

https://dspace.org/

Discovery of pancreatic cancer biomarker(s) using focused pathway analyses

Nweke, Ekene Emmanuel

2017

http://hdl.handle.net/10539/23156

Nweke, Ekene Emmanuel (2017) Discovery of pancreatic cancer biomarker(s) using focused pathway analyses, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/23156>

Downloaded from a DSpace repository. https://dspace.org/



Discovery of pancreatic cancer biomarker(s) using focused pathway analyses

Ekene Nweke

A Thesis submitted to the faculty of health sciences, University of Witwatersrand, in fulfilment of the requirements for the degree of Doctor of philosophy.

Johannesburg, 2017

DECLARATION

I, Ekene Nweke declare that this Thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

_____

(Signature of candidate)

_____day of_____20_____in_____

DEDICATION

To every Black African child out there, NOTHING IS IMPOSSIBLE.

To Late Dr Boitelo Letsolo, I am sure you are smiling down now 😊

PRESENTATIONS

Meeting	Presentation Title	Authors	Venue	Date	Type of presentation
Surgical Research Forum	Proteomics of pancreatic cancer	**Ekene Nweke**, and Geoffrey Candy	9th floor, Faculty of health sciences, University of Witwatersrand.	August 2015	Oral
SAGS/SASBI Joint conference	Gene expression profiling of pancreatic tumours using next generation sequencing	**Ekene Nweke**, Monde Ntwasa, Martin Brand, John Devar, Martin Smith, Geoffrey Candy	Nelson Mandela teaching hospital, UKZN, Durban	20th- 24th September 2016	Poster
Bert Myburg Surgical Forum	Gene and protein expression profiling of pancreatic tumours in the search for novel potential biomarkers	**Ekene Nweke**, Monde Ntwasa, Martin Brand, John Devar, Martin Smith, Geoffrey Candy	9th floor, Faculty of health sciences, University of Witwatersrand.	30th November, 2016	Oral
ACGT Proteomics Conference	Proteomic profiling of pancreatic tumours in the search for novel potential biomarkers.	**Ekene Nweke**, Monde Ntwasa, Martin Brand, John Devar, Martin Smith, Geoffrey Candy	Irene Farm Conference center, Pretoria	5th December, 2016	Oral
MBRT research day	Gene expression profiling of pancreatic tumours using next generation sequencing	**Ekene Nweke**, Monde Ntwasa, Martin Brand, John Devar, Martin Smith, Geoffrey Candy	WITS Club, West campus, University of Witwatersrand.	8th December 2016	Poster
Cancer Symposium	Gene and Protein expression profiling reveals novel potential biomarkers	**Ekene Nweke**, Monde Ntwasa, Martin Brand, John Devar, Martin Smith, Geoffrey Candy	Faculty of health sciences, University of Witwatersrand.	9th February, 2017	Oral

PUBLICATIONS

Transcriptomic profiling of pancreatic tumours reveal novel potential biomarkers (2017). **Ekene Nweke**, Monde Ntwasa, Martin Brand, John Devar, Martin Smith, Geoffrey Candy---- In Preparation

Proteomic profiling of pancreatic tumours reveal novel potential biomarkers (2017). **Ekene Nweke**, Monde Ntwasa, Martin Brand, John Devar, Martin Smith, Geoffrey Candy---- In Preparation

Publications not related to study

The *Drosophila* Retinoblastoma Binding Protein 6 family member has two isoforms and is potentially involved in embryonic patterning. Rodney Hull, Brent Oosthuysen, Umar-Faruq Cajee, Lehlogonolo Mokgohloa, **Ekene Nweke**, Ricardo Jorge Antunes, Theresa HT Coetzer and Monde Ntwasa.

SUMMARY

Pancreatic cancer (PDAC) is a deadly type of cancer with almost an equal amount of new cases and deaths observed yearly. It accounts for about 7% of cancer related deaths worldwide with less than 5% of PDAC patients living up to 5 years. The lack of specific and sensitive diagnostic tests is strongly responsible for this poor statistic. The discovery of differentially expressed genes and proteins associated with PDAC is crucial to elucidating this condition and may lead to biomarker finding and further understanding of the disease. This in turn may lead to improved diagnostic tests for early diagnosis. The aim of this study was to identify novel potential biomarkers for PDAC. 18 tissue samples (9 tumour and 9 normal) were obtained from PDAC patients at Chris Hani Baragwanath Johannesburg, South Africa during the Whipple procedure. Using focused gene arrays, differences in expression profiles of pancreatic tumours and cell line (MIA PaCa-2) were shown. Furthermore, RNA sequencing and antibody arrays were used for the identification of differentially expressed genes and proteins respectively. Utilizing multiple bioinformatics tools, functional enrichment, biomarker filtering, and variant analyses were performed. After discovery, novel potential gene biomarkers were then validated using real time PCR. Also, protein-protein interaction (PPI) analysis on potential biomarkers was done. In the current study, the overexpression of necrotic pathways was observed in tumours but not in normal and MIA-PaCa-2. Furthermore, differential gene and protein expression in tumour was shown. Functional enrichment analyses of the differentially expressed genes and proteins showed the upregulation of several pathways including: IGF/MAPK, Ras, and the immune system pathways. Importantly, *JUP*, ANGPTL4 and p27/kip1 were identified as novel potential biomarkers. Also, I have shown novel variants (*SPARC*, *ZFHX3*, *TGM2* and *COLIA1*) predicted to result in loss of gene function. This study has identified novel potential biomarkers which can help in the diagnosis of PDAC. PPI analyses of these biomarkers indicate they may play important roles in inducing tumorigenesis. Information obtained from dysregulated signalling pathways and biological processes help in further understanding the biology of PDAC. Additionally, identified variations may enable tumour growth. In future, the identified novel potential biomarkers should be further validated in a larger sample number using easily accessible samples like blood.

ACKNOWLEGEMENTS

I thank God Almighty for his Infinite mercy, unshakeable love, grace, and strength.

To my family, for being my emotional and financial support.

To Profs Geoffrey candy and Monde Ntwasa, thank you for academic and life mentorship.

To Dr Martin Brand for input and help with samples.

To Dr John Devar for help with sample collection and funding for RNA sequencing.

To Dr Marietha Nel, for immense help with cell culturing.

To every member of the Department of Surgery University of Witwatersrand especially Ritchie, Brenda, and Jeanette, who helped in sample collection and storage. Thanks for all your amazing contributions.

Finally, I thank everyone in the Flylab (Philemon, Umar, Bernice, Ntombi, Zanele and Charmy) including Siyanda and Anza. You are all an awesomely amazing bunch! And I'm forever grateful to have worked with you all.

LIST OF ABBREVIATIONS

µL-microliter

AKT- Protein Kinase B

ANGPTL4- Angiopoietin-like 4

BRCA2- Breast Cancer type 2

BSA- Bovine serum albumin

CA-19-9- Cancer antigen 19-9

CCKR- Cholecystokinin receptor

CDKN1B- Cyclin-dependent kinase inhibitor 1B

COLIA1- Collagen type I alpha 1

COPG- Coatomer subunit gamma

CXXC4- CXXC Finger Protein 4 (C stands for cysteine and X for any amino acid)

DAVID- Database for annotation, visualization, and integrated discovery

DEG- Differentially expressed genes

DMEM-Dulbecco's modified eagle medium

DNA- Deoxyribonucleic acid

DNase- Deoxyribonuclease

EDTA- Ethylenediaminetetraacetic acid

EGFR- Epidermal Growth Factor Receptor

EGR1- Early growth response protein

GABPA- GA-binding protein alpha chain

GALNT1- Polypeptide N-actylgalactosaminyltransferase 1

Hnf 6- Hepatocyte nuclear factor 6

HNF1 β- Hepatocyte nuclear factor 1 beta

HS3ST4- Heparan Sulphate (Glucosamine) 3-O-Sulfotransferase 4

IGF- Insulin-like Growth Factor 1

IPMN- Intraductal papillary mucinous neoplasm

JUP- Junction Plakoglobin

K-RAS- Kirsten rat sarcoma

KLF7- Kruppel-like factor 7

L1CAM- L1 Cell Adhesion Molecule

MHC- Major Histocompatibility complex

MOPS- (3-(N-morpholino) propanesulfonic acid)

mRNA- Messenger RNA

MRPL19- Mitochondrial Ribosomal Protein L19

NCBI- National center for biotechnology information

NeuroD- Neurogenic differentiation 1

Ngn 3- Neurogenin 3

PALB 2- Partner and localizer of BRCA2

Pax 4- Paired box gene 4

PCR- Polymerase chain reaction

PDAC- Pancreatic ductal adenocarcinoma

PDGF- Platelet-Derived Growth Factor

Pdx 1- Insulin promoter factor 1

PMSF- Phenylmethylsulfonyl fluoride

PNETS- Pancreatic neuroendocrine tumours

PPI- Protein-protein interaction

REST- Relative expression software tool

RNA- Ribonucleic acid

RNase- Ribonuclease

Rpm- Revolutions per minute

SLC15A4- Solute Carrier Family 15 member 4

SOX 9- SRY-Box 9

SP1- Specificity protein 1

SP4- Specificity protein 4

SPARC- Secreted Protein Acidic and Rich in Cysteine

TBE- Tris borate EDTA buffer

TCIRG1- T-Cell Immune Regulator 1

TGM2- Transglutaminase 2

TSPAN7- Tetraspanin- 7

VEGF- Vascular epidermal growth factor

YY1- Yin Yang 1

ZFHX3- Zinc Finger Homeobox 3

α- Alpha

β- Beta

δ- Delta

ε- Epsilon

TABLE OF CONTENTS

DECLARATION i

DEDICATION ii

PRESENTATIONS iii

PUBLICATIONS iv

SUMMARY v

ACKNOWLEGEMENTS vi

LIST OF ABBREVIATIONS vii

TABLE OF CONTENTS ix

LIST OF FIGURES xiv

LIST OF TABLES xvi

CHAPTER 1- LITERATURE REVIEW 1

1.1 Introduction 1

1.1.1 The pancreas 1

1.1.2 Location of the pancreas 2

1.1.3 Development of the pancreas 2

1.1.4 Regulation of pancreatic development 2

1.1.5 Pancreatic ductal cell lineage regulation during development 3

1.1.6 Functions of the pancreas 3

1.1.7 Diseases of the pancreas 4

Exocrine tumours 4

Endocrine tumours 5

1.1.8 Incidence of PDAC 5

1.1.9 Risk factors identified with PDAC 6
1.1.10 Signs and symptoms of PDAC 6
1.1.11 Diagnosis of PDAC 7
1.1.12 Staging in PDAC 7
1.1.13 Treatment options for pancreatic cancer 8
1.2 Discovering biomarkers of PDAC 9
1.2.1 Gene expression profiling (Transcriptomics) in PDAC 10
1.2.2 Protein expression profiling (proteomics) in pancreatic cancer 12
Protein-Protein interaction (PPI) networks- Role in pancreatic cancer research 13
1.2.3 Alternative splice variants as biomarkers of PDAC 13
1.2.4 Signalling Pathways altered during PDAC 14
1.3 Justification of study 16
1.4 Aim 16
1.5 Objectives 16
CHAPTER 2- METHODS 18
2.1 Materials 18
2.2 Methods 18
2.2.1 Summary of experimental strategy 18
2.2.2 Ethics statement and Funding 19
2.2.3 Sample collection, storage, and processing 20
Sample homogenization 20
2.2.4 Total RNA and Protein isolation 21
Total RNA isolation 21
DNase Digestion 21
RNA quantification and purity assessment 21
RNA Agarose gel electrophoresis 21
Protein Extraction 22

Protein quantification 22
2.2.5 cDNA First strand synthesis 22
2.2.6 Focused gene expression profiling using the Human Cancer PathwayFinder RT^2 Profiler PCR Arrays 23
Brief description of kit 23
Procedure 23
Data analysis 23
2.2.7 Gene expression profiling using RNA sequencing 24
Brief description of RNA sequencing 24
Procedure 24
Data analysis 24
2.2.8 Biomarker filtering analysis 25
2.2.9 Biomarker (gene) Validation by Real time PCR using RT^2 qPCR primer assays 26
Procedure 26
Data analysis 26
2.2.10 Protein-protein interaction network analysis 27
2.2.11 Variant filtering analysis 27
2.2.12 Focused proteomic profiling using Human oncology proteome profiler array 29
Brief Description of kit 29
Procedure 29
Data analysis 30
2.2.13 Biomarker filtering analysis 30
2.2.14 Protein-protein interaction network analysis 30
CHAPTER 3- RESULTS 31
3.1 Summary of results 31
3.2 Gene expression profile in tumours and MIA PaCa-2 32
Visualization of gene expression profiles 32

Pathway analysis of Group 1 and Group 2 DEGs....33
Biomarker filtering analysis....37
3.3 Differential gene expression using RNA sequencing....38
Functional enrichment analysis of upregulated genes....40
Gene network analysis of upregulated genes....43
Functional enrichment of downregulated genes....45
Gene network analysis of downregulated genes....46
3.4 Gene biomarker discovery and validation....47
Functional enrichment analysis of the JUP PPI network....49
3.5 Discovering genetic variants involved in PDAC....51
Functional enrichment analysis of variants....52
3.6 Differential protein expression in occurring in pancreatic tumours....55
Significantly differentially expressed proteins....57
Pathway analysis of significantly upregulated proteins....59
3.7 Protein biomarker discovery and PPI network analysis....60
Biomarker discovery....60
PPI network analysis of novel potential protein biomarkers....60
CHAPTER 4- DISCUSSION....65
4.1 Overview....65
4.2 Novel potential biomarkers of PDAC....66
Junction plakoglobin (JUP)....66
Angiopoietin-like 4 (ANGPTL4)....67
p27/kip1....69
4.3 Genetic variants might induce tumorigenesis....70
Secreted protein acidic and rich in cysteine (SPARC)....70
Zinc finger homeobox 3 (ZFHX3)....72
Transglutaminase 2 (TGM2)....72

Collagen type I alpha 1 (COLIA1) 73
4.4 Dysregulated pathways that induce PDAC progression 73
Insulin/IGF pathway-mitogen activated protein kinase/MAP kinase cascade 74
Integrin signalling pathway 75
Ras pathway 75
CCKR signalling pathway 76
PDGF signalling pathway 76
EGF receptor signalling 77
p13k/Akt signalling pathway 78
Angiogenesis- the role of VEGF signalling pathway 79
4.5 Downregulation of crucial genes aid PDAC progression 80
CXXXC Finger protein 4 (CXXC4) 80
Polypeptide N-actylgalactosaminyltransferase 1 (GALNT1) 81
Heparan sulfate 3-O-sulfotransferase (HS3ST4) 81
Solute carrier family 15 member 4 (SLC15A4) 82
Tetraspanin (TSPAN7) 82
4.6 Necrosis as a prognostic marker in pancreatic cancer 83
4.7 The immune system is active during PDAC 84
CONCLUSION 86
APPENDICES 88
REFERENCES 115

LIST OF FIGURES

Figure 1.1: Diagram of the pancreas showing its four major regions- head, neck, body, and tail. Adapted from (Skandalakis et al., 2004.). 1

Figure 2.1: Flowchart of methods undertaken in the study. 19

Figure 3.1: Heat map of fold changes observed in Group 1 (a) and Group 2 (b) for the various genes within the array. Colour bar indicating expression levels is also shown. Fold regulation of genes can be observed as colours indicated in the colour bar. Negative values indicate downregulation and positive values, upregulation. Individual fold change values are shown in figure A3 in the appendix section. 33

Figure 3.2: Diagrammatic representation of enriched pathways of downregulated genes unique to Group 1 (a) and Group 2 (b). Orange coloured lines indicate enriched pathways. Group 1 is normal tissue vs tumour tissue and Group 2 is normal tissue vs MIA PaCa-2 cell line. Generated from REACTOME 34

Figure 3.3: Diagrammatic representation of enriched pathways of upregulated genes unique to Group 1 (a) and Group 2 (b). Orange coloured lines indicate enriched pathways. Group 1 is normal tissue vs tumour tissue and Group 2 is normal tissue vs MIA PaCa-2 cell line. Generated by REACTOME. 36

Figure 3.4: Venn diagrams showing differentially expressed genes. (a) Genes upregulated in sample sets. 4911 genes are common to both sets. (b) Genes downregulated in sample sets. 5 genes common to both sets. 39

Figure 3.5: Heatmap showing differentially expressed genes for both sample sets. N1vsT1 indicates normal vs tumour from patient 1 while N2vsT2 indicates normal vs tumour from patient 2. Blue colour represents down regulation and Red indicates upregulation. Heatmap was generated from MORPHEUS software. 40

Figure 3.6: Diagram showing intricate gene network of upregulated genes. Red nodes indicate upregulated genes. Red edges (lines) indicate interactions. A magnified section of the gene network shows a clearer interaction of the genes. Generated from FunRich. 44

Figure 3.7: Gene network of downregulated genes including interactions. Nodes on the left indicate downregulated genes. On the right are interacting partners of downregulated genes. The bigger the sizes of nodes on the left, the greater the certainty of interaction. Key to colour edges can be seen at the bottom left of the diagram. 46

Figure 3.9: Protein-protein interaction network of JUP. The network has a PPI enrichment p- value of 1.38E-05. Edges represent various types of interactions as indicated above. Generated from STRING. 49

Figure 3.10: Schematic diagram showing the number of resulting variants and genes as each filter is applied. 3453 variants and 359 genes were reduced to 87 and 30 respectively after 6 filters. 52

Figure 3.12: Bar charts of upregulated proteins. Proteins in chart (a) have maximum spot intensity of 40000 while those in chart (b) have a maximum of about 6000. Blue bar indicates normal expression and Red indicates tumour tissue expression. Standard error bars are also shown..... 58
Figure 3.13: Diagrammatic representation pathways in which significantly upregulated proteins are enriched. Orange lines indicate enriched pathways. Generated from REACTOME software..... 59
Figure 3.14: Protein-Protein interaction network of ANGPTL4. The network has a PPI enrichment p-value of 4.44E-16. Edges represent various types of interactions as indicated above. Generated from STRING..... 61
Figure 3.15: Protein-Protein interaction network of CDKN1B/p27. The network has a PPI enrichment p- value of 3.03E-06. Edges represent various types of interactions as indicated above. Generated from STRING. 63
Figure A1: Samples passed all tests (Reverse transcription control, Genomic DNA contamination and PCR array reproducibility). The Inquiry indicated on the PCR array reproducibility test was due to the average C_T of all samples exceeding 22. However, due to differences in sensitivity of real time machines this can be acceptable provided values did not vary more than 2 cycles between the arrays which was the case in this experiment..... 94
Figure A2: The software automatically did normalization. The software chooses the most stable housekeeping gene. The C_T values were geometrically averaged and used for normalization. B2M was the most stable. 95
Figure A3: Array plate layout showing fold changes of Group 1(a) and Group 2(b) genes. Genes marked with “A” indicate that the gene is observed to relatively low in one sample but relatively high in the other..... 96
Figure A4: Bar chart showing top transcription factors in which our upregulated genes are enriched. Percentages of genes enriched are shown on top of each bar. Yellow dots indicate p-value of less than 0.001 and red indicates p-values of 0.05. 101
Figure A6: Pie chart showing pathways associated with the observed genetic variants. The number of variants enriched in each pathway is seen in parenthesis. 106
Figure A7: Pie chart showing diseases associated with the observed genetic variants. The number of variants enriched in each disease is seen in parentheses. 107

LIST OF TABLES

Table 1.1: Pathways and regulatory processes genetically altered in pancreatic cancer….....15
Table 2.1: Number of reads obtained per sample…… 24
Table 3.1: Top 10 pathways enriched in downregulated genes unique to Group 1 and Group 2……35
Table 3.2: Top 10 pathways enriched in upregulated genes unique to Group 1 and Group 2….....37
Table 3.3: Showing top 10 pathways enriched by genes upregulated in tumour tissues …….41
Table 3.4: Showing top biological processes enriched by genes upregulated in tumour tissues …….42
Table 3.5: Showing pathways enriched by downregulated gene network components……47
Table 3.6: Results of real time PCR performed on potential biomarkers……..48
Table 3.7: Predicted functional partners of *JUP* PPI network……...50
Table 3.8: Top 5 biological processes enriched in *JUP* PPI network……50
Table 3.9: Top 5 pathways enriched in *JUP* PPI network…….51
Table 3.10: Top 10 pathways enriched in the observed genetic variants……53
Table 3.11: Top 10 biological processes enriched in the observed genetic variants……..54
Table 3.12: Top 10 diseases enriched in the observed variants…….....54
Table 3.13: Showing manually filtered variants resulting in loss of gene function……...55
Table 3.14: Significantly upregulated proteins with the p-values…….….57
Table 3.15: Showing Top 10 enriched pathways their p-values and candidate proteins……60
Table 3.16: Predicted functional partners of ANGPTL4……...61
Table 3.17: Top 5 biological processes in which the ANGPTL4 PPI network is enriched……...…62
Table 3.18: Top 5 pathways in which the ANGPTL4 PPI network is enriched…….62
Table 3.19: Predicted functional partners of CDKN1B/p27……..63
Table 3.20: Top 5 biological processes in which the CDKN1B/p27 PPI network is enriched. …….64
Table 3.21: Top 5 pathways in which the CDKN1B PPI network is enriched……64
Table A1: Genetic biomarkers of Pancreatic cancer……..88
Table A2: Metabolomic biomarkers of pancreatic cancer…….89
Table A3: Proteomic biomarkers of pancreatic cancer……..……...90-92
Table A4: Circulating Endothelial Cells as biomarkers of Pancreatic cancer……..…….93
Table A5: Salivary Transcriptomic Markers of Pancreatic cancer……..…..93
Table A6: List of genes uniquely upregulated and downregulated in Group 1 and 2……97
Table A7: Showing top 10 molecular functions enriched by upregulated genes……98
Table A8: Showing top 10 cellular components enriched by upregulated genes……99
Table A9: Showing top 10 protein classes enriched by upregulated genes……100
Table A10: Showing functional enrichment analysis of downregulated genes……103-104

Table A11: Top 5 molecular functions enriched in JUP PPI network……...105
Table A12: Top 5 molecular functions in which the ANGPTL4 PP1 network is enriched……..108
Table A13: Top 5 molecular functions in which the CDKN1B/p27 PP1 network is enriched……109
Table A14: Showing patients used in study with relevant information……...110
Table A15: List of kits used with their manufacturers and catalogue numbers……111
Table A16: List of Instruments used with their manufacturers and catalogue numbers……...112
Table A17: List of software used with the link to their websites……..113
Table A18: List of Reagents/Chemicals used with their manufacturers and catalogue numbers…….114

CHAPTER 1- LITERATURE REVIEW

1.1 Introduction

1.1.1 The pancreas

The pancreas is an organ of the digestive system. It is approximately 10 to 25 cm long with a flattened pear shape that stretches across the abdomen. Generally, the pancreas is divided into four regions; head, neck, body, and tail (figure 1.1).

1. Head: The widest region of the pancreas and is found in the right side of the abdomen, sitting on the curve of the duodenum (first part of the small intestine). The head also has subdivisions: head proper and the uncinate process. Using the hand as an analogy the head proper is the palm whilst the uncinate process correlates to the pinkie and thumb.
2. Neck: This is the region between the head and the body.
3. Body: The mid region of the pancreas found between the neck and the tail. The vein and the superior mesenteric artery pass behind this region.
4. Tail: Located on the left side of the abdomen (close to the spleen) and found at the end of the pancreas forming a thin tip.

Importantly, 60-70% of Pancreatic ductal adenocarcinomas (PDAC)- the cancer of interest in this study- are found at the head of the pancreas while 20-25% are observed at the body/tail (Modolell et al., 1999).

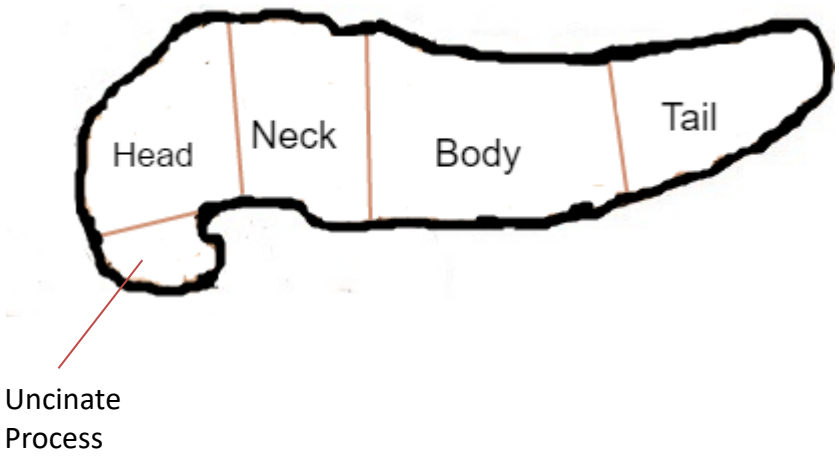


Figure 1.1: Diagram of the pancreas showing its four major regions- head, neck, body, and tail. Adapted from (Skandalakis et al., 2004.).

1.1.2 Location of the pancreas

The pancreas is situated at the back of the abdomen behind the stomach. Its location makes it difficult for early proper diagnosis of PDAC. Furthermore, the pancreas is surrounded by different blood vessels that supply blood to it and other neighbouring organs. This means that it is easy for PDAC to invade blood vessels thereby making treatment more difficult and complicated These blood vessels include the superior mesenteric artery, the celiac axis, the superior mesenteric vein, and the portal vein. The head of the pancreas can be seen at the interface between the early part of the small intestine and the stomach. The location of the pancreas at this interface is crucial as it can release digestive enzymes to partially digested food passing from the stomach into the and small intestine. Furthermore, this location is critical as it masks PDAC symptoms in the patient consequently leading to misdiagnosis at its early stages.

1.1.3 Development of the pancreas

The pancreas originates from the embryonic foregut. Embryonic development of the pancreas commences with the origination of the dorsal and ventral pancreatic buds. Both buds are linked to the foregut by ducts. The dorsal bud develops into the head, body, and tail while the ventral bud is responsible for the formation of the ucinate process. The rotation and fusion of dorsal and ventral pancreatic buds produces the pancreas.

Differentiation of pancreatic cells occurs in 2 divergent pathways which giving rise to the endocrine and exocrine characteristics of the pancreas. In the exocrine pancreas, progenitor cells contain crucial molecules that initiate differentiation. These molecules include fibroblast growth factors, those of the Notch receptor system and follistatin. In PDAC, these molecules have been observed to be overexpressed. This study seeks to identify more important molecules and pathways involved in pancreatic development that could be overexpressed during PDAC. NEUROG3 is expressed by endocrine progenitor cells facilitating the development of the endocrine pancreas region (Gradwohl et al., 2000).

1.1.4 Regulation of pancreatic development

A multipotent group of cells located in the foregut endoderm give rise to the pancreas, liver and biliary duct (Puri and Hebrok, 2010; Wandzioch and Zaret, 2009). In humans, the pancreas is first observed at four weeks of gestation as the dorsal bud emerging from the proximal end of the duodenum whilst in mice, the pancreas can be observed around E8.5 to E9.5. The

endocrine section consists of α, β, δ, ε and polypeptide cells that are arranged into islets. These cells produce the pancreatic hormones. Pancreatic development is strongly regulated with the exocrine and endocrine parts developing from the same progenitor cells. This regulation involves a coaction of Hedgehog signalling (Hebrok et al., 2000; Lau and Hebrok, 2010), Notch signalling, KRAS signalling and integrin signalling. These pathways are known to be dysregulated during PDAC and play a role in inducing tumorigenesis (Apelqvist et al., 1999; Esni et al., 2001; Murtaugh et al., 2003). Also, studies have described some transcription factors crucial for pancreatic development. These include: Pdx1 which is crucial for the stipulation of every pancreatic lineage (Gannon et al., 2000), NeuroD, ngn3, Pax4 and Hnf6, all contributing to propagation, differentiation and the commitment of endocrine lineage (Gannon et al., 2001; Offield et al., 1996). The differentiation of the exocrine lineage is influenced by both the absence of endocrine-favouring transcription factors and by the existence of permissive signals equipped by the adjoining pancreatic mesenchyme (Esni et al., 2001).

1.1.5 Pancreatic ductal cell lineage regulation during development

Markers such as cystic fibrosis transmembrane receptor, cytokeratin19 (K19) carbonic anhydrase II (CA II), DBA lectin and transcriptional factors which include HNF1β, SOX9 and HNF6 are expressed by ductal cells. During pancreatic ductal development, the HNF6 plays an important role as *HNF6*$^{-/-}$ mice develop cysts in intralobular and interlobular pancreatic ducts (Pierreux et al., 2006). Also critical to pancreatic ductal formation is HNF1β, whose activation occurs downstream of HNF6 (Maestro et al., 2003) and is decreased in pancreatic cystic structures of *HNF6*$^{-/-}$ mice. Sox9, another crucial transcription factor, is responsible for the stimulation of proliferation and survival of progenitors thereby being critical for maintaining the pancreatic progenitor pool. The overexpression of these markers might play a crucial role in enhancing excessive cellular proliferation, consequently leading to carcinogenesis. (Seymour et al., 2007).

1.1.6 Functions of the pancreas

The pancreas is made up of the exocrine and endocrine gland. Both glands have different but critical functions. The exocrine pancreas is made up of 2 types of cells: acinar and duct cells. The acinar cells are responsible for the production of digestive enzymes. These enzymes include those that digest carbohydrate (amylase), proteins (trypsin and chymotrypsin) and fats (lipase). The enzymes are produced in their inactive forms to prevent the digestion of the

pancreas by them. The duct cells are responsible for the manufacture of fluids that combine with the inactive enzymes (proenzymes) to produce pancreatic juice. From the acinar cells, pancreatic juice flows into the duodenum. In the duodenum, the pancreatic juice then mixes with ingested food during its exit from the stomach. When pancreatic juice reaches the small intestine, it causes the proenzymes to be activated. Elevated amounts of bicarbonate in the pancreatic juice neutralize the acidic nature of the stomach.

The endocrine pancreas contains millions of clustering cells called the Islets of Langerhans (Pancreatic islets). The pancreatic islets consist of 4 different cell types. Each of these cell types is responsible for the production of different hormones and play divergent roles. They include;

(i) Alpha (α) cells secrete glucagon. The pancreas releases glucagon when there is a decrease in blood glucose levels. It increases the blood glucose levels by inducing the conversion of glycogen to glucose by the liver.

(ii) Beta (β) cells which secrete insulin. Insulin is released as a response to elevated blood glucose levels. During increased blood glucose levels insulin is released and promotes the absorption of glucose into fats or glycogen in the liver or fats in skeletal muscles.

(iii) Delta (δ) cells secrete somatostatin. Somatostatin is responsible for the regulation of α and β cells. It inhibits α and β cells from producing glucagon and insulin respectively when not needed.

(iv) Gamma cells also referred to as pancreatic polypeptide cells act in the control the control of blood glucose levels by secreting glucagon and insulin when necessary.

(v) Epsilon cells constitute about 1% of the entire islet cells. They secret ghrelin which play a role in stimulating hunger.

1.1.7 Diseases of the pancreas

Pancreatic cancer (the disease of interest in this study) is a lethal and aggressive cancer. It initiates when pancreatic cells begin to grow and proliferate uncontrollably.

Exocrine tumours- They make up more than 95% of pancreatic cancer. These tumours begin in the exocrine cells that produce pancreatic enzymes which aid in digestion. Adenocarcinoma is the most prevalent of the exocrine tumours. When the tumour originates from the pancreatic ducts, it is referred to as pancreatic ductal adenocarcinoma (PDAC). PDAC constitutes approximately 90% of all pancreatic tumours making it the most frequently occurring one. Rare

types of exocrine tumours include acinar adenocarcinoma (tumours originating from acinar cells-cells responsible for the production of pancreatic juices), mucinous cystic tumours (that result fluid filled sacs or cysts in the pancreas), serous cystadenoma, colloid carcinoma, adenosquamous carcinoma, signet ring cell carcinoma, squamous cell carcinoma, solid and pseudopapillary tumours, hepatoid carcinoma, giant cell tumour and undifferentiated carcinoma (Herbst, 2016).

Endocrine tumours- They are less common than exocrine tumours. They begin in the endocrine region of the pancreas which is responsible for the production of hormones like insulin and glucagon. These tumours are also referred to as islet cell tumours or pancreatic neuroendocrine tumours (PNETS). About one-third of these tumours produce hormones which can cause unusual symptoms. They are named per the hormone they produce: insulinomas (produce insulin), glucagonomas (produce glucagon), VIPomas (produce vasoactive intestinal peptide), gastrinomas (produce gastrin), PPomas (produces pancreatic polypeptide hormone) and somastostatinomas (produce somatostatin). Two-thirds of the endocrine tumours do not produce hormones or symptoms and are referred to as non-functioning tumours. Other rare types of pancreatic cancer include lymphomas, pancreatoblastomas and sarcomas (Herbst, 2016).

In this study, only tissue samples obtained from PDAC patients was utilized.

1.1.8 Incidence of PDAC

PDAC accounts for 2.8% of new cancer cases in USA and about 7% of cancer related deaths worldwide (Siegel et al., 2012). In developed countries, it is the 4th leading cause of cancer related deaths. Pancreatic cancer has an overall 5-year survival rate which remains below 5% (Ferrone et al., 2008). In the US, African Americans have the highest incidence rate compared to any other ethnic group followed by Ashkenazi Jews with Asians having the least incidence (Lynch et al., 2004).

In South Africa, black males and females have the highest incidence rate, followed by Caucasians, Coloureds and Asians (National institute of occupational health, 2016). Due to the high incidence rate of pancreatic cancer (PDAC) amongst the black population, and thus ease of availability I have opted to use biological samples from this group only in this study.

1.1.9 Risk factors identified with PDAC

Several risk factors have been associated with PDAC. These risk factor are important as they can be utilised in screening purposes for the screening of high risk groups. Risk factors include:

Family history and inherited genetic disorders- Approximately 80% of pancreatic cancers occur sporadically but a small percentage is genetic and results from an inherited germ line mutation. Patients with a family history are 3 times more likely to develop PDAC (Fernandez et al., 1994). The most common inherited germ-line mutation in familial PDAC was identified as *BRCA2* deleterious mutations (Murphy et al., 2002). Also, a germline point mutation was observed in *PALB2* gene in 3% of familial PDAC patients (Jones et al., 2009).

Smoking- Approximately 25% to 35% of PDAC cases have been linked to smoking making it a crucial risk factor (Lowenfels and Maisonneuve, 2004). It's been suggested that smokers have a 1.74-fold risk of pancreatic cancer.

Dietary factors- Acute alcohol consumption, consumption of meats and fatty foods and increased serum folate levels enhances the risk for developing PDAC.

Pancreatitis- Chronic pancreatitis is responsible for up to 3% of PDAC cases (Lowenfels et al., 1999). Risk of developing PDAC ranges from 2.3% to 18.5% with patients suffering from chronic pancreatitis.

Helicobacter pylori- Gastric colonization by *H. pylori* is the main cause of hyperacidity and gastric diseases and is implicated in PDAC (Trikudanathan et al., 2011).

Obesity- This may lead to the formation of tumours through an abnormality in glucose metabolism, inflammation and hyperinsulinemia (Bracci, 2012). *Diabetes*: Obesity is often associated with type 2 diabetes which is associated with a doubling in the risk of PDAC. About 1% of PDAC patients with early onset diabetes which manifests about 1-3 years develop PDAC.

1.1.10 Signs and symptoms of PDAC

One of the major reasons pancreatic cancer is difficult to detect and diagnose is that it is asymptomatic at its early stages. As the cancer matures signs and symptoms begin to appear but these are dependent on the type of pancreatic cancer and its location in the pancreas. Although symptoms occur, they could be non-specific and might appear and disappear. Also, PDAC symptoms can include those illnesses that are common and not severe, consequently,

patients are misdiagnosed several times. Signs and symptoms include abdominal pain, weight loss, jaundice, nausea and vomiting, diabetes, fever and shivering as well as diarrhoea, anaemia, weakness and high blood pressure (Herbst, 2016).

1.1.11 Diagnosis of PDAC

PDAC is mostly diagnosed at its late stages. The doctor takes this following step for diagnosis:

- Physical examination to check for any growth especially in the abdominal region.
- Laboratory tests e.g liver function tests to check for bile blockage
- Ultrasound, CT (computed tomography) scan and MRI (Magnetic resonance imaging) for detailed imaging of abdominal organs.
- Biopsies, for example ERCP (Endoscopic retrogade cholangiopancreatography) in which a tube attached to a brush and a camera is passed through the mouth taking biopsies (Herbst, 2016). Other procedures include: percutaneous and EUS guided biopsies.

In most cases, before these steps are done the cancer has progressed. This emphasizes the need for biomarkers (biological markers) present in easily accessible body fluids (like blood). This study seeks to identify novel potential biomarkers for the diagnosis of pancreatic cancer.

1.1.12 Staging in PDAC

Worldwide (y the American Joint Committee on Cancer) PDAC is grouped into stages from stage 0 to stage IV. They include:

Stage 0: Tumour is located at superficial layers of the pancreatic ductal cells and has not spread outside the pancreas. These tumours can be referred to as pancreatic intraepithelial neoplasia III (PanIN III) or pancreatic carcinoma i*n situ.*

Stage I: Tumour has still not spread outside of the pancreas and is larger than 2cm.

Stage II: Tumour is growing outside pancreas and has spread to the lymph nodes but has not spread into major blood vessels or nerves and distant sites

Stage III: Tumour is growing outside pancreas and has spread to major blood vessels but not distant sites.

Stage IV: Tumour has spread to distant sites such as lungs, liver and other surrounding organs and is therefore metastatic (Herbst, 2016).

In South Africa, PDAC is classified in 3 stages:

Stage I: Tumour is resectable. This can also include borderline resectable tumours.

Stage II: Tumour is locally advanced (non-metastatic).

Stage III: Tumour is metastatic (Herbst, 2016).

In this study, only stage I tumours were utilized since I was investigating the discovery of biomarkers for early diagnosis.

1.1.13 Treatment options for pancreatic cancer

Immediately after diagnosis of pancreatic cancer, the doctor needs to ascertain the stage of the tumour. That is, the size of the tumour and the degree of which it has spread must be determined. This assessment enables the doctor to decide on which treatment option to undertake. Treatment options are more effective when PDAC is diagnosed early, this again, emphasizes the need for biomarkers for early detection. Below are the pancreatic cancer treatment options available.

1. *Surgery*- The type of surgery will depend size and location of the tumour, the patient's wellbeing and how much the tumour has spread. The different types of surgeries that can be performed include;

- Whipple procedure- This procedure is also referred to as pancreaticoduodenectomy and it is the most common type of surgery performed for the removal of tumours from the pancreas. In this case the surgeon removes the head of the pancreas, parts of the stomach called the pylorus and duodenum, gall bladder, distal common bile duct, and lymph nodes situated at the head of the pancreas. Thereafter the remaining part of the pancreas and digestive organs are reconnected to allow digestive enzymes, bile, and contents of the stomach to pass through into the small intestine during digestion. This surgery usually takes about 5-6 hours.
- Distal pancreatectomy- This type of surgery is performed when the tumour is located at the body and tail of the pancreas. The surgeon proceeds to remove these parts including the spleen.
- Total pancreatectomy- In this case the whole pancreas, parts of the stomach and small intestine, the spleen, bile duct and neighbouring lymph nodes are removed.

Neoadjuvant therapy includes the use of fluorouracil and adjuvant therapy could be gemcitabine or fluorouracil based. Adjuvant therapy increases survival of patients that have underwent surgery.

2. *Chemotherapy*- This treatment option requires the use of drugs to kill the tumour. Side effects may vary depending on the drug and patient. Gemcitabine is a widely used chemotherapeutic drug for pancreatic cancer. Lack of efficacy of gemcitabine and other chemotherapeutic agents is problematic (Lopez-Casas and Lopez-Fernandez, 2010).Other widely used chemotherapeutic agents include: 5- Fluorouracil and Folfirinox.
3. *Radiotherapy*- This treatment uses energy rays to kill the tumour. Radiation is directed to the abdomen region by a machine called the Linear accelerator (LINAC). Side effects may include vomiting, nausea, diarrhoea and indigestion.

In some cases, the tumour cannot be removed completely, for example, if the tumour blocks the small intestine or common bile duct. This situation requires the surgeon to create a bypass that allows fluid to pass through the digestive tract.

1.2 Discovering biomarkers of PDAC

At the early stages of pancreatic cancer, it is usually misdiagnosed due to the lack of efficient diagnostic tests and the similarity of its symptoms to other more common ailments. Early diagnosis is critical for increasing survival rates for patients with PDAC; making the discovery of biomarkers urgent. The characteristics of an ideal biomarker includes, being able to provide prognostic and predictive information in order to help guide decisions of clinicians and to provide adequate information on the progression of the disease (Pepe et al., 2008). It can also be used to assess efficacy of treatment administered to a patient over a period. Also, a good biomarker should be expressed in bodily fluids especially in the blood. This makes for patients to be diagnosed without any need for invasive procedures. Furthermore, a biomarker must be very specific and sensitive for the disease condition. Over the years several potential biomarkers of pancreatic cancer have been discovered and studied using various techniques such as transcriptomics, metabolomics, and proteomics but most have not passed validation as they lack adequate sensitivity and specificity. Appendix A1-A5 contains some potential biomarkers discovered over the years. Amongst them is the most widely studied, CA 19-9. CA

19-9 is being used mainly in assessing prognosis and in the monitoring of patients under therapy but has disadvantages. CA 19-9 is elevated in several other cancers including hepatocellular, colorectal and oesophageal. It has also been shown to be increased in several other diseases including pancreatitis. Its non-specificity makes it unsuitable for diagnosis and screening of pancreatic cancer. Additionally, CA-19-9 is not expressed in patients lacking Lewis antigen and thus cannot be used to diagnose this group (Chan et al., 2013). These disadvantages hint for the discovery of specific and sensitive biomarkers for the diagnosis of early asymptomatic PDAC.

1.2.1 Gene expression profiling (Transcriptomics) in PDAC

Gene expression profiles vary across individuals and population groups. Approximately 83 % and 17% of genes have been observed to be differentially expressed amongst individuals and populations respectively (Storey et al., 2007). This hints for gene expression studies to be conducted amongst several population groups. The emergence of high-throughput genomic technologies has enabled gene expression profiling in pancreatic cancer studies. These studies have allowed different groups to pursue the discovery and validation of biomarkers for the early detection (diagnostic) and classification of PDAC. Furthermore, these technologies might help to identify novel therapeutic targets for pancreatic cancer treatment (Lopez-Casas and Lopez-Fernandez, 2010). Analysing the gene expression profiles aids the description of how genes react to pancreatic cancer progression and this helps in the formulation of preliminary hypothesis regarding the disease condition. Using various techniques, researchers can identify differentially expressed genes by comparing normal pancreatic gene profiles to their cancerous counterparts. One of the first published studies used the serial analysis of gene expression (SAGE) technique to analyse thousands of transcripts using bulk tissues obtained from pancreatic cancer patients (Argani et al., 2001). The study identified mesothelin as being overexpressed in most PDAC tissue samples making it a putative marker. One major problem observed in pancreatic cancer studies involving gene expression profiling of bulk tissues is extracting good quality RNA. Therefore, stringent RNA extraction techniques must be utilized to achieve this.

Grutzmann *et al*. performed one of the first meta-analyses studies in which they profiled genes involved in PDAC. 568 genes were identified to be significantly up/down regulated during PDAC (Grutzmann et al., 2005). This study uniquely identified more genes dysregulated during pancreatic cancer as only 22% were previously described in published literature. Jones and his

colleagues provided comprehensive work in which they integrated genomic and transcriptomic (SAGE) data. They showed essential signalling pathways and biological processes occurring in many pancreatic cancers. These pathways include Wnt/Notch signalling, KRAS signalling, TGF-β signalling, regulation of invasion, DNA repair, small non-KRAS GTPase-dependent signalling, homophilic cell adhesion, c-Jun N-terminal kinase signalling, Hedgehog signalling, regulation of G1/S phase transition and integrin signalling (Jones et al., 2008a). The authors claim that the identification of several core pathways indicates the involvement of many genes in PDAC and this hints to pancreatic cancer progression might be patient specific. This point of view encourages the need for personalized medicine in the treatment of PDAC. Nevertheless, in the discovery of biomarkers for prognostic, diagnostic, disease staging and monitoring treatment outcomes, gene expression profiling will help.

Diagnostic biomarkers aid in the early detection of disease and enable for appropriate treatment to be administered. The early detection of PDAC is crucial as it increases the chances of survival in patients. Over the years, gene expression profiling has been used for the discovery of novel potential biomarkers. One study on IPNM (Intraductal papillary mucinous neoplasm), a PDAC precursor lesion, samples utilizing microarrays revealed four genes (*GAS6*, *GPC1*, *CAV1* and *CYR61*) that were differentially expressed and proposed as potential biomarkers (Terris et al., 2002). In another study using microarrays, Buchholz and colleagues identified 30 differentially expressed genes occurring in PDAC. Amongst these genes are, *S100P*, *IF127*, *IL-13RA1* and *RABAC1* all of which were significantly upregulated (Buchholz et al., 2005). *S100P* has been further used in some cases as a diagnostic biomarker (Deng et al., 2008). Although, several studies have identified potential biomarkers (like *S100P*), they fail to pass validation experiments. This failure is mostly attributed to lack of specificity and sensitivity to pancreatic cancer.

Prognostic biomarkers help determine the progression of PDAC and gene profiling helps identify these markers. Using SW1990 (a pancreatic cancer cell line) and SW1990HM (a liver metastasized pancreatic cancer clone), 40 genes were shown to be involved during pancreatic cancer associated liver metastasis. Campagna and colleagues did a comparative study of stage 4 and stage 2/3 pancreatic cancers and observed 173 genes dysregulated between the two groups (Campagna et al., 2008). Amongst these genes, four were further validated using real-time PCR but only *MXI1* was confirmed as upregulated in stage 4 compared to stage2/3. Further validation studies on the use of *MXI1* as a prognostic marker is yet to be conducted.

Although a good amount of work has been done to identify prognostic biomarkers of pancreatic cancer, a lot more must be done to identify better predictors of prognosis.

1.2.2 Protein expression profiling (proteomics) in pancreatic cancer

Several studies have indicated that although genomic level mutations can affect protein structure and function, mRNA expression does not correspond to protein expression(Pan et al., 2013). Furthermore, although investigations into mRNA expression levels help provide a better understanding of any disease, they do not accurately correspond to progression of the ailment. Protein profiles however provide a more accurate understanding of disease states and progression (Grantzdorffer et al., 2008). Also, crucial signalling molecules in important protein molecules cannot be identified by studying mRNA expression profiles.

Studying the protein expression profiles in diseased conditions such as cancer has become critical in elucidating the initiation and progression of such diseases. Proteins are essential biomolecules for the induction and regulation of biological processes. In cancer studies, focus is mostly on the identification of important proteins, protein interactions and significant pathways involved in tumorigenesis. Proteomics enable the analysing of the whole proteome of a cell, tissue, body fluids and organs. It facilitates the understanding of the structure and function of crucial proteins (Baak et al., 2003; Grantzdorffer et al., 2008). Also, proteomic tools are beneficial for the discovery of novel potential biomarkers and therapeutic targets in pancreatic cancer treatment. Recent proteomics studies using varying sample types- serum, tissue, pancreatic juice, pancreatic cancer cell lines-, have identified some proteins differentially expressed during PDAC. However, they still lack the specificity and sensitivity to adequately detect PDAC in asymptomatic patients (Chen et al., 2005a).

Proteomic studies have revealed some potential diagnostic and prognostic biomarkers of PDAC. Table A3 provides a list of these potential biomarkers including the various proteomic technologies utilized for their discovery. One of the first published studies on the discovery of protein biomarkers for pancreatic cancer utilized immunoassays. These studies identified CEA and CA-50 as potential biomarkers (Haglund et al., 1987; Molina et al., 1990). Furthermore, mass spectroscopy is one of the most utilized technologies for the discovery of protein biomarkers. Some of the biomarkers identified using this method include; galectin-1, gelsolin, lumican, cathepsin D, plectin-1, moesin, cofilin, (Chen et al., 2005b; Lu et al., 2004; Shen et al., 2004), pancreatic lipase related protein-1, pre-proprotein and serine proteinase-2 (Gao et al., 2010). More recently, Glypican 1 was identified as a potential biomarker for pancreatic

cancer (Melo et al., 2015). Enzyme-linked immunosorbent assays (ELISA) have also been used for protein biomarker discovery in pancreatic cancer studies. Some of the potential biomarkers identified using ELISA include Cytokeratin 18 (Dive et al., 2010) and serum REG4 (Takayama et al., 2010). To determine the efficacy of these potential biomarkers, they would have to be independently tested multiple times using large number of samples obtained from PDAC patients.

Protein-Protein interaction (PPI) networks- Role in pancreatic cancer research

Most proteins carry out their functions within an intricate network consisting of other proteins. One component within this network can affect many others. The upregulation or downregulation of a protein might result to more or lesser connections being made. This consequently might lead to the initiation and propagation of pathological conditions (1). PPI networks help to understand functions of recently discovered proteins of interest or provide clue of new functions for already known ones (Sevimoglu and Arga, 2014) . These networks also aid to identify protein sets within its environment and discovers known/predicted interactors. Furthermore, they help to identify biological processes, pathways and protein domains enriched within the network (Dai et al., 2015). The analysis of these enriched pathways and biological processes help to elucidate disease conditions.

PPI network analysis have been applied to a few pancreatic cancer studies. Yuan and colleagues constructed a large network utilizing information from protein-protein interactions (Yuan et al., 2015). They then used a shortest part method to mine 13 novel genes involved in pancreatic cancer. They proposed that these genes might have therapeutic benefits and as future drug targets. In their gene expression analysis of normal and cancerous pancreatic tissues, Long *et al*., used PPI network analysis to predict effectively protein interactors, biological processes and pathways in which their differentially expressed genes were enriched (Long et al., 2016). PDAC is a disease whose aggressive progression is still poorly understood, PPI network analysis might help to further elucidate it.

1.2.3 Alternative splice variants as biomarkers of PDAC

Alternative splice variants are promising for use as biomarkers for PDAC (Omenn et al., 2010). They have been identified to be common all over the human genome. Alternative splicing enables a gene to give rise to multiple proteins. Splice variants that affect the protein coding region of the mRNA results in proteins with aberrant structure and function. Alternative splice

variants affect gene regulation and/or structure of proteins (Hartel et al., 2008; Omenn et al., 2010). Splice variants are generally stable and can be quantified.

Using exon array analysis, splicing variants associated with cancers of the bladder, colon and prostrate have been identified at the mRNA level (Thorsen et al., 2008). These variants are observed to have both diagnostic and prognostic consequences. Hayes and colleagues postulated the use of splice variants as potential biomarkers of pancreatic cancer (Hayes et al., 2007). Carrigan and colleagues utilized spliceoform-specific microarrays and PCR to identify 28 genes encoding spliceosomal proteins that are underexpressed during pancreatic cancer (Carrigan et al., 2011). Also, abnormal expressions of AP-2α have been identified in pancreatic cancer. Using RT-PCR (for gene level analysis), immunofluorescence and western blotting (for protein level analysis), Carriere and colleagues showed that various isoforms of AP-2α are expressed in multiple pancreatic cancer cell lines (Carriere et al., 2011). Amongst these isoforms identified was AP-2α variant 6 which seemed to pancreatic cancer specific. Furthermore, using similar techniques as Carriere *et al.*, Arafat and colleagues identified alternative splicing in various tumour specific exons of stromal COL6A3. They postulate that the resulting spliced proteins might be involved in the initiation and development of pancreatic cancer and serve as potential biomarkers.

1.2.4 Signalling Pathways altered during PDAC

The importance of signalling pathways in the elucidation of tumour progression has been identified (Hanahan and Weinberg, 2000; Vogelstein and Kinzler, 2004). The study of pathways involved in tumour progression helps in understanding the disease (Chittenden et al., 2008; Edelman et al., 2008; Lin et al., 2007). During pancreatic cancer, genes involved in core signalling pathways are altered. These alterations consequently lead to the progression of pancreatic cancer.

Jones and colleagues have identified 12 core signalling pathways in which some genes are altered during pancreatic cancer (Table 1.1), although there was no shown certainty that all the altered genes play functional roles (Jones et al., 2008a).

Table 1.1: Pathways and regulatory processes genetically altered in pancreatic cancer.

Pathway or regulatory process	**Some altered genes**
Wnt/Notch signalling	*TCF4, GATA6MYC, PPP2R3A, TSC2, MAP2, WNT9A*
KRAS signalling	*KRAS, RASGRP3, MAP2K4*
Hedgehog signalling	*CREBBP, TBX5, LRP2, SOX3, GL13, GL11, BOC, BMPR2*
DNA damage control	*TP53, EP300, ERCC6, ERCC4, RANBP2*
Regulation of G_1/S phase transition	*APC2, CHD1, FBXW7, CDKN2A*
Homophilic cell adhesion	*CDH1, CDH7, CDH10, PCDH17, PCDH9*
Integrin signalling	*LAMA1, LAMA4, LAMA5, ILK, FN1, ITGA4, ITGA11*
Regulation of invasion	*DPP6, PCSK6, APG4A, PRSS23, ADAM11, ADAM12, ADAMTS15*
TGF-β signalling	*SMAD4, SMAD3, BMPR2, TGFBR2*
Epidermal growth factor receptor (EGFR)	*ErbB-2, ErbB-1, ErbB-4*
NF-kB	*Rel A, RelB, c-Rel, P100/p52*
Small GTPase-dependent signalling	*RP1, PLCB4, DEPDC2, PRKCG, AGHGEF7, CDC42BPA*

Source: Adapted from (Jones et al., 2008a)

Furthermore, Waddell and colleagues performed whole genome sequencing and copy number variation analysis using 100 PDAC samples. They observed chromosomal arrangements leading to the disruption of key genes (*TP53*, *CDKN2A*, *SMAD4*, *ROBO2*, and *ARID1A*) that are crucial to pancreatic cancer (Waddell et al., 2015). They also identified structural variations in novel drivers of pancreatic tumorigenesis (*PREX2* and *KDM6A*). Most of these variants identified were duplications, deletions, inversions, and chromosomal arrangements. The most frequently identified genetic alterations are those of *K-Ras*, *Tp53*, *CDKN2A*, *BRCA2* and *DPC4*(Jones et al., 2008b). Mutations of the *K-Ras* gene are known to initiate tumour growth in the pancreas. This is then followed by a loss mutations of p16, Tp53 and SMAD4 (Polireddy and Chen, 2016). These alterations result in the suppression of apoptosis, induction of cellular proliferation and the enhancement of tumour invasion & metastasis (Korc, 1998). Studies identifying more genetic alterations are crucial in further elucidating the progression of PDAC.

The overexpression of growth factors in signalling pathways have been observed in pancreatic cancer cell lines. These growth factors include; Epidermal growth factor (Korc et al., 1986), fibroblast growth factor (Kornmann, 1997) and platelet derived growth factor (Weissmueller et al., 2014). However, studies with pancreatic tumours are necessary to corroborate these findings.

1.3 Justification of study

Pancreatic cancer (PDAC) accounts for 2.8% of new cancer cases and about 7% of cancer related deaths worldwide (Siegel et al., 2012). In the United states, it is projected to be the 2nd leading cause of cancer-related deaths by 2030 (Polireddy and Chen, 2016). PDAC is usually diagnosed at late stages which is responsible for its low survival rate. This emphasizes the need for biomarkers that can detect PDAC at asymptomatic stage. Profiling genes and proteins that play a role during pancreatic cancer provide a good way to identify novel potential biomarkers to aid in the early diagnosis of pancreatic cancer. Previous studies have identified potential biomarkers for pancreatic cancer, however, the challenge for lack of sensitivity and specificity to detect early stage PDAC is still prevalent. This study seeks to overcome this challenge in four major ways. First, in using stringent sample handling and processing techniques to retain integrity of samples. This is crucial as poor sample quality can be responsible for misleading and/or non-reproducible results. Second, by utilizing both transcriptomic and proteomic discovery methods. This gives us a broader pool of candidates to analyse for biomarker discovery. Third, with the use highly sensitive, specific, and high throughput gene and protein expression methods which include focused array panels and next generation sequencing. These highly sensitive methods reduce the occurrence of false positive results and enable the detection of low quantity occurring gene/proteins. Pathway arrays enabled for cancer-focused analyses and the identification of dysregulated biological processes/pathways. Fourth, by employing an array of bioinformatics tools, for example, the Ingenuity pathway analysis (IPA) biomarker filtering tool help to identify only the best candidate biomarkers relevant to PDAC. Furthermore, this study also uniquely utilizes samples from a black African population. It has been observed that approximately 17% of genes vary within populations (Storey et al., 2007), as there is currently no published PDAC gene expression data from any African population, it is pertinent to conduct this study.

1.4 Aim

The overall aim of this study was to identify novel potential biomarkers of pancreatic cancer.

1.5 Objectives

1. Comparative focused gene expression profiling of pancreatic tumours and cell line (MiaPaca-2) using RT^2 Profiler PCR arrays to identify differentially dysregulated pathways.

2. Transcriptome profiling of pancreatic tissues (tumour and normal) using RNA sequencing to identify differentially expressed genes and dysregulated pathways.

3. Gene biomarker discovery using Ingenuity pathway analysis tool to identify potential biomarkers.

4. Protein-protein interaction network analysis of potential gene biomarker using STRING software.

5. Validation of potential gene biomarkers using real-time PCR.

6. Genetic variant discovery using Ingenuity variant analysis tool.

7. Focused protein expression profiling of pancreatic tissues (tumour and normal) using antibody arrays to identify differentially expressed proteins and dysregulated pathways.

8. Protein biomarker discovery using Ingenuity pathway analysis tool to identify potential biomarkers.

9. Protein-protein interaction network analysis of potential protein biomarker using STRING software.

CHAPTER 2- METHODS

2.1 Materials

All materials including kits, reagents, instruments, and software used in this study can be found in the appendix section (Tables A15- A18)

2.2 Methods

2.2.1 Summary of experimental strategy

Except for sample collection, sample storage, RNA sequencing and differential gene expression analysis, every method utilized was done by me. Tissue samples (normal and tumour) were obtained from black South African pancreatic cancer patients at Chris Hani Baragwanath hospital, Johannesburg, South Africa by means of the Whipple procedure. Samples were processed (as described in Section 2.2.3) and total RNA and protein were extracted (see 2.2.4) from tissues. The Human cancer PathwayFinder RT^2 Profiler PCR array – a focused pathway analysis platform- was then used for gene expression profiling of tissue samples. The Human Cancer PathwayFinder RT^2 Profiler PCR array was also used to compare the gene expression profile of tissue samples to that of a widely used pancreatic cancer cell line (MIA PaCa-2). After obtaining differentially expressed genes (DEGs) in tissue samples and the cell line, functional enrichment analyses were performed to identify dysregulated pathways. Furthermore, RNA sequencing was performed using tissue samples and differentially expressed genes (DEGs) obtained. Utilizing a range of bioinformatics tools (Table 2.3), functional enrichment analyses were then performed on DEGs obtained from the RNA sequencing. Also, variant and biomarker filtering analyses were achieved using the Ingenuity variant analysis and Ingenuity pathways analysis tools respectively. Having identified novel potential biomarkers (from the biomarker filtering analysis), real-time PCR was used for preliminary validation studies. Additionally, the Human Oncology Proteome profiler array – a focused proteomic platform- was used for protein expression profiling of tissue samples. Differentially expressed proteins were observed and functional enrichment analysis performed. Biomarker filtering analysis was then performed to identify novel potential protein biomarkers. The protein-protein interaction network analyses of identified biomarkers were also done using STRING software.

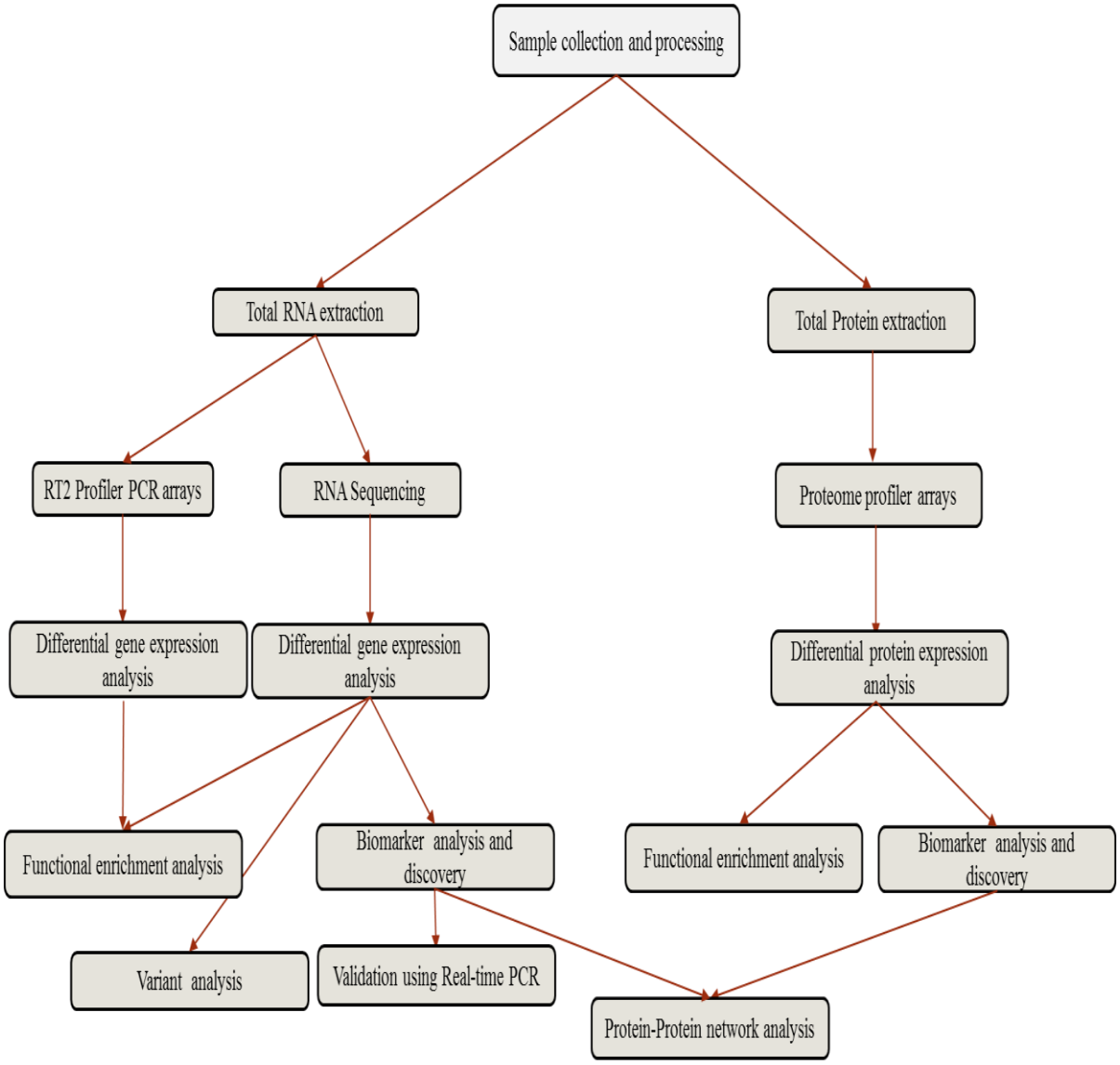


Figure 2.1: Flowchart of methods undertaken in the study.

2.2.2 Ethics statement and Funding

The Human ethics committee of the University of Witwatersrand, Johannesburg, South Africa, approved the collection of biological samples from pancreatic cancer patients (Ethics clearance certificate number- M150778). The Director of Chris Hani Baragwanath Johannesburg, South Africa also approved sample collection on behalf of the hospital board of directors. Patients were required to sign consent forms (included in ethics application).

The Medical Research Council of South Africa provided funding for most of this study. Dr John Devar also made funds to be available for RNA sequencing. There was no conflict of interest.

2.2.3 Sample collection, storage, and processing

A total of 9 (nine) patients were used for this study (see Table A14 for information). A total of 18 tissues (9 tumour and corresponding normal tissues). All tissues were obtained from black South African pancreatic cancer patients undergoing Whipple procedure at Chris Hani Baragwanath Johannesburg, South Africa. Procedure was conducted by Drs Martin Brand, Omoshoro Jones, and John Devar. Samples were collected and stored by staff at the Department of Surgery, University of Witwatersrand. Tissues of about 20-50 mg were then transferred into 1.5 ml cryotubes containing about 1 ml RNA*later* RNA stabilization reagent. Biopsies were kept at 4°C overnight and then stored at -70°C till it was ready to be processed. Tissue samples stored using RNA*later* are stable for 1 week at 25°C, 1 month at 4°C and 1 year at -70°C.

Sample homogenization.

Tissues

For gene expression profiling using the Human Cancer PathwayFinder RT^2 Profiler PCR arrays, 1 normal and 2 tumour tissue samples were used. 30 mg of each tissue sample was homogenized in 600 µl of lysis buffer (provided in the Qiagen Allprep DNA/RNA Mini kit) using a TissueRuptor. The lysate was then centrifuged for 3 min at maximum speed.

For gene expression profiling using RNA sequencing, samples from 2 patients (patients 1 and 2) was used. From each patient 1 tumour and corresponding normal tissue was obtained, that is, 2 normal and 2 tumour samples (same samples used for the Human Cancer PathwayFinder RT^2 Profiler PCR arrays) were used. Homogenization was performed as previously described.

For Validation of potential gene biomarkers using real- time PCR, samples from 5 patients (patients 3,4,5,6 and 7) was used. From each patient 1 tumour and corresponding normal tissue was obtained, that is, 5 normal and 5 tumour samples (different samples from the ones previously used) were utilized. Homogenization was performed as previously described.

For proteomic profiling using the Proteome profiler arrays, 3 normal and 3 tumour tissues (patients 6,8 and 9) were used. 40 mg of tissue was placed in lysis buffer (0.5% igepal, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulphate, 150 mM sodium chloride and 50 mM Tris-HCL (pH 7.5) containing a protease inhibitor cocktail of Aprotinin (0.5 µg/ml) and PMSF

(1 mM). Tissues were then homogenized using TissueRuptor and centrifuged at 2000 xg for 10 mins at 4°C.

Cell Culture

For this study, we used the pancreatic cancer cell line, MIA PaCa-2, which was obtained from a 65-year-old Caucasian male. Pancreatic cancer cells **(**MIA PaCa-2) were cultured with 10% FBS in DMEM and 1% penicillin-streptomycin. For RNA extraction, media was carefully removed from flask and adherent cells washed with 1 ml 1X PBS. Cells were trypsinized to detach adherent cells using 700 µl of Trypsin-EDTA. Fresh media was then added to resuspend cells. Resuspended cells were transferred to 15 ml tubes and centrifuged at 1000 xg for 5 mins to pellet cells. The supernatant was carefully discarded and 350 µl of Buffer RLT Plus (provided in the Qiagen Allprep DNA/RNA Mini kit). The cells were then lysed by pipetting up and down severally.

2.2.4 Total RNA and Protein isolation

Total RNA isolation

Total RNA was extracted using the Qiagen Allprep DNA/RNA Mini kit and procedure performed following the manufacturer's instruction.

DNase Digestion

This was done using a DNase kit. The kit contains a 10X Reaction buffer, DNase 1 and stop solution. 5 µl each of 10X Reaction buffer and DNase 1 was added to 50 µl of RNA sample. The combination was gently mixed and incubated for 15 mins at room temperature. 5 µl of stop solution was then added to bind magnesium and calcium ions thereby inactivating DNase 1. The mixture is then heated at 70°C for 10 mins and then chilled on ice.

RNA quantification and purity assessment

After DNase digestion, Total RNA was quantified using the NanoDrop spectrophotometer. A ratio of absorbance at 260 nm and 280 nm of above 1.8 was observed in every sample. Furthermore, an RNA agarose gel was run for each sample to ascertain integrity.

RNA Agarose gel electrophoresis

A sterile 500 ml beaker was placed in a fume hood and 1.95 g agarose, 108.23 ml RNase free water, and 13 ml 10X MOPS added. The mixture was boiled in a microwave until dissolved completely. 8.78 ml formaldehyde and 6.5 µl ethidium bromide were then added to the

dissolved mixture in a fume hood. Once the mixture cooled, it was poured into a gel tray with comb already inserted. The mixture took about 20 mins to solidify. While gel was solidifying the running buffer (1 X MOPS) and loading samples were prepared. For a 20 µl sample, 2 µl of RNA loading dye was added. This made up the loading samples. After the gel had solidified, the comb was carefully removed. The running buffer was poured into the gel tank and the gel placed inside. Samples were then loaded into the wells and the gel covered. Gel tank was plugged into a power supply and voltage set at 100 V. After about 120 mins the gel was viewed using a Gel Doc ^{TM} XR+ imager to observe RNA.

Protein Extraction

The supernatant obtained from the tissue sample homogenization technique was transferred into a new tube. For each sample 10 µl was transferred for a new tube to be used for protein quantification. The remaining samples were stored at -70°C for proteome profiling assays.

Protein quantification

Protein quantification was performed using the Quick start™ Bradford protein assay kit. 20 µl of the standard (BSA 2 mg/ml) and unknown samples were pipetted into 1 ml cuvettes. 1 ml of the 1X Dye reagent was then added and mixed repeatedly. The mixture was then incubated at room temperature for 10 mins. The spectrophotometer was set to 595 nm and blanked using a mixture of 20 µl water and 1 ml of 1X Dye reagent. The absorbance of standard and unknown samples at 595 nm were measured. A standard curve was created by plotting the absorbance values at 595 nm values (y-axis) versus their concentrations in µg/ml (x-axis). The unknown sample concentrations were then determined from the plot.

2.2.5 cDNA First strand synthesis

The RT^2 First strand kit was used in the synthesis of cDNA from mRNA in the total RNA. A genomic elimination mix was prepared by adding 2 µg of RNA, 2 µl of Buffer GE and RNase free water to a total volume of 10 µl. The mix was then incubated for 5 min at 42° C and then placed immediately on ice for at least 2 mins. Afterwards the reverse transcription mix was prepared. For 1 reaction, this mix included 4 µl of 5X Buffer BC3, 1 µl Control P2, 2 µl of RE3 Reverse Transcriptase Mix and 3 µl RNase-free water to make a total volume of 10 µl. The reverse transcription mix and genomic DNA mix were combined and incubated at 42°C

for 15 mins and then 95°C for 5 mins. 91 µl of RNase-free water was then added and the mix stored at -20°C until when needed.

2.2.6 Focused gene expression profiling using the Human Cancer PathwayFinder RT^2 Profiler PCR Arrays

Brief description of kit

This is a reliable and sensitive technology for gene expression profiling. It can be used in analysing concentrated panels of genes involved in pathways, biological processes, and signal transduction with the use of real-time PCR. In this study, the Human Cancer PathwayFinder RT^2 Profiler PCR Array was utilised. This array contains 84 genes involved in various pathways- Angiogenesis, Apoptosis, Cell cycle, Cellular senescence, DNA Damage& Repair, Epithelial-to-Mesenchymal Transition (EMT), Hypoxia signalling, Metabolism and Telomeres & Telomerases- that have been identified or suspected to be involved in carcinogenesis. The 96-well array also contains 5 housekeeping genes, 1 human genomic contamination control, 3 reverse transcription controls and 3 positive PCR controls. Each 96-well plate contains SYBR® green optimized primer assays for a thoroughly researched pathway focused panel of genes. This allows for simplicity, high sensitivity and specificity in gene expression study and analysis.

Procedure

For each array, the PCR components mix was first prepared by adding 1350 µl of 2X RT^2 SYBR Green Mastermix, 102 µl of cDNA synthesis reaction and 1248 µl of RNase-free water to a sterile 5 ml tube. Using the Integra 8-well multi-pipette, 25 µl of the mix was dispensed in each of the 96-well in the array. The array was then tightly sealed using an optical adhesive film and centrifuged for 1 min at 1000 g at room temperature. The array was then placed in the real-time cycler machine. The cycling condition was set as follows: 95°C for 10 min (1 cycle), 95°C for 15 secs (45 cycles) and 60°C for 1 min (45 cycles). At the end of the cycles the Cp values were obtained to be used for analysis.

Data analysis

The Qiagen RT^2 Profiler PCR array analysis software was used to generate differentially expressed genes. Data obtained from the Roche LC-480 was uploaded onto Microsoft excel. This data was then carefully arranged into adequate columns and rows as instructed on the website and further uploaded for analysis. Also, to identify novel potential biomarkers, manual filtering was performed using genes upregulated in both tumour tissues and MIA PaCa-2 cell

line. Manual filtering involved going through the pancreatic cancer databases. Furthermore, the Reactome software was used (Croft et al., 2011) to indicate pathways in which the identified genes of interest were enriched. Reactome is a peer-reviewed database that allows for data sets to be mapped, assigned pathways and overrepresentation analysis performed.

2.2.7 Gene expression profiling using RNA sequencing

Brief description of RNA sequencing

RNA sequencing uses deep sequencing technology for profiling the transcriptome. RNA sequencing is known to provide much more accuracy in the detection and measurement of transcripts compared to other methods (Wang et al., 2009). Interestingly, RNA sequencing can also reveal sequence variation which makes it an exciting tool for gene expression studies.

Procedure

RNA sequencing was done by Inqaba-biotec, South Africa. Total RNA was extracted from tissue samples and DNase digestion performed as previously described. Total RNA for each sample was sent to Inqaba-biotec for further processing, sequencing, and differential gene analysis.

Library preparation included the depletion of ribosomal RNA (rRNA) from total RNA samples which was done using rRNA reduction library kit (IB Miseq RNA-0). The rRNA depleted RNA was then converted to complimentary DNA and 150 bp sequenced. Sequencing was done using the Illumina miSeq (which can perform up to 2 X 300 paired-end reads in a single run). Read numbers were obtained for each sample (Table 2.1). The CLC genomics workbench software was used for data normalization and differential gene analysis.

Table 2.1: Number of reads obtained per sample

Sample	Number of reads
Tumour patient 1	263325
Tumour patient 2	154409
Normal patient 1	1504525
Normal patient 2	964390

Data analysis

Differential gene analysis

The CLC-BIO genomics workbench (www.qiagenbioinformatics.com/products/clc-genomics-workbench) was used for differential gene expression analysis following the procedure shown

available at http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=RNA_Seq_analysis.html. HG19 was used as the reference genome.

Functional enrichment analysis

The Morpheus software was used to generate a heat map of the observed differentially expressed genes. The PANTHER version 6 (Mi et al., 2007) was then used for functional enrichment analysis of identified upregulated genes. PANTHER enabled the identification of molecular functions, biological processes, and pathways in which upregulated genes were enriched. For this enrichment analysis, our cut off was a p-value of 0.05. Furthermore, FunRich (Pathan et al., 2015) was used to observe gene interaction network, protein domains, transcription factors, clinical phenotypes in which upregulated genes were enriched.

The DAVID tool (Huang et al., 2007) for functional enrichment analysis of downregulated genes and the GeneMania tool used to identify gene interaction networks.

2.2.8 Biomarker filtering analysis

The Biomarker filtering analysis too available in the Ingenuity pathway analysis (IPA) software was used to identify novel potential biomarkers. Excel files containing the differentially expressed genes (DEGs) obtained from RNA sequencing of tissue samples were uploaded onto the IPA platform. The Biomarker filter analysis tool on the IPA platform was then utilized. The Biomarker filtering tool by IPA enables for the identification of relevant and novel biomarkers. Several filters available on the platform were used, including:

(i) Species- To indicate species of the organism of interest
- ✓ Human

(ii) Tissues and cell lines- To indicate where the biomarker expressed
- ✓ Pancreatic cancer cell lines

(iii) Node types
- ✓ All nodes except complexes and groups

(iv) Diseases- To indicate type of disease
- ✓ Cancer

(v) Biofluids- To indicate where biofluids biomarker can be found
- ✓ Blood

(vi) Molecules

- ✓ All types

(vii) Biomarker- To indicate novelty of biomarker

- ✓ Not a known biomarker

To further refine biomarker analysis, manual filtering was conducted. Manual filtering involved searching through multiple databases including NCBI and the pancreatic cancer databases bearing in mind the hallmarks of a good biomarker (see Section 1.2).

2.2.9 Biomarker (gene) Validation by Real time PCR using RT^2 qPCR primer assays

Procedure

For this study, the target genes were *JUP*, *COPG* and *TCIRG1* because they were identified as novel potential biomarkers after several Biomarker filtering analyses identified them as novel potential biomarkers. *MRPL19*, a housekeeping gene was used as a control. *MRPL19* was chosen since its expression remains unchanged during pancreatic cancer (Mohelnikova-Duchonova et al., 2012). All MIQE guidelines were observed including performing 3 independent experiments (Bustin et al., 2009).

In a 1.5 ml tube the PCR components mix was prepared. The mixture included 12.5 µl RT^2 SYBR Green Mastermix, 1 µl cDNA synthesis reaction, 1µl RT^2 qPCR Primer assay and 10.5 µl RNase-free water for one reaction. The mixture was then thoroughly mixed and transferred into a 96-well plate with a pipette. The 96-well plate was covered with an adhesive and centrifuged for 1 min at 1000g to remove bubbles. The plate was then placed into the Roche LightCycler-480. The cycling condition was set as follows: 95°C for 10 min (1 cycle), 95°C for 15 secs (45 cycles) and 60°C for 1 min (45 cycles). At the end of the cycles the Cp values were obtained to be used for analysis.

Data analysis

Data obtained from the Roche LC-480 were copied into Microsoft excel which was then uploaded into the REST software (Pfaffl et al., 2002). The REST software calculates the confidence interval, standard error, and the probability of alternate hypothesis that difference between sample and control groups is due only to chance. From these calculations, it produces the result which includes a box-plot and a statement indicating whether genes were up, unchanged, or down regulated.

2.2.10 Protein-protein interaction network analysis

The STRING v9.1 (Franceschini et al., 2013) was used for the protein-protein interaction network analysis of the identified novel potential biomarker. STRING enabled us to identify known and predicted biological processes and pathways in which components of the network are enriched

2.2.11 Variant filtering analysis

The Ingenuity Variant Analysis tool (www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/) was used to identify novel SNP variants. The CLC Genomics work bench (which includes the Basis variant detection tool, fixed ploidy variant detection tool, and the low frequency variant detection tool) was used for variant calling for all samples (2 normal and 2 tumour tissues). Variant calling files were then uploaded into the Ingenuity Variant Analysis (IVA) platform and analysed.

Various filters were used to identify gene variants. They include;

(i) *Confidence*- This was used to remove suspicious or low quality variants. These types of variants are likely to be false positives.
Only variants that satisfied the following criteria was kept
- ✓ Call quality is at least 20 in both case and control
- ✓ Variant passed upstream pipeline filtering
- ✓ Outside top 5% most exonically variable 100 base windows in healthy public genomes.

(ii) *Common variants*- This was used to exclude common variants.
Exclusion criteria included variants that are observed in any of the population with the allele frequency of;
- ✓ At least 0.5% in 1000 Genomes project (The 1000 Genomes Project Consortium, 2015).
- ✓ At least 1% in the public complete genomics genome
- ✓ At least 0.5% of all NHLBI ESP exomes.
- ✓ At least 0.5% in the Allele frequency community.

(iii) *Predicted deleterious*- This filter refers to variants that confer deleterious effects to gene function or expression. For this filter, only variants that are experimentally observed to be associated with a phenotype were kept:

- ✓ Disease-associated per the complete ACMG (American college of genetics and genomics) guidelines classification.
    - ➢ Pathogenic
    - ➢ Likely pathogenic
- ✓ Listed in HGMD® Human gene mutation database)
- ✓ Are associated with loss of function of a gene
    - ➢ Frameshift, inframe indel, start/stop codon change.
    - ➢ Missense unless predicted tolerated by SIFT or polypen-2
    - ➢ Splice site loss up to 2 bases into intron or as predicted by MaxEntScan.

(iv) *Genetic analysis*- This filter was used to keep or exclude variants based on some gene analysis criteria.

- ✓ Pair matches samples from the same individuals.
- ✓ And in both case and control samples, variations that were; Homozygous, Heterozygous, Het-ambiguous, Haploinsufficient, Hemizygous and Nullizygous

(v) *Cancer driver variants*- This filter specifically targets variants and genes associated with cancer.

Only variants found in the following were kept

- ✓ cancer-associated mouse knockout phenotypes.
- ✓ cancer-associated cellular processes with any directionality
- ✓ cancer therapeutic targets
- ✓ published cancer literature (Variant and gene level findings)
- ✓ known or predicted cancer sub network regulatory sites
- ✓ COSMIC at the frequency greater than or equal to 0.01%
- ✓ TCGA at a frequency greater than or equal to 0.01%

(vi) Biological context- This enabled variants that are associated with a specific condition (cellular proliferation) to be filtered.

After these filters were applied, analysis was then performed.

Furthermore, manual filtering was performed to identify novel genetic variants. This involved a search though several databases as indicated in 2.2.8

2.2.12 Focused proteomic profiling using Human oncology proteome profiler array

Brief Description of kit

The R&D systems Proteome Profiler™ Human Oncology Array kit was used to ascertain differential protein expression between normal and tumour samples. It is a sensitive, rapid and cost-effective tool. It contains 84 cancer-related proteins on the array membrane. An antibody for each protein is present as a spot on the membrane. Each protein spot is duplicated to serve as technical replicate.

Procedure

Protein expression profiling was carried out using the Proteome Profiler Human XL Oncology Array kit. Samples and reagents were prepared per the instructions described by the manufacturer. 2 ml of the Array buffer 6 (Block buffer) was pipetted into each well of the 4-well multi-dish. Each membrane was then placed into separate wells in such a way that the numbers on the membranes faced upwards. Membranes were incubated on a rocking platform shaker for 1 hr. During this time, 500 µl of the samples were added to 500 µl of Array Buffer 4 and the final volume adjusted to 1.5 ml with Array Buffer 6. The Array Buffer 6 placed in the wells of the multi-dish was removed and the sample mix added and the lid placed on the dish. This was then incubated overnight at 4°C on a rocking platform shaker. Each membrane was carefully removed from the well using appropriate forceps and placed into separate plastic containers with 20 ml of 1X wash buffer for 10 mins on a rocking platform shaker. This wash step was repeated 3 times. For each array, 30 µl of detection antibody cocktail was added to 1.5 ml of 1X Array buffer 4/6 (a mixture of 4 ml Array buffer 4 and 8 ml Array buffer 6). 1.5 ml of the diluted detection antibody mixture was then added to each well. Each membrane was carefully removed from its wash plastic container and placed in the multi-dish containing the diluted detection antibody cocktail and covered with the lid. This was then incubated on a rocking platform shaker for 1 hr. After incubation, the membrane underwent the washing steps described previously. 2 ml of the diluted streptavidin-HRP was added to each well. The membrane was then placed in the diluted streptavidin-HRP and incubated for 30 mins on a rocking platform shaker. After incubation, the membrane underwent the washing steps described previously. The membrane was then placed on a plastic sheet protector. 1 ml of the Chemi reagent mix was pipetted onto the membrane, another plastic protector placed on top

and the membrane incubated for 1 min. The membrane was using the Bio-Rad Chemi-doc system.

Data analysis

Differential protein analysis

The wvision (western vision) QuickSpots software was used to analyse the captured array membranes and identify differentially expressed proteins. This software measures the intensity of each spot. The average spot intensities of tumour and normal tissue samples was ascertained. These intensity values were then exported to Microsoft excel. Standard error and p-values were then calculated using Microsoft excel.

Functional enrichment analysis

After identification of statistically significant ($p< 0.05$) differentially expressed proteins, the REACTOME (Croft et al., 2011) software was used to classify pathways in which they were enriched. Only pathways and biological processes with $P < 0.05$ were considered.

2.2.13 Biomarker filtering analysis

Biomarker filtering analysis was performed as described in 2.2.8

2.2.14 Protein-protein interaction network analysis

After identification of potential protein biomarkers protein-protein interaction network analysis was done as described in 2.2.10.

CHAPTER 3- RESULTS

3.1 Summary of results

Using the Human Cancer PathwayFinder RT^2 Profiler PCR arrays differential gene expression (DEGs) profiles of pancreatic tissue samples-obtained from PDAC patients- and the pancreatic cancer cell line, MIA PaCa-2, was shown. Pathway analysis was then performed to compare DEGs identified in tissue samples to those observed in the pancreatic cancer cell line. Upregulated genes in tumour tissues were observed to be uniquely enriched in pathways associated with necrosis. After biomarker analysis, RT^2 Profiler PCR arrays did not identify any novel potential biomarker thus RNA sequencing was performed.

Differential gene analysis of RNA sequencing data revealed 4911 upregulated and 5 downregulated genes. I performed gene ontology analysis and observed that the upregulated genes were enriched in biological processes associated with nuclear transport. Also, these genes enriched signalling pathways included the insulin/IGF/MAPK cascade, integrin signalling pathway and *Ras* signalling pathway. Gene network analysis showed an intricate interaction network of most upregulated genes. Furthermore, downregulated genes were observed to be enriched in several processes including metabolism, embryonic patterning, glycosylation, and protein transport.

Using the IPA biomarker tool, 3 novel potential biomarkers- *JUP*, *COPG* and *TCIRG1* were identified. To further validate these biomarkers, real time PCR was done and the upregulation of *JUP* across additional tissue samples confirmed. Protein-protein interaction (PPI) network analysis showed that *JUP* and functional interactors were enriched in pathways associated with cancer.

In this study, 87 genetic variants associated with PDAC which were then filtered to identify those predicted to result in loss of gene function and are only present in tumour samples. This filtering process identified 6 variants present in 4 genes- *SPARC, ZFHX3, COLIA1,* and *TGM2.*

Furthermore, 18 significantly ($p<0.05$) upregulated proteins in tumour tissue samples were identified. These proteins are observed to be enriched in pathways associated with signal transduction and the immune system. ANGPTL4 and p27kip1- were further identified as novel potential biomarkers. Also, ANGPTL4 PPI network analysis identified lipid metabolism and PPAR signalling enriched biological process and pathway respectively. Similarly, p27/kip1 PPI network analysis identified cell cycle as the most enriched biological process.

3.2 Gene expression profile in tumours and MIA PaCa-2

MIA PaCa-2 is a widely used pancreatic cancer cell line obtained from a Caucasian male (as described in 2.2.3) and tissue samples used in this study were obtained from black South African patients. As gene expression profiles can differ between biological samples and population, I have compared the differential expressed genes in pancreatic tumours to those in the cell line. Two pancreatic tumour tissues were obtained from different PDAC patients (patient 1 and 2) and normal tissue obtained from patient 1. RNA obtained from biological samples were converted to cDNA and pipetted into the human cancer PathwayFinder RT^2 Profiler array. The array was then placed in the Roche lightcycler-480 and analysed. Cp values were generated and further analysed. Cp values of both tumour samples were averaged and then compared to those of the normal tissue (this was referred to as Group 1). Similarly, Cp values from MIA PaCa-2 were generated and compared to those from the normal tissue (referred to as Group 2).

All samples and arrays used passed data quality control tests (Figure A1 in appendix). Amongst the reference genes in the array, B2M was observed to be the most stable across all samples and was therefore used for normalization (Figure A2 in appendix).

Visualization of gene expression profiles

The QIAGEN RT^2 Profiler PCR analysis software was used to generate heat maps to help visualize differential gene expression profile for both Group 1 and Group 2 (Figures 3.1). Visual inspection of heatmaps for Groups 1 and 2 showed differences in gene expression between both groups. Differential gene expression analysis for Group 1 showed the upregulation of 20 genes and downregulation of 20 genes. For Group 2, 32 genes were upregulated while 41 genes down regulated.

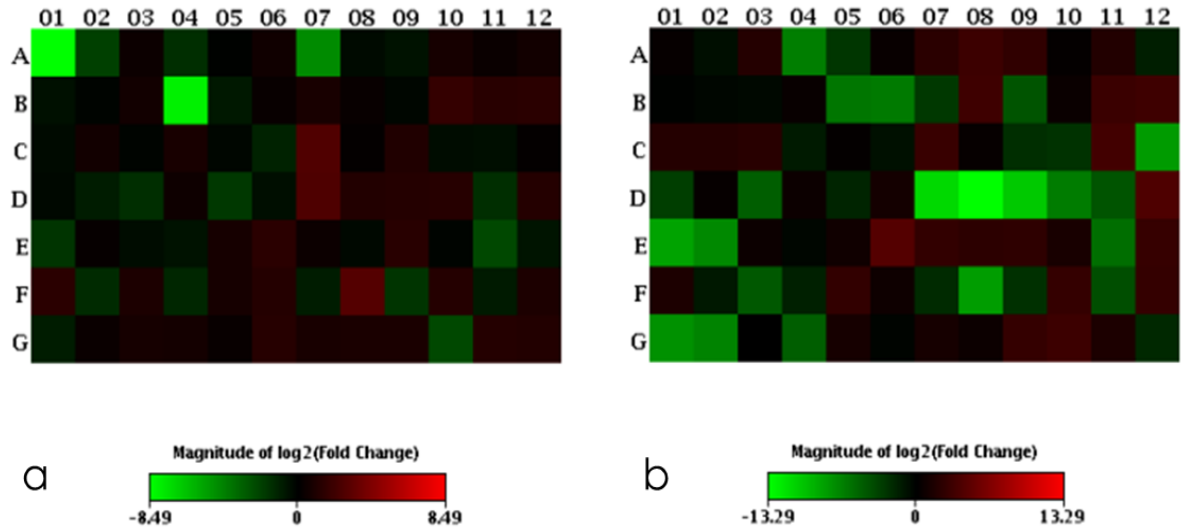


Figure 3.1: Heat map of fold changes observed in Group 1 (a) and Group 2 (b) for the various genes within the array. Colour bar indicating expression levels is also shown. Fold regulation of genes can be observed as colours indicated in the colour bar. Negative values indicate downregulation and positive values, upregulation. Individual fold change values are shown in figure A3 in the appendix section.

Pathway analysis of Group 1 and Group 2 DEGs

Heatmap representations showed variances in differential gene expression between Group 1 and 2, I then sought to identify dysregulated genes unique in both groups. Unique in the sense that they were dysregulated in Group 1 but not in Group 2 and vice versa. In Group 1, 10 genes were uniquely upregulated and 6 downregulated. In Group 2, 23 were uniquely upregulated and 25 downregulated (Table A6 in appendix section).

To identify pathways in which uniquely DEGs for Group 1 and 2 were enriched in, pathway analysis using REACTOME software was utilized. Downregulated genes unique to each group were input separately into the system and an output of a diagram and table showing enriched pathways were generated (Figure 3.2, Table 3.1). This process was replicated for upregulated genes (Figure 3.3, Table 3.2)

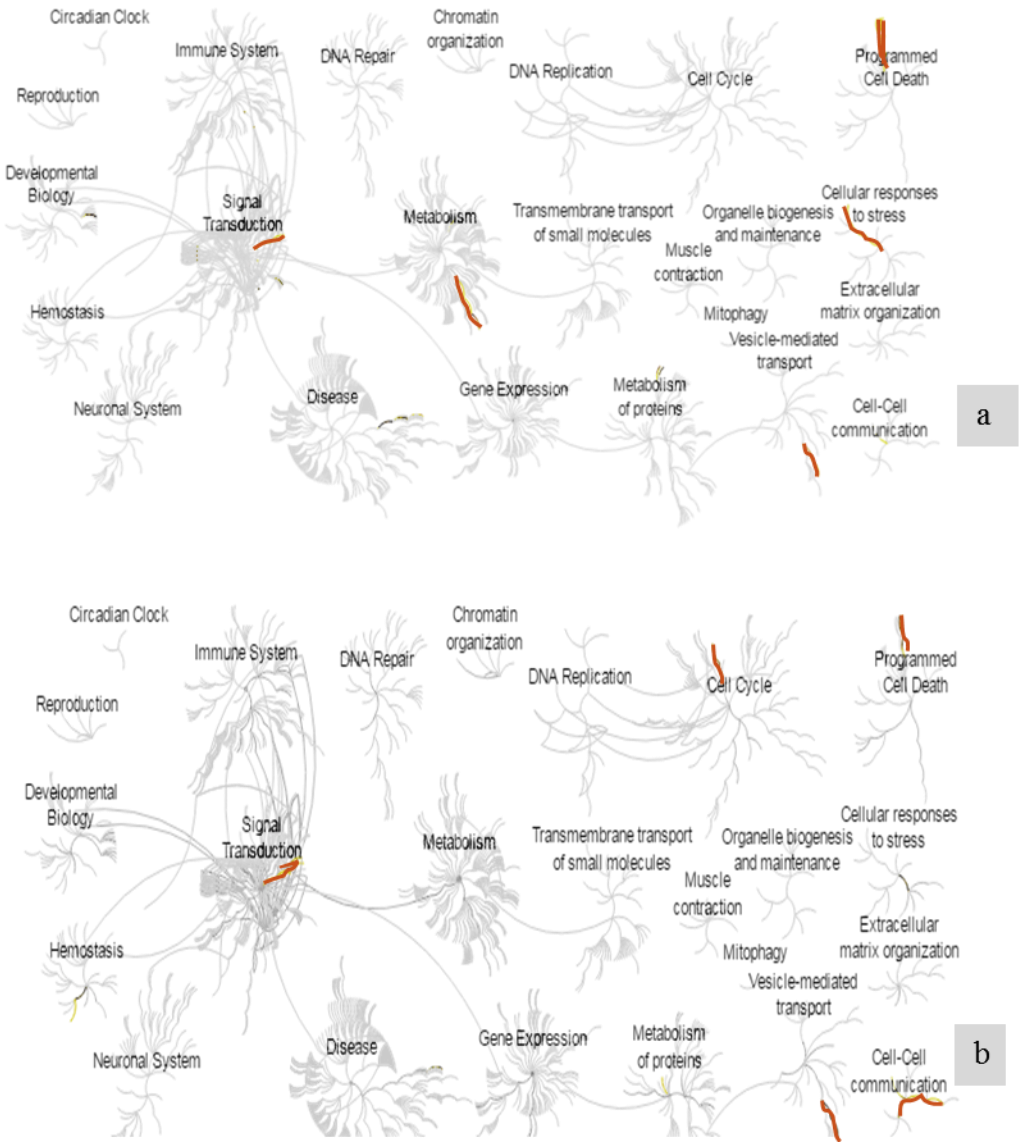


Figure 3.2: Diagrammatic representation of enriched pathways of downregulated genes unique to Group 1 (a) and Group 2 (b). Orange coloured lines indicate enriched pathways. Group 1 is normal tissue vs tumour tissue and Group 2 is normal tissue vs MIA PaCa-2 cell line. Generated from REACTOME

Table 3.1: Top 10 pathways enriched in downregulated genes unique to Group 1 and Group 2

Group 1 (tumour vs normal tissue)		**Group 2 (MIA PaCa-2 vs normal tissue)**	
Pathway	**p-value**	**Pathway**	**p-value**
Formation of cornified envelope	3.733E-5	VEGF ligand-receptor interactions	3.01E-6
VEGF ligand-receptor interactions	6.71E-5	VEGF binds to VEGFR leading to receptor dimerization	3.01E-6
VEGF binds to VEGFR leading to receptor dimerization	6.71E-5	Tie2 signalling	6.05E-5
Keratinization	2.71E-4	Neurophilin interactors with VEGF & VEGFR	8.84E-5
Developmental biology	5.33E-3	Oncogene induced senescence	3.75E-4
Formation of apoptosome	7.44E-3	Cellular senescence	5.12E-4
SMAC-mediated apoptotic response	7.44E-3	Signalling by VEGF	1.69E-3
SMAC-mediated dissociation of IAP: caspase complexes	7.44E-3	TBS3 regulates transcription of death receptors and ligands	1.74E-3
SMAC binds to IAPs	7.44E-3	Regulation of insulin-like growth factor transport uptake by insulin-like growth factor binding proteins	2.57E-3
Cell junction organization	8.23E-3	Developmental biology	4.38E-3

Pathways enriched by downregulated genes in both groups can be observed in figure 3.2. Amongst the top 10 most enriched pathways, VEGF binding and developmental biology were common to both groups. The formation of cornified envelope was observed to be the pathway in which Group 1 downregulated genes were most significantly enriched (p = 3.733 E-5). VEGF ligand-receptor interactions followed by VEGF binds to VEGFR leading to receptor dimerization (p values of 6.71E-5) (Table 3.1). Group 2 downregulated genes were mostly enriched in VEGF ligand-receptor interactions followed by VEGF binds to VEGFR leading to receptor dimerization both with a p values of 3.01E-6. Group 1 was observed to have some unique pathways including; formation of cornified envelope, keratinization, formation of apoptosome, SMAC-mediated apoptotic response and cell junction organization. In comparison, Group 2 also had some unique significantly downregulated pathways including; cellular senescence, oncogene induced senescence and regulation of insulin-like growth factor transport uptake by insulin-like growth factor binding proteins.

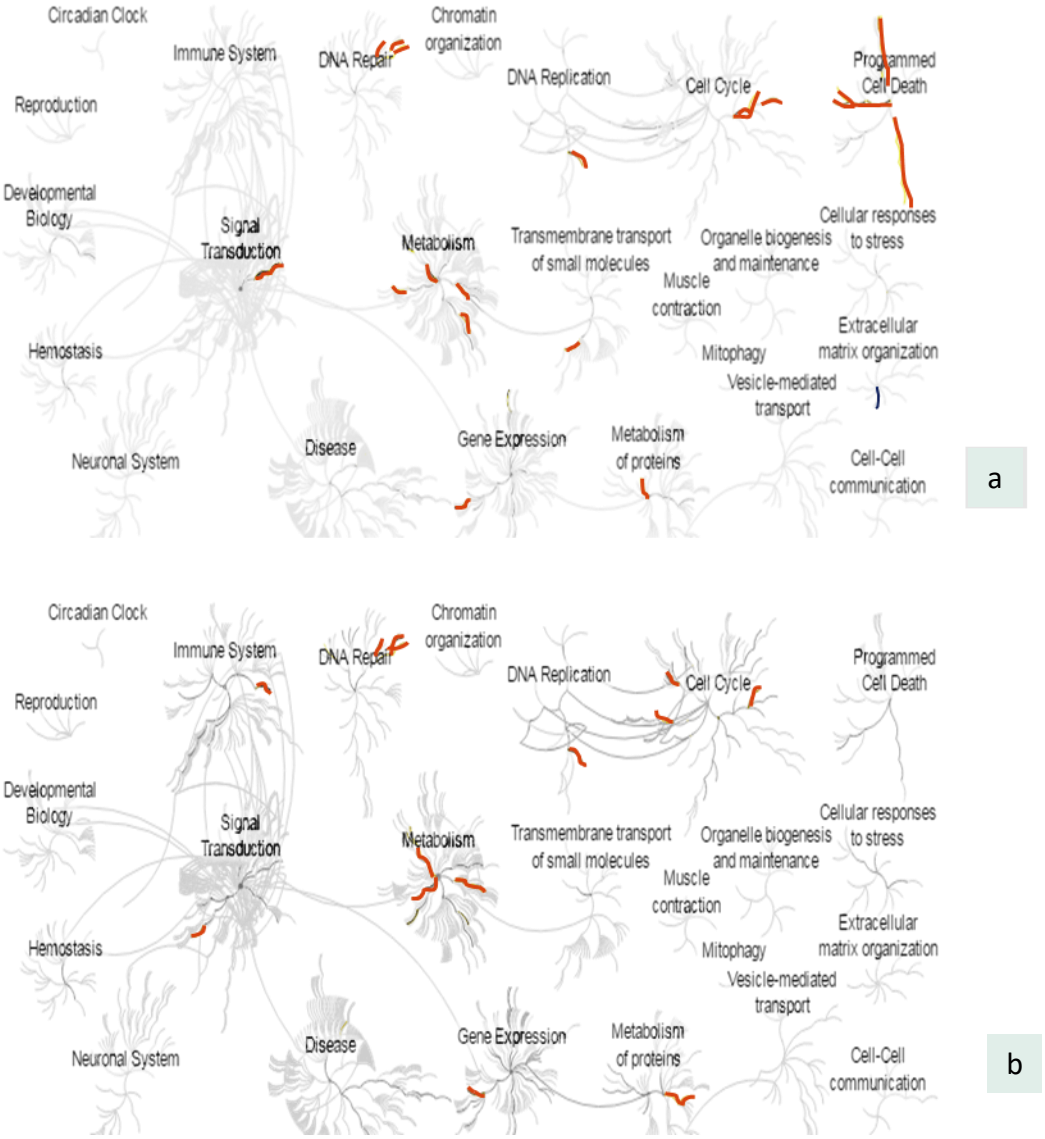


Figure 3.3: Diagrammatic representation of enriched pathways of upregulated genes unique to Group 1 (a) and Group 2 (b). Orange coloured lines indicate enriched pathways. Group 1 is normal tissue vs tumour tissue and Group 2 is normal tissue vs MIA PaCa-2 cell line. Generated by REACTOME.

Table 3.2: Top 10 pathways enriched in upregulated genes unique to Group 1 and Group 2.

Group 1 (Tumour vs normal tissue)		**Group 2 (MIA PaCa-2 vs normal tissue)**	
Pathway	**p-value**	**Pathway**	**p-value**
Unwinding of DNA	3.69E-6	Ub-specific processing proteases	6.59E-6
Regulation of necroptotic cell death	1.69E-5	Unwinding of DNA	1.08E-5
RIPK1-mediated regulated necrosis	4.1E-5	XAV939 inhibits tankyrase stabilizing AXIN	5.22E-5
Regulated necrosis	4.1E-5	Deubiquitination	5.34E-5
Activation of pre-replicative complex	8.82E-5	G1/S transition	4.88E-2
DNA strand elongation	1.12E-4	TCA & respiratory electron transport	6.32E-3
Activation of ATR in response to replication stress	1.21E-4	S phase	7.13E-3
Oncogene induced senescence	1.4E-4	Mitotic G1-G1/S	7.13E-3
Pyruvate metabolism	1.97E-4	Activation of the pre-replicative complex	9.11E-3
Cellular response to stress	2.15E-4	Cell cycle	9.33E-3

Upregulated genes for both groups were enriched in pathways associated with metabolism and signal transduction (Figure 3.3). The top 10 enriched pathways for both groups were observed to have one pathway in common; Unwinding of DNA. However, Group 1 genes are also observed to be enriched in pathways involved in programmed cell death. This enrichment was identified to be in pathways involved in necrosis (Table 3.2). Unwinding of DNA is the most enriched pathway in Group 1 genes (p value=3.69E-6), while Ub-specific processing proteases is for Group 2. Pathways enriched by Group 2 genes were mostly associated with cell cycle progression. These pathways include; mitotic G1-G1/S, S phase, G1/S transition, activation of pre-replicative complex and cell cycle. On the other hand, Group 1 genes were mostly associated with necrosis and response to stress pathways. These pathways include; regulated necrosis, regulation of necroptotic cell death, activation of ATR in response to replication stress, and cellular response to stress.

Biomarker filtering analysis

To identify potential novel biomarkers, biomarker analysis was conducted. This analysis included using the Ingenuity pathway analysis tool and manual analysis. Differentially

expressed genes identified from both Groups 1 and 2 were used in this analysis. As no potential novel biomarkers were identified from the DEGs; thus, RNA sequencing was conducted.

3.3 Differential gene expression using RNA sequencing

RNA sequencing was conducted to enable gene expression analysis of a much larger pool of genes. Four samples from 2 patients were used and were those previously used for the RT^2 profiler PCR array. From each patient one normal and one tumour sample was utilized. Both tumour samples were obtained from stage 1 PDAC patients. Patient 1 was a 62-year-old female and patient 2 was a 68-year-old male. After RNA extraction samples were sent to Inqaba-biotec for sequencing and differential gene expression analysis. Each patients tumour samples were compared to their corresponding normal tissues. That is, for patient 1, tumour (T1) was compared to normal tissue (N1) and for patient 2, tumour (T2) was compared to normal tissues (N2).

Differential gene expression analysis using the CLC genomics workbench identified 8032 upregulated in the tumour from patient 1 while 7456 were overexpressed in that of patient 2. Amongst the upregulated genes, 4911 were observed to be upregulated in both patients. Furthermore, 54 and 558 genes were identified to be downregulated in patient 1 and 2 respectively. Amongst the downregulated genes, only 5 were common to both (Figure 3.4).

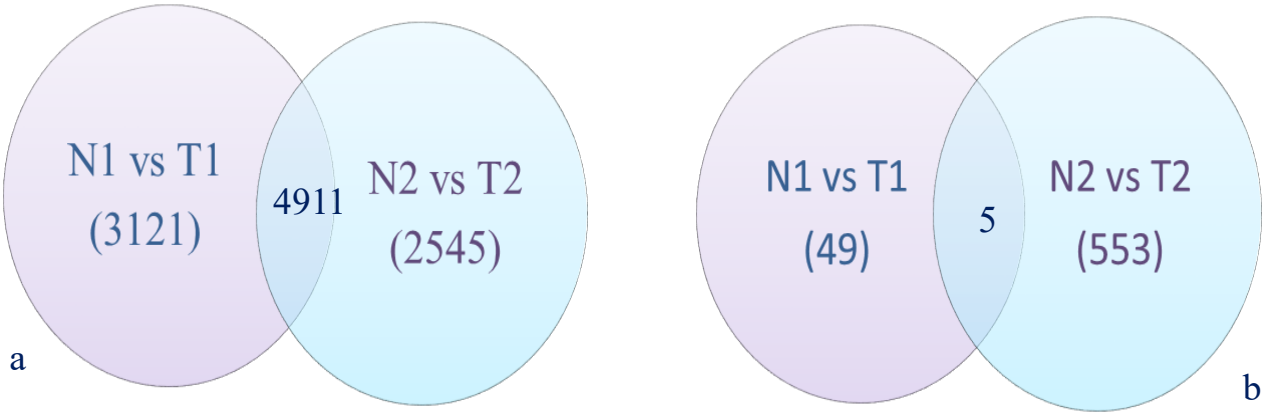


Figure 3.4: Venn diagrams showing differentially expressed genes. (a) Genes upregulated in sample sets. 4911 genes are common to both sets. (b) Genes downregulated in sample sets. 5 genes common to both sets.

To visualize differential gene expression in both patients the MORPHEUS software was used. Differentially expressed genes for both patients were uploaded into the software and heatmap generated (Figure 3.5). Expected variations between gene expression patterns of both patients were observed from the heatmap (see Section 1.2.1).

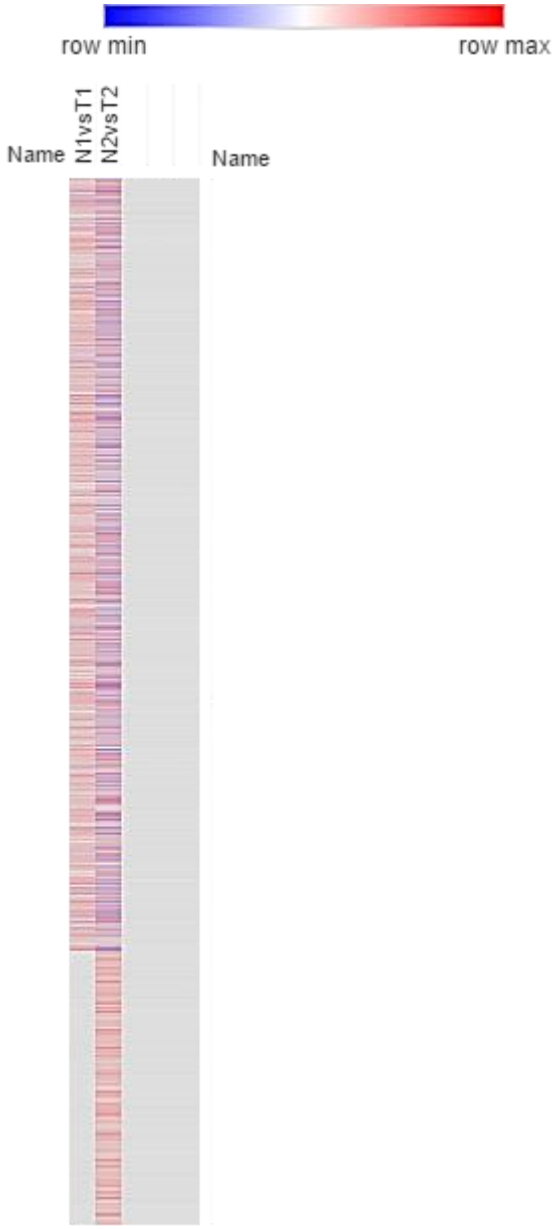


Figure 3.5: Heatmap showing differentially expressed genes for both sample sets. N1vsT1 indicates normal vs tumour from patient 1 while N2vsT2 indicates normal vs tumour from patient 2. Blue colour represents down regulation and Red indicates upregulation. Heatmap was generated from MORPHEUS software.

Functional enrichment analysis of upregulated genes

To group upregulated genes per roles they play, functional enrichment analysis was performed. The 4911 upregulated genes common to both sample sets were uploaded into the PANTHER software and analysed. The PANTHER software was used to analyse the pathways, biological processes, molecular functions, and protein classes enriched by the upregulated genes.

The Insulin/IGF pathway-mitogen activated protein kinase/MAP kinase cascade is observed to be the most enriched pathway. This pathway had a fold enrichment of 2.62 and p-value of 2.19E-02. It is followed by the integrin signalling pathway and Ras signalling pathway (p-values 9.65E-15 and 1.10E-03 respectively). Furthermore, the PDGF and EGFR signalling pathway were also observed to be enriched. Examination of the top 10 list indicate that our upregulated genes were mostly enriched in pathways involved in several signalling pathways (Table 3.3).

Table 3.3: Showing top 10 pathways enriched by genes upregulated in tumour tissues.

Pathways	**Homo sapiens (Reference)**	**Client input**	**Expected**	**Fold enrichment**	**p-value**
Insulin/IGF pathway-mitogen activated protein kinase/MAP kinase cascade	33	20	7.64	2.62	2.19E-02
Integrin signalling pathway	192	110	44.45	2.47	9.65E-15
Ras pathway	76	39	17.59	2.22	1.10E-03
PDGF signalling pathway	149	73	34.49	2.12	1.06E-06
EGF receptor signalling pathway	139	67	32.18	2.08	7.72E-06
Angiogenesis	176	83	40.74	2.04	5.58E-07
CCKR signalling map	173	81	40.05	2,02	1.16E-06
VEGF signalling pathway	72	33	16.67	1.98	4.11E-02
B cell activation	72	33	16.67	1.98	4.11E-02
Gonadotropin-replacing hormone receptor pathway	235	106	54.40	1.95	4.70E-08

The most enriched biological process observed was nuclear transport (p =3.8E-02) (Table 3.4). An over representation of upregulated genes in nuclear transport was observed with a fold enrichment of 1.82 is shown. This is followed by nuclear localization (p = 1.17E-06).

.

Table 3.4: Showing top biological processes enriched by genes upregulated in tumour tissues.

GO ID	Biological process	Homo sapiens (Reference)	Client input	Expected	Fold enrichment	p-value
GO:0051169	Nuclear transport	107	45	24.77	1.82	3.80E-02
GO:0051179	Nuclear localization	2144	622	496.33	1.25	1.17E-06
GO:0000398	mRNA splicing, via spliceosome	174	70	40.28	1.74	3.03E-03
GO:0006397	mRNA processing	242	91	56.02	1.62	2.31E-03
GO:0006139	Nucleobase-containing compound metabolic process	3160	835	731.54	1.14	5.91E-03
GO:0044238	Primary metabolic process	5773	1536	1336.44	1.15	3.42E-08
GO:0008152	Metabolic process	6817	1837	1578.13	1.16	6.48E-13
GO:0000375	RNA splicing via transcription reaction	134	53	31.02	1.71	4.71E-02
GO:0006325	Chromatin organization	263	93	60.88	1.53	1.69E-02

The most enriched molecular functions of the upregulated genes were investigated and it was determined that most genes were involved in Rac GTPase binding. Rac GTPase had a fold enrichment of 3.35. It was followed by Rho GTPase binding and Ras GTPase binding. Most of the upregulated genes were observed to function in various types of binding. Table A7 in the appendix section shows the top 10 most enriched molecular functions.

Cellular location of dysregulated genes can be crucial in elucidating the biology of PDAC. The cellular components enriched by the upregulated genes were investigated. It was observed that the ribososme is the most enriched cellular component with a p-value of 3.01E-04 and fold enrichment of 1.96. Furthermore, the genes were also enriched in ribonucleoprotein and macromolecular complexes. These complexes come in 2nd and 3rd in the list with fold enrichments of 1.42 and 1.23 respectively. A closer look at Table A8 (appendix section) shows that many the upregulated genes were located intracellularly.

Grouping of genes-of-interest per protein class help in further understanding their role in PDAC. Protein classes were observed and nucleic acid binding identified as the most enriched with a p-value of 5.83E-10 and fold enrichment of 1.29. Table A9 (Appendix) shows that 2706

of the upregulated genes are enriched in the protein class involved with nucleic acid binding. Enzyme modulator and transferase follow (p-values 8.68E-12 and 1.35E-03 respectively).

To investigate the transcription factors and clinical phenotypes enriched by the upregulated genes the FunRich software was used. The 4911 upregulated genes were uploaded onto the software and then analysed. Figure A4 (Appendix) shows the top 6 transcription factors enriched by the upregulated genes. SP1(Specificity protein 1) shown to be the most enriched with 58.8% of the genes. SP4 followed with 37.4 % of genes and EGR1 with 28.3%. Members of the SP family- SP1 and SP4 are observed to play a role in the transcription of the upregulated genes. The KLF7, YY1 and GABPA are also significantly enriched. All 6 transcription factors are observed to have p-values less than 0.001 except for YY1.

Furthermore, more than 72% of the upregulated genes were observed to be expressed in the skin. The joint laxity and nerve conduction velocity were also enriched with 3.9% and 1.1% respectively of the upregulated genes (Figure A5).

Gene network analysis of upregulated genes

To examine the inter-connectedness of upregulated genes, gene network analysis was performed using FunRich. Majority of upregulated genes were observed to be part of an intricate network (Figure 3.6) with exception to just a minute number (right corner of Figure 3.6). A close look at figure 3.7 show that most of these genes were linked together by diverse interactions. These interactions might be linked to similar biological processes or pathways.

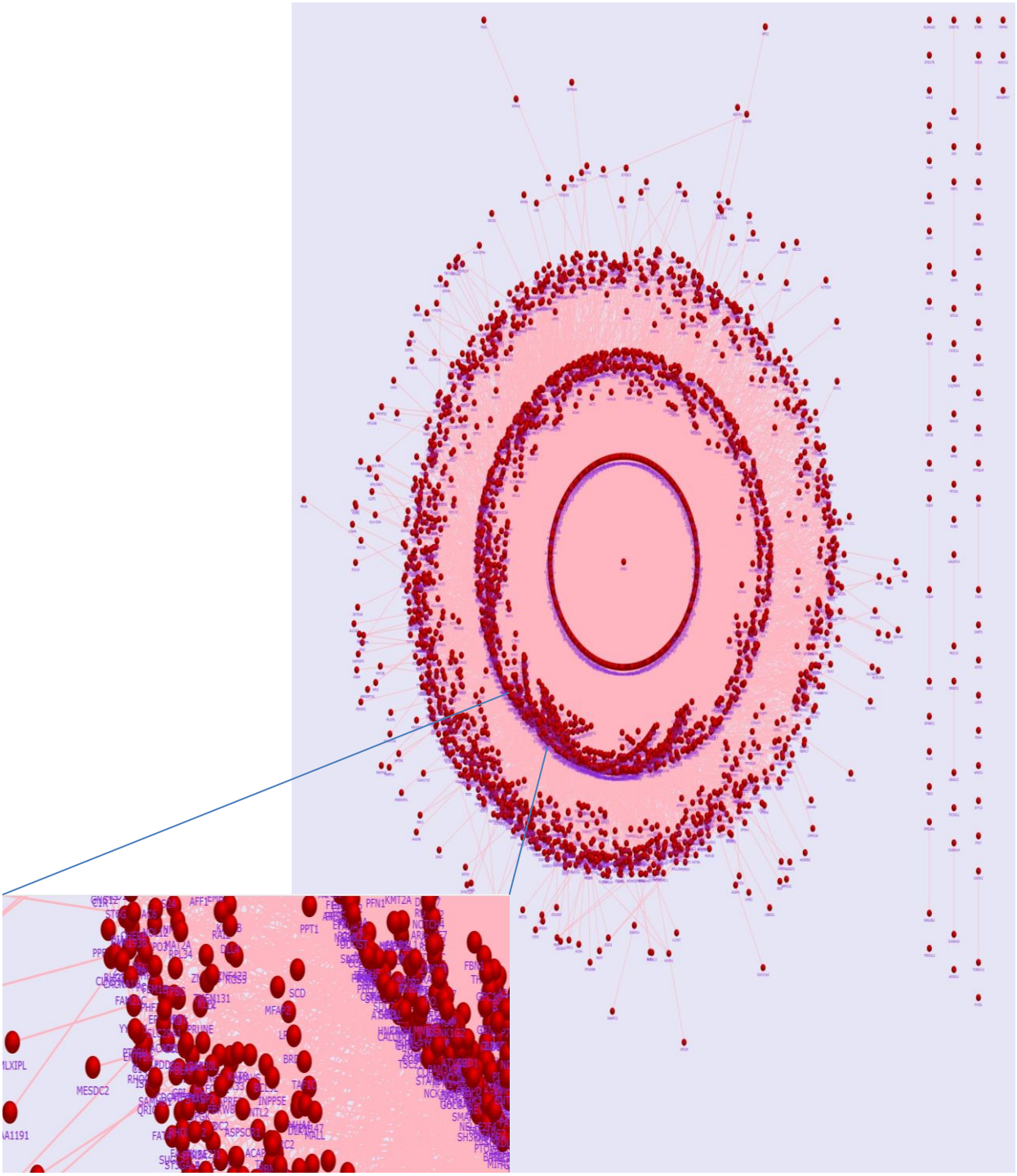


Figure 3.6: Diagram showing intricate gene network of upregulated genes. Red nodes indicate upregulated genes. Red edges (lines) indicate interactions. A magnified section of the gene network shows a clearer interaction of the genes. Generated from FunRich.

Functional enrichment of downregulated genes

To understand the roles played by the downregulated genes, functional enrichment analysis was performed. Only 5 common genes were downregulated in both patients, they included: *CXXC4*, *GALNT1*, *HS3ST4*, *SLC15A4* and *TSPAN7*. The 5 genes were uploaded onto the DAVID software and analysed. Attempt to group downregulated genes per molecular functions, biological processes, and pathways in which they are involved did not yield any result. Downregulated genes were then analysed individually. A comprehensive table highlighting diverse roles and properties of these genes can be observed in Table A10, however, they are summarized below:

CXXC4 was observed to be enriched in biological processes involved in embryonic patterning, axis formation and negative regulation of signal transduction. It is mostly enriched in cytoplasmic and membrane bound vesicles. *CXXC4* functions mostly in DNA and ion binding. Its most enriched pathway is the Wnt signalling pathway.

GALNT1 was mostly enriched in biological processes involving glycosylation and can be found in the membrane, Golgi apparatus and extracellular matrix. GALNT1 functions in acetylgalactosaminyltransferase activity, sugar, and ion binding. Its most enriched pathway is O-glycan biosynthesis.

HS3ST4 was identified to be involved in proteoglycan and glycoprotein metabolic processes. Just like *GALNT1*, it is found to be localized in the membrane, Golgi apparatus and extracellular matrix. HS3ST4 belongs to the protein class of sulfotransferases and therefore functions as such.

SLC15A4 was mostly enriched in biological processes involving protein transport and localization. It is found mostly in the membrane, vacuole, and lysosome where it functions in symporter and transporter activities.

TSPAN7 belongs to the protein class of tetraspanins and is mostly localized in the plasma membrane. It is involved in several biological processes including; protein amino acid glycosylation, glycoprotein metabolic process, and glycoprotein biosynthetic process.

Gene network analysis of downregulated genes

To identify genes interacting with the downregulating genes, gene network analysis was performed using GeneMania. It was observed that altogether the five downregulated genes had 20 interacting partners (Figure 3.7).

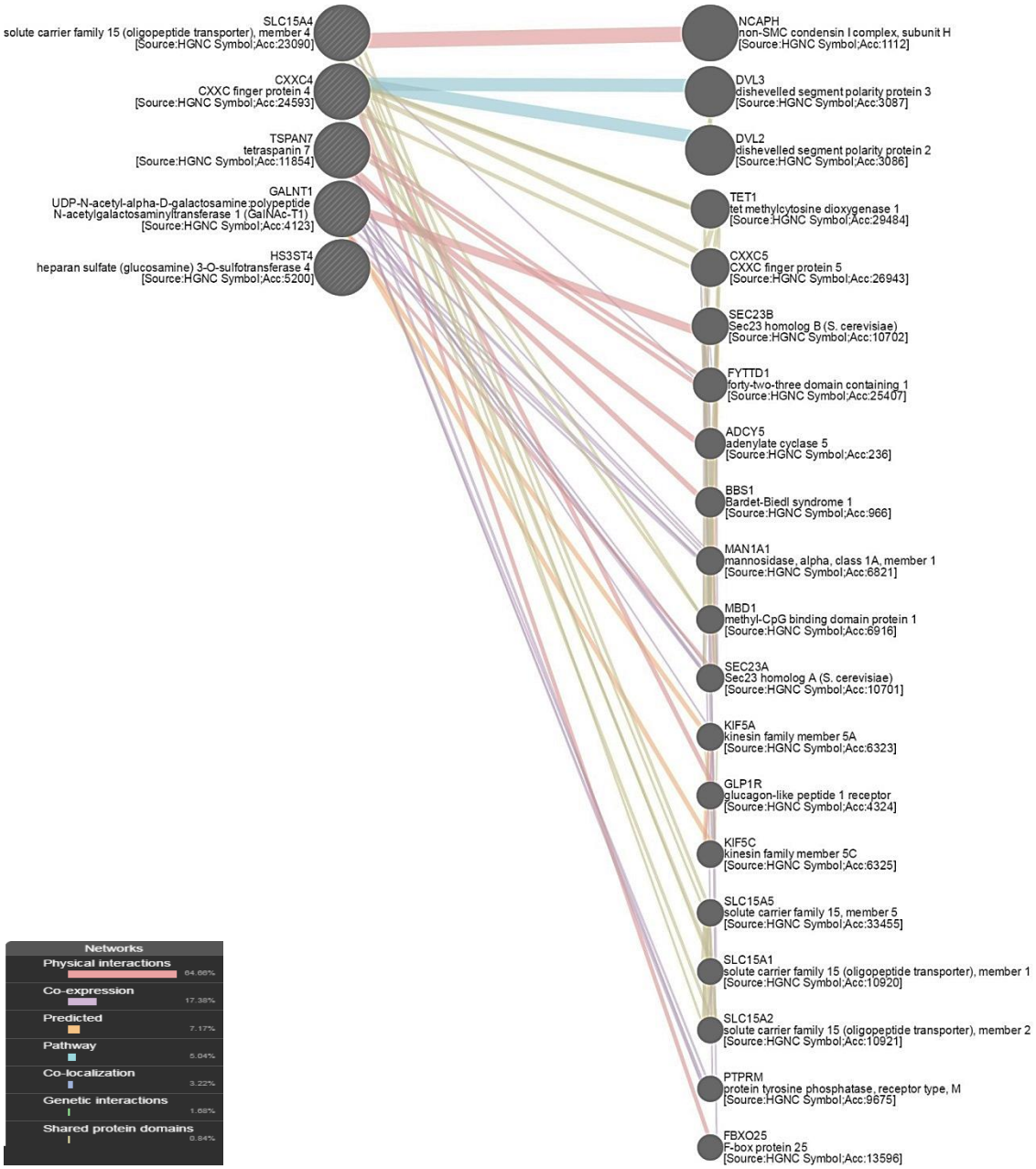


Figure 3.7: Gene network of downregulated genes including interactions. Nodes on the left indicate downregulated genes. On the right are interacting partners of downregulated genes. The bigger the sizes of nodes on the left, the greater the certainty of interaction. Key to colour edges can be seen at the bottom left of the diagram.

To further elucidate the roles which these genes play in the network, pathway analysis was performed in REACTOME using the downregulated genes and their interacting partners as input. Table 3.5 shows the top 5 most downregulated pathways. Downregulated genes were observed to be involved in the Wnt signalling and membrane trafficking pathways. NCAPH observed in figure 3.7 as the most prominent interacting partner seem to be absent in the top enriched pathways

Table 3.5: Showing pathways enriched by downregulated gene network components

Pathways	**Genes involved**	**p-value**
Proton/oligopeptide co-transportation	SLC15A1, SLC15A2, SLC15A4	1.21E-07
Wnt signalling	DVL2, DVL3, CXXC4	2.36E-07
Intra-golgi and retrograde Golgi-to-ER traffic	KIFC, GALNT1, KIF5A, MANIA1	1.12E-03
Membrane trafficking	SEC23B, KIF5C, GALNT1, KIF5A, DVL2, MANIA1	2.82E-03
MHC Class II antigen presentation	SEC23B, KIF5C, KIF5A	3.87E-03

3.4 Gene biomarker discovery and validation

For identification of novel potential biomarkers, the 4911 upregulated genes common to patients 1 and 2 were used for biomarker filtering analysis. After filtering with the IPA biomarker analysis software, 101 genes were identified as potential biomarkers of pancreatic cancer. The IPA analysis tool filtered genes based on several criteria describe in section 2.2.8 Additional manual biomarker filtering identified 3 novel potential biomarkers – *JUP*, *COPG*, and *TCIRG1*. As mentioned in 2.2.8 manual filtering involved a search through several databases to ascertain the novelty of genes as potential biomarkers. Furthermore, the characteristics of a good biomarker (see section 1.2) were taken into consideration in selecting these biomarkers.

An additional ten samples (5 tumour and 5 normal) were obtained from 6 different patients. These patients' samples were not included in those used for RNA sequencing. All samples were obtained from stage I PDAC patients. Real-time PCR was performed and Cp values obtained. Analysis using the REST software analysed the Cp values of the 3 genes and normalized to the reference gene- *MRPL19*.

Table 3.6 and Figure 3.8 show *JUP* was UP-regulated in tumour (in comparison to normal group); p(H1) of 0.012. The expression of *COPG* in tumour was not different to normal tissues (p(H1) =0.703). TCIRG1 was also unchanged in normal and tissue samples (P(H1) of 0.597). Although the expression of *JUP* was identified as upregulated in tumour tissues it is observed to have a wide 95% confidence interval (CI) of 1.25 - 1,856.39. Future validation studies with the use of a lot more samples would be appropriate to correct the large CI.

Table 3.6: Results of Real-time PCR performed on potential novel biomarkers.
Legend:
p(H1) - Probability of alternate hypothesis that difference between tumour and normal groups is due only to chance.
TRG - Target
REF – Reference

Gene	Type	Reaction Efficiency	Expression	Std. Error	95% C.I.	P(H1)	Result
JUP	TRG	1.0	33.778	3.925 - 407.131	1.246 - 1,856.393	0.012	UP
MRPL19	REF	1.0	1.000				
COPG1	TRG	1.0	1.887	0.132 - 43.829	0.006 - 96.567	0.703	
TCIRG1	TRG	1.0	2.378	0.052 - 77.719	0.003 - 348.572	0.597	

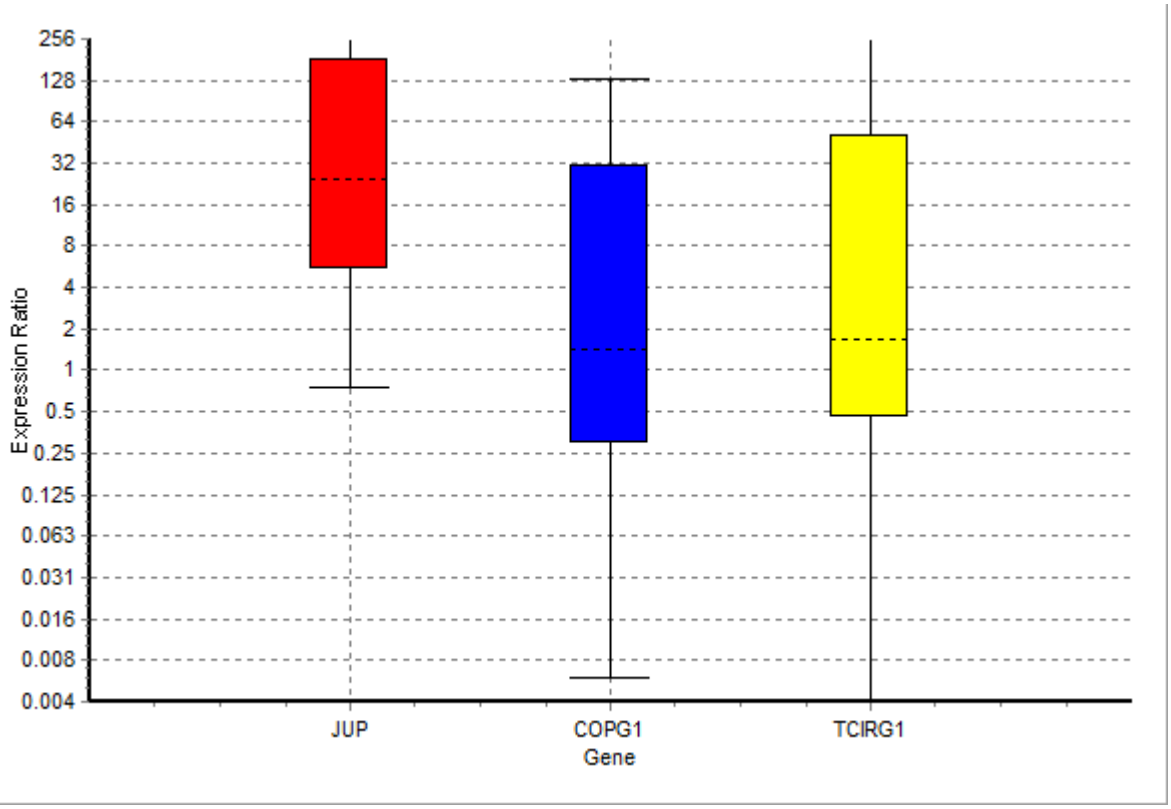


Figure 3.8: Box plot of expression ratio of *JUP*, *COPG*, and *TCIRG1* in tumours (n=6) compared to corresponding normal tissues (n=6). Boxes represents the interquartile range, or the middle 50% of observations. The dotted line represents the median gene expression. Whiskers represent the minimum and maximum observations. Boxplot generated from REST software.

Functional enrichment analysis of the JUP PPI network

The preliminary validation of JUP as a novel potential biomarker, prompted the analysis of PPI network it is involved in. Using the STRING software tool, the PPI network of *JUP* was observed (Figure 3.9). All functional interactors had a confidence score greater than 0.9 (Table 3.7). The STRING software tool uses confidence score to describe the degree of certainty of interaction between molecules. Confidence score rages from 0-1, thus, any score of >0.9 is considered good. Amongst these partners, CTNNA1 is shown to be with the highest confidence score of 0.993. TCFL7 and CDH2 follow with confidence scores of 0.991 and 0.987 respectively.

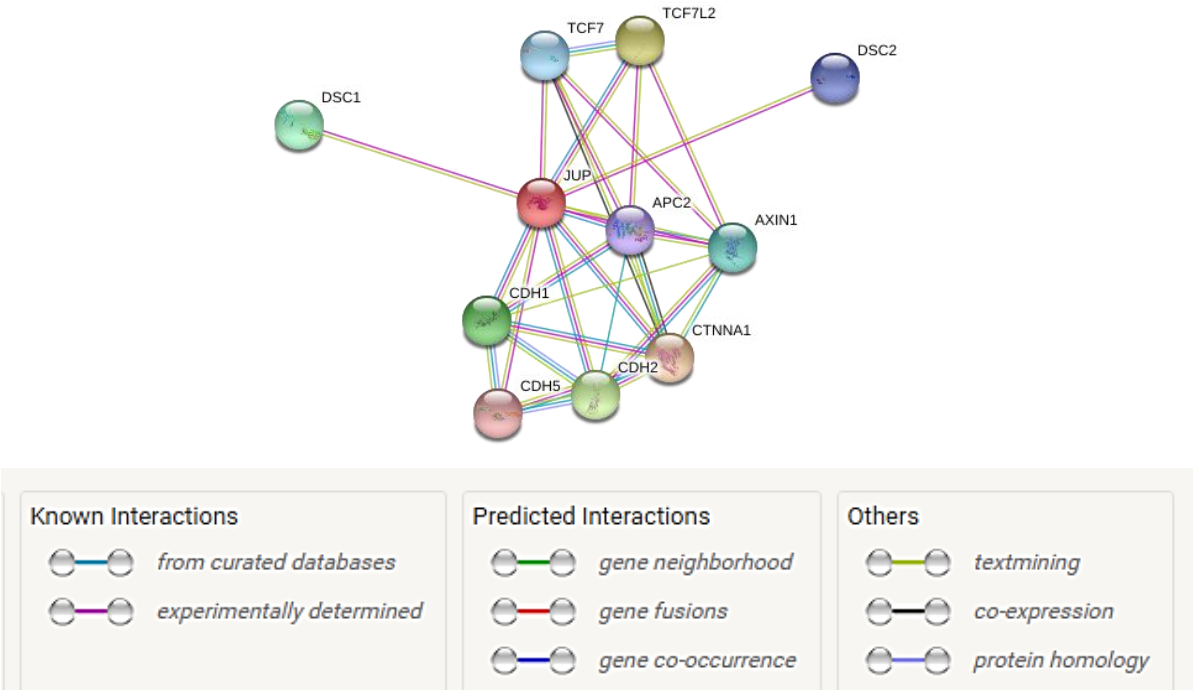


Figure 3.9: Protein-protein interaction network of *JUP*. The network has a PPI enrichment p- value of 1.38E-05. Edges represent various types of interactions as indicated above. Generated from STRING.

Table 3.7: Predicted functional partners of *JUP* PPI network

Protein symbol	**Protein name**	**Confidence Score**
CTNNA1	Catenin alpha 1	0.993
TCF7L2	Transcription factor 7-like 2	0.991
CDH2	Cadherin 2, type 1, N-Cadherin	0.987
CDH1	Cadherin 1, type 1, E-Cadherin	0.986
DSC1	Desmocollin 1	0.949
AXIN1	Axin 1	0.945
TCF7	Transcription factor 7	0.943
DSC2	Desmocollin 2	0.941
APC2	Adenomatosis polyposis coli 2	0.936
CDH5	Cadherin 5	0.935

To further understand the roles played by *JUP* and its interaction partners, functional enrichment analysis of the PPI network was conducted. The biological processes, molecular functions and pathways enriched by the network of proteins were observed.

Cellular response to indole-3-methanol was shown to be the most enriched biological process in the *JUP* PPI network with 3 genes being involved in the process and a false discovery rate (FDR) of 7.97E-06 (Table 3.8). Cell-cell adhesion followed with the most number of genes-7-involved in this process and a false discovery rate of 1.044E-05.

Table 3.8: Top 5 biological processes enriched in JUP PPI network

Pathway ID	**Pathway description**	**Count in gene set**	**False discovery rate**
GO: 0071681	Cellular response to indole-3-methanol	3	7.97E-06
GO:0098609	Cell-cell adhesion	7	1.044E-05
GO:0007156	Homophilic cell adhesion via plasma membrane adhesion molecules	5	2.15E-05
GO:0016337	Single organismal cell-cell adhesion	6	5.9E-05
GO:0046822	Regulation of nucleocytoplasmic transport	5	5.9E-05

Furthermore, genes in the *JUP* PPI network are also implicated in endometrial (FDR= 2.05E-11) and colorectal cancers (FDR= 8.42E-07). Seven of the genes within the network are also shown to be associated with pathways in cancer (FDR= 4.82E-09).

Table 3.9: Top 5 pathways enriched in JUP PPI network

Kegg Pathway ID	Pathway description	Count in gene set	False discovery rate
05213	Endometrial cancer	6	2.05E-11
05412	Arrhythmogenic right ventricular cardiomyopathy	6	1.09E-10
04390	Hippo signalling pathway	6	4.82E-09
05200	Pathways in cancer	7	4.82E-09
05210	Colorectal cancer	4	8.42E-07

Further analysis of the *JUP* PPI network functions showed their involvement in different types of binding (Table A11). Beta-catenin binding was the most enriched with 8 out of 11 genes involved and a false discovery rate of 7.87E-16. 10 genes in the *JUP* PPI network are observed to function in protein binding.

3.5 Discovering genetic variants involved in PDAC

Variant analysis is crucial for elucidating the progression of PDAC as one of the hallmarks of cancer is mutation in certain crucial genes. To identify genetic variants, we used the Ingenuity Variant analysis tool. Using the CLC genomics workbench, variant calling was performed from the RNA sequencing data. Each of the 4 samples had a VCF (Variant calling file) which was loaded into the IVA software. Stringent variant filtering analysis was then undertaken (see 2.2.12).

Figure 3.10 shows a schematic flowchart of the variant filtering done using the Ingenuity variant analysis software. Prior to commencement of filtering, there were 3453 variants in 359 genes observed in our samples. Every filter applied decreased the number of variants. The first filter applied was the confidence filter which reduced the variants to 3268. Five more filters- common variants, predicted deleterious, genetic variants, cancer driver variants and biological context- were also applied. This extensive filtering processes identified 87 variants present in 30 genes. Filtering parameters have been explained in section 2.2.12

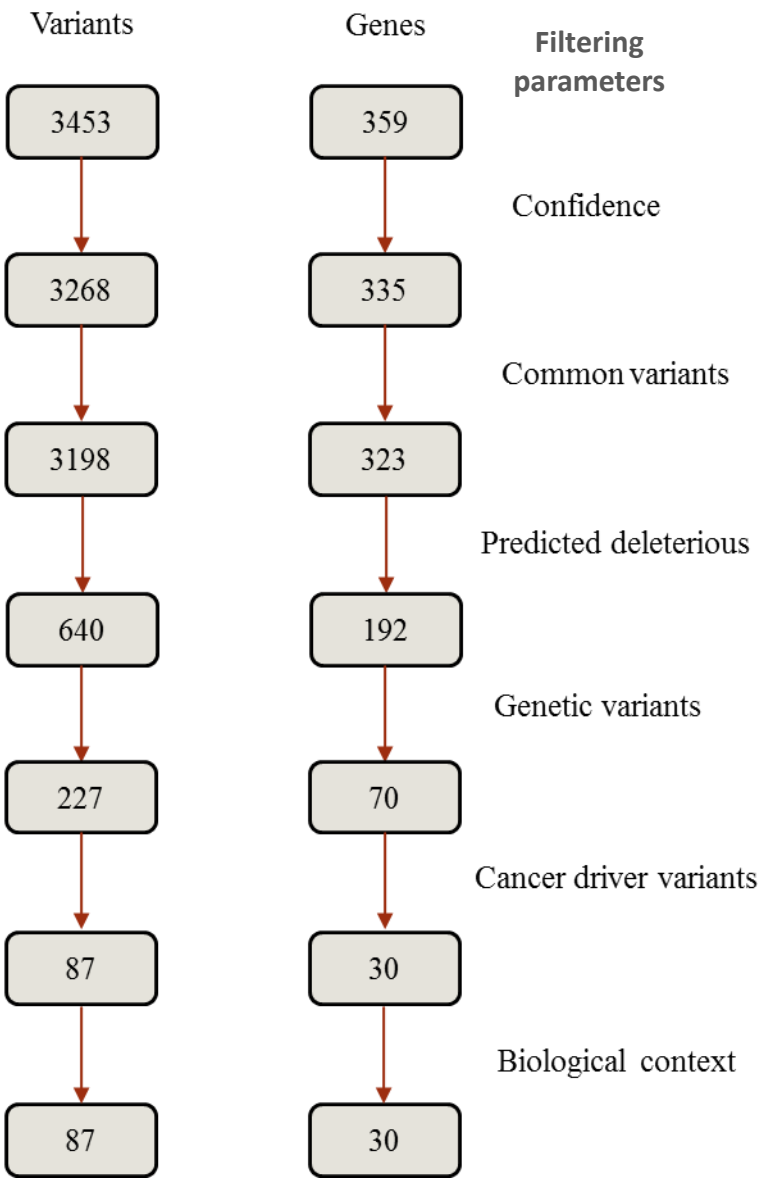


Figure 3.10: Schematic diagram showing the number of resulting variants and genes as each filter is applied. 3453 variants and 359 genes were reduced to 87 and 30 respectively after 6 filters.

Functional enrichment analysis of variants

To further elucidate the roles these genetic variants play in the progression of PDAC, functional enrichment analysis was performed. This analysis was performed using the Ingenuity variant analysis software and it included; grouping genetic variants based on disease phenotype, biological processes, and enriched pathways.

Several crucial pathways were affected by the observed genetic variants (see in Appendix for pie chart). The top 10 pathways in which the variants are enriched are shown in (Table 3.10). The Aryl hydrocarbon signaling pathway was the most enriched (p=1.081E-6) with a total number of 6 variants. Followed by the protein ubiquitination pathway with 12 variants (p=2.859E-06).

Table 3.10: Top 10 pathways enriched in the observed genetic variants.

Pathways	**Number of genes**	**Number of variants**	**P-value**
Aryl hydrocarbon receptor signalling	4	6	1.081E-06
Protein ubiquitination pathway	4	12	2.859E-06
Phagosome maturation	3	4	1.545E-05
Small cell lung cancer signalling	3	6	1.678E-05
Antigen presentation pathway	2	2	1.200E-04
cdc 42 signalling	3	3	1.746E-04
Neuroprotective role of THOP1 in Alzheimer's disease	2	2	1.816E-04
Xenobiotic metabolism signalling	3	3	2.046E-04
Dendritic cell maturation	3	6	2.227E-04
Huntington's disease signalling	3	3	2.264E-04

Furthermore, the identified variants were enriched in several biological processes (see in appendix section). Genetic variants would impact crucially on these processes. The top 10 biological processes enriched by the 87 variants were also observed (Table 3.11). Interestingly, apoptosis was identified as the most enriched (p= 1.307E-25) with 69 variants associated in the biological process. Cell death followed with 71 variants enriched (p=1.702E-24). Other enriched biological processes include; necrosis, developmental biology, and the proliferation of cells.

Table 3.11: Top 10 biological processes enriched in the observed genetic variants.

Biological processes	**Number of genes**	**Number of variants**	**P-value**
Apoptosis	26	69	1.307E-25
Cell death	27	71	1.702E-24
Cell death and survival	27	71	8.987E-24
Necrosis	24	50	1.356E-21
Proliferation of cells	25	66	3.538E-21
Molecular biology	26	70	5.182E-21
Disorder of gentinourinary system	28	83	6.579E-21
Cell biology	30	87	7.131E-21
Developmental biology	25	68	2.695E-20
Immunology	27	78	3.484E-20

The diseases in which our 87 variants were enriched were investigated and 30 diseases identified (see Appendix section). The top 10 diseases in which the variants are known to be enriched are shown in (Table 3.12). Gastrointestinal tract cancer was observed to be the top disease with 86 variants being enriched (p=1.375E-23). Diseases of the pancreas was also observed to be amongst the top 10 most enriched diseases with the involvement of 71 variants (p= 2.130E-22).

Table 3.12: Top 10 diseases enriched in the observed variants.

Diseases	**Number of genes**	**Number of variants**	**p-value**
Gastrointestinal tract cancer	29	86	1.375E-23
Gastrointestinal tract cancer & tumours	29	86	2.042E-23
Abnormalities of abdominal organ	29	86	9.574E-23
Gastrointestinal disturbance	29	86	1.133E-22
Reproductive system disease	28	83	1.365E-22
Diseases of the gastrointestinal tract	29	86	1.380E-22
Diseases of the pancreas	24	71	2.130E-22
Adenocarcinoma	29	86	3.622E-22
Endocrine system disorder	27	79	8.876E-22
Digestive system cancer	29	86	1.169E-21

Novel variant discovery

The 87 variants were further filtered manually to identify novel ones. The manual filtering process involved; the selection of genetic variants present in both tumour samples only, those resulting in loss of gene function and that have not been previously reported on in any publication. It is important to note that it was the IVA software tool that predicted these variations to result in loss of gene function. This manual filtering led to the identification of 6 variants present in 4 genes- *SPARC, ZFHX3, COL1A1* and *TGM2*. Table 3.13 shows the variants, region, and type of variations, including their impact on translation.

The *COL1A1* contains 3 variants leading to splice site loss. A deletion in the exonic region of *ZFH3* and *SPARC* results in frameshift. Furthermore, a single nucleotide variation at the splice site of TGM2 results in a splice site loss.

Table 3.13: Showing manually filtered variants resulting in loss of gene function.

Gene symbol	**Chromosome**	**Position**	**Gene region**	**Type of variation**	**Translation impact**
SPARC	5	151043159	Exonic	Insertion	Frameshift
ZFHX3	16	72821648	Exonic	Deletion	Frameshift
COL1A1	17	48263999	Splice site	SNV	Splice site loss
COL1A1	17	48265999	Splice site	Insertion	Splice site loss
COL1A1	17	48268176	Splice site	SNV	Splice site loss
TGM2	20	36784493	Intronic, splice site	SN	Splice site loss

3.6 Differential protein expression in occurring in pancreatic tumours

To identify differentially expressed proteins occurring during PDAC, human oncology proteome profiler arrays were used. As described in section 2.2.13, these arrays are highly sensitive antibody arrays with in-built technical replicates. Three patients, samples were used. From each patient, one normal and one tumour tissue was obtained. All samples were obtained from stage I PDAC patients. Total protein was extracted from tissues and array experiment conducted. The arrays were viewed in a Chemi Doc instrument and images captured (Figure 3.11).

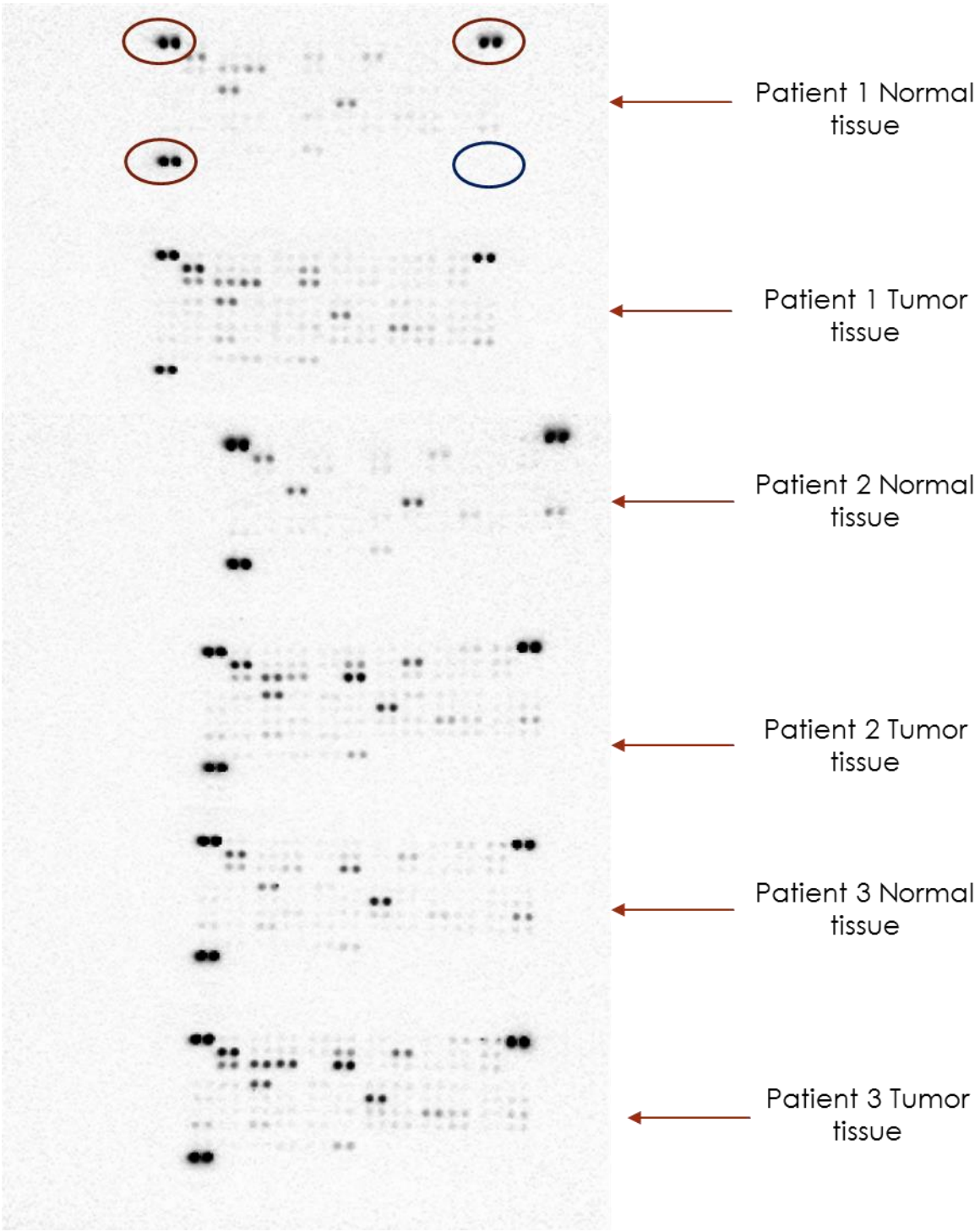


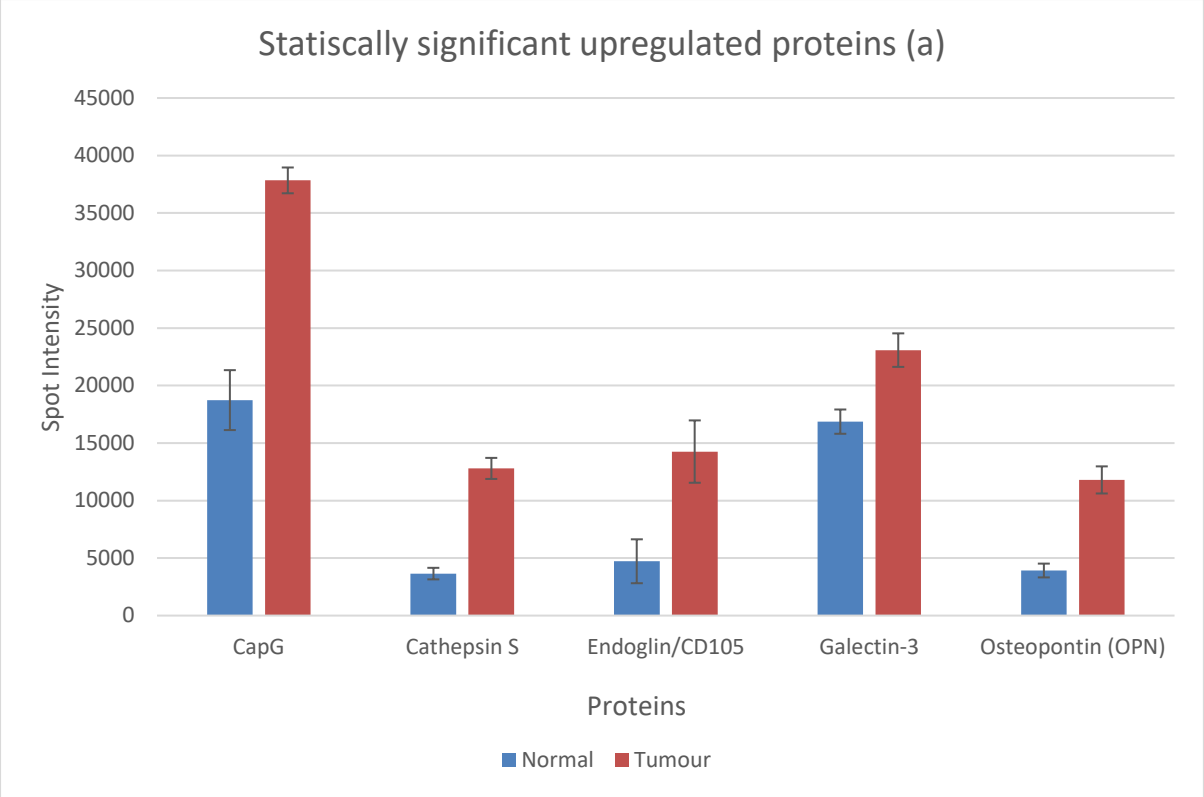
Figure 3.11: Arrays showing protein expression profile for normal and tumour tissues from 3 patients. Each paired spot represents a protein. As spot intensity increases protein expression is said to increase. Paired spots at the top left, bottom left and top right (all shown by red circles on patient 1 normal tissue) serve as positive control while bottom right (blue circle) serve as negative control. These positive and negative controls are present in same position (indicated in patient 1 normal) in all arrays.

Significantly differentially expressed proteins

Captured images were further analysed using the wvision Quickspot software. The wvision Quickspot software provided spot intensity values for each protein. These values were then exported to Microsoft Excel. The averages and standard error values of each protein in both normal and tumour tissues were calculated. The spot intensity values for each protein in the tumour were then compare to those of normal counterparts. Only differentially expressed genes with $p < 0.05$ were considered. No significantly downregulated proteins were observed in tumour samples compared to normal tissues. However, 19 significantly upregulated proteins (with p-value >0.05) were observed in tumour tissues compared to normal tissues (Table 3.14). These proteins were further plotted on a bar chart. Nectin -4 was identified to be the most significantly upregulated with a p-value of 4.852E-03. This was followed by P27/Kip1 (CDKN1B) and MUC-1 with p-values 5.966E-03 and 8.420E-03. The spot intensity for CapG is seen to be the highest with about 40000 (Figure 3.12), however, when compared to its normal samples it places 4th on the list of significantly upregulated proteins.

Table 3.14: Significantly upregulated proteins with their p-values

Protein	**p-value**
Nectin-4	4.852E-03
P27/Kip 1 (CDKN1B)	5.966E-03
MUC-1	8.420E-03
CapG	1.021E-02
Cathepsin D	1.229E-02
ERalpha/NR3A1	1.575E-02
Cathepsin S	1.702E-02
ErbB3/Her3	1.741E-02
Hif-1 alpha	1.820E-02
Endoglin/NR3A1	1.864E-02
EGFR/ErbB1	2.099E-02
Galectin-3	2.378E-02
Cathepsin B	2.712E-02
ENPP-2/Autotaxin	2.829E-02
SPARC	3.125E-02
Angiopoietin-like 4	3.958E-02
CEACAM-5	4.112E-02
Osteopontin (OPN)	4.296E-02
DLL1	4.873E-02



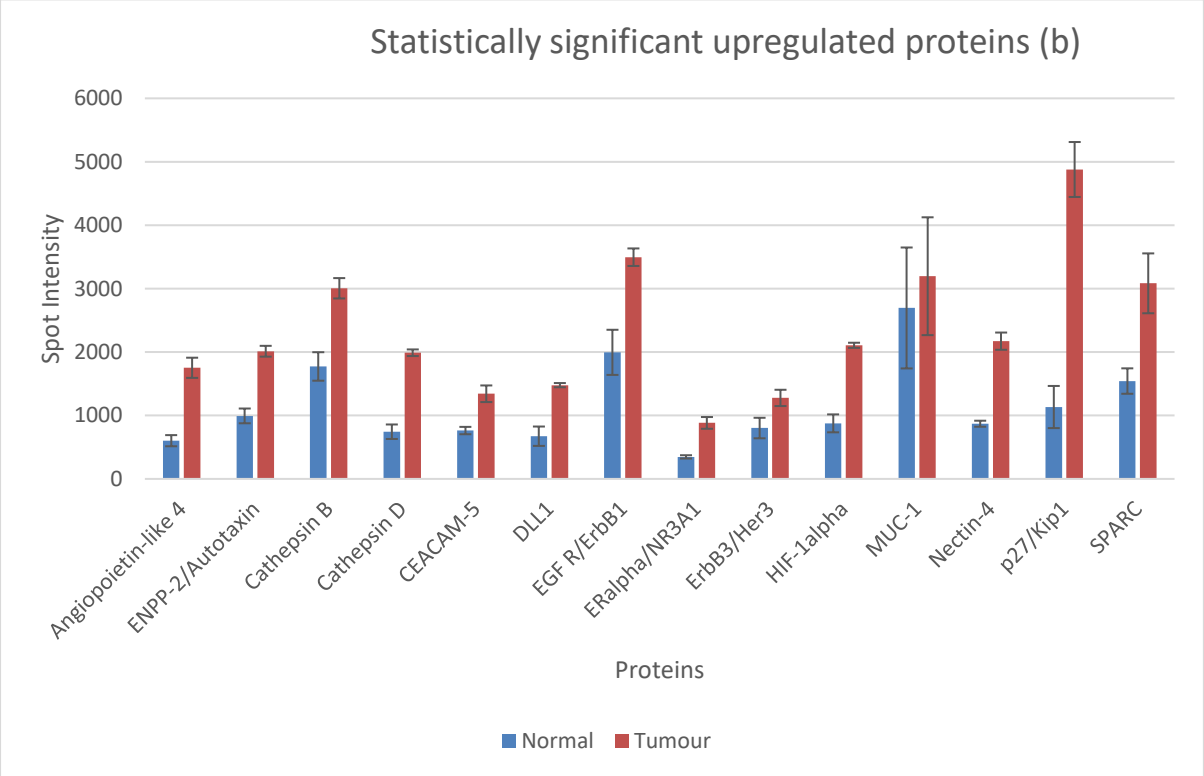


Figure 3.12: Bar charts of upregulated proteins. Proteins in chart (a) have maximum spot intensity of 40000 while those in chart (b) have a maximum of about 6000. Blue bar indicates normal expression and Red indicates tumour tissue expression. Standard error bars are also shown.

Pathway analysis of significantly upregulated proteins

To identify the pathways in which upregulated proteins were enriched, pathway analysis using REACTOME was performed. The 18 significantly upregulated proteins were input into REACTOME. Visual assessment of Figure 3.13 show that the significantly upregulated proteins were enriched mostly in pathways associated with signal transduction and immune system. These pathways include the ErbB and p13/Akt signalling pathways. Interestingly, enrichment in signal transduction pathways were also observed in upregulated genes (Table 3.15). The ErbB signalling pathway was observed to the most upregulated (p= 4.9E-05). The proteins, EGFR, ERBB3, and p27 are crucially involved in this signalling pathway.

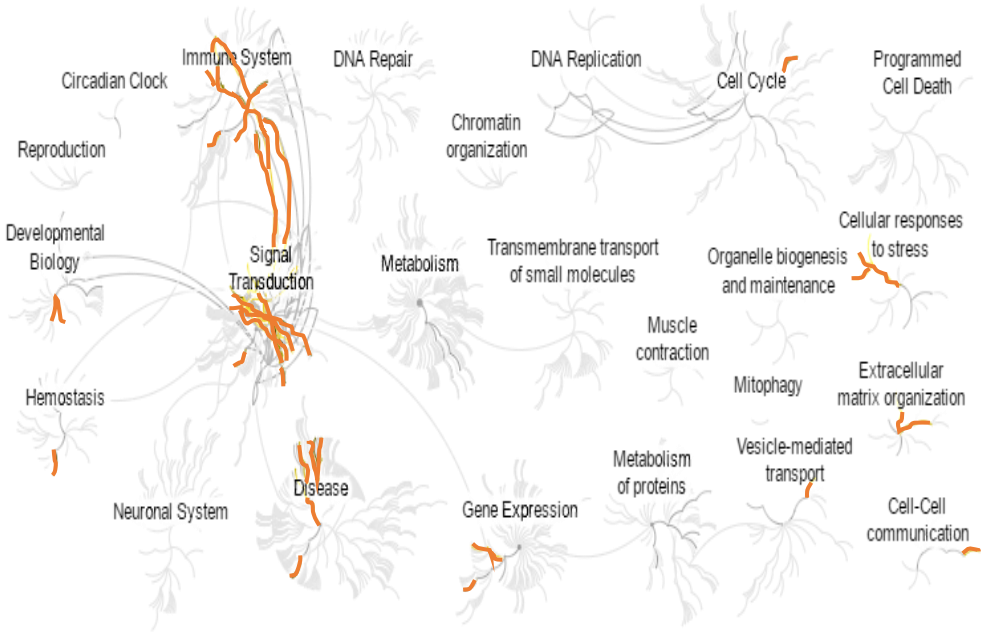


Figure 3.13: Diagrammatic representation pathways in which significantly upregulated proteins are enriched. Orange lines indicate enriched pathways. Generated from REACTOME software.

Table 3.15: Showing Top 10 enriched pathways their p-values and candidate proteins.

Pathway	**p-value**	**Candidate proteins**
ErbB signaling pathway	4.9E-05	EGFR, ERBB3, CDKN1B
MHC class II antigen presentation	5.74E-05	CTSS, CTSB, CTSD
Diseases of signal transduction	7.68E-05	ERBB3, EGFR, DLL1, CDKN1B
Adaptive immune system	8.14E-05	CTSD, CDKN1B, ERBB3, CTSB, CTSS, EGFR
miR-target in muscle cell-Tarbase	1.18E-04	EGFR, CAPG, ERBB3, SPARC, CDKN1B
Trafficking & processing of endosomal TLR	1.3E-04	CTSS, CTSB
Signalling pathway in endosomal TLR	1.68E-04	EGFR, CDKN1B, ERBB3
Signalling pathway in Glioblastoma	1.74E-04	EGFR, ERBB3
P13/AKT signalling in cancer	1.8E-04	ERBB3, EGFR, CDKN1B
Innate immune system	1.87E-04	CDKN1B, ERBB3, LALS3, CTSB

3.7 Protein biomarker discovery and PPI network analysis

Biomarker discovery

To identify novel potential protein biomarkers for PDAC the 18 upregulated proteins underwent the same filtering process described in section 3.4. This process included the use of the Ingenuity pathway analysis software tool in combination with manual filtering process. After the filtering processes, ANGPTL4 and p27/kip1 (CDKN1B) were identified as potential biomarkers as they bear the hallmarks of a good biomarker (see Section 1.2).

PPI network analysis of novel potential protein biomarkers

To understand the roles ANGPTL4 and p27/kip1 play in pancreatic carcinogenesis their PPI networks were analysed using the STRING software. The STRING software also generated the biological processes, pathways, and molecular functions in which both PPI networks were enriched. This analysis further helped in making informed hypothesis on the roles the ANGPTL4 and p27/Kip1 may play in PDAC.

The PPI network show ANGPTL4 is involved in an intricate network of proteins (Figure 3.14). Ten proteins were observed to be involved within this PPI network (Table 3.16). All 10 functional partners are shown to have confidence scores above 0.9. The strongest functional interacting partner of ANGPTL4 was observed to be PPARA with a confidence score of 0.990.

PPARG and FABP4 follow closely with confidence scores of 0.949 and 0.939 respectively. LPL was also shown to be a functional partner with a confidence score of 0.937.

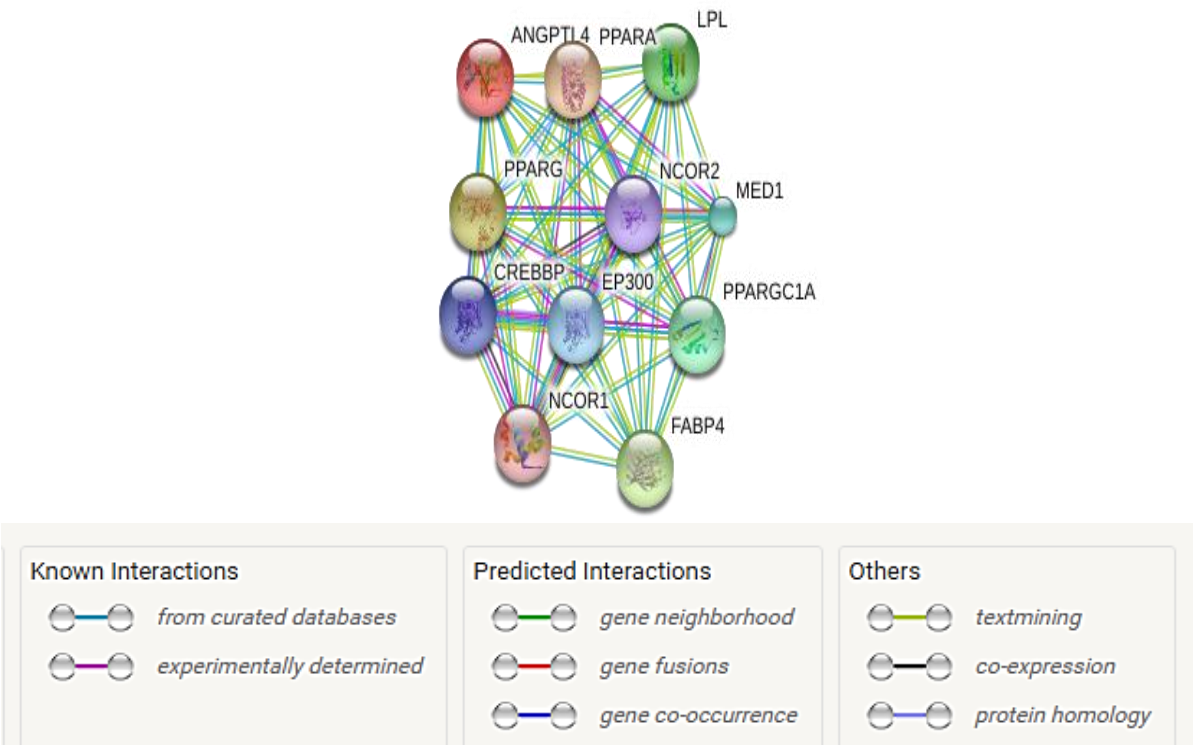


Figure 3.14: Protein-Protein interaction network of ANGPTL4. The network has a PPI enrichment p-value of 4.44E-16. Edges represent various types of interactions as indicated above. Generated from STRING.

Table 3.16: Predicted functional partners of ANGPTL4

Protein symbol	**Protein name**	**Confidence Score**
PPARA	Peroxisome proliferator-activated receptor alpha	0.990
PPARG	Peroxisome proliferator-activated receptor gamma	0.949
FABP4	Fatty acid binding protein 4	0.939
LPL	Lipoprotein lipase	0.937
PPARGC1A	Peroxisome proliferator-activated receptor gamma, cofactor 1 alpha	0.917
MED1	Mediator complex subunit 1	0.914
EP300	E1A binding protein p300	0.908
CREBBP	CREB binding protein	0.907
NCOR2	Nuclear receptor corepressor 2	0.906
NCOR1	Nuclear receptor corepressor 1	0.905

Cellular lipid metabolism was identified to be the most enriched biological process in the ANGPTL4 PP1 network with a false discovery rate of 5.99E-10 (Table 3.17). Ten proteins in the PPI network are known or predicted to be involved in cellular lipid metabolism. Transcription initiation from RNA polymerase II promoter is 2nd most enriched with a false discovery rate of 6.89E-08. Also, important to note is the biological process involve response to oxygen levels in which 6 out of 11 PPI network proteins are known or predicted involved.

Table 3.17: Top 5 biological processes in which the ANGPTL4 PPI network is enriched

Pathway ID	**Pathway description**	**Count in gene set**	**False discovery rate**
GO: 0044255	Cellular lipid metabolism process	10	5.99E-10
GO: 0006367	Transcription initiation from RNA polymerase II promoter	7	6.89E-08
GO: 0044281	Small molecule metabolic process	10	2.08E-06
GO: 0055088	Lipid homeostasis	5	2.42E-06
GO: 0070482	Response to oxygen levels	6	6.38E-06

Furthermore, pathways enriched by the ANGPTL4 PPI network were observed. The PPAR signalling pathway wasobserved to be the most enriched with 5 proteins being involved (Table 3.18). The Notch signalling pathways is also shown to be enriched with a false discovery rate of 1.34E-04.

Table 3.18: Top 5 pathways in which the ANGPTL4 PPI network is enriched.

Kegg Pathway ID	**Pathway description**	**Count in gene set**	**False discovery rate**
03320	PPAR signalling pathway	5	4.56E-08
04919	Thyroid hormone signalling pathway	4	4.35E-05
04330	Notch signalling pathway	3	1.34E-04
05016	Huntington's disease	4	1.34E-04
05169	Epstein-Barr virus infection	3	6.96E-03

Transcription factor activity and binding was observed to be the most enriched molecular function with this PPI network with a false discovery rate of 8.09E-08 (Table A11). Transcription factor binding followed with 7 out of 11 PPI network proteins being known or predicted to be implicated.

Similarly, the p27/kip1 PPI network was investigated and 11 functional interacting proteins identified (Figure 3.15). All functional interactions of p27/kip1 have confidence scores of 0.999 (Table 3.19).

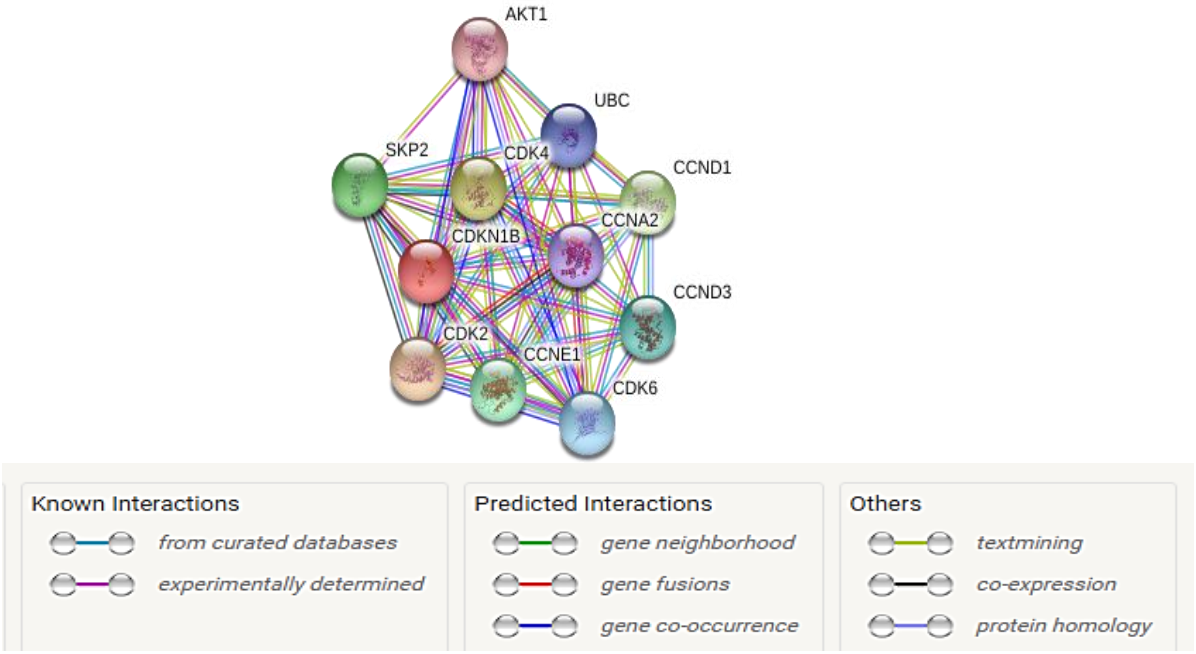


Figure 3.15: Protein-Protein interaction network of CDKN1B/p27. The network has a PPI enrichment p- value of 3.03E-06. Edges represent various types of interactions as indicated above. Generated from STRING.

Table 3.19: Predicted functional partners of CDKN1B/p27

Protein symbol	**Protein name**	**Confidence Score**
CDK2	Cyclin-dependent kinase 2	0.999
CDK4	Cyclin-dependent kinase 4	0.999
CCND1	Cyclin D1	0.999
SKP2	S-phase kinase associated protein 2	0.999
CCNE1	Cyclin E1	0.999
CCND3	Cyclin D3	0.999
CDK6	Cyclin-dependent kinase 6	0.999
UBC	Ubiquitin C	0.999
CCNA2	Cyclin A2	0.999
AKT1	V-akt murine thymoma viral oncogene homolog 1	0.999

The CDKN1B PPI network is mostly enriched in biological processes associated with mitosis (Table 3.20). G1/S transition of mitotic cell cycle was observed to be the top most enriched biological processes in the PPI network with a false discovery rate of 5.41E-14.

Table 3.20: Top 5 biological processes in which the CDKN1B/p27 PPI network is enriched.

Pathway ID	Pathway description	Count in gene set	False discovery rate
GO: 0000082	G1/S transition of mitotic cell cycle	9	5.41E-14
GO: 0044772	Cell cycle phase transition	9	6.16E-12
GO: 0000278	Cell cycle	9	1.69E-08
GO: 0051726	Regulation of cell cycle	9	4.71E-08
GO: 0045787	Positive regulation of cell cycle	7	6.7E-08

The cell cycle was the most enriched in the CDKN1B PPI network with 8 interactors involved in the pathway and a false discovery rate of 5.53E-14 (Table 3.22). Pathways involved in viral carcinogenesis and small cell lung cancer were also observed to be enriched both with false discovery rates of 4.96E-13.

Table 3.21: Top 5 pathways in which the CDKN1B PPI network is enriched.

Kegg Pathway ID	Pathway description	Count in gene set	False discovery rate
04110	Cell cycle	8	5.53E-14
05203	Viral carcinogenesis	8	4.96E-13
05222	Small cell lung cancer	7	4.96E-13
05162	Measles	7	7.8E-12
05161	Hepatitis B	7	1.06E-11

The top 5 molecular functions enriched in the CDKN1B PPI network were also observed. Amongst these, the cyclin-dependent protein serine/threonine kinase regulator activity was the most enriched with a false discovery rate of 1.11E-06 and 4 interactors involved (Table A13). Furthermore, 7 members of the CDKN1B PPI network were shown to function as kinases.

CHAPTER 4- DISCUSSION

4.1 Overview

Due to the devastating nature of PDAC, the discovery of biomarkers for pancreatic cancer is crucial for diagnostic, prognostic, and therapeutic purposes. In this study, three (ANGPTL4, p27 and *JUP*) novel potential biomarkers of PDAC were identified. Furthermore, validation tests performed in this study utilizing multiple tissue samples have shown the overexpression of *JUP*, providing further evidence for its use as a potential biomarker. Also, the use of highly sensitive proteome profiler arrays for discovery of ANGPTL4 and p27 validates their overexpression in tumours. These identified potential biomarkers exhibit fitting characteristics such as: significantly overexpressed in only tumour samples compared to corresponding normal tissues and secretion into the bloodstream for easy accessibility in diagnostic testing. Importantly, these potential biomarkers are overexpressed in stage I PDAC, suggesting they may be useful for early diagnosis. This study suggests that the use of these biomarkers in combination may provide increase specificity and sensitivity to PDAC.

Genetic variations identified by the Ingenuity variant analysis software in *SPARC*, *ZFHX3*, *TGM2* and *COLIA1* that are predicted to result to loss of gene function occurring during PDAC. It is important to note that this is the first pancreatic cancer study that reports variations in some of these gene (i.e. *ZFHX3*, *TGM2* and *COLIA1)*. These genes have been observed to play tumour suppressive roles. In several other cancers, they have been identified to induce cell death. I propose that the reinstating of the normal functions in these genes by providing targeted therapies, might result to the inhibition of pancreatic tumorigenesis.

Furthermore, the present study shows the dysregulation of several signalling pathways such as IGF/MAPK, integrin, and Ras. Dysregulation would result in aberrant biological processes. It is possible that these pathways play critical roles in propagating pancreatic tumorigenesis and that targeting them might be beneficial for PDAC treatment.

I also identified the downregulation of *CXXC4*, *GALNT1*, *HS3ST4*, *SLC15A4*, and *TSPAN7* during pancreatic cancer and hypothesize that this drives tumorigenesis. These genes are known to be involved in the inhibition of cellular proliferation and survival. Targeting them for potential drug discovery might prove beneficial

Finally, this study reveals the upregulation of genes involved in the necroptic and immune system pathways in tumours. Necrosis in tumours has been observed to result in poor outcomes

for cancer patients. Therefore, necrosis in tissues may be way of determining outcome in pancreatic cancer patients thereby enabling the use of targeted therapies for treatment.

The immune system is activated during PDAC; however, cancer cells have developed several mechanisms to circumvent detection. This study has identified the activation of some of these mechanisms including: the upregulation of VEGF and cathepsin S corroborating others that have identified this during tumorigenesis. These upregulations may lead to blocking of anti-tumorigenic functions of the immune system.

In the following sections I discuss in detail the above findings of this study

4.2 Novel potential biomarkers of PDAC

A biomarker (biological marker) functions as an indicator of a diseased state. It can also help in the monitoring of the progression of a disease and a patient's response to treatment. The characteristics of a good biomarker have been well described (Section 1.2). Due to the aggressive nature of PDAC, there is urgent need for biomarkers in diagnosis and prognosis. In this study, I have identified 3 novel potential biomarkers of PDAC using highly sensitive and high throughput discovery and validation techniques. Importantly, I have employed a series of processing methods to ensure high sample quality. Furthermore, the identified potential biomarkers meet the hallmarks of a good biomarker (also described in Section 1.2) including; being secreted in the blood (for easy accessibility) and differentially expressed in normal and tumour samples. I describe these identified potential biomarkers below stating supporting reasons for their proposition as candidate for early diagnostic testing.

Junction plakoglobin (JUP)

In this study, RNA sequencing was utilized for differential gene expression analysis. The 4911 upregulated genes then went through a biomarker filtering process involving the use of the IPA software that identified potential biomarkers. Validation test with additional tumours (n=5) and their corresponding normal tissues (n=5) was also conducted using real time PCR. This led to the identification of *JUP* as being upregulated in tumour samples but not in normal tissues.

Junction plakoglobin (*JUP*) gene encodes an important cytoplasmic protein called plakoglobin (also known as γ catenin). *JUP* functions in cell-cell adhesion as an important component of desmosomes and adherens junction. Furthermore, *JUP* is expressed in the blood making it easily accessible to be measured (Cooksley-Decasper et al., 2012). The easy accessibility of biomarkers is crucial as it enables for less invasive ways of testing for a diseased state.

Additionally, the observed overexpression of *JUP* observed in this study in all tumours but not in corresponding normal tissues might be indicative of its specificity to PDAC, further signifying its role as a biomarker of PDAC. Importantly, tumour samples used in this study for both discovery (RNA sequencing) and validation (real time PCR) tests were from early stage (stage I) PDAC patients. The overexpression of *JUP* in early stage PDAC is a critical characteristic for its proposed role as a potential diagnostic biomarker.

This study contradicts others in which the underexpression of JUP was observed in pancreatic tumours and cell lines (Pryczynicz et al., 2010; Toyoda et al., 2005). This contradiction might be because in this study, samples from black South African population were used and gene expression profiles are known to vary amongst population groups (Storey et al., 2007). As this seem to be the first study to profile extensively genes in pancreatic tumours obtained from any black African PDAC patient, I suggest that this observed overexpression might be unique to the population group, however, more validation tests are needed to confirm this. Also, the method of discovery used in other studies differ with the present one and this might result in contradicting results. This study however resonates with others done with testicular germ cell tumours in which an overexpression of *JUP* was identified (Skotheim et al., 2003). Although, JUP was observed to be overexpressed in testicular germ cell tumours, it could still be used in combination with other potential biomarkers identified in this study for early diagnostic PDAC testing. This combination might enhance sensitivity and specificity for PDAC detection

This study has identified *JUP* as a potential biomarker for PDAC. Future studies with a much larger sample amount should be done to further verify this potential biomarker. Additionally, the use biological samples such as blood from different stages of PDAC to observe the expression pattern of *JUP* as the tumour progresses might prove useful to corroborate these findings.

Angiopoietin-like 4 (ANGPTL4)

The Human Oncology Proteome Profiler Array allowed for the focused analysis of protein expression. The highly sensitive and specific array panel was used to compare protein expression profile in normal and tumour samples. The 18 significantly upregulated proteins observed in this study were then subjected to the filtering process (Section 2.2.8). This process then identified ANGPTL4 as a potential biomarker for PDAC.

ANGPTL4 is a member of the ANGPTL protein family. The *ANGPTL4* gene encodes a glycoprotein with a mass of approximately 45-65kDa. Importantly, this study showed that

ANGPTL4 was significantly (p=3.958E-02) overexpressed in all tumour samples but not in corresponding normal tissues. As the samples used were obtained from early stage (stage I) PDAC patients, this overexpression of ANGPTL4 in tumour tissues might indicate its role as a diagnostic biomarker. Also, the use of the highly sensitive proteome profiler array and the overexpression of ANGPTL4 in all tumour samples further supports its specificity and sensitivity to PDAC. Additionally, ANGPTL4 has been observed to be secreted and expressed in the blood (Mandard et al., 2004). This makes it easily accessible and a good candidate biomarker.

Furthermore, the significant upregulation of ANGPTL4 in early stage pancreatic tumour suggests that it plays an important role in tumorigenesis. Although there have been conflicting reports over the years growing evidence show that ANGPTL4 plays a pro-tumorigenic role (Shibata et al., 2010). Shibita and colleagues observed the overexpression of ANGPTL4 oesophageal carcinoma and the levels increase as tumour develop from benign to malignant status. This suggests that ANGPTL4 expression levels could be used to predict prognosis in carcinoma cells.

Tumours produce an increased amount of reactive oxygen species. However, these tumours develop a strategy to escape death by anoikis. Anoikis is a type of programmed cell death in which certain cells die when separated from the extracellular matrix. This anoikis-resistance in crucial to tumorigenesis. A study showed that ANGPTL4 instigated anoikis resistance to tumour cells (Shibata et al., 2010). Overexpression of ANGPTL4 implies increased anoikis resistance in tumour cells enabling cells to survive and migrate uncontrollably.

Additionally, PPI network analysis using STRING software helped to identify possible roles of ANGPTL4 in PDAC. The analysis showed that ANGPTL4 plays a role in PPAR signalling pathway (Table 3.18). The PPAR family members are known to regulate the expression of ANGPTL4 (Georgiadi et al., 2010; Kersten et al., 2009) and are known to play crucial roles in lipid homeostasis and metabolism (Kersten, 2005). They have also been observed to be involved in tumorigenesis. With its interaction with the PPAR family, ANGPTL4 is implicated in tumorigenesis, however, there is still no evidence that the PPAR family induces the overexpression of ANGPTL4.

PPI network analysis also showed ANGPTL4 plays a role in lipid metabolism (Table 3.17). ANGPTL4 is known to inhibit lipoprotein (LPL) the enzyme responsible for the hydrolysis of triacylglycerols (TAGs). This inhibition by ANGPTL4, prevents the removal of circulating

triacylglyerols. ANGPTL4 impedes LPL by converting active lipoprotein dimers to their inactive form- monomers. The upregulation of ANGPTL4 results in an increase in circulating triacylglycerols whereas, a downregulation results in a decrease of TAGs (Zhu et al., 2012). An increase in triacylglycerol level can result in the inflammation of the pancreas observed in pancreatitis which is known to increase the risk of pancreatic cancer.

The significant upregulation of ANGPTL4 in early stage PDAC hints on its role as a potential diagnostic biomarker. Although, highly a sensitive technique was used for the discovery of ANGPTL4 as a potential biomarker, future validation tests are crucial to corroborating these findings. Future validation tests should include the use of easily accessible blood samples like blood and a much larger sample number.

p27/kip1

As described above in the ANGPTL4 section, the Human Oncology Proteome Profiler Array was used for differential protein expression analysis. From 18 significantly upregulated proteins p27 was identified as potential biomarker by the IPA software. This study show that p27 is significantly overexpressed in tumour tissues compared to normal ones. This overexpression was observed in all tumour samples indicating its sensitivity to PDAC. Also, p27 has been identified as a secretory protein (Fero et al., 1998) and its presence in the plasma makes it a potential candidate biomarker. Also, as described previously all tumour samples were obtained from stage I PDAC patients and the overexpression of p27 might indicate its used as an early indicator of PDAC.

p27 is encoded by the *CDKN1B* gene and acts as tumour suppressor protein. It functions in regulating the cell cycle at the G0 to S phase transitions. p27 carries out its roles by binding and modulating the actions of cyclin-dependent kinases (CDKs) (Sherr and Roberts, 1999) . Increased levels of active tumour suppressor proteins like p27 usually correlate to cancerous cell deaths. Most tumours in several human organs (including lung, stomach, liver, colon, and breast) show a reduction in p27 expression due to proteasomal degradation (Chu et al., 2008). This reduction usually correlates to poor prognosis. In this study, I have observed the significant (p=5.966E-03) upregulation of p27 in all stage I PDAC tumour tissues (Table 3.14, Figure 3.12) and propose its function as a novel potential diagnostic biomarker.

In most tumours, p27 is inactivated by mislocalization from the nucleus to the cytoplasm and phosphorylation (Lee and Kim, 2009). Phosphorylation of p27 at the Thr157 and Thr198 by Akt enables it to form a complex with cyclin D and CDK4. In tumour cells, active p27-cyclin

D-CDK4 accumulates in the cytoplasm (Soos et al., 1996). Oncogenic stimulation of MAP2K, Akt, and *Ras* (Cheng et al., 1998; Liu et al., 2000; Viglietto et al., 2002) signalling pathways is responsible for this. Interestingly, this present study showed that these pathways are upregulated in pancreatic tumours (Table 3.7, 3.21) indicating that they might be involved in the inactivation of p27. Although p27 is observed to be upregulated in this study, the upregulation of its inactivators may render it non-functional and therefore as good as a reduced one. Additionally, germline mutations in p27 can render it inactive. These mutations have been observed in breast cancers (Chappuis et al., 2000) and endocrine neoplasms (Pellegata et al., 2006). The suggested inactivation of p27 may be responsible for its overexpression in PDAC tumours which makes it a potential biomarker.

The significant upregulation of p27 in all stage I tumours observed in this study, is characteristic of a potential diagnostic biomarker, however, further validation studies are necessary to support this claim.

4.3 Genetic variants might induce tumorigenesis

Variations in genes might affect the roles they play in biological processes and signalling pathways. Variations in certain genes may lead to uncontrolled cellular proliferation and tumour invasion. For example, the inactivation of tumour suppressive genes due to loss of gene function is characteristic of cancerous cells (Hanahan and Weinberg, 2000). These mutations can result from genetic single nucleotide polymorphisms, histone modifications and DNA methylations (Berdasco and Esteller, 2010; Esteller, 2007; Jones and Baylin, 2007). In PDAC, some gene alterations result in loss of gene function driving carcinogenesis.

In this study, variant calling was performed from RNA sequencing to produce VCF (variant calling files) for each sample which were interrogated to identified novel genetic variants. Files were then uploaded into Ingenuity Variant Analysis software tool. Six variants were identified in 4 genes and in all tumour samples. These variants were only present in tumour samples and not in their corresponding normal tissues. They include: insertions, deletions, and single nucleotide variations, all of which are predicted (by the IVA software tool) to result in a loss of gene function. The identified genes have been described in detail below.

Secreted protein acidic and rich in cysteine (SPARC)

An insertion at the exonic region of *SPARC* at position 151043159 (Chr 5) causes a frameshift mutation predicted to result in loss of gene function has been observed in pancreatic tumours

(Table 3.13). Notably, this mutation was observed in stage I PDAC tumours suggesting that it might play a role in inducing and enhancing pancreatic tumorigenesis. This result resonates with several studies suggesting that a deletion or mutation in *SPARC* is a critical feature premalignant lesions of the pancreas (Vaz et al., 2015). Importantly, variant analysis in this study identify SPARC as a cancer therapeutic target suggesting it might be of significance during carcinogenesis.

SPARC gene encodes osteonectin (SPARC), a glycoprotein belonging to a group of matricellular proteins. It is multifunctional and involved in an array of biological processes including development, cell adhesion, cellular differentiation, and proliferation. Although its role in cancer is not fully understood, it has been reported to function in angiogenesis, apoptosis, and cellular proliferation (Rossi et al., 2016; Vaz et al., 2015).

Furthermore, the hypermethylation of *SPARC*, which silences the gene, at the CPG islands in pancreatic cancer cell lines has been identified (Nagaraju and El-Rayes, 2013). Hahn and colleagues also observed that a loss of heterozygosity in *SPARC* results in the inactivation of the gene and this was found to occur in 20% of pancreatic tumours (Hahn et al., 1995). This further reiterates the role anti-tumorigenic role of *SPARC* in pancreatic cancer and that a loss of its function drives tumorigenesis.

Additionally, Mao and colleagues hypothesized that SPARC induces the overexpression of p53 consequently resulting in cell cycle arrest at the G1/S phase (Mao et al., 2014). *SPARC* was also shown to be involved in the inhibition of angiogenesis in pancreatic cancer thereby resulting in decreased tumour growth. It carries out this function by downregulating the oncogenic VEGF (Jendraschak and Sage, 1996). However, in this study I have observed an upregulation of the VEGF signalling pathway suggesting that this might not be the case (Table 3.3).

This study has identified a frameshift variation in SPARC that may result in loss of gene function. Due to its anti-tumorigenic effect (observed in other studies), the loss of *SPARC* can be linked to the induction of carcinogenesis and this makes it a potential therapeutic target. Thus, future studies with a large sample number is crucial to validating this mutation and further linking it to carcinogenesis.

Zinc finger homeobox 3 (ZFHX3)

ZFHX3 was identified to be a suppressor of AFP (alpha- fetoprotein) by binding the AP-rich region of the gene (Zhang et al., 2005). *ZFHX3* has been observed to play anti-tumorigenic roles in different cancers. I have shown that a deletion in the exonic region of *ZFHX3* causes a frameshift mutation that may result in loss of gene function (Table 3.13). In this study, this mutation was only observed in pancreatic tumours and not in normal tissues. It is possible that a loss of gene function in *ZFHX3* might help to induce carcinogenesis.

The inactivation of *ZFHX3* has been observed in several cancers. Studies in mice have done by Sun *et al* showed that a deletion of *ZFHX3* resulted in growth of pancreatic tumours (Sun et al., 2014). They observed that deleting *ZFHX3* resulted in the dysregulation of multiple signalling pathways including Akt - which I observed to be upregulated in our study (Table 3.21). The dysregulation of the Akt pathway can enhance tumorigenesis (see Section 4.4 below)

Christopher and colleagues observed the silencing of *ZFHX3* in endometrial cancers. They observed that a loss of *ZFHX3* results in poor prognosis in endometrial cancers (Walker et al., 2015). In gastric cancer, the absence of ZFHX3 resulted in the expression of AFP (Kataoka et al., 2001). AFP is known to be involved in invasive and malignant gastric cancer. *ZFHX3* was also found to inhibit the transcription of the oncoprotein-Myb (Kaspar et al., 1999). Therefore, a loss of *ZFHX3* might enhance carcinogenesis. There is evidence that by interacting with PIAS3, *ZFHX3* induces the downregulation of the STAT3 signalling pathway (Nojiri et al., 2004). This pathway plays a role in inducing oncogenesis.

As far as can be determined, this study is the first of its kind to identify a mutation in *ZFHX3* in human pancreatic tumours. The loss of gene function resulting from this mutation might help to drive carcinogenesis. Therefore, the role of *ZFHX3* in PDAC development should be investigated further as it might have therapeutic benefits.

Transglutaminase 2 (TGM2)

In this study, it was observed that in tumours a single nucleotide variation occurring at the splice site if *TGM2* was predicted to result in a loss of gene function (Table 3.13). I propose that this loss of gene function (predicted by the IVA software) might help drive pancreatic carcinogenesis. This proposition stems from previous cancer studies indicating the pro-apoptotic function of *TGM2*. Importantly, TGM2 was observed in this study to be overexpressed, however, the identified mutation seems to have rendered it non-functional. An

overexpression in TGM2 might hinder pancreatic tumorigenesis but if the gene has lost its function, it cannot play its pro-apoptotic role.

TGM2 gene encodes a multifunctional protein (TGM2) playing several roles within and outside the cell (Mehta, 2008). It is known to play pro-apoptotic (Rodolfo et al., 2004) roles in carcinogenesis. Fok and Mehta showed that *TGM2* induced apoptosis in an AIF (apoptosis inducing factor) and caspase independent manner (Fok and Mehta, 2007). They observed that *TGM2* induced swift apoptosis in pancreatic cancer cell lines and this was linked with translocation of AIF from the mitochondria to the nucleus driving caspase dependent apoptosis. Furthermore, *TGM2* was observed to interacting with the pro-apoptotic Bax to induce apoptosis in neuroblastoma cells (Rodolfo et al., 2004). They further observed that mutations in TGM2-BH3 domain resulted in the loss of inducing ability of TGM2. These studies show that *TGM2* induces apoptosis in cancer cells, its non-functionality may result in an inhibition of apoptosis.

TGM2 is also associated with chemotherapeutic resistance in pancreatic cancer. It is highly expressed in pancreatic tumours and this upregulation is associated with invasion, metastasis, and poor clinical outcomes. Thus, it has been suggested to be a potential therapeutic target in pancreatic cancer (Mehta, 2008).

TGM2 has been observed to play pro-apoptotic roles in carcinogenesis. Its predicted loss of function in pancreatic tumours observed in this study suggests for further investigation.

Collagen type I alpha 1 (COLIA1)

I have shown that in pancreatic tumours there are variations in the *COLIA1* gene at 3 different splice sites may result in a loss of gene function (Table 3.13). *COLIA1* is a gene that encodes the protein collagen type I, alpha 1. Collagen supports and toughens multiple tissues including bones, skin, sclera, and cartilages. There have been no studies on *COLIA1* in cancer, however, its deficiency has been linked to several bone diseases. For example, polymorphisms in *COLIA1* were observed to relate to osteoporotic fractures (Alvarez-Hernandez et al., 2003). The role of COLIA1 in carcinogenesis has not been elucidated, it might prove useful to investigate its function during tumour progression.

4.4 Dysregulated pathways that induce PDAC progression

RNA sequencing showed the upregulation of 4911 genes. Pathway analysis using the PANTHER software identified significantly upregulated pathways. Similarly, differential protein expression data obtained from the Human Oncology Proteome Profiler Array identified

18 significantly upregulated proteins. Pathway analysis of these proteins also identified upregulated pathways. It is important to note that the samples used for these expression profiles where obtained from stage I PDAC patients indicating that aberrancy in the identified pathways occur at the early stages of PDAC.

Amongst the pathways identified from the analysis of the gene and protein expression profiles are those involved in the hallmarks of cancer. These six hallmarks include: avoiding growth suppressors, opposing cell death, triggering angiogenesis, sustaining proliferative signalling, aiding replicative immortality, and activating invasion and metastasis (Hanahan and Weinberg, 2000). I describe these dysregulated pathways and the role they may play in the induction and enhancement of pancreatic tumorigenesis

Insulin/IGF pathway-mitogen activated protein kinase/MAP kinase cascade

IGF (Insulin growth factor) signalling is important for tissue development and integrity. It functions in inducing cellular proliferation and inhibiting apoptosis (LeRoith and Roberts Jr, 2003). It can cascade with MAPK (Mitogen activated protein kinase) to further induce the above-mentioned biological processes. MAPK signalling is crucial to the cell as it connects external signals to the system that regulates cell growth, differentiation, proliferation and migration (Denduluri et al., 2015). Dysregulation of the IGF/MAPK signalling pathway have been observed in various cancers including pancreatic cancer (Denduluri et al., 2015; Sachdev and Yee, 2007). In this study, I observed that the IGF/MAPK signalling pathway was the most enriched by the upregulated genes (Table 3.3) suggesting its involvement during pancreatic tumour progression. A previous study observed that IGF signalling enhances cellular proliferation and growth through the Ras/MAPK pathway. In glioma stem cells, it was shown that IGF signalling activates MAPK leading to heightened cellular proliferation (LeRoith and Roberts Jr, 2003). Furthermore, MAPK signalling is known to induce the expression of downstream genes associated with cellular growth, proliferation and mitosis (Brunet et al., 1999; Furukawa et al., 2006).

One major reason for the poor survival statistics of PDAC patients is drug resistance during chemotherapeutic treatment. PDAC patients (including those used for this study) are treated with gemcitabine even though their survival rate is poor. IGF signalling has been implicated in resistance to the drug, gemcitabine, used in pancreatic cancer treatment (Tian et al., 2013). Drawing from the observations of Tian and colleagues, the upregulation observed in this study is indicative of gemcitabine resistance in PDAC patients undergoing treatment.

This study corroborates the amplification of the IGF/MAPK signalling pathway during pancreatic cancer progression observed in previous studies. Other studies have linked this amplification to the enhancement of tumourigenesis and drug resistance. Thus, I suggest that this pathway plays a crucial role during pancreatic tumorigenesis and that the discovery and utilization of inhibitors of this pathway might be crucial in finding better ways of pancreatic cancer treatment.

Integrin signalling pathway

The integrins are a group of cell adhesion regulators involved in an array of biological processes crucial for tumour initiation, advancement, and metastasis (Jones and Walker, 1999). In this study, I observed the upregulation of the integrin signalling pathway (Table 3.3) signifying its possible involvement in tumorigenesis. This upregulation was also observed in pancreatic cancer studies where it was associated with poor prognosis and chemotherapeutic resistance (Desgrosellier and Cheresh, 2010). When pancreatic cancer cells undergo chemotherapy, they induce the expression of L1CAM-an anti-apoptotic cell adhesion molecule. However, L1CAM binds α5β1 integrin-a component of the integrin signalling pathway resulting to chemotherapeutic resistance (Kiefel et al., 2010). Furthermore, the integrin signalling pathway induces the protection of pancreatic cancer cells by pancreatic stellate cells from radiotherapy-induced apoptosis (Mantoni et al., 2011) signifying it also plays a role in radiotherapeutic resistance.

The upregulation of the integrin signalling pathway in pancreatic tumours observed in this study suggests its role in PDAC progression. This role has been observed in other studies. Additionally, might be a good target for future pancreatic cancer treatment because of its role in inducing therapeutic resistance.

Ras pathway

The Ras family (H-Ras, K-Ras, and N-Ras) are proto-oncogenes known to be involved in multiple cellular processes. K-Ras is required for cellular growth control and its mutation have been identified in about 90% of pancreatic cancers (LeBleu et al., 2013; Nicke et al., 1999). Ras becomes activated when it binds to GTP (guanidine tri-phosphate) and is inactivated when ATP is hydrolysed (Logsdon and Lu, 2016). This hydrolysis is however impaired in Ras mutants. The Ras signalling pathway can be regulated by multiple external stimuli including; inflammation (Logsdon and Lu, 2016), receptor tyrosine kinases (Buday and Downward, 2008), and G-protein coupled receptors (Almendro et al., 2010). Importantly, in this study I

have observed the overexpression of one of these external stimuli- receptor tyrosine kinases (see EGFR signalling pathway). The receptor tyrosine kinases induce the Ras signalling pathway and an overexpression of these kinases suggests stimulation of the Ras signalling pathway. Furthermore, I observed the upregulation of the Ras signalling pathway in pancreatic tumours (Table 3.3) indicating they might play a role in enhancing tumorigenesis. This corroborates the study by Delire and Starkel in which they showed that the upregulation of the Ras signalling pathway, not just mutation of K-Ras, induces carcinogenesis (Delire and Starkel, 2015).

This present study observed the upregulation of the *Ras pathway* in PDAC tumours and I propose it aids in enhancing tumorigenesis. Future studies targeting components of this pathway might prove a good strategy for the inhibition pancreatic tumorigenesis.

CCKR signalling pathway

CCKR signalling has been observed to play a role in enhancing pancreatic tumour growth and development (Smith et al., 2016). CCKR (Cholecystokinin) is highly upregulated in pancreatic cancer cell lines (Smith et al., 1994). When CCKR is activated by CCK or gastrin (Smith et al., 1995, 1990), it can induce the Akt pathway which results in cellular proliferation (Dufresne et al., 2006). Interestingly, I have observed the upregulation of both the CCKR and p13k/Akt signalling pathways in pancreatic tumours (Table 3.7, 3.15). The upregulation of CCKR signalling pathway might be involved in the induction of the p13k/Akt signalling pathway which might then drive tumorigenesis.

Other studies have linked the overexpression of the CCKR signalling pathway to pancreatic tumorigenesis (Weinberg et al., 1997). In this study, I have corroborated this overexpression.

Numerous antagonist of CCKR are being developed -with promising effects- to block binding to CCK and gastrin, thereby inhibiting the CCKR signalling pathway (Berna and Jensen, 2007). These developments further show the role of the CCKR signalling pathway in enhancing tumour growth and may be a good step to improving strategies in treating PDAC.

PDGF signalling pathway

The isoforms of PDGF (platelet derived growth factor) induce cell survival, growth, and migration. In tumorigenesis, they play a role in stimulating cellular proliferation and thus its overexpression correlates with tumour invasion (Pietras et al., 2003). In this study, it was observed that the PDGF signalling pathway is upregulated in pancreatic tumour (Table 3.3)

indicating they might be involved in tumour initiation. This upregulation was also observed by Matthias and colleagues in pancreatic cancer cell lines (Longati et al., 2013). Blocking $PDGFR_b$ prevents pancreatic cancer invasion and metastasis further reiterating the importance of the PDGF signalling pathway in pancreatic cancer progression (Weissmueller et al., 2014). Furthermore, studies in breast cancer cells, have linked the PDGF signalling pathway to tumorigenesis. PDGF induces the proliferation of myofibroblast-like cells and overproduction of ECM (extracellular matrix) resulting in tumorigenesis (Shao et al., 2000).

Taken together observations from this study and other ones, the upregulation of the PDGF signalling pathway in PDAC tumours shows it might be involved in tumour advancement; therefore, targeting the components of this pathway might prove useful in therapy.

EGF receptor signalling

In this study, it was observed that the EGFR signalling pathway is overexpressed in PDAC tumours (Table 3.3) indicating it might play a role in cancer development. Epidermal growth factor receptor (EGFR) also known as ErbB1 is a transmembrane tyrosine kinase. It is activated by the binding of its ligands, TGF-α (Transform growth factor α) and EGF. EGFR is known to be upregulated in approximately 90% of pancreatic tumours (Troiani et al., 2012). Several pancreatic cancer studies have targeted the EGFR signalling pathway for chemotherapeutic purposes. When pancreatic cancer cell lines are exposed to gemcitabine, it resulted to increased phosphorylation (Morgan et al., 2008) which can be hindered by EGFR inhibitors leading to apoptosis (Bruns et al., 2000). Erlotinib is the only FDA approved EGFR inhibitor administered to manage pancreatic cancer progression, however, there is little progress in patient prognosis. Multiple mechanisms initiate this resistance including EGFR initiated activation of downstream mediators (Guix et al., 2008). In this study, I observed two of these mediators, they include: activating mutations in KRas and upregulating P13K-Akt signalling pathways- which I have observed in this study (Table 3.15).

Furthermore, the upregulation of ErbB3 -a component of the EGFR signalling pathway-was also observed in pancreatic tumours (Table 3.15) indicating it might be involved in driving tumour growth. This observation supports those of other studies in which the ErbB signalling pathway was observed to be upregulated. For example, the overexpression of ErbB2 has been observed in breast, ovarian and pancreatic cancers- where it also confers chemoresistance to cancer cells (de Graeff et al., 2008; Hynes and MacDonald, 2009; Skrypek et al., 2015). The

upregulation of ErbB3 has also been observed in pancreatic cancer cell lines (Friess, 1995) and tumours (Liles et al., 2010) where it induces tumorigenesis.

The ErbB receptor kinase family function crucially in various physiological process in normal and cancerous cells. They include EGFR, ErbB2, ErbB3 and ErbB4 (Hynes and MacDonald, 2009). The expression of members of the ErbB signalling pathway have been observed to enhance carcinogenesis. It was commonly thought that ErbB3 played a role in cellular proliferation by acting as a scaffold agent to EGFR and ErbB2 (Sithanandam and Anderson, 2008). This was mostly due to the evidence that the EGFR-ErbB3 heterodimer drives tumour progression (Engelman et al., 2005; Soler et al., 2009; Ueno et al., 2008). It is important to note that I have also observed the overexpression of EGFR indicating that it might function together with ErbB3 to induce tumorigenesis. However, Liles and colleagues showed that the overexpression ErbB3 alone is sufficient to induce cellular proliferation (Liles et al., 2010) and that its silencing decreases cell growth proliferation (Frolov et al., 2007). Furthermore, ErbB3 was identified to be involved in the P13K/Akt signalling pathway and this association further induces cellular proliferation (Jones et al., 2006).

Additionally, ErbB3 also confers radioresistance on cancer cells induced by HSP90 inhibitors- a strategy used in chemotherapy for cancers in which ErbB2 is overexpressed (Dote et al., 2005). ErbB3 is chiefly responsible for conferring cellular resistance to ErRB2 and EGFR inhibitor therapies.

The expression of components of the EGFR signalling pathway in pancreatic cancer play a role in metastasis and tumour occurrence (Tobita et al., 2003) and taken together the above observations more studies are needed to produce better EGFR inhibitors.

p13k/Akt signalling pathway

In this study, I have observed the upregulation of the p13K/Akt signalling pathway in pancreatic tumours (Table 3.15) indicating it plays a role in tumour progression. Amongst the components of this pathway I observed to be overexpressed is EGFR and ErbB3. Both proteins have been identified to induce the p13K/Akt signalling pathway (Guix et al., 2008; Jones et al., 2006). Furthermore, the CCKRs can also induce P13K/Akt signalling (Dufresne et al., 2006) and I have shown in this study the enrichment of the CCKR signalling pathway (see CCKR signalling pathway section for further description).

p13k (Phosphoinositide 3-kinases) are known to be hyperactive during tumorigenesis. Activation of p13k occur downstream GPCR (G-protein coupled receptors) and tyrosine receptor kinases (RTK). p13Ks induce mTOR (mammalian target of rapamycin) via Akt (Martini et al., 2014). The aberrant signalling of the p13K/Akt pathway is observed in majority of human tumours and is known to be responsible for drug resistance in chemotherapy. The overexpression of various components of p13k/Akt pathway has also been shown (Martini et al., 2014). Akt2- a member of the Akt kinase family- was shown to be upregulated in pancreatic cancer (Ndubaku et al., 2013) suggesting its role in the induction of carcinogenesis.

The overexpression of the p13K/Akt signalling pathway during pancreatic tumorigenesis is corroborated in this study. Its role in enhancing tumour growth and chemoresistance has also been observed in other studies. Thus, the development of antagonists for the p13K/Akt signalling pathway and its inducers is a promising area for the treatment and management for PDAC.

Angiogenesis- the role of VEGF signalling pathway

In the current study, I have observed the overexpression of the VEGF signalling pathways in pancreatic tumours indicating it might play a role in propagating pancreatic tumorigenesis (Table 3.3). Since the VEGF signalling pathway is known to regulate angiogenesis, an overexpression of this pathway leads to increased angiogenesis. This increase consequently results in tumour invasion and metastasis.

Inducing angiogenesis- the production of new blood vessels- is one of the hallmarks of cancer Tumours as in normal tissues require the supply of oxygen and adequate nutrients to survive and they require an avenue for the removal of waste products. Blood vessels formed by the process of angiogenesis help provide this avenue. During tumorigenesis, the angiogenic process is always active, consequently leading to the uncontrollable production of new blood vessels which aids in tumour invasion and metastasis (Hanahan and Folkman, 1996). Angiogenesis in tumour cells is mostly mediated by VEGF-A (Folkman, 1990). The production of VEGF-A (vascular permeability growth factor) has been shown to be involved in tumorigenesis (Ferrara, 1999) and pancreatic cancer cells secrete bioactive VEGF A (Luo, 2001). Furthermore, the expression of VEGF-A is upregulated in several carcinomas including pancreatic cancer (Itakura, 1997) and has been observed to confer radioresistance to cancer cells (Gupta, 2002).

These findings reiterate those of other studies observing the overexpression of the VEGF signalling pathway in pancreatic tumorigenesis. This pathway is known to induce angiogenesis thereby enhance tumour growth. The pro-angiogenic, pro-tumorigenic and radioresistance role of VEGF signalling pathway makes it a viable target for therapy. Indeed, anti-angiogenic therapy have been shown to decrease tumour growth and metastasis is pancreatic cancer animal models (Hotz, 2001).

4.5 Downregulation of crucial genes aid PDAC progression

Using RNA sequencing, differential gene expression analysis was performed. I observed 5 downregulated genes in pancreatic tumours but not in their corresponding normal tissues. The downregulation of genes that play anti-tumorigenic roles can occur during carcinogenesis. Below, I have described the identified genes observed to be downregulated in PDAC tumours.

CXXXC Finger protein 4 (CXXC4)

This study has observed the downregulation of *CXXC4* in pancreatic tumours (Table A10) and further propose they might play a role in inhibiting tumorigenesis. This proposition stems from work done on gastric cancer in which it was observed that *CXXC4* functions as a tumour suppressor and might be a prognostic factor for the disease (Lu et al., 2013). Kojima and colleagues also observed that in renal cell carcinoma, a downregulation of CXXC4 resulted in the induction of cellular proliferation, survival, and malignancy by the activation of the Wnt signalling pathway (Kojima et al., 1993). *CXXC4* encodes a CXXC-type zinc finger domain that inhibits the WNT signalling pathway (implicated in cancer progression) by binding to dishevelled (dvl) and consequently stabilizing the β catenin destruction complex.

Since *CXXC4* blocks the WNT pathway (which is known to drive carcinogenesis), thus, its downregulation might enable the upregulation of this pathway. Possibly, the upregulation of *CXXC4* might lead to a down regulation of Wnt pathway subsequently enabling the progression of pancreatic cancer was observed by Lu *et al* in gastric cancer and Kojima *et al* in renal cell carcinoma (Kojima et al., 1993; Lu et al., 2013).

This study has shown the downregulation of *CXXC4* in pancreatic tumours. Drawing from observations from work done on other cancers this downregulation might aid tumorigenesis. Future studies on the effect of the expression of *CXXC4* in pancreatic cancer are needed.

Polypeptide N-actylgalactosaminyltransferase 1 (GALNT1)

I observed a downregulation of *GALNT1* (Table A10 Appendix) in pancreatic tumours. Although there are no studies on *GALNT1* in pancreatic cancer cells however, some work has been done in other cancers. These studies observe the overexpression of GALNT1 and its link it to carcinogenesis. Ding and colleagues observed the overexpression of *GALNT1* in bladder cancer and suggested its role as a potential biomarker. Furthermore, silencing *GALNT1* in bladder cancer resulted in the inhibition of cellular proliferation and tumour growth (Ding et al., 2012). In hepatocellular carcinoma, knockdown of *GALNT1* inhibited malignancy via the downregulation of the EGFR signalling pathway suggesting its role as an oncogene (Huang et al., 2015). Huang and colleagues observed that the knockdown of *GALNT1* results in anti-malignancy but this is through the downregulation EGFR pathway. Importantly, I have observed the upregulation of the EGFR signalling pathway in pancreatic tumours (Table 3.7) suggesting that the downregulation of *GALNT1* might have not influenced the pathway in PDAC.

As the mechanism of action of *GALNT1* in pancreatic cancer is still unclear; future studies to underscore its roles in pancreatic cancer are required.

Heparan sulfate 3-O-sulfotransferase (HS3ST4)

In this study the downregulation of *HS3ST4* in pancreatic tumours was observed (Table A10). While no work has been done on *HS3ST4* in pancreatic cancer cells, a study in breast tumours has linked its underexpression to metastasis. Fernandez and colleagues showed that HS3ST4 was underexpressed (8-fold) in 70% metastatic breast tumours and a more than 30-fold underexpression in 90% metastasis (Fernández-Vega et al., 2015) suggesting that the downregulation of *HS3ST4* might play a role in enhancing metastasis and it might serve as a prognostic biomarker.

HS3ST4 encodes the enzyme heparin sulphate D-glucosaminyl 3-O- sulfotransferase 4. It functions as a receptor for herpes simplex virus 1 (HSV1) and thus plays an inducing role in HSV1 pathogenesis (Fernández-Vega et al., 2015). Its role in pancreatic tumorigenesis is yet to be ascertained.

The downregulation observed in breast (Fernández-Vega et al., 2015) and pancreatic tumours (Table 3.9) is indicative that *HS3ST4* might be required to prevent metastasis. Drawing from the observations of Fernandez and colleagues, it would be interesting to study the effects of overexpression of *HS3ST4* in tumours as it might anti-metastasis results.

Solute carrier family 15 member 4 (SLC15A4)

I observed the downregulation of *SLC15A4* in pancreatic tumours (Table A10). This appears to be the first study that identifies the expression of *SLC15A4* in PDAC. Although little is known about the function of SLC15A4 in carcinogenesis, it has been identified to play a role in innate immune response via toll-like receptor 9 and NOD1 (Sasawatari et al., 2011). This suggests that its downregulation in pancreatic tumours might negatively affect innate immune response during tumorigenesis. *SLCI5A4* is a member of the family of solute carriers known to be intrinsic membrane transproteins. *SLCI5A4* transports oligonucleotides from the lysosome to the cytosol (Blasius et al., 2010; Kobayashi et al., 2014). This transport is crucial for immune response against pathogens (Crotzer and Blum, 2008). The impairment of this transport by the downregulation of *SLC15A4* might result in the weakening of the immune system against cancerous cells.

The role of *SLC15A4* is not clearly defined in cancer. Other studies have linked it to the activation of the immune response. This study showed its downregulation in pancreatic tumours but not in normal tissues which might impair immune function during carcinogenesis. Future studies to investigate its role in pancreatic tumorigenesis might prove useful.

Tetraspanin (TSPAN7)

In this study, I observed the underexpression of *TSPAN7* in pancreatic tumours (Table A10) and I hypothesize it might have a role in inhibiting tumorigenesis. This hypothesis is supported by Cheong and colleagues who showed that the overexpression of *TSPAN7* in multiple myeloma patients was associated with better patient outcomes in up to 50% of patients (Cheong et al., 2015). Furthermore, in renal cell carcinoma patients the overexpression of *TSPAN7* was associated with good outcome and disease-free survival (Wuttig et al., 2012). This suggests that its overexpression might increase patient survival whereas its underexpression may decreases it.

TSPAN7 (also known as CD231, TM4SF2 and TALLA-1) belongs to a family of tetraspanins which are integral membrane proteins associated with an array of biological processes including the regulation of cell development, growth, and migration. Some cancer studies have identified the anti-metastatic role of the overexpression of *TSPAN7*. Studies in pancreatic cancer is necessary to elucidate its role in PDAC and maybe corroborate findings from other cancer studies.

4.6 Necrosis as a prognostic marker in pancreatic cancer

The RT^2 Human CancerPathway PCR Profiler array enabled us compare gene expression profiles in tumour tissues to those of the pancreatic cell line, MIA PaCa-2. Pathway analysis (using REACTOME software) of dysregulated genes show the upregulation of necrotic pathways in tumour tissues only indicating necrosis. This upregulation was not observed in the pancreatic cancer cell line- MIA PaCa-2. As this upregulation was observed in stage I tumour tissues, necrosis might be an indicator of PDAC induction. Hiraoka and colleagues observed the presence of tumour necrosis in about 62% of pancreatic tumours after histological analysis. They concluded that tumour necrosis in pancreatic tissues was a fast, simple, and efficient way of predicting postoperative out comes in pancreatic cancer patients (Hiraoka et al., 2010).

Necrosis is a type of cell death in which cells expand and burst giving off their contents to the tissue microenvironment. This process contrasts with apoptosis (programmed cell death) in which a dying cell shrinks and is devoured by neighbouring cells. Necrosis has recently been recognized as a tightly regulated process occurring under certain conditions (Galluzzi and Kroemer, 2008; Zong and Thompson, 2006). Unlike other types of programmed cell death such as autophagy and apoptosis, necrosis causes the release of proinflammatory indicators into the surrounding tissue microenvironment (Galluzzi and Kroemer, 2008; Grivennikov et al., 2010; White et al., 2010). This release consequently leads to the recruitment of the immune inflammatory cells, which act to remove necrotic debris in damaged tissue. However, during carcinogenesis these inflammatory cells can be pro-tumorigenic as they can induce cellular proliferation, angiogenesis and tumour invasion. Furthermore, necrotic cells release regulatory factors like the interleukin-1α, that can induce adjoining cells to proliferate uncontrollably(Grivennikov et al., 2010).

Taken together observations from this study and those of Hiraoka *et al.*, suggest that the presence of tumour necrosis may play a role in the induction and advancement pancreatic tumorigenesis and may be used as a prognostic marker. Furthermore, targeting tumour necrosis may prove beneficial therapeutically in the management of pancreatic cancer.

4.7 The immune system is active during PDAC

Having analysed the18 significantly upregulated proteins in tumour samples using REACTOME it appears that several immune system pathways are upregulated. This suggests that the immune system is activated during PDAC.

This present study corroborates others in which the immune system is observed to be activated during carcinogenesis. The other studies have observed that the innate and adaptive immune responses are both active having crucial and diverse effects on cancer cells (Balkwill and Mantovani, 2012). The innate immune system is primarily involved in removal of dead cells and wound healing while the adaptive one identifies and aid in the destruction of infectious foreign agents. These functions are carried out by different types of immune inflammatory cells some of which are crucial sources of growth and angiogenic factors (as described above). These factors are required and thus recruited for tumour progression (Egeblad et al., 2010; Johansson et al., 2008; Mantovani, 2010).

To function effectively against cancer cells, the immune system requires mature dendritic cells, CD8 T cells, activated macrophages, T-helper 1 cells and natural killer cells. Cancer cells have developed several mechanisms to avoid recognition by the immune system (Amedei et al., 2014). Avoiding immune recognition and destruction has been identified as an emerging hallmark of cancer (Hanahan and Weinberg, 2000). Mechanisms used by tumour cells to avoid recognition include the secretions of factors that suppress the immune system such as VEGF. Importantly, in this study I have observed the overexpression of components of the VEGF signalling pathway in pancreatic tumours indicating VEGF might be exerting immunosuppressive abilities on pancreatic cancer cells (Table 3.3). Other mechanisms include; interfering with MHC class II antigen presentation and expressing ligands of the immune system like PD-L1(Amedei et al., 2014).

In this study, the MHC II antigen presentation pathway was observed to be downregulated and suggest tumour cells may evade the immune system via this mechanism (Table 3.5). MHC class II molecules are present in only antigen presenting cells like B cells and dendritic cells which are crucial in the initiation of immune responses. The downregulation of the MHC class II antigen presentation pathway has been observed to enable tumours avoid identification and eradication by the immune system (Pandha et al., 2007).

Interestingly, the MHC II antigen presentation pathway was also enriched by upregulated proteins (Table 3.15). This enrichment is by the cathepsins- a family of proteases. Cathepsin S (a member of the cathepsin family) plays a role in tumour invasion, metastasis and in the regulation of MHC II (Beers et al., 2005; Tan et al., 2013). Thus, an upregulation of cathepsin S result to a downregulation of the MHC class II protein. This downregulation in turn may provide a way for tumour to evade the immune system as described above. The overexpression of cathepsin S in pancreatic tumours might inhibit the expression and function of MHC II as they are known regulators.

This study showed the activation of the immune response during PDAC. I observed some of the mechanisms reported in the literature as to how tumours evade the immune system. Targeting these mechanisms might prove beneficial for the development of immune-based therapeutic strategies for PDAC treatment.

CONCLUSION

PDAC is a lethal disease and early detection by identifying of diagnostic biomarkers and further understanding of the disease is critical. The present study has identified three potential biomarkers-ANGPTL4, P27 and *JUP* for PDAC. Their significant upregulation in pancreatic tumours but not in corresponding normal tissues and their abilities to be secreted into the bloodstream, may make them potential candidates for use in diagnosis tests. The implication of this is that these candidate markers may be useful in the early diagnosis of asymptomatic PDAC. Even though these biomarkers could be overexpressed in other cancers such as the upregulation of JUP in testicular germ cell tumours, the combination of 3 of them may provide increased sensitivity and specificity to PDAC. Future studies using high throughput methods like ELISA and mass spectrometry with a much larger sample amount should be done to confirm these novel potential biomarkers. Also, it would be interesting to observe the expression patterns of these potential biomarkers at different stages of PDAC as this might provide useful information of their probable in prognosis too.

This study identified novel genetic variants *SPARC*, *ZFHX3*, *TGM2* and *COLIA1* predicted to result in loss of gene function. These genes have been observed to play anti-tumorigenic roles, thus their loss of function may enable tumorigenesis. Future validation studies utilizing a larger sample size may be crucial to validating the occurrence of these variants in PDAC. It might also be important to investigate extensively the roles of these genes in pancreatic tumorigenesis.

The dysregulation of several signalling pathways such as IGF/MAPK, Integrin, and Ras signalling pathway in tumours were observed in this study. These dysregulated pathways may induce cellular proliferation, survival and migration which results in tumorigenesis. Thus, they might possess capabilities of being good targets in pancreatic cancer treatment. Future studies in investigating the dysregulation of these pathways at different stages of PDAC is pertinent n elucidating their roles.

This study has identified 5 genes downregulated in pancreatic tumours. Other studies indicate these genes may play tumour suppressive roles and that their upregulation might be crucial to inhibit tumorigenesis. It is however crucial that the downregulation of these genes be validated using a much larger sample amount. Future studies targeting these genes for therapy might prove helpful.

Furthermore, the upregulation of necrotic pathways in pancreatic tumours observed in this study might be indicative of poor prognosis. Interestingly, these pathways were observed only in pancreatic tumours but not in their corresponding normal tissues and the MIA PaCa-2 cell line. Presence or absence of necrosis in tumours might provide a way of predicting prognosis of early stage PDAC patients.

Finally, the upregulation of immune system pathways in pancreatic tumours observed in this study, suggest the activation of immune response. However, cancer cells seem to develop mechanisms to circumvent immune detection and destruction. This study identified some of these mechanisms such as the upregulation of the VEGF signalling pathway. Future studies, targeting these mechanisms may enable the immune system plays its role in the eradication of tumour cells.

APPENDICES

Table A1: Genetic biomarkers of Pancreatic cancer

Potential biomarker	**Sample(s) used**	**Discovery method(s) used**	**Indicator**	**Function(s)**	**Reference(s)**
K-ras	Tumor	Global genomic analyses	Gain-of-function mutation frequently occurs in pancreas cancer.	Performs a critical function in cell growth and signalling	(Jones et al., 2008a)
miR-210	Plasma	qRT-PCR, quantified using TaqMan MicroRNA Assays	Plasma levels are elevated in pancreatic cancer	Implicated in several cellular mechanisms	(Ho et al., 2010)
MUCI	Tissue	Tissue microarrays	Pam4 reactive MUC1 expressed in pancreatic cancer	Hydrates cell surfaces and protects against microorganisms.	(Gold et al., 2007)
CA 19-9	Serum Blood	Immunoradiometric assay, Enzyme immunoassay	Serum levels are elevated in pancreatic cancer	Functions in adhering cells to endothelium	(Tian et al., 1992) (Molina et al., 1990)
Htert	Pancreatic juice Tissue	Tissue microarray	Overexpressed in pancreatic cancer tissue samples	Provides information on the manufacture of a component of the enzyme-telomerase	(Hashimoto et al., 2008)
L1-CAM	Tissue	Tissue microarray	Overexpressed in pancreatic cancer tissue samples	Provides information on the manufacture of the L1 protein that can be found on the surfaces of neurons	(Ben et al., 2010)
B7-H3	Tissue	qRT-PCR	Significantly upregulated in pancreatic cancer	Function unclear	(Loos et al., 2009)
PALB2	Tissue	Exomic sequencing	Germline mutation	Encodes a protein that functions in genome maintenance	(Jones et al., 2009)

Table A2: Metabolomic biomarkers of pancreatic cancer

Potential biomarker	**Sample(s)used**	**Discovery method(s) used**	**Indicator**	**Function(s)**	**Reference(s)**
Dimethylamine	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels elevated in pancreatic cancer	Carcinogen	(Zhang et al., 2012)
Acetone	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels elevated in pancreatic cancer		(Zhang et al., 2012)
3-hydroxybutyrate	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer		(Zhang et al., 2012)
Lactate	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Critical source of energy	(Zhang et al., 2012)
Citrate	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Critical metabolite in the TCA cycle	(Zhang et al., 2012)
3-hydroxyisovalerate	Serum	(1)H nuclear magnetic resonance spectra	Serum levels reduced in pancreatic cancer	Suppresses the breakdown of protein, and the damage of cells after exercising	(OuYang et al., 2011)
Trimethyamine-N-oxide	Serum	(1)H nuclear MR spectra	Serum levels reduced in PC		(OuYang et al., 2011)

Table A3: Proteomic biomarkers of pancreatic cancer

Potential biomarker	**Sample(s) used**	**Method(s) used**	**Indicator**	**Function(s)**	**Reference(s)**
Vimentin	Serum, Pancreatic cancer cell line	2-D PAGE, Western blotting	Overexpressed in pancreatic cancer cell line (Panc-1)	Anchorage and positioning of organelles within the cytosol	(Hong et al., 2006)
Rad51	Tissue, Pancreatic cancer cell lines	3-D cell culture, Western blotting	Overexpressed in pancreatic cancer tissue samples	Functions in the repair of DNA double strands break	(Maacke et al., 2000)
Osteopontin	Serum	Surface-enhanced laser desorption and ionization (SELDI) protein chip mass spectrometry	Elevated levels in serum of pancreatic cancer pantients	Multiple roles in the remodelling of the bone and in the immune system	(Koopmann et al., 2004)
DEAD-box protein 48	Serum	2-D PAGE, Immunoblot analysis, liquid chromatography-electrospray ionization tandem mass spectrometry	Expressed in serum of pancreatic cancer patients	Role in RNA metabolism	(Xia et al., 2005)
Calreticulin	Serum, Pancreatic cancer cell line	2-D PAGE, Western blotting	Expressed in tissue of pancreatic cancer patients	Multiple roles including transcription regulation	(Hong et al., 2004)
Anti-thrombin-III	Pancreatic juice	2-D PAGE, MALDI-TOF-MS	Upregulated in pancreatic juice of pancreatic cancer patients.	Functions in inhibiting the coagulating of lysed thrombin and factor Xa	(Zhou et al., 2007)
Lipocalin 2	Pancreatic juice	1-D gel electrophoresis, liquid chromatography tandem mass spectrometry (LC-MS/MS).	Overexpressed in pancreatic juice of pancreatic cancer patients	Functions in innate immunity response and as a growth factor	(Gronborg et al., 2004)
PAP-2	Pancreatic juice	1-D gel electrophoresis, liquid chromatography tandem mass spectrometry (LC-MS/MS).	Overexpressed in pancreatic juice of pancreatic cancer patients	Stimulation of macrophage activity and in the modulation of the inflammatory environment in pancreatitis	(Gronborg et al., 2004)
Cytokeratin 18	Tissue, Plasma	Enzyme-linked immunoassays	Plasma levels are elevated in pancreatic cancer	Component of the intracellular cytoskeleton	(Dive et al., 2010)
Serum REG4	Serum	Enzyme-linked immunosorbent assay (ELISA).	Serum levels are elevated in pancreatic cancer	Maintains CHO recognition activity even in an acidic condition	(Takayama et al., 2010)

Serine proteinase-2(PRSS2) pre protein	Pancreatic juice	2-D PAGE, Mass spectrometry, qRT-PCR	Upregulated in pancreatic juice in pancreatic cancer	Protein precursor of PRSS2 PRSS2 is involved in cleaving peptide bonds where serine functions as the nucleophillic amino acid at the active site of the enzyme	(Gao et al., 2010)
Pancreatic lipase related protein-1 (PLRP1)	Pancreatic juice	2-D PAGE, Mass spectrometry, qRT-PCR	Upregulated in pancreatic juice in pancreatic cancer	Function unclear	(Gao et al., 2010)
Mucins- MUC2, MUC4, MUC5B,MUC5C	Tissue, Pancreatic cancer cell line	Immunohistochemistry, *in situ* hybridization	Blood levels elevated	Formation of chemical barriers, lubrication and involved in cell signalling	(Swartz et al., 2002)
S100P	Tissue, Pancreatic cancer cell line	Far Western screening and in vitro interaction assays	Overexpression in pancreatic cancer	Calcium binding protein	(Dowen et al., 2005)
CEA	Blood	Radioimmunoassays	Overexpression in pancreatic cancer	Involved in the adhesion of cells	(Molina et al., 1990)
CA-50	Blood	Time-resolved flourescence immunoassay, immunoradiometric assay	Overexpression in pancreatic cancer		(Molina et al., 1990) (Haglund et al., 1987)
TROP2	Tissue	Immunohistochemistry	Overexpression in tissue samples of pancreatic cancer patients	Transmembrane protein- function unclear	(Fong et al., 2008)
CDCP1	Tissue, Pancreatic cancer cell line	Western blotting, Cell migration and Matrigel invasion assay, ECM degradation assays, Immunocytochemical staining	Overexpressed in pancreatic cancer	regulates anoikis resistance, Cell migration and the degradation of ECM	(Miyazawa et al., 2010)
N-acetyl glycoprotein	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels elevated in pancreatic cancer		(Zhang et al., 2012)
VLDL	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels elevated in pancreatic cancer	Body's transport mechanism for lipids	(Zhang et al., 2012)

HDL	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Protection against heart disease by reverse transporting cholesterol	(Zhang et al., 2012)
LDL	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Transport of cholesterol to the cells	(Zhang et al., 2012)
Alanine	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Helps in protein production, neurotransmitters formation and the proper functioning of the nervous system	(Zhang et al., 2012)
Glutamate	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Involved in normal brain functioning and in the CNS	(Zhang et al., 2012)
Glutamine	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Source of cellular energy, production of proteins, regulating acid-base balance	(Zhang et al., 2012)
Histidine	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Transport of metals and enhances calcium absorption	(Zhang et al., 2012)
Isoleucine	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Source of energy	(Zhang et al., 2012)
Lysine	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Helps in bone formation, collagen formation, calcium absorption in the intestine	(Zhang et al., 2012)
Valine	Plasma	Pattern recognition techniques applied to (1)H nuclear magnetic resonance spectra	Plasma levels reduced in pancreatic cancer	Prevents muscle breakdown, helps in the functioning of the nervous system	(Zhang et al., 2012)

Table A4: Circulating Endothelial Cells as biomarkers of Pancreatic cancer

Potential biomarker	**Sample(s) used**	**Discovery method(s) used**	**Indicator**	**Function(s)**	**References**
Circulating endothelial cells	Blood	Flow cytometry, Immunohistochemistry, ELISA	Blood levels increased in pancreatic cancer	Endothelial cells maintain the strength of the vascular membrane and its permeability	(Sabbaghian et al., 2010)

Table A5: Salivary Transcriptomic Markers of Pancreatic cancer

Potential biomarker	**Sample(s) used**	**Discovery method(s) used**	**Indicator**	**Function(s)**	**References**
Salivary transcriptomic biomarkers (MBD3L2, ACRV1, DMPM1, KRAS)	Saliva	Affymetrix array, qPCR	Expressed in pancreatic cancer	Varying	(Zhang et al., 2010)

Test Performed	Normal	Tumour	MIPACA 2
PCR Array Reproducibility			
Average C_T (PPC)	22.45	22.49	23.39
Result	Inquiry	Inquiry	Inquiry
Reverse Transcription Control (RTC)			
Delta C_T (Average RTC - Average PPC)	1.04	2.46	1.52
Result	Pass	Pass	Pass
Genomic DNA Contamination (GDC)			
C_T(GDC)	35.00	35.00	35.00
Result	Pass	Pass	Pass

Figure A1: Samples passed all tests (Reverse transcription control, Genomic DNA contamination and PCR array reproducibility). The Inquiry indicated on the PCR array reproducibility test was due to the average C_T of all samples exceeding 22. However, due to differences in sensitivity of real time machines this can be acceptable provided values did not vary more than 2 cycles between the arrays which was the case in this experiment.

Groups	Samples	B2M	Geometric Mean	Average Geometric Mean
Control Group	Normal	23.5	23.50	23.50
Group 1	Tumour	22.83	22.83	22.83
Group 2	MIPACA 2	22.13	22.13	22.13

Figure A2: The software automatically did normalization. The software chooses the most stable housekeeping gene. The C_T values were geometrically averaged and used for normalization. B2M was the most stable.

Layout	01	02	03	04	05	06	07	08	09	10	11	12
A	ACLY -359.54 A	ACSL4 -4.44	ADM 1.30	ANGPT1 -2.89 A	ANGPT2 -1.03 A	APAF1 1.46	ARNT -24.93	ATP5A1 -1.27	AURKA -1.52	BCL2L11 1.68	BIRC3 1.23	BMI1 1.51
B	CA9 -1.45	CASP2 -1.11	CASP7 1.53	CASP9 -272.48 A	CCL2 -1.72	CCND2 1.21	CCND3 1.71	CDC20 1.22	CDH2 -1.16	CFLAR 3.36	COX5A 2.48	CPT2 2.68
C	DDB2 -1.26	DDIT3 1.48	DKC1 -1.11	DSP 1.74	E2F4 -1.13	EPO -2.19	ERCC3 6.41 A	ERCC5 1.11	ETS2 2.13	FASLG -1.35	FGF2 -1.39	FLT1 1.09
D	FOXC2 -1.20	G6PD -1.88	GADD45G -2.93 A	GPD2 1.40	GSC -3.68	HMOX1 -1.39	IGFBP3 5.98	IGFBP5 2.16	IGFBP7 2.33	KDR 2.38	KRT14 -2.83	LDHA 2.31
E	LIG4 -3.32	LPL 1.16	MAP2K1 -1.34	MAP2K3 -1.52	MAPK14 1.64	MCM2 2.64 A	MKI67 1.32	NOL3 -1.22	OCLN 2.57	PFKL -1.09	PGF -5.28 A	PINK1 -1.61
F	POLB 2.68	PPP1R15A -2.62	SERPINB2 1.91	SERPINF1 -2.39	SKP2 1.64	SLC2A1 2.33	SNAI1 -1.91	SNAI2 6.87	SNAI3 -3.25	SOD1 2.27	SOX10 -1.83	STMN1 1.87
G	TBX2 -1.88	TEK 1.26	TEP1 1.65	TERF1 1.61	TERF2IP 1.20	TINF2 2.41	TNKS 1.74	TNKS2 1.79	UQCRFS1 1.84	VEGFC -5.28	WEE1 2.33	XIAP 2.20

a

Layout	01	02	03	04	05	06	07	08	09	10	11	12
A	ACLY 1.28	ACSL4 -1.66	ADM 3.71	ANGPT1 -97.68 A	ANGPT2 -7.11	APAF1 1.34	ARNT 4.41 A	ATP5A1 8.46	AURKA 5.74	BCL2L11 1.19	BIRC3 3.46	BMI1 -2.89
B	CA9 -1.07	CASP2 -1.25	CASP7 -1.33	CASP9 1.32	CCL2 -73.52	CCND2 -81.57 A	CCND3 -7.78	CDC20 9.58	CDH2 -21.41	CFLAR 1.40	COX5A 8.22	CPT2 9.71 A
C	DDB2 3.56	DDIT3 3.56	DKC1 4.29	DSP -2.58	E2F4 1.18	EPO -1.82	ERCC3 7.73 A	ERCC5 1.25	ETS2 -5.39	FASLG -6.15	FGF2 11.00 A	FLT1 -278.20 A
D	FOXC2 -9.32	G6PD 1.26 A	GADD45G -29.65 A	GPD2 1.56	GSC -3.73	HMOX1 2.03	IGFBP3 -2646.74 A	IGFBP5 -10015.87 A	IGFBP7 -1488.87	KDR -88.03 A	KRT14 -21.26	LDHA 16.22
E	LIG4 -382.68 A	LPL -151.17 A	MAP2K1 1.56	MAP2K3 -1.23	MAPK14 1.78	MCM2 20.39 A	MKI67 6.02	NOL3 4.76	OCLN 5.43	PFKL 2.36	PGF -54.19 A	PINK1 6.92 A
F	POLB 2.83	PPP1R15A -2.43	SERPINB2 -23.75	SERPINF1 -3.27	SKP2 6.11	SLC2A1 1.64	SNAI1 -4.41	SNAI2 -306.55 A	SNAI3 -5.82	SOD1 6.50	SOX10 -16.56	STMN1 6.54
G	TBX2 -200.85 A	TEK -113.77 A	TEP1 -1.01	TERF1 -30.48	TERF2IP 2.25	TINF2 -1.18	TNKS 2.16	TNKS2 1.53	UQCRFS1 6.32	VEGFC 8.51 A	WEE1 2.77	XIAP -4.26

b

Figure A3: Array plate layout showing fold changes of Group 1(a) and Group 2(b) genes. Genes marked with “A” indicate that the gene is observed to relatively low in one sample but relatively high in the other.

Table A6: List of genes uniquely upregulated and downregulated in Group 1 and 2

Upregulated		**Down-regulated**	
Group 1	**Group 2**	**Group 1**	**Group 2**
SNAI2	*FGF2*	*ACLY*	*IGFBIP5*
IGFBP3	*CDC20*	*CASPD*	*IGFBIP3*
CFLAR	*HPRT1*	*ARNT*	*IGFBIP7*
TINF2	*VEGFC*	*VEGFC*	*SNAI2*
KDR	*ATPSA1*	*ACSL4*	*FLT1*
SLC2A1	*PINX1*	*EPO*	*TBX2*
IGFBP7	*STMN1*		*LPL*
XIAP	*UQCRFS1*		*TEK*
IGFBP5	*SKP2*		*KDR*
ETS2	*MK167*		*CCND2*
	RPLPO		*CCL2*
	AURKA		*TERF1*
	NOL3		*SERPINB2*
	ARNT		*CDH2*
	DKC1		*SOX10*
	ADM		*FOXC2*
	DDIT3		*CCND3*
	DDB2		*ANGPT2*
	BIRC3		*FASLG*
	PFKL		*ETS2*
	TERF21P		*SNAI1*
	TNKS		*XIAP*
	HMOX1		*BMI1*
			DSP
			HGDC

Table A7: Showing top 10 molecular functions enriched by upregulated genes.

GO ID	Molecular function	Homo sapiens (Ref)	Client input	Expected	Fold enrichment	p-value
GO:0048365	Rac GTPase binding	40	31	9.26	3.35	3.69E-05
GO:0017048	Rho GTPase binding	80	50	18.52	2.7	2.62E-06
GO:0017016	Ras GTPase binding	254	121	58.8	2.06	1.48E-09
GO:0031267	Small GTPase binding	272	131	62.97	2.08	7.78E-09
GO:0051020	GTPase binding	295	143	68.29	2.09	3.20E-12
GO:0019899	Enzyme binding	1750	762	405.12	1.88	2.71E-59
GO:0005515	Protein binding	1075	3422	2488.85	1.37	4.63E-160
GO:0005488	Binding	14268	4057	3303.03	1.23	1.27E-130
GO:0050840	Extracellular matrix binding	53	36	12.27	2.93	7.23E-05
GO:0005518	Collagen binding	65	40	15.05	2.66	1.71E-04

Table A8: Showing top 10 cellular components enriched by upregulated genes.

GO ID	Cellular component	Homo sapiens (Ref)	Client input	Expected	Fold enrichment	p-value
GO:0005840	Ribosome	119	54	27.55	1.96	3.01E-04
GO:1990904	Ribonucleoprotein complex	335	110	77.55	1.42	1.61E-02
GO:0032991	Macromolecular complex	1454	415	336.6	1.23	5.98E-04
GO:0031012	Extracellular matrix	185	73	42.83	1.7	9.29E-04
GO:0005794	Golgi apparatus	130	50	30.09	1.66	3.19E-02
GO:0043226	Organelle	2656	765	614.86	1.24	1.16E-08
GO:0005829	Cytosol	328	126	75.93	1.66	4.30E-06
GO:0005737	Cytoplasm	2088	664	483.37	1.37	5.75E-15
GO:0005622	Intracellular	4020	1204	930.63	1.29	1.86E-20
GO:0044464	Cell part	4115	1222	952.62	1.28	1.42E-19

Table A9: Showing top 10 protein classes enriched by upregulated genes.

Protein class	**Homo sapiens (Ref)**	**Client input**	**Expected**	**Fold enrichment**	**p-value**
Nucleic acid binding	2367	2706	547.96	1.29	5.83E-10
Enzyme modulator	1353	449	313.22	1.43	8.68E-12
Transferase	1164	342	269.47	1.27	1.35E-03
RNA binding protein	887	308	205.34	1.50	9.59E-10
Cytoskeletal protein	778	248	180.11	1.38	1.17E-04
G-protein modulator	481	201	111.35	1.81	1.27E
Actin-family cytoskeletal protein	389	152	90.05	1.69	2.31E-07
Ribosomal protein	228	90	52.78	1.71	3.45E-04
Non-motor actin binding protein	179	85	41.44	2.05	3.40E-07
Guanyl-nucleotide exchange factor	179	79	41.44	1.91	2.38E-05

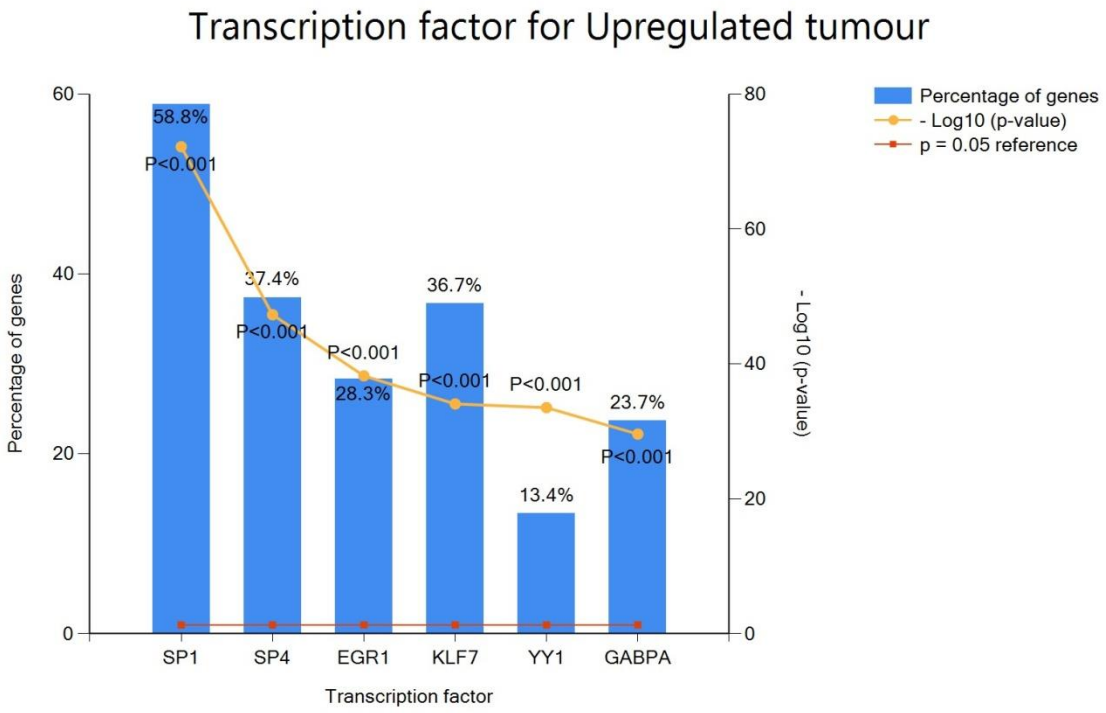


Figure A4: Bar chart showing top transcription factors in which our upregulated genes are enriched. Percentages of genes enriched are shown on top of each bar. Yellow dots indicate p-value of less than 0.001 and red indicates p-values of 0.05.

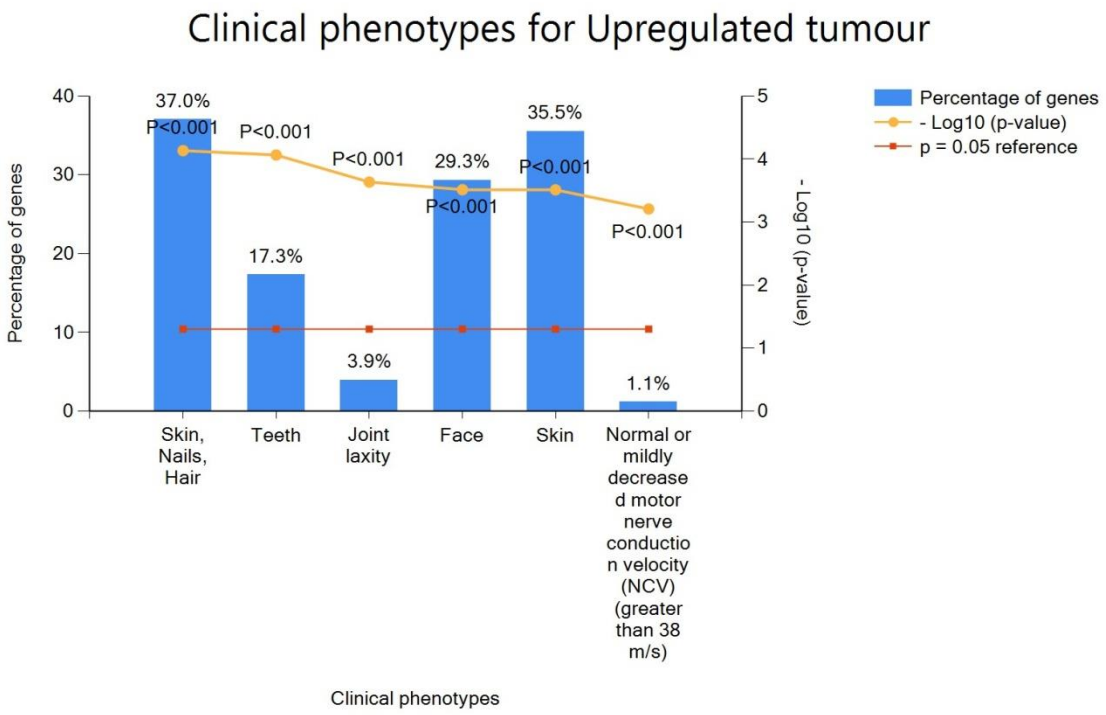


Figure A5: Bar chart showing top clinical phenotypes in which our upregulated genes are enriched. Percentages of genes enriched are shown on top of each bar. Yellow dots indicate p-value of less than 0.001 and red indicates p-values of 0.05.

Table A10: Showing functional enrichment analysis of downregulated genes.

CXXC4	**CXXC finger 4**
Biological process	embryonic axis specification, regionalization, cell surface receptor linked signal transduction, zygotic determination of dorsal/ventral axis, pattern specification process, axis specification, embryonic pattern specification, dorsal/ventral axis specification, dorsal/ventral pattern formation, negative regulation of signal transduction, negative regulation of cell communication, Wnt receptor signalling pathway, regulation of Wnt receptor signalling pathway, negative regulation of Wnt receptor signalling pathway,
Cellular component	cytoplasmic membrane-bounded vesicle, cytoplasmic vesicle, vesicle, membrane-bounded vesicle,
Molecular function	DNA binding, zinc ion binding, protein domain specific binding, PDZ domain binding, ion binding, cation binding, metal ion binding, transition metal ion binding,
Protein class	Zinc finger, CXXC-type,
Pathway	Wnt signalling pathway,
GALNT1	**UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 13 (GalNAc-T13); UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 1 (GalNAc-T1)**
Biological process	protein amino acid glycosylation, protein amino acid O-linked glycosylation, glycoprotein metabolic process, glycoprotein biosynthetic process, peptidyl-serine modification, peptidyl-threonine modification, protein amino acid O-linked glycosylation via serine, protein amino acid O-linked glycosylation via threonine, biopolymer glycosylation, glycosylation,
Cellular component	extracellular region, Golgi apparatus, integral to membrane, intrinsic to membrane, perinuclear region of cytoplasm,
Molecular function	polypeptide N-acetylgalactosaminyltransferase activity, calcium ion binding, sugar binding, acetylgalactosaminyltransferase activity, manganese ion binding, carbohydrate binding, ion binding, cation binding, metal ion binding, transition metal ion binding,
Protein class	Ricin B lectin, Glycosyl transferase, family 2
Pathway	O-Glycan biosynthesis
HS3ST4	**Heparan sulphate (glucosamine) 3-O-sulfotransferase 4**
Biological process	proteoglycan metabolic process, sulfur metabolic process, glycoprotein metabolic process, heparan sulphate proteoglycan metabolic process,
Cellular component	Golgi apparatus, integral to membrane, intrinsic to membrane,
Molecular function	sulfotransferase activity, [heparan sulphate]-glucosamine 3-sulfotransferase 1 activity, transferase activity, transferring sulfur-containing groups, heparan sulphate sulfotransferase activity,
Protein class	Sulfotransferase
SLC15A4	**solute carrier family 15, member 4**

Biological process	oligopeptide transport, protein localization, protein transport, peptide transport, establishment of protein localization,
Cellular component	lytic vacuole, lysosome, lysosomal membrane, vacuole, vacuolar membrane, integral to membrane, organelle membrane, intrinsic to membrane, vacuolar part,
Molecular function	peptide transporter activity, symporter activity, solute:cation symporter activity, solute:hydrogen symporter activity, peptide:hydrogen symporter activity, proton-dependent peptide secondary active transmembrane transporter activity
Protein class	TGF-beta receptor, type I/II extracellular region, TGF-beta receptor, type I/II extracellular region.
TSPAN7	**Tetraspanin 7**
Biological process	protein amino acid glycosylation, protein amino acid N-linked glycosylation, glycoprotein metabolic process, glycoprotein biosynthetic process, biopolymer glycosylation, glycosylation,
Cellular component	plasma membrane, integral to plasma membrane, integral to membrane, intrinsic to membrane, intrinsic to plasma membrane, plasma membrane part,
Protein class	Tetraspanin, subgroup, Tetraspanin, Tetraspanin, conserved site,

Table A11: Top 5 molecular functions enriched in JUP PPI network

Pathway ID	**Pathway description**	**Count in gene set**	**False discovery rate**
GO:0008013	Beta-catenin binding	8	7.87E-16
GO:0045295	Gamma-catenin binding	4	3.06E-08
GO:0005515	Protein binding	10	1.94E-03
GO:0086083	Cell adhesive protein binding involved in bundle of His cell-Purkinje myocyte communication	2	2.7E-03
GO:0045294	Alpha-catenin binding	2	5.55E-03

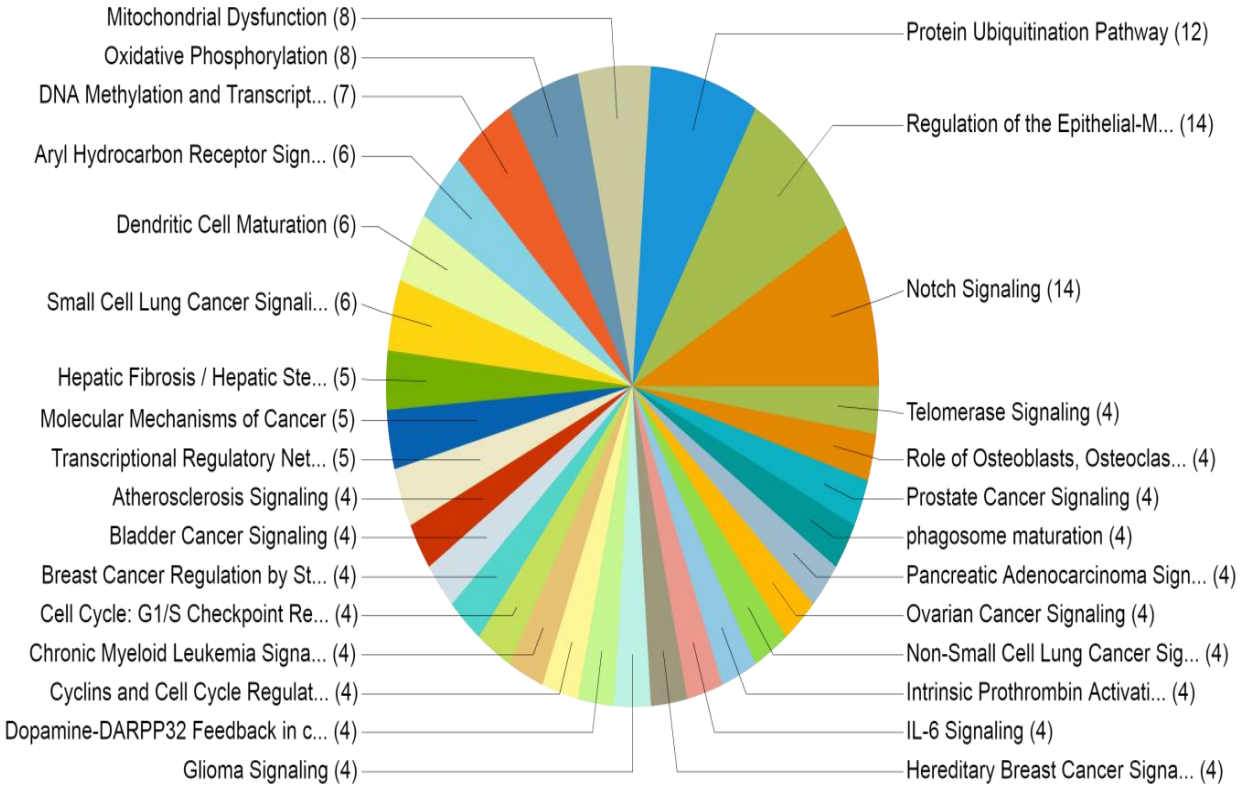


Figure A6: Pie chart showing pathways associated with the observed genetic variants. The number of variants enriched in each pathway is seen in parenthesis.

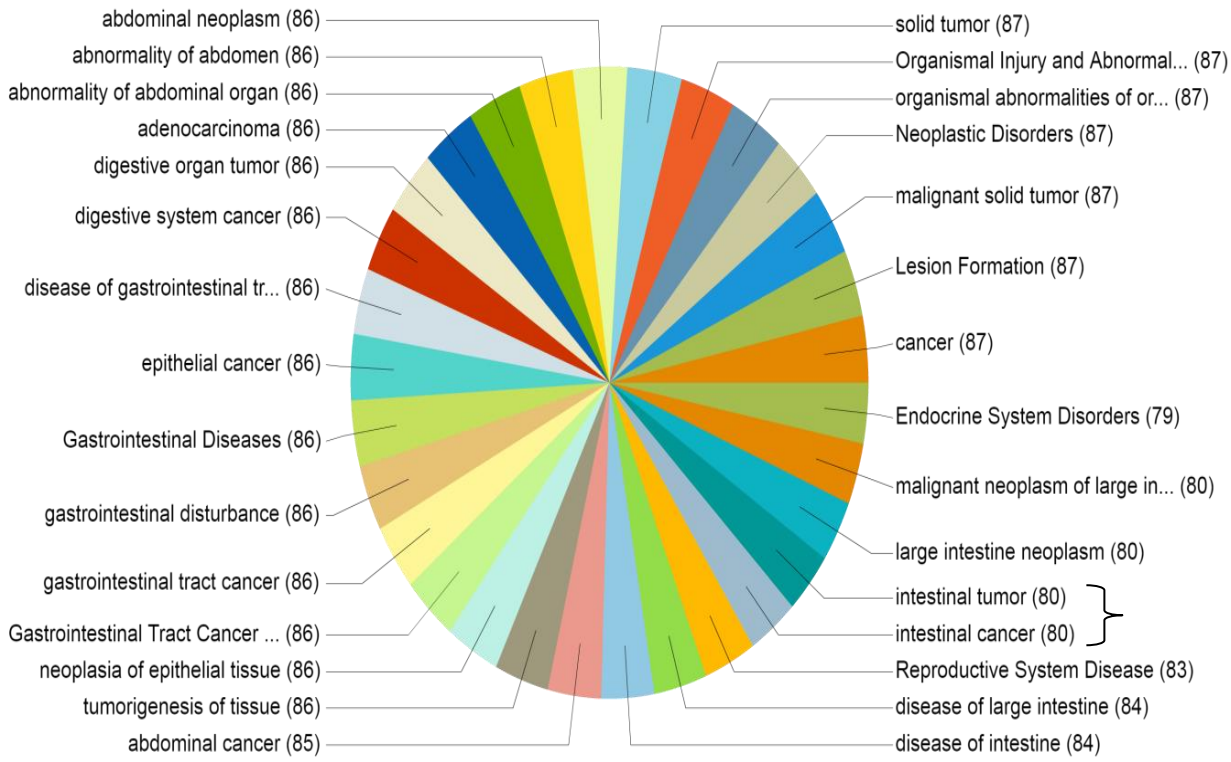


Figure A7: Pie chart showing diseases associated with the observed genetic variants. The number of variants enriched in each disease is seen in parentheses.

Table A12: Top 5 molecular functions in which the ANGPTL4 PP1 network is enriched.

Pathway ID	**Pathway description**	**Count in gene set**	**False discovery rate**
GO: 0000989	Transcription factor activity	8	8.09E-08
GO: 0008134	Transcription factor binding	7	3.48E-07
GO: 0003712	Transcription cofactor activity	7	1.47E-06
GO: 0035257	Nuclear hormone receptor binding	5	1.47E-06
GO: 0033613	Activating transcription factor binding	4	6.05E-06

Table A13: Top 5 molecular functions in which the CDKN1B/p27 PP1 network is enriched.

Pathway ID	**Pathway description**	**Count in gene set**	**False discovery rate**
GO: 0016538	Cyclin-dependent protein serine/threonine kinase regulator activity	4	1.11E-06
GO: 0004693	Cyclin-dependent protein serine/threonine kinase activity	4	2.21E-06
GO: 0016301	Kinase activity	7	1.69E-05
GO: 0030332	Cyclin binding	3	4.42E-05
GO: 0004672	Protein kinase activity	6	8.34E-05

Table A14: Showing patients used in study with relevant information

Patient	**Patient number**	**Age**	**Gender**	**Smoked**	**Alcohol**	**Crp levels**	**Stage**
1	CHMBAD000009	62	Female	No	No	----	1
2	CHMBAD000031	69	Male	Yes	Yes	----	1
3	CHMBAD000066	41	Male	Yes	Yes	12	1
4	CHMBAD000074	60	Male	Yes	Yes	279	1
5	CHMBAD000091	62	Male	Yes	Yes	----	1
6	CHMBAD000096	52	Female	No	Yes	117	1
7	CHMBAD000097	61	Female	No	No	23	1
8	CHMBAD000118	67	Male	Yes	No	31	1
9	CHMBAD000120	56	Male	Yes	Yes	116	1

Table A15: List of kits used with their manufacturers and catalogue numbers.

Kit	**Manufacturer and Catalogue number**
AllPrep DNA/RNA mini kit	Qiagen 80201
DNase	Sigma-AMPDL IKT
Bradford protein assay	Bio-Rad 5000006
RT^2 First strand kit	Qiagen 330401
2X PCR Mastermix	NEB M0207L
Human Cancer PathwayFinder RT^2 Profiler PCR array	Qiagen 330231 PAHS-033Z
SYBR green mix	Qiagen 7540246
Proteome profiler Human Oncology array kit	RND systems ARY026
JUP Primer assay	Qiagen 330001 PPH07138F
COPG Primer assay	Qiagen 330001 PPH07138F
TCIRG1 Primer assay	Qiagen 330001 PPH09294A
MRPL19 Primer assay	Qiagen 330001 PPH09294A

Table A16: List of Instruments used with their manufacturers and catalogue numbers.

Instrument	**Manufacturer and Catalogue number**
TissueRuptor	Qiagen 9001272
Nanodrop	ND-1000 UV
Electrophoresis tank	Scieplus VG Sys 30020529
Power supply	Sigma techware PS250-2
Spectrophotometer	Shimadza UV-160A
Chemi-Doc	Bio-Rad Gel Doc XR^{+}
8-well Multi pipette	Integra 4624
Real-time cycler machine	Roche LC-480
Room temperature centrifuge	Heraeus T5003280
4°C centrifuge	Eppendorf 22331
96-well plate centrifuge	Grant-bio LMC-3000
96-well plate	4titude-4ti-0955
PCR machine	Geneamp PCR system 2400

Table A17: List of software used with the link to their websites.

Software used	**Website**
Wvision Quickspots	www.wvision.com/Quickspots.html
Uniprot	www.uniprot.com
FunRich	www.funrich.org
Panther	www.pantherdb.org
GeneMania	www.genemania.com
Gene Ontology	Geneontology.org
DAVID	DAVId.ncifcrf.gov
Qiagen RT^2 Profiler PCR array software	pcrdataanalysis.sabiosciences.com
Reactome	www.reactome.org
Ingenuity Pathway Analysis (IPA)-Biomarker filtering	www.ingenuity.com/ipa
Ingenuity Variant Analysis (IVA)	www.qiagenbioinformatics.com/products/ingenuity-variant-analysis/
Morpheus	https://software.broadinstitute.org/morpheus/
CLC genomics workbench	www.qiagenbioinformatics.com/products/clc-genomics-workbench
REST software	http://www.wzw.tum.de/gene-quantification/
STRING analysis tool	http://string-db.org/
Pancreatic cancer database	http://pancreaticcancerdatabase.org/publications.php
NCBI	https://www.ncbi.nlm.nih.gov/pubmed
MEDICALC	https://www.medcalc.org/calc/diagnostic_test.php

Table A18: List of Reagents/Chemicals used with their manufacturers and catalogue numbers.

Reagent/Chemical	**Manufacturer and Catalogue number**
RNA*later*	Qiagen 76106
Igepal	Sigma l3021
Sodium chloride	Ace 29249
Sodium dodecyl sulphate	VWR Chemicals 442444H
Sodium deoxy cholate	BDH 43035
Aprotinin	Sigma A6279
PMSF	Sigma P7626
Trizma base	Sigma T1503-500G
FBS	Biowest S1815-500
Penicillin-Streptomycin	Sigma P433-20ml
DMEM	Sigma D5796-500ml
Trypsin-EDTA	Sigma T3924-10ml
BSA	Armour laboratories CH3170
RNase-free water	Qiagen 129317
Formaldehyde	Sigma F8775-50
Ethidium bromide	Sigma E1510
DNA Molecular weight marker	Thermo scientific SM1333
RNA loading dye	Thermo scientific
6X DNA loading dye	Thermo scientific R0611
Agarose	Lonza 50004
TBE	Sigma T4415
MOPS	Sigma 69947-100G

REFERENCES

Almendro, V., Garcia-Recio, S., Gascon, P., 2010. Tyrosine kinase receptor transactivation associated to G protein-coupled receptors. Curr. Drug Targets 11, 1169–1180.

Alvarez-Hernandez, D., Naves, M., Diaz-Lopez, J.B., Gomez, C., Santamaria, I., Cannata-Andia, J.B., 2003. Influence of polymorphisms in VDR and COLIA1 genes on the risk of osteoporotic fractures in aged men. Kidney Int. Suppl. S14-18. doi:10.1046/j.1523-1755.63.s85.5.x

Amedei, A., Niccolai, E., Prisco, D., 2014. Pancreatic cancer: role of the immune system in cancer progression and vaccine-based immunotherapy. Hum. Vaccines Immunother. 10, 3354–3368. doi:10.4161/hv.34392

Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D.J., Honjo, T., Hrabe de Angelis, M., Lendahl, U., Edlund, H., 1999. Notch signalling controls pancreatic cell differentiation. Nature 400, 877–881. doi:10.1038/23716

Argani, P., Iacobuzio-Donahue, C., Ryu, B., Rosty, C., Goggins, M., Wilentz, R.E., Murugesan, S.R., Leach, S.D., Jaffee, E., Yeo, C.J., Cameron, J.L., Kern, S.E., Hruban, R.H., 2001. Mesothelin is overexpressed in the vast majority of ductal adenocarcinomas of the pancreas: identification of a new pancreatic cancer marker by serial analysis of gene expression (SAGE). Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 7, 3862–3868.

Baak, J.P.A., Path, F.R.C., Hermsen, M.A.J.A., Meijer, G., Schmidt, J., Janssen, E.A.M., 2003. Genomics and proteomics in cancer. Eur. J. Cancer Oxf. Engl. 1990 39, 1199–1215.

Balkwill, F.R., Mantovani, A., 2012. Cancer-related inflammation: common themes and therapeutic opportunities. Semin. Cancer Biol. 22, 33–40. doi:10.1016/j.semcancer.2011.12.005

Beers, C., Burich, A., Kleijmeer, M.J., Griffith, J.M., Wong, P., Rudensky, A.Y., 2005. Cathepsin S controls MHC class II-mediated antigen presentation by epithelial cells in vivo. J. Immunol. Baltim. Md 1950 174, 1205–1212.

Ben, Q.-W., Wang, J.-C., Liu, J., Zhu, Y., Yuan, F., Yao, W.-Y., Yuan, Y.-Z., 2010. Positive expression of L1-CAM is associated with perineural invasion and poor outcome in pancreatic ductal adenocarcinoma. Ann. Surg. Oncol. 17, 2213–2221. doi:10.1245/s10434-010-0955-x

Berdasco, M., Esteller, M., 2010. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. Dev. Cell 19, 698–711. doi:10.1016/j.devcel.2010.10.005

Berna, M.J., Jensen, R.T., 2007. Role of CCK/gastrin receptors in gastrointestinal/metabolic diseases and results of human studies using gastrin/CCK receptor agonists/antagonists in these diseases. Curr. Top. Med. Chem. 7, 1211–1231.

Blasius, A.L., Arnold, C.N., Georgel, P., Rutschmann, S., Xia, Y., Lin, P., Ross, C., Li, X., Smart, N.G., Beutler, B., 2010. Slc15a4, AP-3, and Hermansky-Pudlak syndrome proteins are required for Toll-like receptor signaling in plasmacytoid dendritic cells. Proc. Natl. Acad. Sci. U. S. A. 107, 19973–19978. doi:10.1073/pnas.1014051107

Bracci, P.M., 2012. Obesity and pancreatic cancer: overview of epidemiologic evidence and biologic mechanisms. Mol. Carcinog. 51. doi:10.1002/mc.20778

Brunet, A., Roux, D., Lenormand, P., Dowd, S., Keyse, S., Pouyssegur, J., 1999. Nuclear translocation of p42/p44 mitogen-activated protein kinase is required for growth

factor-induced gene expression and cell cycle entry. EMBO J. 18, 664–674. doi:10.1093/emboj/18.3.664

Bruns, C.J., Solorzano, C.C., Harbison, M.T., Ozawa, S., Tsan, R., Fan, D., Abbruzzese, J., Traxler, P., Buchdunger, E., Radinsky, R., Fidler, I.J., 2000. Blockade of the epidermal growth factor receptor signaling by a novel tyrosine kinase inhibitor leads to apoptosis of endothelial cells and therapy of human pancreatic carcinoma. Cancer Res. 60, 2926–2935.

Buchholz, M., Braun, M., Heidenblut, A., Kestler, H.A., Kloppel, G., Schmiegel, W., Hahn, S.A., Luttges, J., Gress, T.M., 2005. Transcriptome analysis of microdissected pancreatic intraepithelial neoplastic lesions. Oncogene 24, 6626–6636. doi:10.1038/sj.onc.1208804

Buday, L., Downward, J., 2008. Many faces of Ras activation. Biochim. Biophys. Acta 1786, 178–187. doi:10.1016/j.bbcan.2008.05.001

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. Clin. Chem. 55, 611–622. doi:10.1373/clinchem.2008.112797

Campagna, D., Cope, L., Lakkur, S.S., Henderson, C., Laheru, D., Iacobuzio-Donahue, C.A., 2008. Gene Expression Profiles Associated with Advanced Pancreatic Cancer. Int. J. Clin. Exp. Pathol. 1, 32–43.

Carriere, C., Mirocha, S., Deharvengt, S., Gunn, J.R., Korc, M., 2011. Aberrant expressions of AP-2alpha splice variants in pancreatic cancer. Pancreas 40, 695–700. doi:10.1097/MPA.0b013e31821f2715

Carrigan, P.E., Bingham, J.L., Srinvasan, S., Brentnall, T.A., Miller, L.J., 2011. Characterization of alternative spliceoforms and the RNA splicing machinery in pancreatic cancer. Pancreas 40, 281–288. doi:10.1097/MPA.0b013e31820128d2

Chan, A., Diamandis, E.P., Blasutig, I.M., 2013. Strategies for discovering novel pancreatic cancer biomarkers. J. Proteomics 81, 126–134. doi:10.1016/j.jprot.2012.09.025

Chappuis, P.O., Kapusta, L., Begin, L.R., Wong, N., Brunet, J.S., Narod, S.A., Slingerland, J., Foulkes, W.D., 2000. Germline BRCA1/2 mutations and p27(Kip1) protein levels independently predict outcome after breast cancer. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 18, 4045–4052. doi:10.1200/jco.2000.18.24.4045

Chen, R., Pan, S., Brentnall, T.A., Aebersold, R., 2005a. Proteomic profiling of pancreatic cancer for biomarker discovery. Mol. Cell. Proteomics MCP 4, 523–533. doi:10.1074/mcp.R500004-MCP200

Chen, R., Yi, E.C., Donohoe, S., Pan, S., Eng, J., Cooke, K., Crispin, D.A., Lane, Z., Goodlett, D.R., Bronner, M.P., Aebersold, R., Brentnall, T.A., 2005b. Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape. Gastroenterology 129, 1187–1197. doi:10.1053/j.gastro.2005.08.001

Cheng, M., Sexl, V., Sherr, C.J., Roussel, M.F., 1998. Assembly of cyclin D-dependent kinase and titration of p27Kip1 regulated by mitogen-activated protein kinase kinase (MEK1). Proc. Natl. Acad. Sci. U. S. A. 95, 1091–1096.

Cheong, C.M., Chow, A.W.S., Fitter, S., Hewett, D.R., Martin, S.K., Williams, S.A., To, L.B., Zannettino, A.C.W., Vandyke, K., 2015. Tetraspanin 7 (TSPAN7) expression is upregulated in multiple myeloma patients and inhibits myeloma tumour development in vivo. Exp. Cell Res. 332, 24–38. doi:10.1016/j.yexcr.2015.01.006

Chittenden, T.W., Howe, E.A., Culhane, A.C., Sultana, R., Taylor, J.M., Holmes, C., Quackenbush, J., 2008. Functional classification analysis of somatically mutated genes in human breast and colorectal cancers. Genomics 91, 508–511. doi:10.1016/j.ygeno.2008.03.002

Chu, I.M., Hengst, L., Slingerland, J.M., 2008. The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. Nat. Rev. Cancer 8, 253–267. doi:10.1038/nrc2347

Cooksley-Decasper, S., Reiser, H., Thommen, D.S., Biedermann, B., Neidhart, M., Gawinecka, J., Cathomas, G., Franzeck, F.C., Wyss, C., Klingenberg, R., Nanni, P., Roschitzki, B., Matter, C., Wolint, P., Emmert, M.Y., Husmann, M., Amann-Vesti, B., Maier, W., Gay, S., Luscher, T.F., von Eckardstein, A., Hof, D., 2012. Antibody phage display assisted identification of junction plakoglobin as a potential biomarker for atherosclerosis. PloS One 7, e47985. doi:10.1371/journal.pone.0047985

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., Stein, L., 2011. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 39, D691–D697. doi:10.1093/nar/gkq1018

Crotzer, V.L., Blum, J.S., 2008. Cytosol to Lysosome Transport of Intracellular Antigens During Immune Surveillance. Traffic 9, 10–16. doi:10.1111/j.1600-0854.2007.00664.x

DAI, Y., JIANG, J.-B., WANG, Y.-L., JIN, Z.-T., HU, S.-Y., 2015. Functional and protein-protein interaction network analysis of colorectal cancer induced by ulcerative colitis. Mol. Med. Rep. 12, 4947–4958. doi:10.3892/mmr.2015.4102

de Graeff, P., Crijns, A.P.G., ten Hoor, K.A., Klip, H.G., Hollema, H., Oien, K., Bartlett, J.M., Wisman, G.B.A., de Bock, G.H., de Vries, E.G.E., de Jong, S., van der Zee, A.G.J., 2008. The ErbB signalling pathway: protein expression and prognostic value in epithelial ovarian cancer. Br. J. Cancer 99, 341–349. doi:10.1038/sj.bjc.6604471

Delire, B., Starkel, P., 2015. The Ras/MAPK pathway and hepatocarcinoma: pathogenesis and therapeutic implications. Eur. J. Clin. Invest. 45, 609–623. doi:10.1111/eci.12441

Denduluri, S.K., Idowu, O., Wang, Z., Liao, Z., Yan, Z., Mohammed, M.K., Ye, J., Wei, Q., Wang, J., Zhao, L., Luu, H.H., 2015. Insulin-like growth factor (IGF) signaling in tumorigenesis and the development of cancer drug resistance. Genes Dis. 2, 13–25. doi:10.1016/j.gendis.2014.10.004

Deng, H., Shi, J., Wilkerson, M., Meschter, S., Dupree, W., Lin, F., 2008. Usefulness of S100P in diagnosis of adenocarcinoma of pancreas on fine-needle aspiration biopsy specimens. Am. J. Clin. Pathol. 129, 81–88. doi:10.1309/5D76NDE81LE8G545

Desgrosellier, J.S., Cheresh, D.A., 2010. Integrins in cancer: biological implications and therapeutic opportunities. Nat Rev Cancer 10, 9–22. doi:10.1038/nrc2748

Ding, M.-X., Wang, H.-F., Wang, J.-S., Zhan, H., Zuo, Y.-G., Yang, D.-L., Liu, J.-Y., Wang, W., Ke, C.-X., Yan, R.-P., 2012. ppGalNAc T1 as a potential novel marker for human bladder cancer. Asian Pac. J. Cancer Prev. APJCP 13, 5653–5657.

Dive, C., Smith, R.A., Garner, E., Ward, T., George-Smith, S.S., Campbell, F., Greenhalf, W., Ghaneh, P., Neoptolemos, J.P., 2010. Considerations for the use of plasma cytokeratin 18 as a biomarker in pancreatic cancer. Br. J. Cancer 102, 577–582. doi:10.1038/sj.bjc.6605494

Dote, H., Cerna, D., Burgan, W.E., Camphausen, K., Tofilon, P.J., 2005. ErbB3 expression predicts tumor cell radiosensitization induced by Hsp90 inhibition. Cancer Res. 65, 6967–6975. doi:10.1158/0008-5472.CAN-05-1304

Dowen, S.E., Crnogorac-Jurcevic, T., Gangeswaran, R., Hansen, M., Eloranta, J.J., Bhakta, V., Brentnall, T.A., Luttges, J., Kloppel, G., Lemoine, N.R., 2005. Expression of S100P and its novel binding partner S100PBPR in early pancreatic cancer. Am. J. Pathol. 166. doi:10.1016/S0002-9440(10)62234-1

Dufresne, M., Seva, C., Fourmy, D., 2006. Cholecystokinin and gastrin receptors. Physiol. Rev. 86, 805–847. doi:10.1152/physrev.00014.2005
Edelman, E.J., Guinney, J., Chi, J.-T., Febbo, P.G., Mukherjee, S., 2008. Modeling cancer progression via pathway dependencies. PLoS Comput. Biol. 4. doi:10.1371/journal.pcbi.0040028
Egeblad, M., Nakasone, E.S., Werb, Z., 2010. Tumors as organs: complex tissues that interface with the entire organism. Dev. Cell 18, 884–901. doi:10.1016/j.devcel.2010.05.012
Engelman, J.A., Janne, P.A., Mermel, C., Pearlberg, J., Mukohara, T., Fleet, C., Cichowski, K., Johnson, B.E., Cantley, L.C., 2005. ErbB-3 mediates phosphoinositide 3-kinase activity in gefitinib-sensitive non-small cell lung cancer cell lines. Proc. Natl. Acad. Sci. U. S. A. 102, 3788–3793. doi:10.1073/pnas.0409773102
Esni, F., Johansson, B.R., Radice, G.L., Semb, H., 2001. Dorsal pancreas agenesis in N-cadherin- deficient mice. Dev. Biol. 238, 202–212. doi:10.1006/dbio.2001.0405
Esteller, M., 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 8, 286–298. doi:10.1038/nrg2005
Fernandez, E., La Vecchia, C., D'Avanzo, B., Negri, E., Franceschi, S., 1994. Family history and the risk of liver, gallbladder, and pancreatic cancer. Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol. 3, 209–212.
Fernández-Vega, I., García-Suárez, O., García, B., Crespo, A., Astudillo, A., Quirós, L.M., 2015. Heparan sulfate proteoglycans undergo differential expression alterations in right sided colorectal cancer, depending on their metastatic character. BMC Cancer 15, 742. doi:10.1186/s12885-015-1724-9
Fero, M.L., Randel, E., Gurley, K.E., Roberts, J.M., Kemp, C.J., 1998. The murine gene p27Kip1 is haplo-insufficient for tumour suppression. Nature 396, 177–180. doi:10.1038/24179
Ferrara, N., 1999. Molecular and biological properties of vascular endothelial growth factor. J Mol Med 77. doi:10.1007/s001099900019
Ferrone, C.R., Brennan, M.F., Gonen, M., Coit, D.G., Fong, Y., Chung, S., Tang, L., Klimstra, D., Allen, P.J., 2008. Pancreatic adenocarcinoma: the actual 5-year survivors. J. Gastrointest. Surg. Off. J. Soc. Surg. Aliment. Tract 12, 701–706. doi:10.1007/s11605-007-0384-8
Fok, J.Y., Mehta, K., 2007. Tissue transglutaminase induces the release of apoptosis inducing factor and results in apoptotic death of pancreatic cancer cells. Apoptosis Int. J. Program. Cell Death 12, 1455–1463. doi:10.1007/s10495-007-0079-3
Folkman, J., 1990. What is the evidence that tumors are angiogenesis dependent? J Natl Cancer Inst 82. doi:10.1093/jnci/82.1.4
Fong, D., Moser, P., Krammel, C., Gostner, J.M., Margreiter, R., Mitterer, M., Gastl, G., Spizzo, G., 2008. High expression of TROP2 correlates with poor prognosis in pancreatic cancer. Br. J. Cancer 99, 1290–1295. doi:10.1038/sj.bjc.6604677
Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., Jensen, L.J., 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41, D808–D815. doi:10.1093/nar/gks1094
Friess, H., 1995. Enhanced erbB-3 expression in human pancreatic cancer correlates with tumor progression. Clin Cancer Res 1.
Frolov, A., Schuller, K., Tzeng, C.-W.D., Cannon, E.E., Ku, B.C., Howard, J.H., Vickers, S.M., Heslin, M.J., Buchsbaum, D.J., Arnoletti, J.P., 2007. ErbB3 expression and dimerization with EGFR influence pancreatic cancer cell sensitivity to erlotinib. Cancer Biol. Ther. 6, 548–554.

Furukawa, T., Kanai, N., Shiwaku, H.O., Soga, N., Uehara, A., Horii, A., 2006. AURKA is one of the downstream targets of MAPK1/ERK2 in pancreatic cancer. Oncogene 25, 4831–4839. doi:10.1038/sj.onc.1209494

Galluzzi, L., Kroemer, G., 2008. Necroptosis: a specialized pathway of programmed necrosis. Cell 135, 1161–1163. doi:10.1016/j.cell.2008.12.004

Gannon, M., Gamer, L.W., Wright, C.V., 2001. Regulatory regions driving developmental and tissue-specific expression of the essential pancreatic gene pdx1. Dev. Biol. 238. doi:10.1006/dbio.2001.0359

Gannon, M., Herrera, P.L., Wright, C.V., 2000. Mosaic Cre-mediated recombination in pancreas using the pdx-1 enhancer/promoter. Genes. N. Y. N 2000 26, 143–144.

Gao, J., Zhu, F., Lv, S., Li, Z., Ling, Z., Gong, Y., Jie, C., Ma, L., 2010. Identification of pancreatic juice proteins as biomarkers of pancreatic cancer. Oncol. Rep. 23, 1683–1692.

Georgiadi, A., Lichtenstein, L., Degenhardt, T., Boekschoten, M.V., van Bilsen, M., Desvergne, B., Muller, M., Kersten, S., 2010. Induction of cardiac Angptl4 by dietary fatty acids is mediated by peroxisome proliferator-activated receptor beta/delta and protects against fatty acid-induced oxidative stress. Circ. Res. 106, 1712–1721. doi:10.1161/CIRCRESAHA.110.217380

Gold, D.V., Karanjawala, Z., Modrak, D.E., Goldenberg, D.M., Hruban, R.H., 2007. PAM4-reactive MUC1 is a biomarker for early pancreatic adenocarcinoma. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 13, 7380–7387. doi:10.1158/1078-0432.CCR-07-1488

Gradwohl, G., Dierich, A., LeMeur, M., Guillemot, F., 2000. neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. Proc. Natl. Acad. Sci. U. S. A. 97, 1607–1611.

Grantzdorffer, I., Carl-McGrath, S., Ebert, M.P., Rocken, C., 2008. Proteomics of pancreatic cancer. Pancreas 36, 329–336. doi:10.1097/MPA.0b013e31815cc452

Grivennikov, S.I., Greten, F.R., Karin, M., 2010. Immunity, inflammation, and cancer. Cell 140, 883–899. doi:10.1016/j.cell.2010.01.025

Gronborg, M., Bunkenborg, J., Kristiansen, T.Z., Jensen, O.N., Yeo, C.J., Hruban, R.H., Maitra, A., Goggins, M.G., Pandey, A., 2004. Comprehensive proteomic analysis of human pancreatic juice. J. Proteome Res. 3, 1042–1055. doi:10.1021/pr0499085

Grutzmann, R., Boriss, H., Ammerpohl, O., Luttges, J., Kalthoff, H., Schackert, H.K., Kloppel, G., Saeger, H.D., Pilarsky, C., 2005. Meta-analysis of microarray data on pancreatic cancer defines a set of commonly dysregulated genes. Oncogene 24, 5079–5088. doi:10.1038/sj.onc.1208696

Guix, M., Faber, A.C., Wang, S.E., Olivares, M.G., Song, Y., Qu, S., Rinehart, C., Seidel, B., Yee, D., Arteaga, C.L., Engelman, J.A., 2008. Acquired resistance to EGFR tyrosine kinase inhibitors in cancer cells is mediated by loss of IGF-binding proteins. J. Clin. Invest. 118, 2609–2619. doi:10.1172/JCI34588

Gupta, V.K., 2002. Vascular endothelial growth factor enhances endothelial cell survival and tumor radioresistance. Cancer J 8. doi:10.1097/00130404-200201000-00009

Haglund, C., Kuusela, P., Jalanko, H., Roberts, P.J., 1987. Serum CA 50 as a tumor marker in pancreatic cancer: a comparison with CA 19-9. Int. J. Cancer J. Int. Cancer 39, 477–481.

Hahn, S.A., Seymour, A.B., Hoque, A.T., Schutte, M., da Costa, L.T., Redston, M.S., Caldas, C., Weinstein, C.L., Fischer, A., Yeo, C.J., 1995. Allelotype of pancreatic adenocarcinoma using xenograft enrichment. Cancer Res. 55, 4670–4675.

Hanahan, D., Folkman, J., 1996. Patterns and emerging mechanisms of the angiogenic switch during tumorigenesis. Cell 86. doi:10.1016/S0092-8674(00)80108-7

Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. Cell 100.
Hartel, M., Narla, G., Wente, M.N., Giese, N.A., Martignoni, M.E., Martignetti, J.A., Friess, H., Friedman, S.L., 2008. Increased alternative splicing of the KLF6 tumour suppressor gene correlates with prognosis and tumour grade in patients with pancreatic cancer. Eur. J. Cancer Oxf. Engl. 1990 44, 1895–1903. doi:10.1016/j.ejca.2008.06.030
Hashimoto, Y., Murakami, Y., Uemura, K., Hayashidani, Y., Sudo, T., Ohge, H., Fukuda, E., Sueda, T., Hiyama, E., 2008. Detection of human telomerase reverse transcriptase (hTERT) expression in tissue and pancreatic juice from pancreatic cancer. Surgery 143, 113–125. doi:10.1016/j.surg.2007.07.042
Hayes, G.M., Carrigan, P.E., Dong, M., Reubi, J.-C., Miller, L.J., 2007. A novel secretin receptor splice variant potentially useful for early diagnosis of pancreatic carcinoma. Gastroenterology 133, 853–861. doi:10.1053/j.gastro.2007.06.013
Herbst, M., 2016. Facts sheet on pancreatic cancer.
Hiraoka, N., Ino, Y., Sekine, S., Tsuda, H., Shimada, K., Kosuge, T., Zavada, J., Yoshida, M., Yamada, K., Koyama, T., Kanai, Y., 2010. Tumour necrosis is a postoperative prognostic marker for pancreatic cancer patients with a high interobserver reproducibility in histological evaluation. Br. J. Cancer 103, 1057–1065. doi:10.1038/sj.bjc.6605854
Ho, A.S., Huang, X., Cao, H., Christman-Skieller, C., Bennewith, K., Le, Q.-T., Koong, A.C., 2010. Circulating miR-210 as a Novel Hypoxia Marker in Pancreatic Cancer. Transl. Oncol. 3, 109–113.
Hong, S.-H., Misek, D.E., Wang, H., Puravs, E., Giordano, T.J., Greenson, J.K., Brenner, D.E., Simeone, D.M., Logsdon, C.D., Hanash, S.M., 2004. An autoantibody-mediated immune response to calreticulin isoforms in pancreatic cancer. Cancer Res. 64, 5504–5510. doi:10.1158/0008-5472.CAN-04-0077
Hong, S.-H., Misek, D.E., Wang, H., Puravs, E., Hinderer, R., Giordano, T.J., Greenson, J.K., Brenner, D.E., Simeone, D.M., Logsdon, C.D., Hanash, S.M., 2006. Identification of a Specific Vimentin Isoform That Induces an Antibody Response in Pancreatic Cancer. Biomark. Insights 1.
Hotz, H.G., 2001. Angiogenesis inhibitor TNP-470 reduces human pancreatic cancer growth. J Gastrointest Surg 5. doi:10.1016/S1091-255X(01)80024-X
Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 35, W169–W175. doi:10.1093/nar/gkm415
Huang, M.-J., Hu, R.-H., Chou, C.-H., Hsu, C.-L., Liu, Y.-W., Huang, J., Hung, J.-S., Lai, I.-R., Juan, H.-F., Yu, S.-L., Wu, Y.-M., Huang, M.-C., 2015. Knockdown of GALNT1 suppresses malignant phenotype of hepatocellular carcinoma by suppressing EGFR signaling. Oncotarget 6, 5650–5665.
Hynes, N.E., MacDonald, G., 2009. ErbB receptors and signaling pathways in cancer. Curr. Opin. Cell Biol. 21, 177–184. doi:10.1016/j.ceb.2008.12.010
Itakura, J., 1997. Enhanced expression of vascular endothelial growth factor in human pancreatic cancer correlates with local disease progression. Clin Cancer Res 3.
Jendraschak, E., Sage, E.H., 1996. Regulation of angiogenesis by SPARC and angiostatin: implications for tumor cell biology. Semin. Cancer Biol. 7, 139–146.
Johansson, M., Denardo, D.G., Coussens, L.M., 2008. Polarized immune responses differentially regulate cancer development. Immunol. Rev. 222, 145–154. doi:10.1111/j.1600-065X.2008.00600.x

Jones, J.L., Walker, R.A., 1999. Integrins: a role as cell signalling molecules. Mol. Pathol. 52, 208–213.
Jones, P.A., Baylin, S.B., 2007. The epigenomics of cancer. Cell 128, 683–692. doi:10.1016/j.cell.2007.01.029
Jones, R.B., Gordus, A., Krall, J.A., MacBeath, G., 2006. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 439, 168–174. doi:10.1038/nature04177
Jones, S., Hruban, R.H., Kamiyama, M., Borges, M., Zhang, X., Parsons, D.W., Lin, J.C.-H., Palmisano, E., Brune, K., Jaffee, E.M., Iacobuzio-Donahue, C.A., Maitra, A., Parmigiani, G., Kern, S.E., Velculescu, V.E., Kinzler, K.W., Vogelstein, B., Eshleman, J.R., Goggins, M., Klein, A.P., 2009. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 324. doi:10.1126/science.1171202
Jones, S., Zhang, X., Parsons, D.W., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S.-M., Fu, B., Lin, M.-T., Calhoun, E.S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., Smith, D.R., Hidalgo, M., Leach, S.D., Klein, A.P., Jaffee, E.M., Goggins, M., Maitra, A., Iacobuzio-Donahue, C., Eshleman, J.R., Kern, S.E., Hruban, R.H., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V.E., Kinzler, K.W., 2008a. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801–1806. doi:10.1126/science.1164368
Jones, S., Zhang, X., Parsons, D.W., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Kamiyama, H., Jimeno, A., Hong, S.-M., Fu, B., Lin, M.-T., Calhoun, E.S., Kamiyama, M., Walter, K., Nikolskaya, T., Nikolsky, Y., Hartigan, J., Smith, D.R., Hidalgo, M., Leach, S.D., Klein, A.P., Jaffee, E.M., Goggins, M., Maitra, A., Iacobuzio-Donahue, C., Eshleman, J.R., Kern, S.E., Hruban, R.H., Karchin, R., Papadopoulos, N., Parmigiani, G., Vogelstein, B., Velculescu, V.E., Kinzler, K.W., 2008b. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science 321, 1801–1806. doi:10.1126/science.1164368
Kaspar, P., Dvorakova, M., Kralova, J., Pajer, P., Kozmik, Z., Dvorak, M., 1999. Myb-interacting protein, ATBF1, represses transcriptional activity of Myb oncoprotein. J. Biol. Chem. 274, 14422–14428.
Kataoka, H., Miura, Y., Joh, T., Seno, K., Tada, T., Tamaoki, T., Nakabayashi, H., Kawaguchi, M., Asai, K., Kato, T., Itoh, M., 2001. Alpha-fetoprotein producing gastric cancer lacks transcription factor ATBF1. Oncogene 20, 869–873. doi:10.1038/sj.onc.1204160
Kersten, S., 2005. Regulation of lipid metabolism via angiopoietin-like proteins. Biochem. Soc. Trans. 33, 1059–1062. doi:10.1042/BST20051059
Kersten, S., Lichtenstein, L., Steenbergen, E., Mudde, K., Hendriks, H.F.J., Hesselink, M.K., Schrauwen, P., Muller, M., 2009. Caloric restriction and exercise increase plasma ANGPTL4 levels in humans via elevated free fatty acids. Arterioscler. Thromb. Vasc. Biol. 29, 969–974. doi:10.1161/ATVBAHA.108.182147
Kiefel, H., Bondong, S., Erbe-Hoffmann, N., Hazin, J., Riedle, S., Wolf, J., Pfeifer, M., Arlt, A., Schafer, H., Muerkoster, S.S., Altevogt, P., 2010. L1CAM-integrin interaction induces constitutive NF-[kappa]B activation in pancreatic adenocarcinoma cells by enhancing IL-1[beta] expression. Oncogene 29, 4766–4778.
Kobayashi, T., Shimabukuro-Demoto, S., Yoshida-Sugitani, R., Furuyama-Tanaka, K., Karyu, H., Sugiura, Y., Shimizu, Y., Hosaka, T., Goto, M., Kato, N., Okamura, T., Suematsu, M., Yokoyama, S., Toyama-Sorimachi, N., n.d. The Histidine Transporter

SLC15A4 Coordinates mTOR-Dependent Inflammatory Responses and Pathogenic Antibody Production. Immunity 41, 375–388. doi:10.1016/j.immuni.2014.08.011
Kojima, S., Nara, K., Rifkin, D.B., 1993. Requirement for transglutaminase in the activation of latent transforming growth factor-beta in bovine endothelial cells. J Cell Biol 121. doi:10.1083/jcb.121.2.439
Koopmann, J., Zhang, Z., White, N., Rosenzweig, J., Fedarko, N., Jagannath, S., Canto, M.I., Yeo, C.J., Chan, D.W., Goggins, M., 2004. Serum diagnosis of pancreatic adenocarcinoma using surface-enhanced laser desorption and ionization mass spectrometry. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 10, 860–868.
Korc, M., 1998. Role of growth factors in pancreatic cancer. Surg Oncol Clin N Am 7.
Korc, M., Meltzer, P., Trent, J., 1986. Enhanced expression of epidermal growth factor receptor correlates with alterations of chromosome 7 in human pancreatic cancer. Proc Natl Acad Sci U A 83. doi:10.1073/pnas.83.14.5141
Kornmann, M., 1997. Fibroblast growth factor-5 stimulates mitogenic signaling and is overexpressed in human pancreatic cancer: evidence for autocrine and paracrine actions. Oncogene 15. doi:10.1038/sj.onc.1201307
LeBleu, V.S., Taduri, G., O'Connell, J., Teng, Y., Cooke, V.G., Woda, C., Sugimoto, H., Kalluri, R., 2013. Origin and function of myofibroblasts in kidney fibrosis. Nat. Med. 19, 1047–1053. doi:10.1038/nm.3218
Lee, J., Kim, S.S., 2009. The function of p27 KIP1 during tumor development. Exp. Mol. Med. 41, 765–771. doi:10.3858/emm.2009.41.11.102
LeRoith, D., Roberts Jr, C.T., 2003. The insulin-like growth factor system and cancer. Cancer Lett. 195, 127–137. doi:10.1016/S0304-3835(03)00159-9
Liles, J.S., Arnoletti, J.P., Tzeng, C.-W.D., Howard, J.H., Kossenkov, A.V., Kulesza, P., Heslin, M.J., Frolov, A., 2010. ErbB3 expression promotes tumorigenesis in pancreatic adenocarcinoma. Cancer Biol. Ther. 10, 555–563. doi:10.4161/cbt.10.6.12532
Lin, J., Gan, C.M., Zhang, X., Jones, S., Sjoblom, T., Wood, L.D., Parsons, D.W., Papadopoulos, N., Kinzler, K.W., Vogelstein, B., Parmigiani, G., Velculescu, V.E., 2007. A multidimensional analysis of genes mutated in breast and colorectal cancers. Genome Res. 17, 1304–1318. doi:10.1101/gr.6431107
Liu, X., Sun, Y., Ehrlich, M., Lu, T., Kloog, Y., Weinberg, R.A., Lodish, H.F., Henis, Y.I., 2000. Disruption of TGF-beta growth inhibition by oncogenic ras is linked to p27Kip1 mislocalization. Oncogene 19, 5926–5935.
Logsdon, C.D., Lu, W., 2016. The Significance of Ras Activity in Pancreatic Cancer Initiation. Int. J. Biol. Sci. 12, 338–346. doi:10.7150/ijbs.15020
LONG, J., LIU, Z., WU, X., XU, Y., GE, C., 2016. Gene expression profile analysis of pancreatic cancer based on microarray data. Mol. Med. Rep. 13, 3913–3919. doi:10.3892/mmr.2016.5021
Longati, P., Jia, X., Eimer, J., Wagman, A., Witt, M.-R., Rehnmark, S., Verbeke, C., Toftgård, R., Löhr, M., Heuchel, R.L., 2013. 3D pancreatic carcinoma spheroids induce a matrix-rich, chemoresistant phenotype offering a better model for drug testing. BMC Cancer 13, 95. doi:10.1186/1471-2407-13-95
Loos, M., Hedderich, D.M., Ottenhausen, M., Giese, N.A., Laschinger, M., Esposito, I., Kleeff, J., Friess, H., 2009. Expression of the costimulatory molecule B7-H3 is associated with prolonged survival in human pancreatic cancer. BMC Cancer 9. doi:10.1186/1471-2407-9-463
Lopez-Casas, P.P., Lopez-Fernandez, L.A., 2010. Gene-expression profiling in pancreatic cancer. Expert Rev. Mol. Diagn. 10, 591–601. doi:10.1586/erm.10.43

Lowenfels, A.B., Maisonneuve, P., 2004. Epidemiology and prevention of pancreatic cancer. Jpn. J. Clin. Oncol. 34, 238–244.
Lowenfels, A.B., Maisonneuve, P., Lankisch, P.G., 1999. CHRONIC PANCREATITIS AND OTHER RISK FACTORS FOR PANCREATIC CANCER. Gastroenterol. Clin. North Am. 28, 673–685. doi:10.1016/S0889-8553(05)70080-7
Lu, H., Sun, J., Wang, F., Feng, L., Ma, Y., Shen, Q., Jiang, Z., Sun, X., Wang, X., Jin, H., 2013. Enhancer of zeste homolog 2 activates wnt signaling through downregulating CXXC finger protein 4. Cell Death Dis 4, e776.
Lu, Z., Hu, L., Evers, S., Chen, J., Shen, Y., 2004. Differential expression profiling of human pancreatic adenocarcinoma and healthy pancreatic tissue. Proteomics 4, 3975–3988. doi:10.1002/pmic.200300863
Luo, J., 2001. Pancreatic cancer cell-derived vascular endothelial growth factor is biologically active in vitro and enhances tumorigenicity in vivo. Int J Cancer 92. doi:10.1002/ijc.1202
Lynch, H.T., Deters, C.A., Lynch, J.F., Brand, R.E., 2004. Familial pancreatic carcinoma in Jews. Fam. Cancer 3, 233–240. doi:10.1007/s10689-004-9549-8
Maacke, H., Jost, K., Opitz, S., Miska, S., Yuan, Y., Hasselbach, L., Luttges, J., Kalthoff, H., Sturzbecher, H.W., 2000. DNA repair and recombination factor Rad51 is over-expressed in human pancreatic adenocarcinoma. Oncogene 19, 2791–2795. doi:10.1038/sj.onc.1203578
Maestro, M.A., Boj, S.F., Luco, R.F., Pierreux, C.E., Cabedo, J., Servitja, J.M., German, M.S., Rousseau, G.G., Lemaigre, F.P., Ferrer, J., 2003. Hnf6 and Tcf2 (MODY5) are linked in a gene network operating in a precursor cell domain of the embryonic pancreas. Hum. Mol. Genet. 12, 3307–3314. doi:10.1093/hmg/ddg355
Mandard, S., Zandbergen, F., Tan, N.S., Escher, P., Patsouris, D., Koenig, W., Kleemann, R., Bakker, A., Veenman, F., Wahli, W., Muller, M., Kersten, S., 2004. The direct peroxisome proliferator-activated receptor target fasting-induced adipose factor (FIAF/PGAR/ANGPTL4) is present in blood plasma as a truncated protein that is increased by fenofibrate treatment. J. Biol. Chem. 279, 34411–34420. doi:10.1074/jbc.M403058200
Mantoni, T.S., Lunardi, S., Al-Assar, O., Masamune, A., Brunner, T.B., 2011. Pancreatic Stellate Cells Radioprotect Pancreatic Cancer Cells through β1-Integrin Signaling. Cancer Res. 71, 3453. doi:10.1158/0008-5472.CAN-10-1633
Mantovani, A., 2010. Molecular pathways linking inflammation and cancer. Curr. Mol. Med. 10, 369–373.
Mao, Z., Ma, X., Fan, X., Cui, L., Zhu, T., Qu, J., Zhang, J., Wang, X., 2014. Secreted protein acidic and rich in cysteine inhibits the growth of human pancreatic cancer cells with G1 arrest induction. Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med. 35, 10185–10193. doi:10.1007/s13277-014-2315-0
Martini, M., De Santis, M.C., Braccini, L., Gulluni, F., Hirsch, E., 2014. PI3K/AKT signaling pathway and cancer: an updated review. Ann. Med. 46, 372–383. doi:10.3109/07853890.2014.912836
Mehta, K., 2008. Biological and therapeutic significance of tissue transglutaminase in pancreatic cancer. Amino Acids 36, 709. doi:10.1007/s00726-008-0128-4
Melo, S.A., Luecke, L.B., Kahlert, C., Fernandez, A.F., Gammon, S.T., Kaye, J., LeBleu, V.S., Mittendorf, E.A., Weitz, J., Rahbari, N., Reissfelder, C., Pilarsky, C., Fraga, M.F., Piwnica-Worms, D., Kalluri, R., 2015. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature 523, 177–182.

Mi, H., Guo, N., Kejariwal, A., Thomas, P.D., 2007. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Res. 35, D247–D252. doi:10.1093/nar/gkl869

Miyazawa, Y., Uekita, T., Hiraoka, N., Fujii, S., Kosuge, T., Kanai, Y., Nojima, Y., Sakai, R., 2010. CUB domain-containing protein 1, a prognostic factor for human pancreatic cancers, promotes cell migration and extracellular matrix degradation. Cancer Res. 70, 5136–5146. doi:10.1158/0008-5472.CAN-10-0220

Modolell, I., Guarner, L., Malagelada, J.R., 1999. Vagaries of clinical presentation of pancreatic and biliary tract cancer. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 10 Suppl 4, 82–84.

Mohelnikova-Duchonova, B., Oliverius, M., Honsova, E., Soucek, P., 2012. Evaluation of reference genes and normalization strategy for quantitative real-time PCR in human pancreatic carcinoma. Dis. Markers 32, 203–210. doi:10.3233/DMA-2011-0875

Molina, L.M., Diez, M., Cava, M.T., Maestro, M.L., Ortega, M.D., Mendiz, J.G., Prego, A., De Diego, J.A., Balibrea, J.L., 1990. Tumor markers in pancreatic cancer: a comparative clinical study between CEA, CA. Int. J. Biol. Markers 5, 127–132.

Morgan, M.A., Parsels, L.A., Kollar, L.E., Normolle, D.P., Maybaum, J., Lawrence, T.S., 2008. The combination of epidermal growth factor receptor inhibitors with gemcitabine and radiation in pancreatic cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 14, 5142–5149. doi:10.1158/1078-0432.CCR-07-4072

Murphy, K.M., Brune, K.A., Griffin, C., Sollenberger, J.E., Petersen, G.M., Bansal, R., Hruban, R.H., Kern, S.E., 2002. Evaluation of candidate genes MAP2K4, MADH4, ACVR1B, and BRCA2 in familial pancreatic cancer: deleterious BRCA2 mutations in 17%. Cancer Res. 62, 3789–3793.

Murtaugh, L.C., Stanger, B.Z., Kwan, K.M., Melton, D.A., 2003. Notch signaling controls multiple steps of pancreatic differentiation. Proc. Natl. Acad. Sci. U. S. A. 100, 14920–14925. doi:10.1073/pnas.2436557100

Nagaraju, G.P., El-Rayes, B.F., 2013. SPARC and DNA methylation: possible diagnostic and therapeutic implications in gastrointestinal cancers. Cancer Lett. 328, 10–17. doi:10.1016/j.canlet.2012.08.028

Naghibalhossaini, F., Yoder, A.D., Tobi, M., Stanners, C.P., 2007. Evolution of a Tumorigenic Property Conferred by Glycophosphatidyl-Inositol Membrane Anchors of Carcinoembryonic Antigen Gene Family Members during the Primate Radiation. Mol. Biol. Cell 18, 1366–1374. doi:10.1091/mbc.E06-10-0884

National institute of occupational health, 2016. Summary stastistics of cancer diagnosed histologically in 2012.

Ndubaku, C.O., Heffron, T.P., Staben, S.T., Baumgardner, M., Blaquiere, N., Bradley, E., Bull, R., Do, S., Dotson, J., Dudley, D., Edgar, K.A., Friedman, L.S., Goldsmith, R., Heald, R.A., Kolesnikov, A., Lee, L., Lewis, C., Nannini, M., Nonomiya, J., Pang, J., Price, S., Prior, W.W., Salphati, L., Sideris, S., Wallin, J.J., Wang, L., Wei, B., Sampath, D., Olivero, A.G., 2013. Discovery of. J. Med. Chem. 56, 4597–4610. doi:10.1021/jm4003632

Nicke, B., Tseng, M.-J., Fenrich, M., Logsdon, C.D., 1999. Adenovirus-mediated gene transfer of Ras^{N17} inhibits specific CCK actions on pancreatic acinar cells. Am. J. Physiol. - Gastrointest. Liver Physiol. 276, G499.

Nojiri, S., Joh, T., Miura, Y., Sakata, N., Nomura, T., Nakao, H., Sobue, S., Ohara, H., Asai, K., Ito, M., 2004. ATBF1 enhances the suppression of STAT3 signaling by interaction with PIAS3. Biochem. Biophys. Res. Commun. 314, 97–103.

Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., Wright, C.V., 1996. PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Dev. Camb. Engl. 122, 983–995.

Omenn, G.S., Yocum, A.K., Menon, R., 2010. Alternative splice variants, a new class of protein cancer biomarker candidates: findings in pancreatic cancer and breast cancer with systems biology implications. Dis. Markers 28, 241–251. doi:10.3233/DMA-2010-0702

OuYang, D., Xu, J., Huang, H., Chen, Z., 2011. Metabolomic profiling of serum from human pancreatic cancer patients using 1H NMR spectroscopy and principal component analysis. Appl. Biochem. Biotechnol. 165, 148–154. doi:10.1007/s12010-011-9240-0

Pan, S., Brentnall, T.A., Kelly, K., Chen, R., 2013. Tissue proteomics in pancreatic cancer study: discovery, emerging technologies, and challenges. Proteomics 13, 710–721. doi:10.1002/pmic.201200319

Pandha, H., Rigg, A., John, J., Lemoine, N., 2007. Loss of expression of antigen-presenting molecules in human pancreatic cancer and pancreatic cancer cell lines. Clin. Exp. Immunol. 148, 127–135. doi:10.1111/j.1365-2249.2006.03289.x

Pathan, M., Keerthikumar, S., Ang, C.-S., Gangoda, L., Quek, C.Y.J., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., Bacic, A., Hill, A.F., Stroud, D.A., Ryan, M.T., Agbinya, J.I., Mariadason, J.M., Burgess, A.W., Mathivanan, S., 2015. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 15, 2597–2601. doi:10.1002/pmic.201400515

Pellegata, N.S., Quintanilla-Martinez, L., Siggelkow, H., Samson, E., Bink, K., Hofler, H., Fend, F., Graw, J., Atkinson, M.J., 2006. Germ-line mutations in p27Kip1 cause a multiple endocrine neoplasia syndrome in rats and humans. Proc. Natl. Acad. Sci. U. S. A. 103, 15558–15563. doi:10.1073/pnas.0603877103

Pepe, M.S., Feng, Z., Janes, H., Bossuyt, P.M., Potter, J.D., 2008. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. J. Natl. Cancer Inst. 100, 1432–1438. doi:10.1093/jnci/djn326

Pfaffl, M.W., Horgan, G.W., Dempfle, L., 2002. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 30, e36–e36.

Pierreux, C.E., Poll, A.V., Kemp, C.R., Clotman, F., Maestro, M.A., Cordi, S., Ferrer, J., Leyns, L., Rousseau, G.G., Lemaigre, F.P., 2006. The transcription factor hepatocyte nuclear factor-6 controls the development of pancreatic ducts in the mouse. Gastroenterology 130, 532–541. doi:10.1053/j.gastro.2005.12.005

Pietras, K., Sjoblom, T., Rubin, K., Heldin, C.-H., Ostman, A., 2003. PDGF receptors as cancer drug targets. Cancer Cell 3, 439–443.

Polireddy, K., Chen, Q., 2016. Cancer of the Pancreas: Molecular Pathways and Current Advancement in Treatment. J. Cancer 7, 1497–1514. doi:10.7150/jca.14922

Pryczynicz, A., Guzinska-Ustymowicz, K., Kemona, A., Czyzewska, J., 2010. Expression of the E-cadherin-catenin complex in patients with pancreatic ductal adenocarcinoma. Folia Histochem. Cytobiol. 48, 128–133. doi:10.2478/v10042-008-0089-1

Puri, S., Hebrok, M., 2010. Cellular plasticity within the pancreas--lessons learned from development. Dev. Cell 18, 342–356. doi:10.1016/j.devcel.2010.02.005

Rodolfo, C., Mormone, E., Matarrese, P., Ciccosanti, F., Farrace, M.G., Garofano, E., Piredda, L., Fimia, G.M., Malorni, W., Piacentini, M., 2004. Tissue transglutaminase is a multifunctional BH3-only protein. J. Biol. Chem. 279, 54783–54792. doi:10.1074/jbc.M410938200

Rossi, M.K., Gnanamony, M., Gondi, C.S., 2016. The “SPARC” of life: Analysis of the role of osteonectin/SPARC in pancreatic cancer (Review). Int. J. Oncol. 48, 1765–1771. doi:10.3892/ijo.2016.3417
Sabbaghian, M.S., Rothberger, G., Alongi, A.P., Gagner, J.-P., Goldberg, J.D., Rolnitzky, L., Chiriboga, L., Hajdu, C.H., Zagzag, D., Basch, R., Shamamian, P., 2010. Levels of elevated circulating endothelial cell decline after tumor resection in patients with pancreatic ductal adenocarcinoma. Anticancer Res. 30, 2911–2917.
Sachdev, D., Yee, D., 2007. Disrupting insulin-like growth factor signaling as a potential cancer therapy. Mol. Cancer Ther. 6, 1. doi:10.1158/1535-7163.MCT-06-0080
Sasawatari, S., Okamura, T., Kasumi, E., Tanaka-Furuyama, K., Yanobu-Takanashi, R., Shirasawa, S., Kato, N., Toyama-Sorimachi, N., 2011. The solute carrier family 15A4 regulates TLR9 and NOD1 functions in the innate immune system and promotes colitis in mice. Gastroenterology 140, 1513–1525. doi:10.1053/j.gastro.2011.01.041
Sevimoglu, T., Arga, K.Y., 2014. The role of protein interaction networks in systems biomedicine. Comput. Struct. Biotechnol. J. 11, 22–27. doi:10.1016/j.csbj.2014.08.008
Seymour, P.A., Freude, K.K., Tran, M.N., Mayes, E.E., Jensen, J., Kist, R., Scherer, G., Sander, M., 2007. SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc. Natl. Acad. Sci. U. S. A. 104, 1865–1870. doi:10.1073/pnas.0609217104
Shao, Z.M., Nguyen, M., Barsky, S.H., 2000. Human breast carcinoma desmoplasia is PDGF initiated. Oncogene 19, 4337–4345. doi:10.1038/sj.onc.1203785
Shen, J., Person, M.D., Zhu, J., Abbruzzese, J.L., Li, D., 2004. Protein expression profiles in pancreatic adenocarcinoma compared with normal pancreatic tissue and tissue affected by pancreatitis as detected by two-dimensional gel electrophoresis and mass spectrometry. Cancer Res. 64, 9018–9026. doi:10.1158/0008-5472.CAN-04-3262
Sherr, C.J., Roberts, J.M., 1999. CDK inhibitors: positive and negative regulators of G1-phase progression. Genes Dev. 13, 1501–1512.
Shibata, K., Nakayama, T., Hirakawa, H., Hidaka, S., Nagayasu, T., 2010. Clinicopathological significance of angiopoietin-like protein 4 expression in oesophageal squamous cell carcinoma. J. Clin. Pathol. 63, 1054–1058. doi:10.1136/jcp.2010.078600
Siegel, R., Naishadham, D., Jemal, A., 2012. Cancer statistics, 2012. CA. Cancer J. Clin. 62. doi:10.3322/caac.20138
Sithanandam, G., Anderson, L.M., 2008. The ERBB3 receptor in cancer and cancer gene therapy. Cancer Gene Ther. 15, 413–448. doi:10.1038/cgt.2008.15
Skandalakis, J., Colborn, G.L., Weidman, T.A., n.d. Skandalakis’ surgical anatomy.
Skotheim, R.I., Abeler, V.M., Nesland, J.M., Fosså, S.D., Holm, R., Wagner, U., Flørenes, V.A., Aass, N., Kallioniemi, O.P., Lothe, R.A., 2003. Candidate Genes for Testicular Cancer Evaluated by In Situ Protein Expression Analyses on Tissue Microarrays. Neoplasia N. Y. N 5, 397–404.
Skrypek, N., Vasseur, R., Vincent, A., Duchene, B., Van Seuningen, I., Jonckheere, N., 2015. The oncogenic receptor ErbB2 modulates gemcitabine and irinotecan/SN-38 chemoresistance of human pancreatic cancer cells via hCNT1 transporter and multidrug-resistance associated protein MRP-2. Oncotarget 6, 10853–10867. doi:10.18632/oncotarget.3414
Smith, J.P., Fantaskey, A.P., Liu, G., Zagon, I.S., 1995. Identification of gastrin as a growth peptide in human pancreatic cancer. Am. J. Physiol. 268, R135-141.
Smith, J.P., Fonkoua, L.K., Moody, T.W., 2016. The Role of Gastrin and CCK Receptors in Pancreatic Cancer and other Malignancies. Int. J. Biol. Sci. 12, 283–291. doi:10.7150/ijbs.14952

Smith, J.P., Liu, G., Soundararajan, V., McLaughlin, P.J., Zagon, I.S., 1994. Identification and characterization of CCK-B/gastrin receptors in human pancreatic cancer cell lines. Am. J. Physiol. 266, R277-283.

Smith, J.P., Solomon, T.E., Bagheri, S., Kramer, S., 1990. Cholecystokinin stimulates growth of human pancreatic adenocarcinoma SW-1990. Dig. Dis. Sci. 35, 1377–1384.

Soler, M., Mancini, F., Meca-Cortes, O., Sanchez-Cid, L., Rubio, N., Lopez-Fernandez, S., Lozano, J.J., Blanco, J., Fernandez, P.L., Thomson, T.M., 2009. HER3 is required for the maintenance of neuregulin-dependent and -independent attributes of malignant progression in prostate cancer cells. Int. J. Cancer 125, 2565–2575. doi:10.1002/ijc.24651

Soos, T.J., Kiyokawa, H., Yan, J.S., Rubin, M.S., Giordano, A., DeBlasio, A., Bottega, S., Wong, B., Mendelsohn, J., Koff, A., 1996. Formation of p27-CDK complexes during the human mitotic cell cycle. Cell Growth Differ. Mol. Biol. J. Am. Assoc. Cancer Res. 7, 135–146.

Storey, J.D., Madeoy, J., Strout, J.L., Wurfel, M., Ronald, J., Akey, J.M., 2007. Gene-Expression Variation Within and Among Human Populations. Am. J. Hum. Genet. 80, 502–509.

Sun, X., Fu, X., Li, J., Xing, C., Frierson, H.F., Wu, H., Ding, X., Ju, T., Cummings, R.D., Dong, J.-T., n.d. Deletion of Atbf1/Zfhx3 In Mouse Prostate Causes Neoplastic Lesions, Likely by Attenuation of Membrane and Secretory Proteins and Multiple Signaling Pathways. Neoplasia 16, 377–389. doi:10.1016/j.neo.2014.05.001

Swartz, M.J., Batra, S.K., Varshney, G.C., Hollingsworth, M.A., Yeo, C.J., Cameron, J.L., Wilentz, R.E., Hruban, R.H., Argani, P., 2002. MUC4 Expression Increases Progressively in Pancreatic Intraepithelial Neoplasia. Am. J. Clin. Pathol. 117, 791–796. doi:10.1309/7Y7N-M1WM-R0YK-M2VA

Takayama, R., Nakagawa, H., Sawaki, A., Mizuno, N., Kawai, H., Tajika, M., Yatabe, Y., Matsuo, K., Uehara, R., Ono, K., Nakamura, Y., Yamao, K., 2010. Serum tumor antigen REG4 as a diagnostic biomarker in pancreatic ductal adenocarcinoma. J. Gastroenterol. 45, 52–59. doi:10.1007/s00535-009-0114-y

Tan, G.-J., Peng, Z.-K., Lu, J.-P., Tang, F.-Q., 2013. Cathepsins mediate tumor metastasis. World J. Biol. Chem. 4, 91–101. doi:10.4331/wjbc.v4.i4.91

Terris, B., Blaveri, E., Crnogorac-Jurcevic, T., Jones, M., Missiaglia, E., Ruszniewski, P., Sauvanet, A., Lemoine, N.R., 2002. Characterization of gene expression profiles in intraductal papillary-mucinous tumors of the pancreas. Am. J. Pathol. 160, 1745–1754. doi:10.1016/S0002-9440(10)61121-2

The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. Nature 526, 68–74.

Thorsen, K., Sorensen, K.D., Brems-Eskildsen, A.S., Modin, C., Gaustadnes, M., Hein, A.-M.K., Kruhoffer, M., Laurberg, S., Borre, M., Wang, K., Brunak, S., Krainer, A.R., Torring, N., Dyrskjot, L., Andersen, C.L., Orntoft, T.F., 2008. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. Mol. Cell. Proteomics MCP 7, 1214–1224. doi:10.1074/mcp.M700590-MCP200

Tian, F., Appert, H.E., Myles, J., Howard, J.M., 1992. Prognostic value of serum CA 19-9 levels in pancreatic adenocarcinoma. Ann. Surg. 215, 350–355.

Tian, X., Hao, K., Qin, C., Xie, K., Xie, X., Yang, Y., 2013. Insulin-like growth factor 1 receptor promotes the growth and chemoresistance of pancreatic cancer. Dig. Dis. Sci. 58, 2705–2712. doi:10.1007/s10620-013-2673-2

Tobita, K., Kijima, H., Dowaki, S., Kashiwagi, H., Ohtani, Y., Oida, Y., Yamazaki, H., Nakamura, M., Ueyama, Y., Tanaka, M., Inokuchi, S., Makuuchi, H., 2003.

Epidermal growth factor receptor expression in human pancreatic cancer: Significance for liver metastasis. Int. J. Mol. Med. 11, 305–309.
Toyoda, E., Doi, R., Koizumi, M., Kami, K., Ito, D., Mori, T., Fujimoto, K., Nakajima, S., Wada, M., Imamura, M., 2005. Analysis of E-, N-cadherin, alpha-, beta-, and gamma-catenin expression in human pancreatic carcinoma cell lines. Pancreas 30, 168–173.
Trikudanathan, G., Philip, A., Dasanu, C.A., Baker, W.L., 2011. Association between Helicobacter pylori infection and pancreatic cancer. A cumulative meta-analysis. JOP J. Pancreas 12.
Troiani, T., Martinelli, E., Capasso, A., Morgillo, F., Orditura, M., De Vita, F., Ciardiello, F., 2012. Targeting EGFR in pancreatic cancer treatment. Curr. Drug Targets 13, 802–810.
Ueno, Y., Sakurai, H., Tsunoda, S., Choo, M.-K., Matsuo, M., Koizumi, K., Saiki, I., 2008. Heregulin-induced activation of ErbB3 by EGFR tyrosine kinase activity promotes tumor growth and metastasis in melanoma cells. Int. J. Cancer 123, 340–347. doi:10.1002/ijc.23465
Vaz, J., Ansari, D., Sasor, A., Andersson, R., 2015. SPARC: A Potential Prognostic and Therapeutic Target in Pancreatic Cancer. Pancreas 44, 1024–1035. doi:10.1097/MPA.0000000000000409
Viglietto, G., Motti, M.L., Bruni, P., Melillo, R.M., D'Alessio, A., Califano, D., Vinci, F., Chiappetta, G., Tsichlis, P., Bellacosa, A., Fusco, A., Santoro, M., 2002. Cytoplasmic relocalization and inhibition of the cyclin-dependent kinase inhibitor p27(Kip1) by PKB/Akt-mediated phosphorylation in breast cancer. Nat. Med. 8, 1136–1144. doi:10.1038/nm762
Vogelstein, B., Kinzler, K.W., 2004. Cancer genes and the pathways they control. Nat. Med. 10, 789–799. doi:10.1038/nm1087
Waddell, N., Pajic, M., Patch, A.-M., Chang, D.K., Kassahn, K.S., Bailey, P., Johns, A.L., Miller, D., Nones, K., Quek, K., Quinn, M.C.J., Robertson, A.J., Fadlullah, M.Z.H., Bruxner, T.J.C., Christ, A.N., Harliwong, I., Idrisoglu, S., Manning, S., Nourse, C., Nourbakhsh, E., Wani, S., Wilson, P.J., Markham, E., Cloonan, N., Anderson, M.J., Fink, J.L., Holmes, O., Kazakoff, S.H., Leonard, C., Newell, F., Poudel, B., Song, S., Taylor, D., Waddell, N., Wood, S., Xu, Q., Wu, J., Pinese, M., Cowley, M.J., Lee, H.C., Jones, M.D., Nagrial, A.M., Humphris, J., Chantrill, L.A., Chin, V., Steinmann, A.M., Mawson, A., Humphrey, E.S., Colvin, E.K., Chou, A., Scarlett, C.J., Pinho, A.V., Giry-Laterriere, M., Rooman, I., Samra, J.S., Kench, J.G., Pettitt, J.A., Merrett, N.D., Toon, C., Epari, K., Nguyen, N.Q., Barbour, A., Zeps, N., Jamieson, N.B., Graham, J.S., Niclou, S.P., Bjerkvig, R., Grutzmann, R., Aust, D., Hruban, R.H., Maitra, A., Iacobuzio-Donahue, C.A., Wolfgang, C.L., Morgan, R.A., Lawlor, R.T., Corbo, V., Bassi, C., Falconi, M., Zamboni, G., Tortora, G., Tempero, M.A., Gill, A.J., Eshleman, J.R., Pilarsky, C., Scarpa, A., Musgrove, E.A., Pearson, J.V., Biankin, A.V., Grimmond, S.M., 2015. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 518, 495–501. doi:10.1038/nature14169
Walker, C.J., Miranda, M.A., O'Hern, M.J., McElroy, J.P., Coombes, K.R., Bundschuh, R., Cohn, D.E., Mutch, D.G., Goodfellow, P.J., 2015. Patterns of CTCF and ZFHX3 Mutation and Associated Outcomes in Endometrial Cancer. JNCI J. Natl. Cancer Inst. 107, djv249. doi:10.1093/jnci/djv249
Wandzioch, E., Zaret, K.S., 2009. Dynamic signaling network for the specification of embryonic pancreas and liver progenitors. Science 324, 1707–1710. doi:10.1126/science.1174497
Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. doi:10.1038/nrg2484

Weinberg, D.S., Ruggeri, B., Barber, M.T., Biswas, S., Miknyocki, S., Waldman, S.A., 1997. Cholecystokinin A and B receptors are differentially expressed in normal pancreas and pancreatic adenocarcinoma. J. Clin. Invest. 100, 597–603. doi:10.1172/JCI119570

Weissmueller, S., Manchado, E., Saborowski, M., Morris, J.P. 4th, Wagenblast, E., Davis, C.A., Moon, S.-H., Pfister, N.T., Tschaharganeh, D.F., Kitzing, T., Aust, D., Markert, E.K., Wu, J., Grimmond, S.M., Pilarsky, C., Prives, C., Biankin, A.V., Lowe, S.W., 2014. Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor beta signaling. Cell 157, 382–394. doi:10.1016/j.cell.2014.01.066

White, E., Karp, C., Strohecker, A.M., Guo, Y., Mathew, R., 2010. Role of autophagy in suppression of inflammation and cancer. Curr. Opin. Cell Biol. 22, 212–217. doi:10.1016/j.ceb.2009.12.008

Wuttig, D., Zastrow, S., Fussel, S., Toma, M.I., Meinhardt, M., Kalman, K., Junker, K., Sanjmyatav, J., Boll, K., Hackermuller, J., Rolle, A., Grimm, M.-O., Wirth, M.P., 2012. CD31, EDNRB and TSPAN7 are promising prognostic markers in clear-cell renal cell carcinoma revealed by genome-wide expression analyses of primary tumors and metastases. Int. J. Cancer 131, E693-704. doi:10.1002/ijc.27419

Xia, Q., Kong, X.-T., Zhang, G.-A., Hou, X.-J., Qiang, H., Zhong, R.-Q., 2005. Proteomics-based identification of DEAD-box protein 48 as a novel autoantigen, a prospective serum marker for pancreatic cancer. Biochem. Biophys. Res. Commun. 330, 526–532. doi:10.1016/j.bbrc.2005.02.181

Yuan, F., Zhang, Y.-H., Wan, S., Wang, S., Kong, X.-Y., 2015. Mining for Candidate Genes Related to Pancreatic Cancer Using Protein-Protein Interactions and a Shortest Path Approach. BioMed Res. Int. 2015, 623121. doi:10.1155/2015/623121

Zhang, L., Farrell, J.J., Zhou, H., Elashoff, D., Akin, D., Park, N.-H., Chia, D., Wong, D.T., 2010. Salivary Transcriptomic Biomarkers for Detection of Resectable Pancreatic Cancer. Gastroenterology 138. doi:10.1053/j.gastro.2009.11.010

Zhang, L., Jin, H., Guo, X., Yang, Z., Zhao, L., Tang, S., Mo, P., Wu, K., Nie, Y., Pan, Y., Fan, D., 2012. Distinguishing pancreatic cancer from chronic pancreatitis and healthy individuals by (1)H nuclear magnetic resonance-based metabonomic profiles. Clin. Biochem. 45, 1064–1069. doi:10.1016/j.clinbiochem.2012.05.012

Zhang, Z., Yamashita, H., Toyama, T., Sugiura, H., Ando, Y., Mita, K., Hamaguchi, M., Kawaguchi, M., Miura, Y., Iwase, H., 2005. ATBF1-a messenger RNA expression is correlated with better prognosis in breast cancer. Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 11, 193–198.

Zhou, L., Lu, Z., Yang, A., Deng, R., Mai, C., Sang, X., Faber, K.N., Lu, X., 2007. Comparative proteomic analysis of human pancreatic juice: methodological study. Proteomics 7, 1345–1355. doi:10.1002/pmic.200600086

Zhu, P., Goh, Y.Y., Chin, H.F.A., Kersten, S., Tan, N.S., 2012. Angiopoietin-like 4: a decade of research. Biosci. Rep. 32, 211–219. doi:10.1042/BSR20110102

Zong, W.-X., Thompson, C.B., 2006. Necrotic death as a cell fate. Genes Dev. 20, 1–15. doi:10.1101/gad.1376506