# *YAP 1* and its Mediators as Key Players in Pancreatic Ductal Adenocarcinoma (PDAC): An Investigation through Whole-Exome Sequencing.




**By**

**Arooba Azeem**

Department of Biochemistry

Faculty of Biological Sciences,

Quaid-i-Azam University

Islamabad, Pakistan

2022

# *YAP 1* and its Mediators as Key Players in Pancreatic Ductal Adenocarcinoma (PDAC): An Investigation through Whole-Exome Sequencing.


A thesis submitted in partial fulfillment of the requirements for the

**Degree of Master of Philosophy**

**In**

**Biochemistry**

**By**

**Arooba Azeem**

Department of Biochemistry

Faculty of Biological Sciences

Quaid-i-Azam University

Islamabad, Pakistan

2020-2022

# CERTIFICATE

This thesis, submitted by **Ms. Arooba Azeem** to the Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan, is accepted in its present form as satisfying the thesis requirement for the Degree of Master of Philosophy in Biochemistry/Molecular Biology.

Supervisor:

Dr. Aneesa Sultan
Associate Professor

External Examinar:

Dr. Ishrat Mahjabeen
Assistant Professor
Department of Biosciences
COMSATS University, Islamabad.

Chairperson:

Dr. Iram Murtaza
Professor

Dated: November 1, 2022

## Declaration

I, Arooba Azeem, hereby declare that all the material and information present in this thesis is my original research work. It has been submitted only for the degree of Master of Philosophy in Biochemistry. I have not previously presented any part of this work elsewhere for any other degree.

**Arooba Azeem**

*DEDICATION TO*

*My Caring Parents, Loving Siblings*

*And*

*Untiringly Supportive Supervisor*

## Acknowledgments

First and foremost, I am grateful to **Allah Almighty**, the most beneficent and the merciful, for giving me strength, patience, and good health to complete my research work. Countless salutations upon our **Holy Prophet Hazrat Muhammad (PBUH)**, for showing the right path of life to mankind.

It's a pleasure to show my gratitude to my respected supervisor and a great mentor "**Dr. Aneesa Sultan**" for providing me the opportunity to do the research work and for her guidance, motivation, reassurance, timely provision of chemicals, and valuable suggestions throughout the research. I am very thankful to the Chairperson, Department of Biochemistry "**Dr. Iram Murtaza**" for providing better research facilities for the accomplishment of our research project. I am also very thankful to "**Dr. Tahir Waheed**" for teaching us research planning and report writing course as it helped me a lot in writing my thesis.

I pay my deepest gratitude to my lab seniors "**Maryam Naeem**", "**Usman Tareen**", "**Asim Hassan**", and "**Maryam Jahan**" for their kind, friendly and supportive attitude toward me. I am thankful for their guidance, support, and valuable suggestions throughout the research period which helped me to learn a lot of new things. I am grateful to my friends "**Iqra Bibi**", "**Aroma Tahir**", and "**Amna Sajid**" and lab fellows "**Anis-ur-Rehman**" and who supported me and was always available for me in every thick and thin of life. I am thankful to the lab assistant, **"Mr. Saeed**", and all the clerical staff of the Department of Biochemistry for their help.

Last but not the least, I would like to pay my deepest gratitude to my parents "**Azhar Mehmood**", "**Nighat Aziz**" and siblings "**Engr. Ashar Azeem**", "**Atiqa Azhar**" for their unconditional love, moral support, efforts, prayers, and the trust they have in me. At last, I am thankful to every person who helped me during this whole tenure.

**Arooba Azeem**

# Table of Contents

**List of Figures**.....vi

**List of Tables**..... viii

**List of Abbreviations**.....ix

**Abstract**..... xiii

**1 Introduction**.....1

**1.1 Cancer:**.....1

**1.1.1 Hallmarks of Cancer:**.....1

**1.2 Pancreatic Cancer (PC):**.....2

**1.3 Types of Pancreatic Cancer:**.....3

**1.3.1 Exocrine Pancreatic Cancers:**.....4

**1.3.2 Endocrine Pancreatic Cancer:**.....5

**1.3.3 Other Rare Types of Pancreatic Cancer:**.....6

**1.4 Molecular Subtypes of Pancreatic Cancer:**.....6

**1.4.1 Squamous Subtype:**.....7

**1.4.2 Pancreatic Progenitor Subtype:**.....7

**1.4.3 Immunogenic Subtype:**.....7

**1.4.4 ADEX Subtype:**.....8

**1.5 Signs and Symptoms of Pancreatic Cancer:**.....8

**1.5.1 Upper Abdominal and Back Pain:**.....10

**1.5.2 Jaundice**.....10

**1.5.3 Weight Loss:**.....10

**1.5.4 Gastrointestinal (GI) Problems:**.....11

**1.5.5 Diabetes:**.....11

**1.5.6 Others:**.....11

**1.6 Etiological Factors of Pancreatic Cancer:**.....11

**1.6.1 Age:** ..... 11

**1.6.2 Gender:** ..... 11

**1.6.3 Area:** ..... 11

**1.6.4 Alcohol:** ..... 12

**1.6.5 Diabetes:** ..... 12

**1.6.6 Smoking:** ..... 12

**1.6.7 Family History:** ..... 12

**1.6.8 Genetic Susceptibility:** ..... 12

**1.6.9 Chronic Pancreatitis (CP):** ..... 13

**1.6.10 Blood Group:** ..... 13

**1.6.11 Obesity:** ..... 13

**1.6.12 Others:** ..... 13

**1.7 Pancreatic Cancer Stages:** ..... 13

**1.7.1 TNM Staging of Pancreatic Cancer:** ..... 14

**1.8 Progression of PDAC Carcinogenesis:** ..... 15

**1.9 Incidence and Mortality Rate of PC Worldwide:** ..... 17

**1.10 Incidence and Mortality Rate of Pancreatic Cancer in Pakistan** ..... 18

**1.11 Diagnosis of Pancreatic cancer:** ..... 19

**1.11.1 Computerized Tomography (CT) Scan:** ..... 19

**1.11.2 Positron Emission Tomography (PET) Scan:** ..... 18

**1.11.3 Magnetic Resonance Imaging (MRI):** ..... 20

**1.11.4 Endoscopic Ultrasonography (EUS):** ..... 20

**1.11.5 Endoscopic Retrograde Cholangiopancreatography (ERCP):** ..... 20

**1.11.6 Histopathological Analysis:** ..... 20

**1.11.7 Blood Tests and Tumor Biomarkers:** ..... 219

**1.12 Treatment and Therapeutics for Pancreatic Cancer:** ..... 219

**1.12.1 Surgical Resection:** ..... 219

**1.12.2 Chemotherapy:** ..... 219

**1.12.3 Radiography:** ..... 22

**1.12.4 Immunotherapy:** ..... 22

**1.12.5 Other Treatment Strategies:** ..... 22

**1.13 Altered Signaling Pathways in Pancreatic Cancer:** ..... 22

**1.14 Role of the driver and various other genes in PDAC:** ..... 23

**1.15 *p53* gene and its normal function:** ..... 24

**1.16 Polymorphism of *p53*:** ..... 26

**1.17 *p53* role in PDAC:** ..... 27

**1.18 *YAP1* gene and its normal function:** ..... 27

**1.19 Polymorphism of *YAP1*:** ..... 29

**1.20 *YAP1* role in PDAC:** ..... 30

**1.21 Cross-linking role of *YAP1* and *p53* in PDAC:** ..... 31

**1.22 Significance of Study:** ..... 32

**1.23 Aims and Objectives:** ..... 33

**2 Materials and Methods** ..... 34

**2.1 Study Plan and Ethical Approval:** ..... 34

**2.2 Tissue and Blood Sample Collection:** ..... 34

**2.3 Extraction of Genomic DNA from Tissue Samples:** ..... 35

**2.3.1 Preparation of Solutions for gDNA Extraction:** ..... 35

**2.3.2 Procedure:** ..... 35

**2.4 Genomic DNA Quantification:** ..... 37

**2.5 Genomic DNA Qualitative Analysis:** ..... 37

**2.5.1 Agarose Gel Electrophoresis Procedure:** ..... 38

**2.6 Whole Exome Sequencing (WES):** ..... 38

**2.7 Bioinformatics Analysis:** ..... 39

**2.7.1 Raw Data Input:** ..... 40

**2.7.2 Quality checking of Samples:** ..... 40

**2.7.3 Removal of Adaptors:** ..... 40

**2.7.4 Mapping against Reference Genome:** ..... 40

**2.7.5 Marking the Duplicates:** ..... 40

**2.7.6 Variants Calling by MuTect2:** ..... 39

**2.7.7 Variants Annotation:** ..... 41

**2.8 AvSNP 150:** ..... 39

**2.9 Damage Pred Count:** ..... 42

**2.10 SIFT, PROVEAN, and PolyPhen2:** ..... 42

**2.11 MutationTaster, Likelihood Ratio Test (LRT), and Mutation Assessor:** ..... 43

**2.12 Variant Effect Scoring Tool (VEST4):** ..... 43

**2.13 CADD Raw, CADD Phred:** ..... 43

**2.14 Cosmic 92:** ..... 43

**2.15 Clinical Significance (CLINSIG):** ..... 44

**2.16 Phyre2:** ..... 44

**3 Results** ..... 45

**3.1 Study Design and Demographic Detail:** ..... 45

**3.1.1 Details of PDAC Patients:** ..... 45

**3.1.2 Graphical Representation of Stages of PDAC Patients:** ..... 46

**3.2 Quantitative Analysis of genomic DNA:** ..... 47

**3.3 Qualitative Analysis of Genomic DNA:** ..... 47

**3.4 WES Result Analysis:** ..... 48

**3.4.1 Quality Checking of Samples Raw Reads:** ..... 48

**3.4.2 Adapter Content Analysis:** ..... 49

**3.4.3 Detection of PDAC Variants:** ..... 50

**3.5 Annotated Variants Data Analysis:** ..... 50

**3.5.1 Nucleotide Variations in *p53* gene:** ..... 51

**3.5.2 Nucleotide Variations in *p53TG1* gene:** ..... 52

**3.5.3 Nucleotide Variations in *YAP1* gene:** ..... 53

**3.5.4 Nucleotide Variations in Driver Genes:** ..... 54

**3.5.5 Variant Probing by AvSNP150 Database:** ..... 54

**3.5.6 Variants Damage Potential Analysis by ClinPred:** ..... 54

**3.5.7 Amino Acid Substitution Analysis of SNVs using SIFT, PROVEAN, and PolyPhen2:** ..... 55

**3.5.8 Damaging and Tolerant Variant Analysis using MutationTaster, LRT, Mutation Assessor:** ..... 55

**3.5.9 Pathogenic Effect of Variants on Protein Activity by VEST4:** ..... 55

**3.5.10 Deleteriousness of Variants using CADD raw and CADD Phred:** ..... 56

**3.5.11 Somatic Mutations Data Analysis by Cosmic 92:** ..... 56

**3.5.12 Role of Variants in other Diseases using CLINSIG:** ..... 56

**3.5.13 Protein Homology Modeling using Phyre2:** ..... 57

**4 Discussion** ..... 58

**5 Conclusion** ..... 62

**6 References:** ..... 63

# List of Figures

**Figure 1.1:** Cancer Hallmarks……...2

**Figure 1.2:** Anatomy of Pancreatic Cancer……..2

**Figure 1.3:** Obstruction of ducts by adenocarcinoma (PDAC)……....4

**Figure 1.4:** Acinar cell carcinoma…….....5

**Figure 1.5:** Pancreatic Neuroendocrine Tumor (PanNETs)…….6

**Figure 1.6:** Pancreatic cancer Stages……...13

**Figure 1.7:** PanINs in ductal epithelium of pancreas leading to PDAC…….14

**Figure 1.8:** Driver genes involved in the progression of pancreatic carcinogenesis…….14

**Figure 1.9:** Research publication regarding pancreatic cancer from different regions of Pakistan……17

**Figure 1.10:** Altered signaling pathways in PC……..21

**Figure 1.11:** Genomic location of *p53* on chromosome 17…….23

**Figure 1.12:** p53 protein domains……...23

**Figure 1.13:** Genomic location of YAP on chromosome 11……...26

**Figure 1.14:** YAP1, 2 (YAP, TAZ) protein domains……..26

**Figure 1.15:** Regulation of YAP1 protein in Hippo pathway…….27

**Figure 1.16:** PDAC development by activation of *YAP1* through KRAS-dependent pathway……28

**Figure 1.17:** Mutant *p53* and *YAP1* interaction for activating pro- oncogenes…….30

**Figure 1.18:** Inhibition of *YAP1* by increased expression of PTPN14 via stabilized *p53*……30

**Figure 2.1:** Flow Chart of Data Analysis using Bioinformatics Tools**……37**

**Figure 3.1:** PDAC Patients Details……**……..43**

**Figure 3.2:** Cancer Stages Frequency in PDAC Patients**……44**

**Figure 3.3:** 1% Agarose Gel Stained by EtBr Showing Genomic DNA Bands from Tissue Samples**…….46**

**Figure 3.4:** Basic Statistics of Raw Reads produced by Base Calling using fastQ**…46**

**Figure 3.5:** Quality Score Graph Represented in FastQC Result Report**……47**

**Figure 3.6:** Adapter Content in FastQC Result Report**……...47**

**Figure 3.7:** Location of exonic variant of *p53* gene**……49**

**Figure 3.8:** Nucleotide Variation in Exon 6 of *p53* gene on Chromosome 17**……49**

**Figure 3.9:** Location of exonic variant on p53 protein domain**…….…..49**

**Figure 3.10:** Various Single nucleotide variations (SNVs) in the *p53* gene**……..….50**

**Figure 3.11:** Location of 3 non-coding exonic variants of *p53TG1* gene**….....50**

**Figure 3.12:** Nucleotide Variation in Exon 1, 2 of *p53TG1* gene**……...…51**

**Figure 3.13:** Various Single nucleotide variations (SNVs) in the *YAP1* gene**……...51**

**Figure 3.14:** Nucleotide Variation in Exon 5 of *KRAS* gene**……..……52**

**Figure 3.15:** Nucleotide Variation in Exon 6 of *SMAD4* gene**……...52**

**Figure 3.16:** 3D structure of p53 protein by Phyre2. **A:** Truncated p53 protein structure **B:** Normal p53 protein DBD structure**……..55**

**Figure 3.17:** Superimposition of truncated p53 protein region on normal p53 DBD (**C**), Chain B of normal p53 protein DBD having sequence homology with truncated p53 protein (D)**……55**

# List of Tables

**Table 1.1:** Incidence and Mortality rate of PC among the different populations……**16**

**Table 1.2:** Association of *p53* gene SNPs with Pancreatic Cancer risk……..**24**

**Table 1.3:** Role of SNPs of *YAP1* gene in various diseases……....**28**

**Table 2.1:** Reagents List and their Chemical Composition……..…**33**

**Table 2.2:** Chemical Composition of Reagent used in Agarose Gel Electrophoresis……**35**

**Table 2.3:** Reference genomes of all organisms used for alignment…….....**38**

**Table 3.1:** PDAC Patients Details used for WES……..**44**

**Table 3.2:** Optical Density (O.D) value and Concentration in ng/ul of Nucleic Acid……..**45**

**Table 3.3:** No. of variants identified in each PDAC sample……...**48**

**Table 3.4:** Pathogenic effect of Variants on Protein Activity……**54**

# List of Abbreviations

| | |
|---|---|
| ACC | Acinar Cell Carcinoma |
| ACS | American Cancer Society |
| AD2 | Activation Domain 2 |
| ADM | Acinar to Ductal Metaplasia |
| ADEX | Aberrantly Differentiated Endocrine Exocrine |
| AJCC | American Joint Committee on Cancer |
| ASCO | American Society of Cancer Oncology |
| BWA | Burrows-Wheeler Aligner |
| CA-19 | Cancer Antigen 19 |
| CADD | Combined Annotation Dependent Depletion |
| CBC | Complete Blood Count |
| CC | Coiled-Coil Domain |
| CEA | Carcinoembryonic Antigen |
| ClinSig | Clinical Significance |
| Cosmic92 | Catalogue of Somatic Mutations in Cancer Database |
| CP | Chronic Pancreatitis |
| CT-Scan | Computed Tomography Scan |
| CVD | Cardiovascular Diseases |
| DBD | DNA Binding Domain |
| DM | Diabetes Mellitus |
| EDTA | Ethylenediamine Tetra Acetic Acid |
| EMT | Epithelial to Mesenchymal Transition |

| | |
|---|---|
| ERCP | Endoscopic Retrograde Cholangiopancreatography |
| EtBr | Ethidium Bromide |
| EUS | Endoscopic Ultrasound |
| FNA | Fine Needle Aspiration |
| FU | 5-Fluorouracil |
| GATK | Genome Analysis Toolkit |
| GEM | Gemcitabine |
| GPCRs | G-Protein Coupled Receptor |
| HbA1c | Glycated Hemoglobin |
| HSP90 | Heat Shock Protein 90 |
| IFNG | Interferon Gamma |
| IGRT | Image Guided Radiotherapy |
| IMRT | Intensity Modulated Radiography |
| IPMN | Intraductal Papillary Mucinous Neoplasm |
| IRB | International Review Board |
| LFT | Liver Function Test |
| LncRNA | Long Non Coding RNA |
| MCN | Mucinous Cystic Neoplasm |
| MRCP | Magnetic Resonance Cholangiopancreatography |
| MRI | Magnetic Resonance Imaging |
| NGS | Next Generation Sequencing |
| OD | Optical Density |
| PanINs | Pancreatic Intraepithelial Neoplasia |

| | |
|---|---|
| PanNETs | Pancreatic Neuroendocrine Tumor |
| PC | Pancreatic Cancer |
| PDAC | Pancreatic Ductal Adenocarcinoma |
| PDGFRβ | Platelet Derived Growth Factor Receptor β |
| PET | Positron Emission Tomography |
| PolyPhen | Polymorphism Phenotyping |
| PONs | Panel of Normals |
| PROVEAN | Protein Variation Effect Analyzer |
| *P53TG1* | *P53* Target Gene 1 |
| RFT | Renal Function Test |
| SBRT | Stereotactic Body Radiotherapy |
| SCC | Squamous Cell Carcinoma |
| SDS | Sodium Dodecyl Sulphate |
| SIFT | Scale Invariant Feature Transform |
| SNP | Single Nucleotide Polymorphism |
| SNV | Single Nucleotide Variation |
| TAD | Transcription Activation Domain |
| TAE Buffer | Tris Acetic Acid Buffer |
| TCGA | The Cancer Genome Atlas |
| TD | Tetramerization Domain |
| TID | TEAD Interaction Domain |
| TLR | Toll Like Receptor |
| *TP53* | Tumor Protein 53 |

| | |
|---|---|
| TRP53 | Transformation Related Protein 53 |
| UICC | Union for International Cancer Control |
| VEST4 | Variant Effect Scoring Tool 4 |
| WES | Whole Exome Sequencing |
| YAP-1 | Yes-Associated Protein 1 |

## Abstract

Pancreatic cancer (PC) is a highly pernicious malignancy with a very poor survival rate of >5 years in 10% of its cases and is reported to be 7$^{th}$ leading cause of death worldwide. It initiates in pancreatic ducts but in later stages metastasize to adjacent tissues. PDAC is the most prevalent type of PC involve in 90% of its cases. In progression of PDAC by PanIN formation, mutation of 4 driver genes including *KRAS*, *CDKN2A*, *p53*, and *SMAD4* plays a significant role. Along with these driver genes, other genes are also altered in PDAC including *YAP1*, *ZEB1*, and *BRCA 1/2*. The main objective of this research is the genetic characterization of PDAC sample with aim to analyze the role of *YAP1* and *p53* in PDAC cases. Molecular profiling of both tumor and control samples was done through WES which provided data regarding variants. Later on, using various bioinformatics tools sequencing data was observed that shown 3 novel nc-RNA variants (87341661; T>C, 87341667; A>T, 87345180; G>C) of *p53TG1* and 1 exonic reported variant of *p53* gene. While, no variation have been observed in the exonic region of *YAP1* gene. However, several SNPs and indels were reported in the intronic region of both genes. Furthermore, ClinPred, LRT, MutationTaster, and Mutation Assessor tools confirmed the novelty of 3 nc-RNA variants and damaging effect of 1 exonic variant of *p53* gene. No results were obtained for these 3 nc-RNA variants from VEST4, SIFT, PROVEAN, and PolyPhen2 which further validate their novelty. No data regarding these 3 variants were available in the AvSNP150, Cosmic92, and ClinSig database which confirms the uniqueness of these variants but SNP ID (rs1137282) and Cosmic ID (ID=COSV52663748) was assigned to 1 exonic variant of *p53* gene. These variations might have role in regulating their own genetic expression along with other genes causing PanINs formation through ADM leading to PDAC progression. In future, to validate this study transcriptomic analysis and molecular docking can be done that will further pave way to develop diagnostic marker that will help in early detection and treatment of PDAC.

**Keywords:** Pancreatic Ductal Adenocarcinoma (PDAC), Yes-Associated Protein 1 (YAP-1), Pancreatic Intraepithelial Neoplasia (PanIN), Whole Exome Sequencing (WES), long non-coding RNA (lnc-RNA), *p53* Target Gene 1 (*p53TG1*).

# 1 Introduction

## 1.1 Cancer:

Cancer is considered as the 2nd leading death cause after cardiovascular disorders worldwide (Nagai and Kim, 2017). In the United States (U.S), almost 1/2th of men and 1/3rd of women experience cancer during their life. Cancer was first discovered in Egypt, back in 3000BC, and was reported as an untreatable disease in Edwin Smith Papyrus (A textbook on Trauma Surgery). Later on, the term "Cancer" was devised from the Greek word "Karkinos" which means "tumor or crab" by Hippocrates (American Cancer Society, 2018). In the 14th century, the Latin word "cancer" was introduced for the tumor in Modern English (Fleming *et al*., 2019). Cancer develops due to uncontrolled growth of normal cells in a specific region of the body and forms new abnormal cells which lead to the death of the individual by metastasis. If these abnormal cells will remain at their place of origin then the tumor is benign while some tumor cells travel to distal regions of the body through blood circulation then the cancer is malignant. Almost 14 million people would be affected with a mortality rate of approximately 9 million each year (Chen *et al*., 2018).

### 1.1.1 Hallmarks of Cancer:

The Cancer Hallmarks comprises a set of functional capabilities adapted by human normal cells to convert them to neoplastic cells, more specifically to develop malignant tumors. The initial six hallmarks required by normal cells for acquiring tumor include sustainable proliferative signaling, evading growth-suppressors, enabling replicative immortality, resist cell death, acquiring vasculature, and initiating invasion and metastasis (Weinberg and Hanahan, 2000). Later in 2011, two additional emerging (Avoiding immune destruction, Deregulating cellular metabolism) and two more enabling (Tumor-promoting inflammation, Genome instability, and mutation) hallmarks were introduced (Hanahan and Weinberg, 2011). Recently, additional emerging and enabling characteristics were introduced as the core hallmarks of cancer including unlocking phenotypic plasticity, senescent cells, non-mutational epigenetic reprogramming, and polymorphic microbiomes (Hanahan, 2022) in Figure 1.1. These emerging and enabling characteristics are important in carcinogenesis but still more research is required to establish their link with the other hallmarks of cancer.

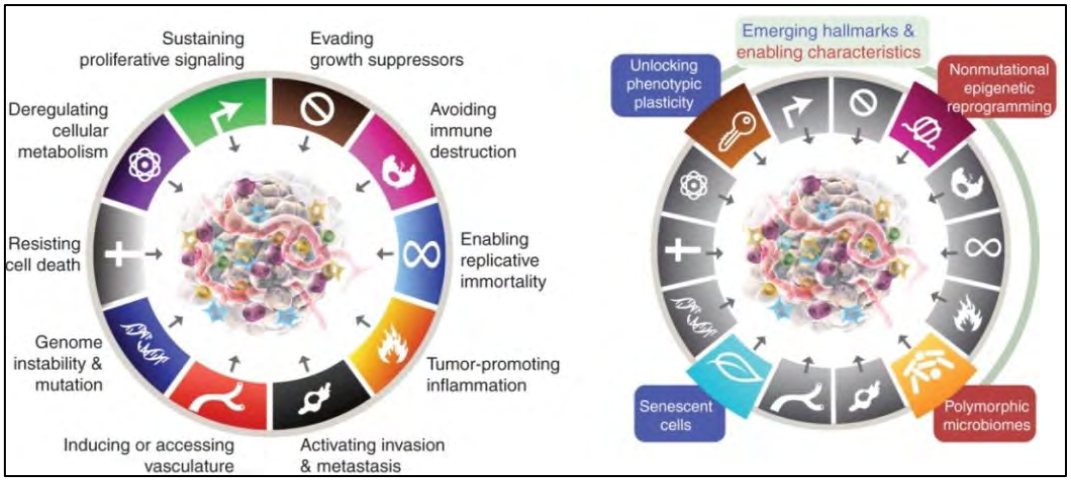


**Figure 1.1:** Cancer Hallmarks (Hanahan, 2022).

## 1.2 Pancreatic Cancer (PC):

Pancreatic cancer (PC) is considered one of the lethal malignancies that originate when normal pancreatic cells start to divide due to DNA mutations and form a tumor mass that can metastasize to various other regions of the body. Due to advanced stage detection, most of the treatments are not effective and the survival rate is also less (Ansari *et al*., 2016). Despite various types of PC, the most common type responsible for 90% of cases is pancreatic ductal adenocarcinoma (PDAC) (Pishvaian and Brody, 2017). Usually, adenocarcinomas start from the pancreatic region which produces digestive enzymes and several other non-adenocarcinomas originate from these cells. The pancreatic hormone-producing cells give rise to pancreatic neuroendocrine tumors (PanNETs) that account for 1–2% of cases of PC and are less aggressive than pancreatic adenocarcinoma.



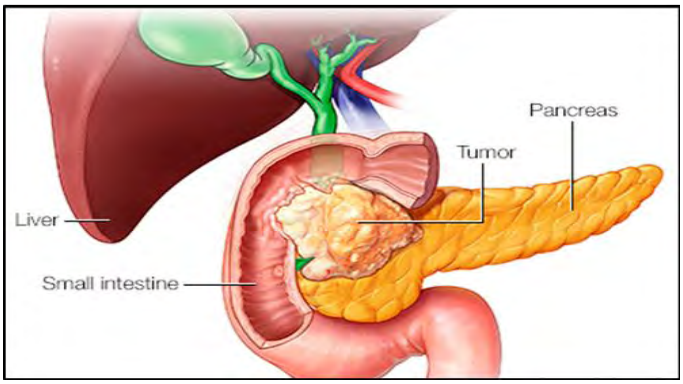


**Figure 1.2:** Anatomy of Pancreatic Cancer (Adapted from Web).

The term pancreatic cancer (PC) was first described by an Italian scientist, G.B Morgagni in 1761 but the microscopic diagnosis of ductal adenocarcinoma was undetermined. Later in 1858, the microscopic diagnosis of adenocarcinoma was explained by an American clinician, J.M Da Costa, and in the 20$^{th}$ century beginning pancreatic cancer that originates from the head was diagnosed (Ansari *et al*., 2016). The PanNETs were first discovered in 1888 and in 1982 Japanese researchers determined the intraductal papillary mucinous neoplasm (IPMN).

In 1898, the first partial pancreaticoduodenectomy was performed by A. Codivilla but unfortunately, the patient didn't survive due to some complications. After many unsuccessful attempts finally in 1935, A.O Whipple did pancreaticoduodenectomy which involve the removal of the pancreas head and the duodenum, and thus gain the interest of many surgeons in pancreatic surgery. However, for PDAC in 1937, the first successful pancreaticoduodenectomy was done by Alexander Brunschwig. Nowadays this procedure of pancreatic surgery is the safest way with a 3% of operative mortality rate (Ansari *et al*., 2014; Yoshioka *et al*., 2014). Although much advancement was made in this procedure still the survival rate of pancreatic cancer patients can't be enhanced.



### 1.3 Types of Pancreatic Cancer:

Based on pancreatic cell type there are various types of pancreatic cancer (PC). The onset of PC can be from the head, body, or tail but almost 60-70% of PC starts in the head of the pancreas. PC is further divided into two main categories:

- Exocrine pancreatic cancer
- Endocrine pancreatic cancer

Almost 95% of the PC arises in the exocrine tissue that makes exocrine glands and ducts (produces digestive enzymes), while only 1-2% of PC is reported in the endocrine tissue of the pancreas (produces hormones) (Zhang *et al*., 2016). Both types are common among men as compared to women (Öberg *et al*., 2012; Harris, 2019).

### 1.3.1 Exocrine Pancreatic Cancers:

- **Pancreatic Ductal Adenocarcinoma (PDAC):**

According to Ryan *et al*, (2014), 85% of the exocrine PC is pancreatic ductal adenocarcinoma (PDAC) that initiates in cells lining ducts of the pancreas that makes only 10% of the pancreatic cell volume. The ducts are responsible for carrying the digestive enzymes produced by the exocrine cells and then secrete them away from the pancreas. PDAC occurs in the head of the pancreas. Most frequently mutated genes in PDAC include *KRAS*, *CDKN2A*, *TP53*, and *SMAD4* (Sun *et al*., 2020; Saiki *et al*., 2021).


**Figure 1.3:** Obstruction of ducts by adenocarcinoma (PDAC) (Adapted from Web).

- **Acinar Cell Carcinoma (ACC):**

The second rare type of exocrine pancreatic cancer is acinar cell carcinoma (ACC) which constitutes 5% of PC cases (Tobias and Hochhauser, 2014). ACC originates from the acinar cells that are present at the end of ducts and produce digestive enzymes and make about 80% of pancreatic volume (Zhou *et al*., 2020). ACC causes an increase in the production of some enzymes causing joint pain and rashes on the skin. The growth of ACC is slower but can be diagnosed earlier compared to PDAC. All the genes mutated in PDAC are not commonly mutated in ACC however *APC* and *β-catenin* from the Wnt signaling pathway are likely to be mutated in ACC (Zhou *et al*., 2020).

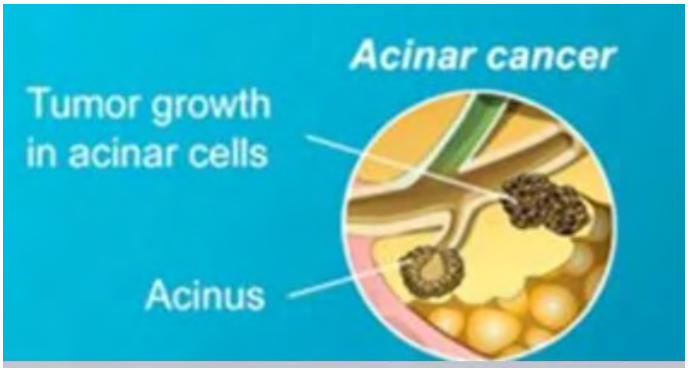


**Figure 1.4:** Acinar cell carcinoma (Adapted from Web).

- **Squamous Cell Carcinoma (SCC):**

Another rarely occurring exocrine pancreatic cancer type is SCC which develops in the squamous cells of pancreatic ducts. Recent studies suggested that in most cases SCC is diagnosed after metastasis (Zhang *et al*., 2016).

- **Adenosquamous Carcinoma:**

Adenosquamous carcinoma is also one of the rarely occurring types of cancer and accounts for 1-4% of exocrine pancreatic cancer cases (Zhang *et al*., 2016). It is a more aggressive type of exocrine cancer as compared to adenocarcinoma. In Adenosquamous carcinoma the characteristics of both PDAC and SCC are present.

- **Colloid Carcinoma:**

Colloid carcinoma is another rare type that accounts for only 1-3% of exocrine PC (Zhang *et al*., 2016). This type of tumor will be formed from a benign cyst called intraductal papillary mucinous neoplasm (IPMN). As malignant cells of the colloid tumor are present in mucin so they don't metastasize easily that's why colloid carcinoma is easy to diagnose and treat.

### 1.3.2 Endocrine Pancreatic Cancer:

- **Pancreatic Neuroendocrine Tumor (PanNETs):**

Almost 10% of PC originates in the body and tail and they are endocrine pancreatic cancer. Endocrine pancreatic cancer is a very less frequently occurring cancer type that appears only in 1-2% of PC cases (Zhang *et al*., 2016). Endocrine pancreatic cancer initiates from the cells producing hormones of the pancreas and is also known as pancreatic neuroendocrine tumor (PanNETS) because it originates from

neuroendocrine cells that integrate the endocrine and nervous system. PanNETs are classified into two main categories based on the level of hormone production:

- **Functioning PanNETs:** In such types of PanNETs large amounts of hormones (insulin, gastrin, glucagon) are released resulting in low sugar levels in the blood that make it favorable for early detection (Burns and Edil, 2012). Insulinomas and Gastrinomas are the common types of functional PanNETs.
- **Non-functioning PanNETs:** Such types of PanNETs do not release hormones in appropriate amounts so they are diagnosed when cancer metastasizes (Burns and Edil, 2012).

Almost 60% of the PanNETs are non-functional. The majority of the PanNETs are benign and are called Islets cells tumor while some are malignant and aggressive and are termed Islets cell carcinoma (Xiao *et al*., 2019).

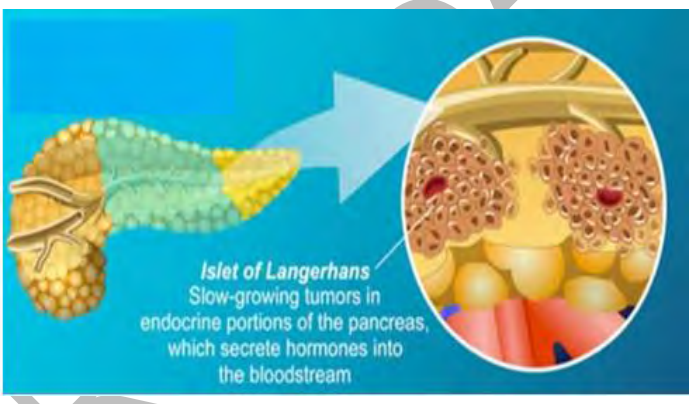


**Figure 1.5:** Pancreatic Neuroendocrine Tumor (PanNETs) (Adapted from Web).

### 1.3.3 Other Rare Types of Pancreatic Cancer:

According to Xiao *et al*, (2019), some other rare types of pancreatic cancer include: Cystic tumor, Sarcoma, Lymphoma of the pancreas, Cystadenocarcinomas, Pancreatoblastoma, Signet ring cell carcinomas, Hepatoid carcinomas, IPMN, and Mucinous cystic neoplasm (MCN)

## 1.4 Molecular Subtypes of Pancreatic Cancer:

According to Bailey *et al*, (2016), transcriptomic profiling of 232 PCs sample reported four main molecular subtypes of PC including squamous, pancreatic progenitor, immunogenic, and aberrantly differentiated endocrine exocrine (ADEX).

These subtypes are devised due to variable transcription factors (TFs) expression and other mediators involved in the development and progression of the pancreas. Out of 26 differentially expressed genes identified in transcriptional analysis 10 are involved in classifying these four subtypes of PC (Bailey *et al*., 2016; Collisson *et al*., 2019). Based on histological characterization squamous subtype is related to adenosquamous PDAC, both pancreatic progenitor and immunogenic (classical) subtypes are related to mucinous and IPMN-linked PDAC (Collisson *et al*., 2019), while ADEX is related to rare acinar cell carcinomas.

### 1.4.1 Squamous Subtype:

The Squamous tumor subtype of PC was enriched in genes involved in inflammation, metabolism reprogramming system, *TGF-β* signaling, activated *EGF* signaling, *MYC* activating pathway, and autophagy (Collisson *et al*., 2019). According to TCGA, elevated expressions of these genes are observed in many tumors other than pancreatic including breast, bladder, and lung cancer. In the pancreatic squamous subtype, mutation of *p53* along with enhanced expression of *p63* was found and is involved in regulating carcinogenicity and EMT. Complete endodermal identity loss was observed in this subtype due to hypermethylation and decreased expression of genes involved in fate-determining endodermal cells of the pancreas (Bailey *et al*., 2016).

### 1.4.2 Pancreatic Progenitor Subtype:

In pancreatic progenitor tumor subtype, expression of transcription factors (TFs) including PDX1, HNF4G/4A, HNF1B/1A, and FOXA2/A3 were observed (Bailey *et al*., 2016). These TFs are important for determining the fate of endodermal cells of the pancreas especially PDX1 is involved in the development of ductal, exocrine and endocrine cells of the pancreas. In this tumor subtype apomucins including MUC1/5AC are expressed and associated with the clustering of PDAC-linked IPMNs (Li *et al*., 2021).

### 1.4.3 Immunogenic Subtype:

The immunogenic tumor subtype possesses characteristics similar to the pancreatic progenitor subtype. Genes of B-cell signaling pathway, antigen-presenting cells, $CD_4^+$ and $CD_8^+$ T cells and Toll-like receptor (TLRs) signaling pathways are linked with this subtype. In this subtype, tumor-immune suppression pathways can be activated

by increasing the expression of CTLA4 and PD1 which further serve as a therapeutic opportunity against PC (Collisson *et al*., 2019).

### 1.4.4 ADEX Subtype:

The ADEX tumor subtype involves pathways that are required for late-stage development and differentiation of the pancreas. Upregulated genes associated with this class include *NR5A2*, *MIST1*, *INS*, *NEUROD1*, and *MAFA* and their downstream regulators that are involved in endocrine and acinar cells of pancreas differentiation. Furthermore, in ADEX tumors pattern of methylation is clustered with other PCs and different from the normal pancreas (Bailey *et al*., 2016; Collisson *et al*., 2019).

## 1.5 Signs and Symptoms of Pancreatic Cancer:

The appearance of the PC symptoms depends on the locality of the tumor-like if the tumor is in the pancreatic head then jaundice is the dominating symptom. While if the pancreatic head or tail has tumor mass then pain in the abdomen and loss of weight are common symptoms and there the tumor will take more time to grow and metastasize. Moreover, PDAC is the most frequently occurring type of PC and has slightly different symptoms than neuroendocrine tumors (Mizrahi *et al*., 2020). Not all the symptoms of PDAC appear in each case but the most common symptoms are as follows:

### 1.5.1 Upper Abdominal and Back Pain:

If the pancreas body or tail area has a tumor then it exerts pressure on the spine and causes upper abdomen and upper back pain. This pain gets even worse by lying down (Khyade, 2018).

### 1.5.2 Jaundice

Jaundice is also one of the most common symptoms in PDAC and is not frequently seen in ACC (Zhou *et al*., 2020). It usually appears when the tumor mass blocked the bile duct and ultimately tumor mass grows in the pancreas head and results in yellowing of the eyes, skin, dark and pale urine, and stools, irritating skin, and weight loss (Walter *et al*., 2016).

### 1.5.3 Weight Loss:

Unintentional weight loss is also considered as most common symptom that appears because of the tumor presence in the pancreas head (Bond-Smith *et al*., 2012).

### 1.5.4 Gastrointestinal (GI) Problems:

When the tumor metastasizes and presses the stomach or other regions of the digestive tract then these symptoms appear and ultimately result in loss of appetite, indigestion, nausea, vomiting, abdominal bloating, and fatigue (Zhou *et al*., 2020).

### 1.5.5 Diabetes:

According to Ryan *et al*, (2014), almost 50% of PDAC patients have this symptom. During the progression of PC, beta cells are also destroyed which leads to diabetes so the early onset of diabetes is also one of the important symptoms of pancreatic cancer.

### 1.5.6 Others:

Other symptoms include weakness, dry mouth, sleeping problems, fever, chills, and extreme tiredness (Zhou *et al*., 2020).

## 1.6 Etiological Factors of Pancreatic Cancer:

The risk of developing PC throughout the life of a person is 1.49% which accounts for 1 case in 67 individuals (Becker *et al*., 2014) and is mostly reported at the age of 60 to 80 years (Ansari *et al*., 2016). Some of the main risk factors causing PC include:

### 1.6.1 Age:

Pancreatic cancer mostly originates in older adults and is a very rarely occurring type of cancer in young people at 30 years. Almost 90% of newly diagnosed cases are in individuals having the age over 55 years, mostly between 70-80 years old (McMenamin *et al*., 2017)

### 1.6.2 Gender:

Pancreatic cancer is less common in women as compared to males throughout the world (Öberg *et al*., 2012; Harris, 2019). One study suggested that women have a high level of steroids than men so it protects them against PC (Masoudi *et al*., 2017).

### 1.6.3 Area:

In the U.S the incidence rate of PC is higher in African Americans than Caucasians while the incidence rate is lower among both Asian Americans and Pacific Islanders (Midha *et al*., 2016). In China, the mortality rate due to PC is more in the urban area

as compared to rural areas (Li *et al*., 2017). Moreover, in Africa, the incidence rate is the lowest (Rawla *et al*., 2019).

### 1.6.4 Alcohol:

Alcohol and Chronic pancreatitis have a close relation in causing PC. Recent studies suggest that high consumption of alcohol consumption can enhance the risk of getting pancreatic cancer (Lucenteforte *et al*., 2012).

### 1.6.5 Diabetes:

Long-term diabetes mellitus is considered an important risk factor for PC (Ansari *et al*., 2016) but the onset of diabetes mellitus (DM) in individuals at an age of 50 years may also lead to PC. A recent study suggested that newly diagnosed patients with diabetes have a 7-fold enhanced chance of getting PC as compared to nondiabetic patients. In diabetic patients with PC, the level of glucose in the blood and glycated hemoglobin (HbA1c) increased significantly so HbA1c can be used as a potential biomarker for determining PC (Huang *et al*., 2020).

### 1.6.6 Smoking:

In the causation of PC, smoking is one of the most common risks (Nitsche *et al*., 2011). A study suggests that an almost 74% enhanced risk of getting PC exists among smokers than in non-smokers (Molina-Montes *et al*., 2020). According to Nimmakayala *et al*, (2018) smoking cigarette along with their components enhances the stemness characteristics of pancreatic cells and makes the pancreatic cells to renew themselves on their own and differentiate to develop PC.

### 1.6.7 Family History:

Almost 90% of PC cases are sporadic, while only 10% are hereditary (Nitsche *et al*., 2011). Studies suggest that all mutations inherited in the *BRCA2* gene enhance PC risk. Approximately 30–40% of individuals having hereditary pancreatitis are at high risk of getting PC in their early 70s (Wolfgang *et al*., 2013).

### 1.6.8 Genetic Susceptibility:

The highly mutated genes in PC include *KRAS*, *CDKN2A/p16*, *TP53*, and $SMAD_4$ (Makohon-Moore *et al*., 2013; Zhang *et al*., 2016) while, point mutation of other

genes that increased the risk of PC include *BRCA1*, *BRCA2*, *PRSS1*, *STK11/LKB1*, *hMLH1*, *hMSH2*, *FANC-C*, and *FANC-G* (Del Chiaro *et al*., 2014).

### 1.6.9 Chronic Pancreatitis (CP):

Chronic Pancreatitis (CP) is the pancreatic inflammation leading to pancreatic fibrosis and is one of the risks for PC (Yadav and Lowenfels, 2013; Samokhvalov *et al*., 2015). The CP leads to the production of pancreatic enzymes that are abnormal and deteriorates the endoplasmic reticulum (ER), mitochondria, and autophagy systems of lysosome of cells of the pancreas leading to damage of DNA, mutations in chromosomes, and oncogenes activation (Yadav and Lowenfels, 2013).

### 1.6.10 Blood Group:

According to recent studies, the blood group antigen is considered a prominent risk for PC (Risch *et al*., 2013). Diabetic patients having blood groups A, AB, or B have an increased chance of getting PC unlike the type O blood group (Li *et al*., 2018).

### 1.6.11 Obesity:

Obesity is also responsible for the high incidence rate of pancreatic cancer. Obese people have 20% more chances of getting PC as compared to people with normal weight (Bracci, 2012).

### 1.6.12 Others:

Some other risk factors that are responsible for causing PC include intestinal microbiota, dietary factors, hepatitis B and bacterial infection, pancreatic cyst, cystic fibrosis, chemicals, and liver cirrhosis (Li *et al*., 2018).

## 1.7 Pancreatic Cancer Stages:

Staging is very important for measuring the size and spread of cancer and is usually done by using Computed Tomography (CT scan) (De La Cruz *et al*., 2014). Preferably pancreatic cancers are classified into three main groups based on their spread and resectability.

- **Resectable:**

In this type, the tumor can be removed through surgery because the tumor is in early stages (I, II) and is present in the pancreas and not spread to neighboring arteries or veins. At this stage, only 10% to 15% of PC cases can be diagnosed.

- **Borderline-Resectable:**

In this type, the tumor is difficult to remove through surgery but if it gets diagnosed earlier than by chemotherapy and radiotherapy the tumor can be shrinked and then removed (Wolfgang *et al*., 2013).

- **Un-Resectable:**

In this type, the tumor can't be removed surgically because the tumor is in stage IV and spread to arteries, veins, and neighboring organs like the liver, lungs, or parts of the abdomen. At this stage, 35% to 40% of patients are diagnosed (Wolfgang *et al*., 2013).

### 1.7.1 TNM Staging of Pancreatic Cancer:

The American Joint Committee on Cancer (AJCC) along with the Union for International Cancer Control (UICC) devised four stages for pancreatic cancer from early to advanced disease based on TNM staging shown in Figure 1.6.

- **Stage I:**

Cancer is present only in the pancreas and based on the size of the tumor mass stage I is categorized into the following categories:

  - **Stage IA:** The tumor mass is less than 2cm (T1, N0, M0).
  - **Stage IB:** The tumor mass is larger than 2cm (T2, N0, M0).

- **Stage II**

In this stage, cancer may spread to neighboring tissues, organs, and lymph nodes, and based on cancer metastasis Stage II is categorized into the following types:

- **Stage IIA:** Tumor mass spread to various tissue and organs present close to the pancreas but not to lymph nodes, blood vessels, and distant regions (T3, N0, M0).
- **Stage IIB:** Tumor mass spread to lymph nodes, tissues, and organs but not metastasized to distant sites and blood vessels (T1/2/3; N1, M0).

➢ **Stage III**

In this stage, cancer metastasizes to blood vessels and lymph nodes but is not moved to distant regions (T4, Any N, M0).

➢ **Stage IV**

In this stage, cancer is larger and metastasizes to distant organs like the liver, lung, and parts of the abdomen along with the lymph nodes, organs, and tissues near the pancreas (Any T, Any N, M1).

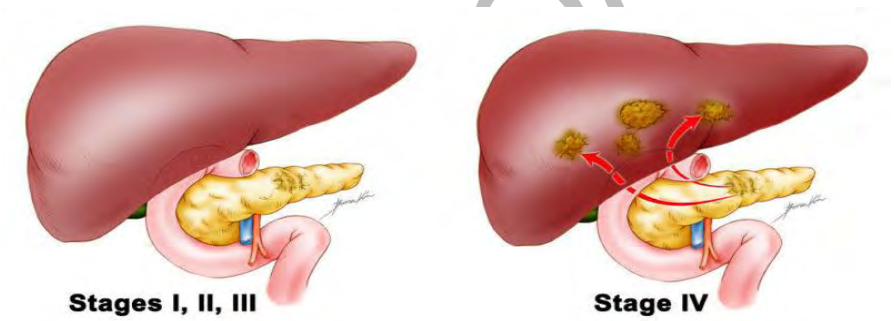


**Figure 1.6:** Pancreatic cancer Stages (Adapted from Web).

## 1.8 Progression of PDAC Carcinogenesis:

The PC progression requires various events including the oncogenes activation, tumor suppressor genes inactivation, and the cell cycle genes deregulation. Noninvasive pancreatic neoplasia involves three morphologic forms including IPMN, MCN, and pancreatic intraepithelial neoplasia (PanIN). Out of these, the frequently occurring precursor lesion of PDAC is PanIN involving the pancreatic ductal epithelium (Zhang *et al*., 2016) as shown in Figure 1.7.

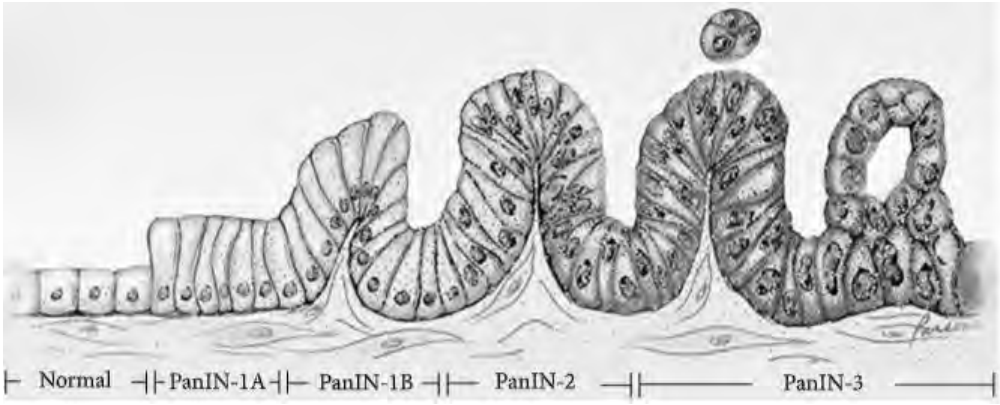


**Figure 1.7:** PanINs in ductal epithelium of pancreas leading to PDAC (Zhang *et al*., 2016).

The four major driver gene mutations involved in the progression of PC include *KRAS*, *CDKN2A*, *TP53*, and *SMAD4* (Makohon-Moore *et al*., 2013) as mentioned in Figure 1.8. An activation mutation in KRAS at codons 12 and 13 can lead to the development of PanIN-1 lesion from normal acinar cells with large columnar epithelium lining the ductal system and have minute nuclear atypia (Makohon-Moore *et al*., 2013). The second inactivating mutation by allelic loss, homozygous deletion, or hypermethylation occurs in *CDKN2A* which leads to the PanIN-1 progression to PanIN-2 lesion causing polarity loss, and formation of papillary and nuclear atypia (Zhang *et al*., 2016). Then the inactivating mutation of *TP53* (point mutation) and *SMAD4* (allelic loss, homozygous deletion) causes a complete polarity loss, clear nuclear atypia, increased nuclear to cytoplasmic ratio, and formation of pseudopapillary leads to PanIN-3 lesions which than metastasize and develop PDAC (Makohon-Moore *et al*., 2013).



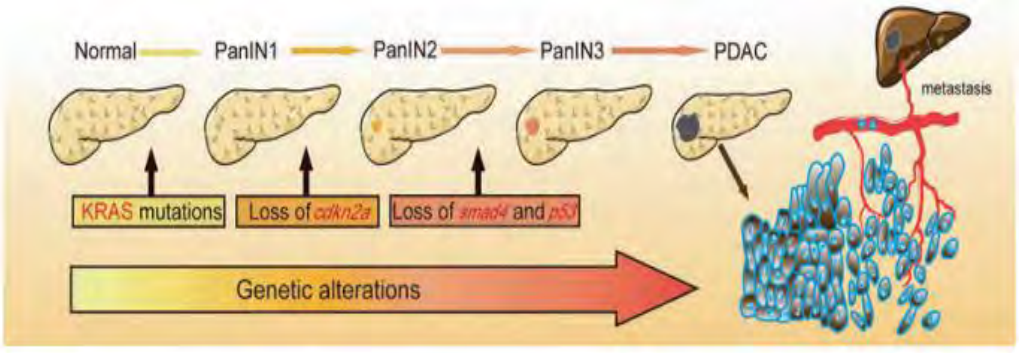


**Figure 1.8:** Driver genes involved in the progression of pancreatic carcinogenesis (Hu *et al*., 2021).

## 1.9 Incidence and Mortality Rate of PC Worldwide:

Pancreatic cancer (PC) is considered a highly fatal malignancy and is reported to be the seventh leading cause of death related to cancer among both men and women throughout the world. Recently, a report by the American Society of cancer oncology (ASCO) showed in 2020 approximately 495,773 people were diagnosed and 466,003 people died from PC worldwide. According to GLOBOCAN 2020 estimate approximately 495773 new cases and 466 003 deaths of PC were reported worldwide (GLOBOCAN, 2020). It is also predicted that in European Union PC will be going to replace the place of breast cancer as the 3$^{rd}$ leading cause of death because of cancer. Furthermore, it is reported as the fourth leading cause of death in the U.S (Giannis *et al*., 2021). In the next twenty to thirty years, in the USA, PC is projected to be the 2$^{nd}$ leading cause of death related to cancer (Mizrahi *et al*., 2020). When diagnosed the overall survival rate of 5 years among PC patients is just 10% in the USA, because in almost 80–85% of pancreatic cancer cases either it becomes a non-resectable or metastasized disease (Mizrahi *et al*., 2020). Besides this, if a small ratio of patients can be diagnosed with a localized and resectable tumor, still cancer prognosis is poor with only a 20% of 5 years survival rate even after surgery (Mizrahi *et al*., 2020).

As reported by the American Cancer Society (ACS), in 2019 approximately 56 000 new pancreatic cancer cases were diagnosed in the USA with an estimated 45000 deaths, ranking third highest death rate after lung and colorectal cancer. Furthermore, they proposed that approximately 62210 people will be diagnosed and 49830 people will die due to PC in 2022. Among females, PC is the 8$^{th}$ highly frequent type of cancer and is the 10$^{th}$ most prominent cancer type among men.

Moreover, Cancer Research UK reported PC as the tenth most common cancer with approximately a 10% increase in the incidence rate over 10 years (Mizrahi *et al*., 2020). Over the past few years, several advancements were made in diagnostic approaches, and therapies but still new strategies are required for detecting pancreatic tumors at a very early stage to produce a clinically significant effect.

**Table 1.1:** Incidence and Mortality rate of PC among the different populations.

| Population | Number of Incidences | Number of Deaths |
|---|---|---|
| **Asia** | 233701 | 224034 |
| **Europe** | 140116 | 132134 |
| **North America** | 62643 | 53277 |
| **Latin America and the Caribbean** | 37352 | 36030 |
| **Africa** | 17070 | 16549 |
| **Oceania** | 4891 | 3979 |



### 1.10 Incidence and Mortality Rate of Pancreatic Cancer in Pakistan:

A recent review showed, that approximately 97.8% of pancreatic cancer cases lead to death in Pakistan (Ali *et al*., 2021). In 2020, the rate of incidence of PC increased by two-fold in Pakistan as compared to 2012. A recent GLOBOCAN report proposed that the incidence rate will increase twice in Pakistan by 2040 with a high mortality rate. Exocrine pancreatic cancer accounts for 93% of all PC cases, the most common of which is PDAC. In the last 14 years, only 23 studies have been reported in Pakistan out of them not a single study was done for molecular analysis of pancreatic cancer. In 2004, Aga Khan University reported the three most common symptoms of pancreatic cancer among the Pakistani population including smoking, diabetes mellitus, and chronic pancreatitis (Ali *et al*., 2021). From various studies conducted in Pakistan, the most frequently targeted area of origin of tumor mass is the pancreatic head while some cases have been reported from the body of the pancreas. In Pakistan, endoscopic ultrasound (EUS) is preferably used as a diagnostic technique. FDA approved cancer antigen (CA-19) as the marker for pancreatic cancer detection but in Pakistan, no study has been reported involving the use of CA-19 as a diagnostic

marker (Ali *et al*., 2021). Currently, in Pakistan, no study has been conducted involving drug trials against pancreatic cancer.

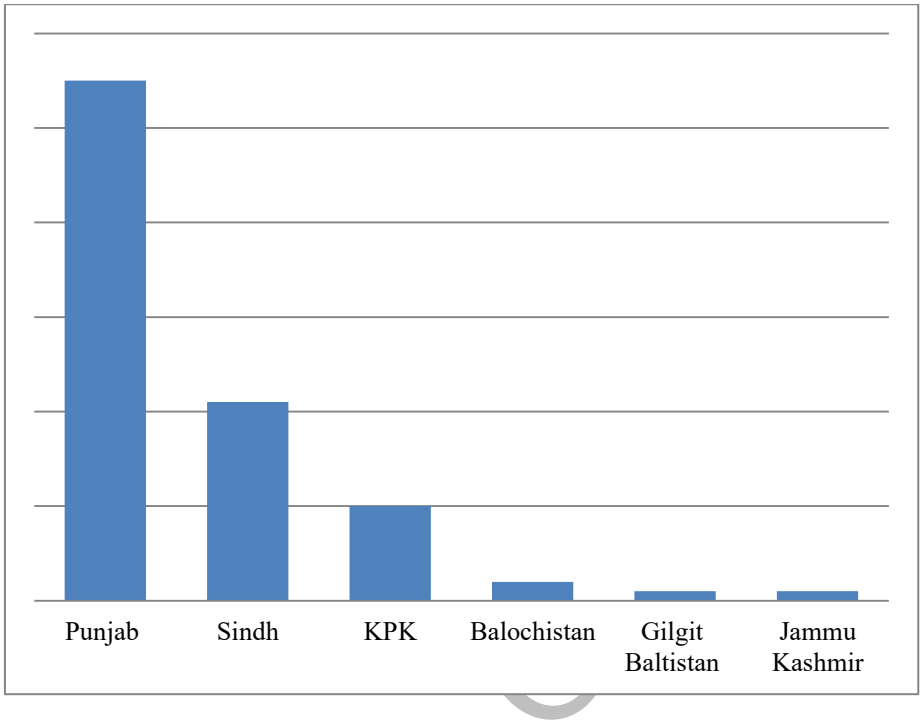


**Figure 1.9:** Research publication regarding pancreatic cancer from different regions of Pakistan (Ali *et al*., 2021).

### 1.11 Diagnosis of Pancreatic cancer:

Early diagnosis of PC can be difficult but it is very important because it will help the doctors to devise an effective treatment for the patients timely. Diagnosis of PC can be done by using various imaging techniques, blood tests, and tumor marker analysis (Zhang *et al*., 2016).

#### 1.11.1 Computerized Tomography (CT) Scan:

To detect PanINs lesions, multidetector computed tomography (MDCT) can be used. It also helps in assessing the size, location, resectability, malignancy, and vascular invasion of pancreatic tumors (Zhang *et al*., 2016). As MDCT not only provide details regarding lesion but also show the morphological changes of tumor that help in defining the stage of PC and are significant in finding out the resectability status of the tumor.

### 1.11.2 Positron Emission Tomography (PET) Scan:

The PET scan is used for diagnosing tumors and other diseases at a very early stage. The metabolism of tumor cells is quite active which helps them to take more imaging agents as compared to normal cells and results in the production of more light spots on the image. Hence, this technique will help in the early detection of PC even before the tumor expresses any anatomical changes.

### 1.11.3 Magnetic Resonance Imaging (MRI):

Nowadays, MRI is preferably used for detecting PC staging and used to measure tumor size. It is also used to observe whether the tumor can be removed through surgery. The results of MRI for PC staging show a sensitivity and specificity of 1.00 and 0.67 respectively (Zhou *et al*., 2020). In MRI magnetic fields helps to observe detailed body images. One type of MRI, Magnetic resonance cholangiopancreatography (MRCP) uses computer software to observe ductal blockages and cysts in the pancreas and to detect pancreatic cancer.

### 1.11.4 Endoscopic Ultrasonography (EUS):

Several studies reported that the sensitivity of EUS is more as compared to MDCT for detecting pancreatic cancer. During this procedure, a biopsy may also be done. EUS-guided fine needle aspirations (FNA) of pancreatobiliary tumors have more accuracy and can be used for examining the involvement of blood vessels and lymph nodes in the staging of tumors in the pancreatic head (Zhang *et al*., 2016).

### 1.11.5 Endoscopic Retrograde Cholangiopancreatography (ERCP):

The ERCP is used for detecting cancer of the pancreatic head. It assists us to detect bile duct obstruction and stenting can also be done to reduce jaundice symptoms. During this technique, small tissue can be taken that will help in the diagnosis of PC. This technique is most commonly used for bile duct stenting rather than the diagnosis of PC.

### 1.11.6 Histopathological Analysis:

Histopathological analysis can also be used for the diagnosis of PC. Through EUS-guided biopsy tissue will be collected that will, later on, be used for morphological analysis of PC (Zhou *et al*., 2020).

### 1.11.7 Blood Tests and Tumor Biomarkers:

For the early diagnosis of PC, blood tests are always preferred. Common tests include complete blood count (CBC), liver function test (LFT), renal function test (RFT), and analysis of bilirubin level. Along with them, the level of tumor biomarkers will also be observed that are helpful for the diagnosis of PC. Currently, carbohydrate antigen 19-9 (CA19-9) and carcinoembryonic antigen (CEA) are the tumor biomarkers whose level is very important for the detection of pancreatic cancer (Zhang *et al*., 2016; Zhou *et al*., 2020). However, their level can also be elevated due to some other medical conditions so the level of these tumor biomarkers is sometimes not so reliable.

## 1.12 Treatment and Therapeutics for Pancreatic Cancer:

Over the past few years, several advancements were made in diagnostic tests but still, the rate of diagnosis and survival is quite low. Despite of chemotherapy and radiotherapy, surgery is the best treatment for PC patients (Zhang *et al*., 2016).

### 1.12.1 Surgical Resection:

For treating PC, surgical resection is one of the best methods which involves either total pancreatectomy, distal pancreatectomy plus splenectomy, or pancreaticoduodenectomy (Zhou *et al*., 2020). Based on the location, size, and staging of tumor mass the resection procedure will be decided for PC. The most common surgical procedure for pancreatic cancer of the head is pancreatoduodenectomy. While if the pancreatic body and tail are involved in cancer, then distal pancreatectomy will be performed. Still, several improvements have to be made in surgical techniques to reduce the mortality rate.

### 1.12.2 Chemotherapy:

Despite low efficacy in metastatic PC cases, chemotherapy is still used before and after surgery to reduce the symptoms and improve the survival rate. For unresectable PC, gemcitabine (GEM) is used. Several other drugs that have been approved by U.S Food and Drug Administration (FDA) for PC include 5-fluorouracil (5FU), capecitabine (Xeloda), cisplatin, leucovorin, oxaliplatin, epirubicin, and irinotecan. In 2011, FOLFIRINOX was used for advanced PC cases and it is a combination of oxaliplatin, 5FU, leucovorin, and irinotecan. Oxaliplatin is also administered as

primary therapy for reducing the intensification of advanced PC (Springfeld *et al*., 2019).

### 1.12.3 Radiography:

Radiotherapy is considered the best treatment option when the tumor is advanced and unresectable. Radiotherapy uses X-rays to kill cancerous cells and recent forms of radiotherapy that made this treatment more effective include intensity-modulated radiotherapy (IMRT) and image-guided radiotherapy (IGRT) (Son *et al*., 2012). Radiotherapy is further divided into two main types, external beam radiotherapy and brachytherapy (internal radiotherapy) (Zhou *et al*., 2020). Stereotactic body radiotherapy (SBRT) is an advanced radiotherapy type that targets the cells of tumor more precisely without damaging the neighboring normal cells (Zhang *et al*., 2016).

### 1.12.4 Immunotherapy:

PC is not a highly immunogenic cancer because it creates an immunosuppressive microenvironment due to which no immunotherapy has been approved to treat PC. CTLA-4 or PD1 inhibitors monotherapy is also not effective against PC (Sahin *et al*., 2017). In one study it is reported that in cells of tumor mass by inhibiting the interferon-gamma (IFNG) signaling the ability of the body to kill tumor cells can be enhanced and ultimately increase the efficacy of immunotherapy (Benci *et al*., 2019).

### 1.12.5 Other Treatment Strategies:

New therapies including Microbial, Stem Cells, and Gene therapy have not been yet applied for PC treatment but have successful *in-vitro* and *in-vivo* trials (Ansari *et al*., 2016; Zhou *et al*., 2020). Gene therapy involves the replacement, modification, and blockade of targeted genes including *CDKN2A*, *KRAS*, *VEGF*, *HER-2*, and *LSM1*. Recently a new vaccine using a live attenuated Listeria strain has been prepared against PC and is in the early phase of the clinical trial. In the future, new strategies will be devised that will produce a promising effect on PC (Zhang *et al*., 2016).

## 1.13 Altered Signaling Pathways in Pancreatic Cancer:

Studies reported almost 63 genetic alterations in major signaling pathways involved in pancreatic cancer including the alteration in cellular processes like an apoptotic pathway, DNA damage control, transition of G1/S cell cycle, and invasion regulation

pathway (Eltawil *et al*., 2012) as shown in Figure 1.10. Based on whole-exome sequencing analysis, the following are the highly altered pathways involving most of the mutated genes in PC.

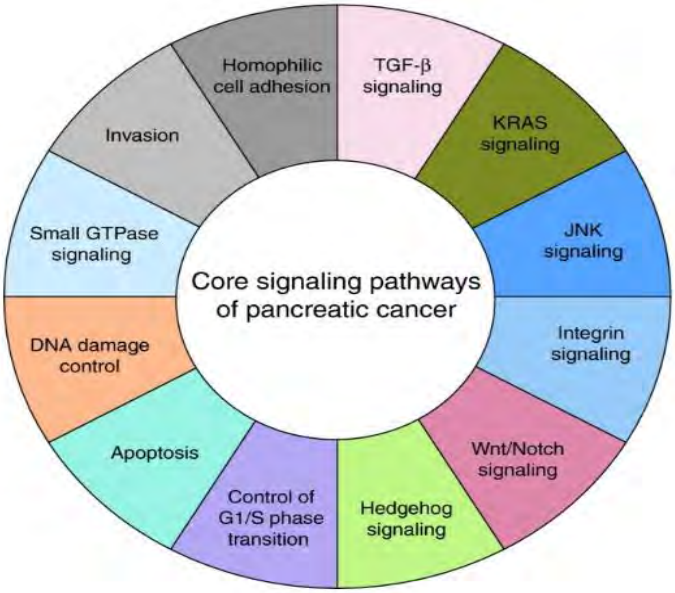


**Figure 1.10:** Altered signaling pathways in PC (Makohon-Moore *et al*., 2013; Bailey *et al*., 2016).

Some mutated genes play the role only in a single pathway like *KRAS* while other altered genes are involved in multiple pathways like *p53*. Specific targeting of altered genes within various signaling pathways further paves the way for treating pancreatic cancer (Makohon-Moore *et al*., 2013; Bailey *et al*., 2016).

### 1.14 Role of the driver and various other genes in PDAC:

According to Hu *et al*, (2021) in PDAC the highly mutated driver genes include *KRAS* (77%), *CDKN2A* (63%), *TP53* (22%), and *SMAD4* (16%) with missense and nonsense as the most common types of mutations (Hu *et al*., 2021). Along with these several other genes have been discovered whose mutation is significant in the progression of pancreatic cancer.

- ***KRAS*:**

Almost 93% of PDAC cases involve *KRAS* mutation mostly in codon G12 (G12C/D) of exon 2. Due to codon 12-point mutations, conversion of GTP to GDP is halted, which leads to constant activation of signaling pathways leading to various types of cancer (Hu *et al*., 2021). *KRAS* mutation along with several other factors including

oxidative stress and inflammation causes the formation of PanINs and lead to the development and progression of PDAC.

- ***CDKN2A*:**

The *CDKN2A* is a tumor suppressor gene, deletion mutation, or promoter hypermethylation of which is involved in 30-50% of PDAC cases. Along with *KRAS* mutations, *CDKN2A* inactivation leads to pancreatic malignancy (Singh and Ellenrieder, 2013). *CDKN2A* encodes $p14^{ARF}$ and $p16^{INK4A}$, whose mutation is involved in PDAC progression and differentiation (Hu *et al*., 2021).

- ***TP53*:**

*TP53* gene mutation is involved in 60%-70% of PC cases. Homozygous mutation of *p53* was observed in the formation of PanIN-3 leading to PDAC progression. Studies suggested that invasion and metastasis of pancreatic tumor cells to lymph nodes are enhanced by *p53* mutation (Hu *et al*., 2021).

- ***SMAD4*:**

Out of four driver genes, one is *SMAD4* which is involved in almost half of the PDAC cases. Homozygous deletion and chromosome allelic loss are involved in most of the cases. Along with *KRAS* and other mutations, the inactivation of *SMAD4* is also involved in PanINs progression and tumor metastasis (Hu *et al*., 2021).

- **Other Genes:**

*BRCA1/2*, *ATM*, *PALB2*, and *BRAF* are other genes that are most commonly mutated in PDAC. In almost 4%–7% of pancreatic cancer cases germline mutation of *BRCA1/2* was also observed (Golan *et al*., 2019). *BRAF* gene point mutation has been identified in 1.4-3% of PDAC cases (Davis *et al*., 2018).

### 1.15 *p53* gene and its normal function:

The Human *p53* gene is a tumor suppressor gene, found in the 13.1 region of the short arm (p-arm) of chromosome 17 (17p13.1) (Gbadegesin *et al*., 2021). This gene spans over 20 kb and has 11 exons including a non-coding exon 1 and a very large $1^{st}$ intron of 10 kb size. In vertebrates, a highly conserved sequence is present in the coding

sequence of the *p53* gene including exons 2, 5, 6, 7, and 8. The homolog genes like *TP53* in humans and *Trp53* in mice encode many protein isoforms including tumor protein P53 (TP53), phosphoprotein p53, and transformation-related protein 53 (TRP53). Moreover, *p53* has a significant role in stability maintenance by avoiding mutation in the genome that's why referred to as "the guardian of the genome".

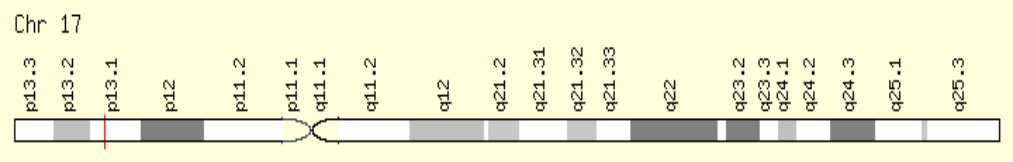


**Figure 1.11:** Genomic location of *p53* on chromosome 17 (Adapted from GeneCards).

As shown in Figure 1.12, p53 protein is comprised of 7 domains including an acidic transcription activation domain (TAD) located at N-terminus, activation domain 2 (AD2), a domain rich in proline, central DNA-binding core domain (DBD), nuclear localization signaling (NLS) domain, the tetramerization domain (TD), and regulatory domain located at C-terminal.

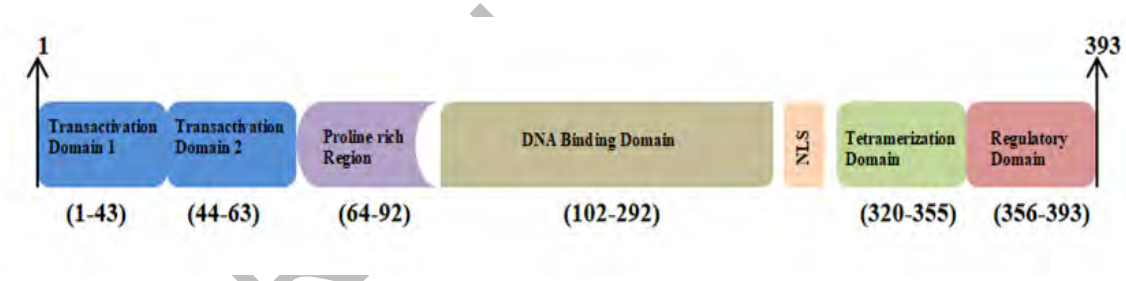


**Figure 1.12:** p53 protein domains.

Due to alternative splicing, there are almost 12 human p53 isoforms have been known and their expression is dependent on different tissues. In multicellular organisms, p53 along with its other isoforms including p63 and p73 play a significant role in DNA damage detection and repair. In humans, *p53* forms a barrier against carcinogenesis and is also involved in apoptosis, regulation of cell cycle, inhibition of angiogenesis, and DNA repair mechanism (Zawacka-Pankau, 2020).

These processes regulated by *p53* require an expression of *p53*-target genes (*p53TGs*). One of the p53 target genes is *TP53TG1* which is an RNA Gene and is associated with the long non-coding RNA (lncRNA) class.

*TP53TG1* is a wild-type *p53*-activated lncRNA, that act as a tumor suppressor (Diaz-Lagares *et al*., 2016) and under stress conditions, *p53* induces the expression of *p53TG1* ultimately showing its role in damage repair. However, *p53TG1* decreased expression is involved in gliomas and various human cancer progressions (Xiao *et al*., 2018). A certain level of *p53* is significant for stem cell maintenance throughout human life (Jain *et al*., 2012). DBD is the most frequently mutated domain of *p53* involved in many types of cancer. Furthermore, *p53* is the highly mutated or deleted gene in almost 50% of cancers including colorectal, ovarian, esophageal, and lung cancers, *etc*. The recessive loss-of-function mutations of the *p53* gene prevent attachment of the protein to DNA sequences; as a result, prevent the transcriptional activation of various genes.

### 1.16 Polymorphism of *p53*:

Recently a study suggested some most common single nucleotide polymorphisms (SNPs) located at chromosome 17 on the *P53* gene that has a very significant role in pancreatic cancer risk (Feng *et al*., 2019) as shown in Table 1.2.

**Table 1.2:** Association of *p53* gene SNPs with Pancreatic Cancer risk (Feng *et al*., 2019).

| SNP | Chromosomes Position | Allele (Reference Allele/ Effect Allele) | Effect Allele Frequency (EAF) |
|---|---|---|---|
| rs17884306 | 7572101 | C/T | 0.06 |
| rs9891744 | 7574864 | C/T | 0.06 |
| rs9895829 | 7578679 | A/G | 0.06 |
| rs17883323 | 7579619 | G/T | 0.06 |
| rs8079544 | 7580052 | C/T | 0.06 |

| | | | |
|---|---|---|---|
| rs75732100 | 7576348 | C/T | 0.06 |
| rs17879377 | 7574721 | C/T | 0.05 |

### 1.17 *p53* role in PDAC:

Almost 75% of pancreatic cancer cases have mutated *p53* which is commonly observed in PDAC and then in adenosquamous carcinoma while no mutation is observed in pancreatic neuroendocrine neoplasms (Simtniece *et al*., 2015; Jahedi *et al*., 2019). Most of the mutations of *p53* are missense and are present in the DBD which helps the tumor cells to survive and proliferate. Due to mutation in *p53* their ability to induce carcinogenesis is increased by tumor microenvironment remodeling and increasing cellular metabolism. Heat shock protein (Hsp90) binds to mutated *p53* and blocks MDM2 resulting in the accumulation of mutated *p53* leads to pancreatic cancer aggressiveness (Weissmueller *et al*., 2014; Mantovani *et al*., 2019).

Various studies on MiaPaca-2 and BxPC-3 cell lines of PC suggested that platelet-derived growth factor receptor β (PDGFRβ) interacts with mutant *p53* (mut-*p53*) and enhance the growth of pancreatic cancer cells (Weissmueller *et al*., 2014). Mut-*p53* interacts with their isoforms (*p63, p73*) and inactivates them resulting in tumor invasion and metastasis. Mutant *p53* also interacts with transcription factors to alter cell cycle regulation.

According to one study, in a mouse model mutant *p53* helps in the metastasis of pancreatic tumor cells to lymph nodes and also facilitates the conversion of neoplasms to PDAC. Recently, a prominent homozygous mutation of *p53* was also observed in PanIN-3 leading to PDAC (Hu *et al*., 2021). In mouse pancreas, combined mutation of *p53* and *KRAS* (Trp53R172H, KrasG12D) leads to metastatic PDAC with the instability of a high degree in the genome.

### 1.18 *YAP1* gene and its normal function:

The Yes Associated Protein 1 (*YAP1*)/ *YAP65* is a transcriptional coregulator that has a significant role in the tumor suppressor Hippo pathway. This gene is located at the

22.1 regions of the long arm (q-arm) of chromosome 11 (11q22.1) and has 122978 bases and 9 exons. YAP along with its paralog transcriptional coactivator TAZ plays a vital role in the Hippo pathway (Liu *et al*., 2018).

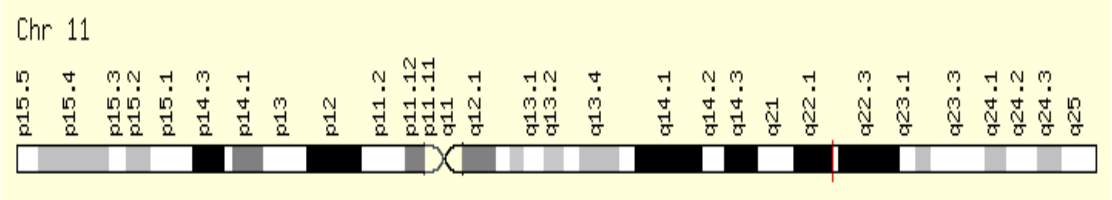


**Figure 1.13:** Genomic location of YAP on chromosome 11 (Adapted from GeneCards).

As shown in Figure 1.14, YAP protein constitutes of 7 domains including a proline-rich domain located at N-terminal, a TEAD interaction domain (TID), two WW domains, an SH3-binding motif, and a coiled-coil domain (CC), a transcription activation domain (TAD), and a PDZ-binding motif located at C-terminal. Moreover, the basic variation between YAP and TAZ protein is in their WW domain (which has 38 amino acids) as YAP has two WW domains while TAZ has just one WW domain (Liu *et al*., 2018). *YAP* gene has 2 main isoforms *YAP1, 2* that include many subtypes based on their difference in their WW and transcription activation domain (TAD) including YAP1-1α/1β/1γ/1σ, and YAP1-2α/2β/2γ/2σ (Sudol, 2013).

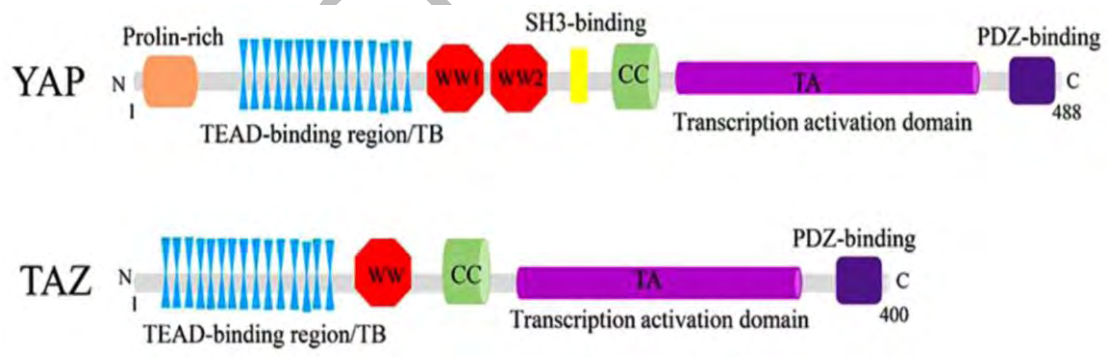


**Figure 1.14:** YAP1, 2 (YAP, TAZ) protein domains (Liu *et al*., 2018).

In the Hippo signaling pathway, the *YAP* gene act as a coactivator and a corepressor, and as they don't have a DNA binding domain so it promotes the growth, and development of stem cells, control organ size, maintain tissue homeostasis, suppress tumor, and promote apoptosis by interacting and enhancing the expression of various nuclear effectors (Szulzewsky *et al*., 2021). MST1/2 and LATS1/2 are two types of

kinases that phosphorylate YAP at S127 residue located in the TEAD interacting domain that causes the cytoplasmic retention of YAP and facilitates apoptosis. While, if any residue other than S127 is phosphorylated it causes YAP ubiquitination and proteosomal degradation, while if YAP is not phosphorylated then it localizes to the nucleus, and along with TEAD family members they form a complex leading to growth-promoting and anti-apoptotic genes transcription (Hayashi *et al*., 2021). The WW domains of YAP1 promote its interaction with other proteins including p63, p73, RUNX2, and SMADs (Szulzewsky *et al*., 2021). However, the *YAP* gene enhanced expression is involved in tumor progression of different organs (Sudol *et al*., 2012).

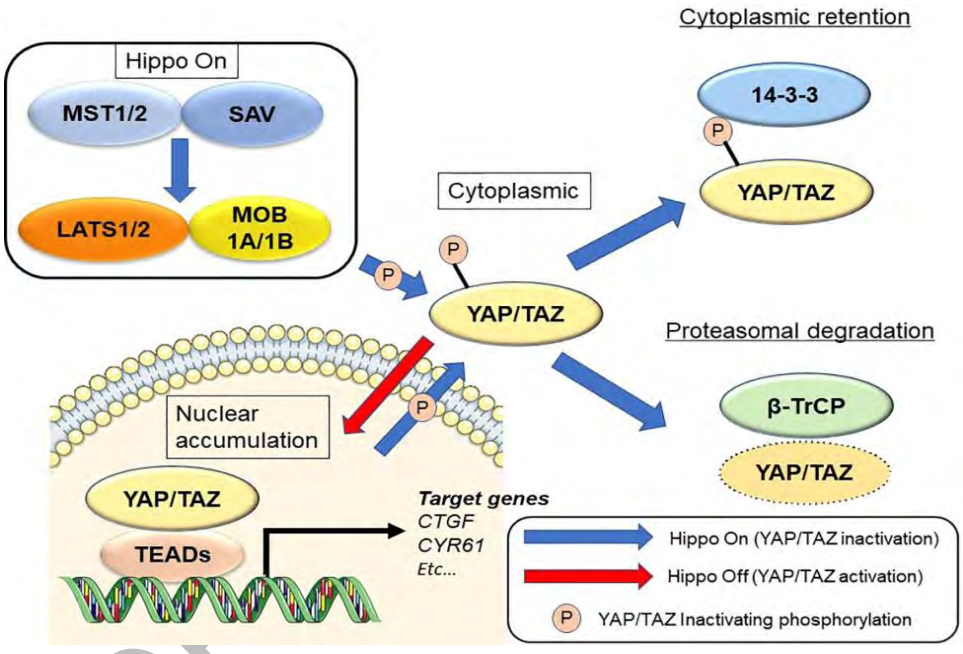


**Figure 1.15:** Regulation of YAP1 protein in Hippo pathway (Hayashi *et al*., 2021).

## 1.19 Polymorphism of *YAP1*:

Some of the main single nucleotide polymorphism (SNPs) located at a different position in the *YAP 1* gene at chromosome 11 that is responsible for causing different diseases as shown in Table 1.3 include the following:

**Table 1.3:** Role of SNPs of *YAP1* gene in various diseases (Adapted from SNPedia).

| SNP | Chromosomes position | Disease |
|---|---|---|
| rs11225138 | 102,123,167 | Polycystic ovary syndrome |
| rs11225161 | 102,199,763 | Polycystic ovary syndrome |
| rs11225163 | 102,200,112 | Cutaneous melanoma (CM) |
| rs11225166 | 102,219,736 | Polycystic ovary syndrome |
| rs1820453 | 102,109,604 | Small-cell lung cancer |
| rs1894116 | 102,199,908 | Polycystic ovary syndrome |



### 1.20 *YAP1* role in PDAC:

In several cancer types including PDAC, the *YAP* and *TAZ* are considered important oncogenes and are involved in the several gene regulations involved in various processes (Yang *et al*., 2015). Alone increased expression of YAP1 is not sufficient for inducing tumors but the inactivation of several other mediators involved in the hippo pathway facilitates YAP in cancer progression (Szulzewsky *et al*., 2020; Szulzewsky *et al*., 2021). As shown in Figure 1.16, one of the main targets of the KRAS signaling pathway, YAP1 is involved in acinar-to-ductal metaplasia (ADM) which progressively develops PanIN leading to PDAC among mouse models that were altered genetically (Gruber *et al*., 2016).


**Figure 1.16:** PDAC development by activation of YAP1 through KRAS-dependent pathway.

*YAP1* and *TAZ* also interact with mutant *p53* and are involved in PC development (Di Agostino *et al*., 2016). Studies showed that *YAP* deletion has no effect on the normal development of the pancreas but PDAC progression through KRAS is blocked (Zhang *et al*., 2014). G-protein coupled receptors (GPCRs) interact with the IGF-1 receptor and upregulates the expression of YAP1 through PI3K in various cancers including PDAC, ovarian, breast, lung, and colon cancer (Hao *et al*., 2017; Luo and Yu, 2019; Szulzewsky *et al*., 2021). By regulating various pathways, YAP1 is involved in epithelial-to-mesenchymal transition (EMT) in PC cells. ZEB1, SMADs, and TGF-β are important nuclear factors that interact with YAP/TAZ and involve in EMT (Lehmann *et al*., 2016; Narimatsu *et al*., 2016). TGF-β induces EMT in cancer cells by promoting the localization and stabilization of YAP to the nucleus however in PDAC cells this activity can be blocked by AKT pathway inactivation (Gao *et al*., 2021).



### 1.21 Cross-linking role of *YAP1* and *p53* in PDAC:

According to Raj and Bam, (2019), in response to DNA damage both the *YAP1* and *p53* corporate and help in maintaining the integrity of the genome by reprogramming transcriptional processes. However, any genetic disturbances in these genes contribute to the tumorigenesis of different organs. Due to the absence of the PPxY motif in the *p53* gene, YAP1 can't directly interact with wild-type p53 therefore YAP1 regulates the functioning of *p53* through other mediators such as MDM2, PTPN14, and LATS1/2 (Raj and Bam, 2019). In response to various stresses (like an expression of an oncogene, failure in cytokinesis, and replication stress), YAP1 coordinates with p53 through LATS1/2, which binds to MDM2 and causes its inhibition which helps in p53 stabilization leading to apoptosis (Furth and Aylon, 2017). YAP1 can also bind to the promoter of the *p53* gene and enhanced its expression and vice versa (Bai *et al*., 2013). Recently in one study, YAP1 was found to be physically interacting with mutant *p53* leading to increased expression of various oncogenes and also improving the proliferative ability of mutant *p53* (Di Agostino *et al*., 2016; Raj and Bam, 2019) as shown in Figure 1.17.

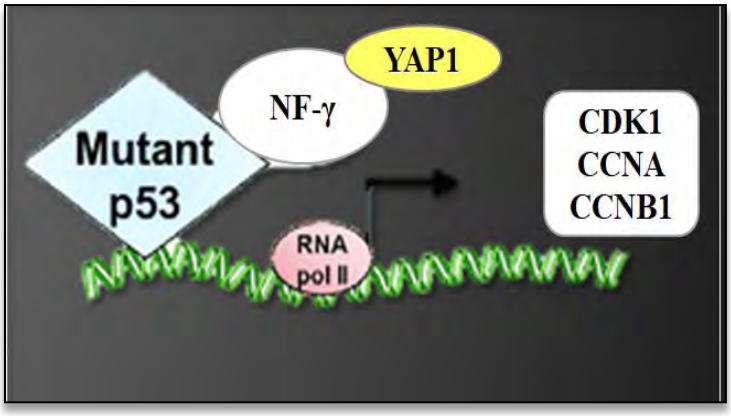


**Figure 1.17:** Mutant p53 and YAP1 interaction through NF-γ for activating various pro- oncogenes (Raj and Bam, 2019).

Under stress conditions, stabilized wild-type *p53* increases the PTPN14 level resulting in inhibition of YAP1 in PC. However, a study reported that deficiency of p53 reduces PTPN14 levels leading to YAP1 activation in KRAS mutant PDAC (Mello *et al*., 2017; Murakami *et al*., 2019) as shown in Figure 1.18.

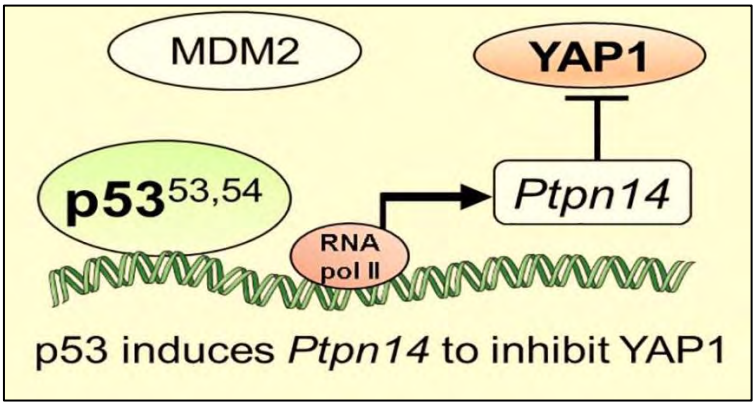


**Figure 1.18:** Inhibition of YAP1 by increased expression of PTPN14 via stabilized p53 (Raj and Bam, 2019).

### 1.22 Significance of Study:

Pancreatic cancer is a fatal malignancy but the research in Pakistan is almost negligible. Currently, there is a need for more research on the molecular aspect of PC in Pakistan (Ali *et al*., 2021). For PC early detection, it is very important to determine different markers. So our study is mainly based on genetic analysis of pancreatic cancer tissue using whole-exome sequencing which will help to determine the role of different genes including *p53* and *YAP1* in causing pancreatic ductal adenocarcinoma.

This will further help for the determination of different markers for the development of diagnostic tools and for conducting drug trials that can be used for early cancer detection and also help to plan better treatment strategies.

### 1.23 Aims and Objectives:

The main aim of this research is to identify the role of *p53* and *YAP1* in PDAC with the following objectives:

1. Genetic characterization of PDAC sample through whole exome sequencing among Pakistani population.
2. Identification and Analysis of disease-causing variants using bioinformatics tools.
3. To determine the role of variants of *p53* and *YAP1* genes in PDAC.

# 2 Materials and Methods

The current study was designed to find out novel single nucleotide variations (SNVs) among different genes and to determine their probability of causing pancreatic cancer using insilico tools among the Pakistani population. This research work was conducted at Cancer Genetics Lab (CGL), Department of Biochemistry and Molecular Biology, Quaid-i-Azam University, Islamabad, Pakistan.

## 2.1 Study Plan and Ethical Approval:

For conducting this study the approval was acquired from the Institutional Review Board (IRB) of Quaid-i-Azam University Islamabad, Pakistan. Along with this, sample collection approval was also obtained from the board of directors of the Pakistan Institute of Medical Sciences (PIMS). A consent form was designed for getting approval from patients to use their reports and details for research purposes. All the patients were requested to fill out the questionnaire that was then kept as a record. For this research, the study plan is a Case-Control Study for which fresh tissue samples, (tumor, T and control, C) of Pancreatic Cancer (PC) patients were collected from September 2021 to April 2022 from the Pakistan Institute of Medical Sciences (PIMS), Holy Family, Quaid-e-Azam International Hospital (QIH).

## 2.2 Tissue and Blood Sample Collection:

For this study, fresh tissue samples of both tumor and control were collected soon after the Whipple surgery or biopsy of the PC patient. Due to the rapid prognosis of PC to advance stages and late diagnosis, the frequency of getting the sample is quite low so we collected 20 samples (n=20) (including 10 Tumor and 10 Control samples) for this study. These samples were collected in Cryo-Tubes which were then placed in the flask containing liquid nitrogen so that DNA and RNA degradation can be prevented by freezing the samples immediately. Then, these samples were stored at -80˚C until used for further experiments. Out of these 20 samples, 6 (3 Tumor, 3 Control) was used for whole exome sequencing rest 14 samples was kept for further analysis. By using 5ml syringes blood samples from PC patients were collected in anticoagulant Ethylenediamine Tetra Acetic Acid (EDTA) containing Vacutainers (Atlas-labovac, Italiano) that were then kept at -20˚C.

## 2.3 Extraction of Genomic DNA from Tissue Samples:

### 2.3.1 Preparation of Solutions for gDNA Extraction:

**Table 2.1:** Reagents List and their Chemical Composition.

| Sr.No | Solutions | Chemical Composition |
|---|---|---|
| 1. | **70% Ethanol** | • Absolute Ethanol (70mL)<br>• Distilled water (30mL) |
| 2. | **Tail Lysis Buffer** | • 100mM EDTA (2.92g ) , pH 8.0<br>• 10mM Tris HCl (0.12g), pH 8.0<br>• 50mM NaCl (0.29g)<br>• 0.5 %SDS (0.5g)<br>• Distilled water (100ml) |
| 3. | **Proteinase K** | • Proteinase K (0.1g)<br>• 10mM Tris HCl (50µl)<br>• 20mM CaCl2 (2000µl)<br>• Glycerol (5ml)<br>• Nuclease-free Water (5ml) |
| 4. | **Phenol: Chloroform: Isoamyl Alcohol** | • 25:24:1 |
| 5. | **Sodium Acetate (pH 5.2)** | • 3M Sodium Acetate (12.31g)<br>• Distilled water (50ml) |
| 6. | **Tris-EDTA buffer (pH 8.0)** | • 1mM EDTA (0.029g)<br>• 10mM Tris hydroxyl (methylamino) methane (0.12g) in Distilled water (100ml) |
| 7. | **Sodium Dodecyl sulphate (SDS)** | • 20% SDS (10g)<br>• Distilled Water (50ml) |



### 2.3.2 Procedure:

DNA extraction was done by using the phenol-chloroform method (Ghatak *et al*., 2013) from 20mg of both control and tumor tissue samples by

using the following steps:

- **Addition of Lysis Buffer:**

For this organic method, 20 mg of both the control and tumor tissue samples were added in different Eppendorf tubes. All samples were homogenized after adding 500µl of lysis buffer and then placed in an incubator shaker for 20-30 minutes at room temperature for proper homogenization. After incubation centrifugation was done for separating the phase at 13000 rpm for 3min. Soon after centrifugation, two layers appeared, the upper layer was aqueous and the lower layer contains organic content. From each eppendorf upper layer was discarded while the lower layer containing DNA was further washed multiple times using lysis buffer to avoid any contamination. In the pellet 400 µl of lysis buffer was again added along with 13 µl of 20% SDS (Sodium Dodecyl sulphate) and 25 µl proteinase K. After adding these reagents tubes were incubated at 37°C overnight.

- **DNA Clean Up (Phenol: Chloroform: Isoamyl Alcohol Treatment):**

For complete digestion of cells, 500 µl of phenol, chloroform, and isoamyl alcohol solution were added to each tube and inverted several times. Then centrifugation was done for 10min at 13000 rpm for proper mixing. The upper layer was carefully picked and transferred to other eppendorf tubes by micropipette for extraction and purification of DNA while the pellet was discarded. Again 500 µl of chloroform and isoamyl alcohol (24:1) were added to the aqueous layer and centrifuged at 13000 rpm for 10min and the upper layer was picked and transferred into the centrifuge tube.

- **Precipitation of DNA:**

In tubes containing an aqueous layer, 55 µl of sodium acetate and 500 µl of chilled isopropanol were added and inverted for precipitation of DNA. Then, for incubation samples were kept at -20˚C for 45min. Later on, centrifugation was done at 13000 rpm for 10min and supernatant was discarded gently leaving the pellet behind.

- **DNA Washing:**

For washing, in each tube containing pellet 500 µl of 70% ethanol was added and centrifugation was done at 7500 rpm for 5min. Then, the supernatant was discarded

and in the pellet, this process was repeated 2-3 times to wash off all impurities. After that, the pellet was air dried.

- **Storage of DNA:**

After completely drying the pellet TE Buffer (Tris EDTA) was added for pellet resuspension and tubes were stored at 4˚C. For longer use, these tubes can be stored at -20˚C (Ghatak *et al*., 2013). This freshly isolated DNA was then used for further quantitative and qualitative analysis and after a quality check was used for Whole Exome Sequencing (WES).

**2.4 Genomic DNA Quantification:**

For quantitative analysis of DNA, Thermo scientific Multi Skan Go Instrument was used. Absorbance was taken at 260nm and 280nm and if the optical density (O.D) value for DNA falls between 2-1.8 then the DNA was selected as of good quality.

**2.5 Genomic DNA Qualitative Analysis:**

The extracted genomic DNA qualitative analysis was done by agarose gel electrophoresis in which all 6 samples were run twice (12 bands have to appear) in 1% agarose gel.

**Table 2.2:** Chemical Composition of Reagent used in Agarose Gel Electrophoresis.

| Sr. No. | Chemicals | Composition |
|---|---|---|
| 1. | **1% Agarose Gel** | • 1g Agarose<br>• 1X TAE buffer (100 ml) |
| 2. | **10X TAE** | • Tris-Base<br>• Acetic Acid<br>• EDTA<br>• Distilled Water |
| 3. | **1X TAE** | • 10X TAE (10ml)<br>• Distilled Water (90ml) |

| | | |
|---|---|---|
| 4. | **Ethidium Bromide (EtBr)** | • EtBr (400mg)<br>• Distilled Water (20ml) |
| 5. | **Bromophenol Blue** | • Bromophenol blue (0.25%)<br>• Sucrose (40%)<br>• Distilled water (100 ml) |

### 2.5.1 Agarose Gel Electrophoresis Procedure:

1. For qualitative analysis gel electrophoresis was done by using 1% agarose gel.
2. Agarose gel was prepared by dissolving 1g of agarose in 100ml of 1X Tris Acetic acid EDTA (TAE) buffer.
3. To dissolve agarose the solution was heated in the microwave for 2min.
4. A clear solution was obtained in which 7 µl of EtBr was added which helps in locating DNA in the gel.
5. This gel solution was poured into the gel casting tray and after inserting combs gel was left at room temperature for 25-30min for complete solidification.
6. After solidification of gel, combs were removed carefully and gel caster was picked and placed in the gel tank having 1X TAE buffer.
7. 6 µl of extracted genomic DNA was mixed with 2 µl of 6X bromophenol blue dye (loading dye) and was loaded in wells after thorough mixing.
8. After setting the parameters, 500mA current with 75 volts the gel was run for 60 minutes.
9. UV Trans-Illuminator Bio Doc Analyzer was used to visualize the bands that appeared on the gel.



## 2.6 Whole Exome Sequencing (WES):

For genome-wide analysis, WES was done using Illumina HiSeq X Platform which is a very convenient and cost-effective method as compared to whole genome sequencing. WES is the sequencing technique that only target protein-coding regions (exome) of genes having susceptibility to carry variants related to diseases. This technique can be used for various studies including complex disease, cancer research, or human population studies. This method is based on next-generation sequencing

(NGS) and provides detailed coverage of coding region. It produces high quality and manageable data for convenient analysis.

**2.7 Bioinformatics Analysis:**

First of all, by base calling raw sequence data (short reads) was obtained in fastq format. This raw data was then used to detect variants and for their annotation using the GATK variant calling pipeline in which the MuTect2 tool was used for somatic variants annotation. Later on, after variants annotation, the vcf file was obtained having results of various tools that were then used for further analysis. The Figure 2.1, flow chart provides an outlook of how the data was interpreted starting from raw short reads to variants annotation:

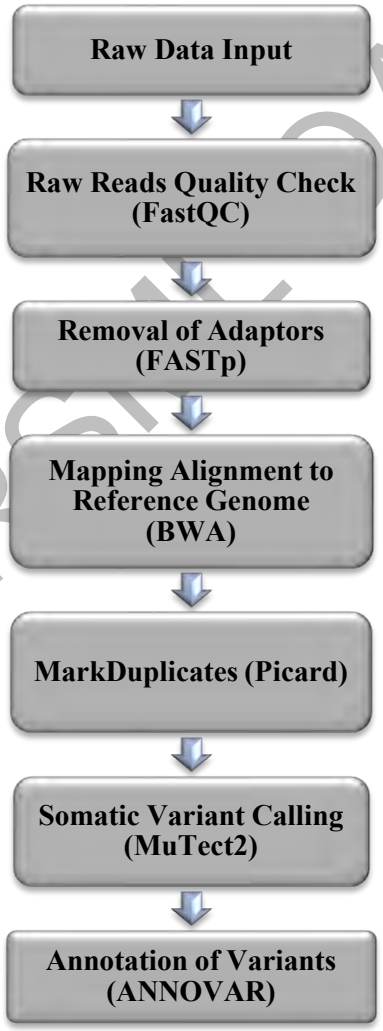




**Figure 2.1:** Flow Chart of Data Analysis using Bioinformatics Tools.

### 2.7.1 Raw Data Input:

Raw data in the form of short reads were obtained in FastQ format by base calling in the Linux environment for convenient data processing. Sequencing can be either single-end or pair-end sequencing.

### 2.7.2 Quality checking of Samples:

Once data was obtained in FastQ format then quality checking of samples was done by the FastQC tool. After applying this tool an html report was generated including details regarding different parameters analyzed for quality check.

### 2.7.3 Removal of Adaptors:

After quality checking FastP tool was applied to the sequence file for removing any adapter's content present in sequences (Chen *et al*., 2018). Adaptor's content can also be removed by using a timometric tool.

### 2.7.4 Mapping against Reference Genome:

Mapping alignment of sample sequence to the indexed reference genome is very important which was done by using Burrows-Wheeler Aligner (BWA) tool. Hg38 was the human reference genome used in this study which was in fasta format and the sample sequence is in Fastq format. After mapping an aligned BAM format file was obtained.

**Table 2.3:** Reference genomes of all organisms used for alignment

| Species | Reference Genome | |
|---|---|---|
| **Human** | **Hg38** | Homo_sapiens.GRCh38.dna.primary_assembly.fa (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.39/) |

### 2.7.5 Marking the Duplicates:

After alignment, PCR duplicate marking was performed using MarkDuplicates (Picard) tool that used BAM aligned files as an input and provides a new BAM file as an output which had a SAM flags field for each read identifying the duplicates.

### 2.7.6 Variants Calling by MuTect2:

For the detection of somatic variants, the MuTect2 tool of Genome Analysis Toolkit (GATK) software was used. For the detection of somatic mutation, this tool provides the highest rate of validation (90%). After sample sequence alignment with the reference genome and other initial processing steps data was analyzed by MuTect2 by comparing the tumor sample data with their matched normal control sample data (Cibulskis *et al*., 2013).

For normal control samples Panel of Normals (PoN) was created which was used for better analysis of somatic variants. Different PONs will be created for different variants type but all of them will be created only for normal samples (healthy tissues that don't have any somatic variations) not for tumor samples (Chen *et al*., 2018). To enhance variant calling analysis results PONs capture recurrent technical artifacts (Benjamin *et al*., 2019). For each tumor sample, a separate raw vcf file was generated as a final output file by the MuTect2 tool of variant calling.

### 2.7.7 Variants Annotation:

By using the ANNOVAR tool, variants annotation was performed in which the raw vcf file was converted into annotated vcf file. To interpret the functional effects of the genetic variation from high throughput sequencing data this tool is very rapid and efficient. As an output three files was generated including a .vcf file, .txt format, and .avinput file. In the vcf file, the field is populated with ANNOVAR annotations, while the .txt file has limited tabs and is generated in an excel sheet that is easier to interpret, and the .avinput file has some extra columns that include comments on different variants. An annotated vcf .txt file includes results from many tools including Scale-invariant feature transform (SIFT), Polymorphism Phenotyping (PolyPhen-2), Protein Variation Effect Analyzer (PROVEAN), Mutation Tester, Mutation Assessor, and Combined Annotation Dependent Depletion (CADD raw and phred) that was then used for further analysis.

### 2.8 AvSNP 150:

AvSNP150 datasets are the reformatted form of dbSNPs dataset that helps to match user data indels (insertions and deletions) against dbSNPs assigned rs ID. It is involved in assigning rs IDs to new SNPs variants with detailed research. AvSNP150

was used to observe whether any rs ID was assigned to novel variants or not in this study.

**2.9 Damage Pred Count:**

Damage pred count/ ClinPred were used to identify non-synonymous single nucleotide variants (SNVs) that are related to PC disease. This tool is highly specific and sensitive and is based on the ClinVar database that permits the use of disease-causing variants as a training set. If the ClinPred score/ Damage pred count is less than 0.5 then the disease-causing variant is considered tolerant (T) while if it is more than 0.5 then the variant is damaging (D) (Alirezaie *et al*., 2018).

**2.10 SIFT, PROVEAN, and PolyPhen2:**

- **SIFT** was used to analyze the effect of amino acid substitution on the functioning of protein based on amino acid's physical properties and homology. If SIFT score is lower than 0.05 then amino acid substitution is considered to have a damaging (D) effect on protein functioning while if the score is greater than 0.05 then amino acid substitution will be tolerant (T) and have no effect on protein functioning (Sim *et al*., 2012).
- **PROVEAN** software tool was used to observe how the function of protein is affected by variation in amino acid sequence. It provides information for all types of variations including single or multiple amino acid substitutions (Choi *et al*., 2012). In PROVEAN, based on threshold function category of protein is decided either as neutral (N) or deleterious (D) (Sandell and Sharp, 2022).
- **PolyPhen2** was also used to further validate the results which explore the effect of amino acid substitution on the basic structure and function of protein considering physical characteristics (Adzhubei *et al*., 2013). PolyPhen2 includes two training sets (HumDiv, HumVar) for which if the score is 0.85-1 then the substitution is considered as most probably damaging (D), if the score is 0.15-1 then substitution is possibly damaging (P) while if the score is 0.0-0.15 than substitution is benign (B). SIFT and PolyPhen2 prediction scoring are contrary to one another like a variant is benign if the PolyPhen2 score is 0 while the SIFT score is 1.

**2.11 MutationTaster, Likelihood Ratio Test (LRT), and Mutation Assessor:**

- **MutationTaster** was used to evaluate the nucleotide variants including synonymous or intronic variants for their disease-causing ability. It represents the variations in four ways deleterious/ disease-causing automatic (A), disease-causing (D), probably harmless polymorphism (N), and harmless polymorphism automatic (P).
- **LRT** was used to observe nucleotide variants that can affect highly conserved amino acids. This tool helps to predict variants such as deleterious (D), neutral (N), and unknown (U).
- **Mutation assessor** was used for studying the amino acid substitution effect on the functionality of proteins based on the conservation of the affected amino acid in homologs of protein. If the substitution has a functional impact then represented as high (H) or medium (M) while if there is no functional impact of substitution then represented as low (L) or neutral (N) (Reva *et al*., 2011).

**2.12 Variant Effect Scoring Tool (VEST4):**

VEST4 was used for identifying the effect of missense mutations on the activity of protein based on the pathogenicity of mutations. If the score is high (ranging 0-1) then the change will be more deleterious (Carter *et al*., 2013).

**2.13 CADD Raw, CADD Phred:**

The combined Annotation Dependent Depletion (CADD) tool was used to analyze the deleteriousness of SNVs including indels in the human genome. CADD raw scores are positive if the variant is deleterious (D) while scores are negative if the variant is neutral (N) (Rentzsch *et al*., 2019). CADD Phred score is 10 or more than raw score is in top 10% of reference genome SNVs while if the Phred score is 20 or more than raw score is in top 1%of reference genome (Rentzsch *et al*., 2019).

**2.14 Cosmic 92:**

The catalogue of somatic mutation in cancer (Cosmic) 92 is the world's largest database that was used to explore the effect of somatic mutation on various types of cancers. A genomic mutation ID is assigned to reported variations that have an impact on any type of cancer.

### 2.15 Clinical Significance (CLINSIG):

The clinical significance database was used to check the possible role of variants in other diseases. It is based on the ClinVar database and different clinical significance values will be assigned to variants including benign, likely benign, pathogenic, likely pathogenic, protective, and several others.

### 2.16 Phyre2:

Phyre2 tool was used for protein 3D structure modeling that helps to analyze its function and also helps to observe how variations in protein amino acid sequence can affect the structure and function of proteins (Kelley *et al*., 2015). ChimeraX was used to analyze and compare the structure of a mutated protein by using normal protein 3D structure.

# 3 Results

## 3.1 Study Design and Demographic Detail:

For this study, 20 Tissue samples (10 Tumor, 10 Control) were collected from PDAC patients. All the patients were requested to fill out the questionnaire. Data including demographic details were collected from hospital records. From tissue samples, DNA was extracted using the Phenol-chloroform method, and DNA bands were visualized on 1% agarose gel that was used for qualitative analysis of DNA

### 3.1.1 Details of PDAC Patients:

The following graph in Figure 3.1 shows the details of all 10 samples of PDAC patients collected from questionnaires and reports data.

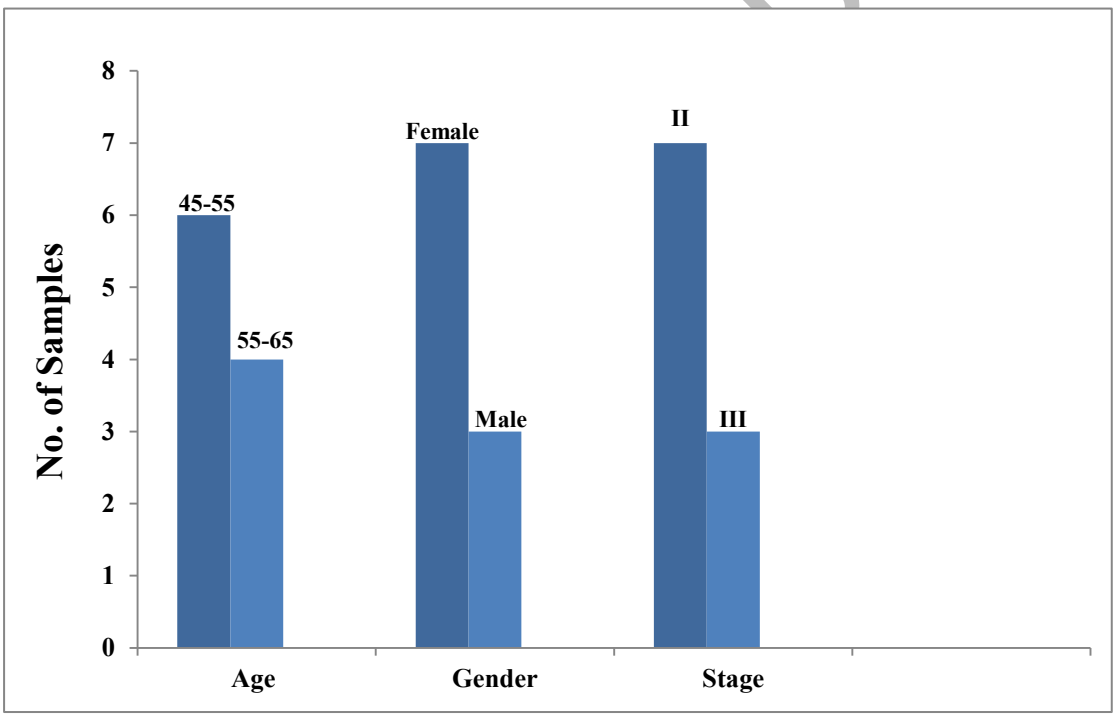


**Figure 3.1:** PDAC Patients Details.

The following Table 3.1 shows the details of 3 PDAC patients sample that was used for whole exome sequencing.

**Table 3.1:** PDAC Patients Details used for WES.

| Sample No. | Gender | Age | Histo-pathological Difference | Symptoms | Comorbidity | Hereditary Diseases | Treatment |
|---|---|---|---|---|---|---|---|
| S1 | M | 56 | Poor | Abdominal Pain, Jaundice, Vomiting | None | None | Whipple (Folforinox) |
| S2 | F | 50 | Poor | Weight Loss, Abdominal Pain | None | None | Colostomy, Whipple Surgery |
| S3 | F | 60 | Poor | Weight Loss, Abdominal Pain, Jaundice | Diabetes, Jaundice | Diabetes | Stenting, Whipple Surgery |

### 3.1.2 Graphical Representation of Stages of PDAC Patients:

Depending on the PDAC disease progressiveness and resectability, it was divided into four stages. As shown in Figure 3.2, Out of three patients, 2 patients were at stage II while 1 patient was diagnosed at stage III. Due to failed early diagnosis, no sample was collected from stage I. Moreover, no sample was collected from stage IV PDAC patients because the tumor is non-resectable due to its aggressiveness and the patient survival rate is less.

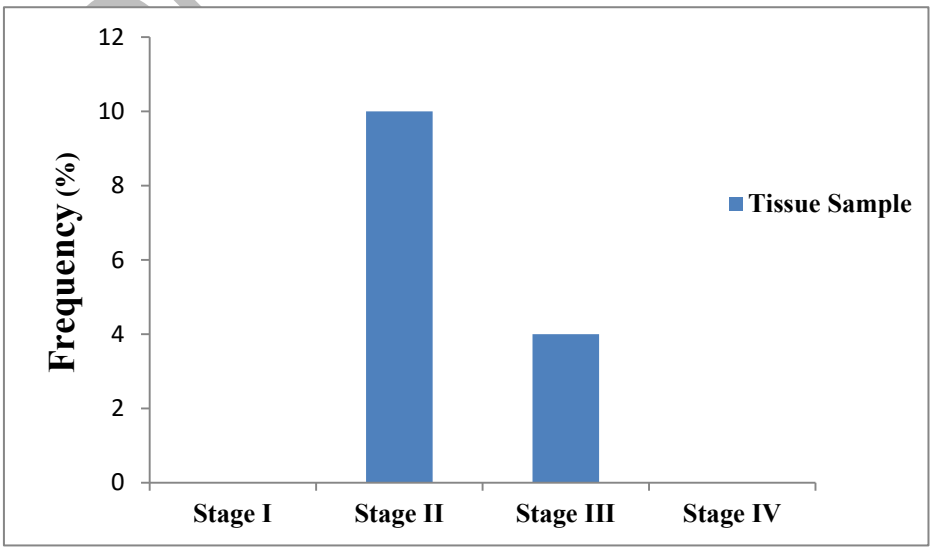


**Figure 3.2:** Cancer Stages Frequency in PDAC Patients.

### 3.2 Quantitative Analysis of genomic DNA:

The 260/280 absorbance ratio taken by the DNA Thermo scientific Multi Skan Go Instrument was between 1.8-1.9 which means that the DNA is of good quality and the nucleic acid concentration was between 650-850ng/ul as shown in Table 3.2. All 6 samples (including 3 tumors and 3 control) were run in duplicates (n=12) to enhance results accuracy.

**Table 3.2:** Optical Density (O.D) value and Concentration in ng/ul of Nucleic Acid.

| Sr.No | Nucleic Acid 260/280 | Nucleic Acid Conc. in (ng/ul) |
|---|---|---|
| 1 | 1.89 | 770 |
| 2 | 1.83 | 780 |
| 3 | 1.79 | 720 |
| 4 | 1.81 | 773 |
| 5 | 1.88 | 780 |
| 6 | 1.82 | 719 |
| 7 | 1.85 | 831 |
| 8 | 1.83 | 777 |
| 9 | 1.84 | 704 |
| 10 | 1.88 | 810 |
| 11 | 1.80 | 790 |
| 12 | 1.87 | 650 |



### 3.3 Qualitative Analysis of Genomic DNA:

On 1% agarose gel, along with ladder and control DNA, 12 bands appeared because all 6 samples (3T, 3C) were run in duplicates. 1KB ladder was loaded in 1$^{st}$ wells with

DNA samples in adjacent wells. High-quality bands can be seen on gel because extracted DNA from tissue samples was highly intact and had a size of more than 20kb. Clear DNA bands with comparison to 1KB Ladder were shown in Figure 3.3 below:

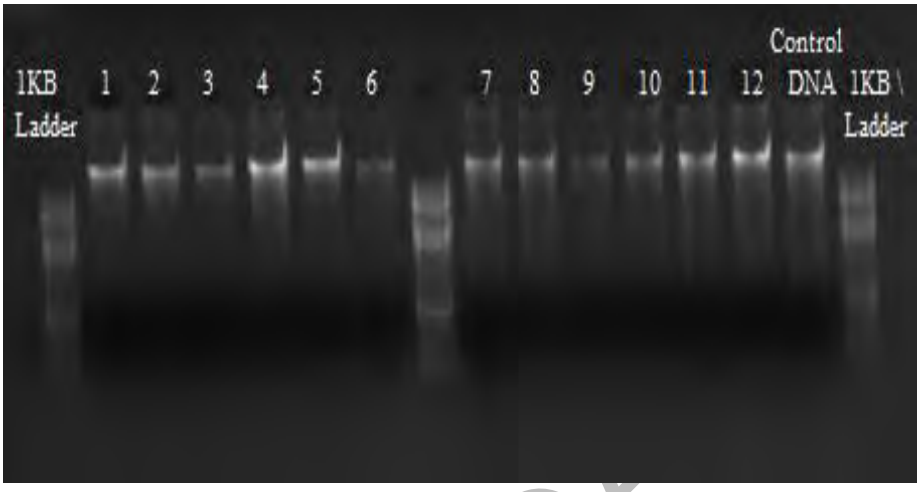


**Figure 3.3:** 1% Agarose Gel Stained by EtBr Showing Genomic DNA Bands from Tissue Samples.

### 3.4 WES Result Analysis:

WES produces a very high quality and manageable data for convenient analysis. For analyzing WES results variety of bioinformatics tools were used.

#### 3.4.1 Quality Checking of Samples Raw Reads:

After pair-wise sequencing once raw reads were obtained in which adapters sequence was present. By using the fastp tool raw reads quality was checked and adaptors were removed from the sequence files. Given below was the report of FastQC results which showed that the total sequences were 47627388 and the GC content was 56%.

**Basic Statistics**

| Measure | Value |
|---|---|
| Filename | S1ACHN-1_R1.fastq.gz |
| File type | Conventional base calls |
| Encoding | Sanger / Illumina 1.9 |
| Total Sequences | 47627388 |
| Sequences flagged as poor quality | 0 |
| Sequence length | 31-150 |
| %GC | 56 |

**Figure 3.4:** Basic Statistics of Raw Reads produced by Base Calling using fastQ Tool.

In the FastQC result report, this quality score graph was used to represent per base sequence quality in which the y-axis shows the quality score, the higher score of which shows a better base call. As shown in Figure 3.5, in our case, the blue line was in the green part of the graph which represented very good quality base calls.

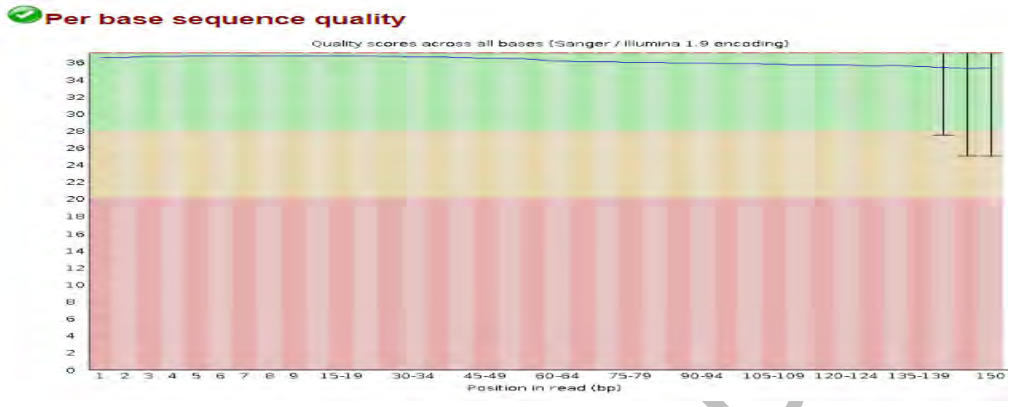


**Figure 3.5:** Quality Score Graph Represented in FastQC Result Report.

### 3.4.2 Adapter Content Analysis:

In the FastQC result report, the adaptor content graph in Figure 3.6 showed a straight line at the x-axis means that no adapter content was present which is further improving the quality of data.

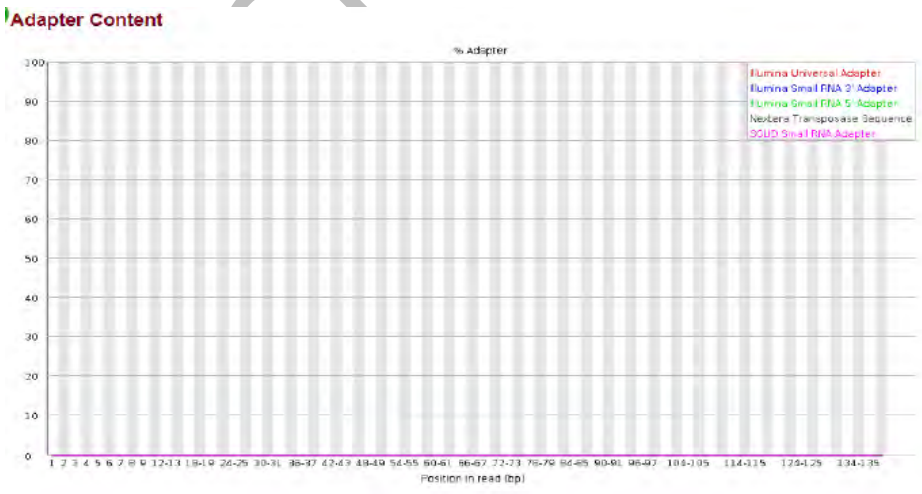


**Figure 3.6:** Adapter Content in FastQC Result Report.

### 3.4.3 Detection of PDAC Variants:

Sample IDs including S1APC, S7APC, and S8APC was assigned when both tumor and control PDAC tissue sample was aligned by Mutect2 (GATK) to find out novel variants in all samples. For alignment Panel of Normals was first created using a control tissue sample. In total 2581210 variants were identified in all 3 samples along with that the number of variants in each sample was given below:

**Table 3.3:** No. of variants identified in each PDAC sample

| Species | Sample ID | Number of Variants |
|---|---|---|
| **Homo Sapiens** | S1APC | 1590779 |
| | S7APC | 356593 |
| | S8APC | 633838 |

## 3.5 Annotated Variants Data Analysis:

Three annotated vcf files were generated for 3 samples by the ANNOVAR tool that has many variants in the exonic region of different genes including *KRAS*, *SMAD4*, and *BRCA1/2*. Overall 4 variants were reported in the exonic region of the *p53* gene, out of which 3 variants (87341661; T>C, 87341667; A>T, 87345180; G>C) were novel found in *p53TG1* coding for long non-coding RNA (nc-RNA), while 1 variant found in the exonic region of *p53* gene was already reported (rs397516435) at position 7674945 on chromosome 17 in which G>A substitution results in stop-gain mutation replacing arginine (R) with a stop codon. However, approximately 280 variations were reported in the intronic region of the *p53* gene including 182 single nucleotide variations (SNVs) and 98 InDels (Insertion and Deletions). These intronic variations include 146 variants in the intronic region, 113 variants in the intergenic region, 15 variants in 5′UTR, and 6 variants in the downstream region of a gene.

On other hand annotated variants analysis has not shown any variation in the exonic region of *YAP1* in our population but more than 330 variations were seen in the intronic region of the *YAP 1* gene including 242 SNVs and 88 InDels and have 212 variants in the intergenic region, 107 variants in the intronic region, 2 variants in 3′ UTR, 7 variants in the downstream region, 1 variant in upstream and 5′UTR.

### 3.5.1 Nucleotide Variations in *p53* gene:

In the *p53* gene, 1 nucleotide variant at position 7674945 as shown in Figure 3.7 were reported in an exon 6 of the *p53* gene in which G is substituted by A as shown in Figure 3.8 and 3.9:


**Figure 3.7:** Location of exonic variant of *p53* gene.

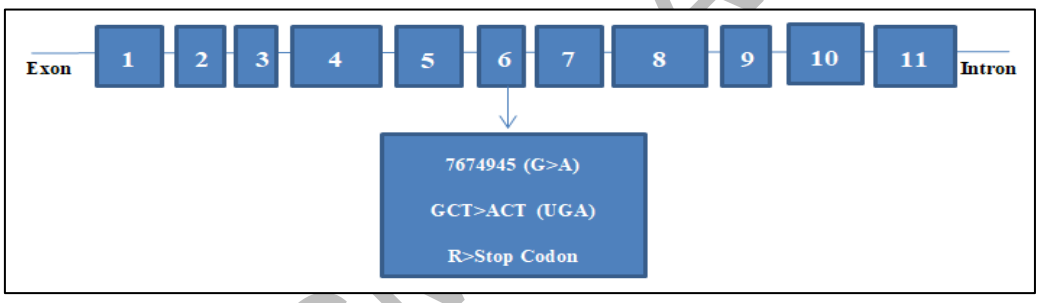


**Figure 3.8:** Nucleotide Variation in Exon 6 of *p53* gene on Chromosome 17.

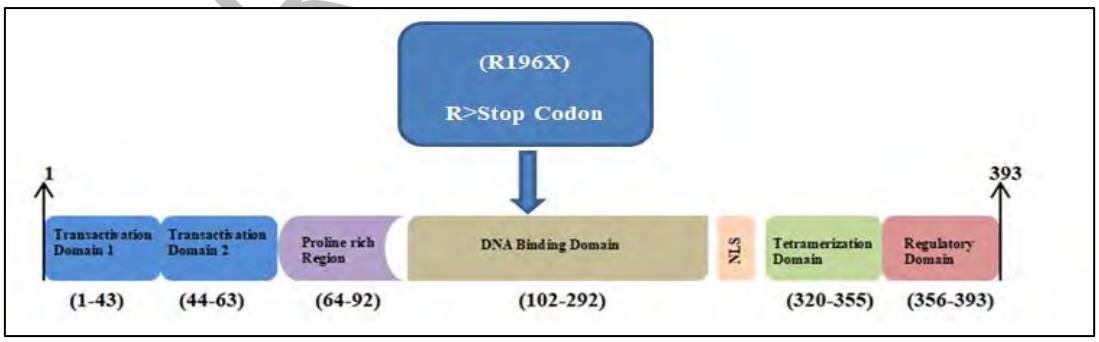


**Figure 3.9:** Location of exonic variant on *p53* protein domain.

The following graph represents the frequency of SNVs in the intronic, intergenic, upstream, the downstream regions of the *p53* gene and showed that the majority of variations were reported in the intronic region. The frequently observed nucleotide substitution in the intonic region of the *p53* gene was of C>T as shown in Figure 3.10.

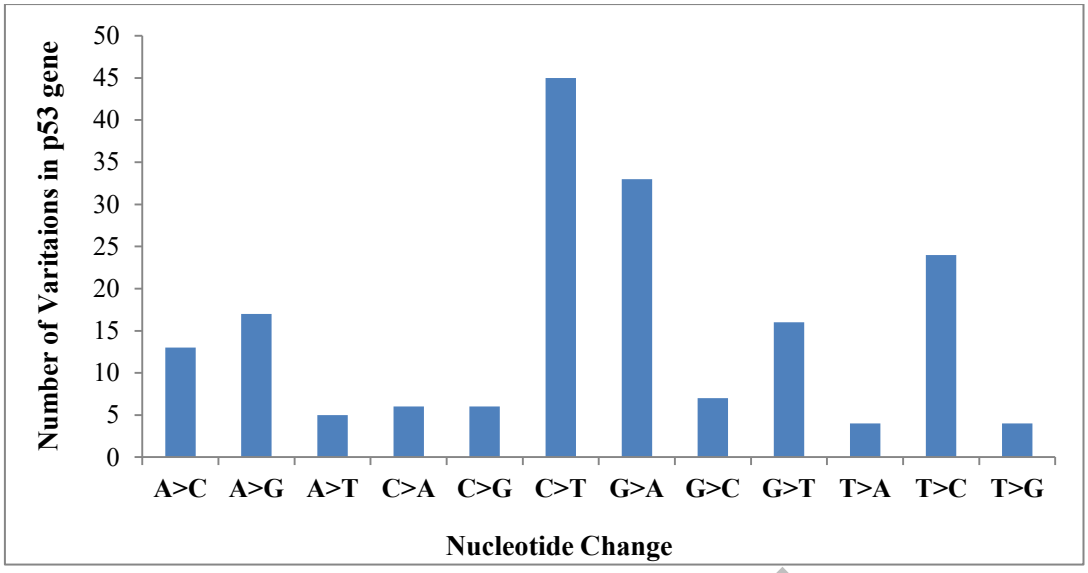


**Figure 3.10:** Various Single nucleotide variations (SNVs) in the *p53* gene

### 3.5.2 Nucleotide Variations in *p53TG1* gene:

Overall, 3 nucleotide variants were reported in the non-coding RNA exonic region of the *p53TG1* gene on chromosome 7 as mentioned below in Figure 3.11 and 3.12:

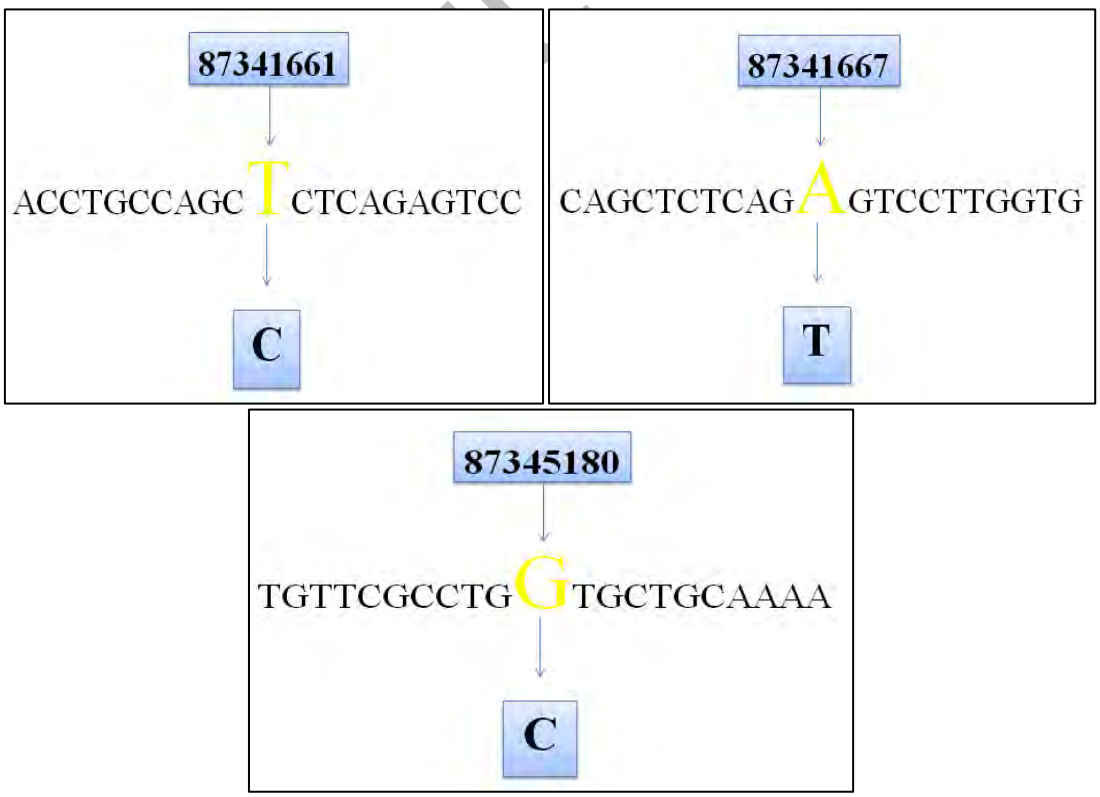


**Figure 3.11:** Location of 3 non-coding exonic variants of *p53TG1* gene.

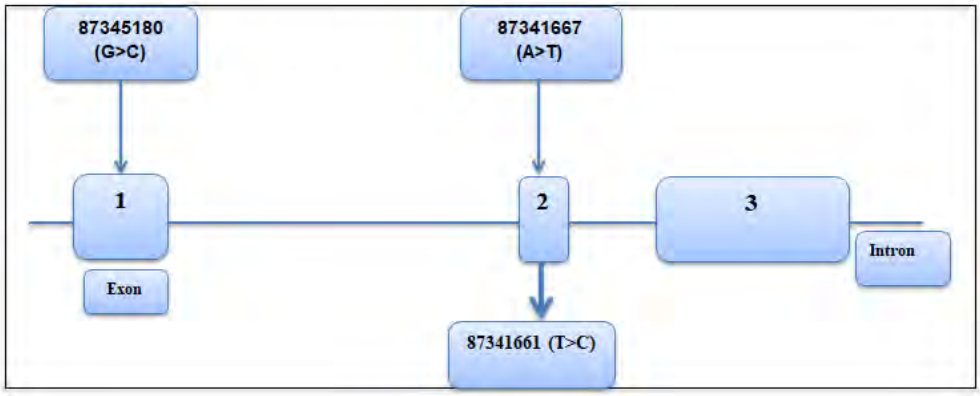


**Figure 3.12:** Nucleotide Variation in Exon 1, 2 of *p53TG1* gene on Chromosome 7.

### 3.5.3 Nucleotide Variations in *YAP1* gene:

In an exonic region of the *YAP1* gene, no nucleotide variation was observed but many variations were seen in its intronic region. The following graph shows the frequency of SNVs in the intronic, intergenic, upstream, and downstream regions of the *YAP1* gene and showed that most variations were in the intronic region. The frequently observed nucleotide variation in the intonic region of the *YAP1* gene was G>A as shown in Figure 3.13.

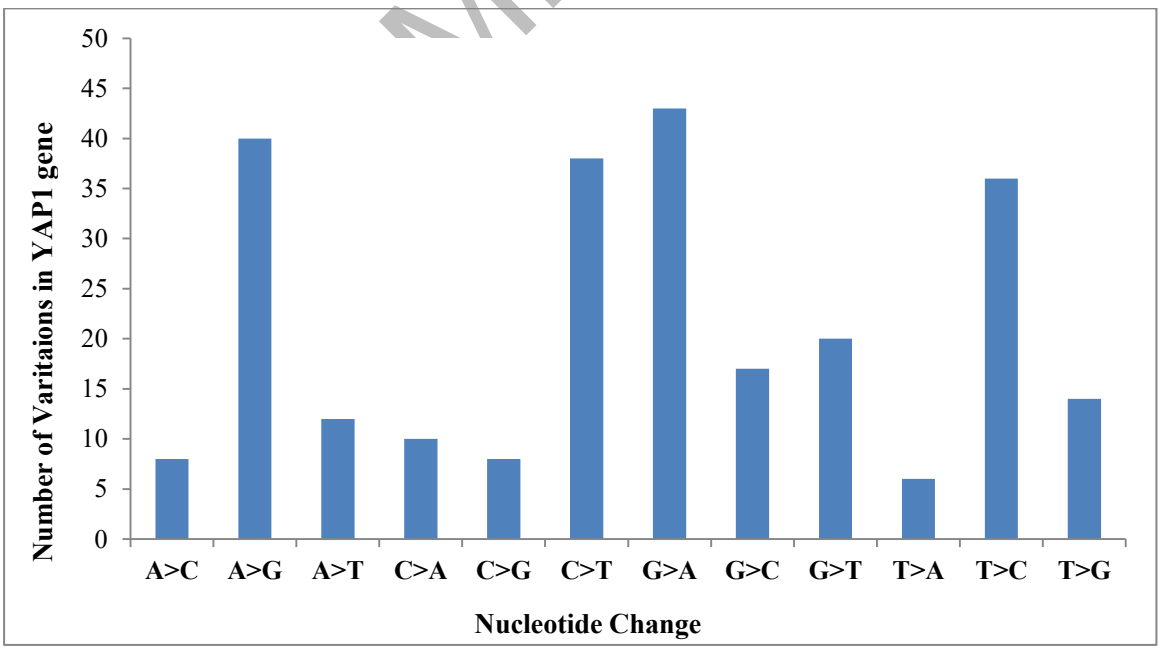


**Figure 3.13:** Various Single nucleotide variations (SNVs) in the *YAP1* gene.

### 3.5.4 Nucleotide Variations in Driver Genes:

Out of 4 driver genes, no variations were observed in the exonic region of the *CDKN2A* gene while 1 variation was observed in the exonic region of the *p53* gene as mentioned above (Figure 3.8). Furthermore, as shown in Figure 3.14 and 3.15, 1 variation (25209843; A>G) was observed in exon 5 of the *KRAS* gene which were already reported with an SNP ID rs1137282, and 1 variation (51058160; G>A) was reported in exon 6 of the *SMAD4* gene,.

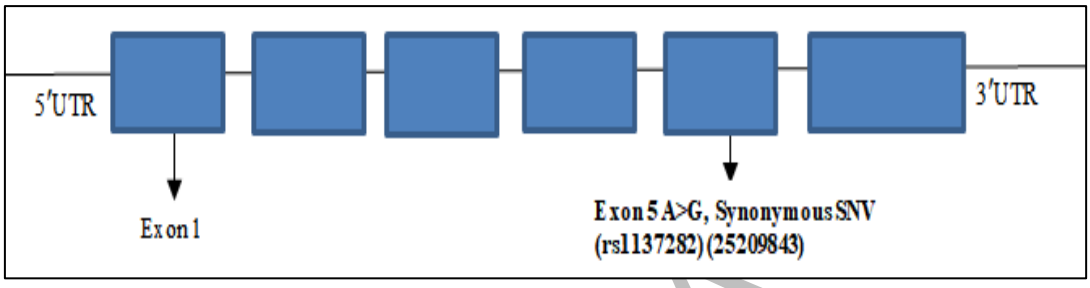


**Figure 3.14:** Nucleotide Variation in Exon 5 of *KRAS* gene.

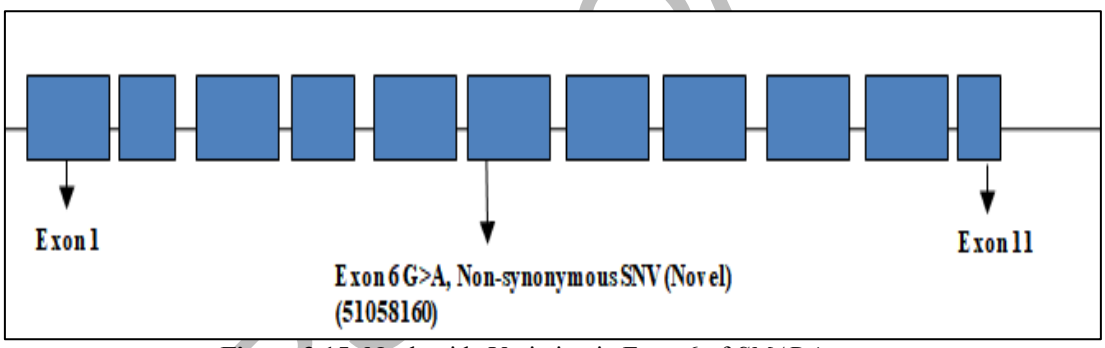


**Figure 3.15:** Nucleotide Variation in Exon 6 of *SMAD4* gene.

### 3.5.5 Variant Probing by AvSNP150 Database:

Out of 4 variants (3 nc-RNA and 1 exonic variant) of the *p53* gene, no ID was assigned to 3 nc-RNA variants by the AvSNP150 database which shows that no data regarding these variants is available in this database, thus these 3 variants of *p53* gene are considered as novel. While the SNP ID for 1 exonic variant of the *p53* gene (rs397516435) is available on the database means that this variant was already reported.

### 3.5.6 Variants Damage Potential Analysis by ClinPred:

For 3 of 4 variants of the *p53* gene no data regarding damaging or tolerant nature was found in ClinPred representing that no information is available in the ClinVar

Database regarding these variants hence depicting the novelty of these variants. However, the 1 exonic variant (rs397516435) has a ClinPred Score of more than 0.5 showing that the variant is damaging and disease-causing.

### 3.5.7 Amino Acid Substitution Analysis of SNVs using SIFT, PROVEAN, and PolyPhen2:

No amino acid substitution in the exonic variant of the *p53* gene has been provided using SIFT, PROVEAN, and PolyPhen2 because these tools involve different databases for analysis. As no information regarding the identified variants is available in the database that's why no result was obtained confirming the novelty of *p53* variants. However, the *p53* variant located at the 7674945 position on chromosome 17 showed amino acid substitution from Arginine (R) to stop codon which further confirms that this variation will affect the protein structure and functioning.

### 3.5.8 Damaging and Tolerant Variant Analysis using MutationTaster, LRT, Mutation Assessor:

MutationTaster, LRT, and Mutation Assessor tools have not provided any result for 3 out of 4 variants of *p53* genes as these tools also require data available in the database for analysis of nucleotide variation, hence it further confirms the *p53* nc-RNA variant novelty. However, for the 1 exonic variant of the *p53* gene (rs397516435), the LRT result were showing "D" (deleterious variant) means that this variant affects conserved amino acid sequence while the mutationtaster result was representing "A" which means that this variant is deleterious and have the ability of diseases causing.

### 3.5.9 Pathogenic Effect of Variants on Protein Activity by VEST4:

For 1 single exonic variant (rs397516435), of the *p53* gene, the VEST4 score was given, whose value was 0.973 as shown in Table 3.4. As this value was between 0-1 so it means that the variation is deleterious and has an impact on the activity of a protein. While the remaining 3 nc-RNA variants of the *p53TG1* gene don't have any VEST4 results which further validates their novelty.

**Table 3.4:** Pathogenic effect of Variants on Protein Activity.

| Gene | Exonic Variants | VEST4 |
|---|---|---|
| *p53* | 1 | 0.973 (protein activity effected) |
| *p53TG1* | 3 | - |
| *YAP1* | - | - |

### 3.5.10 Deleteriousness of Variants using CADD raw and CADD Phred:

The CADD raw score for the rs397516435 variant of the *p53* gene was 7.664 which represents that the variant was deleterious as it has a positive value. However, for the remaining 3 nc-RNA variants of *p53*, no CADD raw score was given to depict the deleteriousness of SNVs. The CADD Phred score was 39 for rs397516435 variant of *p53* and as its value is above 20 so it means that the raw score was in the top 1% of the reference genome. While for the remaining three nc-RNA variants no CADD Phred score was given as it uses algorithms to compare the variation with a well-reported deleterious variation. So, no CADD Phred score means that no data was available regarding the deleteriousness of these variants which further confirms the novelty of these 3 variants in the *p53* gene.

### 3.5.11 Somatic Mutations Data Analysis by Cosmic 92:

The cosmic92 ID assigned to the rs397516435 variant of the *p53* gene was ID=COSV52663748 which means that data regarding this variant was already reported in a cosmic database while no cosmic92 IDs were assigned to the remaining *p53* nc-RNA variants which further supports their novelty in this gene.

### 3.5.12 Role of Variants in other Diseases using CLINSIG:

ClinSig provided a pathogenic status for the rs397516435 variant of the *p53* gene by using the ClinVar database while no data was given for the remaining three nc-RNA variants of the *p53* gene further validating the novelty of these variants.

### 3.5.13 Protein Homology Modeling using Phyre2:

Due to stop gain mutation in 1 exonic variant (rs397516435) of *p53* a truncated protein structure (Figure 3.16, A) was formed using the Phyre2 tool. Upon superimposition, with normal p53 protein DNA Binding Domain (DBD) structure (Figure 3.16, B) using ChimerX it was observed that this truncated protein structure has sequence homology with chain B (Yellow in Figure 3.16, B) of normal p53 DBD.

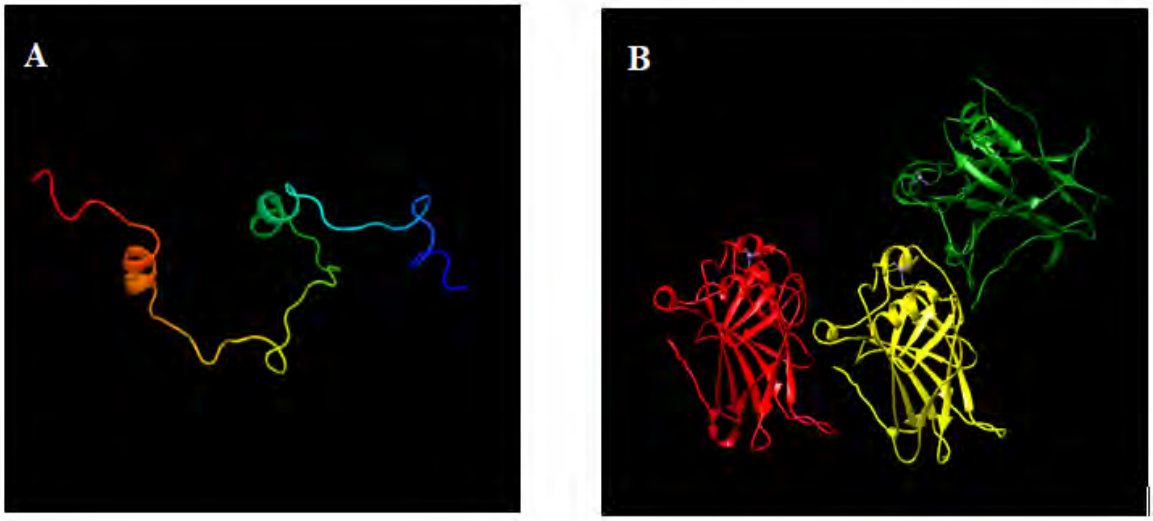


**Figure 3.16:** 3D structure of p53 protein by Phyre2. **A:** Truncated p53 protein structure **B:** Normal p53 protein DBD structure.

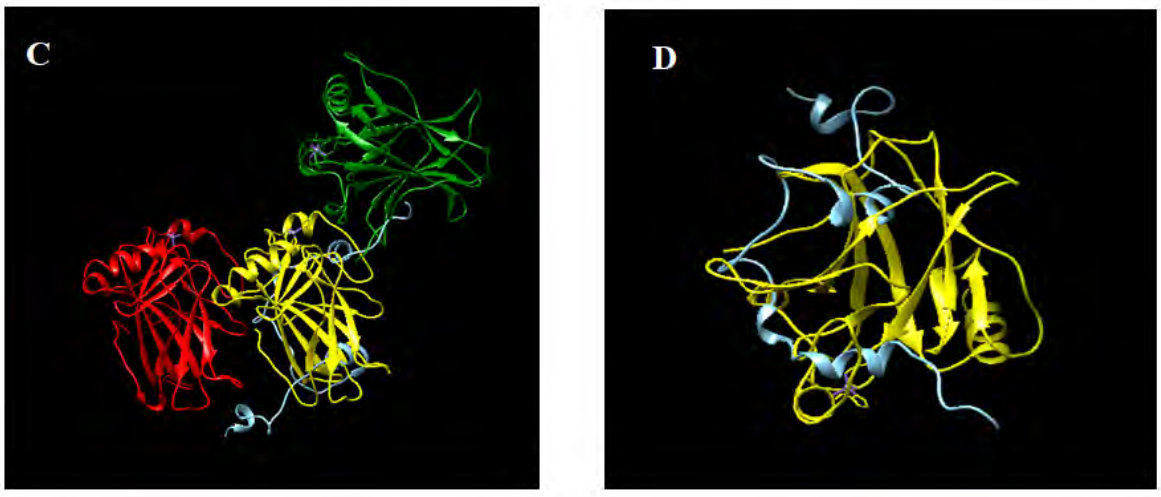


**Figure 3.17:** Superimposition of truncated p53 protein region on normal p53 DBD (**C**), Chain B of normal p53 protein DBD having sequence homology with truncated p53 protein (D).

# 4 Discussion

Pancreatic cancer (PC) is one of the fatal malignancy drastically affecting individuals across the world with a poor survival rate of 5 years in 10% of cases (Mizrahi *et al*., 2020). Due to the asymptomatic nature of this cancer, early-stage detection isn't possible that's why the survival rate of pancreatic cancer is quite low. However, in Pakistan, the situation is worst because there are no proper screening kits available, and can't be detected at early stages. Recently a report showed that in Pakistan 97.8% of pancreatic cancer cases lead to death (Ali *et al*., 2021). Researchers have been working on molecular profiling of PC to find out biomarkers that can be targeted for early diagnosis and treatment. Due to an incomplete data repository, no genetic characterization was done previously in Pakistan. As an Asian population, we have a slightly different genetic makeup as compared to European countries so for devising proper treatment, PC molecular analysis is very important.

Out of all types of PC, PDAC is the most prevalently occurring type in almost 90% of PC cases (Pishvaian and Brody, 2017). Most commonly mutated driver genes reported in PDAC progression include *KRAS*, *CDKN2A*, *TP53*, and *SMAD4* (Sun *et al*., 2020; Saiki *et al*., 2021). Studies reported the overexpression of YAP1 and mutant p53 in PDAC cases. In almost 75% of pancreatic cancer cases, the *p53* gene is highly mutated and commonly observed in PDAC (Simtniece *et al*., 2015; Jahedi *et al*., 2019). Mutant p53 induces carcinogenesis by remodeling the tumor microenvironment and inducing cellular metabolism (Weissmueller *et al*., 2014; Mantovani *et al*., 2019). By interacting with PDGFRβ, mutant p53 enhances the growth of PC cells (Weissmueller *et al*., 2014). Studies reported that a prominent homozygous mutation of *p53* was also seen in PDAC through PanIN-3 formation (Hu *et al*., 2021).

The YAP1 acts as a transcriptional coactivator and has a significant role in the Hippo pathway. YAP1 by interacting with other TEAD family members induces expression of growth-promoting and anti-apoptotic genes that facilitate in cell proliferation, organ development, and tissue homeostasis maintenance. YAP1 by interacting with ZEB1, SMADs, and TGF-β promotes EMT in cancer cells (Lehmann *et al*., 2016; Narimatsu *et al*., 2016). Along with KRAS, YAP1 is involved in PanIN formation

from ADM leading to PDAC in genetically altered mouse models (Gruber *et al*., 2016). According to Di Agostino *et al*, (2016), mutant p53 and YAP1 interacts leading to enhanced expression of oncogenes ultimately causing PC development.

The current study aimed to probe genetic variations among *p53* and *YAP1* genes by molecular profiling of PDAC patients using WES, an efficient method for the genetic characterization of Mendelian diseases and cancers. After quality checking, identified sample variants were annotated using the ANNOVAR tool which generates a vcf file for each sample that was further used for analysis. We have found multiple variations in exonic and intronic regions of different genes including driver genes. Among 4 driver genes, no exonic variation was observed in the *CDKN2A* gene, and 4 variations were observed in the *p53* gene out of which 3 were novel mutations of the nc-RNA region of *p53TG1* while 1 exonic variation was already reported as it has SNP ID (rs397516435) and its data is also available in databases. Furthermore, 1 exonic variation (25209843; A>G) was seen in the *KRAS* gene with an SNP ID rs1137282 and 1 novel exonic variation (51058160; G>A) was observed in the *SMAD4* gene.

Upon annotated variants analysis, we have found 4 substitutive variations in the coding region of the *p53* gene, out of which 3 variants (87341661; T>C, 87341667; A>T, 87345180; G>C) were novel present in nc-RNA region of *p53TG1* while 1 exonic variant (7674945; G>A) (rs397516435) found was already reported in the literature. However, approximately 280 variations were seen in the intronic region of the *p53* gene including both SNVs and InDels. The novelty of these 3 variants of *p53TG1* was confirmed by bioinformatics analysis using various tools. As explained earlier *p53TG1* is an lnc-RNA, that acts as a tumor suppressor, and upon stress, it is activated in p53 dependent manner (Xiao *et al*., 2018). Decreased expression of *p53TG1* is involved in cancer progression, especially in the lung and colon (Diaz-Lagares *et al*., 2016). As in our study 3, nc-RNA variants were observed in *p53TG1* that might downregulate their expression and leads to cancer progression.

AvSNP150 database was used to check whether any IDs were assigned to these variants or not which showed that no ID was assigned to 3 nc-RNA variants while SNP ID (rs397516435) was assigned to 1 exonic variant of the *p53* gene which clearly

shows that this variant was already reported. To further validate the novelty of these 3 variants several other bioinformatics tool was applied. ClinPred has not provided any score regarding the damaging or tolerant nature of these 3 nc-RNA variants however the ClinPred score for 1 exonic variant of the *p53* gene was more than 0.5 which depicts this variant as damaging and disease-causing. As the ClinPred tool uses ClinVar Database to predict the damage potential of a variant and as no score was given for 3 nc-RNA variants it means that no information regarding them is present in the database that shows the uniqueness of these variants.

Furthermore, SIFT, PROVEAN, and PolyPhen2 tools have not provided any information regarding the effect of these 3 nc-RNA variants on amino acid substitution. As these tools use different databases for variant detection so, no result confirms that no data is available regarding this newly identified variant in the database which further supports the novelty of these variants. While for 1 exonic variant amino acid substitution from Arginine (R) to stop codon as a result of which truncated protein will be formed that will affect the protein functioning. LRT, MutationTaster, Mutation Assessor, and VEST 4 also provided no result for these 3 nc-RNA variants which further validates the novelty of these variants. However, for 1 exonic variant LRT was giving "D" which means that this variant is deleterious and the MutationTaster result was showing "A" which also confirms this variant as a deleterious one with a damaging effect. The VEST4 score for 1 exonic variant was 0.973 which also confirms the deteriorative and pathogenic effect of this variant.

As CADD raw provided a 7.664 score and CADD Phred showed a 39 score for 1 exonic variant which represents the deleteriousness and damaging potential of this variant and CADD Phred score also validates that the raw score for this variant was in the top 1% of the reference genome. However, no score was given for the rest of 3 nc-RNA variants which further confirms their unfamiliarity.

As mentioned earlier that no SNP ID was assigned to these 3 nc-RNA variants similarly no Cosmic 92 ID was assigned to these variants which further confirms their novelty however, ID=COSV52663748 Cosimc92 ID was assigned to 1 exonic variant which means that data regarding this variant is present in literature. ClinSig status was

pathogenic for 1 exonic variant while no data was provided for the remaining 3 nc-RNA variants. Phyre2 tool showed truncated protein structure for 1 exonic variant and when superimposed on normal p53 structure it showed sequence homology with DNA binding domain sequence. As no database have data regarding these 3 nc-RNA variants and no result was provided by any of the bioinformatics tools so it validates the fact that these 3 variants reported in the annotated vcf file were novel and need to be explored more.

Upon annotated variants analysis no variation was observed in the exonic region of *YAP1* in our population but more than 330 variations were seen in the intronic region of the *YAP 1* gene including both SNVs and InDels. These variations might have a role in the increased expression of both genes in PDAC. As 3 nc-RNA novel variants were reported and we already know *p53* is one of the driver genes involved in PanIN-3 formation so might be these variants will also be involved in its formation leading to PDAC development.

In the future, further validation is required for these findings through *invitro* and *invivo* studies using animal models and cell lines, transcriptomic analysis, molecular docking, and *in silico* evaluation. Moreover, by molecular docking, several drug targeting domains can be obtained that will be targeted for treatment against PDAC. These studies will further help in determining the root cause of disease and different biomarkers that are important for diagnostic kits development which will help in early cancer detection and assist in devising better therapeutic strategies to improve PDAC patients' survival rate.

# 5 Conclusion

In conclusion, we summarized that novel substitutive variation (87341661; T>C, 87341667; A>T, 87345180; G>C) in the nucleotide sequence of nc-RNA, *p53TG1* (tumor suppressor) along with 1 exonic reported variant (rs397516435) of *p53* gene might be responsible for *p53TG1* downregulation and facilitate the progression of cancer. Furthermore, 1 exonic variant of the *p53* (mut-*p53*) gene probably can interact with *YAP1* and promote the expression of pro-tumorigenic genes. In addition, no exonic region variations have been found in the *YAP1* gene hence its higher expression in PDAC would not be credited to nucleotide variation in its exonic region. However, in the intronic and intergenic region of both genes, several variations were seen that possibly have a link with enhanced expression of mutant *p53* and *YAP1* in PDAC and poor survival rate. To validate these findings and to explore diagnostic and therapeutic targets further studies are still required that will help in formulating a cure for this pernicious cancer.

## 6 References:

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current protocols in human genetics*, *76*(1), 7-20.

Ali, S. M., Adnan, Y., & Ali, S. A. (2021). Pancreatic Cancers: A Review of Studies from Pakistan and Comparison with Global Trends. Pakistan Journal of Public Health, 11(2), 120-127.

Alirezaie, N., Kernohan, K. D., Hartley, T., Majewski, J., & Hocking, T. D. (2018). ClinPred: prediction tool to identify disease-relevant nonsynonymous single-nucleotide variants. *The American Journal of Human Genetics*, *103*(4), 474-483.

Ansari, D., Tingstedt, B., Andersson, B., Holmquist, F., Sturesson, C., Williamsson, C., ... & Andersson, R. (2016). Pancreatic cancer: yesterday, today, and tomorrow. Future Oncology, 12(16), 1929-1946.

Ansari, D., Williamsson, C., Tingstedt, B., Andersson, B., Lindell, G., & Andersson, R. (2014). Pancreaticoduodenectomy–the transition from a low-to a high-volume center. Scandinavian journal of gastroenterology, 49(4), 481-484.

Bai, N., Zhang, C., Liang, N., Zhang, Z., Chang, A., Yin, J., ... & Liu, C. (2013). Yes-associated protein (YAP) increases chemosensitivity of hepatocellular carcinoma cells by modulation of p53. *Cancer biology & therapy*, *14*(6), 511-520.

Bailey, P., Chang, D. K., Nones, K., Johns, A. L., Patch, A. M., Gingras, M. C., ... & Grimmond, S. M. (2016). Genomic analyses identify molecular subtypes of pancreatic cancer. Nature, 531(7592), 47-52.

Becker, A. E., Hernandez, Y. G., Frucht, H., & Lucas, A. L. (2014). Pancreatic ductal adenocarcinoma: risk factors, screening, and early detection. World journal of gastroenterology: WJG, 20(32), 11182.

Benci, J. L., Johnson, L. R., Choa, R., Xu, Y., Qiu, J., Zhou, Z., ... & Minn, A. J. (2019). Opposing functions of interferon coordinate adaptive and innate immune responses to cancer immune checkpoint blockade. Cell, 178(4), 933-948.

Benjamin, D., Sato, T., Cibulskis, K., Getz, G., Stewart, C., & Lichtenstein, L. (2019). Calling somatic SNVs and indels with Mutect2. *BioRxiv*, 861054.

Bond-Smith, G., Banga, N., Hammond, T. M., & Imber, C. J. (2012). Pancreatic adenocarcinoma. Bmj, 344.

Bracci, P. M. (2012). Obesity and pancreatic cancer: overview of epidemiologic evidence and biologic mechanisms. Molecular carcinogenesis, 51(1), 53-63.

Burns, W. R., & Edil, B. H. (2012). Neuroendocrine pancreatic tumors: guidelines for management and update. Current treatment options in oncology, 13(1), 24-34.

Carter, H., Douville, C., Stenson, P. D., Cooper, D. N., & Karchin, R. (2013). Identifying Mendelian disease genes with the variant effect scoring tool. *BMC genomics*, *14*(3), 1-16.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884-i890.

Chen, T., Ren, L., Liu, X., Zhou, M., Li, L., Xu, J., & Zhu, X. J. I. j. o. m. s. (2018). DNA nanotechnology for cancer diagnosis and therapy. *19*(6), 1671.

Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels.

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., ... & Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology*, *31*(3), 213-219.

Collisson, E. A., Bailey, P., Chang, D. K., & Biankin, A. V. (2019). Molecular subtypes of pancreatic cancer. *Nature reviews Gastroenterology & hepatology*, *16*(4), 207-220.

Davis, E. J., Johnson, D. B., Sosman, J. A., & Chandra, S. (2018). Melanoma: What do all the mutations mean? Cancer, 124(17), 3490-3499.

De La Cruz, M. S. D., Young, A. P., & Ruffin, M. T. (2014). Diagnosis and management of pancreatic cancer. American family physician, 89(8), 626-632.

Del Chiaro, M., Segersvärd, R., Lohr, M., & Verbeke, C. (2014). Early detection and prevention of pancreatic cancer: is it really possible today? World journal of gastroenterology: WJG, 20(34), 12118.

Di Agostino, S., Sorrentino, G., Ingallina, E., Valenti, F., Ferraiuolo, M., Bicciato, S., ... & Blandino, G. (2016). YAP enhances the pro-proliferative transcriptional activity of mutant p53 proteins. EMBO reports, 17(2), 188-201.

Diaz-Lagares, A., Crujeiras, A. B., Lopez-Serra, P., Soler, M., Setien, F., Goyal, A., ... & Esteller, M. (2016). Epigenetic inactivation of the p53-induced long noncoding RNA TP53 target 1 in human cancer. *Proceedings of the National Academy of Sciences*, *113*(47), E7535-E7544.

Eltawil, K. M., Renfrew, P. D., & Molinari, M. (2012). Meta-analysis of phase III randomized trials of molecular targeted therapies for advanced pancreatic cancer. HPB, 14(4), 260-268.

Feng, Y., Liu, H., Duan, B., Liu, Z., Abbruzzese, J., Walsh, K. M., ... & Wei, Q. (2019). Potential functional variants in SMC2 and TP53 in the AURORA pathway genes and risk of pancreatic cancer. Carcinogenesis, 40(4), 521-528.

Fleming, R., Fleming, M., Chaudhuri, T., & Dooley, W. J. C. R. i. O. C. (2019). What is Cancer. 108.

Furth, N., & Aylon, Y. (2017). The LATS1 and LATS2 tumor suppressors: beyond the Hippo pathway. *Cell Death & Differentiation*, *24*(9), 1488-1501.

Gao, C., Quan, M. Y., Chen, Q. J., Yang, R., Wu, Y., Liu, J. Y., ... & Zhang, J. S. (2021). Yap1-2 isoform is the primary mediator in TGF-β1-induced EMT in pancreatic cancer. Frontiers in Oncology, 1241.

Gbadegesin, M. A., Omotoso, O. E., Oluwasola, T. A., Okolo, C. A., Soremekun, O., Ogun, G. O., ... & Odunola, O. A. (2021). Mutational analysis of p53 gene in cervical cancer and useful polymorphic variants in exons 3 and 4. Egyptian Journal of Medical Human Genetics, 22(1), 1-8.

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

Giannis, D., Moris, D., & Barbas, A. S. (2021). Diagnostic, Predictive and Prognostic Molecular Biomarkers in Pancreatic Cancer: An Overview for Clinicians. *Cancers*, *13*(5), 1071.

Golan, T., Hammel, P., Reni, M., Van Cutsem, E., Macarulla, T., Hall, M. J., ... & Kindler, H. L. (2019). Maintenance olaparib for germline BRCA-mutated metastatic pancreatic cancer. New England Journal of Medicine, 381(4), 317-327.

Gruber, R., Panayiotou, R., Nye, E., Spencer-Dene, B., Stamp, G., & Behrens, A. (2016). YAP1 and TAZ control pancreatic cancer initiation in mice by direct up-regulation of JAK–STAT3 signaling. Gastroenterology, 151(3), 526-539.

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

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

Hao, F., Xu, Q., Zhao, Y., Stevens, J. V., Young, S. H., Sinnett-Smith, J., & Rozengurt, E. (2017). Insulin Receptor and GPCR Crosstalk Stimulate YAP via PI3K and PKD in Pancreatic Cancer CellsYAP Activation by INSR and GPCR Crosstalk. Molecular Cancer Research, 15(7), 929-941.

Harris, R. E. (2019). Epidemiology of chronic disease: global perspectives. Jones & Bartlett Learning.

Hayashi, H., Uemura, N., Zhao, L., Matsumura, K., Sato, H., Shiraishi, Y., & Baba, H. (2021). Biological Significance of YAP/TAZ in Pancreatic Ductal Adenocarcinoma. *Frontiers in Oncology*, 2941.

Hu, H. F., Ye, Z., Qin, Y., Xu, X. W., Yu, X. J., Zhuo, Q. F., & Ji, S. R. (2021). Mutations in key driver genes of pancreatic cancer: Molecularly targeted therapies and other clinical implications. Acta Pharmacologica Sinica, 42(11), 1725-1741.

Huang, B. Z., Pandol, S. J., Jeon, C. Y., Chari, S. T., Sugar, C. A., Chao, C. R., ... & Setiawan, V. W. (2020). New-onset diabetes, longitudinal trends in metabolic markers, and risk of pancreatic cancer in a heterogeneous population. Clinical gastroenterology and hepatology, 18(8), 1812-1821.

Jahedi, H., Fahud, A. L., & Lim, C. L. (2019). Role of p53 family isoforms in enhancing aggressiveness and chemoresistance in pancreatic cancer. World Academy of Sciences Journal, 1(5), 236-246.

Jain, A. K., Allton, K., Iacovino, M., Mahen, E., Milczarek, R. J., Zwaka, T. P., ... & Barton, M. C. (2012). p53 regulates the cell cycle and microRNAs to promote the differentiation of human embryonic stem cells. PLoS biology, 10(2), e1001268.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, *10*(6), 845-858.

Khyade, V. B. https://scientificrc. com/journals/international-journal-of-social-science-and-humanities-research/volume-6-issue-4-October-December-2018/Pancreatic Cancer on Occasion of November: Pancreatic Cancer Awareness Month.

Lehmann, W., Mossmann, D., Kleemann, J., Mock, K., Meisinger, C., Brummer, T., ... & Brabletz, T. (2016). ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types. Nature communications, 7(1), 1-15.

Li, B. Q., Wang, L., Li, J., Zhou, L., Zhang, T. P., Guo, J. C., & Zhao, Y. P. (2017). Surgeons' knowledge regarding the diagnosis and management of pancreatic cancer in China: a cross-sectional study. BMC health services research, 17(1), 1-9.

Li, X., Xu, H., & Gao, P. (2018). ABO blood group and diabetes mellitus influence the risk for pancreatic cancer in a population from China. Medical science monitor: international medical journal of experimental and clinical research, 24, 9392.

Li, R., He, Y., Zhang, H., Wang, J., Liu, X., Liu, H., ... & Liang, Z. (2021). Identification and validation of immune molecular subtypes in pancreatic ductal adenocarcinoma: implications for prognosis and immunotherapy. *Frontiers in immunology*, 2829.

Liu, H., Du, S., Lei, T., Wang, H., He, X., Tong, R., & Wang, Y. (2018). Multifaceted regulation and functions of YAP/TAZ in tumors. Oncology reports, 40(1), 16-28.

Luo, J., & Yu, F. X. (2019). GPCR-Hippo signaling in cancer. Cells, 8(5), 426.

Lucenteforte, E., La Vecchia, C., Silverman, D., Petersen, G. M., Bracci, P. M., Ji, B. A., ... & Duell, E. J. (2012). Alcohol consumption and pancreatic cancer: a pooled analysis in the International Pancreatic Cancer Case–Control Consortium (PanC4). Annals of oncology, 23(2), 374-382.

Makohon-Moore, A., Brosnan, J. A., & Iacobuzio-Donahue, C. A. (2013). Pancreatic cancer genomics: insights and opportunities for clinical translation. Genome medicine, 5(3), 1-11.

Mantovani, F., Collavin, L., & Del Sal, G. (2019). Mutant p53 as a guardian of the cancer cell. Cell Death & Differentiation, 26(2), 199-212.

Masoudi, S., Sanat, Z. M., Saleh, A. M., Nozari, N., Ghamarzad, N., & Pourshams, A. (2017). Menstrual and reproductive factors and risk of pancreatic cancer in women. Middle East Journal of Digestive Diseases, 9(3), 146.

McMenamin, Ú. C., McCain, S., & Kunzmann, A. T. (2017). Do smoking and alcohol behaviors influence GI cancer survival? Best Practice & Research Clinical Gastroenterology, 31(5), 569-577.

Mello, S. S., Valente, L. J., Raj, N., Seoane, J. A., Flowers, B. M., McClendon, J., ... & Attardi, L. D. (2017). A p53 super-tumor suppressor reveals a tumor suppressive p53-Ptpn14-Yap axis in pancreatic cancer. *Cancer cell*, *32*(4), 460-473.

Midha, S., Chawla, S., & Garg, P. K. (2016). Modifiable and non-modifiable risk factors for pancreatic cancer: A review. Cancer letters, 381(1), 269-277.

Mizrahi, J. D., Surana, R., Valle, J. W., & Shroff, R. T. (2020). Pancreatic cancer. *The Lancet*, *395*(10242), 2008-2020.

Molina-Montes, E., Van Hoogstraten, L., Gomez-Rubio, P., Löhr, M., Sharp, L., Molero, X., ... & PanGenEU Study Investigators. (2020). Pancreatic Cancer Risk in Relation to Lifetime Smoking Patterns, Tobacco Type, and Dose-Response Relationships Smoking and Pancreatic Cancer Risk. Cancer epidemiology, biomarkers & prevention, 29(5), 1009-1018.

Murakami, S., Nemazanyy, I., White, S. M., Chen, H., Nguyen, C. D., Graham, G. T., ... & Yi, C. (2019). A Yap-Myc-Sox2-p53 regulatory network dictates metabolic homeostasis and differentiation in Kras-driven pancreatic ductal adenocarcinomas. *Developmental cell*, *51*(1), 113-128.

Nagai, H., & Kim, Y. H. (2017). Cancer prevention from the perspective of global cancer burden patterns. *Journal of thoracic disease*, *9*(3), 448.

Narimatsu, M., Labibi, B., Wrana, J. L., & Attisano, L. (2016). Analysis of Hippo and TGFβ signaling in polarizing epithelial cells and mouse embryos. Differentiation, 91(4-5), 109-118.

Nimmakayala, R. K., Seshacharyulu, P., Lakshmanan, I., Rachagani, S., Chugh, S., Karmakar, S., ... & Ponnusamy, M. P. (2018). Cigarette smoke induces stem cell features of pancreatic cancer cells via PAF1. Gastroenterology, 155(3), 892-908.

Nitsche, C., Simon, P., Weiss, F. U., Fluhr, G., Weber, E., Gärtner, S., ... & Lerch, M. M. (2011). Environmental risk factors for chronic pancreatitis and pancreatic cancer. Digestive Diseases, 29(2), 235-242.

Öberg, K., Knigge, U., Kwekkeboom, D., & Perren, A. (2012). Neuroendocrine gastro-entero-pancreatic tumors: ESMO Clinical Practice Guidelines for diagnosis, treatment, and follow-up. Annals of oncology, 23, vii124-vii130.

Pishvaian, M. J., & Brody, J. R. (2017). Therapeutic implications of molecular subtyping for pancreatic cancer. Oncology (Williston Park, NY), 31(3), 159-66.

Raj, N., & Bam, R. (2019). Reciprocal crosstalk between YAP1/hippo pathway and the p53 family proteins: mechanisms and outcomes in cancer. *Frontiers in cell and developmental biology*, *7*, 159.

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

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research*, *39*(17), e118-e118.

Risch, H. A., Lu, L., Wang, J., Zhang, W., Ni, Q., Gao, Y. T., & Yu, H. (2013). ABO blood group and risk of pancreatic cancer: a study in Shanghai and meta-analysis. American journal of epidemiology, 177(12), 1326-1337.

Ryan, D. P., Hong, T. S., & Bardeesy, N. (2014). Pancreatic adenocarcinoma. New England Journal of Medicine, 371(11), 1039-1049.

Sahin, I. H., Askan, G., Hu, Z. I., & O'Reilly, E. M. (2017). Immunotherapy in pancreatic ductal adenocarcinoma: an emerging entity. Annals of Oncology, 28(12), 2950-2961.

Saiki, Y., Jiang, C., Ohmuraya, M., & Furukawa, T. (2021). Genetic Mutations of Pancreatic Cancer and Genetically Engineered Mouse Models. Cancers, 14(1), 71.

Samokhvalov, A. V., Rehm, J., & Roerecke, M. (2015). Alcohol consumption as a risk factor for acute and chronic pancreatitis: a systematic review and a series of meta-analyses. EBioMedicine, 2(12), 1996-2002.

Sandell, L., & Sharp, N. P. (2022). Fitness effects of mutations: An assessment of PROVEAN predictions using mutation accumulation data. *Genome biology and evolution*, *14*(1), evac004.

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

Simtniece, Z., Vanags, A., Strumfa, I., Sperga, M., Vasko, E., Prieditis, P., ... & Gardovskis, J. (2015). Morphological and immunohistochemical profile of pancreatic neuroendocrine neoplasms. Polish Journal of Pathology, 66(2), 176-194.

Singh, S. K., & Ellenrieder, V. (2013). Senescence in pancreatic carcinogenesis: from signaling to chromatin remodeling and epigenetics. Gut, 62(9), 1364-1372.

Son, S. H., Song, J. H., Choi, B. O., Lee, M. A., Kang, K. M., & Jang, H. S. (2012). The technical feasibility of image-guided intensity-modulated radiotherapy (IG-IMRT) to perform a hypofractionated schedule in terms of toxicity and local control for patients with locally advanced or recurrent pancreatic cancer. Radiation Oncology, 7(1), 1-8.

Springfeld, C., Jäger, D., Büchler, M. W., Strobel, O., Hackert, T., Palmer, D. H., & Neoptolemos, J. P. (2019). Chemotherapy for pancreatic cancer. La Presse Medicale, 48(3), e159-e174.

Sudol, M., Shields, D. C., & Farooq, A. (2012, September). Structures of YAP protein domains reveal promising targets for the development of new cancer drugs. In

Seminars in cell & developmental biology (Vol. 23, No. 7, pp. 827-833). Academic Press.

Sudol, M. (2013). YAP1 oncogene and its eight isoforms. Oncogene, 32(33), 3922-3922.

Szulzewsky, F., Arora, S., Hoellerbauer, P., King, C., Nathan, E., Chan, M., ... & Holland, E. C. (2020). A comparison of tumor-associated YAP1 fusions identifies a recurrent set of functions critical for oncogenesis. Genes & development, 34(15-16), 1051-1064.

Szulzewsky, F., Holland, E. C., & Vasioukhin, V. (2021). YAP1 and its fusion proteins in cancer initiation, progression, and therapeutic resistance. Developmental Biology, 475, 205-221.

Tobias, J. S., & Hochhauser, D. (2014). Cancer and its management. John Wiley & Sons.

Walter, F. M., Mills, K., Mendonça, S. C., Abel, G. A., Basu, B., Carroll, N., ... & Emery, J. D. (2016). Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study. The Lancet Gastroenterology & Hepatology, 1(4), 298-306.

Weinberg, R., & Hanahan, D. J. C. (2000). The hallmarks of cancer. *100*(1), 57-70.

Weissmueller, S., Manchado, E., Saborowski, M., Morris IV, J. P., Wagenblast, E., Davis, C. A., ... & Lowe, S. W. (2014). Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor β signaling. Cell, 157(2), 382-394.

Wolfgang, C. L., Herman, J. M., Laheru, D. A., Klein, A. P., Erdek, M. A., Fishman, E. K., & Hruban, R. H. (2013). Recent progress in pancreatic cancer. CA: a cancer journal for clinicians, 63(5), 318-348.

Xiao, H., Liu, Y., Liang, P., Wang, B., Tan, H., Zhang, Y., ... & Gao, J. (2018). TP53TG1 enhances cisplatin sensitivity of non-small cell lung cancer cells through regulating miR-18a/PTEN axis. *Cell & bioscience*, *8*(1), 1-13.

Xiao, Y., Yang, Y., Wang, Y., Li, X., & Wang, W. (2019). Five novel genes related to the pathogenesis and progression of pancreatic neuroendocrine tumors by bioinformatics analysis with RT-qPCR verification. Frontiers in neuroscience, 13, 937.

Yadav, D., & Lowenfels, A. B. (2013). The epidemiology of pancreatitis and pancreatic cancer. Gastroenterology, 144(6), 1252-1261.

Yang, S., Zhang, L., Purohit, V., Shukla, S. K., Chen, X., Yu, F., ... & Dong, J. (2015). Active YAP promotes pancreatic cancer cell motility, invasion, and tumorigenesis in a mitotic phosphorylation-dependent manner through LPAR3. Oncotarget, 6(34), 36019.

Yoshioka, R., Yasunaga, H., Hasegawa, K., Horiguchi, H., Fushimi, K., Aoki, T., ... & Kokudo, N. (2014). Impact of hospital volume on hospital mortality, length of stay, and total costs after pancreaticoduodenectomy. Journal of British Surgery, 101(5), 523-529.

Zawacka-Pankau, J. E. (2020). The undervalued avenue to reinstate tumor suppressor functionality of the p53 protein family for improved cancer therapy-drug repurposing. Cancers, 12(9), 2717.

Zhang, W., Nandakumar, N., Shi, Y., Manzano, M., Smith, A., Graham, G., ... & Yi, C. (2014). Downstream of mutant KRAS, the transcription regulator YAP is essential for neoplastic progression to pancreatic ductal adenocarcinoma. Science signaling, 7(324), ra42-ra42.

Zhang, Q., Zeng, L., Chen, Y., Lian, G., Qian, C., Chen, S., ... & Huang, K. (2016). Pancreatic cancer epidemiology, detection, and management. Gastroenterology research and practice, 2016.

Zhou, W., Han, X., Fang, Y., Han, S., Cai, Y., Kuang, T., ... & Wang, D. (2020). Clinical analysis of acinar cell carcinoma of the pancreas: a single-center experience of 45 consecutive cases. Cancer Control, 27(1), 1073274820969447.