Integrative Proteomics Profiling for Understanding the Role of Genetic Markers in Liver Cancer

By:

Aleena Abid (00000279434)

Moel Abaidullah (00000279517)

Wardah Shakil (00000280524)

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

National University of Sciences and Technology (NUST)

2022

Integrative Proteomics Profiling for Understanding the Role of Genetic Markers in Liver Cancer

By:

Aleena Abid (00000279434)

Moel Abaidullah (00000279517)

Wardah Shakil (00000280524)

Supervised by: Dr. Yasmeen Badshah

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

National University of Sciences and Technology (NUST)

2022

Integrative Proteomics Profiling for Understanding the Role of Genetic Markers in Liver Cancer

By:
Aleena Abid (00000279434)
Moel Abaidullah (00000279517)
Wardah Shakil (00000280524)

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. Yasmeen Badshah

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 Ms Aleena Abid, Moel Abaidullah and Wardah Shakil (Registration No. 00000279517, aleena's reg no. and 00000280524), of '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

We, Aleena Abid, Moel Abaidullah, Wardah Shakil hereby state that our BS FYP thesis titled "Integrative Proteomics Profiling for Understanding the Role of Genetic Markers in Liver Cancer" is our own work and has not been submitted previously by us for taking any degree from this university 'National University of Sciences and Technology' or anywhere else in the country/world.

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

Name of student:

Aleena Abid:

_____Aleena Abid___

Moel Abaidullah: __________________

Wardah Shakil: __Wardah___________

Date:

Certificate for Plagiarism

It is certified that BS FYP Thesis Titled "Integrative Proteomics Profiling for Understanding the Role of Genetic Markers in Liver Cancer" by Aleena Abid, Moel Abaidullah, Wardah Shakil 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

Aleena Abid: Aleena Abid

Moel Abaidullah: __________________

Wardah Shakil: Wardah____________

Date:

Signature & Stamp of Supervisor

This thesis is dedicated to our beloved parents, friends, and family who have been a constant source of love, support, encouragement and guidance.

ACKNOWLEDGEMENTS

Firstly, we would like to thank our supervisors, Dr Yasmin Badshah and Dr Maria Shabbir for their valuable support and guidance throughout our final year project and their involvement in our academic journey over the course of our program.

We are also immensely grateful to Dr Hussnain A. Janjua and the department ASAB for providing us the opportunities and facilities which allowed us to execute our final year project.

Ms. Khushbakhat Khan helped us immensely in making progress in the lab and beyond through her expertise, knowledge, and patience. We are also thankful to MS students Amna Hafeez, Sameen and Kanza for their supervision in our lab work. We want to appreciate the staff of the Cancer Biology lab.

Finally, we want to express our gratitude to our parents, friends and family members for their unwavering support and encouragement through our lives and the project.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS 1

LIST OF ABBREVIATIONS 6

ABSTRACT 10

INTRODUCTION 11

1.1 Liver Cancer 11

1.2 Viral Liver Cancer 12

1.3 Hepatitis C Led Hepatocellular Carcinoma 12

1.4 Potential Protein Biomarkers 13

1.5 GC-MS (Gas chromatography/Mass spectrometry) 13

1.6 Protein Identification through GCMS 14

1.7 Insilico analysis 15

CHAPTER 2 16

LITERATURE REVIEW 16

2.1 Liver Cancer 16

2.2 Global Trends in Incidence 16

2.3 Prevalence of Liver Cancer in Pakistan 17

2.4 Prevalence and Distribution of Hepatitis C 17

2.5 Viral Liver Cancer 18

2.5.1 Global Prevalence 18

2.5.2 Hepatitis C Led Viral Liver Cancer 18

2.5.3 Mechanism of HCV Led Viral Liver Cancer 19

2.6 Liver Cancer Diagnosis 20

2.7 Screening Tests 20

2.7.1 Limitations of Current Diagnostics 21

2.8 Cancer Biomarkers 22

2.9 Protein Biomarkers 22
2.10 Blood Biomarkers 23
2.11 Gas chromatography Mass spectrometry (GCMS) 24
2.11.1 GCMS of biofluids and extracts 25
2.11.2 Components of GCMS 25
2.12 Data Analysis tools for GC-MS 26
2.12.1 Cytoscape 26
2.12.2 Panther 26
2.12.3 Biorender 26
2.13 Aims and Objectives 27
CHAPTER 3 28
METHODOLOGY 28
3.1 Samples Collection, Handling and Transportation 28
3.2 Serum Extraction 28
3.3 Samples Preparation 29
3.3.1 Chemicals required 29
3.4 Steps involved in the preparation of serum samples 29
3.4.1 Samples lysis 29
3.4.2 Centrifugation 29
3.4.3 Sample drying 29
3.4.4 Derivatization 30
3.4.5 Silylation 30
3.4.6 Incubation 30
3.5 GC/MS 30
3.6 GCMS analysis 31
3.7 GC-MS data Analysis 32

3.7.1 Cytoscape 32

3.7.2 PANTHER Tool 33

3.7.3 Protein Atlas 33

CHAPTER 4 36

RESULTS 36

4.1 GC-MS Data 36

4.2 Cellular localization of proteins 36

4.3 Identification of protein role in liver cancer 39

- ADP/ATP translocase 2; SLC25A5; ortholog. 39
- Creatine kinase B-type; CKB;ortholog. 39
- Malate Dehydrogenase (MDH), Isocitrate Dehydrogenase (IDH) and Serine Hydroxymethyltransferase 2 (SHMT 2) 40
- Fatty acid synthase; FASN; ortholog 41
- Sodium/Potassium-Transporting ATPase Subunit Alpha 1 (ATP1A1) 42
- Clathrin Heavy Chain (CLTC) 42
- Adenine Nucleotide Translocase (ANT) 3. 42
- Peroxisomal Multifunctional Enzyme (DHB4) 43
- Prelamin-(A/C) 43
- Heterogeneous Nuclear Ribonucleoprotein A1 (HNRNPA1) 43
- 60S acidic ribosomal protein P0; RPLP0; ortholog. 45

4.4 Evaluation of Filtered Proteins as Diagnostic and Prognostic Markers 46

CHAPTER 5 53

DISCUSSION 53

Future Prospects 58

REFERENCES 59

APPENDIX 68

TABLE OF FIGURES

Figure 1 Model of HCC caused by HCV from (Mitchell, Lemon, & McGivern, 2015). 18
Figure 2 The cytoscape interface. 31
Figure 3 Different classes of obtained proteins. 34
Figure 4 Molecular functions of the proteins 35
Figure 5 CLTC interactions obtained from cytoscape. CLTC controls transport of Caspase 8 (CASP 8) and Fas Associated Death Domain (FADD). CLTC controls state change of Fas ligand (FASLG) and Fas. 37
Figure 6 Pathways of proteins locaslised in the mitochondrion. 39
Figure 7 proteins linked to the P13k/Akt and JAK/STAT pathway 40
Figure 8 Kaplan-Meier graphs 46
Figure 9 Type and number of non-coding variants 48
Figure 10 Number of variants with each rank 49
Figure 11 Changes in transcription sites introduced by SNPs. 50

LIST OF ABBREVIATIONS

%	Percentage
(m/z)	Mass number / charge number
°C	Degree Celcius
β	Beta
μL	Microlitre
AAV	Adeno Associated Virus
ADP	Adenosine Diphosphate
AFP	Alpha Fetoprotein
AFU	α-L-fucosidase
Akt	RAC(Rho family)-alpha serine/threonine-protein kinase
ALDH3A2	Aldehyde Dehydrogenase Family 3 Member A2
ANT	Adenine Nucleotide Translocase
ASIR	age standardised incidence rate
ASR	Age Standardised Rate
ATP	Adenosine Triphosphate
ATP1A1	Sodium/Potassium-Transporting ATPase Subunit Alpha 1
CD4+	Cluster of Differentiation 4 positive

CD8+	Cluster of Differentiation 8 positive
CDK	Cyclin Dependent Kinase
CKB	Creatine kinase B-type
CLTC	Clathrin Heavy Chain
CT	Computerised tomography
DCP	Des-gamma carboxy-prothrombin
DHB4	Peroxisomal Multifunctional Enzyme
DKK-1	Dickkopf-1
DNA	Deoxyribonucleic Acid
EGFR	Epidermal Growth Factor Receptor
ERK	Extracellular Signal regulated Kinase
FASN	Fatty acid synthase
GCMS	Gas Chromatography Mass Spectrometry
GPC3	glypican-3
HCB	Hepatitis B
HCC	Hepatocellular Carcinoma
HCL	Hydrochloric acid
HCV	Hepatitis C
HIV	Human Immunodeficiency Virus
hnRNPK	Heterogeneous nuclear ribonucleoprotein K
ICC	intrahepatic cholangiocarcinoma

IFN-γ	Interferon Gamma
IL-4	Interleukin-4
IRF-3/7	Interferon Regulatory Factor 3 and Interferon Regulatory Factor 7
JAK	Janus Kinase
Kc	distribution coefficient
KDa	Kilo Dalton
MAPK	Mitogen Activated Protein Kinase
MDK	heparin binding growth factor Midkine
MEK	Mitogen activated protein kinase
miR	Micro RNA
ml	millilitre
ml/min	Millilitre per minute
MRI	Magnetic Resonance Imgaing
MSTFA	N-methyl-N-(trimethylsilyl)trifluoroacetamide
NF-κB	Nuclear Factor Kappa-B
OPN	osteopontin
PI3K	phosphatidylinositol 3-kinase
PIVKA II	protein induced by vitamin K absence or antagonist-II
PLAAT4	Phospholipase A and acyl transferase 4

PON1	serum paraoxonase 1
RNA	Ribonucleic Acid
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
RPL12	Ribosomal Protein L-12
RPLP0	Ribosomal Protein Lateral Stalk Subunit P0
rpm	rotations per minute
SCCA	squamous cell carcinoma antigen
SNP	Single Nucleotide Polymorphism
Src	Proto-oncogene tyrosine-protein kinase
STAT	Signal transducer and activator of transcription
TGF-β	Transforming Growth Factor β
TMCS	1% trimethylchlorosilane
WHO	World Health Organisation

ABSTRACT

Liver cancer is a broad term encompassing varying types. The focus of the current research will be on Hepatocellular Carcinoma (HCC) primarily in the context of viral causes. HCC is the sixth most common cancer worldwide, attributing to 626 000 or 5.7% of new cancer cases annually. HCCs have a high incidence to mortality ratio because the tumours are resistant to chemotherapy. The low effectiveness of treatment may also be due to the fact that less than 40 percent of HCC cases are diagnosed at an early stage. The study uses GCMS analysis and multiple insilico tools for the identification of potential protein biomarkers in the blood samples of patients and healthy controls. After identification and filtration of potential protein biomarkers, their insilico analysis was performed. Computational analysis was performed for a target gene for liver cancer PPP1R163. About 800 proteins were identified in the blood samples through GC-MS. The proteins with a fold change greater than 50 were 16 which were found to have a role in cancer progression. Majority were playing their role in metastasis and there were few involved in anti-apoptosis. The computational analysis of PPP1R163 showed that 15 Unique transcription sites were added out of which the transcription factor WT1-I acted as an activator to the complimentary transcription factor binding site in the PPP1R163 gene. The proteomics approach provides richer information about the cancer than existing approaches. A combination of proteins and gene biomarkers provides more accurate diagnosis of cancer. These identified biomarkers will help in the early detection of tumour and the designing of effective therapeutics for patients.

CHAPTER 1

INTRODUCTION

1.1 Liver Cancer

Liver Cancer is a broad term encompassing varying types and sub types of liver cancer. The focus of the current research will be on Hepatocellular Carcinoma (HCC) primarily in the context of viral causes. HCC is the sixth most common cancer worldwide, attributing to 626 000 or 5.7% of new cancer cases annually. It is the third most common cause of cancer related deaths globally and carries an overall survival rate of only 3-5%. The major burden of HCC lies in developing countries; up to 82% of HCC cases are reported from developing countries. Liver cancer is the fifth most commonly occurring cancer in men and the ninth most commonly occurring cancer in women. In 2018 alone, there were over 840,000 new cases (Amna Subhan Butt, 2012).

Thus, it is especially important to cater to liver cancer diagnosis and effective treatment options in Pakistan. Hepatocarcinogenesis involves complex genetic and cellular dysregulations which drive the formation of HCC, with extensive heterogeneity. The lack of representative molecular and cellular signatures in the heterogeneous HCC tumours that can effectively guide the choice of the most appropriate treatment among the patients unavoidably limits the treatment outcome.

Advancement and wide availability of the next-generation sequencing technologies have empowered us to examine and capture not only the detailed genetic alterations of the HCC cells but also the precise composition of different cell types within the tumour microenvironment and their interactions with the HCC cells at an unprecedented level (Lo-KongChan, 2021). Henceforth, in the current research project, we will be employing the Gas chromatography Mass Spectrometry (GCMS) to elucidate proteome analysis of the genetic biomarkers in viral liver cancer through integrative profiling. (Ranjbar1, 2015)

1.2 Viral Liver Cancer

Hepatitis B virus (HBV), Hepatitis C virus (HCV), and Adeno Associated Virus (AAV) 2 can cause inflammation in the liver which can ultimately lead to Hepatocellular Carcinoma (HCC) formation (Tu, Bühler, & Bartenschlager, 2017). HCV and HCB are the leading causes of primary liver cancer. They may even directly drive cancer. HCV causes liver cirrhosis which leads to HCC while HBV can cause liver cancer even in patients without end stage liver disease (Ringehan, McKeating, & Protzer, 2017). HBV and HCV infect in a non-cytopathic manner but the activity of CD 4+ and CD 8+ T cells, Natural Killer cells can lead to oncogenesis (de Mattos et al., 2021). AAV 2 may be involved in liver cancer as well. It integrates its genome at multiple sites in human DNA so it can trigger overexpression of oncogenes leading to cancer (La Bella et al., 2020).

1.3 Hepatitis C Led Hepatocellular Carcinoma

World over, 58 million people are estimated to have chronic Hepatitis C. 1.5 million new infections are estimated to occur every year. Most deaths from Hepatitis C occur due to cirrhosis and HCC. Most people with the infection remain undiagnosed till later stages of liver disease. There is limited access to diagnosis and treatment options (World Health Organisation, 2021b).

Persons with HCV associated liver cirrhosis are 17 times more likely to develop HCC than non-infected control group (Viral hepatitis and liver cancer, 2017). Around one fifth of chronically infected HCV patients develop liver cirrhosis. Out of these, 1-4% go on to develop HCC every year. Coinfections with HBV and HIV substantially increase the risk of developing HCC in Hepatitis C patients (Axley, Ahmed, Ravi, & Singal, 2018).

Although data regarding Pakistani population is insufficient, it was found that 58% of HCC patients in Pakistan were infected with HCV. This points to significance of HCV infections potentially causing HCC. A shift was seen from HBV infections

being more commonly present in HCC patients to HCV infections being the most common from mid 1990s onwards (Subhan Butt, Abbas, & Jafri, 2012).

1.4 Potential Protein Biomarkers

Protein profiling of samples from diseased versus healthy individuals can help discover candidate proteins that can act as diagnostic and prognostic biomarkers. Our aim is to study protein in blood samples of patients with HCV induced liver cancer as well as healthy individuals, both qualitatively and quantitatively. Such a protein profile can aid in the development of a better understanding of HCV induced liver cancer pathways in identifying possible biomarkers. While it is recognised that cancer pathways and biomarkers may differ between populations of different ethnicities (Liu et al., 2017), no study has yet been performed on the Pakistani population to elucidate its unique proteomic profile for HCV induced liver cancer.

Based on existing studies, there are several proteins that can be expected to be notably high in concentration in the blood of cancer patients. Alpha-fetoprotein (AFP) has been added to the screening criteria for HCC (Tzartzeva et al., 2018). It is also upregulated in HCV led HCC. Other potential diagnostic biomarkers are osteopontin (OPN), heparin binding growth factor Midkine (MDK) and glypican-3 (GPC3) (Malov et al., 2021). Dickkopf-1 (DKK-1), glycoprotein called serum paraoxonase 1 (PON1) and squamous cell carcinoma antigen (SCCA), all are elevated in patients with HCC (Parikh et al., 2020).

1.5 GC-MS (Gas chromatography/Mass spectrometry)

GC-MS is used for the study of solid, liquid and gaseous samples. The first step of analysis includes the gas chromatograph, which is involved in vaporizing the sample into the gas phase and then by using a capillary column coated with a stationary (solid or liquid) phase, separates the sample into its various components. An inert carrier gas such as hydrogen, helium or nitrogen will be used to propel the compounds. After separation of components, each compound elutes from the column at a different time on the basis of its boiling point and polarity. The elution time of

each component is referred to its retention time. GC has the capacity to separate complex mixtures which contain hundreds of components.

Components leaving the GC column will be Ionized and fragmented by Mass spectrometer. These Ionized molecules and fragments will then be passed through the mass analyzer of the instrument, which quite often is a quadrupole or ion trap. Here the ions are separated on the basis of their different mass-to-charge (m/z) ratio. The data acquisition of GCMS will be performed in either full scan mode which covers the wide range of m/z ratios or on selected ion monitoring mode which gathers data of selected masses of interest.

The final step of the process comprises of ion detection and analysis. During these steps the fragmented ions appear as a function of their mass to charge ratios. Peak areas correspond to the quantity of the corresponding compound. Many different peaks in the gas chromatogram will be obtained if the complex mixtures are separated by GCMs. Each peak then generates a unique mass spectrum which will be used for the identification of compound. These unknown compounds and target analytes will be identified and quantified by using extensive commercially available libraries of mass spectra.

1.6 Protein Identification through GCMS

In GCMS, the complex mixture of proteins is first subjected to cleavage by enzymes. The peptide products obtained through this cleavage will be analysed through a mass spectrometer. It comprises of the following steps:

1. Protein extraction from the sample.
2. Contaminants and the proteins that are not of interest will be removed through fractionation.
3. Proteins of interest will be digested into peptides.
4. For obtaining the homogeneous mixture of peptides the proteins of interest will be subjected to post digestion separation.
5. Peptides analysis through Mass spectrometry.
6. Protein identification will be carried out by comparing the observed constituents to the one in the database of the previously identified peptides.

1.7 Insilico analysis

Analytical tools are required to process and interpret the data obtained from GCMS. Some of the tools that can be used are:

Protein sequence databases

These databases store information on proteins. There are two categories of universal protein sequence databases: simple records of sequence data and annotated databases in which additional information is also included in the sequence record.

PIR (Protein information resource)

It is the old and the first protein sequence database to be used. Protein sequences from a wide range of species are present in PIR.

SWISS-PROT

This is a detailed protein sequence database which provides a high level of description related to the protein function, the structure of its domain, post translational modifications and variants. It contains data of a wide variety of organisms. Entries in Swiss prot also contain data relevant to genetic disease with medical and biomedical basis of the diseases. These databases also store information linked with the mutations of the genetic diseases and polymorphisms.

TrEMBL (Translation of EMBL nucleotide sequence database)

This is a computer annotated database - created to make new sequences available. All the coding sequences in the Swiss prot like format are included in the EMBL sequence database except the CDs already included in the Swiss prot.

CHAPTER 2

LITERATURE REVIEW

2.1 Liver Cancer

Liver cancer was not among the six most common cancers in the world, but it was the third most common cause of death in 2020. Incidence rates of liver cancer have been growing in many countries (Petrick et al., 2016) and are expected to continue to rise over the next decade. Hepatocellular carcinoma (HCC) and intrahepatic cholangiocarcinoma (ICC) are the two most common histologic forms of primary liver cancer. HCC is caused by oxidative stress, inflammation, and underlying liver disease in hepatocytes (Ambade & Mandrekar, 2012), whereas ICC is caused by cholangiocytes that line the intrahepatic bile duct. On a global scale, HCC accounts for around 75% of all liver malignancies, while ICC accounts for approximately 12–15%.

About 90% of primary liver cancers are hepatocellular carcinomas (HCC) (World Health Organisation, 2022). HCCs have a high incidence to mortality ratio because the tumours are resistant to chemotherapy. The low effectiveness of treatment may also be due to the fact that less than 40 percent of HCC cases are diagnosed at an early stage (Ringehan et al., 2017).

2.2 Global Trends in Incidence

Asia and Africa have the highest incidence rates of liver cancer in the world. However, there is a great deal of variation within specific geographic regions. For example, between 2008 and 2012, the male age-standardised rate (ASR) per 100,000 people in Asian regions with cancer registries ranged from 1.8 in Dindigul, India, to 69.1 in Yanting County, China. While China accounts for nearly half of the global burden, Mongolia has the highest single incidence rate, with an estimated 2018 rate

of 93.7. Outside of Asia and Africa, northern Central America has the highest rates in the world, with Guatemala having the highest estimated rate (ASR = 14.9).

Between 1978–1982 and 2008–2012, the incidence of liver cancer increased in many parts of the world, most notably in Oceania, North and South America, and much of Europe. In contrast, incidence rates in many Asian countries have decreased. In the United States, the 5-year relative survival rate for liver cancer is only 18%. In less developed regions, the prognosis is even worse; thus, incidence and mortality rates are roughly equivalent in all countries (Petrick & McGlynn, 2019).

2.3 Prevalence of Liver Cancer in Pakistan

Liver and intrahepatic bile duct cancer is the fifth most common malignancy in Pakistani adults according to the Collective Cancer Registry report of 2004 to 2020 by Shaukat Khanum Memorial Hospital and Research Centre. 4 percent of cancer patients belong to this category. Liver cancer is more common in males (Cancer Registry and Clinical Data Management (CRCDM), 2021). This trend is echoed by data collected by Karachi Cancer Registry from 2017 to 2019; liver cancer is the second most common cancer in male adults above 20 years of age. The age standardised incidence rate (ASIR) of liver cancer is 16.1 and 7.8 per 100,000 for male adults and female adults respectively (Pervez et al., 2020). A recent study from Karachi found an increase in the number of admissions attributable to HCC, as well as the highest prevalence of viral-associated HCC. According to the paper, the majority of cases are detected at a late stage of the disease, when treatment choices are limited, and less than 1% of cases receive resection (Om Parkash, Saeed Hamid, 2016).

2.4 Prevalence and Distribution of Hepatitis C

58 million people were living with a chronic HCV infection in 2019 with 1.5 million people becoming newly infected with HCV over the year 2019-2020 (World Health Organization, 2021). Low- and middle-income countries, such as Pakistan, are home to 75% of the population infected with Hepatitis C Virus (HCV) (World Health Organization, 2021). Only 21% of HCV patients are aware of their disease.

Pakistan lies in the Eastern Mediterranean region as defined by World Health Organisation (WHO). This region has one of the highest prevalence of HCV as well as the highest rates of infection in the world. There were 470,000 new cases of HCV and 31,000 deaths due to HCV in the Eastern Mediterranean region in 2019 (World Health Organization, 2021).

Pakistan is one of the four countries that account for 40 percent of all HCV cases (World Health Organisation, 2021a). According to a study in which more than 3000 blood samples were collected from the general Pakistani population and screened, 5.73 percent of the samples were seropositive for anti-HCV antibodies (Usman, Wadood, & Nazir, 2018). Thus, more than 1 in every 20 people is infected with HCV in Pakistan. Statistical analyses via meta-regressions provide evidence for an increase in HCV prevalence in Pakistan over the coming years (Mahmud, Chemaitelly, Al Kanaani, Kouyoumjian, & Abu-Raddad, 2020).

2.5 Viral Liver Cancer

2.5.1 Global Prevalence

In 2018, the total number of new HCC cases was 841,000. Deaths from HCC were estimated to be 782,000 (Samant, Amiri, & Zibari, 2021). Out of this total, HCC cases attributable to Hepatitis C and Hepatitis B (HBV) infections numbered 500,000 in 2018. Mortality by viral hepatitis is led by liver cancer (World Health Organization, 2021). As seen with HCV, HCC is also more prevalent in underdeveloped regions (Ringehan et al., 2017).

2.5.2 Hepatitis C Led Viral Liver Cancer

According to a study conducted in a hospital in Rawalpindi from 2017 to 2018, 82% of the patients affected with HCC were HCV patients, and 3.1% had and HCV HBV co-infection. Notably, 68% of the patients had never received any treatment for the infection of cancer (Rashid, Alam, Farooq, & Ali, 2021). Hence it appears that there is a strong association between being infected with HCV and developing liver cancer in the Pakistani population.

2.5.3 Mechanism of HCV Led Viral Liver Cancer

Double stranded RNA replication intermediates of HCV viral particles are detected by Pathogen Associated Molecular Pattern receptors which then activate transcription factors such as NF-κB and IRF-3/7. These receptors also cause the production of interferons which in turn activate Interferon Stimulated Genes. HCV is able to counter such pathways and so becomes a chronic infection. Similarly, HCV can counter the innate immune response pathways of cells. Adaptive antiviral immune pathways which remain activated can clear out the infection only in 30-40 percent of cases. In the rest of the patients, they simply cause more inflammation (Mitchell et al., 2015). In the process, hepatocytes are repeatedly destroyed and regenerated. It is known that HCV causes HCC via liver cirrhosis (Ringehan et al., 2017).

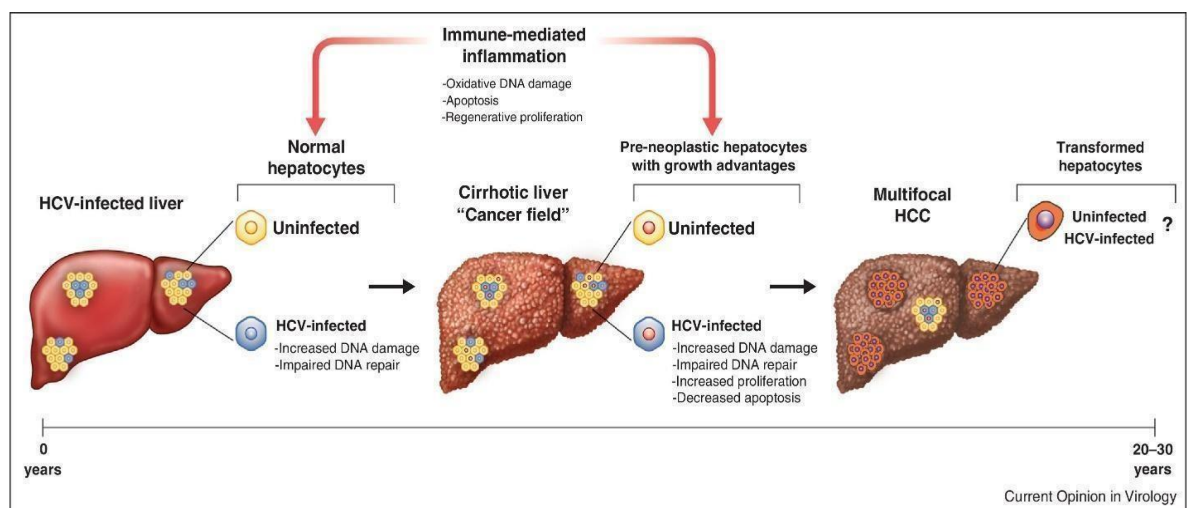


Figure 1 Model of HCC caused by HCV from (Mitchell, Lemon, & McGivern, 2015). Inflammation results in DNA damage, apoptosis of hepatocytes and subsequent proliferation to replace those hepatocytes. HCV itself leads to DNA damage as well. It inhibits DNA repair. Hence a 'Cancer Field' of pre-neoplastic hepatocytes is formed. During regeneration, cells which have a growth advantage may be selected. HCV compounds this affect via the promotion of cell proliferation and prevention of apoptosis. Inflammation due to the immune response, coupled with direct damage by HCV thus leads to formation of cancer over decades of chronic infection

Inflammation and the destruction-regeneration cycle appear to be the causative agents for liver cancer. Immune cells such as macrophages involved in inflammation

release Reactive Oxygen Species (ROS), Reactive Nitrogen Species (RNS) and cause peroxidation of lipids. Peroxidation of lipids, ROS and RNS damage DNA. Such damaged DNA could become heritable when cell regeneration pathways are activated to replace hepatocytes necrotised by HCV. Consequently, genetically changed hepatocytes will be more likely to become dedifferentiated and proliferative (Mitchell et al., 2015).

2.6 Liver Cancer Diagnosis

HCC is usually diagnosed after clinical deterioration, at which time survival is measured in months. Long-term survival requires detection of small tumours, which are often present in asymptomatic individuals, which may be more relevant to invasive treatment options. Monitoring of people at high risk for HCC is usually done using the serum marker alpha-fetoprotein (AFP) often in combination with ultrasound. Many other serological markers are currently being tested to help improve monitoring accuracy. Diagnosis of HCC often requires more sophisticated imaging modalities such as CT and MRI, which are capable of multiphasic contrast enhancement. Serum AFP used alone may be helpful if levels are markedly elevated, occurring in less than half of cases when diagnosed. Confirmation by liver biopsy may be performed in cases where the diagnosis of HCC remains uncertain.

2.7 Screening Tests

AFP is a serum glycoprotein that has been reported to detect preclinical HCC since it was first recognized as a marker for HCC over 40 years ago. The yolk sac and liver of the fetus produce high levels of AFP, which drop to <10 ng / dl within 300 days after birth (Kashyap et al., 2001). Subsequent elevated serum suggests an underlying condition that may be malignant. HCC can generate AFP levels ranging from normal to> 100 000 ng / ml. Normal AFP levels are present in up to 30% of patients at diagnosis and usually remain low even at advanced HCC. AFP > 400–500 ng / mL is considered a diagnosis of HCC, but less than half of patients can produce such high levels (Colombo, 2001). At this magnitude level, AFP's specificity is close to 100%, but at the expense of sensitivity which is below 45 % (Gupta et al., 2003).

Des-gamma carboxy-prothrombin (DCP), also known as PIVKA II (a protein induced by the lack of vitamin K), is a broadly used tumor marker that is highly specific for HCC as an abnormal form of prothrombin (Liebman et al., 1984). Recent studies suggest that DCP levels can be a prognostic indicator for patients with HCC (Nagaoka et al., 2003).

Ultrasound (US) imaging is commonly used in addition to, or in place of AFP, to help detect small liver tumors smaller than 3 cm. Its widespread use as a monitoring tool is associated with its non-invasiveness, high availability, and low cost. However, there are restrictions on the experience of the operator and when imaging subjects with obesity or cirrhosis. Sensitivity and positive predictions can be as low as 35% and 15% for cirrhosis, respectively (Pateron et al., 1994) (Murakami et al., 2001). CT imaging has not been well studied in the context of surveillance testing and is more commonly used for further diagnostic purposes. A study of patients with hepatitis C cirrhosis showed that CT scan imaging was more sensitive to HCC detection than US or AFP when used alone (88%, 59% and 62%, respectively) (Chalasani et al., 1999). Low availability and high cost limits the use of CT. There is little data on magnetic resonance imaging (MRI) that can be used as a surveillance tool for HCC.

2.7.1 Limitations of Current Diagnostics

Des-y-carboxy prothrombin, AFP concentration and the L3 fraction of AFP are the serological screening tests. There sensitivity is low and are not able to pick more than 60% of small HCCs. The likelihood of any of the tests being elevated increases with the increasing size of HCC. Their sensitivity is less than optimal even at a tumour size of more than 5cm.The most commonly used of these is the AFP but it is the least sensitive. All of these three can detect HCC but have poor prognosis. Ultrasound is recommended because its sensitivity in noncirrhotic liver is more than 80 percent. But for small HCC this test is not suitable as its sensitivity falls below 60 percent. Certain factors are known to limit the sensitivity of ultrasound such as liver with fats accumulation and obesity. One of the causes of failure is also the poor technique.

Also CT and MRI are more sensitive than Ultrasound but these are not for the screening of HCC. A four phase CT scan is required for good sensitivity, specificity and significant dose of radiations are not used in MRI and recently for the diagnosis of HCC MRI hepatobiliary phase imaging contrast agents have been used. But the evaluation of these agents as screening tests has not done. Both these test have very high incremental cost-efficacy ratio which make these tests economically unfeasible.

Another reason of failure of screening tests to detect small HCCs is irregular screening and the ideal time between screenings is still not known. Since the growth rates of tumours vary and it is also not possible to 100% predict the incurability of larger lesions and curability of smaller ones. Hence it is difficult to identify the appropriate screening interval for the patients. However it is suggested that a 6 months interval is the most suitable balance between the alot of missed opportunities for cure and too many false positives (Sherman, 2014).

2.8 Cancer Biomarkers

The discovery and usage of biomarkers can play very important roles in cancer related healthcare. They can help identify sections of the population which are at a greater risk for cancer. They can help detect cancer before it becomes symptomatic. They can be used for cancer staging, as prognostic markers and to distinguish a benign tumour from a malignant one. They can be used to evaluate the benefits versus harms of a particular therapy for a particular patient and to determine the risk of recurrence in recovered patients (Henry & Hayes, 2012).

2.9 Protein Biomarkers

WHO recognises the importance of early detecting cancer. Through early detection, mortality due to cancer and the socio-economic burden of the disease can be reduced (World Health Organisation). Biomarkers can help realise this goal of early detection.

Protein biomarkers can be more beneficial for early detection of cancer because they give tangible proof of oncogenic expression (Srivastava & Gopal-Srivastava, 2002). Protein biomarkers are especially intriguing candidates because of multiple reasons.

Firstly, proteomics analysis has become more accessible and better developed in recent years (Yates, 2019). Several techniques such as Mass Spectrometry (MS), Enzyme Linked Immunosorbent Assay (ELISA) and high-density antibody microarrays have been utilised. Secondly, protein biomarkers should be identifiable in samples of blood or other tissues which are easy and less invasive to obtain than conventional biopsies. If proteomic studies are used to develop biomarker signatures, they would have a greater chance of accurately guiding clinical decisions. Furthermore, a combination of protein and gene biomarkers can provide more accurate diagnosis of cancer. Thirdly, the proteomics approach provides richer information about the cancer than existing approaches (Borrebaeck, 2017).

2.10 Blood Biomarkers

Blood biomarkers are very important in the early detection of liver cancer and include categories primarily related to proteins, cytokines, enzymes, isozymes, and transcripts of related genes. Regardless of the different molecular indicators recognized for liver cancer, each cannot be a specific biomarker for different types of liver cancer alone, and the combination of them can be used in the identification of liver cancer type.

Apart from alpha-fetoprotein (AFP) and Des-gamma carboxy-prothrombin (DCP), α-L-fucosidase (AFU) is a potent biomarker. AFU has been shown to function as a useful biomarker for HCC, with serum levels of AFU being higher in patients with HCC than in patients with benign liver disease (Sarmad F, 2012). AFU is considered an early biomarker capable of diagnosing 85% of HCC patients 6 months prior to ultrasound detection (Waidely et al., 2017). The AFU level is positively associated with tumor size in patients with HCC and can be combined with AFP for the early diagnosis of HCC (Bertino et al., 2012).

Golgi Protein 73 (GP73) is a Golgi type II transmembrane glycoprotein with a molecular weight of approximately 70 kDa and is expressed primarily in epithelial cells of large numbers of human tissues (Wei et al., 2019). GP73 is rarely expressed in healthy subjects, but its expression is moderately increased in patients with cirrhosis and viral infections and significantly increased in patients with HCC

(*Knockdown of Golgi Phosphoprotein 73 Blocks the Trafficking of Matrix Metalloproteinase-2 in Hepatocellular Carcinoma Cells and Inhibits Cell Invasion - PubMed*, n.d.) (Block et al., 2005). One study reported that the sensitivity and specificity of GP73 to HCC was 74.6% and 97.4%, respectively, significantly higher than AFP's 58.2 and 85.3% (Mao et al., 2010).

Osteopontin (OPN) is a highly modified, phosphorylated, glycosylated and an extracellular matrix protein that binds to integrins and is expressed in a variety of cells such as immune system, epithelial tissue, smooth muscle cells, osteoblasts, and tumors (Ying et al., 2014). In one study, the prevalence of autoantibodies to OPN in the sera of patients with HCC, cirrhosis, and chronic hepatitis was 12.8, 15.6, and 3.1%, respectively, compared to 0% in the sera of healthy subjects. Diagnosing HCC by combining OPN with AFP led to the sensitivity being increased to 65% (Gao et al., 2020).

Limitations of Current

2.11 Gas chromatography Mass spectrometry (GCMS)

Different analytical tools including mass spectrometry (MS), fourier transform-infrared spectroscopy (ft-ir), nuclear magnetic resonance (nmr), and high performance liquid chromatography (hplc) have been used in metabonomics. Every technique being used has its own advantages and some shortcomings. Each of these techniques can be used for a particular set of analysis. None of these can be used for all the metabolites. however, analytical techniques which are most commonly used in metabolomics profiling studies include ms and nmr.recent developments in nmr machine includes the introduction of dynamic nuclear polarization (dnp) but the detection of secondary metabolites with nmr and its low sensitivity are still an issue. On the other hand, ms is a powerful analytical technique, has high sensitivity and can efficiently detect the low concentration molecules. ms can also be combined with other efficient and powerful techniques such as liquid chromatography (lc) or gas spectrometry (gc) which further enhances its efficiency to separate metabolites.

2.11.1 GCMS of biofluids and extracts

GCMS can be used for the analysis of metabolites in biofluids and this provides a wide range of information that can be in various research fields such as the study of organism response to environmental stresses, toxicology and the potential of various biomarkers discovery together with the monitoring of metabolic responses in different disease conditions. Metabolomics has found its application in clinical diagnostics for a wide variety of human diseases, investigation of genetic disorders, studying response to toxins and evaluation of the effects of various environmental Biofluids most commonly used for these purposes are blood and urine. However, there are some challenges that are faced with the blood samples. Blood contains a wide variety of chemically diverse classes of metabolites having different concentration levels and stability. These blood metabolites are usually bound to proteins through non covalent interactions. These non-covalently bonded proteins made measurements and the analysis of the obtained spectra complicated. These intact proteins are also responsible for the fluctuations in the concentrations of metabolites in blood plasma and serum. Therefore, for protein separation and deproteinization it is important to optimize and standardize the protocols when designing studies of metabolomics. Methods commonly used to separate blood proteins prior to MS measurements are liquid and solid phase extraction ultrafiltration. For the precipitation of proteins various organic solvents such as ethanol, methanol and acetonitrile have been used. However, methanol has been proved to be the most effective with less than 2 percent residual protein.

2.11.2 Components of GCMS

GCMS instrument has three main components which includes an ionization chamber which is involved in generating multiple ions by using different investigation samples, mass analyser which separates the ions on the basis of their mass to charge (m/z) ratio and lastly a high vacuum detector which is involved in determining the m/z values of ions and recording their relative abundance. In gcms the compounds are detected on the basis of a mass spectrum and a retention time. Despite the identification of metabolites, the technique has made possible the quantitative

analysis of analytes by using appropriate comprehensive profiling of biological samples such as human blood is challenging due to their complexity and dynamics. However, the metabolic profiling of compounds which belong to different chemical classes such as amino acids, amines, organic acids, and carbohydrates is possible by gcms based metabolomics using different methods.

2.12 Data Analysis tools for GC-MS

Molecular biology advancement rate has increased when biological research couples the bioinformatics tools to unravel the scientific question. Numerous approaches are employed to gain in-depth knowledge. Some of the tools employed are discussed below:

2.12.1 Cytoscape

Cytoscape is a java-based tool and it is used to trace the pathways of identified protein biomarkers. The pathways of identified proteins are then interlinked to visualize their connection with each other and their simultaneous effect on the disease (S Cline & Smoot, 2007).

2.12.2 Panther

Panther gives a very comprehensive classification of proteins. This tool combines ontology, pathways, gene functions and statistical analysis tools that allow biologists to analyze large scale genomic data from proteomics, sequencing and expression experiments. This tool gives the evolutionary and functional classification of proteins (Mi & Huang, 2017).

2.12.3 Biorender

This is an illustration tool used mainly to draw and share effective scientific drawings. This is also used to construct pathways of proteins and other scientific figures for presentations, publications, and posters (M. Perkel, 2020).

2.13 Aims and Objectives

The aims of this study were to identify potential metabolic biomarkers by evaluating the metabolite levels in plasma samples from HCC cases and patients with liver cirrhosis. Moreover, another objective included the validation of prognostic markers through insilico studies. Protein biomarkers can be more beneficial for early detection of cancer because they give tangible proof of oncogenic expression They will also help in the designing of effective therapies and treatment plans for cancer patients.

CHAPTER 3

METHODOLOGY

3.1 Samples Collection, Handling and Transportation

The Research project was approved by the Institutional Review Board (IRB) of Atta-ur-Rahman School of Applied Biosciences (ASAB) to ensure it aligns with ethical requisites of research. The blood samples for the project were taken from Combined Military Hospital (CMH), Rawalpindi. Consent was taken from these patients prior to blood collection along with a set of questions that were asked from each one of them. These included gender, age, co-morbidity and previous treatment if any. The patient history and consent form is attached herewith in the Appendix section. Similar collecting and storage tubes for all the study samples were used in order to ensure the comparability of the results. The tubes were labelled properly before collecting the blood. A total of 6 samples were collected – 3 from healthy individuals which acted as controls and 3 from diseased patients of Liver cancer which acted as samples. Samples were labelled accordingly: C1, C2 and C3 for controls and L1, L2, L3 for diseased samples. About 10 ml of blood samples from a peripheral vein were collected in a clotting activator containing serum tubes. The lids were closed, and the sample vials were shaken gently and thoroughly after collection. To allow coagulation, the vials were stored at room temperature in an upright position. In most cases clot formation occurred after 30 to 40 minutes. To transport samples, dry ice in insulated boxes or liquid nitrogen was used in order to avoid melting. All the samples of a single study were sent in one parcel.

3.2 Serum Extraction

The samples were centrifuged at 2,500 × g for 15 minutes to separate the serum. The labelled samples were cool, and the pipetting steps were performed by keeping the samples on ice. Then without aspirating blood cells, 100 µL of serum was transferred into pre-cooled storage vials. The tubes were not filled above 75 percent of their capacity. The samples were frozen at -20°C and then stored at -80°C.

3.3 Samples Preparation

3.3.1 Chemicals required

The chemicals required for sample preparation were ordered from the company. This included Extraction solvent which is methanal: water in a 1:1 v/v ratio, Methoxyamine HCL dissolved in pyridine, 1% trimethylchlorosilane (TMCS) and MSTFA (N-methyl-N-(trimethylsilyl) trifluoroacetamide).

3.4 Steps involved in the preparation of serum samples

3.4.1 Samples lysis

Sample lysis is crucial to separate the polar and non-polar metabolites from other cellular substances. To avoid the effect of variable cell number, the same number of cells were used for each sample. About 300 microliters of an extraction solvent (methanol: water, 1:1 v/v) and 100 microliters of serum samples were mixed. The extraction solvent separates the polar metabolites. In order to separate the non-polar metabolites, a non-polar solvent is required. The samples were vortexed for 2 minutes in order to ensure the quantitative extraction of metabolites. The samples were kept on ice for 1 hour. During this period, the samples were vortexed again and again after every 15 minutes for 2 minutes.

3.4.2 Centrifugation

The samples were centrifuged at 13000 rpm for 10 minutes at -4°C in order to separate the cell insoluble matrices. After centrifugation the supernatants were collected in separate GC vials and the pellets discarded.

3.4.3 Sample drying

After collecting the supernatants in separate GC vials the samples were dried via vacuum concentrators at 37°C for 1 hour. These concentrators use heat, centrifugal force and vacuum to remove moisture from the samples. This is commonly used in the preparation of samples. When vacuum is applied, the centrifugal force created in

it prevents the liquids from bumping out of the tube. The application of heat is indirect, and it is through the walls of the vacuum chamber.

3.4.4 Derivatization

Derivatization is done to transform the non-volatile or polar metabolites into volatile or non-polar metabolites. It also enhances the thermal stability of the products, improves the response of the detector by introducing functional groups and enhances the separation performance of the GC. To derivative the dry samples, 25 microliters of 20mg/ml methoxamine hydrochloride was dissolved in pyrimidine and used to derivatize the samples. After addition of this derivatization reagent the samples were subjected to vortex mixing for 2 minutes and then stored for about 6 hours at 25° C.

3.4.5 Silylation

Silylation is another method of derivatization used prior to GC. This is also done to reduce the polarity of the metabolites, to improve their stability and also their behaviour in GC. For silylation 25 microliters of MSTFA + 1% TMCS was dissolved in pyridine and this mix was used for silylation. After addition of this silylation reagent the samples were subjected to vortex mixing for 2 minutes.

3.4.6 Incubation

The samples were then subjected to an incubation at the temperature of 50°C for 30 minutes. After this they were shifted into 200 microliters micro inserts. The samples were then ready to undergo GC analysis.

3.5 GC/MS

In GCMS the sample first passes through a GC unit where the separation of compounds in a mixture takes place. The molecules then passed through the MS unit where an ionization of molecules through electron ionization takes place. The detector detects the molecular fragments on the basis of their m/z ratio. For the mobile phase the gas chromatography consists of a gas supply (N_2, H_2 or He are used as carrier gases).It also consists of a flow control to control a carrier gas, a heat

controlled oven, a heat controlled injector, a detection system and a data recording device. The stationary phase may have either a polymeric liquid phase Gas-liquid chromatography (GLC) or a solid phase Gas-solid chromatography (GSC). The stationary phase on the capillary column inner wall. In GC the inert carrier gas carries the analytes of the interest present in the mixture of compounds. On the basis of their vapor pressures these analytes are vaporized or pumped through the stationary phase. Because of their physical and chemical natures, they interact with the stationary phase differently and hence elute at different retention time from the column. Distribution of compound between the stationary phase and the mobile phase governs the interaction between an analyte and the stationary phase. This is denoted by the Kc (a distribution coefficient).

$$Kc = Cs/Cm$$

Cs indicates the concentration of analyte present in the stationary phase while Cm indicates the concentration of analyte present in the mobile phase. A large Kc corresponds to a longer retention time of a compound in a stationary phase or inside of the column. The temperature of the column and the chemical properties of the stationary phase controls the distribution coefficient.

3.6 GCMS analysis

For the metabolomics analysis a GC-MS-QP 2010 ultra-system along with lab solutions GCMS software was employed. In order to separate the metabolites a Restek Rtx -5ms column was used.99.9% helium at a flow rate of 1.0 ml/min was used as a carrier gas in the GC column. The oven is initially set at a temperature of 60°C and was kept at this temperature for 2 minutes. During the analysis the temperature of the oven is raised to 310°C by 50° Celsius per minute. The temperature of both the interface and ionization was kept at 250°C.The metabolites were analysed within the range of 50 to 650 amu and this analysis is done in full scan mode. By using an AOC-20i injector 10 microliters of sample was injected in splitless mode. By using the NIST/NIH/EPA Mass spectral library,GC total ion chromatograms and fragmentation patterns of the compound were analysed. Each sample was run for 43.67 minutes.

3.7 GC-MS data Analysis

The raw data from GC-MS was received in the form of a spreadsheet. The p-value for each protein was calculated. Only proteins with p-values less than 0.05 were selected for further filtration. This yielded a pool of 286 proteins. In some cases, the concentration of the protein was 0. To calculate foldchange, an equal small arbitrary value (0.1) was added to all values. The average concentration of these proteins was then calculated for both the control and disease samples. Then foldchange was calculated by this formula

Foldchange = average concentration of proteins in disease samples ÷ average concentration of proteins in the samples.

Proteins with a foldchange greater than 50 were filtered out for further analysis. Filtered proteins numbered 16.

3.7.1 Cytoscape

Cytoscape is a java based tool and has a lot of applications in biological research. It is used mainly to visualise and analyse network or interaction graphs having edges and nodes between molecules. The proteins obtained after filtration were visualised in cytoscape and their pathways were visualised and analysed. The pathways of 16 proteins were integrated to trace their connection with each other and their regulation of disease (T. Doncheva & H. Morris, 2018).

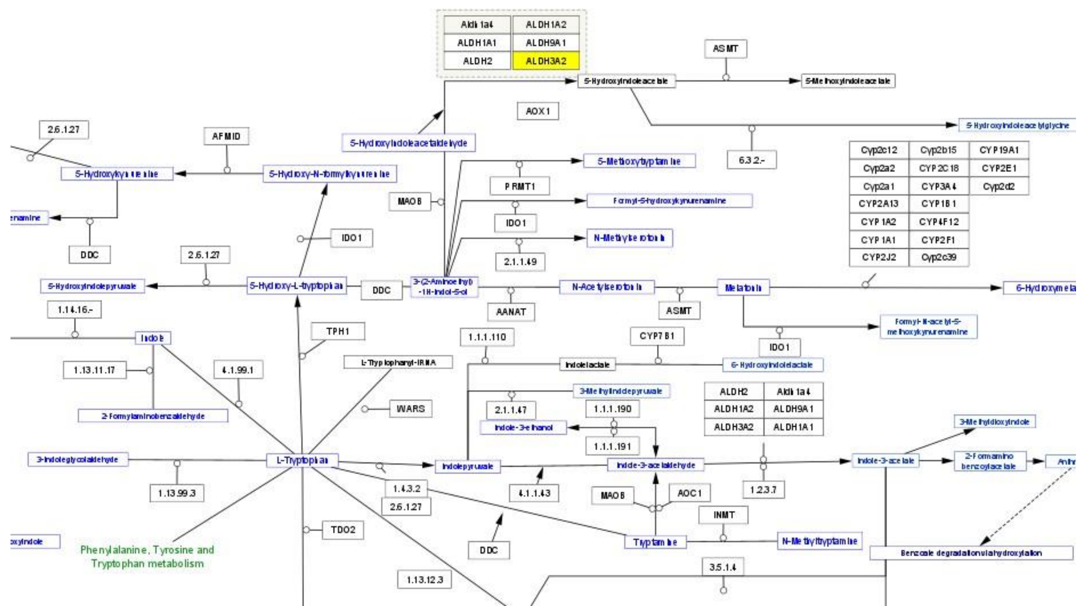


Figure 2 The cytoscape interface representing the role of one of the proteins, ALDH3A2 in metabolic pathways.

3.7.2 PANTHER Tool

PANTHER (protein analysis through evolutionary relationships) is an important bioinformatics tool typically employed for classification of gene/protein families via their biological databases system. This can further be used to classify and identify the function of gene products. The 16 proteins after filtration, in our project, were subjected to PANTHER analysis which gave the localization and exact location of proteins in the mammalian cells. This is helpful when generating pathways as well as for understanding their function and regulation (Mi & Muruganujan, 2013).

3.7.3 Protein Atlas

The Human Protein Atlas (www.proteinatlas.org) is an omics-based knowledge source. It was used to evaluate the status of filtered proteins as prognostic and diagnostic markers. Kaplan-meier graphs for some proteins were also obtained from the “pathology” section of each protein.

Methodology for the Analysis of the Role of PPP1R163

3.9 Software and Databases:

Data of the variants of the gene PPP1R163 was obtained from Ensembl. Transcript with the transcript ID ENST00000549336.6 was selected from the transcript table. Then the variant table of the transcript was retrieved. Data of variants was filtered by selecting "intron variants" in the "consequences" category of the variant table web page and downloading the relevant file with the filtered variants. 15952 variants' data was thus obtained (Table 1).

Table 1 Type and number of variants of gene PPP1R163.

Coding consequences	**Number**	**Non coding consequences**	**Number**
Synonymous variant	869	5 prime UTR variant	602
Missense variant	1809	Start retained variant	0
Inframe insertion	0	Start lost	37
Inframe deletion	22	Incomplete terminal codon	0
Stop gained	49	Upstream gene variant	10296
Frameshift variant	45	Mature miRNA variant	0
Coding sequence variant	301	NMD transcript variant	16756
Splice donor variant	45	Transcript ablation	0
Splice acceptor variant	45	Splice poly pyrimidine tract variant	0
Splice region variant	497	Intron variant	115965
Stop lost	2	3 prime UTR variant	3615
Stop retained variant	1	Downstream gene variant	10630
		Transcript amplification	0

3.9.1 Regulome DB

All of the 15952 variant IDs were searched through RegulomeDB (Boyle et al., 2012), an SNP annotation database. It provided the ranks and scores of the variants. These ranks were used to filter out SNPs with regulatory potential. The ranks and scores were obtained for 11564 variants. The rest did not yield any results in RegulomeDB. dbSNP IDS with a score of 0.99 or higher were filtered out to obtain 261 variants. A text file was compiled with the wild type and mutated sequences of these variants.

3.9.2 Alibaba

Alibaba2 (Alibaba2) is an online tool which was used to predict transcription factor binding sites for filtered dbSNP IDs wild type and mutated sequences. The transcription factor binding sites in the wild type and mutated sequences were tabulated and compared. Sequences which had an addition of binding sites after the mutation were chosen for further analysis.

3.9.3 KEGG

The binding sites which were unique to mutated sequences; sites that were not present in the wild type sequences studied, were searched through Kyoto Encyclopedia of Genes and Genomics (KEGG) Pathway database (Kanehisa & Goto, 2000) to identify the cellular signaling pathways in which the transcription factors complementary to the selected sites may be involved. Through KEGG and existing literature, the relevance of SNPs of PPP1R163 with the liver cancer proteins was investigated.

CHAPTER 4

RESULTS

4.1 GC-MS Data

A total of 800 proteins were identified in the blood samples through GC-MS. They were matched with the Uniprot database to obtain Protein IDs (The UniProt Consortium et al., 2021). About 285 of these proteins had p-values less than 0.05. Averages were calculated for each of the 285 proteins for healthy and disease samples. These averages were used to calculate foldchange values. The lowest foldchange value obtained was 0.01 while the highest was 180. The proteins with a foldchange greater than 50 were chosen. They numbered 16 (Table 2). They were filtered and then studied further via software and literature.

4.2 Cellular localization of proteins

To identify the location of each protein a software called PANTHER was used. All of the 16 proteins of interest were checked in order to determine the exact point in the cell where they are present and performing their role as shown in Table 2

Table 2 Subcellular localisation of the filtered proteins.

Protein ID	Protein Name	Location
P05141	ADP/ATP translocase 2;SLC25A5;ortholog	Mitochondrial inner membrane
P30050	60S ribosomal protein L12;RPL12;ortholog	Cytosol
P05388	60S acidic ribosomal protein P0;RPLP0;ortholog	Cytosol
P12277	Creatine kinase B-type;CKB;ortholog	Mitochondrion, Cytoplasm

P51648	Aldehyde dehydrogenase family 3 member A2;ALDH3A2;ortholog	Cytoplasm, Peroxisomes
P51659	Peroxisomal multifunctional enzyme type 2;HSD17B4;ortholog	Peroxisome
P49327	Fatty acid synthase;FASN;ortholog	Cytoplasm, Melanosome
P12236	ADP/ATP translocase 3;SLC25A6;ortholog	Mitochondrial inner membrane
P05023	Sodium/potassium-transporting ATPase subunit alpha-1;ATP1A1;ortholog	Plasma Membrane, Melanosome
Q00610	Clathrin heavy chain 1;CLTC;ortholog	endocytic vesicle, extracellular space, spindle, clathrin-coated vesicle, membrane coat
Q99715	Collagen alpha-1(XII) chain;COL12A1;ortholog	extracellular matrix
P09651	Heterogeneous nuclear ribonucleoprotein A1;HNRNPA1;ortholog	Cytoplasm, Nucleus
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic;IDH1;ortholog	Membrane, Peroxisome, Cytosol, Mitochondrion
P34897	Serine hydroxymethyltransferase, mitochondrial;SHMT2;ortholog	Cytoplasm
P02545	Prelamin-A/C;LMNA;ortholog	Nucleus, Intermediate Filament
P40926	Malate dehydrogenase, mitochondrial;MDH2;ortholog	Cytoplasm

All the proteins were different and involved in different pathways. However, there were few which linked with each other through a common pathway intermediate or

by having a similar site of action or function in the cell. Panther gives a brief picture showing the different classes to which the identified proteins belong. Of all the proteins about 38 percent belongs to the class of proteins involved in metabolite interconversion (Figure 3).

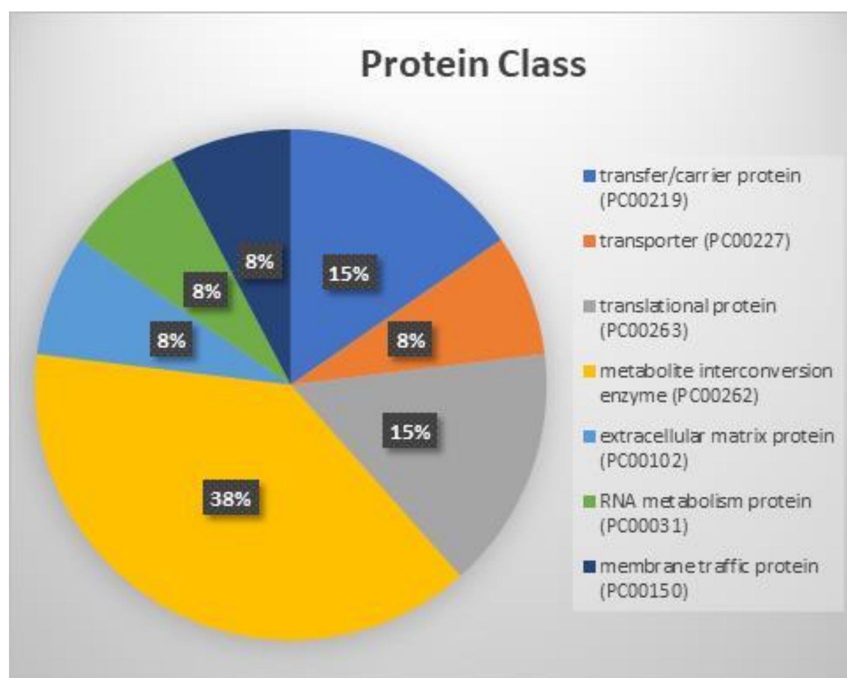


Figure 3 Different classes of obtained proteins.

It also gives a functional classification of proteins in terms of percentage and also clearly giving us the number of proteins performing or are involved in specific functions within the cells. About 43 percent, the largest proportion of proteins, are catalytic enzymes (Figure 4). The second largest group, 29 percent, has the molecular function of binding (Figure 4).

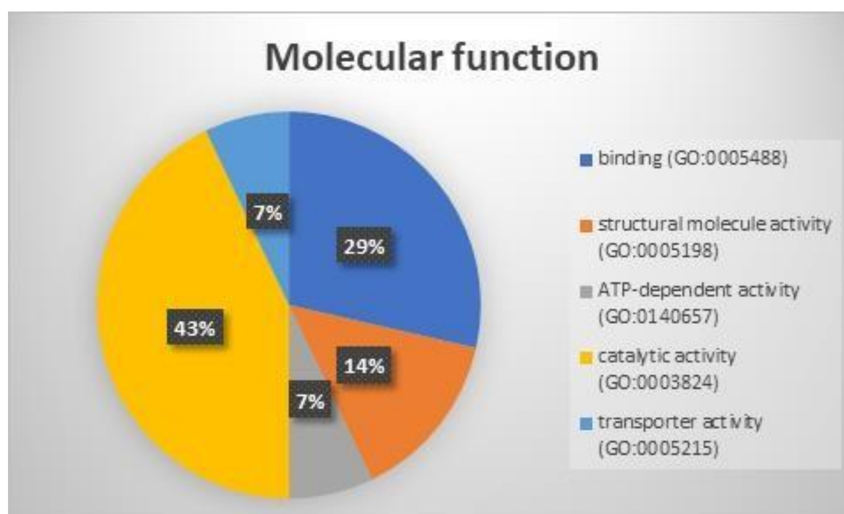


Figure 4 Molecular functions of the proteins

4.3 Identification of protein role in liver cancer

- **ADP/ATP translocase 2; SLC25A5; ortholog.**

In tumor cells mitochondria produces abnormal numbers of ATPs from aerobic glycolysis. An enzyme called cytoplasmic hexokinase generates Glucose 6 phosphate which is then utilized by an isoform of hexokinase HK ll (Figure 6) present in the mitochondria for the generation of ATP. Through the Voltage dependent anion channel (VDAC), ATP4- is imported across the outer mitochondrial membrane into the mitochondria and through ANT2 isoform it is transported across the inner mitochondrial membrane. This leads to the maintenance of the membrane potential. This imported ATP4- is directed towards hydrolysis by the F1 domain of an enzyme ATP synthase which releases ADP3- in the mitochondria and the corresponding development of negative charge on the matrix side and then through the F0 component the transfer of a proton into the space between the inner and outer membrane (Chevrollier, 2011).

- **Creatine kinase B-type; CKB;ortholog.**

miR-551a and miR-483 are the micro RNAs which are involved in the suppression of liver colonization and metastasis by targeting the CKB which is involved in the phosphorylation creatine to phosphocreatine (Figure 6). However, cancer cells release this CKB into extracellular space where it starts catalysing the formation of

phosphocreatine. SLC6A8 transporter is then used to import it into the cells and it starts generating ATPs which acts as a fuel for metastatic cells. Higher CKB and SLC6A8 levels are found in metastatic cells relative to normal cells. While reduced miR-551a/miR-483 levels are found in metastatic cells relative to normal cells (Loo & Alexis Scherl, 2015).

- **Malate Dehydrogenase (MDH), Isocitrate Dehydrogenase (IDH) and Serine Hydroxymethyltransferase 2 (SHMT 2)**

The proteins localised in the mitochondria were found to have effects on liver cancer directly or indirectly. These included Malate Dehydrogenase (MDH), Isocitrate Dehydrogenase (IDH) and Serine Hydroxymethyltransferase 2 (SHMT 2). Malate dehydrogenase (MDH) is a proteomic enzyme extensively dispersed in living organisms and plays a key role in the central oxidative pathway (Siavash Mansouri, 2017). It is involved in the catalysis of the conversion between malate and oxaloacetate employing NAD^+ or $NADP^+$ as a cofactor (Figure 6). Serine Hydroxymethyltransferase 2 (SHMT 2) is an important enzyme that transfigures serine into glycine and a tetrahydrofolate-bound one-carbon unit, eventually contributing in thymidine and purine formation promoting tumour growth (Figure 6). Isocitrate dehydrogenase (IDH) is an indispensable enzyme crucial for cellular respiration in the Kreb's cycle or also called the tricarboxylic acid (TCA) cycle (Figure 6). Repeated mutations in IDH1 or IDH2 are rampant in numerous cancers (Borger, 2012). The mutated IDH protein has a neomorphic activity via the gain-of-function mutation, catalysing the reduction reaction of α-ketoglutarate (α-KG) to 2-hydroxyglutarate (2-HG) by NADPH. Cancer-linked mutations of the enzyme IDH block normal cellular differentiation and promote tumorigenesis via the abnormal production of an oncometabolite. (Zeng et al., 2021)

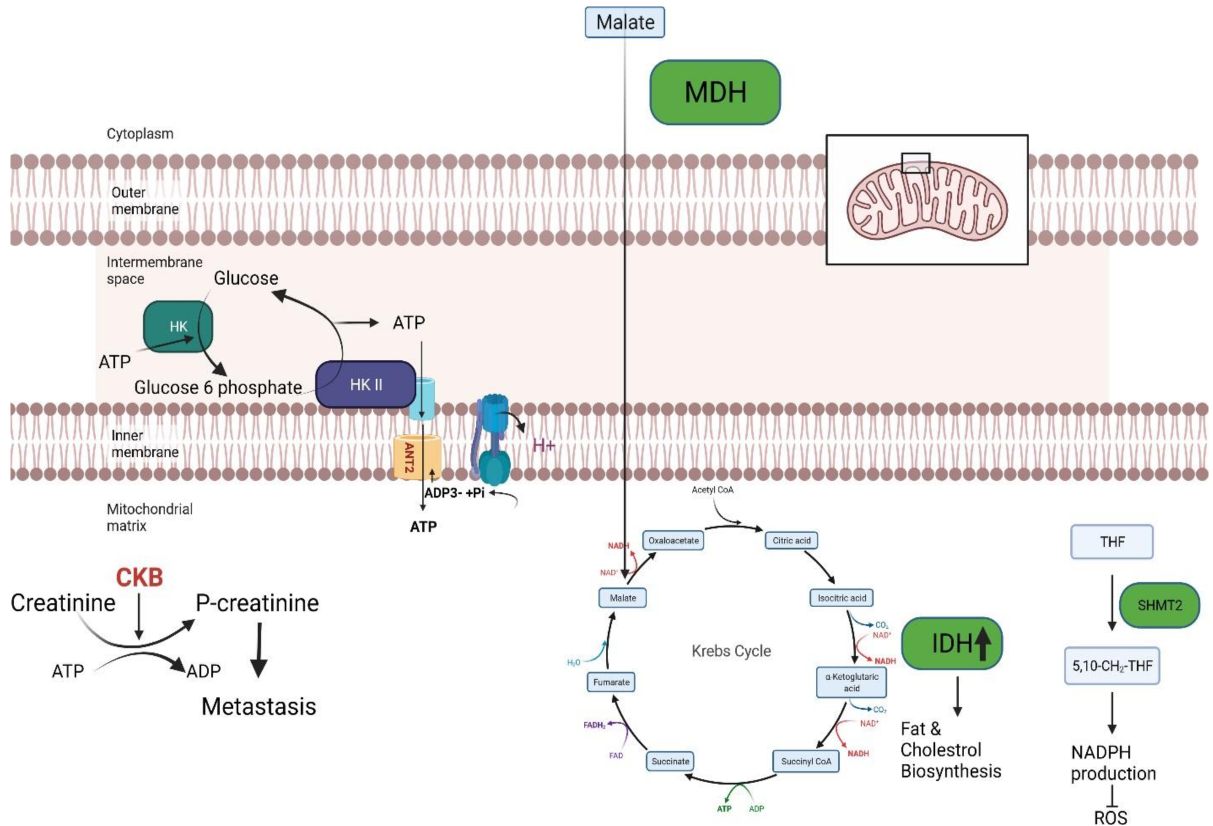


Figure 6 Pathways of proteins localised in the mitochondria. ANT2, IDH and MDH are involved in glycolysis either directly, as is the case for ANT2 or indirectly as shown for IDH and MDH. IDH converts isocitric acid into α Ketoglutaric acid. MDH help transport malate into the mitochondrion and converts malate into oxaloacetate. CKB and SHMT2 are involved in separate metabolic processes which may aid cancer.

ANT2: Adenine Nucleotide Transferase 2, CKB: Creatine Kinase B-type, IDH: Isocitrate Dehydrogenase, MDH: Malate Dehydrogenase, SHMT2: Serine Hydroxymethyltransferase 2

- **Fatty acid synthase; FASN; ortholog**

FASN is an enzyme involved mainly in the metabolism of lipids. Apart from lipids metabolism this enzyme is also involved in the glycolysis and amino acids metabolism. Mitochondria are the site of oxidative phosphorylation. Three main factors primarily determine its function -mitochondrial dynamics, mitochondrial biogenesis and the substrates utilized in mitochondrial respiration. In many cancer types the positive linkage between FASN protein levels and oxidative phosphorylation has been found. Overexpression of FASN generates large amounts of free fatty acids. These fatty acids are then broken down into acetyl CoA which

supports the mitochondrial respiration through fatty acid oxidation (FAO) and generates huge amounts of energy compared to glucose (Figure 7). This relatively huge amount of energy generated through FAO leads to cancer metastasis which demands or requires relatively greater energy (Wai Fhu & Ali , 2020).

- **Sodium/Potassium-Transporting ATPase Subunit Alpha 1 (ATP1A1)**

ATP1A1 engages in protein protein interaction with Thyrotropin receptor.

ATP1A1 activates Src. Src activates MAP kinase which activates Ras. Ras initiates PI3K/Akt signaling as well as MEK/ERK signaling (Figure 7). Both these cascades lead to cell proliferation and growth (Yu et al., 2019). Furthermore, ATP1A1 suppresses IFN mediated JAK/STAT1 and JAK/STAT1/2 pathways which promote apoptosis in gastric cancer cells. Hence ATP1A1 aids tumour progression in gastric cancer (Nakamura et al., 2021).

- **Clathrin Heavy Chain (CLTC)**

Clathrin mediated endocytosis of activated Fas allows transduction of death signals via Fas, leading cells to apoptosis (Chakrabandhu et al., 2008). On the other hand, clathrin veers the Transforming Growth Factor β (TGF-β) signaling pathway towards anti-apoptotic effects in liver cancer (Figure 7). Clathrin expression is upregulated by TGF-β in liver cancer cells. TGF-β also transactivates the Epidermal Growth Factor Receptor (EGFR). Clathrin also plays a role in the correct trafficking of EGFR between the Golgi apparatus and the cell membrane. EGFR, when activated, requires clathrin for internalization and subsequent induction of pro-tumorigenic effects via the PI3k/Akt and ERK pathways. Simultaneously, clathrin has an inhibitory effect on the production of NOX4 protein by TGF-β signaling pathway. ROS production by NOX4 is thus reduced, aiding cell survival (Caballero-Díaz et al., 2020).

- **Adenine Nucleotide Translocase (ANT) 3.**

ANT 3 forms a part of mitochondrial Permeability Transition Pore. This proapoptotic action can be prevented by overexpression of Bcl-2 as Bcl-2 binds to ANT (Belzacq et al., 2003).

Interferon Gamma (IFN-γ) causes STAT1 dimer formation via the JAK/STAT pathway (Figure 7). The STAT1 dimer upregulates ANT 3 production. Interleukin-4 (IL-4) also upregulates ANT 3 production via the JAK/STAT pathway involves the activation of STAT6. Upregulation of ANT-3 promotes cell survival in T-cells, protecting them from apoptosis induced by dexamethasone (Jang & Lee, 2003).

- **Peroxisomal Multifunctional Enzyme (DHB4)**

DHB4 coverts testosterone and dihydrotestosterone, Androgen receptor activating steroids into androstenedione, an inactive steroid, during steroid biosynthesis. This function is suppressed in Castration Resistant Prostate Cancer (Ko et al., 2018).

In cancer cells, it leads to phosphorylation of STAT3 via both Akt and MEK/ERK pathways (Figure 7). Activated STAT3 is translocated to the nucleus and transcribes Cyclin D1 genes. Cyclin D1 binds to Proliferating Cell Nuclear Antigen (PCNA) to form a complex. This complex activates downstream proteins of STAT3 signaling which lead to cell proliferation (Lu et al., 2019).

- **Prelamin-(A/C)**

Prelamin A/C is a protein in the nucleus with various functions in healthy and diseased cells. It plays a role in maintaining structural stability of the cell, cell motility, mechanosensing, organization of chromosome, regulation of gene, cell differentiation, DNA damage repair, and telomere protection. Prelamin A/C deregulation results in various traits, including genomic instability and increased tolerance to mechanical insult, which can ultimately lead to more aggressive cancer and poorer prognosis (Liu, 2020).

- **Heterogeneous Nuclear Ribonucleoprotein A1 (HNRNPA1)**

A known RNA-binding protein, heterogeneous nuclear ribonucleoprotein A1 (HNRNPA1) plays an important part in RNA splicing, stability, and expression. HNRNPA1 has been identified as an oncogene, enhancing deterioration in multiple tumor types. HNRNPA1 primarily regulates the biological activity of the IFNβ

signaling pathway, p53 gene, and AKT signaling pathway, which are associated with tumor proliferation, invasion, and metastasis .

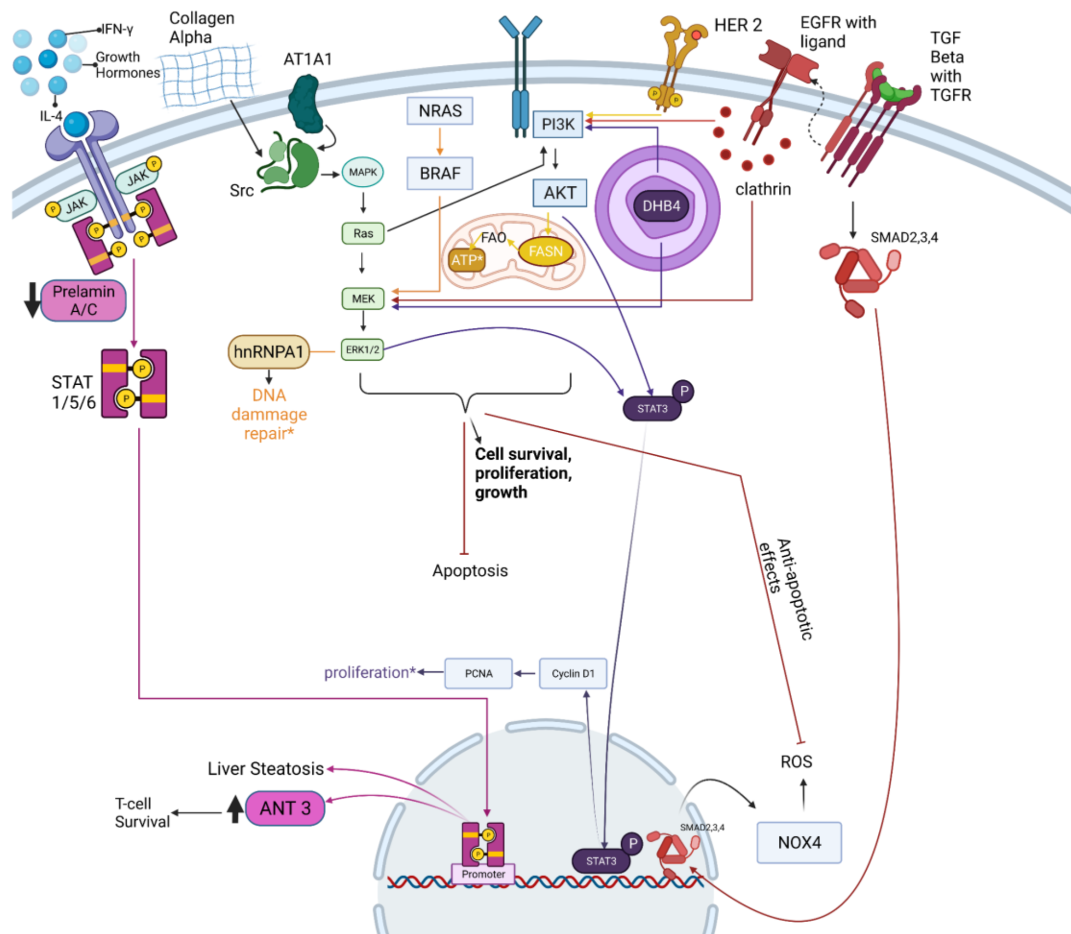


Figure 7 Proteins linked to the P13k/Akt; Clathrin, ATP1A1, DHB4, Collagen Alpha, FASN and HNRNPA1 and JAK/STAT pathway; Prelamin A/C and ANT3.

ATP1A1: Sodium/Potassium-Transporting ATPase Subunit Alpha 1, DHB4: Peroxisomal Multifunctional Enzyme, FASN: Fatty Acid Synthase, HNRNPA1: Heterogeneous Nuclear Ribonucleoprotein A1, ANT3: Adenine Nucleotide Translocase 3.

Integration of the pathways (Figure 7) shows that the PI3k/Akt pathway and the MEK/ERKs pathways play major roles in transducing the effects of many of the studied proteins, especially because several proteins stimulate both the pathways. Both pathways are well known drivers of cancer cell survival and proliferation (Hecquet, Lefevre, Valtink, Engelmann, & Mascarelli, 2002) (Larribere et al., 2004). Multiple initiators of these pathways were revealed at different locations: collagen

activates the pathways from the extracellular matrix via Src protein, ATP1A1 also activates them via Src but it is localised in the cell membrane, clathrin is cytosolic activator of the pathways while DHB4 is a peroxisomal activator. hnRNPA is unique in the scenario because it is a product of the MEK/ERK pathway rather than an initiator. Similarly, FASN is a product of the PI3k/Akt pathway which is localised in the mitochondria where it mediates metastatic effects. In total, 6 out of the 16 filtered proteins are connected to the MEK/ERK or P13k/Akt pathways.

- **60S ribosomal protein L12; RPL12; ortholog**

It has been revealed by phosphoproteomics that in polysomes the phosphorylation of Serine 38 in RPL12/uL11 which acts as a mitotic CDK1 substrate is depleted. However, during mitosis, the phosphorylation of RPL12/uL11 is found to be associated with the translation of specific mRNA subsets. Hence certain post translational modifications which occur in ribosomal proteins can regulate translation (Imami & Milek, 2018).

- **60S acidic ribosomal protein P0; RPLP0; ortholog.**

The expression of ribosomal proteins is associated with transformation and primary immortalization of the cells. The expression of few ribosomal proteins is found high in many cancer types. It has been found that in gastric cancer cells the downregulation of RPLPO protein induces the downregulation of CDK2 which then causes the suppression of cell cycle progression and cell growth. Phospholipase A and acyl transferase 4(PLAAT4) belongs to HREV107, which is a tumor suppressor protein family.PLAAT4 causes the downregulation of the expression of RPLPO by interacting with it.Knockdown of RPLPO leads to cell death mainly by proteins involved in the apoptosis such as Bax and BcI 2.However RPLPO silencing leads to cell cycle arrest and inhibition of cell cycle proliferation by inhibiting the expression of proteins involved in cell cycle. In PLAAT4 mediated cell death the interaction between RPLPO and PLAAT4 plays an important role.

- **Aldehyde Dehydrogenase Family 3 Member A2 (ALDH3A2)**

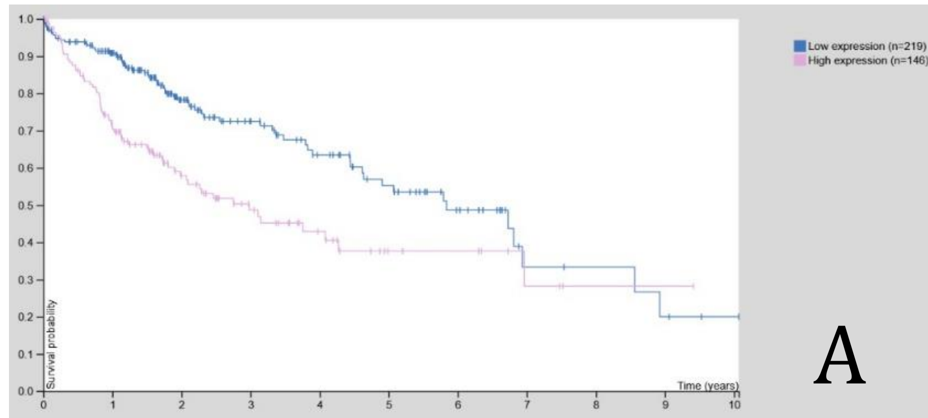
ALDH3A2 is involved in tryptophan metabolism. It converts 3-(2-Aminoethyl) -1H-indol-5-ol and 5-hydroxyindolacetaldehyde into 5-hydroxyindolacetate which leads to the formation of 5-methoxyindoleacetate. It also converts indole-3-aldehyde to indole-3-acetate. This ultimately leads to the formation of glucobrassicin.

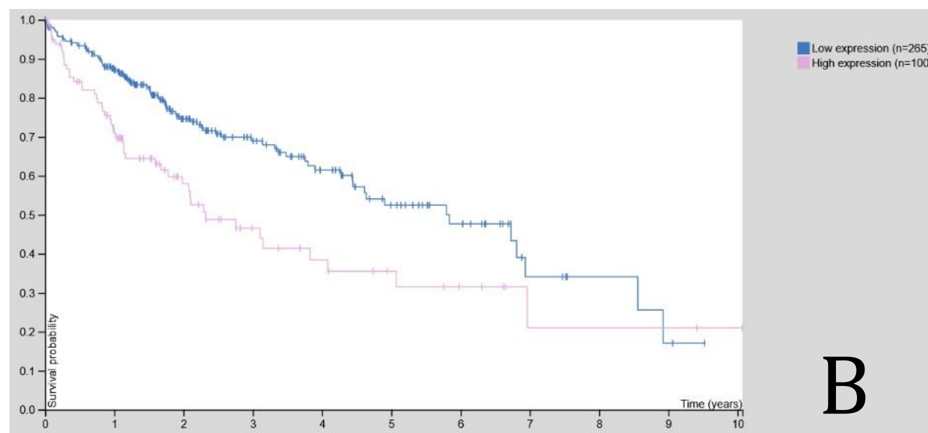
In cancer cells, ALDH3A2 reduces the formation of Reactive Oxygen Species (ROS) and peroxidation of lipids by converting long chain aliphatic aldehydes to fatty acids, aiding cell survival. Without ALDH3A2, fatty aldehydes could accumulate. They would damage DNA and proteins via oxidation (Yusuf et al., 2020).

4.4 Evaluation of Filtered Proteins as Diagnostic and Prognostic Markers

The Human Protein Atlas database was accessed to determine the diagnostic and prognostic significance of the filtered proteins. Half of the filtered proteins, ANT2, RPLP0, DHB4, ANT 3, CLTC, HNRNPA1, Prelamin A/C and MDH had medium or high expression in 7 or more out of 10 patients (Table 3) detected through cytoplasmic positivity or immunoreactivity according to the database (Uhlen et al., 2017). High expression in most patients, combined with the proteins' detectable levels in the current study, point to their potential diagnostic significance.

Three of the filtered proteins can serve as prognostic markers for liver cancer according to the database (Uhlen et al., 2017): ATP1A1, CLTC, and IDH1 since high mRNA expression of these proteins is significantly ($p<0.001$) correlated with changes in probability of survival of patients in the subsequent years. For all three of the proteins', high mRNA expression was correlated with lower survival probability over the years as seen in their Kaplan-Meier graphs (Figure 8).





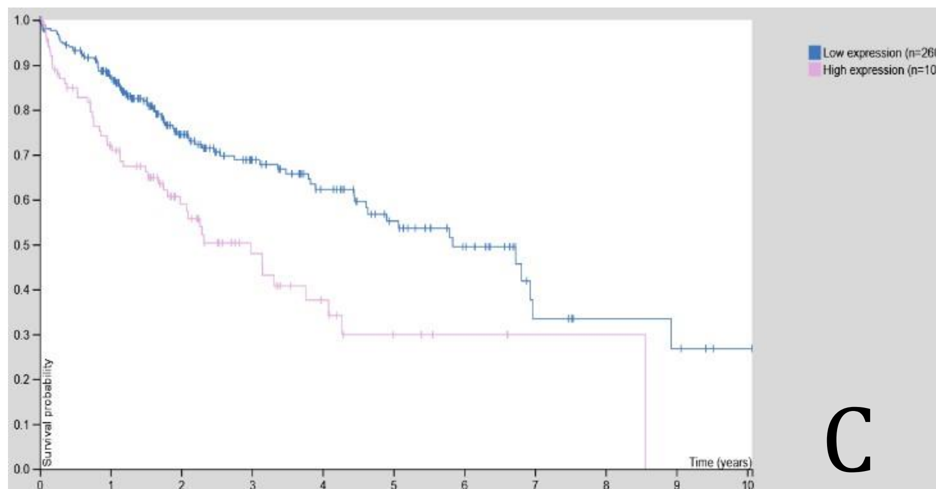


Figure 8 Kaplan-Meier graphs of proteins which can be used as prognostic markers of liver cancer. Probability of survival is plotted on the y-axis. Time in years is plotted on the x-axis. Number of patients in each category are mentioned on the top right corner. 8A: Graph for ATP1A1. Patients with high expression of ATP1A1 have a lower survival probability than the low expression category till 8.5 years. 8B: Graph for CLTC. Patients with high expression of CLTC have a lower survival probability than the low expression category till 9 years. 8C: Graph for IDH1. Patients with high expression of IDH1 have a lower survival probability than the low expression category throughout the time period of the study. The survival probability sharply drops to 0 at 8.5 years.

Table 3 Expression and prognostic status of each filtered protein in liver cancer.

Protein Name	Status as Prognostic Marker for Liver Cancer	Expression in Liver Cancer Patients
ADP/ATP translocase 2	Not considered	7/10 patients show medium or high expression
60S ribosomal protein L12	Not considered	4/10 patients show medium or high expression
60S acidic ribosomal protein P0	Not considered	9/12 patients show medium or high expression
Creatine kinase B-type	Not considered	0/9 patients show medium or high expression
Fatty acid synthase	Not considered	0/11 patients show medium or high expression
Aldehyde dehydrogenase family 3 member A2	Not considered	5/11 patients show medium or high expression
Peroxisomal multifunctional enzyme type 2	Not considered	12/12 patients show medium or high expression, 10 of which show high expression
ADP/ATP translocase 3	Not considered	7/10 patients show medium or high expression
Sodium/potassium-transporting ATPase subunit alpha-1	High expression is unfavourable with $p<0.001$	7/12 patients show medium or high expression
Clathrin heavy chain 1	High expression is unfavourable with $p<0.001$	9/12 patients show medium or high expression
Collagen alpha-1(XII) chain	Not considered	0/11 patients show medium or high expression
Heterogeneous nuclear ribonucleoprotein A1	Not considered	8/10 patients show medium or high expression
Isocitrate dehydrogenase [NADP] cytoplasmic	High expression is unfavourable with $p<0.001$	3/12 patients show medium or high expression
Serine hydroxymethyltransferase, mitochondrial	Not considered	7/12 patients show medium or high expression
Prelamin-A/C	Not considered	10/10 patients show medium or high expression
Malate dehydrogenase, mitochondrial	Not considered	12/12 patients show medium or high expression

4.5 Computational Analysis of Target Gene PPP1R163

Of all the variants of this gene we have filtered the non coding variants. A total of 141,695 non-coding variants were obtained. It has been observed that intron variants, upstream splice variants and downstream splice variants are the most prominent among all the non coding variants. The types of variants obtained, and their number are shown in the figure below.

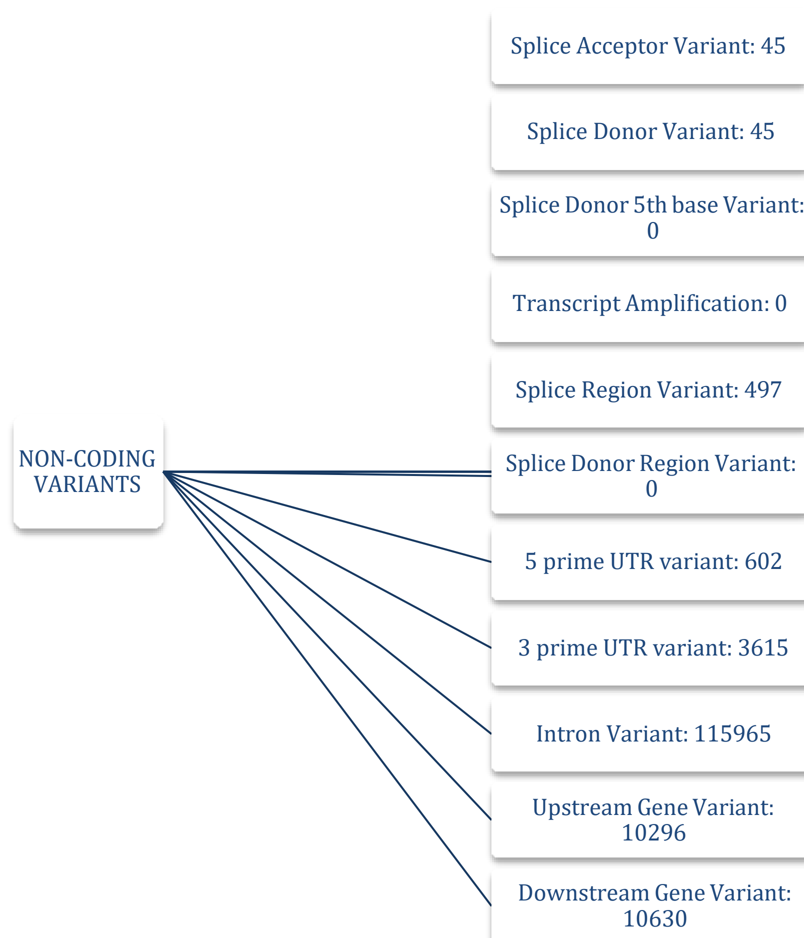


Figure 9 Type and number of Non-coding variants of PPP1R163 (obtained from Ensembl).

In order to obtain the ranks and scores of obtained intron variants a software called Regulome dB was used. For some intron variants Regulome dB showed no data and those were then listed out. Of all the ranks obtained it has been observed that the rank 4 and 5 are the most prominent among intron variants. Hence the focus of study was mainly based on those intron variants with the ranks of 4 and 5. Of all the variants we filtered those which show a score of 0.99. It has been shown in the figure below that the intronic variants with the rank 5 are the most prominent which a show a score of 0.99.

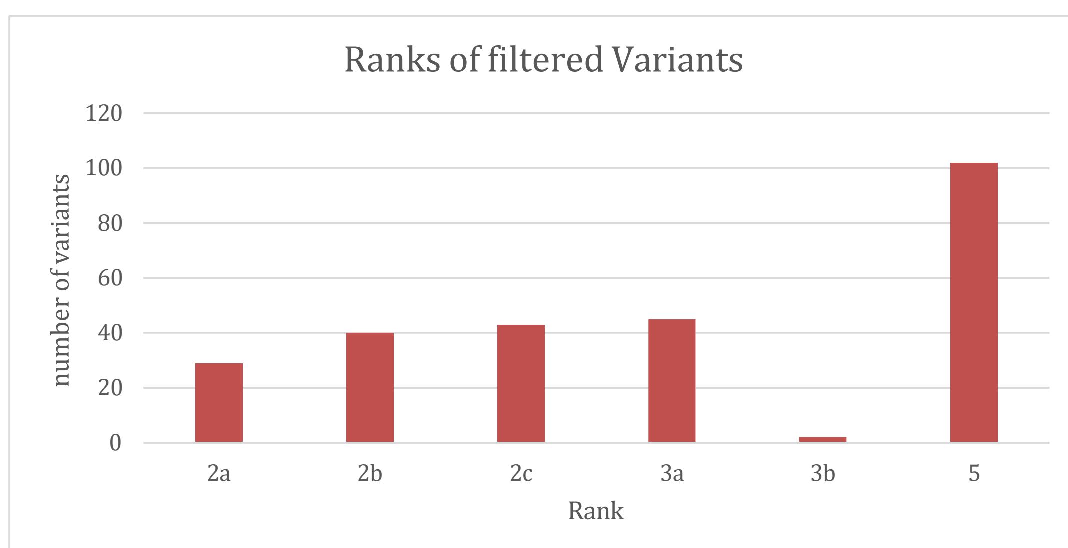


Figure 10 Number of variants (y-axis) with each rank (x-axis) filtered on the basis of >0.99 score from RegulomeDB. Each rank represents regulatory potential on the basis of the presence of certain elements in the sequence. Majority of the filtered SNPs have rank 5 which represents transcription factor binding potential or DNase peak in the sequence.

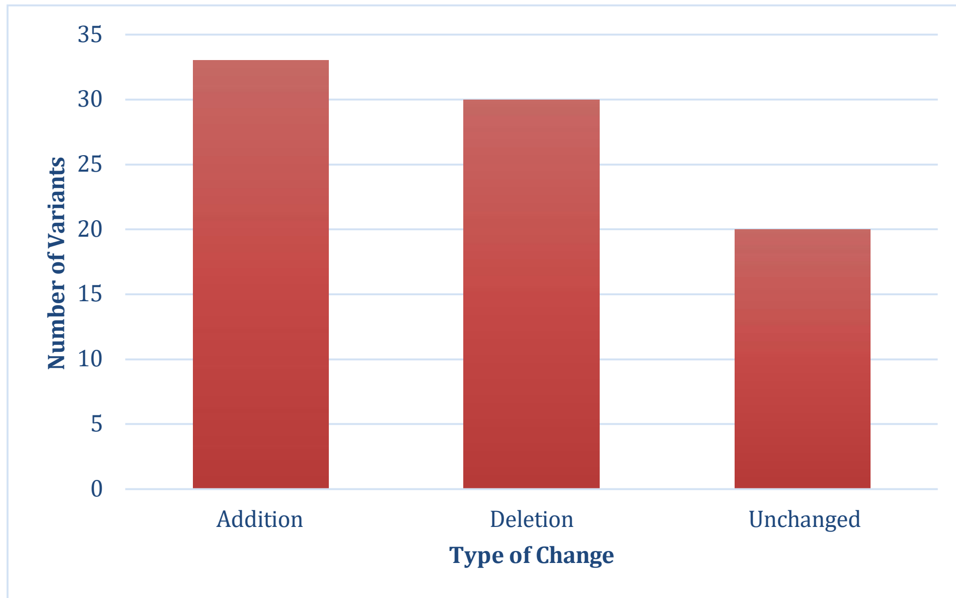


Figure 11 Changes in transcription factor binding sites in wild type sequences due to Single Nucleotide Polymorphisms.

The wild and mutated sequences of all the intronic variants were obtained through in silico software and in order to get their transcription sites a software called AliBaba2.1was used. Of all the transcription sites only the unique addition sites were filtered and their effect on the target gene was studied via a database called KEGG. Unique sites which act as an activator of the gene include HNF-3B, RSRFC4, HEB, YY1, MyoD, ALF-1B,WT1-I and WT1-I-de.while Oct-1A and WT1-del2 act as both the activator and repressor of the gene (table 4). The table given below shows the unique addition site and their effect on the target gene.

Table 4 Functional annotation of unique transcription factor binding sites.

Transcription Factor	Kegg	Pathway	Role
Oct -1A	map05417	PKC alpha pathway	Both a repressor and activator depending on the promoter.
HNF-3B (Hepatocyte Nuclear Factor 3)/ FOXA2	map04950	Differentiation. PDX1 pathway	Allows other transcription factors to bind. Binds to compacted chromatin.
RSRFC4	No entry found.	Cell growth control and apoptosis.	Activator of genes.
CREMdelt	No entry found.	-	-
HEB	No entry found.	(one of E proteins) Regulates lineage specific gene expression.	Activator. Enhancer
Zen-1	No entry found.	No info.	No info.
YY1	No entry found.	vitamin D enhancement of osteocalcin gene transcription in vivo.	Activator.
Id-1	No entry found.	No info.	No info.
Odd	No entry found.	No info.	Info.
MyoD	No entry found.	Spinocerebellar ataxia.	Activator.
ALF-1B	No entry found.	Inflammation	Gene activator
WT1-I	No entry found.	directly regulate the apoptotic pathway through induction of proapoptotic genes such as BAK	Transcriptional activator of Expression for CSF-1, amphiregulin, syndecan-1, E-cadherin, **PPP1R163** and Bcl-2
WT1-I-de	No entry found.	PI3K/Akt pathway.	activator or repressor depending on the cellular or chromosomal context
WT1-I-K	No entry found.	regulator of the Ras/MAPKinase pathway, M	Activator
WT1-del2	No entry found.	Bak/Bax-mediated apoptotic pathway	Activator and repressor

CHAPTER 5

DISCUSSION

Liver cancer is the third most common cause of deaths in 2020. HCC is caused by oxidative stress, inflammation, and underlying liver disease in hepatocytes.90% of primary liver cancers are hepatocellular carcinomas (HCC). The primary focus of our research is also HCC due to its high incidence to mortality ratio. Asia and Africa have the highest incidence rates of liver cancer in the world. Liver and intrahepatic bile duct cancer is the fifth most common malignancy in Pakistani adults according to the Collective Cancer Registry report of 2004 to 2020 by Shaukat Khanum Memorial Hospital and Research Centre. HCC is usually diagnosed after clinical deterioration, at which time survival is measured in months. Long-term survival requires detection of small tumours, which are often present in asymptomatic individuals, which may be more relevant to invasive treatment options. It has also been found that HCC tumors are resistant to chemotherapy. The low effectiveness of treatment may also be due to the fact that less than 40 percent of HCC cases are diagnosed at an early stage (Hsu & Liu, 2018). However, the identification of protein biomarkers allows the early detection and designing of effective treatment plans for the patients. The main objectives on which this research was based are also the identification of potential metabolic biomarkers by evaluating the metabolite levels in plasma samples and their validation through insilico studies.

Blood biomarkers are very important in the early detection of liver cancer and include categories primarily related to proteins, cytokines, enzymes, isozymes, and transcripts of related genes. The identification of biomarkers from patient's and healthy individual's samples has been done through GCMS and their insilico analysis by using different bioinformatics tools. It will help in the early detection and diagnosis of the disease (Kohles & Nagel, 2012).

A total of 16 protein biomarkers of our interest were filtered and studied to identify their role in cancer progression. Majority were found to be directly involved in the metastasis of cancer. There were few not directly involved but have some indirect role in disease progression. An enzyme called ADP/ATP translocase 2 (ANT 2)

located in the inner mitochondrial membrane acts as transporter of ATP from cytoplasm (synthesized through aerobic glycolysis) into the matrix of mitochondria. An enzyme is upregulated in cancer cells and transports a relatively huge number of ATPs which then support the metastatic cell survival and proliferation. Another protein Creatine kinase B-type (CKB) has also been found to upregulate in cancer cells. Cancer cells release this CKB into extracellular space where it starts catalysing the formation of phosphocreatine. SLC6A8 transporter is then used to import it into the cells, and it starts generating ATPs which act as a fuel for metastatic cells. Fatty acid synthase (FASN) is an enzyme involved mainly in the metabolism of lipids. In many cancer types the positive linkage between FASN protein levels and oxidative phosphorylation has been found. Overexpression of FASN generates large amounts of free fatty acids. In mitochondria huge amounts of energy generated through FAO lead to cancer metastasis which demands or requires relatively greater energy (Chee Wai Fhu & Ali, 2020).

A ribosomal protein RPLP0 has been found to have an association with PLAAT4 which belongs to the tumor suppressor protein family. PLAAT4 causes the downregulation of the expression of RPLPO by interacting with it. In gastric cancer cells this protein induces the downregulation of CDK2 which then causes the suppression of cell cycle progression and cell growth. Another ribosomal protein RPL12 is also involved in metastasis. During mitosis the phosphorylation of RPL12/uL11 is found to be associated with the translation of specific mRNA subsets. Hence certain post translational modifications which occur in ribosomal proteins can regulate translation (Milek & Yasuda, 2018).

Hallmarks of cancer are special abilities that cancer cells gain in the process of tumour development. They include resisting apoptosis, uncontrolled cellular growth, invasion of neighbouring or far off tissues, angiogenesis, evading host immune response, replicating indefinitely and altered metabolism (Fouad & Aanei, 2017). Caballero-Díaz et al. showed that clathrin can have anti-apoptotic effects, one of the major hallmarks in liver cancer cells (Caballero-Díaz et al., 2020). ATP1A1 promotes proliferation of Glioma stem cells and ATP1A1 overexpression also suppresses IFN-γ signalling which helps the cell evade an immune response

(Nakamura et al., 2021). FASN and CKB are involved in rewiring the energy metabolism of the cell in a way that promotes metastasis as demonstrated by Fhu et al. (Fhu & Ali, 2020) and Loo et al. (Loo et al., 2015). RPL12 regulates mitotic translation so it may play a role in promoting cell division (Imami et al., 2018). ALDH3A2 as well as IDH normally have a role in lipid metabolism but in cancer cells they also prevent ROS, leading to cell survival in the presence of mechanisms that should produce lipid toxicity (Yusuf et al., 2020).

One of the important classification of our results is the protein association with aerobic glycolysis especially Malate Dehydrogenase (MDH), Isocitrate Dehydrogenase (IDH) and Serine Hydroxymethyltransferase 2 (SHMT 2) (Zeng et al., 2021). Aerobic glycolysis or the Warburg effect happens in liver cancer cells whereby enhanced glucose uptake and lactate production occurs. Unlike normal cells that catabolize glucose by oxidative phosphorylation in the mitochondria, tumor cells tend to convert glucose into lactate even in conditions of sufficient oxygen. The production of lactate also provides an acidic environment to aid the invasion and metastasis of cancer. Cancer cells consume ATP from both the mitochondria OXPHOS and the aerobic glycolysis. In most normal cells, the ATP produced from mitochondria OXPHOS and glycolysis is approximately 90 and 10%, respectively. Whereas cancer cells rely on aerobic glycolysis to provide as much as 60% of the ATP consumption. Although the aerobic glycolysis seems to be energetically inefficient in the production of ATP, this is compensated for by the fact that the aerobic glycolysis process is more rapid and it also generates further downstream bio macromolecules required for cell proliferation (Lee, 2018).

Rapid glucose fermentation by glycolysis causes cancer cells to take up more glucose than normal cells. These findings suggest that the use of aerobic glycolysis in HCC provides advantages during cancer progression. As a new hallmark of HCC, aerobic glycolysis is believed to promote the proliferation, growth, and induce immune evasion in HCC.

There is rapid ATP production during aerobic glycolysis, enabling the tumor to adapt to its microenvironment which is short of energy resources. The enhanced aerobic metabolism is accompanied by activated glycolytic flux, with increased amounts of

metabolic intermediates production. The production of lactate and H^+, causes the acidification of the extracellular environment, which inhibits the function of immunosuppressive cells, including M2 type macrophages and lymphocytes, further facilitating the survival of cancer cells (Feng, 2020).

Metastasis and invasion of HCC are correlated with the enhanced aerobic glycolysis in HCC (Daniel G Tenen, 2021). Because of the hypoxic nature of tumor tissues, cancer cells tend to metastasize to additional sites to enhance energy and blood supplies, thereby enabling its survival. The glycolytic phenotype aids this metastasis and invasion in HCC mainly through lactate and H^+ mediated acidification of the extracellular environment.

Our study has pointed towards potential diagnostic protein biomarkers for HCC. Some of the proteins identified, such as CLTC may aid or inhibit apoptosis so their role as pro-tumorigenic proteins needs to be elucidated further. The statistical models we used account for the interplay between levels of different proteins in the sample. According to the current study, ATP1A1, CLTC, and IDH1 are prognostic markers for liver cancer with high expression being unfavourable. This points to their potential as prognostic markers than can obtained and evaluated from blood samples. DHB4, LMNA AND IDH1 can similarly be used in combination as diagnostic makers because they have a near 100 percent positivity in malignant tissues in liver cancer.

Our study demonstrates that blood samples contain detectable levels of identifiable proteins which can potentially serve as diagnostic biomarkers for HCC after further verification and analysis. The proteins studied for our work have been found to play an important role in cancer at the molecular level by previous studies which show that GC-MS can be effectively used in proteomic studies of tissue samples related to disease.

Conclusion

The current study reports that a total of 16 potential protein biomarkers were identified from the blood samples of healthy individuals and cancer patients through GCMS and several in silico tools. Most of these proteins were directly or indirectly involved in cancer progression. Some are involved in direct metastasis such as ANT2, CKB, and FASN. There are few which are anti-apoptotic and inhibit cell death such as ANT 3 and RPLPO. There are also proteins which are linked to glycolysis such as IDH, and MDH. The identification of these protein biomarkers propose that an early diagnosis of HCC is possible. Intron variant analysis of PPP1R163 gene showed that unique transcription factor binding sites were added due to SNPs, with one of the site's complementary transcription factor, WT1-I being a known activator of the gene.

Future Prospects

The current study should be replicated with a much larger sample size to improve the specificity and accuracy of the proteome analysis. The study should be replicated with different ethnic groups, different age groups, and equal number of samples from both genders. The diagnostics potential of promising targets should be investigated further with both molecular and computational studies. The diagnostic potential can also be studied for a combination of the proteins. The role of certain proteins could be clearly linked to liver cancer so in depth studies could elucidate that role.

The regulatory potential of intron variant SNPs of PPP1R163 can be validated through expression analysis studies.

REFERENCES

AiAxley, P., Ahmed, Z., Ravi, S., & Singal, A. K. (2018). Hepatitis C Virus and Hepatocellular Carcinoma: A Narrative Review. *Journal of Clinical and Translational Hepatology*, *6*(1), 79–84. https://doi.org/10.14218/JCTH.2017.00067

Belzacq, A.-S., Vieira, H. L. A., Verrier, F., Vandecasteele, G., Cohen, I., Prévost, M.-C., … Kroemer, G. (2003). Bcl-2 and Bax Modulate Adenine Nucleotide Translocase Activity1. *Cancer Research*, *63*(2), 541–546.

Borrebaeck, C. A. K. (2017). Precision diagnostics: Moving towards protein biomarker signatures of clinical utility in cancer. *Nature Reviews Cancer*, *17*(3), 199–204. https://doi.org/10.1038/nrc.2016.153

Caballero-Díaz, D., Bertran, E., Peñuelas-Haro, I., Moreno-Càceres, J., Malfettone, A., López-Luque, J., … Fabregat, I. (2020). Clathrin switches transforming growth factor-β role to pro-tumorigenic in liver cancer. *Journal of Hepatology*, *72*(1), 125–134. https://doi.org/10.1016/j.jhep.2019.09.012

Cancer Registry and Clinical Data Management (CRCDM). (2021). *COLLECTIVE CANCER REGISTRY REPORT-DECEMBER 1994 TO DECEMBER 2020: SHAUKAT KHANUM MEMORIAL CANCER HOSPITAL & RESEARCH CENTER, PAKISTAN* (p. 15). Lahore, Peshawar Karachi: SHAUKAT KHANUM MEMORIAL CANCER HOSPITAL & RESEARCH CENTER. Retrieved from SHAUKAT KHANUM MEMORIAL CANCER HOSPITAL & RESEARCH CENTER website:

https://shaukatkhanum.org.pk/wp-content/uploads/2021/09/Collective-Cancer-Registry-Report-Dec.-1994-to-Dec.-2020.pdf

Chakrabandhu, K., Huault, S., Garmy, N., Fantini, J., Stebe, E., Mailfert, S., … Hueber, A.-O. (2008). The extracellular glycosphingolipid-binding motif of Fas defines its internalization route, mode and outcome of signals upon activation by ligand. *Cell Death & Differentiation*, *15*(12), 1824–1837. https://doi.org/10.1038/cdd.2008.115

de Mattos, Â. Z., Debes, J. D., Boonstra, A., Yang, J.-D., Balderramo, D. C., Sartori, G. D. P., & de Mattos, A. A. (2021). Current impact of viral hepatitis on liver cancer development: The challenge remains. *World Journal of Gastroenterology*, *27*(24), 3556–3567. https://doi.org/10.3748/wjg.v27.i24.3556

Fhu, C. W., & Ali, A. (2020). Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules*, *25*(17), 3935. https://doi.org/10.3390/molecules25173935 Fouad, Y. A., & Aanei, C. (2017). Revisiting the hallmarks of cancer. *American Journal of Cancer Research*, *7*(5), 1016–1036.

Hecquet, C., Lefevre, G., Valtink, M., Engelmann, K., & Mascarelli, F. (2002). Activation and role of MAP kinase-dependent pathways in retinal pigment epithelial cells: ERK and RPE cell proliferation. *Investigative Ophthalmology & Visual Science*, *43*(9), 3091–3098.

Henry, N. L., & Hayes, D. F. (2012). Cancer biomarkers. *Molecular Oncology*, *6*(2), 140–146. https://doi.org/10.1016/j.molonc.2012.01.010

Imami, K., Milek, M., Bogdanow, B., Yasuda, T., Kastelic, N., Zauber, H., … Selbach, M. (2018). Phosphorylation of the Ribosomal Protein RPL12/uL11

Affects Translation during Mitosis. *Molecular Cell*, *72*(1), 84-98.e9. https://doi.org/10.1016/j.molcel.2018.08.019

International Agency for Research on Cancer. (2021). Age-standardized rate (World) per 100 000, incidence and mortality, males and females. Retrieved from Cancer Over Time website: https://gco.iarc.fr/overtime/en/dataviz/trends?sexes=1_2&sort_by=value0&cancers=7&types=0_1&years=1943_2018&multiple_populations=1&populations=75200

Jang, J.-Y., & Lee, C.-E. (2003). Mitochondrial adenine nucleotide translocator 3 is regulated by IL-4 and IFN-γ via STAT-dependent pathways. *Cellular Immunology*, *226*(1), 11–19. https://doi.org/10.1016/j.cellimm.2003.11.004

Ko, H.-K., Berk, M., Chung, Y.-M., Willard, B., Bareja, R., Rubin, M., … Sharifi, N. (2018). Loss of an Androgen-Inactivating and Isoform-Specific HSD17B4 Splice Form Enables Emergence of Castration-Resistant Prostate Cancer. *Cell Reports*, *22*(3), 809–819. https://doi.org/10.1016/j.celrep.2017.12.081

La Bella, T., Imbeaud, S., Peneau, C., Mami, I., Datta, S., Bayard, Q., … Zucman-Rossi, J. (2020). Adeno-associated virus in the liver: Natural history and consequences in tumour development. *Gut*, *69*(4), 737–747. https://doi.org/10.1136/gutjnl-2019-318281

Larribere, L., Khaled, M., Tartare-Deckert, S., Busca, R., Luciano, F., Bille, K., … Bertolotto, C. (2004). PI3K mediates protection against TRAIL-induced apoptosis in primary human melanocytes. *Cell Death & Differentiation*, *11*(10), 1084–1091. https://doi.org/10.1038/sj.cdd.4401475

Liu, L., Liu, J., Shao, D., Deng, Q., Tang, H., Liu, Z., … He, J. (2017). Comprehensive genomic profiling of lung cancer using a validated panel to explore therapeutic targets in East Asian patients. *Cancer Science*, *108*(12), 2487–2494. https://doi.org/10.1111/cas.13410

Loo, J. M., Scherl, A., Nguyen, A., Man, F. Y., Weinberg, E., Zeng, Z., … Tavazoie, S. F. (2015). Extracellular metabolic energetics can promote cancer progression. *Cell*, *160*(3), 393–406. https://doi.org/10.1016/j.cell.2014.12.018

Lu, X., Kong, L., Wang, X., Liu, W., Ma, P., & Jiang, L. (2019). 17β- hydroxysteroid dehydrogenase�4 induces liver cancer proliferation-associated genes via STAT3 activation. *Oncology Reports*. https://doi.org/10.3892/or.2019.6981

Malov, S. I., Malov, I. V., Kuvshinov, A. G., Marche, P. N., Decaens, T., Macek-Jilkova, Z., & Yushchuk, N. D. (2021). Search for Effective Serum Tumor Markers for Early Diagnosis of Hepatocellular Carcinoma Associated with Hepatitis C. *Sovremennye Tehnologii v Medicine*, *13*(1), 27. https://doi.org/10.17691/stm2021.13.1.03

Nakamura, K., Shiozaki, A., Kosuga, T., Shimizu, H., Kudou, M., Ohashi, T., … Otsuji, E. (2021). The expression of the alpha1 subunit of Na+/K+-ATPase is related to tumor development and clinical outcomes in gastric cancer. *Gastric Cancer*, *24*(6), 1278–1292. https://doi.org/10.1007/s10120-021-01212-6

PANTHER. (n.d.). About PANTHER [Tool]. Retrieved May 8, 2022, from PANTHER classification system website: http://www.pantherdb.org/about.jsp

Parikh, N. D., Mehta, A. S., Singal, A. G., Block, T., Marrero, J. A., & Lok, A. S. (2020). Biomarkers for the Early Detection of Hepatocellular Carcinoma. *Cancer Epidemiology Biomarkers & Prevention*, *29*(12), 2495–2503. https://doi.org/10.1158/1055-9965.EPI-20-0005

Pervez, S., Jabbar, A. A., Haider, G., Ashraf, S., Qureshi, M., Lateef, F., … Qureshi, H. (2020). Karachi Cancer Registry (KCR): Age-Standardized Incidence Rate by Age-Group and Gender in a Mega City of Pakistan. *Asian Pacific Journal of Cancer Prevention*, *21*(11), 3251–3258. https://doi.org/10.31557/APJCP.2020.21.11.3251

Rashid, H., Alam, L., Farooq, M., & Ali, M. (2021). DEMOGRAPHICS AND CORRELATION OF RISK FACTORS FOR HEPATOCELLULAR CARCINOMA IN PATIENTS PRESENTING TO A TERTIARY CARE FACILITY IN PAKISTAN. *PAFMJ*, *71*(2), 614–618. https://doi.org/10.51253/pafmj.v71i2.2724

Ringehan, M., McKeating, J. A., & Protzer, U. (2017). Viral hepatitis and liver cancer. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1732), 20160274. https://doi.org/10.1098/rstb.2016.0274

Samant, H., Amiri, H. S., & Zibari, G. B. (2021). Addressing the worldwide hepatocellular carcinoma: Epidemiology, prevention and management. *Journal of Gastrointestinal Oncology*, *12*(Suppl 2), S361–S373. https://doi.org/10.21037/jgo.2020.02.08

Srivastava, S., & Gopal-Srivastava, R. (2002). Biomarkers in Cancer Screening: A Public Health Perspective. *The Journal of Nutrition*, *132*(8), 2471S-2475S. https://doi.org/10.1093/jn/132.8.2471S

Wai Fhu , C., & Ali , A. (2020). Fatty Acid Synthase: An Emerging Target in Cancer. *Molecules*, 1420-3049.

Chevrollier, A. (2011). Adenine nucleotide translocase 2 is a key mitochondrial protein in cancer metabolism. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 562-567.

Imami, K., & Milek, M. (2018). Phosphorylation of the Ribosomal Protein RPL12/uL11 Affects Translation during Mitosis. *Molecular Cell*, 84-98.

Loo, J. M., & Alexis Scherl. (2015). Extracellular Metabolic Energetics Can Promote Cancer Progression. *Cell*, 393-406.

Subhan Butt, A., Abbas, Z., & Jafri, W. (2012). Hepatocellular Carcinoma in Pakistan: Where do We Stand? *Hepatitis Monthly*, *12*(10 HCC). https://doi.org/10.5812/hepatmon.6023

Tu, T., Bühler, S., & Bartenschlager, R. (2017). Chronic viral hepatitis and its association with liver cancer. *Biological Chemistry*, *398*(8), 817–837. https://doi.org/10.1515/hsz-2017-0118

Tzartzeva, K., Obi, J., Rich, N. E., Parikh, N. D., Marrero, J. A., Yopp, A., … Singal, A. G. (2018). Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients With Cirrhosis: A Meta-

analysis. *Gastroenterology*, *154*(6), 1706-1718.e1. https://doi.org/10.1053/j.gastro.2018.01.064

What is Cytoscape? (n.d.). Retrieved May 8, 2022, from Cytoscape.org website: https://cytoscape.org/what_is_cytoscape.html

World Health Organisation. (2021a). *Accelerating access to hepatitis diagnostics and treatment*. World Health Organization. Retrieved from World Health Organization website: https://www.jstor.org/stable/resrep30124.7

World Health Organisation. (2021b). Hepatitis C. Retrieved from WHO World Health Organisation website: https://www.who.int/news-room/fact-sheets/detail/hepatitis-c

World Health Organisation. (2022, February 3). Cancer fact sheet. Retrieved March 28, 2022, from Www.who.int website: https://www.who.int/news-room/fact-sheets/detail/cancer

World Health Organisation. (n.d.). Promoting Cancer Early Diagnosis. Retrieved April 29, 2022, from World Health Organisation website: https://www.who.int/activities/promoting-cancer-early-diagnosis

World Health Organization. (2021). *Global progress report on HIV, viral hepatitis and sexually transmitted infections, 2021: Accountability for the global health sector strategies 2016–2021: actions for impact*. Geneva: World Health Organization. Retrieved from World Health Organization website: https://apps.who.int/iris/handle/10665/341412

Yates, J. R. (2019). Recent technical advances in proteomics. *F1000Research*, *8*, F1000 Faculty Rev-351. https://doi.org/10.12688/f1000research.16987.1

Yu, Y., Chen, C., Huo, G., Deng, J., Zhao, H., Xu, R., … Wang, S. (2019). ATP1A1 Integrates AKT and ERK Signaling via Potential Interaction With Src to Promote Growth and Survival in Glioma Stem Cells. *Frontiers in Oncology*, *9*, 320. https://doi.org/10.3389/fonc.2019.00320

Yusuf, R. Z., Saez, B., Sharda, A., van Gastel, N., Yu, V. W. C., Baryawno, N., … Scadden, D. T. (2020). Aldehyde dehydrogenase 3a2 protects AML cells from oxidative death and the synthetic lethality of ferroptosis inducers. *Blood*, *136*(11), 1303–1316. https://doi.org/10.1182/blood.2019001808

The UniProt Consortium, Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Agivetova, R., … Teodoro, D. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., … Ponten, F. (2017). A pathology atlas of the human cancer transcriptome. *Science (New York, N.Y.)*, *357*(6352), eaan2507. https://doi.org/10.1126/science.aan2507

Zeng, Y., Zhang, J., Xu, M., Chen, F., Zi, R., Yue, J., … Chin, Y. E. (2021). Roles of Mitochondrial Serine Hydroxymethyltransferase 2 (SHMT2) in Human Carcinogenesis. *Journal of Cancer*, *12*(19), 5888–5894. https://doi.org/10.7150/jca.60170

Borger, D. R. (2012). Frequent Mutation of Isocitrate Dehydrogenase (IDH)1 and IDH2 in Cholangiocarcinoma Identified Through Broad-Based Tumor Genotyping. *The oncologist,*, 72-79.

Daniel G Tenen, L. C. (2021). Metabolic alterations and vulnerabilities in hepatocellular carcinoma. *Gastroenterology Report*.

Feng, J. (2020). Emerging roles and the regulation of aerobic glycolysis in hepatocellular carcinoma. *J Exp Clin Cancer*.

Lee, N. C. (2018). High Expression of Glycolytic Genes in Cirrhosis Correlates With the Risk of Developing Liver Cancer. *Frontiers in cell and developmental biology*.

Liu, H. (2020). LMNA functions as an oncogene in hepatocellular carcinoma by regulating the proliferation and migration ability. *Journal of cellular and molecular medicine* , 12008-12019.

Sherman, M. (2014). Limitations of screening for hepatocellular carcinoma. *Hepatic Oncology*.

Siavash Mansouri, A. S. (2017). Role of malate dehydrogenase in facilitating lactate dehydrogenase to support the glycolysis pathway in tumors. *Biomedical Reports*, 463-467.

Zhou, Z.-J. (2013). Overexpression of HnRNP A1 promotes tumor invasion through regulating CD44v6 and indicates poor prognosis for hepatocellular carcinoma. *Int J Cancer*, 1080-9.

APPENDIX

<u>Patient Consent Form</u>

Pathogenicity of Proteins in Liver Cancer

Participant ID: _________________

Participants Name: ______________

Study Purpose:

Study aims to identify genetic variation and expression of protein kinases in liver cancer in Pakistani Population. Study commencement will help in pre-diagnosis and prognosis of the disease. Information collected in the study will be kept confidential and blood sample collected in the study will solemnly be used for research purpose.

PI: ____________________

For Donors:

I voluntarily agree to take part in this research project. I have been informed how and where my sample will be used. My blood sample may be collected and used in this study as defined in this consent form. Aim of this study has been conveyed to me in my native language too.

Donor's Signature: __________________ *Date:* ___________________

Patient History Form

Patient details

Name: ________________________________

Age: _____________________________

Gender: Male / Female

Marital status: Single/Married

Ethnic background ___________________

Education: __________________________

Cancer detail chart on the following page:

Please select from chart below:

Cancer localization				
Cancer type	Primary		Secondary	
For primary liver cancer				
Type	Hepatocellular carcinoma Intrahepatic cholangiocarcinoma			
Cancer metastasis	Yes		No	
Cancer stage	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.				

Participants: _______________

Principle Investigator: ___________________