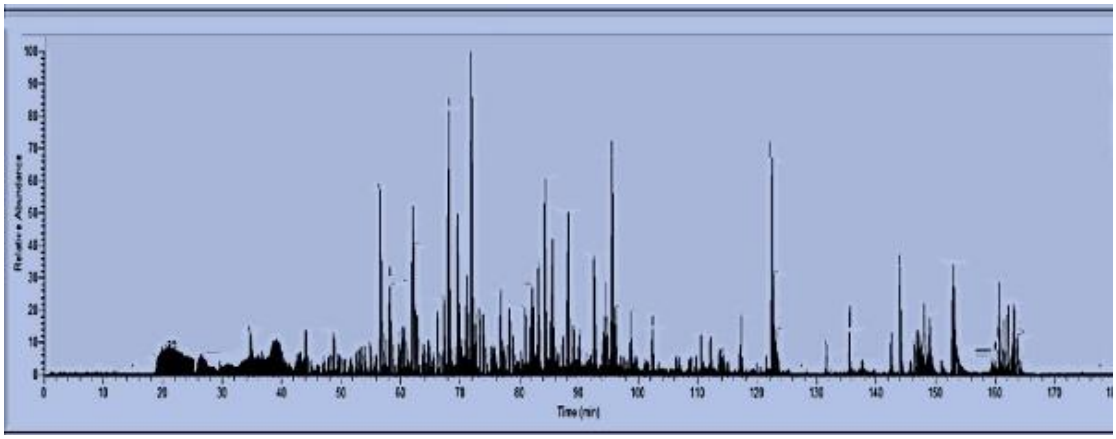

Figure 3: The chromatogram obtained for control samples.

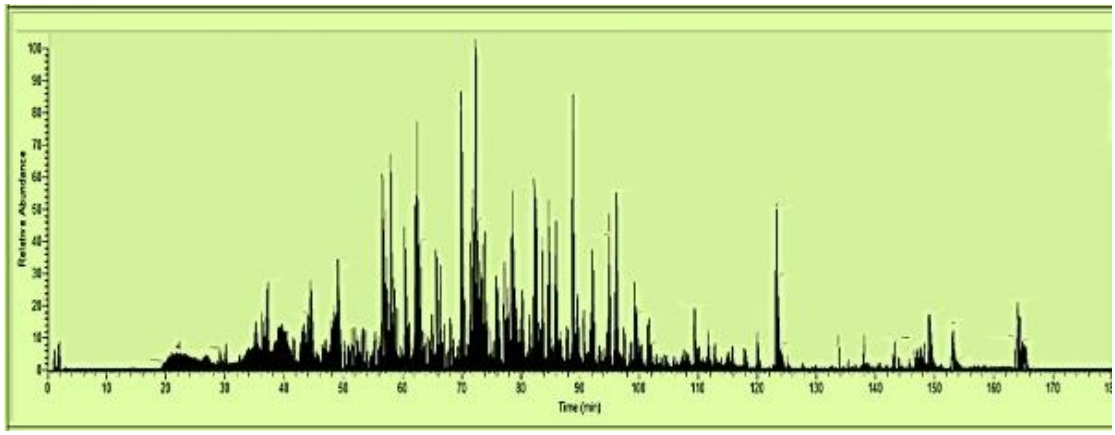

Figure 4: The chromatogram obtained for diseased samples.

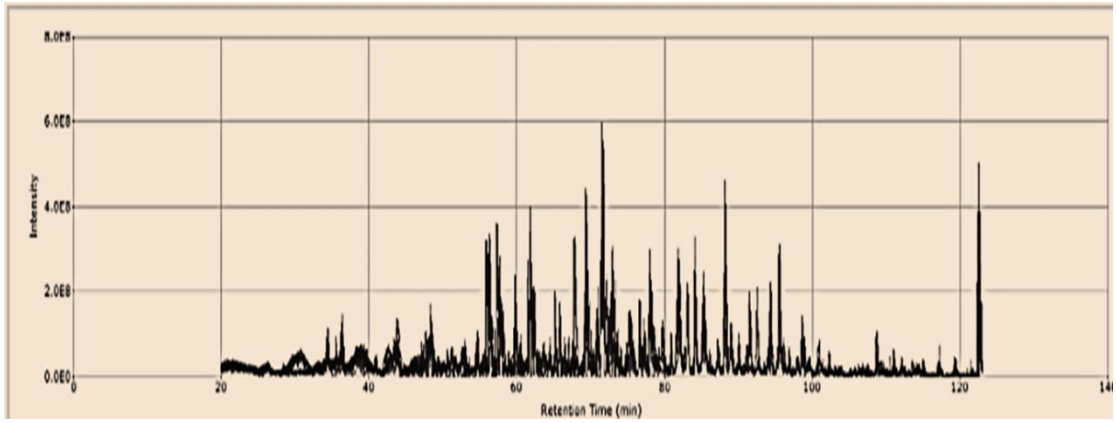

Figure 5: The chromatograms for control and diseased samples were superimposed to produce a final chromatogram with peak differences used for further proteome analysis.

Protein Selection

Upon selection of proteins relevant to our research, T-test on Microsoft excel was applied to the metabolites to confirm and observe the statistical significance of the results. P values less than 0.05 are considered to be statistically significant hence that was the threshold considered in this study (Shrestha, 2019). Additionally, the fold change which is used to calculate the change in gene expression through comparison of the identified proteins in control and cancer samples was calculated by dividing the average of the brain cancer patient samples by the average of the control samples to filter out the proteins needed. A predefined high threshold was selected for filtration so that only those proteins that have a significantly high expression in the cancer samples can be selected for further analysis (Yufei Xiao, 2014).

3.5. Analysis of GC-MS results

Construction of metabolic pathway

Cytoscape (version 3.9.1) was utilized to map, visualize and construct the pathways for each relevant metabolite that was identified using GC-MS. Biorender was then used to generate a collective metabolic signaling pathway by combining those entire constructed using Cytoscape interface. The diagram was constructed to summarize all the proteins' connection to cancer and each other in a signaling pathway.

Cellular localization and Protein class analysis

The PANTHER (Protein Analysis Through Evolutionary Relationships) classification system, which not only allows us to annotate genes in an evolutionary context but also classifies them on the basis of functional relationships (Huaiyu Mi, 2021), was used to retrieve data pertinent to each metabolite which consisted of protein class and the PANTHER derived family and subfamily which were further used to procure the respective cellular localizations of each metabolite. Furthermore, it was used to categorize the metabolites into their respective PANTHER classes.

3.6. In silico analysis of PPP1R163 gene

Human FACT Complex SSRP1 was one of the proteins that had the highest percentage amount in the cells analyzed according to the metabolomics data. According to Cytoscape, this protein gave rise to CDKs that were inhibited by PPP1R163 gene linked to the production of PPP1R163 protein. There have been numerous studies indicating the importance of PPP1R163 as being linked to cancer, in

the sense that it has anti-cancerous effects (Muñoz & Grant, 2022). It inhibits the cancerous cells via several mechanisms by serving multiple purposes; anti-proliferative and pro-differentiation effects, inducing apoptosis, invasion and angiogenesis, immunological effects, etc (Muñoz & Grant, 2022). Hence, this protein and therefore gene is a very important prognostic factor as well as indicator of cancer survivability.

Data retrieval and processing

SNP data was retrieved using ENSEMBL genome browser (gene ID: PPP1R163) which is a software dedicated to the generation and dispensing of genomic annotation (Andrew D Yates, 2020). This included all the variant data that included variant IDs, chromosomal locations, amino acid coordinates, and the relevant wild type to mutant amino acid alteration information. Any pertinent data related to the PPP1R163 protein was derived from Uniprot and Ensembl that included the Uniprot ID's and the respective sequences. PPP1R163 transcript ENST00000549336.6 was used to assess and utilize all SNP related information for the purpose of mapping them. All variant data was retrieved and assessed on February 2022.

Once all the variants were retrieved, all except coding SNPs, missense SNPS, missense splice SNPs, start lost SNPs, start lost – splice region SNPs, stop gained SNPs and stop gained – splice region SNPs were filtered out.

Pathogenicity Analysis

Afterwards, SNPs were subjected to a pathogenicity analysis which was done through prediction of deleterious or tolerant amino acid substitutions using bioinformatics tools SIFT, PolyPhen, Revel, MetaLR, CADD and Mutation Assessor, respectively. The scores of these tools are used classify SNPs as either deleterious or tolerant based on certain score thresholds. A classification score was given for each tool after the variants were exported from ENSEMBL. The score thresholds used for the classification of pathogenic SNPs through identification of deleterious substitutions are illustrated in **Table 3**. The filtered deleterious SNPs were then subjected to further analysis to study their predicted effects on the structure of PPP1R163 protein, the consequent effect on the protein stability as well as the effect of the variation on evolutionary conserved sites.

Table 3: Tools utilized to predict deleterious SNPs

SNP Tools	Score Threshold	References
SIFT	<0.1	(Prateek Kumar, 2009)
PolyPhen	>0.8	(Ivan A Adzhubei, 2010)
Revel	>0.69	(Ioannidis NM, 2016)
MetaLR	>0.5	(Sun H, 2019)
Mutation Assessor	'Medium' and 'High'	(Boris Reva, 2011)
CADD	>30	(Rentzsch P, 2019)

Analysis of evolutionary conservation

Consurf was used to assess the level of conservation of the sites where variations took place through conservation scores that calculate exactly how conserved a particular residue is which directly correlates to its functional importance in the body as well with the higher the score, the higher the degree of conservation. Hence, mutations in evolutionary conserved regions can influence the structure and function of the protein and are more likely to be deemed as pathogenic by the body (Adi Ben Chorin, 2020). Consurf also presented information about the surface accessibility of the protein.

Structural analysis and prediction

Project Hope (Have (y)Our Protein Explained) was employed to predict and assess the changes in the three dimensional structure of the protein due to variations in the sequence through the use of different parameters such as hydrophobicity, the respective charges on the wild type and mutant residue, the domains present as well as the predicted effect of the mutation on the stability of the protein, the interactions between the amino acids such as hydrogen bonding and ionic interactions as well as the interactions between other proteins during signal transduction. HOPE also provides an illustration of the protein so as to visualize the influence of the mutation on the structure of the wild type amino acid (Hanka Venselaar, 2010).

Stability Analysis

The influence of the SNPs on the stability of the protein was also gauged and predicted using I-Mutant3.0 which predicts the effects of single point mutations on the protein based on whether the mutant will have more or less stability depending upon the respective DDG value (protein stability free energy change) of the SNP with a value heading towards less than 0 indicating large destabilization and a value more than 0 indicating increasing stability of the protein (Emidio Capriotti, 2008).

Association with cancer

FATHMM-XF was used to predict the association of PPP1R163 with cancer using a scoring system employed in the website. For every prediction, this bioinformatics tool assigns a confidence score for further interpretation (Mark F Rogers, 2018). Each predicted score directly correlates with each point mutations association with cancer. The residue that was detected as being the most correlated to cancer was then mapped in its respective domain using InterPro that helps analyze proteins on a functional basis by classifying them into families and through the prediction of important sites and domains (Matthias Blum, 2020).

Chapter 4

Results

4.1. GC/MS analysis

The samples underwent GC/MS analysis and the results were compiled according to the T-test on Microsoft excel. The initial results exhibited a variety of identified proteins and T-test was applied to each of them. Initially, there were 852 proteins identified with their calculated P-values from the T-test. The fold change was calculated by dividing the average of brain cancer patient samples by the average of control samples for selected identified proteins, which showed P-values less than or equal to 0.049. After additional filtering, the final results table exhibited identified protein with the specified P-values and fold change greater than 35. These criteria resulted in 25 identified proteins (Table 5).

Identified proteins

The proteins that were identified were grouped according to the effects that they exhibited in relation to cancer. These categories included effects like anti-apoptotic, metastatic, proliferative, angiogenic, Tumor suppressive and immunosuppressive effects detailed in Table 4.

Table 4: Different categories of effects on cancer exhibited by all the proteins involved.

Effects on cancerous tissue	Proteins involved
Anti-apoptotic	RAB5C, ACADV, EIFL3, AT2A1
Metastatic	RPL7, RS8, RAB5B, DHE3, PSA7, SEPT2, RPL3
Proliferative	SSRP1, MET7A, RAB5B, RAB5C, DHE3, SEPT2, LAMC1, SMD3, SEPT7, RPL3, ARPC4, RL18
Angiogenic	RPL7
Tumor suppressive	SSRP1, RPL7, RPS13, RPL4, RPL8, RL13, RAB5B, RL18A, SF3B3, AT2A2, RL18
Immunosuppressive	LAMC1

The information was displayed on a pie chart (Figure 6) to see how many proteins that were identified exhibited each of the effects. Most of the proteins showed proliferative and tumor suppressive effects whereas only one had angiogenic effect and one other had Immunosuppressive properties.

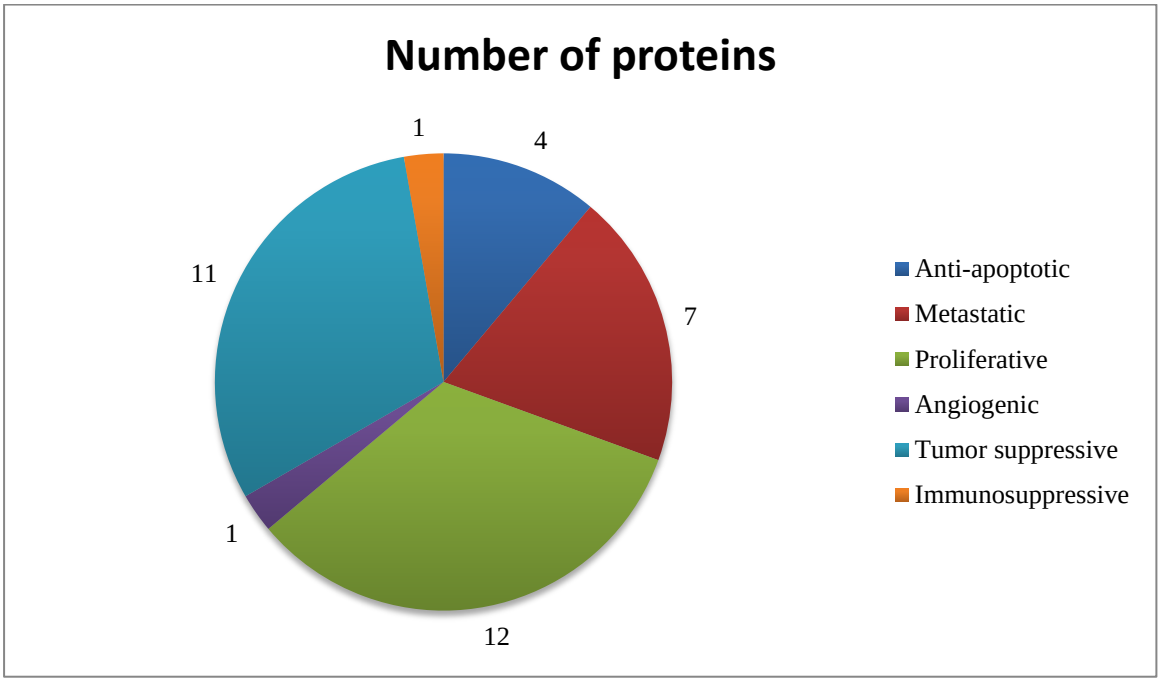

Figure 6: Pie Chart showing the number of proteins that exhibited each type of effect on the cancerous tissue.

Anti-apoptotic

RAB5C

Involved in ESCRT dependent/independent pathway of exocytosis of exosomes, this protein leads to the release of chemicals that lead to inhibited apoptosis (Datta A, 2017).

ACADV

This protein was seen to have an important role in increasing rates of fatty acid beta synthesis (Figure 11) to increase energy production in cancer cells for apoptosis evasion. The end product of this metabolic process; Acetyl CoA, is an important molecule for generating energy through the Krebs Cycle (Poulogiannis, 2019).

EIFL3

This is linked with the activation of hsa-mir-501, which is connected to IGBP1. IGBP1 is seen to have an oncogenic effect in which it exhibits anti-apoptotic activities (Sicong Jiang, 2020).

AT2A1

This is a part of the endoplasmic reticulum interacting with IP3 receptors. It is regulated by the anti-apoptotic protein Bcl-2, which leads to inhibition of apoptosis (Denise Casemore, 2015)**.**

Metastatic

RPL7

It activates hsa-mir-639 that is a tumor suppressor, promoting metastasis (Wang YH, 2018).

RS8

This protein is involved in metastasis in cancer calls and is seen to down regulate hsa-mir-7-1 in brain tissue while being upregulated itself (Korać, 2021)**.**

RAB5B

This protein is involved in migration and invasion in cancer (Kong, Zhang, Li, Shao, & Fang, 2018)**.**

DHE3

This is an important enzyme with a role in amino acid metabolism involved in conversion of glutamate to alpha ketoglutarate which is a key energy source in glioma cells (Renaud Vatrinet, 2017). Increased synthesis of alpha ketoglutarate leads to increased production of citrate which is a pivotal molecule in brain cancer cells as it is involved in their growth and development (Jordan K, 2021).

PSA7

This protein is involved in the activation of hsa-mir-302c and is associated with inhibition of glioma malignancy (Wang Y, 2015) as well as cancer cell metastasis (Yang, 2018).

SEPT2

This protein has a primary role in DNA integrity checkpoints in the cell cycle and it also activates hsa-mir-326 which has a highly important role in cancer as it is involved in cellular invasion and metastasis (Zahra Ghaemi, 2019)**.**

SEPT7

This protein is involved in the activation of hsa-mir-1266, which is linked to cell proliferation. This also correlates to hsa-mir-3647, which is linked to CDKL1 and DDX39, both of which promote cell proliferation and metastasis, respectively (Chunzhi Qin, 2017).

RPL3

This protein is seen to be interacting with MTDH. Overexpression of MTDH leads to activation of the nuclear factor κB (NFκB), and Wnt/β-catenin signaling pathways to increase cell invasion, and cell survival (Feng, 2019).

RL18

This is involved in activation of hsa-mir-28 which is involved in metastasis (Lv, 2019).

Pro-proliferative

SSRP1

Mainly involved in proliferation, it interacts with MDM2 leading to the activation of PPP1R163 gene (Konopleva, 2020)**.**

MET7A

This protein is down regulated by hsa-mir-200c which itself is upregulated and increases proliferation (Yu, 2010).

RAB5B

While this protein itself is involved cellular proliferation, migration and invasion in cancer, it also down regulates hsa-mir 99b which has a role in cellular proliferation (Kong, Zhang, Li, Shao, & Fang, 2018).

RAB5C

Involved in ESCRT dependent/independent pathway of exocytosis of exosomes, this protein leads to the release of chemicals that lead to enhanced tumor growth (Datta A, 2017).

DHE3

An important enzyme with a role in amino acid metabolism involved in conversion of glutamate to alpha ketoglutarate which is a key energy source in glioma cells (Renaud Vatrinet, 2017). Increased synthesis of alpha ketoglutarate leads to increased production of citrate which is a pivotal molecule in brain cancer cells as it is involved in their growth and development (Jordan K, 2021).

SEPT2

This protein has a primary role in DNA integrity checkpoints in the cell cycle and it also activates hsa-mir-326 which has a highly important role in cancer as it is involved in cellular proliferation (Zahra Ghaemi, 2019).

LAMC1

This protein along with other laminins (LAMB1 and LAMA5) is activated by CFB which in turn is activated by complement protein C3 which plays an important role in brain cancer progression by enhancing tumor growth. CFB is also seen to be involved in cellular proliferation (Shimazaki R, 2021). The complement protein production is activated by NFkb which plays a critical role in cancer development (Liu M, 2018).

SMD3

This protein is a component of U1 snRNP which is a part of the spliceosomal E complex. It aims at the start of the pre-mRNA that needs to be discarded. Unregulated splicing activity leads to alterations in mRNA recognition sites. This ends with elevated transcripts of certain genes leading to increased cell proliferation (Pamela Bielli, 2019).

SEPT7

This protein is involved in the activation of hsa-mir-1266, which is linked to cell proliferation. This also correlates to hsa-mir-3647, which is linked to CDKL1 and DDX39, both of which promote cell proliferation (Chunzhi Qin, 2017)**.**

RPL3

This protein is seen to be interacting with MTDH. Overexpression of MTDH leads to activation of the nuclear factor κB (NFκB), and Wnt/β-catenin signaling pathways to increase proliferation and cell invasion (Feng, 2019).

ARPC4

This is a part of a complex with ARPC2 and ARPC3. It is part of the Rac1 pathway and is involved in the production of cytoplasmic cytoskeleton. Brk1 controls the state of ARPC4 and it leads to cell proliferation (Escobar, et al., 2010).

RL18

This is involved in activation of hsa-mir-28 which is involved in tumor progression (Lv, 2019)

Angiogenic

RPL7

It interacts with hsa-mir-632 that negatively regulates TFF1 leading to tumor angiogenesis and endothelial recruitment (Shi, 2019).

Tumor suppressor

SSRP1

It interacts with MDM2 leading to the activation of PPP1R163 gene. MDM2 acts as a tumor suppressor (Konopleva, 2020).

RPL7

It activates hsa-mir-639 that is a tumor suppressor, promoting metastasis (Wang YH, 2018).

RPS13, RPL4 and RPL8

Collectively, these ribosomal proteins activate the NF kappa B pathway leading to the activation of DDX3X, a death domain protein that inhibits cell cycle repressor KLF4 and leads to tumorigenesis (T.C, 2019).

RL13

This is seen to be upregulated in brain cancer and is activated by hsa-mir 378 and 378c which acts as a tumor suppressor (Nauclér, Geisler, & Vetvik, 2019).

RAB5B

While this protein itself is involved in cellular proliferation, migration and invasion in cancer, it also down regulates hsa-mir-99b which acts as a tumor suppressor and has a role in cellular proliferation and apoptosis (Kong, Zhang, Li, Shao, & Fang, 2018).

RL18A

This protein is upregulated in brain cancer and activates hsa-mir-7-1 that is seen to play a characteristic role in brain cancer by acting as a potential tumor suppressor in glioblastomas (Zhenlin Liu, 2014).

SF3B3

This protein is activated by hsa-mir 26b, 98, 19b-1 and can act as a cancer suppressor. (Chen, Xiao, & Zeng, 2017)**.**

AT2A2

This is a part of the endothelin and calcium signaling pathway in the endoplasmic reticulum. It is seen to be a tumor suppressor with IP3 showing apoptotic effects in certain tumors (Wei-Qing Li, 2017).

RL18

This is involved in activation of hsa-mir 1293 which prevents proliferation and cell invasion and hence acts as a tumor suppressor (Wen Luo, 2017).

Immunosuppression

LAMC1

This protein along with other laminins (LAMB1 and LAMA5) is activated by CFB which in turn is activated by complement protein C3 which plays an important role in brain cancer progression by enhancing tumor growth through immune suppression while CFB is seen to be involved in cellular proliferation (Shimazaki R, 2021).The complement protein production is activated by NFkb which plays a critical role in cancer development (Liu M, 2018).

Cellular localization of identified proteins

According to panther software, all proteins had shown different localizations indicating the diversity of the results. Ten of the proteins were localized in the ribosomes, five were a part of the membrane, and four were present in the endoplasmic reticulum. Meanwhile, fewer proteins were present in other compartments of the cell. Three of the proteins were present in the nucleus, three were a part of the cytoskeleton, two resided in the mitochondria, the other two were part of the extracellular matrix, and the remaining one protein was present in the vacuole. These were compiled into a pie chart to illustrate the percentage of proteins in each part of the cell (Figure 7).

Panther class of identified proteins

Panther classes of each protein were derived from the software as well. Except for one, SSRP1, which panther did not classify into any class, all others were distributed into their respective classes. A higher percentage of the proteins were ribosomal, while a lower percentage of the proteins were dehydrogenases, small GTPases, cytoskeletal proteins and primary active transporters (Figure 8).

Signaling pathway construction

Investigation of each protein’s pathway led to understanding its connection to cancer using Cytoscape interface (Figure 9). They exhibited their own link and role in a pathway leading to cancer. Finally, a diagram was drawn to summarize all the proteins' connection to cancer and with each other in a signaling pathway using Biorender. All the proteins lead to cancer either indirectly or directly (Figure 10). However, some of the proteins had shown a link to brain cancers specifically.

Table 5: Identified proteins with P value less than or equal to 0.049 and fold change greater than 35. The analysis was replicated three times to ensure the accuracy of these results.

Identified Proteins (800/815)	T-Test (P-Value)	Foldchange (patient/control)
SSRP1_HUMAN FACT complex subunit SSRP1 OS=Homo sapiens GN=SSRP1 PE=1 SV=1	95% (< 0.00010)	52.791
RL7_HUMAN 60S ribosomal protein L7 OS=Homo sapiens GN=RPL7 PE=1 SV=1	95% (< 0.00010)	59.195
RS13_HUMAN 40S ribosomal protein S13 OS=Homo sapiens GN=RPS13 PE=1 SV=2	95% (< 0.00010)	38.3239
RL4_HUMAN 60S ribosomal protein L4 OS=Homo sapiens GN=RPL4 PE=1 SV=5	95% (0.00011)	91.442
RAB5C_HUMAN Ras-related protein Rab-5C OS=Homo sapiens GN=RAB5C PE=1 SV=2	95% (0.00034)	63.839
RL8_HUMAN 60S ribosomal protein L8 OS=Homo sapiens GN=RPL8 PE=1 SV=2	95% (0.00050)	48.0435
RS8_HUMAN 40S ribosomal protein S8 OS=Homo sapiens GN=RPS8 PE=1 SV=2	95% (0.00051)	119.945
RL13_HUMAN 60S ribosomal protein L13 OS=Homo sapiens GN=RPL13 PE=1 SV=4	95% (0.0015)	78.306
RAB5B_HUMAN Ras-related protein Rab-5B OS=Homo sapiens GN=RAB5B PE=1 SV=1	95% (0.0018)	46.386
RL18_HUMAN 60S ribosomal protein L18 OS=Homo sapiens GN=RPL18 PE=1 SV=2	95% (0.0019)	52.791
SF3B3_HUMAN Splicing factor 3B subunit 3 OS=Homo sapiens GN=SF3B3 PE=1 SV=4	95% (0.0021)	41.742
MET7A_HUMAN Methyltransferase-like protein 7A OS=Homo sapiens GN=METTL7A PE=1 SV=1	95% (0.0022)	46.386
ACADV_HUMAN Very long-chain specific acyl-CoA dehydrogenase, mitochondrial OS=Homo sapiens GN=ACADVL PE=1 SV=1	95% (0.0026)	69.018

Identified Proteins (800/815)	T-Test (P-Value)	Foldchange (patient/control)
DHE3_HUMAN Glutamate dehydrogenase 1, mitochondrial OS=Homo sapiens GN=GLUD1 PE=1 SV=2	95% (0.0036)	51.1329
RL18A_HUMAN 60S ribosomal protein L18a OS=Homo sapiens GN=RPL18A PE=1 SV=2	95% (0.0048)	57.4345
PSA7_HUMAN Proteasome subunit alpha type-7 OS=Homo sapiens GN=PSMA7 PE=1 SV=1	95% (0.0060)	70.244
SEPT2_HUMAN Septin-2 OS=Homo sapiens GN=SEPT2 PE=1 SV=1	95% (0.011)	55.7769
LAMC1_HUMAN Laminin subunit gamma-1 OS=Homo sapiens GN=LAMC1 PE=1 SV=3	95% (0.016)	78.409
SMD3_HUMAN Small nuclear ribonucleoprotein Sm D3 OS=Homo sapiens GN=SNRPD3 PE=1 SV=1	95% (0.018)	49.3724
SEPT7_HUMAN Septin-7 OS=Homo sapiens GN=SEPT7 PE=1 SV=2	95% (0.018)	63.839
EIF3L_HUMAN Eukaryotic translation initiation factor 3 subunit L OS=Homo sapiens GN=EIF3L PE=1 SV=1	95% (0.019)	49.3724
Cluster of AT2A2_HUMAN Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 OS=Homo sapiens GN=ATP2A2 PE=1 SV=1 (P16615)	95% (0.024)	56.1045
AT2A1_HUMAN Sarcoplasmic/endoplasmic reticulum calcium ATPase 1 OS=Homo sapiens GN=ATP2A1 PE=1 SV=1	95% (0.037)	56.1045
RL3_HUMAN 60S ribosomal protein L3 OS=Homo sapiens GN=RPL3 PE=1 SV=2	95% (0.024)	35.338
ARPC4_HUMAN Actin-related protein 2/3 complex subunit 4 OS=Homo sapiens GN=ARPC4 PE=1 SV=3	95% (0.034)	48.0435

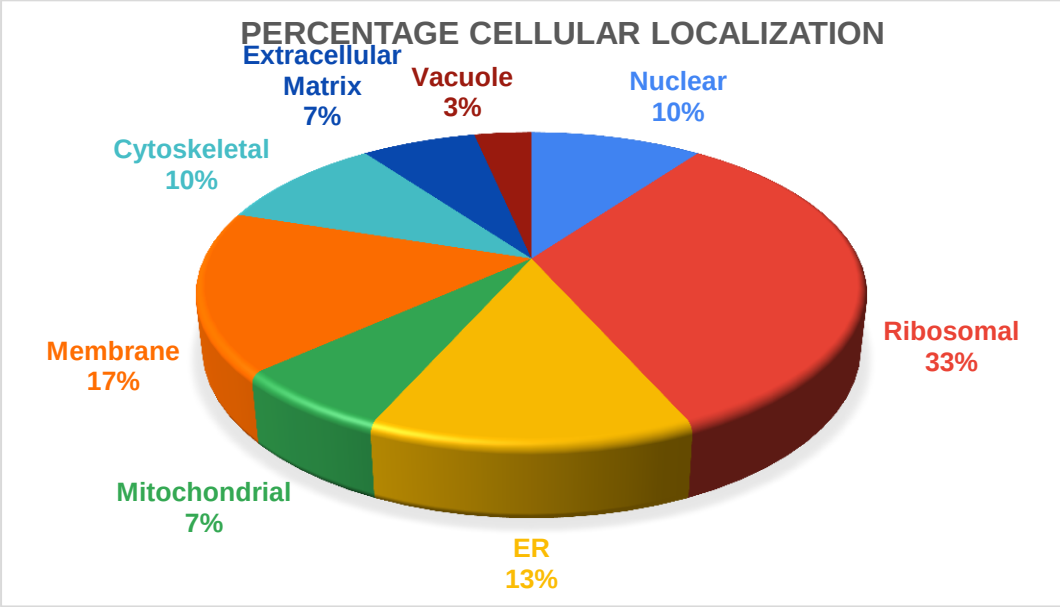

Figure 7: Summarized cellular localization of the 25 proteins shown in percentage format.

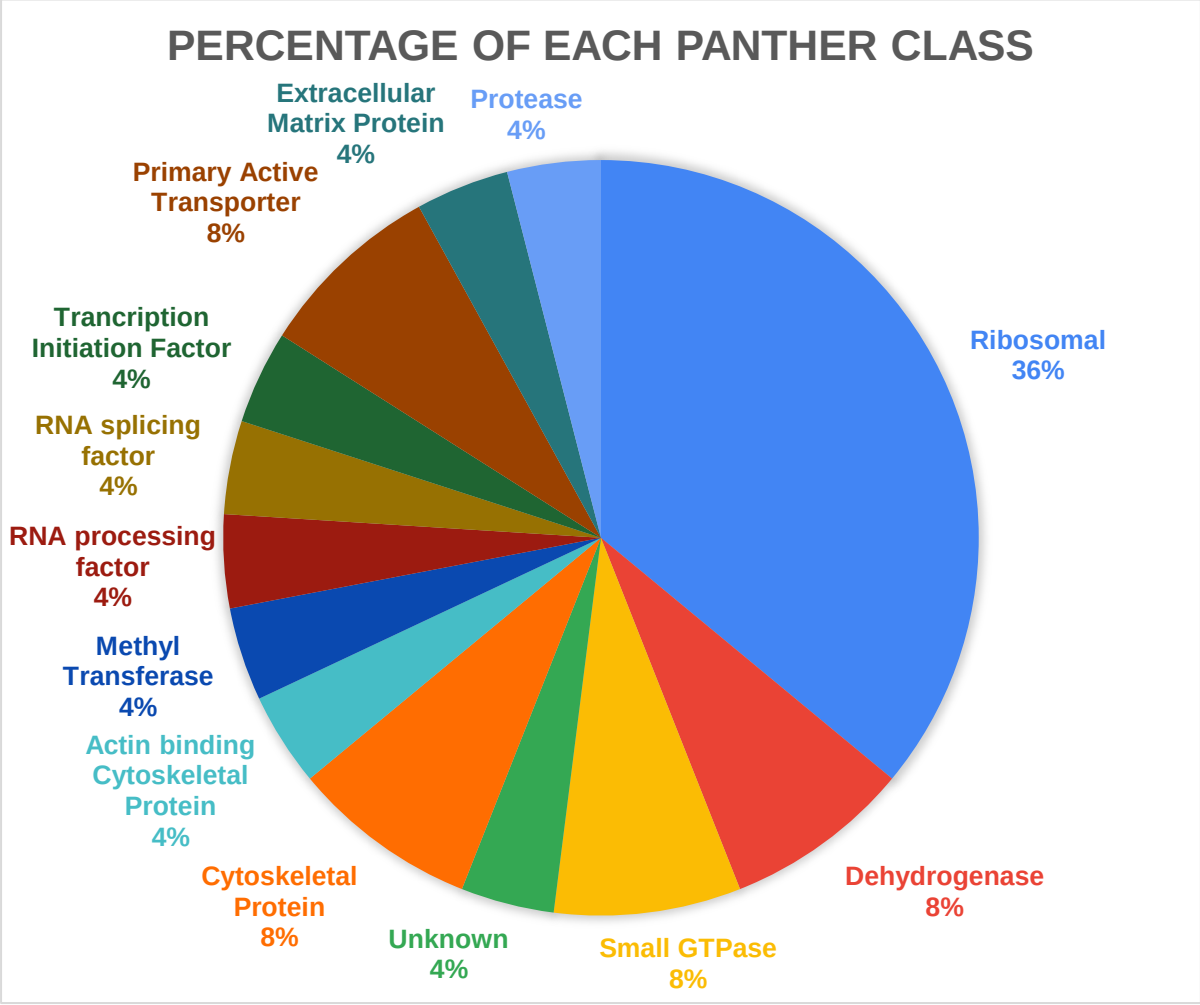

Figure 8: Distribution of percentage Panther class for each identified protein

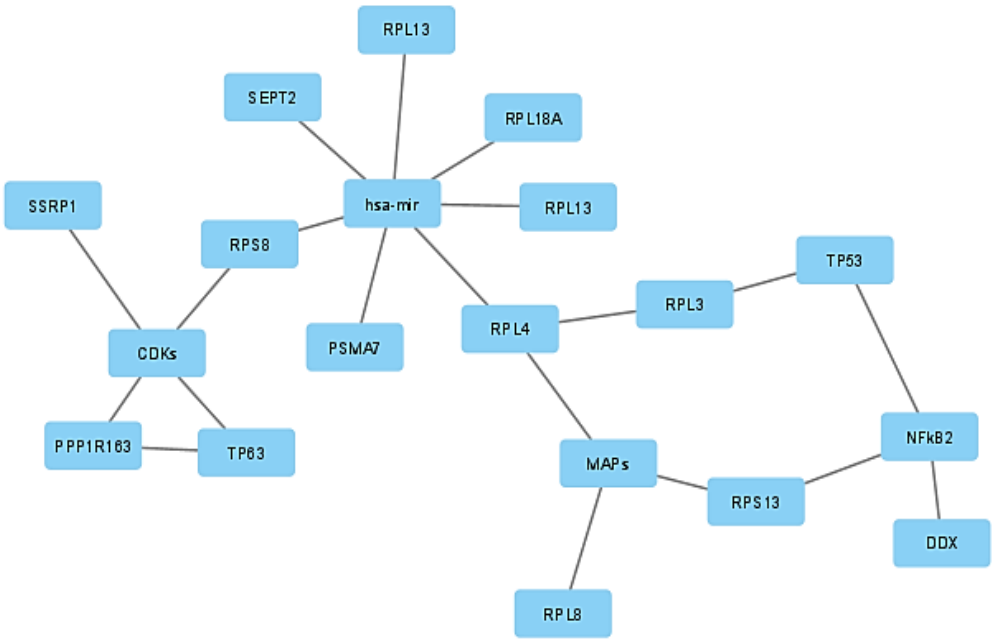

Figure 9**:** Representation Figure of pathways derived from Cytoscape interface

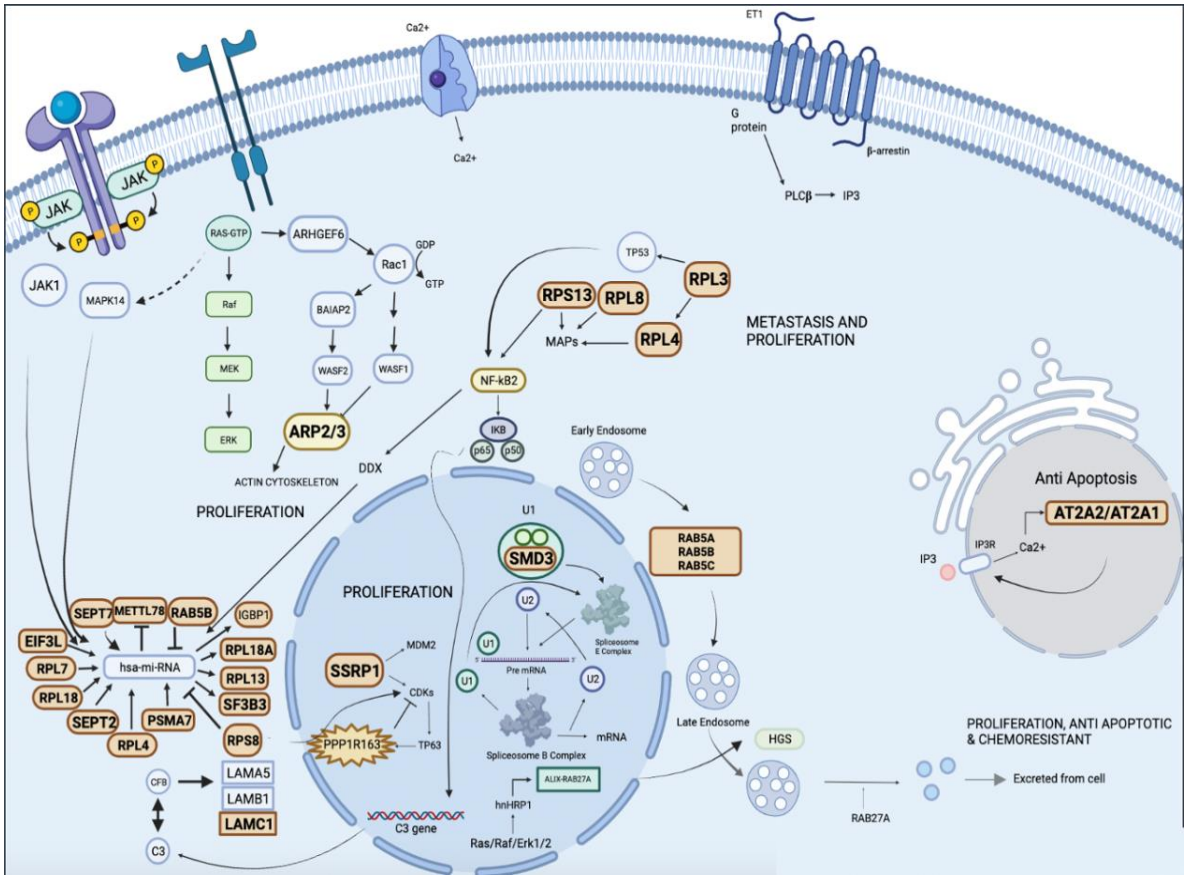

Figure 10**: Cellular localized identified proteins involved in the pathways of the cell and other components like nucleus and the endoplasmic reticulum exhibiting proliferative, anti-apoptotic, and metastatic effects**. The activation from the cell surface receptors like cytokine receptors and GPCRs lead to the activation of multiple signaling molecules. This ends with the main identified proteins activating and indirectly or directly leading to cancer stimulation. Many of the proteins like SEPT 7/2 and RPL4 exhibit a link to hsa-mi-RNA which is linked to proteins leading to proliferation of the cell. The proteins involved in the nucleus, SSRP1 and SMD3, have also shown link to cancer through proliferation of the cells. They also show a link to the gene under further investigation which was PPP1R163. The proteins in the endoplasmic reticulum, AT2A2 and AT2A1, have shown anti-apoptotic abilities along with IP3 and thus indicating tumor suppressive qualities.

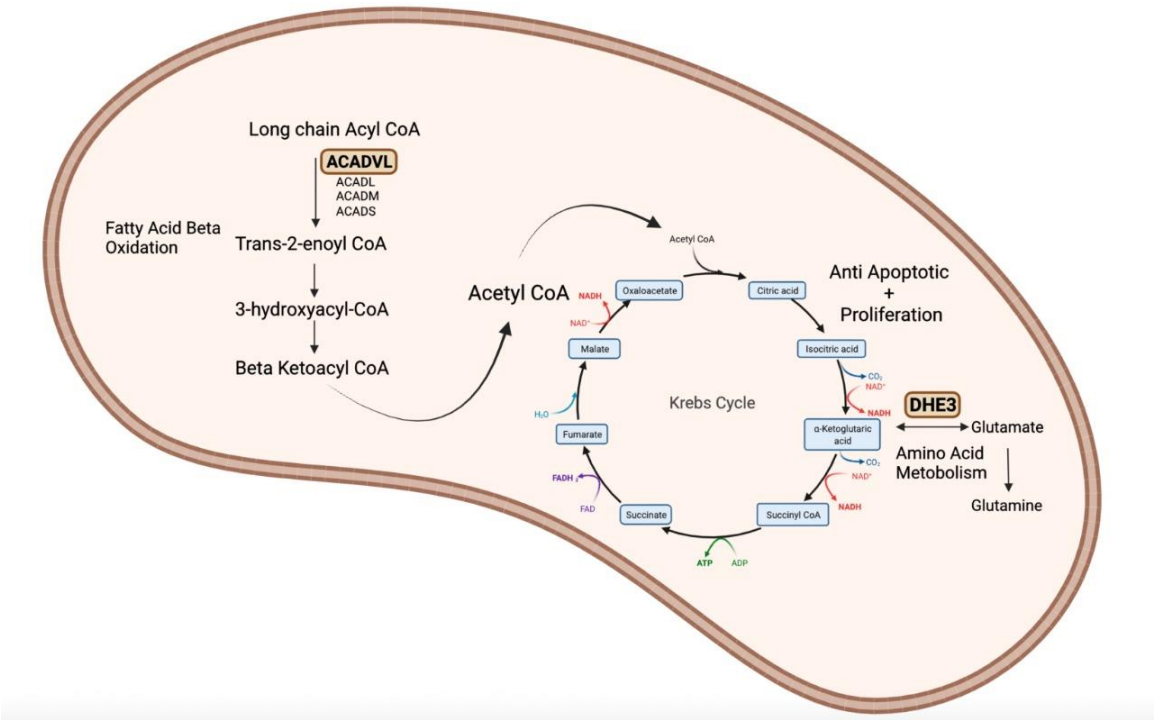

Figure 11: **Mitochondrial proteins exhibiting anti apoptotic effects and proliferation of the cells in brain cancer**. ACADVL is seen to increase rates of fatty acid beta oxidation with its final product; Acetyl CoA, being an important molecule for energy production in the Krebs Cycle. DHE3 has a characteristic role in amino acid metabolism leading to synthesis of alpha-ketoglutarate which in turn leads to production of more citrate; an important organic compound for cell growth in brain cancer.

4.2. Pathogenic SNPs in PPP1R163

In order to envision the pathogenicity of the SNPs, a variety of bio-informatics tools were utilized. A total of 379 nsSNPs were collected from ENSEMBL database. These were then subjected to filtration by 6 tools; SIFT, PolyPhen, REVEL, MetaLR, Mutation Assessor, and CADD. SIFT classified 177 nsSNPs as deleterious with five of them being low confidence deleterious and 25 nsSNPs as tolerated with one being low confidence. PolyPhen narrowed it down to 173 nsSNPs being possibly or probably damaging. REVEL, then, classified 100 nsSNPs as likely disease causing. MetaLR sorted 91 nsSNPs as damaging. Mutation Assessor sorted 22nsSNPs as high and 90 nsSNPs as medium. CADD narrowed it down to 7 nsSNPs as likely deleterious. These 7 nsSNPs were shown to be deleterious by all of the tools applied. Classification of these SNPs was done by selected scores of the tools; SIFT D <0.1, PolyPhen > 0.8, REVEL >0.69, MetaLR >0.5, CADD >30 (Table 7). Upon further investigation, the 7 SNPs were organized on the basis of their Variant ID into a table along with their alleles, amino acid and amino acid coordinates (Table 6).

Table 6: The 7 SNPS with their respective IDs, alleles and amino acid Coordinates.

Serial No	Variant ID	Alleles	AA	AA Coordinates
1	rs1283059621	G/A	S/F	405
2	rs1305311708	A/G	L/P	387
3	rs1057521095	G/A	R/C	343
4	rs757881350	G/A	R/C	274
5	rs113677053	T/C	E/G	269
6	rs946531624	T/A	K/M	246
7	rs1370224730	T/G	E/A	92

Table 7: SNPs classified based on the 6 tools and the scores used.

Variant ID	SIFT	PolyPhen	CADD	REVEL	MetaLR	Mutation Assessor
rs1283059621	D (0.02)	D (0.979)	D (31)	D (0.738)	D (0.884)	M (0.798)
rs1305311708	D (0)	D (0.995)	D (32)	D (0.981)	D (0.96)	H (0.95)
rs1057521095	D (0)	D (0.997)	D (32)	D (0.958)	D (0.943)	M (0.929)
rs757881350	D (0)	D (0.993)	D (32)	D (0.963)	D (0.944)	M (0.9)
rs113677053	D (0)	D (0.914)	D (32)	D (0.949)	D (0.964)	M (0.914)
rs946531624	D (0)	D (0.999)	D (31)	D (0.954)	D (0.967)	M (0.904)
rs1370224730	T (0.05)	D (0.991)	D (33)	D (0.907)	D (0.927)	M (0.875)

4.3. Evolutionary conservation of 7 pathogenic variants

Mutations in sequences that have remained highly conserved in proteins can have a huge negative impact on its function as the degree of conservation directly correlates to the functional importance of the protein.

The conservation scores which were retrieved from Consurf provide insight into the evolutionary conservation status of each of the 7 pathogenic SNPs. Based on the scores as depicted in Table 8, L387P, R343C, R274C, E269G, K246M were more conserved than S405F and E92A. The surface accessibility of each protein was also obtained through which the functional role of a particular amino acid can be

predicted. According to Consurf, L387P was the only residue buried which could be predicting its structural importance in the cell while S405F, R343C, R274C, E269G, K246M and E92A were all exposed meaning their role may lie in signaling pathways as well.

Table 8: Consurf derived conservation Scores, predicted evolutionary conservation and surface accessibility for the 7 SNPs

Residue	Conservation Score	Conservation	Surface Accessibility
S405F	-0.419	Conserved	Exposed
L387P	-0.908	Highly Conserved	Buried
R343C	-0.835	Highly Conserved	Exposed
R274C	-0.835	Highly Conserved	Exposed
E269G	-0.969	Highly Conserved	Exposed
K246M	-1.013	Highly Conserved	Exposed
E92A	-0.500	Conserved	Exposed

4.4. Seven deleterious SNPs mapped across different domains of PPP1R163

In order to further elucidate the structural localization of the seven deleterious SNPs, they were mapped across the different domains of PPP1R163 protein as shown in Figure 12. All 7 of the SNPs were highly conserved with residue E92A being the least conserved comparatively. R343C, E269G, K246M, S405F, R274C and L387P were all located in the Hormone ligand binding domain that is responsible for binding Vitamin-D3. SNP E92A, on the other hand, was present in the zinc-finger DNA binding domain.

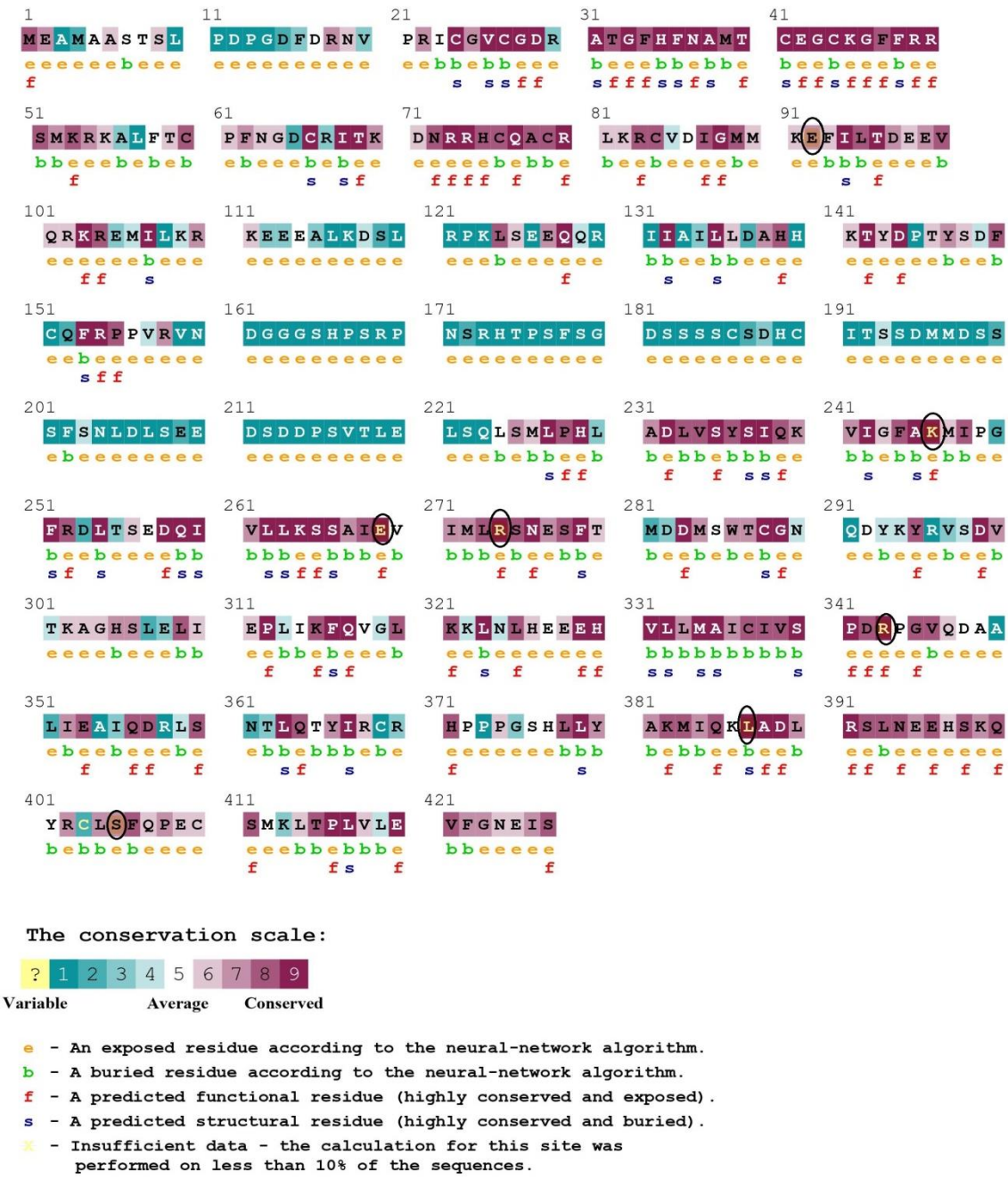

Figure 12: The 7 deleterious SNPs mapped across the full length protein at positions 92, 246, 269, 274, 343, 387 and 405 indicated by a black circle drawn on the residue.

4.5. Influence of PPP1R163 missense variants on its structure

Missense mutations lead to the wrong amino acid being incorporated into the protein, inevitably affecting the structure and hence, the function of that protein. The magnitude of the effect these changes might have depend on what domain the mutation is present in, how much surface accessibility it has and whether the area it's present in is highly evolutionary conserved or not. Seven of the mutated PPP1R163 SNPs were studied by assessing their effect on the structure and function of the

protein. For this purpose, project HOPE was employed that yielded information regarding the properties of mutated amino acids as well as the impact that they might have. Amongst these variants, only one brought about the mutation that was bigger in size than the wild-type residue i.e. S405F. Five of the residues brought about hydrophobicity and five were neutral compared to the charged wild-type residues. Most of the residues affected hydrogen bond formation regardless of their size and charge. The neutral and more hydrophobic residues seemed to cause disturbance in ionic interactions and only two affected local stability both with different charge and size. Four residues in particular seemed to have a significant impact on protein function namely; L387P, R274C, K246M and E92A. The information found using the software has been summarized in Table 9.

Table 9: List of residues and their impact on PPP1R163 structure and function taken from project HOPE

Variant	**Size**	**Charge**		**Hydrophobicity**	**Domain**	**Conservation**	**Impact**
		Wild type	**Mutant**				
S405F	Bigger	.	.	.	Nuclear hormone receptor, ligand binding domain; Nuclear Hormone Receptor-Like Domain Superfamily	.	Affect local stability, ligand contacts and hydrogen bond formation
L387P	Smaller	.	.	.	Nuclear Hormone Receptor, Nuclear Hormone Receptor, Ligand-Binding Domain, Nuclear Hormone Receptor-Like Domain Superfamily	.	Affect signal transduction and protein function
R343C	Smaller	+	Neutral	More	Nuclear Hormone Receptor, Ligand-Binding Domain, Nuclear Hormone Receptor-Like Domain Superfamily	.	Affect hydrogen bond formation and disturb ionic interaction
R274C	Smaller	+	Neutral	More	Nuclear Hormone Receptor, Ligand-Binding Domain, Nuclear Hormone Receptor-Like Domain Superfamily	.	Disturbed protein function due to loss of ligand interaction, affect hydrogen bond and disturb ionic interaction

Variant	Size	Charge		Hydrophobicity	Domain	Conservation	Impact
		Wild type	Mutant				
E269G	Smaller	-	Neutral	More	Nuclear Hormone Receptor, Ligand-Binding Domain, Nuclear Hormone Receptor-Like Domain Superfamily	.	Affect local stability, hydrogen bond formation and disturbed ionic interaction
K246M	Smaller	+	Neutral	More	Nuclear Hormone Receptor, Ligand-Binding Domain, Nuclear Hormone Receptor-Like Domain Superfamily	.	Affect signal transduction and protein function
E92A	Smaller	-	Neutral	More	Nuclear Hormone Receptor, Zinc Finger, Nuclear Hormone Receptor-Type, Zinc Finger, Nhr/Gata-Type, Vdr, Dna-Binding Domain	Very Conserved	Disturbed protein function

4.6. Influence of PPP1R163 missense variants on its stability

All the 7 variants were further investigated to analyze their effect on stability using I-Mutant3.0. This was achieved based on the DDG value (Delta Delta G value) where a value below zero would indicate that the protein is destabilizing. This resulted in 5 of the 7 variants causing destabilization and the rest were stabilizing. Variants L387P, R343C, R274C, E269G, and E92A exhibited the lowest DDG values indicating their destabilizing effect. Variants S405F and K246M exhibited DDG values higher than zero indicating their stabilizing effects (Figure 13).

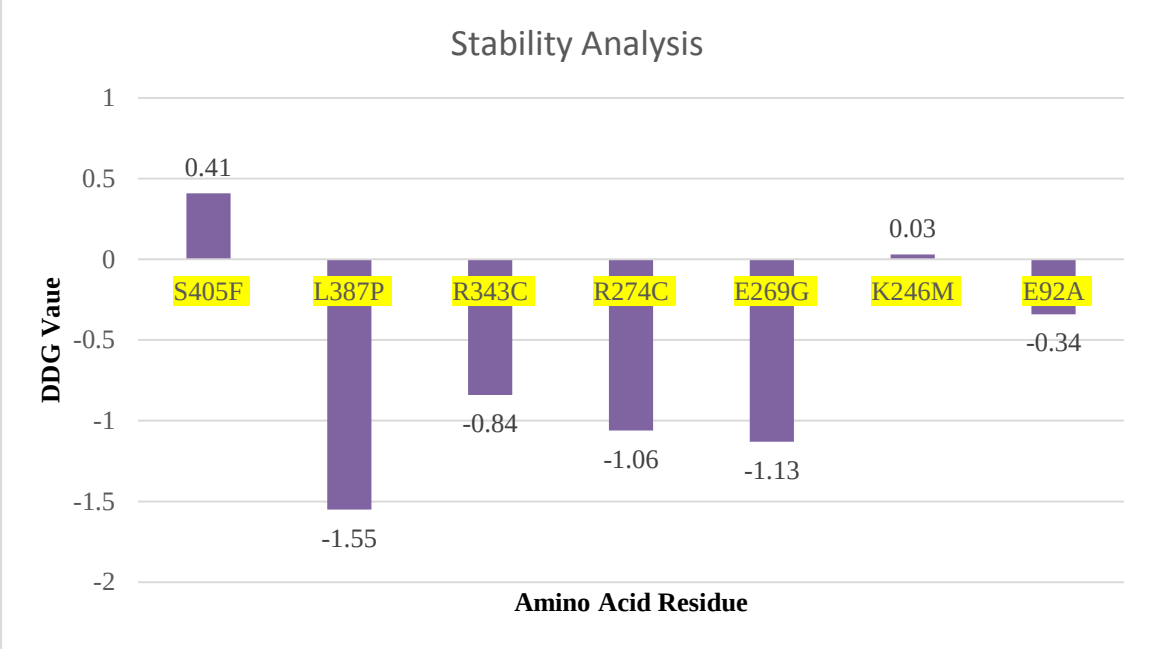

Figure 13: The 7 variants mapped with the DDG values showing their stability effects.

4.7. Association of PPP1R163 with cancer

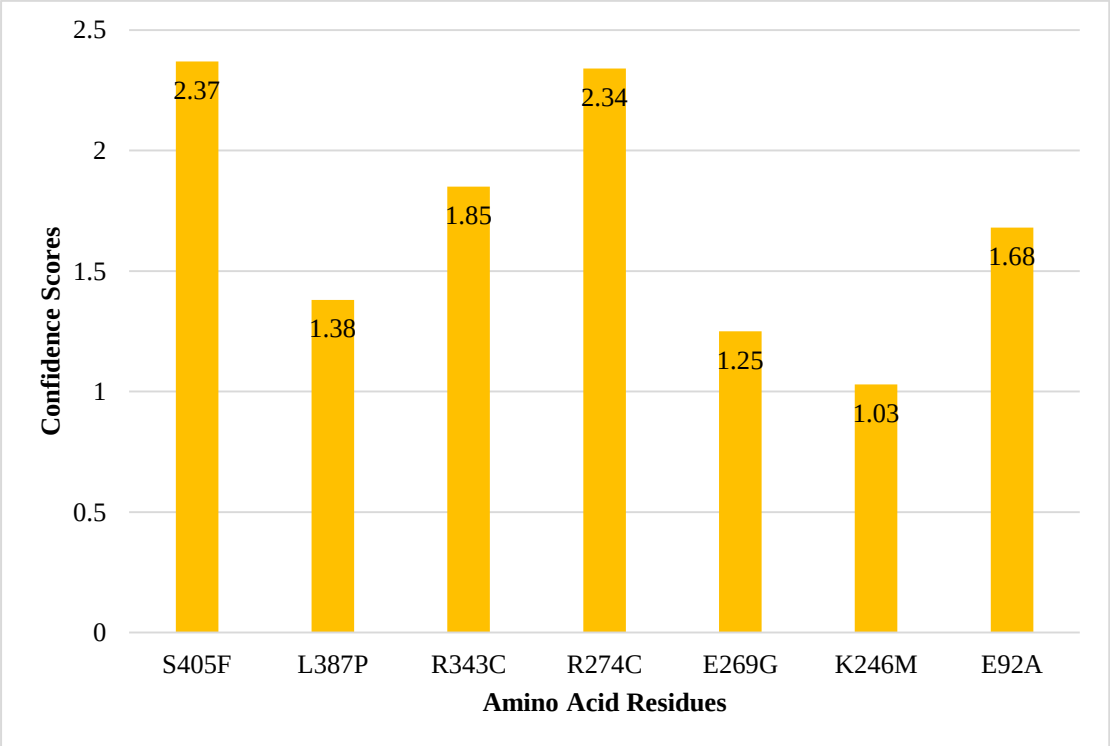

Figure 14: Confidence scores correlating the 7 SNPs association with cancer

According to FATHMM-XF, each prediction correlates to the confidence score assigned to each SNP that in turn relates directly to their respective association with

cancer. The highest score was assigned to S405F hence predicting its highest association to cancer. R343C, E92A, L387P and E269G had moderate association while K246M was seen to be the least likely to be associated with cancer given its lowest score (Figure 14). The deleterious SNPs were mapped across the different domains through InterPro (Figure 15). This was done to identify the position of SNPs on the domains of the gene.

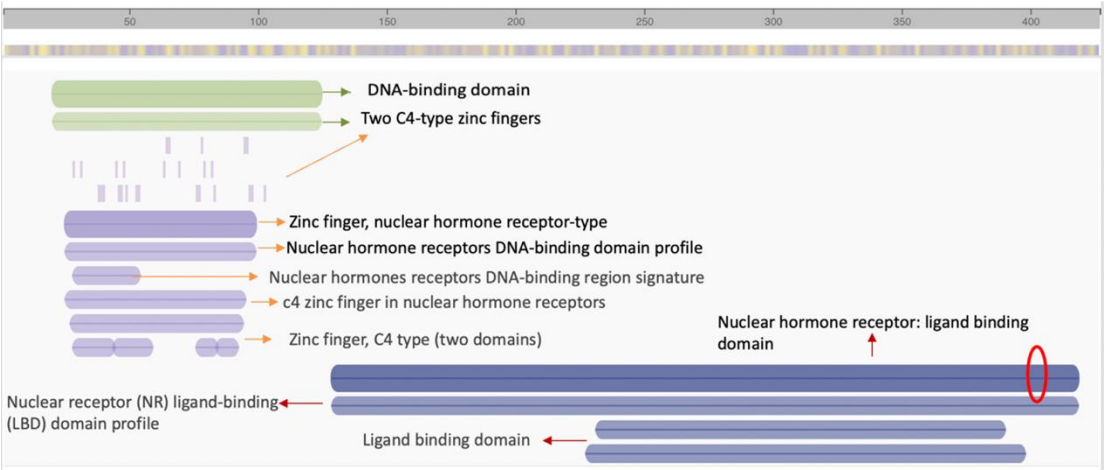

Figure 15: The labeled domains mapped on the PPP1R163 gene. All the SNPs except E92A are located in the ligand binding domain. S405F position is highlighted in red.

Chapter 5

Discussion

Brain tumor is one of the most dangerous and alarming tumors to exist in the world today, mainly identified after a display of symptoms such as headaches, dizziness, nausea, mood swings, and loss of communication (Bondy, Scheurer, & Malmer, 2008). Most of the tumors are recognized at a later stage leading to limited treatment options and increased fatalities. Identification of brain tumor is usually done through MRI and the detection is usually made when the tumor is at the final stage. While the tumor progresses, there are a variety of biochemical and molecular changes that occur, hence the tumor can be identified at an earlier stage via these biomarkers and genetic markers. These biomarkers can be classified into various distinct groups namely molecular biomarkers, circulating free DNA, circulating tumor cells, circulating extracellular vesicles, and microRNA's (Strimbu & Tavel, 2010). As most of the diseases, including brain tumor manifest at the level of protein activity and expression, the aim of this study was proteomic analysis and hence, the identification of genetic markers thereby indicating whether those specific proteins were being over or under expressed leading to the onset and metastasis of brain cancer (Oliveira & Oliveira, 2005).

Brain cancer, out of all the cancers, results in the majority of deaths per year, out of which a lot of them could be prevented by early detection and identification of tumor. Conventional techniques such as Magnetic Resonance Imaging (MRI) and Computed Tomography (CT scan) identify tumors at a later stage that limits the treatment options and hence, this late identification leads to more fatalities. On the contrary, methods and techniques such as proteome analysis can give a cheaper, quicker and easier method of early detection of brain tumors by identifying the under or over expression of biomarkers associated with a certain type of brain tumor (Mayeux, 2004) (Filser & Novak, 2007).

Proteomic analysis helps in the unbiased recovery of the protein constituents, fractionation of the protein mixture and generation of structural data by coupling with mass spectrometry and by matching and comparing this analytical data against

predicted or controlled results in a database which can help us identify over or under expressed proteins in a specific disease (Chalmers & Gaskell, 2000). Proteomic analysis can also help in the identification of post translational modification of proteins, protein-protein interactions and interpretation of signaling pathways. Proteomics coupled with mass spectrometry with high degree of specificity and sensitivity and the ability to search mass spectral data against ever expanding protein databases can result in quick, cheaper and more accurate analysis of changes in protein expression associated with a disease and the role of that specific protein in a cellular pathway (Williams & Addona, 2000).

Proteome analysis of the samples that underwent GC/MS analysis led to the initial identification of 852 proteins with their P values calculated from Microsoft T test. Proteins were first filtered by eliminating proteins with P value greater than 0.049 and fold change was then calculated for rest of the proteins. Additional filtering was done by eliminating proteins with fold change values less than 35. This led to the identification of 25 proteins, out of which ten were located in the ribosomes, five were part of the membrane, while four were present in the endoplasmic reticulum. On the other hand, rest of the proteins were present in other parts of the cell including the nucleus, mitochondria, and cytoskeleton. Through further analysis of these proteins, pathways were constructed using Cytoscape interface which were associated with cancer. Afterwards, they were collectively drawn on Biorender.

Brain tumors are characterized by increasing levels of oncogenic proteins and tumor suppressing proteins. Cancer susceptibility candidate 2 (CASC2) long non-coding RNAs (lncRNA) have shown negative modulation of miR-193a-5p which leads to decreased apoptosis and autophagy (Sung-Hyun Kim, 2021). Moreover, cellular prion proteins have exhibited shielding effect from apoptotic and oxidative stress (Larisa Ryskalin, 2021). Some of the identified proteins, SSRP1 and RAB5C, from the GC/MS analysis have exhibited anti-apoptotic effects. SSRP1 links to p-63 gene which results in anti-apoptotic effects upon expression. Meanwhile, AT2A2 and AT2A1, despite appearing in high levels, exhibit anti-apoptotic effects indicating their suppressive nature in tumors. AT2A2 appear to be increased, along with IDH1 mutations, in secondary glioblastoma and lower grade astrocytoma (Wei-Qing Li, 2017).

Metastasis is a significant level of cancer making the disease even more fatal and invasion is one of the building steps of metastasis. Elevated B7-H6 expression in gliomas revealed its involvement with increased invasion and metastasis (Fengyuan Chea, 2018). ZNFX1 antisense RNA 1 (ZFAS1) lncRNA has also shown role in advancement of metastasis in brain tumors (Sung-Hyun Kim, 2021).

β1,6-Nacetylglucosaminyl transferase (GnT-V) activity has a significant role in metastasis and invasion of tumor cells (Lucas Veillon, 2017). Upon investigation, SEPT 7 overexpression is associated with the activation of the MEK and ERK pathway leading to migration and invasion of the tumor. RPL3 and RPL4 indirectly activate the NF-kB pathway leading to metastasis and cell proliferation.

The beginnings of all cancers start with immense cell proliferation. Dysregulation of splicing activity or splicing factors in the nucleus leads to cell proliferation in glioblastoma multiforme. Spliceosome component SmB/B' (SNRPB) is elevated and leads to increased expression of genes involved in gliomagenesis (Pamela Bielli, 2019). Furthermore, mutated CDK lead to enhanced cell proliferation (Viktorija Juric, 2022). SMD3, one of the proteins from the results, is a part of the spliceosomal E complex where dysregulation could lead to the overexpression of inappropriate genes leading to proliferation. Most of the proteins showed a link to DDX family proteins which exhibited proliferation of cells.

This research has also clearly emphasized the important role of PPP1R163 in the development of brain cancer. SSRP1, identified during the GCMS analysis, activates this gene indirectly through p63 and hence causes inhibition of CDKs (Cyclin dependent kinases). Previous studies have also linked PPP1R163 with CDKs as well as their inhibitors in detailing their active role in the regulation of cell cycle progression as well as in signaling pathways that regulate apoptosis, cell growth and differentiation (Freedman, 1999). All processes which are of crucial significance in cancer growth. The results also support research detailed in previous literature which indicate the role of tumor protein 63 (Tp63) mediated induction of PPP1R163. P63 is the isoform of P53, which has been noted as one of the most significant genes in cancer development by regulating critical processes such as apoptosis, DNA repair, cellular senescence, angiogenesis inhibition and cell cycle arrest (Kommagani R, 2006) and has been specifically implicated in the upregulation of PPP1R163 and

hence implies that it plays a role in cancer emergence through regulation of PPP1R163 pathway (Ramakrishna Kommagani, 2007). Furthermore, calcitriol, a ligand for PPP1R163, has shown to have tumor suppressive and damaged proliferative abilities (Linjie Zhao, 2019).

Considering the importance of PPP1R163 to the results, its variants were retrieved, processed and studied using various bioinformatics tools. The 7 deleterious SNPs of PPP1R163 were finally narrowed down upon application of different tools that study amino acid substitution. The results clearly demonstrated the evolutionary importance of all the SNPs through Consurf particularly L387P, R343C, R274C, E269G, and K246M which highlights their important functional role in the body as key amino acids are usually strongly conserved which coincides with their characteristic importance in the body (Meytal Landau, 2005).

The significance of the SNPs - R343C, E269G, K246M, S405F, R274C and L387P was further illuminated upon mapping them across their respective domains where they were found to be located on the hormone binding site of PPP1R163, which is a site of functional and biological significance with regards to the metabolic activity of the receptor (J. Wesley Pike, 2014). While, L387P, R274C, K246M and E92A may have a characteristic role in impacting the function of PPP1R163. The results also displayed the destabilizing effects of the SNPs most notably variants L387P, R343C, R274C, E269G, and E92A which demonstrated high destabilizing effects on the protein. However, when correlating each SNP with cancer, S405F had the highest confidence score elucidating that it had the closest association with cancer.

There were a few limitations to this study. The results were not specific to the type of brain cancer the patients had and were randomly chosen for the study. Some proteins in the control samples in the GC-MS results were too less to be detected by the instrument, hence showing 0 for that particular protein, that had to be converted to 0.1 for calculating fold/change.

Conclusion

Pakistan lacks a Cancer database that gathers data on the number of cases and deaths related to each type per year. Such limitations have made studying the different types of cancer and gauging their overall effect on the population rather difficult. This study helps elucidate the different biomarkers that would help identify a very under-studied and overlooked type of cancer in Pakistan i.e. Brain cancer. Moreover, it also helps highlight the PPP1R163 gene which has been described in various other studies as being linked to brain tumors and hence, can be considered as a potential therapeutic target in the future.

There is potential for further research through collection and analysis of serum samples as well as samples from tumor resection surgeries from other populations to identify and evaluate the different prognostic markers in brain cancer and link them to its various types. The proteins identified in this study can be further studied through Western blotting and Immunoblotting to gauge their link to brain tumors.

Additionally, further In silico studies can also be conducted on each identified protein particularly PPP1R163 to further validate their role in brain cancer as well as study their characteristic role in further detail.

References

Adi Ben Chorin, G. M. (2020). ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins. *Protein Science, 29*(1), 258-267.

Adip G. Bhargav, S. K.-H. (2020). Nanomedicine Revisited: Next Generation Therapies for Brain Cancer. *Advanced Therapeutics, 3*(10), 1-36.

Ae Kyung Park, P. K. (2019). Subtype-specific signaling pathways and genomic aberrations associated with prognosis of glioblastoma. *Neuro-Oncology, 21*(1), 59-70.

Aldape, K. B. (2019). Challenges to curing primary brain tumours. *Nature reviews*, 509-520.

Anderson ME, D. S. (2020). *Abeloff's Clinical Oncology.* Philadelphia: Elsevier.

Andrew D Yates, P. A. (2020). Ensembl 2020. *Nucleic Acids Research, 48*(D1), 682-688.

Asad ullah Khan, I. M. (2019). Modulation of brain tumor risk by genetic SNPs in PARP1gene: Hospital based case control study. *PLoS One, 14*(10).

Beale, D. P. (2018). Review of recent developments in GC–MS approaches to metabolomics-based research. *Metabolomics, 14*(152).

Benjamin D. Fox, V. J. (2011). Epidemiology of Metastatic Brain Tumors. *Neurosurgery Clinics of North America*.

Board, C. E. (2022, February). *Brain Tumor: Statistics*. Retrieved from Cancer.net: https://www.cancer.net/cancer-types/brain-tumor/statistics#:~:text=Brain%20tumors%20account%20for%2085,year%20in%20the%20United%20States.

Bondy, M. L., Scheurer, M. E., & Malmer, B. (2008). Brain tumor epidemiology: Consensus from the Brain Tumor Epidemiology Consortium. *Cancer*, 1953-1968.

Boris Reva, Y. A. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research, 39*(17), e118.

Campana, D., & Behm, F. G. (2000). Immunophenotyping of Leukemia. *Journal of Immunological methods*, 59-75.

Cancer.Net. (2022). *Brain Tumor: Statistics.* ASCO.org.

Chalmers, M. J., & Gaskell, S. J. (2000). Advances in mass spectrometry for proteome analysis. *ScienceDirect*, 384-390.

Chaurasia, V., Saurabh, P., & Tiwari, B. (2018). Prediction of benign and malignant breast cancer using data mining techniques. *Sage*.

Chen, K., Xiao, H., & Zeng, J. (2017). Alternative Splicing of EZH2 pre-mRNA by SF3B3 Contributes to the Tumorigenic Potential of Renal Cancer. *Clinical Cancer Research*, 3428-3441.

Chetan Bettegowda, M. P. (2022). *Meningioma Recurrence*. Retrieved 4 2, 2022, from Hopkins Medicine: https://www.hopkinsmedicine.org/health/conditions-and-diseases/meningioma-recurrence

Chung, Y.-J. J.-S. (2018). Phagocytic Roles of Glial Cells in Healthy and Diseased Brains. *Biomolecules & Therapeutics*, 350–357.

Chunzhi Qin, L. R. (2017). CDKL1 promotes tumor proliferation and invasion in colorectal cancer. *OncoTargets and Therapy*, 1613–1624.

Datta A, K. H.-M. (2017). Manumycin A suppresses exosome biogenesis and secretion via targeted inhibition of Ras/Raf/ERK1/2 signaling and hnRNP H1 in castration-resistant prostate cancer cells. *Cancer Lett, 408*, 73-81.

Davis, M. E. (2018). EPIDEMIOLOGY AND OVERVIEW OF GLIOMAS. *Seminars in Oncology Nursing,*, 1-10.

Denise Casemore, C. X. (2015). SERCA as a target for cancer therapies. *Integr Cancer Sci Therap*.

Emidio Capriotti, P. F. (2008). A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics, 9*(2).

Emwas, A.-H. &.-T. (2015). Gas Chromatography–Mass Spectrometry of Biofluids and Extracts. Methods in molecular biology. *Clifton, N.J., 1277*, 91-112.

Escobar, B., Cárcer, G. d., Fernández-Miranda, G., Cascón, A., Bravo-Cordero, J. J., Montoya, M. C., et al. (2010). Brick1 Is an Essential Regulator of Actin Cytoskeleton Required for Embryonic Development and Cell Transformation. *Cancer Research*, 9349–9359.

Fabelo H, O. S. (2018). Spatio-spectral classification of hyperspectral images for brain cancer detection during surgical operations. *PLOS ONE, 13*(3).

Farheen Danish, H. S. (2021). Comparative clinical and epidemiological study of central nervous system tumors in Pakistan and global database. *Interdisciplinary Neurosurgery, 25*.

Feng, D. Y. (2019). Metadherin Promotes Malignant Phenotypes And Induces Beta-Catenin Nuclear Translocation And Epithelial–Mesenchymal Transition In Gastric Cancer. *Cancer managaement and research, 11*, 8911–8921.

Fengyuan Chea, b. ,. (2018). B7-H6 expression is induced by lipopolysaccharide and facilitates cancer invasion and metastasis in human gliomas. *International Immunopharmacology*, 318-327.

Filser, D., & Novak, J. (2007). Advances in Urinary Proteome Analysis and Biomarker Discovery. *Journal of the American Society of Nephrology*, 1057-1071.

Freedman, L. P. (1999). Transcriptional Targets of the Vitamin D3 Receptor–Mediating Cell Cycle Arrest and Differentiation. *Journal of Nutrition, 129*(2), 581-586.

Gaia Favero, F. B. (2021). Pineal Gland Tumors: A Review. *Cancers*.

Garcia, C., & Poletti, E. (2009). Bosoquamous carcinoma. *Journal of the American Academy of Dermatology*, 137-143.

Gopal S Tandel, M. B. (2019). A Review on a Deep Learning Perspective in Brain Cancer Classification. *Cancers, 11*(1), 1-32.

Greg James, K. A. (2018). 12 - Craniopharyngiomas. In M. F. Richard G. Ellenbogen, *Principles of Neurological Surgery (Fourth Edition)* (pp. 204-218). Elsevier.

Hanka Venselaar, T. A. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics, 11*.

Hennequin C, G. S. (2019). Combination of chemotherapy and radiotherapy: A thirty years evolution. *Cancer radiother., 23*(6-7), 662-665.

Huaiyu Mi, D. E. (2021). PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research, 49*(D1), 394-403.

Huse, L. Y. (2021). Pathology and Molecular Genetics . In J. C. Joseph Jankovic MD, *Bradley and Daroff's Neurology in Clinical Practice.* Elsevier.

Hyuna Sung PhD, J. F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A cancer journal for Clinicians*.

Idle, J. R., & Gonzalez, F. J. (2007). Metabolomics. *Cell Metabolism*, 348-351.

Ioannidis NM, R. J.-A.-T. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American journal of human genetics, 99*(4), 877-885.

Ivan A Adzhubei, S. S. (2010). A method and server for predicting damaging missense mutations. *Nature Methods, 7*(4), 248-249.

J. Wesley Pike, M. B. (2014). Fundamentals of vitamin D hormone-regulated gene expression. *The Journal of Steroid Biochemistry and Molecular Biology, 144*, 5-11.

Jahangir Moini, N. G. (2021). Chapter 10 - Schwannoma. In M. S. Jahangir Moini, *Epidemiology of Brain and Spinal Tumors* (pp. 179-196). Elsevier.

Jemal, A., Bray, F., & Center, M. M. (2011). Global Cancer Statistics. *A cancer journal for clinicians*.

Jigisha P Thakkar, M. P. (n.d.). *Glioblastoma Multiforme*. Retrieved from American Association of Neurological surgeons: https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Glioblastoma-Multiforme#:~:text=Glioblastoma%20is%20the%20most%20common,men%20as%20compared%20to%20women.

Jordan K, S. E. (2021). Potential involvement of extracellular citrate in brain tumor progression. *Current molecular medicine*.

Kapałczyńska, M., Kolenda, T., Przybyła, W., & Zajączkowska, M. (2018). 2D and 3D cell cultures - a comparison of different types of cancer cell cultures. *Archives of Medical Science*, 910-919.

Kommagani R, C. T. (2006). Identification of vitamin D receptor as a target of p63. *Oncogene*, 3745-3751.

Kong, X., Zhang, J., Li, J., Shao, J., & Fang, L. (2018). MiR-130a-3p inhibits migration and invasion by regulating RAB5B in human breast cancer stem cell-like cells. *Biophysical and Biochemical Resesearch Communications*, 486-493.

Konopleva, M. M. (2020). MDM2 inhibition: an important step forward in cancer therapy. *Leukemia, 34*, 2858–2874.

Korać, P. A. (2021). MiR-7 in Cancer Development. *Biomedicines, 9*(3).

Larisa Ryskalin, F. B. (2021). The Role of Cellular Prion Protein in Promoting Stemness and Differentiation in Cancer. *cancers*, 170.

Linjie Zhao, S. Z.-Å. (2019). Nuclear Receptors: Recent Drug Discovery for Cancer Therapies. *Endocrine Reviews*, 1207–1249.

Liu M, W. H. (2018). NF-κB signaling pathway-enhanced complement activation mediates renal injury in trichloroethylene-sensitized mice. *Journal of immunotoxicology, 15*(1), 63-72.

Liu, X., & Locasale, W. J. (2017). Metabolomics: A Primer. *Trends in Biochemical Sciences*, 274-284.

Lucas Veillon, C. F.-E.-H. (2017). Glycosylation Changes in Brain Cancer. *ACS Chemical Neuroscience*, 51−72.

Lv, Y. Y. (2019). Strand-specific miR-28-3p and miR-28-5p have differential effects on nasopharyngeal cancer cells proliferation, apoptosis, migration and invasion. *Cancer Cell Int, 19*.

Maldonado, K. A., & Alsayouri., K. (2021). *Physiology, Brain.* Treasure Island (FL): StatPearls Publishing.

Marco Prinz, S. J. (2019). Microglia Biology: One Century of Evolving Concepts. *Cell*, 292-311.

Mark F Rogers, H. A. (2018). FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics, 34*(3), 511-513.

Martín-Alonso, P. D.-L.-D.-B. (2020). A Practical Overview on the Molecular Biology of Meningioma. *Current Neurology and Neuroscience Reports*, 62-.

Matthias Blum, H.-Y. C.-L. (2020). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*.

Mayeux, R. (2004). Biomarkers: Potential uses and limitations. *NeuroX*, 182-188.

Meytal Landau, I. M.-T. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Research, 33*(2), 299-302.

Milica M. Badža, M. Č. (2021). Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network. *Applied Sciences, 10*(6), 1999.

Miller, K. D., Ostrom, Q. T., Kruchko, C., & Patil, N. (2021). Brain and other central nervous system tumor statistics, 2021. *ACS Journals*, 381-406.

Mohamed Alshabi, B. V. (2019). Identification of Crucial Candidate Genes and Pathways in Glioblastoma Multiform by Bioinformatics Analysis. *biomolecules, 9*(5), 201.

Mugnani, E. N. (2016). Lymphoma. *Histologic diseases*, 527-702.

Muñoz, A., & Grant, W. (2022). Vitamin D and Cancer: An Historical Overview of the Epidemiology and Mechanisms. *Nutrients, 14*, 1448.

Nauclér, C. S., Geisler, J., & Vetvik, K. (2019). The emerging role of human cytomegalovirus infection in human carcinogenesis: a review of current evidence and potential therapeutic implications. *Oncotarget*, 4333-4347.

Nauman Idris Butt, F. K. (2020). Incidence of Brain Tumours in Children and Adults in Pakistan. *Med. Forum, 31*(9), 111-114.

Nayak L, L. E. (2012). Epidemiology of brain metastases. *Curr Oncol Rep, 14*(1), 48-54.

Oliveira, S. S., & Oliveira, I. M. (2005). Claudins upregulation in human colorectal cancer. *FEBS Letters*, 6179-6185.

Ostrom QT, P. N.-S. (2020). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2013-2017. *Neuro Oncology*, 1-96.

Pamela Bielli, V. P. (2019). Splicing Dysregulation as Oncogenic Driver and Passenger Factor in Brain Tumors. *Cells*, 10.

Paulina Zofia Gorynska, K. C. (2022). Metabolomic Phenotyping of Gliomas: What Can We Get with Simplified Protocol for Intact Tissue Analysis? *Cancers*, 312.

Poojary, M. M. (2016). Improved conventional and microwave-assisted silylation protocols for simultaneous gas chromatographic determination of tocopherols and sterols: Method development and multi-response optimization. *Journal of Chromatography A, 1476*, 88-104.

Poulogiannis, N. K. (2019). Reprogramming of fatty acid metabolism in cancer. *british journal of cancer*, 4-22.

Prateek Kumar, S. H. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols, 4*(7), 1073-1081.

Qian Bu, Y. H.-L. (2012). Metabolomics: a revolution for novel cancer marker identification. *Comb Chem High Throughput Screen, 15*(3), 266-275.

Ramakrishna Kommagani, V. P. (2007). Differential regulation of vitamin D receptor (VDR) by the p53 Family: p73-dependent induction of VDR upon DNA damage. *J Biol Chem*, 29847-29854.

Renaud Vatrinet, G. L. (2017). The α-ketoglutarate dehydrogenase complex in cancer metabolic plasticity. *Cancer and Metabolism*, 1-14.

Rentzsch P, W. D. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, D886-D894.

Renu Pandey, L. C. (2017). Metabolomic signature of brain cancer. *Mol Carcinog, 56*(11), 2355-2371.

Salehiniya, H., & Farmanfarma, K. K. (2020). BRAIN CANCER IN THE WORLD: AN EPIDEMIOLOGICAL REVIEW. *Researchgate*, 1-6.

Sarah Kuhn, L. G. (2019). Oligodendrocytes in Development, Myelin Generation and Beyond. *Cells*.

Sarah Lapointe, A. P. (2018). Primary brain tumours in adults. *The Lancet*.

Shamsi, U. (2020). Cancer Prevention and Control in Pakistan: Review of Cancer Epidemiology and Challenges. *Liaquat National Journal of Primary Care*, 34-38.

Shi, Y. H. (2019). miR-632 promotes gastric cancer progression by accelerating angiogenesis in a TFF1-dependent manner. *BMC Cancer, 19*(14).

Shimazaki R, T. S. (2021). Complement factor B regulates cellular senescence and is associated with poor prognosis in pancreatic cancer. *Complement factor B regulates cellular senescence and is associated with poor prognosis in pancreatic cancer*, 937-950.

Shrestha, J. (2019). P-Value: a true test of significance in agricultural research.

Sicong Jiang, D. L. (2020). High expression of IGBP1 correlates with poor prognosis in esophageal squamous cell carcinoma. *The International Journal of Biological Markers*, 33–40.

Strimbu, K., & Tavel, J. A. (2010). What are Biomarkers? *Curr Open HIV AIDS*, 463-466.

Sun H, Y. G. (2019). New insights into the pathogenicity of non-synonymous variants through multi-level analysis. *Sci Rep, 9*(1), 1667.

Sung-Hyun Kim, K.-H. L.-Y. (2021). Long non-coding RNAs in brain tumors: roles and potential as therapeutic targets. *Journal of Hematology & Oncology* , 77.

T.C, L. (2019). DDX3X Multifunctionally Modulates Tumor Progression and Serves as a Prognostic Indicator to Predict Cancer Outcomes. *International journal of molecular sciences, 21*(1), 281.

Viktorija Juric, B. M. (2022). Cyclin-dependent kinase inhibitors in brain cancer: current state and future directions. *Cancer Drug Resist*, 48-62.

Wang Y, W. Y. (2015). MiR-302c-3p suppresses invasion and proliferation of glioma cells via down-regulating metadherin (MTDH) expression. *Cancer biology and therapy, 16*(9), 1308-1315.

Wang YH, Y. Y. (2018). miR-639 is associated with advanced cancer stages and promotes proliferation and migration of nasopharyngeal carcinoma. *Oncol Lett, 16*(6), 6903-6909.

Weinberg, R. A. (1996). How Cancer Arises. *Scientific American*, 62-70.

Weinberg, R. A. (1996). How Cancer Arises. *Scientific American*, 62-70.

Wei-Qing Li, N.-Z. Z.-M.-J.-M.-Y.-Z.-C. (2017). High ATP2A2 expression correlates with better prognosis of diffuse astrocytic tumor patients. *ONCOLOGY REPORTS*, 2865-2874.

Wen Luo, L. W.-H.-Z.-T.-X. (2017). hsa-mir-3199-2 and hsa-mir-1293 as Novel Prognostic Biomarkers of Papillary Renal Cell Carcinoma by COX Ratio Risk Regression Model Screening. *Journal of Cellular Biochemistry*, 3488-3494.

WHO. (2020). *Pakistan fact sheet.* Globocan.

WHO. (2021). *Pakistan.* Global Cancer Observatory.

Williams, C., & Addona, T. A. (2000). The integration of SPR biosensors with mass spectrometry: possible applications for proteome analysis. *ScienceDirect*, 45-48.

Wojciech Jelski, B. M. (2021). Molecular and Circulating Biomarkers of Brain Tumors. *Int J Mol Sci, 22*(13), 7039.

Yang, L. G. (2018). The tumor suppressive miR-302c-3p inhibits migration and invasion of hepatocellular carcinoma cells by targeting TRAF4. *Journal of Cancer*, 2693-2701.

Ying Zhang, C. D. (2018). The p53 Pathway in Glioblastoma. *Cancers, 10*(9), 1-18.

Yu, J. O. (2010). MicroRNA, hsa-miR-200c, is an independent prognostic factor in pancreatic cancer and its upregulation inhibits pancreatic cancer invasion but increases cell proliferation. *Molecular cancer, 9*.

Yufei Xiao, 1. T.-H. (2014). A novel significance score for gene selection and ranking. *Bioinformatics, 30*(6), 801-807.

Zahra Ghaemi, B. M. (2019). MicroRNA-326 Functions as a Tumor Suppressor in Breast Cancer by Targeting ErbB/PI3K Signaling Pathway. *Frontiers in Oncology, 9*.

Zhenlin Liu, Z. J. (2014). miR-7 inhibits glioblastoma growth by simultaneously interfering with the PI3K/ATK and Raf/MEK/ERK pathways. *International Journal of Oncology*, 1571-1580.