Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Integrative multi-omic analytics for the early detection of pancreatic cancer

Laura E. Kane

A dissertation submitted to Trinity College Dublin for the degree of Doctor of Philosophy

August 2023

Primary Supervisor: Dr. Stephen G. Maher

Secondary Supervisor: Prof. Barbara M. Ryan

Department of Surgery

Trinity College Dublin

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement. I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Summary

Pancreatic cancer (PC) has the lowest 5-year survival rate of any cancer, at just 12% in 2023. The poor survival rates observed in this cancer are the result of an accumulation of shortcomings and gaps in many aspects of PC research, with early detection and diagnosis being at the forefront of concerns. PC symptoms are notoriously vague and everyday in nature. Issues such as appetite loss, abdominal pain or weight loss are early warning signs, however, most patients ignore these common aliments, regarding them as unimportant. As such, the majority of patients diagnosed with PC are found at a late stage of cancer development, further contributing to the poor survival rates of this cancer. Moreover, the only FDA-approved biomarker for PC diagnosis, carbohydrate antigen 19-9 (CA19-9), is known to perform poorly in patients with co-morbidities such as diabetes or pancreatitis, making its utility in this setting extremely limited.

Pancreatic cystic lesions (PCLs) are fluid-filled protrusions either on, or inside, the pancreas, which can be benign or pre-malignant. PCLs in many cases are precursors to PC, and as such could be the key to early detection of this cancer. Unfortunately, the current guidelines used to stratify patients based on their PC risk are imperfect. At present, there are several sets of these guidelines utilised globally, highlighting the lack of consensus among clinicians as to the best approach for these patients. PCLs represent a unique opportunity to identify and monitor patients with a high-risk of PC development, and as such more robust solutions to the stratification of these lesions are of urgent need.

Here, the vast accumulation of research into potential biomarkers for PC is explored. A systematic review and meta-analysis of blood-based biomarkers highlights over 40 years of PC research, examining 250 manuscripts, and puts a spotlight on several promising biomarkers. Multi-biomarker panels are shown to be superior to single biomarkers alone, indicating that panels of multiple biomarkers more adequately adapt to patient-to-patient variability, allowing other biomarkers to make up the shortfall when one is dysregulated; a luxury not shared by single biomarkers. Moreover, it is shown that the current standard biomarker for PC diagnosis, CA19-9, has little utility on its own, but when added to a panel of multiple biomarkers, it can improve the panel. The study cohort utilised for biomarker identification was demonstrated to greatly

affect the results obtained. Indeed, cohorts of both healthy individuals and patients with a benign condition were found to be more clinically relevant and produce more robust results than cohorts of either alone. Importantly, 13 novel biomarkers were identified as being repeatedly examined across the literature, with promising qualitative metrics for PC diagnosis. This systematic review of the literature also allowed the detailed critique of current PC research, with several common issues in study design and data reporting being highlighted across the included studies.

Taking the information and evident flaws in PC research that were revealed via the systematic review and meta-analysis, this study then delved into biomarker identification and validation. Indeed, the performance of those promising diagnostic biomarkers identified in the systematic review was examined in a cohort of patient pancreatic cyst fluid (PCF) and blood sera, to determine their utility for risk stratification of patients with PCLs. While some were found to be differentially expressed in patients at a low- or high-risk of PC, their ability to stratify patients based on expression of these markers was poor. As such, exploratory multi-omic profiling of patient PCF and serum was conducted, focusing on the proteome and transcriptome of both. Strict differential expression analysis identified proteins and miRNA in both biofluids that were differentially expressed in low- and high-risk patients. Alone, these panels of proteins and miRNA perform poorly-to-moderately in stratifying patients based on their risk. However, when coupled to form multi-omic panels, these biomarkers perform risk stratification with high accuracy. Finally, integration of both the PCF-based multi-omic panel and the serum-based multi-omic panel was conducted using CombiROC software. This cross-biofluid multi-omic biomarker panel demonstrated the best stratification performance overall, and was uniquely capable of appropriately controlling for outlier patients with genetic mutations. While validation of these results in an independent patient cohort remains to be conducted, this promising biomarker panel could be the key to early detection of PC.

Lastly, the biological activity of the PCF was interrogated. As PCLs can be precursors to PC, the role that the fluid within these cysts plays in the progression of PCLs to PC remains to be elucidated. Here, normal and intermediary pancreatic cell lines were exposed to patient PCF at low concentrations for 24 h. PCF was demonstrated to

be biologically active, having the ability to alter cell line viability, proliferation, apoptosis, metabolism, phenotypic and functional marker expression levels, and DNA damage levels. Furthermore, PCF was shown to cause functional changes to the cells, resulting in increased invasive potential. However, PCF in many cases was also shown to be extremely cytotoxic to cells, even at low concentrations for just 24 h. Indeed, two distinct effects were seen in the PCF, cytotoxic and stimulatory. These preliminary results demonstrate the biological activity of PCF, and raise questions about its potential role in the progression of PCLs to PC. Further work is urgently needed to interrogate the mechanisms at play during the PCL to PC transformation, and the part that PCF plays in this process. If PCF could be demonstrated to contribute to the malignant transformation of PCLs, the aspiration of all PCLs as a preventative measure could be called for. This would have substantial implications for the management of PCL patients, and also for the advent of potential preventative measures for PC development.

Overall, this body of work highlights the desperate need for better understanding of PCLs and PC, and presents promising avenues for future research in biomarker identification and early detection of PC.

Acknowledgements

This thesis is the culmination, not just of years or work, but a lifetime of support, guidance, friendship and love. I am exceedingly lucky to have too many people in my life worthy of thanks to name, but I'll do my best.

As a zoologist-turned-cancer researcher, my journey to this point would not have been possible without the belief and trust of my primary supervisor, Dr. Stephen G. Maher. Stephen, I don't know what possessed you to take a chance on a student with a zoology undergrad for this highly translational cancer research, but I will be eternally grateful you did. Thank you for teaching me what it means to ask the hard questions, but more importantly, how to answer them. The generosity with which you give your time, your extreme attention to detail, and your stellar sense of humour have made this PhD a fun and rewarding experience. Thank you for giving me the freedom to pursue the hard questions, and for making me the scientist I am today.

To my secondary supervisor, Prof Barbara M. Ryan, thank you for all of your support and encouragement over the last few years, and for always looking out for me. I'm so delighted I was able to work with you, and develop a passion for pancreatic research that I didn't know was inside me. You have been an inspiration to me throughout this research project.

Importantly, I would like to thank both the Meath Foundation and Viatris Ireland for funding this important work. I would also like to extend my thanks to the patients who so generously donate their samples to research, and without whom this work would not be possible.

Dr. Gregory Mellotte, thank you for always being only a text away, and for being such a supportive and calm lab partner. Your stories of Kate and Monty got us through some long, tough days in the lab, and your help along the course of this work has been invaluable. I'm so glad I was able to share this journey with you, and to learn from you.

To the Maher lab group, Dr. Jason McGrath, Rebecca Lyons and David Hackett. Thank you all for being such wonderful lab pals, and for taking turns reminding Stephen about our group meetings. The time during my PhD was made better by knowing all of you.

The Department of Surgery has been an incredible place to be a part of over the last few years, solely due to the wonderful people within it. Thank you to all the PIs, Prof. Jacintha O'Sullivan, Prof. Joanne Lysaght, Dr. Niamh Lynam-Lennon and Dr. Graham Pidgeon, for providing a stimulating and encouraging environment to share research ideas and try new things. It is this atmosphere that allows projects and young researchers to flourish, and for that I am so grateful. I would also like to thank the many researchers who have come through the Dept. of Surgery during my PhD, and that have contributed to the wonderful time I have spent here. Thank you, Klaudia Majcher, Dr. Maria Davern, Dr. Simone Marcone, Dr. Aoife Cannon, Dr. Margaret Dunne, Dr. Aisling Heeran, Dr. Aoife Kilgallon, Dr. Aisling Ui Mhaonaigh, Dr. Noel Donlon, Dr. Croi Buckley, Dr. Marina Zaki, Anshul Bhardwaj, Christine Butler, Dr. Rebecca O'Brien, Andrew Sheppard, Christina Cahill, Maitiú Ó Murchú, Lorraine Smith, Caroline Marion, Dr. Brendan Moran, Kirstan Murphy, Niamh O'Connor, Dr. Fiona Crotty. I would also like to extend a massive thanks to Dr. Melissa Conroy. Melissa, you have not only been a huge inspiration to me over the last few years, but your pep-talks, advice, and extensive knowledge have been an invaluable source of support for me, thank you. Also, to Dr. Martin Barr, thank you for always stopping in the halls for a chat, and for being the upbeat and cheerful person who brightens a long day in the lab. Lastly, Dr. Aidan Meade, thank you for your advice and guidance on all things high-dimensional, and for being such a

pleasure to work with. Also, thank you for teaching me that just because no one has ever done something before, doesn't mean its not possible.

Before the Dept of Surgery, there was Palmerstown Credit Union. PCU enabled me to undertake my master's degree, without which I would not be here. But more importantly, the people that I have had the pleasure to work with and call my friends in PCU, are among the finest there are. Breda, Mags, Gemma, Stacy, Trisha, Betty, Jen, Amy, Teresa, Ken, Peggy, Mary, Lisa, Denise, Ellie, Gary, Seán, Michael Craig and Michael Connolly, thank you for encouraging me and for helping me to grow. You have all been a huge source of support and love in my life. But also, to Molly O'Callaghan, who told me to dream big. I hope this is what you had in mind.

To the best PhD pals a girl could ask for. Dr. Eimear Mylod, my day one PhD pal. I absolutely would not be at this point if I had not been so fortunate as to meet you on my first day. You are the reason for my social coffee addiction, and the rock I have needed to get through some of the hardest times of the PhD. You are an incredible scientist, and somehow an even better friend. You have the best high kicks I have ever seen, and I would follow your lead anywhere. Dr. Fiona O'Connell, ever the help and never the hindrance. I've learned so much from you over the last few years, like how to multichannel absolutely anything, or how to check the oil in my car. You've been my travel buddy and my chauffeur, my ear to vent to, and my biggest fan. You are one of the most thoughtful, caring, above-and-beyond people I've ever encountered, and while I wish you'd learn the word 'no', there is no one I'd rather walk down Frenchmen Street with. I know you'll do outstanding things, just maybe don't go so far. Dr. Laura Bogue Edgerton, my La-in-arms. I don't know how it took us so long to figure out that we're basically the same person, especially with you living just up the road, but I will be eternally grateful that we did. The long walks with you and Raura have been a saving grace, and your ted talk voice notes while driving are some of the best things I receive. Thank you for being there to deliver care packages, and listen to me ramble, and for being such a selfless and loving person. You are stronger than you know, and you inspire me. This PhD has given me friends for life in you three girls, and I promise to always cherish it.

To my Friday wine night crew. Floriane O'Keeffe, joining the TCD basketball team was one of the best decisions of my life, because it lead me to you. You have been an inspiration to me throughout our whole friendship, and have given me an unconditional love that I strive to deserve. From maths tutorials, to science balls, zoology, Kenya, Annecy, PhDs, drag race, wine nights and takeaways; I'm so lucky to have had you by my side through all of it. Thank you for being a wonderful friend, for welcoming me into your home, and for learning to like wine as much as me. Dr. Andrew Mooney, my stats tutor turned best friend. I can't tell you how much your friendship has meant to me these last few years. Thank you for having such impeccable taste and opinions, and for proof-reading like nobody's business. Thank you for being such an incredible friend. Your support and encouragement has spurred me to the end, and I will strive to continue to write paragraphs that you think are fire. Dr. Jenny Bortoluzzi, my stress-filled submission partner. Thank you for offering to proof-read on the week of your viva, and for meeting me for wine on difficult days. Along this whole process, having you there with me has been cathartic, and being able to share this experience with you has meant so much to me. Dr. Paula Tierney, my fellow zoologist-turned-cancer researcher. Thank you for being the incredible hype-woman that you are, and for always listening with interest about my soap opera-based life. Your enthusiasm and boss-woman attitude never ceases to amaze me. You four people have kept me sane, listened endlessly

to my woes, and have always stood by me in support and love. You are the best thing that's ever been mine.

To the Kanes and the Powers, who are vastly too many to name. Thank you for always supporting me, and for fighting over which last name I should go by. I would not be the person I am today without your love, encouragement, and jokes. I'm delighted to say I have finally finished school, and I so sincerely hope this work has made you all proud.

Without my immediate family, I would not be where, or who, I am today.

My parents, Louise and Robbie, are the two pillars which have shaped me into the person I am today. To both of you, thank you for giving me everything you could and more, and for teaching me everything I know. Mam, thank you for talking to me on the phone for hours and hours, and listening to every detail of my life. Having you there as a source of support has meant the world to me. Dad, thank you for always telling me there was room for improvement, and for hugging me like I've been gone for years. Thanks to you, I don't think my education is *so* sadly lacking. At every step of the way, neither of you have ever doubted me, and you have *always* believed in me. I could never truly articulate what you mean to me, but I'm sure you already know, moreya.

To my brothers, Blob and Dave. Growing up with both of you, and now growing as an adult, I could not ask for two more amazing siblings. There whenever I need you, endlessly supporting me, looking after me and just generally assuming the mindful brother roles. Dave, thank you for enduring long phone calls, and then long voice notes, for checking in on me, and for inspiring me from a young age. Thank you for being proud of me, and believing I could do anything. Blob, thank you for never getting too old to think your big sister is cool, and for bragging about being related to me. You are so generous with everything you have to give, and I'm so grateful to be on the receiving end. Thank you for being two wild characters, and for shaping my personality. I'm very proud to have such wonderful brothers who made me the weirdo I am today, even if you're slow at texting back sometimes. To Mel and Aoife, thank you for being the sisters I never had. Mel, you are an incredible, strong, brave, resilient, and caring person. Thank you for always being only a text, or a 20 minute voice note away. Aoife, in the short time I've known you, I feel you've already become such an integral part of the family. Thank you for thinking I'm the cool one, though we both know that's not true.

To my beautiful niece and goddaughter, Evie. You're far too small to read this now, but maybe one day you will. Your love and admiration is infinite, as is mine for you. I hope to inspire you as you get older, and to be the role model you deserve.

Finally, to my partner Ralph. It's been over 12 years of partnership, all of which I've managed to stay in 'school'. Thank you for enabling me to follow my dreams, for encouraging me at every step, for believing in me at every point, and for giving me Raura. You are my rock, and my partner. I promise now I'll get a real job.

This thesis is dedicated to my parents, Louise and Robbie, who taught me everything they know – and more.

Publications

First author publications (2)

- **Diagnostic accuracy of blood-based biomarkers for pancreatic cancer: A systematic review and meta-analysis.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Eimear Mylod, Rebecca M. O'Brien, Fiona O'Connell, Croí E. Buckley, Jennifer Arlow, Khanh Nguyen, David Mockler, Aidan D. Meade, Barbara M. Ryan, Stephen G. Maher. *Cancer Research Communications, 2022.*
- **Multi-Omic biomarkers as potential tools for the characterisation of pancreatic cystic lesions and cancer: Innovative patient data integration.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Kevin C. Conlon, Barbara M. Ryan, Stephen G. Maher. *Cancers, 2021.*

First author abstracts/conference proceedings (6)

- **Establishment of a novel multi-omic biomarker panel in cyst fluid and blood for stratifying patient risk of pancreatic cancer.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Rebecca G. Lyons, Eimear Mylod, Simone Marcone, Paul F. Ridgway, Finbar MacCarthy, Kevin C. Conlon, Joanne Lysaght, Barbara M. Ryan, Stephen G. Maher. *Cancer Research, 2022.*
- **Diagnostic accuracy of blood-based biomarkers for pancreatic cancer.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Eimear Mylod, Rebecca M. O'Brien, Fiona O'Connell, Croí E. Buckley, Jennifer Arlow, Khanh Nguyen, David Mockler, Aidan D. Meade, Barbara M. Ryan, Stephen G. Maher. *BMC Proceedings, 2022.*
- **Multi-omic profiling of patient pancreatic cyst fluid for the identification of a novel biomarker panel of patient cancer risk.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Simone Marcone, Paul F. Ridgway, Kevin C. Conlon, Barbara M. Ryan, Stephen G. Maher. *BMC Proceedings, 2022.*
- **Profiling the immune component of pancreatic cyst fluid to identify novel biomarkers of patient cancer risk.** Authors: **Laura E. Kane**, Gregory S. Mellotte,

Rebecca G. Lyons, Eimear Mylod, Simone Marcone, Paul F. Ridgway, Kevin C. Conlon, Barbara M. Ryan, Joanne Lysaght, Stephen G. Maher. *Cancer Research, 2022.*

- **Multi-omic profiling of patient pancreatic cyst fluid for the identification of a novel biomarker panel of patient cancer risk.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Simone Marcone, Barbara M. Ryan, Stephen G. Maher. *Cancer Research, 2021.*
- **Diagnostic accuracy of blood-based biomarkers for pancreatic cancer: A systematic review and meta-analysis.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Eimear Mylod, Rebecca M. O'Brien, Fiona O'Connell, Khanh Nguyen, Croí E. Buckley, Jennifer Arlow, David Mockler, Aidan D. Meade, Barbara M. Ryan, Stephen G. Maher. *Cancer Research, 2021.*

Manuscripts in preparation (2)

- **Multi-omic biomarker panel in pancreatic cyst fluid and serum predicts patients at a high-risk of pancreatic cancer.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Paul Dowling, Simone Marcone, Eimear Mylod, Caitriona Scaife, Paul F. Ridway, Finbar MacCarthy, Kevin C. Conlon, Aidan D. Meade, Barbara M. Ryan, Stephen G. Maher.
- **Pancreatic cyst fluid drives the development of pancreatic cancer in patients with pancreatic cysts.** Authors: **Laura E. Kane**, Gregory S. Mellotte, Fiona O'Connell, Eimear Mylod, Paul F. Ridway, Finbar MacCarthy, Kevin C. Conlon, Barbara M. Ryan, Stephen G. Maher.

Other publications (3)

- **Investigating the Effects of Olaparib on the Susceptibility of Glioblastoma Multiforme Tumour Cells to Natural Killer Cell-Mediated Responses.** Authors:

Jennifer Moran, Eimear Mylod, **Laura E. Kane**, Caroline Marion, Emily Keenan, Marianna Mekhaeil, Joanne Lysaght, Kumlesh K. Dev, Jacintha O'Sullivan, Melissa J. Conroy. *Pharmaceutics, 2023.*

- **The influence of microRNA-31 on oxidative stress and radiosensitivity in pancreatic ductal adenocarcinoma.** Authors: Jason McGrath, **Laura E. Kane**, Stephen G. Maher. *Cells, 2022.*
- **Platelets, immune cells and the coagulation cascade; friend or foe of the circulating tumour cell?** Authors: Mark P. Ward*, **Laura E. Kane***, Lucy A. Norris, Bashir M. Mohamed, Tanya Kelly, Mark Bates, Andres Clarke, Nathan Brady, Cara M. Martin, Robert D. Brooks, Doug A. Brooks, Stavros Selemidis, Sean Hanniffy, Eric P. Dixon, Sharon A. O'Toole, John J. O'Leary. *Molecular Cancer, 2021.*

**Denotes equal contribution.*

Presentations

Oral presentations

- **Irish Association for Cancer Research**, Prof. Patrick G Johnston Excellence in Cancer Research Outreach session. Cork, 30th March – 1st April 2022. Title: Investigating the performance of blood-based diagnostic molecules for pancreatic cancer.
- **International Conference for Healthcare and Medical Students**, Virtual, 11th February 2022. Title: Profiling the immune component of pancreatic cyst fluid to identify novel biomarkers of patient cancer risk.
- **International Conference for Healthcare and Medical Students**, Virtual, 11th February 2022. Title: Diagnostic accuracy of blood-based biomarkers for pancreatic cancer.
- **International Conference for Healthcare and Medical Students**, Virtual, 11th February 2022. Title: Multi-omic profiling of patient pancreatic cyst fluid for the identification of a novel biomarker panel of patient cancer risk.
- **Trinity Translational Medicine Institute Health Sciences Postgraduate Research Blitz**, Dublin, 11th March 2022. Title: Identification of novel biomarkers of pancreatic cancer risk.
- **Irish Society of Gastroenterology Winter Meeting**, Virtual, 2nd – 3rd December 2021. Title: Multi-omic profiling of pancreatic cyst fluid for the identification of a novel biomarker panel of patient cancer risk.
- **Irish Society of Gastroenterology Winter Meeting**, E-poster session. Virtual, 2nd – 3rd December 2021. Title: Diagnostic accuracy of blood-based biomarkers for pancreatic cancer.
- **Trinity Translational Medicine Institute Seminar Series**, Dublin, 19th October 2021. Title: The search for pancreatic cancer biomarkers – out with the old, in with the new.
- **Ballyfermot Chapelizod Partnership Webinar**, Virtual, 26th October 2021. Title: Reference Managers – how to take the stress out of referencing.
- **American Association for Cancer Research Virtual Special Conference on Pancreatic Cancer**, Virtual, 29th – 30th September 2021. Title: Multi-omic profiling

of pancreatic cyst fluid for the identification of a novel biomarker panel of patient cancer risk.

- **American Association for Cancer Research Virtual Special Conference on Pancreatic Cancer**, Virtual, 29th – 30th September 2021. Title: Diagnostic accuracy of blood-based biomarkers identified through multi-omics for PDAC: a systematic review and meta-analysis.
- **Trinity Translational Medicine Institute Health Sciences Postgraduate Research Blitz**, Virtual, 21st May 2021. Title: Multi-omic profiling of pancreatic cyst fluid for the identification of a novel biomarker panel of patient cancer risk.
- **Irish Association for Cancer Research**, Prof. Patrick G Johnston Excellence in Cancer Research Outreach session. Virtual, 24th – 26th March 2021. Title: Investigating the role of pancreatic cyst fluid in the development of pancreatic cancer.

Poster presentations

- **Irish Association for Cancer Research**, Athlone, 22nd – 24th February 2023. Title: Multi-omic blood-based biomarker panel associated with pancreatic cancer development in a novel patient cohort.
- **Trinity St. James' Cancer Institute Cancer Conference**, Dublin, 15th October 2022. Title: Identification of multi-omic blood-based biomarkers associated with pancreatic cancer development in a novel patient cohort.
- **European Association for Cancer Research Congress**, Seville, Spain 20th – 23rd June 2022. Title: Biochemical and functional characterisation of pancreatic cyst fluid identifies a novel biomarker panel of patient cancer risk.
- **American Association for Cancer Research Annual Meeting**, New Orleans, USA, 8th – 13th April 2022. Title: Establishment of a novel multi-omic biomarker panel in cyst fluid and blood for stratifying patient risk of pancreatic cancer.
- **Irish Association for Cancer Research**, Cork, 30th March – 1st April 2022. Title: Diagnostic accuracy of blood-based biomarkers for pancreatic cancer: a large scale systematic review and meta-analysis.

- **Irish Association for Cancer Research**, Cork, 30th March – 1st April 2022. Title: Multi-omic profiling and cluster analysis of pancreatic cyst fluid identifies a novel biomarker signature of patient cancer risk.
- **Irish Association for Cancer Research**, Virtual, 24th – 26th March 2021. Title: Diagnostic accuracy of blood-based biomarkers identified through multi-omics for PDAC: A systematic review.

Awards and funding

- **Best Cancer Screening and Prevention Poster**, Trinity St. James' Cancer Institute Cancer Conference, 2022.
- **Leaders in Learning Grant**, Dublin City Council, 2022.
- **Trinity Trust Travel Grant**, Trinity College Dublin, 2022.
- **Best Oral Presentation,** Abstract in Basic Science session, International Conference for Healthcare and Medical Students, 2022.
- **Best Oral Presentation**, Scientific Abstract session, Irish Society of Gastroenterology Winter Meeting 2021.
- **Leaders in Learning Grant**, Dublin City Council, 2021.
- **Best Oral Presentation,** Trinity Translational Medicine Institute Health Sciences Postgraduate Research Blitz, Trinity College Dublin, 2021.
- **GESINAS ImmunoTools Award**, ImmunoTools, 2021.
- **Prof. Patrick G Johnston Award for Excellence in Cancer Research Outreach**, Irish Association for Cancer Research, 2021.
- **Leaders in Learning Grant**, Dublin City Council, 2020.
- **Leaders in Learning Grant**, Dublin City Council, 2019.

Table of Contents

Declaration ..... **ii**
Summary ..... **iii**
Acknowledgements ..... **vi**
Dedication ..... **ix**
Publications ..... **x**
Presentations ..... **xiii**
Awards and funding ..... **xvi**
Table of Contents ..... **xvii**
List of figures ..... **xxiv**
List of Tables ..... **xxxi**
List of Appendices ..... **xxxii**
Abbreviations ..... **xxxiii**
Units ..... **xxxv**
Chapter 1. Introduction **1**
1.1 Pancreatic Cancer 2
1.1.1 Overview 2
1.1.2 Clinical presentation 2
1.2 Development of PC and the tumour microenvironment 2
1.3 Hallmarks of PC 5
1.3.1 Sustaining proliferative signalling 7
1.3.2 Evading growth suppressors 9
1.3.3 Resisting cell death 10
1.3.4 Enabling replicative immortality 10
1.3.5 Inducing or accessing vasculature 10
1.3.6 Activating invasion and metastasis 11
1.3.7 Deregulating cellular metabolism 13
1.3.8 Avoiding immune destruction 13
1.3.9 Genome instability and mutation 15
1.3.10 Tumour-promoting inflammation 16
1.3.11 Unlocking phenotypic plasticity 16
1.3.12 Senescent cells 17
1.3.13 Non-mutational epigenetic reprogramming 18
1.3.14 Polymorphic microbiomes 19

1.4 Pancreatic cystic lesions 19
1.5 PCL biology – the little we know 20
1.6 High or low – determining which PCLs are high-risk 23
1.6.1 Subtypes of PCLs 23
1.6.2 Current management of PCLs: a lack of consensus 23
1.7 Identification of biomarkers in PCLs and PC using omics 28
1.7.1 Genomics 28
1.7.2 Transcriptomics 35
1.7.3 Epigenomics 38
1.7.4 Proteomics 40
1.7.5 Metabolomics 44
1.8 Multi-omics as the key to biomarker identification 45
1.9 Rationale 54
1.10 Hypothesis 54
1.11 Specific aims 54
Chapter 2. Materials and Methods 55
2.1 Desk-based materials and methods 56
2.1.1 PRISMA guidelines 56
2.1.2 Search strategy and inclusion criteria 56
2.1.3 Data extraction and risk of bias assessment 58
2.1.4 Data clean-up and meta-analysis 62
2.2 Lab-based materials and methods 63
2.2.1 Reagents and materials 63
2.2.2 Ethical approval 63
2.2.3 Patient recruitment 63
2.2.4 Patient serum collection and storage 63
2.2.5 Patient PCF collection and storage 64
2.2.6 Sonication of patient PCF 64
2.2.7 Proteomic profiling of PCF 65
2.2.7.1 Protein quantification 65
2.2.7.2 Preparation of SP3 magnetic beads 65
2.2.7.3 Sample lysis, reduction, alkylation 66
2.2.7.4 Tryptic digestion 66
2.2.7.5 SP3 peptide clean-up and elution 67
2.2.7.6 Pierce quantitative colorimetric peptide assay 67

2.2.7.7 Peptide preparation for LC-MS 68
2.2.8 Proteomic profiling of patient serum 68
2.2.8.1 Immunodepletion of patient serum 68
2.2.8.2 Acetone precipitation 69
2.2.8.3 Processing and preparation of precipitated proteins for LC-MS 69
2.2.9 Quantification of soluble protein levels by sandwich ELISA 70
2.2.10 HTG Whole Transcriptome Sequencing of PCF and serum 71
2.2.11 QIAGEN qPCR Microarray 72
2.2.11.1 RNA isolation and quantification 72
2.2.11.2 cDNA synthesis 73
2.2.11.3 Custom quantitative RT-PCR microarray setup and analysis 74
2.2.12 Culture of cell lines 75
2.2.12.1 H6c7-normal cell line maintenance 75
2.2.12.2 HPNE-intermediary cell line maintenance 75
2.2.12.3 Sub-culture 76
2.2.12.4 Cell counting 77
2.2.12.5 Cryopreservation and reconstitution of cell line stocks 77
2.2.12.6 Mycoplasma testing 78
2.2.13 Serum starvation of cell lines 79
2.2.14 Treatment of normal pancreatic cell lines with PCF 80
2.2.15 Proliferation assay 80
2.2.16 ApoTox Glo assay 80
2.2.17 Flow cytometry 81
2.2.17.1 Antibody titrations 81
2.2.17.2 Flow cytometric staining 82
2.2.17.3 Processing and analysis of flow cytometric data 82
2.2.18 Seahorse XF ATP rate test 84
2.2.19 Invasion assay 85
2.2.20 DNA damage competitive ELISA 86
2.2.21 Data handling and statistical analysis 87
2.2.21.1 Processing and analysis of omics data 87
2.2.21.2 Processing and analysis of cell line experimental data 89
Chapter 3. Diagnostic accuracy of blood-based biomarkers for pancreatic cancer: A systematic review and meta-analysis 90
3.1 Introduction 91
3.2 Hypothesis 93

3.3 Specific aims 93
3.4 Results 94
3.4.1 Identification of relevant studies 94
3.4.2 Accuracy and RoB assessment 94
3.4.3 Summary of extracted studies 94
3.4.4 Meta-analysis: full dataset 98
3.4.5 Meta-analysis: CA19-9 and novel biomarker subgroups 101
3.4.6 Meta-analysis: CA19-9 and novel biomarkers in different patient cohort subgroups 104
3.4.7 Biomarker efficacy of different omic compartments 104
3.4.8 Biomarkers in the 90th percentile for sensitivity and specificity 108
3.4.9 Most frequently examined novel biomarkers across included studies 108
3.5 Discussion 113
3.5.1 Multi-biomarker panels - the better choice 113
3.5.2 CA19-9 and its role as the current clinical standard biomarker 114
3.5.3 Multi-omics in biomarker identification 116
3.5.4 Promising novel biomarkers for PC diagnosis 116
3.5.5 The state of current PC research 121
3.5.6 Limitations of this systematic review and meta-analysis 122
Chapter 4. Multi-omic profiling of pancreatic cyst fluid for the identification of novel biomarkers of patient pancreatic cancer risk 125
4.1 Introduction 126
4.2 Hypothesis 128
4.3 Specific Aims 128
4.4 Experimental design 129
4.4.1 Patient demographic information for PCF cohort 129
4.4.2 Proteomic profiling of PCF 129
4.4.3 Transcriptomic profiling of PCF 129
4.4.4 Analysis of proteomic and transcriptomic PCF data 129
4.4.5 Quantification of CA19-9 in patient serum by sandwich ELISA 131
4.4.6 Quantification of CEA in patient PCF 131
4.4.7 Quantification of soluble LCN2, REG1A, PIGR and PRSS8 by sandwich ELISA 131
4.4.8 QIAGEN qPCR custom microarray for PCF 131
4.5 Results 133

4.5.1 Four promising biomarkers are significantly increased in high-risk PCF 133
4.5.2 Eight novel proteins were identified as being significantly increased in high-risk PCF 133
4.5.3 8-protein panel in PCF stratifies patients with modest accuracy 137
4.5.4 Three miRNA were identified as being significantly increased in high-risk PCF . 141
4.5.5 3-miRNA panel in PCF stratifies patients with poor accuracy 145
4.5.6 11-feature multi-omic panel in PCF stratifies patients with high accuracy 145
4.5.7 CA19-9 does not improve the accuracy of the 11-feature multi-omic panel 154
4.5.8 CEA does not improve the accuracy of the 11-feature multi-omic panel 159
4.5.9 Proteomic and transcriptomic results could not be validated using other techniques 162
4.6 Discussion 170
Chapter 5. Multi-omic profiling of serum for the identification of novel biomarkers of patient pancreatic cancer risk 177
5.1 Introduction 178
5.2 Hypotheses 180
5.3 Specific Aims (Part One) 180
5.4 Specific Aims (Part Two) 180
5.5 Experimental design 181
5.5.1 Patient demographic information for serum cohort 181
5.5.2 Proteomic profiling of patient serum 181
5.5.3 Transcriptomic profiling of patient serum 181
5.5.4 Analysis of proteomic and transcriptomic serum data 181
5.5.5 QIAGEN qPCR custom microarray for serum 181
5.5.6 Panel reduction and integration using CombiROC software 183
5.6 Results: Part One 184
5.6.1 Three promising biomarkers are not differentially expressed in serum 184
5.6.2 Eight proteins were identified as being decreased in high-risk serum 184
5.6.3 8-protein panel in serum classifies patients with modest accuracy 189
5.6.4 Five miRNA were identified as significantly increased in high-risk serum 192
5.6.5 5-miRNA panel in serum stratifies patients with poor accuracy 197
5.6.6 13-feature multi-omic panel in serum stratifies patients with high accuracy 201
5.7 Results: Part Two 211

5.7.1 10-feature multi-omic cross-biofluid panel stratifies patients with high accuracy 211
5.8 Discussion 216
Chapter 6. Functional characterisation of the role of pancreatic cyst fluid in the development of pancreatic cancer 223
6.1 Introduction 224
6.2 Hypothesis 226
6.3 Specific aims 226
6.4 Experimental design 227
6.4.1 Dose-response curve for normal pancreatic cell line treatment with PCF 227
6.4.2 Flow cytometric staining of PCF treated cells 228
6.4.3 Analysis of experimental data 228
6.5 Results 229
6.5.1 PCF can be cytotoxic and significantly affects apoptosis, proliferation and viability in normal pancreatic cell lines 229
6.5.2 Serum starvation significantly alters the basal respiration of normal pancreatic cell lines 236
6.5.3 PCF does not significantly alter the metabolic profile of normal pancreatic cell lines 236
6.5.4 PCF does not alter the expression of phenotypic and functional markers on normal pancreatic cell lines after 6 h 244
6.5.5 PCF significantly decreases the percentage of Vimentin$^+$ and PD-L1$^+$ H6c7-normal cells after 24 h 249
6.5.6 Low-risk PCF significantly increases the percentage of EGFR$^+$ HPNE-intermediary cells after 24 h 253
6.5.7 PCF does not significantly alter the co-expression of phenotypic and functional markers on normal cells after 6 h 256
6.5.8 PCF significantly alters the percentage of Vimentin$^+$/Slug$^+$, Vimentin$^+$/E-cadherin$^+$, Vimentin$^+$/EGFR$^+$, Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cadherin$^+$, and PD-L1$^+$/EGFR$^+$ H6c7-normal cells after 24 h 260
6.5.9 PCF significantly increases the percentage of Vimentin$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cadherin$^+$ HPNE-intermediary cells after 24 h 267
6.5.10 PCF significantly increases the invasive capacity of normal pancreatic cell lines 270

6.5.11 PCF does not significantly alter the production of 8-OHdG by normal pancreatic cell lines 274
6.6 Discussion 280
Chapter 7. General Discussion 287
7.1 PC – the fight to improve survival 288
7.2 Using the past to guide the future 288
7.3 Data handling - a double-edged sword 291
7.4 Omics and Hallmarks of Cancer – the bigger picture 293
7.5 Limitations 295
7.6 Future directions 298
7.7 Conclusion 299
References 301
Appendices 356

List of figures

Figure 1.1 Progression of PC. 4
Figure 1.2 The Pancreatic TME. 6
Figure 1.3 The Hallmarks of PC. 8
Figure 1.4 Molecular subgroups of PCLs. 22
Figure 1.5 The multi-omic nature of *KRAS* mutations in PC. 49
Figure 2.1 Representative dot plots showing gating strategies for cell lines. 83
Figure 3.1. PRISMA flow diagram of record selection process. 95
Figure 3.2. Summary of results for the QUADAS-2 RoB and study quality assessment. 96
Figure 3.3 Multivariate three-level meta-analysis. 99
Figure 3.4 Comparison of single biomarkers and multi-biomarker panels overall and subdivided by biomarker group. 100
Figure 3.5 Comparison of single biomarkers with multi-biomarker panels for both CA19-9 and novel cohorts. 102
Figure 3.6 Evaluation of platform-to-platform variation in CA19-9 detection. 103
Figure 3.7 Comparison of CA19-9 and novel biomarkers subdivided by patient cohorts. 105
Figure 3.8 Comparison of CA19-9 and novel markers showing omic compartment breakdown. 106
Figure 3.9 Comparison of novel markers showing pooled omic breakdown. 107
Figure 3.10 Details of the biomarkers in the 90th percentile for sensitivity and specificity. 109
Figure 3.11 Details of the risk of bias assessment for biomarkers in the 90th percentile for sensitivity and specificity. 110
Figure 4.1 Four promising biomarkers are significantly increased in high-risk PCF compared to low-risk. 134
Figure 4.2 Expression of the four top occurring biomarkers in PCF clusters patients into risk groups with 58.3% accuracy. 135
Figure 4.3 Eight proteins were identified as being significantly upregulated in high-risk patient PCF compared to low-risk. 136
Figure 4.4 Expression of significant proteins in PCF significantly correlates with several clinical factors. 138

Figure 4.5 8-protein panel in PCF clusters patients into risk groups with 81.25% accuracy. 139
Figure 4.6 PCA analysis of 8-protein panel in PCF shows moderate distinction of the risk groups. 140
Figure 4.7 8-protein panel in PCF classifies patients based on risk with AUC of 0.608. 142
Figure 4.8 Three miRNA were identified as being significantly upregulated in high-risk patient PCF compared to low-risk. 143
Figure 4.9 Expression of significant miRNA in PCF does not significantly correlate with clinical factors. 144
Figure 4.10 3-miRNA panel in PCF clusters patients into risk groups with 60% accuracy. 146
Figure 4.11 PCA analysis of 3-miRNA panel in PCF shows poor distinction of the risk groups. 147
Figure 4.12 3-miRNA panel in PCF classifies patients based on risk with AUC of 0.658. 148
Figure 4.13 11-feature multi-omic panel in PCF clusters patients into risk groups with 95.8% accuracy. 150
Figure 4.14 11-feature multi-omic panel in PCF clusters patients into risk groups with 100% accuracy when VHL patient is reclassified. 151
Figure 4.15 Biomarkers within the 11-feature multi-omic panel in PCF significantly positively correlate with the expression of each other. 152
Figure 4.16 PCA analysis of 11-feature multi-omic panel in PCF shows adequate distinction of the risk groups. 153
Figure 4.17 PCA analysis of 11-feature multi-omic panel in PCF shows improved distinction of the risk groups when VHL patient is reclassified. 155
Figure 4.18 11-feature multi-omic panel in PCF classifies patients based on risk with AUC of 0.806. 156
Figure 4.19 11-feature multi-omic panel in PCF classifies patients based on risk with AUC of 0.867 when VHL patient is reclassified. 157
Figure 4.20 CA19-9 levels are not significantly different in patients with low- or high-risk PCLs. 158

Figure 4.21 11-feature multi-omic panel in PCF plus serum CA19-9 clusters patients into risk groups with 58.3% accuracy. 160
Figure 4.22 Serum CEA levels are significantly increased in high-risk patients compared to low-risk. 161
Figure 4.23 11-feature multi-omic panel in PCF plus serum CEA clusters patients into risk groups with 95% accuracy. 163
Figure 4.24 11-feature multi-omic panel in PCF plus serum CEA classifies patients based on risk with AUC of 0.690. 164
Figure 4.25 11-feature multi-omic panel in PCF plus serum CEA classifies patients based on risk with AUC of 0.636 when VHL patient is reclassified. 165
Figure 4.26 Sandwich ELISA does not validate proteomic results regardless of the normalisation method. 166
Figure 4.27 There is no correlation between LFQ Intensity values obtained via proteomics and sandwich ELISA protein concentration. 167
Figure 4.28 qPCR microarray does not validate transcriptomic results. 168
Figure 5.1 Three promising biomarkers are not significantly differentially expressed high-risk serum compared to low-risk. 185
Figure 5.2 Expression of the three top occurring biomarkers in serum clusters patients into risk groups with 65.5% accuracy. 186
Figure 5.3 Eight proteins were identified as being downregulated in high-risk patient serum compared to low-risk. 187
Figure 5.4 Expression of SHROOM3 in the serum significantly correlates with patient cancer risk. 188
Figure 5.5 8-protein panel in serum clusters patients into risk groups with 76.5% accuracy. 190
Figure 5.6 PCA analysis of 8-protein panel in serum shows poor distinction of the risk groups. 191
Figure 5.7 PCA analysis in 3-dimensions of 8-protein panel in serum shows poor distinction of risk groups. 193
Figure 5.8 8-protein panel in serum classifies patients based on risk with AUC of 0.608. 194

Figure 5.9 Five miRNA were identified as being significantly upregulated in high-risk patient serum compared to low-risk. 195
Figure 5.10 Expression of significant miRNA in the serum does not significantly correlate with clinical factors. 196
Figure 5.11 5-miRNA panel in serum clusters patients into risk groups with 60% accuracy. 198
Figure 5.12 PCA analysis of 5-miRNA panel in serum shows poor distinction of the risk groups. 199
Figure 5.13 5-miRNA panel in serum classifies patients based on risk with AUC of 0.427. 200
Figure 5.14 13-feature multi-omic panel in serum clusters patients into risk groups with 79.3% accuracy. 202
Figure 5.15 13-feature multi-omic panel in serum clusters patients into risk groups with 75.9% accuracy when VHL patient is reclassified. 203
Figure 5.16 Ten biomarkers within the 13-feature biomarker panel in serum significantly correlate with expression of each other..... 204
Figure 5.17 PCA analysis of 13-feature multi-omic panel in serum shows modest distinction of the risk groups. 206
Figure 5.18 PCA analysis of 13-feature multi-omic panel in serum shows no change in the distinction of the risk groups when VHL patient is reclassified. 207
Figure 5.19 PCA analysis in 3-dimensions of 13-feature multi-omic panel in serum shows modest distinction of risk groups. 208
Figure 5.20 13-feature multi-omic panel in serum classifies patients based on risk with AUC of 0.824. 209
Figure 5.21 13-feature multi-omic panel in serum classifies patients based on risk with AUC of 0.948 when VHL patient is reclassified. 210
Figure 5.22 10-feature multi-omic cross-biofluid panel classifies patients based on risk with AUC of 0.927. 213
Figure 5.23 10-feature multi-omic cross-biofluid panel classifies patients based on risk with AUC of 0.909 when VHL patient is reclassified. 215
Figure 6.1 Treatment with PCF affects the proliferation of normal pancreatic cell lines. 230

Figure 6.2 Treatment with PCF for 24 h significantly increases the viability and decreases the apoptosis of H6c7-normal cells. 231
Figure 6.3 Treatment with PCF for 24 h is cytotoxic to HPNE-intermediary cells, significantly increasing apoptosis and decreasing proliferation. 232
Figure 6.4 Treatment with high-risk PCF for 24 h significantly increases the viability of H6c7-normal cells. 234
Figure 6.5 Treatment with high-risk PCF for 24 h significantly decreases the proliferation of HPNE-intermediary cells. 235
Figure 6.6 Serum starvation for 24 h significantly increases the reliance of H6c7-normal cells on oxidative phosphorylation compared to glycolysis. 237
Figure 6.7 Serum starvation for 24 h significantly increases the production of glycoATP in HPNE-intermediary cells. 238
Figure 6.8 Treatment with PCF for 24 h significantly increases the production of glycoATP in H6c7-normal cells. 240
Figure 6.9 Treatment with PCF for 24 h does not significantly alter the metabolic profile of HPNE-intermediary cells. 241
Figure 6.10 Treatment high-risk PCF for 24 h significantly increases the production of glycoATP in H6c7-normal cells. 242
Figure 6.11 Treatment with low- or high-risk PCF for 24 h does not significantly alter the metabolic profile of HPNE-intermediary cells. 243
Figure 6.12 Treatment with PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on H6c7-normal cells. 245
Figure 6.13 Treatment with PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on HPNE-intermediary cells. 246
Figure 6.14 Treatment with low- or high-risk PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on H6c7-normal cells. 247
Figure 6.15 Treatment with low- or high risk PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on HPNE-intermediary cells. 248
Figure 6.16 Treatment with PCF for 24 h significantly decreases the expression of vimentin and PD-L1 on H6c7-normal cells. 250
Figure 6.17 Treatment with high-risk PCF for 24 h significantly decreases the expression of PD-L1 on H6c7-normal cells. 251

Figure 6.18 PCF significantly decreases the percentage of Vimentin$^+$ and EGFR$^+$ H6c7-normal cells at 24 h compared to 6 h. 252
Figure 6.19 Treatment with PCF for 24 h significantly increases the expression of EGFR on HPNE-intermediary cells. 254
Figure 6.20 Treatment with low-risk PCF for 24 h significantly increases the expression of EGFR on HPNE-intermediary cells. 255
Figure 6.21 PCF significantly increases the percentage of EGFR$^+$ HPNE-intermediary cells at 24 h compared to 6 h..... 257
Figure 6.22 Treatment with PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on H6c7-normal cells. 258
Figure 6.23 Treatment with PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on HPNE-intermediary cells. 259
Figure 6.24 Treatment with low- of high-risk PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on H6c7-normal cells. 261
Figure 6.25 Treatment with low- or high-risk PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on HPNE-intermediary cells..... 262
Figure 6.26 Treatment with PCF for 24 h significantly decreases the percentage of Vim$^+$/Slug$^+$, Vim$^+$/E-cad$^+$, Vim$^+$/EGFR$^+$, Vim$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cad$^+$ and PD-L1$^+$/EGFR$^+$ H6c7-normal cells. 263
Figure 6.27 Treatment with high-risk PCF for 24 h significantly decreases the percentage of Vim$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cad$^+$ and PD-L1$^+$/EGFR$^+$ H6c7-normal cells. .. 265
Figure 6.28 Low- and/or high-risk PCF significantly decreases the percentage of Vim$^+$/Slug$^+$, Vim$^+$/E-cad$^+$ and Vim$^+$/EGFR$^+$ H6c7-normal cells at 24 h compared to 6 h. 266
Figure 6.29 Treatment with PCF for 24 h significantly increases the percentage of Vim$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cad$^+$ HPNE-intermediary cells. 268
Figure 6.30 Treatment with low-risk PCF for 24 h significantly increases the percentage of Vim$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cad$^+$ HPNE-intermediary cells. 269
Figure 6.31 PCF significantly increases the percentage of Vim$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cad$^+$ HPNE-intermediary cells at 24 h compared to 6 h. 271
Figure 6.32 Treatment with high-risk PCF for 24 h significantly increases the invasive capacity of H6c7-normal cells. 272

Figure 6.33 Treatment with high-risk PCF for 24 h significantly increases the invasive capacity of HPNE-intermediary cells. 273
Figure 6.34 8-OHdG levels are significantly increased in the supernatants of H6c7-normal cells treated with PCF for 24 h compared to the matched cyst samples. 275
Figure 6.35 8-OHdG levels are significantly increased in the supernatants of HPNE-intermediary cells treated with PCF for 24 h compared to the matched cyst samples. 276
Figure 6.36 8-OHdG levels in H6c7-normal cell supernatants are not significantly altered following 24 h treatment with PCF. 278
Figure 6.37 8-OHdG levels in HPNE-intermediary cell supernatants are not significantly altered following 24 h treatment with PCF. 279

List of Tables

Table 1.1 Indications for EUS or surgery as per the current global clinical guidelines. . 25
Table 1.2 Indications for surveillance as per the current global clinical guidelines. 26
Table 1.3 Overview of biomarkers in PC and PCLs that have been validated in an independent cohort. 29
Table 2.1 Individualised search strategies for academic databases. 57
Table 2.2 Inclusion and exclusion criteria. 59
Table 2.3 Complete list of extracted data fields. 60
Table 2.4 QUADAS-2 quality and risk of bias assessment questions. 61
Table 2.5 Reverse transcription reaction components. 74
Table 2.6 Reaction master mix setup for LNA miRNA Custom PCR panel. 74
Table 2.7 PCR cycling conditions for miRCURY LNA miRNA custom PCR panel. 75
Table 3.1 Summary of extracted data fields from included papers. 97
Table 3.2 Performance of the most frequently examined novel biomarkers. 111
Table 4.1 Patient demographic information for PCF cohorts. 130
Table 4.2 QIAGEN 24-array custom plate targets and quality controls. 132
Table 5.1 Patient demographic information for serum cohorts. 182
Table 5.2 QIAGEN 24-array custom plate targets and quality controls. 183
Table 5.3 CombiROC reduction creates a 10-feature multi-omic cross-biofluid panel. 212
Table 6.1 PCF treatments for dose response curve. 227
Table 6.2 Flow cytometry panel of fluorochrome-conjugated anti-human antibodies. 228

List of Appendices

Appendix 1. PRISMA 2020 Checklist. 356
Appendix 2. Clinical data binary code key for correlations. 360
Appendix 3. QIAGEN custom 24-array plate layouts. 361
Appendix 4. Scree plot showing the variance explained by each principal component for the 8-protein panel in PCF. 362
Appendix 5. Scree plot showing the variance explained by each principal component for the 3-miRNA panel in PCF. 363
Appendix 6. Scree plot showing the variance explained by each principal component for the 11-feature multi-omic panel in PCF. 364
Appendix 7. Scree plot showing the variance explained by each principal component for the 8-protein panel in the serum. 365
Appendix 8. Scree plot showing the variance explained by each principal component for the 5-miRNA panel in the serum. 366
Appendix 9. Scree plot showing the variance explained by each principal component for the 13-feature multi-omic panel in the serum. 367
Appendix 10. qPCR microarray does not validate transcriptomic results in the serum. 368

AGA	American Gastroenterological Association
AIC	Akaike information criterion
AUC	Area under the curve
BD	Branch-duct
BIC	Bayesian information criterion
CA125	Cancer antigen 125
CA19-9	Carbohydrate antigen 19-9
CA242	Carbohydrate antigen 242
CAF	Cancer-associated fibroblast
CDO1	Cysteine dioxygenase 1
CEA	Carcinoembryonic antigen
CHIP	Carboxyl terminus of heat shock protein 70-interacting protein
CI	Confidence interval
CPM	Counts per million
CRC	Colorectal cancer
CSC	Cancer stem cell
CT	Computed tomography
dH_2O	Deionised H_2O
DMEM	Dulbecco's Modified Eagle's Medium
ECM	Extracellular matrix
EDTA	Ethylene diamine tetra acetic acid
EGFR	Epidermal growth factor receptor
EMT	Epithelial to mesenchymal transition
European	European evidence-based
EUS	Endoscopic ultrasound
EZH2	Enhancer of zeste homologue 2
FBS	Foetal bovine serum
FNA	Fine-needle aspiration
FSC-A	Forward scatter area
FSC-H	Forward scatter height
Fukuoka	International Association of Pancreatology Fukuoka
GWAS	Genome-wide association study
IPMN	Intraductal papillary mucinous neoplasm
LC-MS	Liquid chromatography-based mass spectroscopy
LC-MS	Liquid chromatography-mass spectrometry
Le	Lewis blood group
LFQ	Label-free quantification
lncRNA	Long non-coding RNA
LOOCV	Leave-one-out cross validation
LRG1	Leucine-rich-alpha-2-glycoprotein
MCIA	Multiple co-inertia analysis
MCN	Mucinous neoplasm
MD	Main-duct
MDSC	Myeloid-derived suppressor cells
miRNA	microRNA
MMP	Matrix metalloproteinases

MOFA	Multi-omics factor analysis
MRCP	Magnetic resonance cholangiopancreatography
MRI	Magnetic resonance imaging
MS	Mass spectroscopy
NGS	Next generation sequencing
NPP	Nuclease protection probe
NSCLC	Non-small cell lung cancer
OIS	Oncogene-induced senescence
OV	Ovarian cancer
PBS	Phosphate Buffer Saline
PC	Pancreatic cancer
PCA	Principle component analysis
PCF	Pancreatic cyst fluid
PCL	Pancreatic cystic lesion
PCR	Polymerase chain reaction
PD	Pancreatic duct
PD-1	Programmed cell death protein 1
PDAC	Pancreatic ductal adenocarcinoma
PD-L1	Programmed death-ligand 1
PSC	Pancreatic stellate cell
QC	Quality control
RoB	Risk of bias
RT	Room temperature
SASP	Senescence-associated secretory phenotype
SCA	Serous cystadenoma
SEM	Standard error of the mean
SN	Sensitivity
SNP	Single nucleotide polymorphism
SP	Specificity
SPN	Solid-pseudopapillary neoplasm
SSC-A	Side scatter area
TAM	Tumour-associated macrophage
TBE	Tris borate-EDTA
TCGA	The Cancer Genome Atlas
TIMP	Tissue inhibitor of matrix metalloproteinases
TME	Tumour microenvironment
T_{reg}	T regulatory
UTC	Untreated control
VHL	Von Hippel-Lindau

Units

× g	acceleration due to gravity
°C	degrees Celsius
µg	micrograms
µL	microlitres
µm	micrometres
µM	micromolar
bp	base pairs
cm	centimetres
cm^2	centimetres squared
g	grams
h	hours
kV	kilovolts
L	litres
M	molar
mg	milligrams
min	minutes
mL	millilitres
mm	millimetres
mM	millimolar
ng	nanograms
nL	nanolitres
nm	nanometres
RPM	revolutions per minute
sec	seconds
Th	Thomson
U	units
V	volts
v/v	volume per volume
w/v	weight per volume

Chapter 1.

Introduction

Published in part in:

Kane, L.E., Mellotte, G.S., Conlon, K.C., Ryan, B.M. and Maher, S.G., 2021. Multi-Omic biomarkers as potential tools for the characterisation of pancreatic cystic lesions and cancer: Innovative patient data integration. *Cancers*, 13(4), p.769.

1.1 Pancreatic Cancer

1.1.1 Overview

Pancreatic cancer (PC) has the worst 5-year survival rate of any cancer, at just 12% for 2023[1]. PC is the third leading cause of cancer-related mortality in both men and women combined, with incidence rates continuing to rise every year[1, 2]. PC can be divided into two main subtypes: pancreatic ductal adenocarcinoma (PDAC), which is responsible for 85-90% of all pancreatic neoplasms and has a 5-year survival rate of roughly 10%, and pancreatic neuroendocrine tumour, which is far less common and represents less than 5% of PC[3-6]. As PDAC is by far the most prevalent type of PC, it is used synonymously with PC throughout the literature. In order to avoid generalisations, where the subtype of PC is specified in the literature, it is reported here.

1.1.2 Clinical presentation

The poor survival rates seen in PC are attributed not only to the highly aggressive nature of the cancer, and the lack of immune cell infiltrate making it immunologically 'cold', but also to the very late stage of clinical detection[7]. Indeed, the symptoms associated with PC are notoriously vague, issues such as bloating, unexplained weight loss or back pain, with patients experiencing no symptoms at early stage disease[8, 9]. As such, the majority of PC patients tend to present clinically at a late stage of cancer development, limiting the treatment options available and further compounding the problem of poor survival in this cancer. Of critical importance, therefore, is the advent of novel modalities of early detection, which could be key to improving survival of PC patients.

1.2 Development of PC and the tumour microenvironment

PC generally arises from one of several histologically distinct precursor lesions in the pancreas[10]. These precursor lesions include certain subtypes of pancreatic cystic lesions (PCLs) such as intraductal papillary mucinous neoplasms (IPMNs) or mucinous cystic neoplasm (MCN), which are discussed further in section 1.4, as well as pancreatic intraepithelial neoplasia (PanIN) lesions[10, 11]. While PanIN are the most frequently occurring precursor lesion for PDAC, these microscoping lesions are not detectable via

radiological or endoscopic ultrasound examination, and are characteristically asymptomatic, thus hampering the early detection of patients with these precursor lesions[12, 13]. A simulation model using data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program, however, estimated PanIN to have a relatively indolent course to PDAC development, with the probability of progressing from low-grade PanIN 1 lesions to PC over the course of a lifetime being just 1.3% for females and 1.5% for males[14]. Unfortunately, given our inability to effectively screen for PanIN lesions via imaging, the prevalence and subsequent rate of malignant transformation of these precursor lesions in the population is unknown.

Tumorigenesis typically involves the progressive development of driver mutations. In the pancreatic setting, this often includes mutations in the *KRAS* oncogene (93% of PDAC tumours) and the *TP53* tumour suppressor gene (73% of PDAC tumours)[15-17]. These genetic mutations generally occur alongside the development of precursor lesions, and are therefore present at their subsequent progression through increasing histological grades, culminating in invasive carcinoma (Figure 1.1)[18, 19]. A unique feature in the stepwise progression of PC is the desmoplastic reaction to the tumour, which is present in both primary and metastatic PC tumours[20]. This desmoplasia, often termed fibrosis, is a common reaction in normal tissues during the wound healing process, and can aid in the reduction of inflammation and granulation tissue[20]. While this desmoplasia, characterised by a dense stromal microenvironment which limits cell motility, is present in both chronic pancreatitis and PC tissues, in PC desmoplasia has been shown to promote tumour development while inhibiting the penetration of anti-tumour therapies into the tumour microenvironment (TME) (Figure 1.2)[21]. As with any TME, the tissues, cells and factors within this environment are highly heterogeneous and allow for a complex and diverse web of interactions. Of particular importance in the pancreatic TME, are the stromal cells responsible for this desmoplastic reaction, which comprise cancer associated fibroblasts (CAFs), endothelial cells and immune cells[22]. The stromal tissue in which PC cells become embedded, accounts for as much as 80% of the pancreatic tumour volume and consists of an extracellular matrix (ECM) of fibrous proteins such as collagens, fibronectin and glycoproteins[23]. The combination of these cellular components, their respective

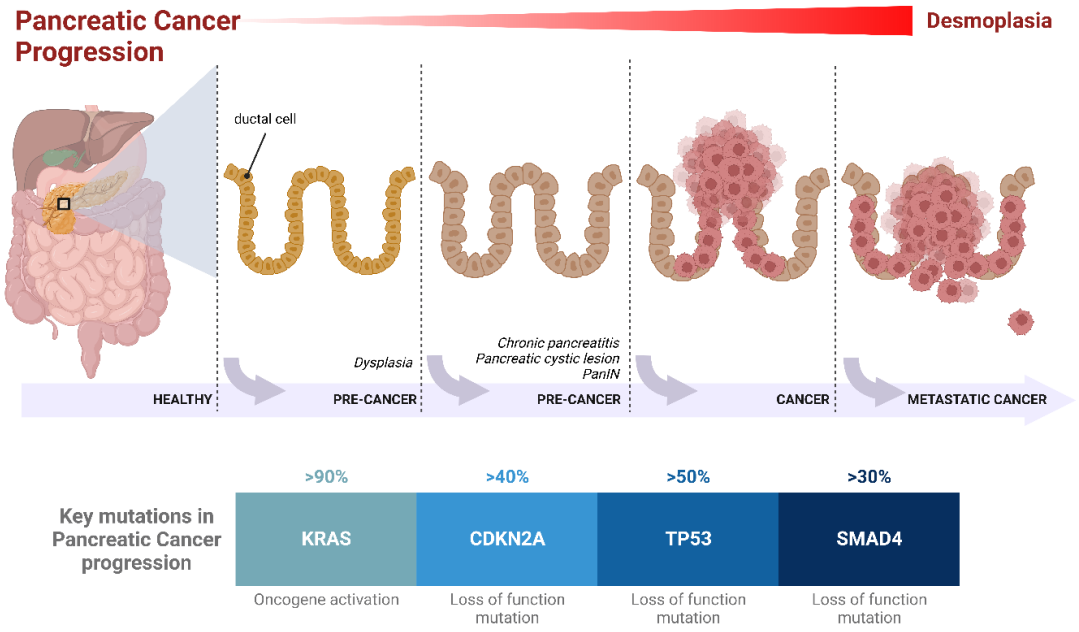


Figure 1.1 Progression of PC. The progression of PC is driven by four key mutations, *KRAS*, *CDKN2A*, *TP53* and *SMAD4*, from pre-cancer to metastatic cancer. These mutations occur in >90%, >40%, >50% and >30% of PC tumours, respectively. Pancreatic tissue progresses step-wise from low-grade dysplasia (pre-cancer) to metastatic cancer while accumulating these genetic mutations. The level of desmoplasia increases with PC progression, with metastatic tumours having the highest level of desmoplasia.

secreted factors, and the dense stromal tissue creates a hypoxic and nutrient deprived TME, which further contributes to cancer progression[24, 25]. This, combined with the immunologically 'cold' TME of PC, and our inability to accurately replicate this dense stromal TME, has negatively impacted the development of novel therapeutics for PC treatment[26, 27].

Pancreatic stellate cells (PSCs) are a subset of CAFs that are responsible for producing the collagenous stroma in PC, and also in chronic pancreatitis[24, 25]. PSCs are resident phagocytic cells of the pancreas that are important players in the normal turnover of the ECM, producing proteins such as collagen I-IV, fibronectin and laminin[21, 28]. Given the role of PSCs in the desmoplastic reaction in PC, the overlap in gene expression seen in PC and chronic pancreatitis tissues, and the increased risk of PC in chronic pancreatitis patients, the role of PSCs in PC progression has been under much scrutiny in recent years[25, 29]. The detection of activated PSCs expressing α-smooth muscle actin, periostin and galectin-1 in PCLs, and chronic pancreatitis tissues has led to the suggestion that PSCs are an early event in PC[21, 30, 31]. Furthermore, PSCs have been shown to interact with PC cells in a bi-directional manner, where each cell type promotes the survival and proliferation of the other[21]. They have also been shown to travel from the primary tumour to distant metastatic sites, suggesting a role in the metastatic cascade[32]. Two subsets of PSCs have been described in PC tumours due to their distinct secretion profiles[33]. Those PSCs that are found in close proximity to cancer cells have been termed myofibroblasts, while those residing at some distance from cancer cells have been called inflammatory PSCs or iPSCs (Figure 1.2A)[33].

In the context of the pancreatic TME, there are a plethora of factors and cell types to be found within its confines which play unique roles in PC development and progression. Some of those most important factors and interactions are discussed below.

1.3 Hallmarks of PC

In 2000, Douglas Hanahan and Robert Weinberg published a seminal paper entitled 'The Hallmarks of Cancer', where they elegantly discussed a set of traits unique

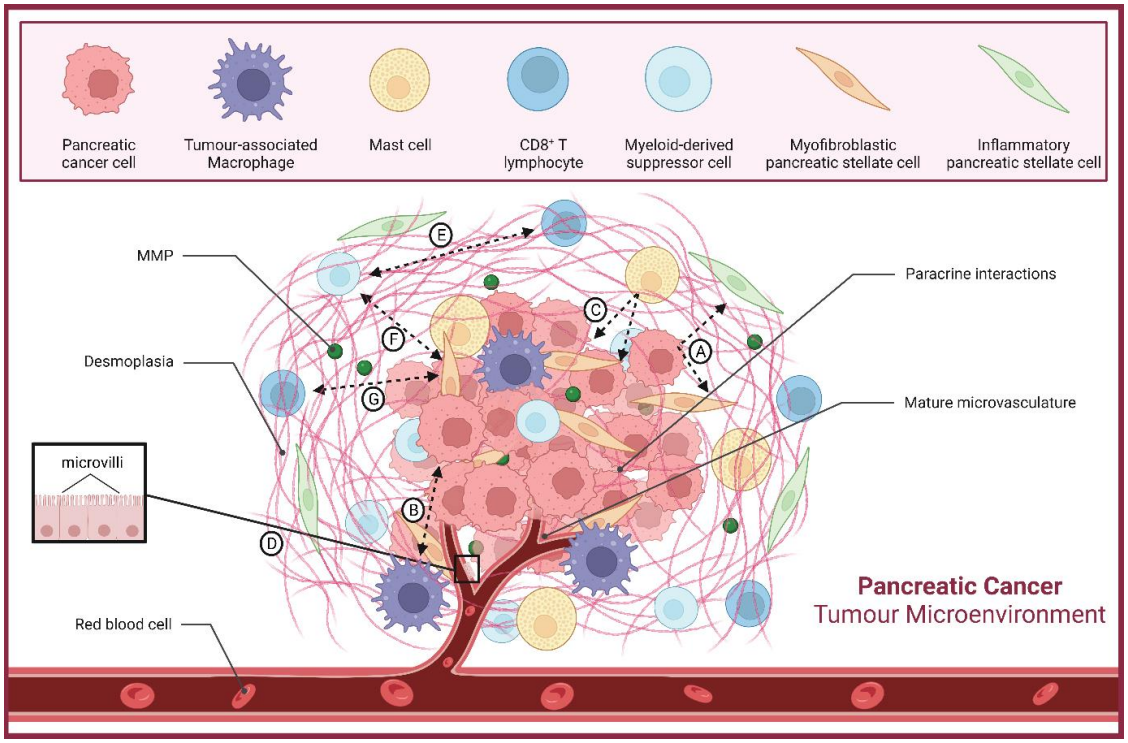


Figure 1.2 The Pancreatic TME. The unique structure of the pancreatic TME is a result of the various cells and factors within it, and their interconnected interactions. (**A**) Aberrant TGF-β signalling by myofibroblastic and inflammatory PSCs promotes PC cell growth, motility, invasion and metastasis. PC cells secrete IL-1α to induce production of VEGF-A and IL-6 by PSCs and other CAFs. PC cells also release CD147 to stimulate the production of MMPs by PSCs. (**B**) PC cells secrete factors to recruit tumour-associated macrophages (TAMs). TAMs produce cytokines and growth factors to stimulate the proliferation of PC cells. (**C**) Mast cells are recruited via VEGF secreted by PCSs. Mast cells promote PC cell proliferation via secretion of tryptase and IL-13. Tryptase and IL-13 activate PSC proliferation via TGF-β signalling. (**D**) The presence of basal microvilli on mature microvasculature within the TME facilities glucose uptake by PC cells. (**E**) MDSCs deplete lymphocyte-required nutrients from the environment and subsequently suppress $CD8^+$ T cell cytotoxic functions. MDSCs also express ADAM17 to interfere with lymphocyte trafficking. (**F**) PSCs stimulate recruitment of MDSCs to PC via IL-16. (**G**) PSCs secrete CXCL12 causing migration of T cells and subsequent sequestration in the stromal tissue.

to cancer cells, which could therefore be designated as hallmarks of the disease[34]. These six original traits are (1) sustaining proliferative signalling; (2) evading growth suppressors; (3) resisting cell death; (4) enabling replicative immortality; (5) inducing or accessing vasculature; and (6) activating invasion and metastasis. In 2011, these characteristics were revisited in 'Hallmarks of Cancer: the next generation', with two new hallmarks and two 'enabling characteristics' (EC) being induced into the mix to give a total of ten features[35]. The new hallmarks were (7) deregulating cellular metabolism; (8) avoiding immune destruction; (9) genome instability and mutation (EC); and (10) tumour-promoting inflammation (EC). In 2022, Douglas Hanahan revised these guidelines for a third time, publishing 'Hallmarks of Cancer: new dimensions'[36]. Here, the ECs from the 2011 version were given full hallmark status, while two new hallmarks and two more ECs were added to give a final total of 14 features. The newest hallmarks of cancer are (11) unlocking phenotypic plasticity; (12) senescent cells; (13) non-mutational epigenetic reprogramming (EC); and (14) polymorphic microbiomes (EC). Together these 14 hallmarks of cancer provide insights into the mechanisms involved in cancer development and progression, and highlight specific characteristics that are unique to the disease. Below, these hallmarks in the context of PC and its unique characteristics are discussed (Figure 1.3).

1.3.1 Sustaining proliferative signalling

All cancer cells have the ability to divide and multiply indefinitely[34]. This acquired characteristic of tumour cells is a result of genetic mutations that cause the activation of oncogenes, which promote growth and survival, and the deactivation of tumour suppressor genes, which enable limitless cell division. As mentioned in section 1.2, common mutations in PC occur in the oncogene *KRAS* and the tumour suppressor gene *TP53*[15-17]. Activation of *KRAS* is an early event in PC and is known to play a large role in the sustained proliferation of PC cells, as well as in the processes of transformation, survival and invasion[37]. Indeed, *KRAS* activation generally results in the subsequent activation of downstream pathways such as the NF-κB pathway, the PI3K/AKT/mTOR pathway and the MAPK pathway[37]. Increased activation of the AKT pathway is seen in 50% of PDAC cases and promotes sustained PC cell proliferation[38]. Overexpression of

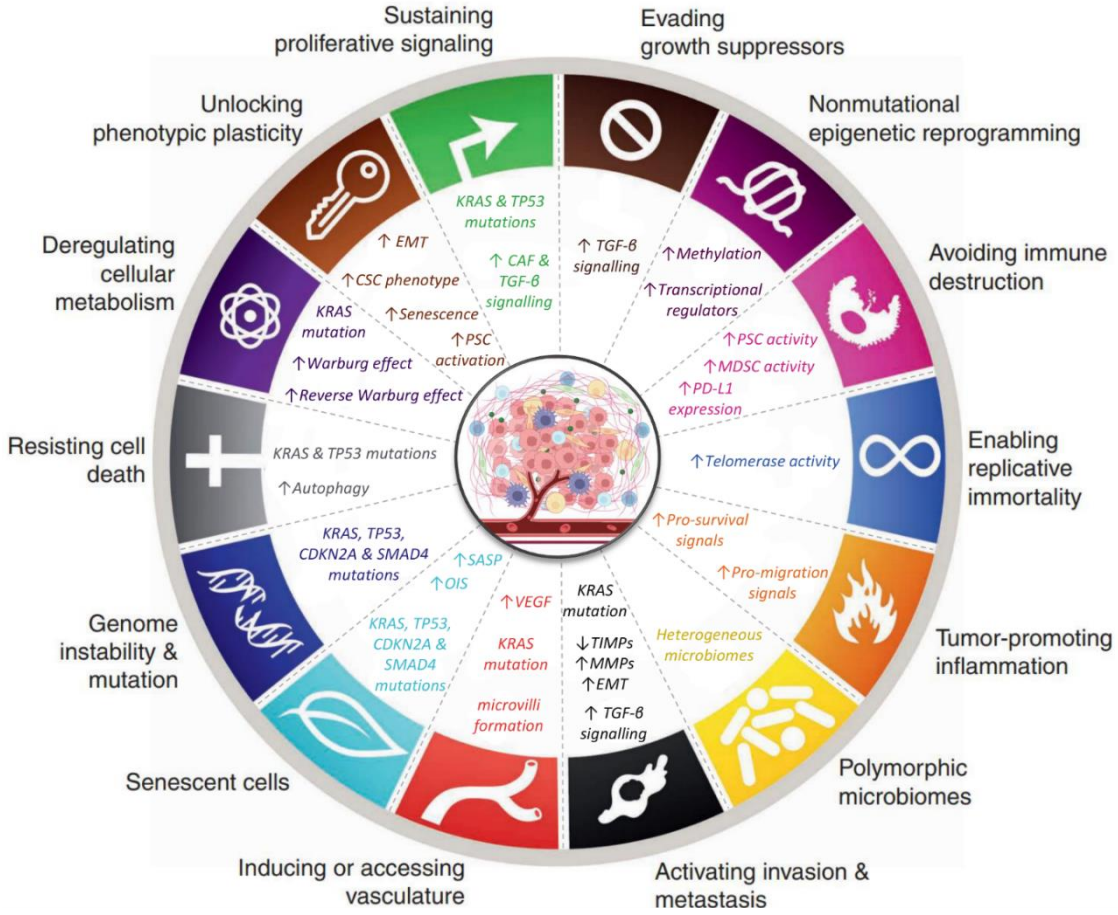


Figure 1.3 The Hallmarks of PC. The specific characteristics of PC in the context of the fourteen hallmarks of cancer, sustaining proliferative signalling; evading growth suppressors; resisting cell death; enabling replicative immortality; inducing or accessing vasculature; activating invasion and metastasis; deregulating cellular metabolism; avoiding immune destruction; genome instability and mutation; tumour-promoting inflammation; unlocking phenotypic plasticity; senescent cells; non-mutational epigenetic reprogramming; and polymorphic microbiomes, are provided. Important events related to each hallmark in the pancreatic setting are listed within the relevant segments, and discussed in detail in sections 1.3.1 – 1.3.14. *Adapted from Hanahan 2022*[36].

TGF-β in PC has also been shown to directly induce PC cell proliferation via the MAPK and AKT pathways[8]. Similarly, *TP53* mutation can occur in as many as 79% of PDAC cases, and this loss-of-function mutation can allow the uncontrolled proliferation of PC cells[39, 40]. Indeed, studies have shown that the induction of high wild-type *TP53* expression in PC cells, both *in vitro* and *in vivo*, can inhibit tumour cell proliferation and apoptosis, thus making *TP53* a potential therapeutic target in PC[41, 42].

Cells within the TME have also been shown to be important players in the uncontrolled proliferation of tumour cells. Indeed, in the early stages of tumour development, tumour-associated macrophages (TAMs) within the PC TME produce cytokines and growth factors which sustain the proliferation of tumour initiating cells (Figure 1.2B)[43]. These cells then secrete factors to recruit more TAMs to the TME, causing a vicious cycle of signalling and growth in favour of cancer initiation[43]. Mast cells are also recruited to the TME through VEGF, stem cell factor, and FGF-2 secretion by PC cells, which in turn promote the proliferation of cancer cells via secretion of factors such as tryptase and IL-13 (Figure 1.2C)[44, 45]. Mast cell-produced IL-13 and tryptase have also been shown to activate proliferation of PSCs via the TGF-β and STAT6 pathways[45]. These PSCs are also key contributors towards the proliferative signalling of cancer cells, as mentioned in section 1.2, where the bi-directional signalling of PSCs with PC cells promotes the proliferation and survival of each cell[46].

1.3.2 Evading growth suppressors

The loss of negative growth constraints in PC can be primarily linked to aberrant TGF-β signalling (Figure 1.2A)[47]. TGF-β is known to promote the progression of many tumour types, such as PC, via paracrine effects within the TME, and is secreted at high levels by PSCs in the PC TME[46, 48]. Indeed, PSC-secreted TGF-β has been shown to cause a more aggressive phenotype in PDAC cells[48]. In early PC, TGF-β acts as a tumour suppressor gene, inhibiting the growth of epithelial cells[49]. However, as the tumour develops, the response to TGF-β switches, with tumour cells utilising this aberrant expression as a potent promotor of cell growth, motility, invasion and metastasis[50].

1.3.3 Resisting cell death

The tumour suppressor *TP53* is a key component in the cellular response to DNA damage in PDAC[51]. Mutations in this gene, therefore, result in the disruption of DNA repair machinery and apoptotic signals. Indeed, mutations in *TP53* have been shown to increase PDAC cell resistance to chemotherapy[52]. Oncogenic *KRAS* mutation in PC has also been shown to upregulate autophagy in tumour cells, enabling proliferation and inhibiting apoptosis in these cells[53]. The activation of downstream anti-apoptotic and pro-survival pathways, such as the STAT3, NF-κB and AKT, pathways, are also important processes in the development of PC[54, 55].

1.3.4 Enabling replicative immortality

Cancer cells have limitless replicative potential[34]. Unlike normal cells, which see their telomeres shortened with each cellular division until a limit is reached and cell death occurs, cancer cells acquire the ability to circumvent this process. This is achieved by either upregulating the expression of the telomerase enzyme such that telomere length is maintained just over a critical threshold, or by extending the length of their telomeres via recombination[56]. Telomere shortening is associated with a higher risk of many cancers, and has been reported in both PC and PCL tissues[56-58]. However, the maintenance of longer telomeres may mean overactive telomerase activity, with longer telomeres in peripheral blood leukocytes having been shown to be associated with an increased risk of PC development[59]. As such, current research into the utilisation of telomere length as a potential biomarker or therapeutic target requires further elucidation of the mechanisms involved in PC.

1.3.5 Inducing or accessing vasculature

Angiogenesis is the process by which new capillaries are formed, as branches from existing vasculature, towards the tumour in order to supply nutrients and oxygen to aid growth and metastasis[34]. The activation of pro-angiogenic molecules is essential for this process, with VEGF-A being a key factor in most cancers, including PC[60]. In PC, angiogenesis is induced by both genetic and epigenetic alterations, such as *KRAS*

mutation, as well as by tumour and stromal cells such as PSCs within the TME[37, 61]. PSC are known to express several angiogenic-regulating molecules, such as Tie-2, angiopoietin-1 and VEGF; and hypoxia has been shown to induce the expression of VEGF in PSCs[62]. Importantly, the unique presence of basal microvilli on mature microvasculature within pancreatic tumours facilities the excessive glucose uptake observed in PC cells (Figure 1.2D)[63]. Despite this, PC tumours are known to be hypovascular, as the desmoplastic TME causes vascular collapse and thus inhibits drug penetration to the tumour[20]. As such, anti-angiogenic therapy in PC has not proven to be successful[61].

1.3.6 Activating invasion and metastasis

The degradation of the ECM is essential for the invasion and metastasis of tumour cells[64]. Matrix metalloproteinases (MMPs) are proteolytic enzymes that are key to this degradation process, and their activity is dependent on the levels of activated MMPs and tissue inhibitors of matrix metalloproteinases (TIMPs) present[64, 65]. In PC, overexpression of MMP2, MMP7, MMP8, MMP9, MMP11, and TIMP-3, alongside concurrent decreases in TIMP-2 expression, have been associated with aggressive phenotypes and poor survival[64, 66]. Elevated MMP2, MMP9 and TIMP-2 expression in PC tumours in particular have been associated with liver metastasis in patients[67]. MMP2 has also been implicated in the perineural invasion of PC cells, as exposure to nerve growth factor, a neurotrophic factor secreted by intra- and extra-pancreatic neural elements, has been demonstrated to stimulate MMP2 expression and increase the invasive potential of PC cells[68]. Perineural invasion is an important and often overlooked method of local invasion that has been observed to be a critical factor in local tumour recurrence[68, 69].

PSCs have been shown to be an important source of both MMPs and TIMPs in the pancreatic TME, with PC cells releasing the glycoprotein CD147 to stimulate the production of MMP2 in these PSCs (Figure 1.2A)[21, 70, 71]. However, both stromal and tumour cells have been shown to be sources of MMPs and TIMPs in PC[72]. Importantly, MMP2 expression has been found to be strongly associated with the extent of the desmoplastic reaction seen in PC tissues[73]. Despite the levels of involvement of MMPs

in the PC TME, efforts in clinical trials to target MMPs in PC via MMP inhibitors both alone and in combination with chemotherapy have been unsuccessful[74]. While these results have been discouraging, issues such as timing of delivery, dosage, and non-selectivity of the inhibitor are among those aspects that could be improved upon in future studies[74].

Also key to tumour cell invasion and metastasis is the process of epithelial to mesenchymal transition (EMT). The downregulation of epithelial markers, such as E-cadherin or EpCAM, coupled with the upregulation of mesenchymal markers and markers of invasion, such as N-cadherin, Vimentin, Slug or Twist, are typical features acquired by cells during EMT[50]. In PC, *KRAS* mutation and TGF-β signalling have been shown to be important in this process[50]. Indeed, the activation of TGF-β signalling in PC is a critical early step in enabling EMT initiation, causing the downregulation of adhesion molecules while enabling the acquisition of a more mesenchymal phenotype, and upregulating the production of MMPs[75]. TGF-β has been shown to stimulate the expression of N-cadherin and Vimentin in PC cells *in vitro*, while simultaneously decreasing E-cadherin expression[76]. *In vivo*, Slug expression was found in 50% of PC tumour tissues, with Slug expression being observed predominantly in the invasive front of the tumour[77].

Lastly, cancer stem cells (CSCs) represent the final phase of invasion and metastasis. The acquisition by cancer cells of stem-like traits is key to the successful formation of secondary metastatic tumours, as CSCs are widely regarded as the tumour-initiating cell population[35]. CSC markers, such as CD133, CD44 and CD24, have been identified on the surface of CSC populations within many solid tumours, including PC[78-80]. A 2015 systematic review and meta-analysis highlighted the significant correlation between CD133 and prognosis, TNM stage (describing the primary tumour type (T), whether the cancer has spread to nearby lymph nodes (N), and whether the cancer has metastasised to other parts of the body (M)), tumour differentiation, and lymph node status[81]. Indeed, PC tumours with $CD133^+$ CSC populations were also shown to be highly resistant to chemotherapy[80]. Using xenograft models of primary human PDAC tumours in immunocompromised mice, it was demonstrated that as few as 100

pancreatic CSCs were required to form a tumour that was histologically indistinguishable from the tumour of origin[79].

1.3.7 Deregulating cellular metabolism

The dysregulation of cellular metabolism emerged as a hallmark of cancer in the second iteration of the seminal paper by Hanahan and Weinberg in 2011[35]. The tendency for tumour cells to metabolise glucose anaerobically instead of aerobically, even in the presence of oxygen, termed the Warburg effect, is a known characteristic of cancer cells[82]. In PDAC, activation of oncogenic *KRAS* causes an upregulation of glucose transporter 1, hexokinase 1/2, phosphofructokinase 1, and lactate dehydrogenase A, which promotes glycolysis[83]. However, PC cells can alter their metabolism in order to adapt to different environments, utilising alternative metabolic methods such as amino acid and lipid metabolism, among others [84]. Indeed, it has long been understood that metabolic pathways are not mutually exclusive, and as such individual tumour cells within the same tumour can utilise alternative methods of ATP production as needed[85, 86]. The paradox, however, of why cancer cells would opt to produce ATP in a less efficient manner when oxygen is available, remains to be understood[85].

In many caners, the cross-talk between tumour cells and CAFs demonstrate a new metabolic phenotype known as the 'reverse Warburg effect', whereby reactive oxygen species are produced by tumour cells and induce oxidative stress in adjacent CAFs[86]. This causes the CAFs to shift towards aerobic glycolysis and as such produce pyruvate, fatty acids and lactate, fuel which feeds oxidative phosphorylation in tumour cells, thus allowing efficient metabolic processes in these cells[85].

1.3.8 Avoiding immune destruction

Evading immune recognition is achieved in part through the secretion of factors produced by both tumour cells and stromal accessory cells, such as growth factors, chemokines and cytokines[87]. The secretion of these inflammatory factors can promote the differentiation of myeloid-derived suppressor cells (MDSCs), as well as their subsequent trafficking into the TME[88]. In the cancer setting, myeloid cell differentiation

can be dysregulated, causing the terminal differentiation of mature macrophages, dendritic cells and granulocytes into immunosuppressive populations of MDSCs[89]. MDSCs can actively supress the abilities of cytotoxic lymphocytes such as T cells and natural killer cells via the depletion of lymphocyte-required nutrients; the generation of oxidative stress (which drives molecular blocks in T cells); interference with lymphocyte trafficking via expression of ADAM17; and the activation and expansion of T regulatory (T_{reg}) cell populations, which can subsequently supress both the innate and adaptive immune response (Figure 1.2E)[87, 89]. $CD4^+$ T_{reg} cells in the stroma are of particular importance in PC, as they have been shown to play a crucial role in warding off the host immune system[8]. Interestingly, the presence of immunosuppressive tumour-associated macrophages, MDSCs and T_{reg} cells has been observed in the earliest stages of PCL development, and has dominated the immune cell infiltrate of the pancreatic microenvironment throughout the subsequent development of invasive carcinoma[90].

PSCs have been shown to stimulate the migration of MDSCs to the tumour via secretion of IL-16 in a STAT3-dependent manner, allowing these MDSCs to exert their immunosuppressive functions in the TME, further aiding PC cell immune evasion (Figure 1.1F)[91]. Activated PSCs have also been shown to secrete CXCL12, a chemokine known to regulate T cell migration, and subsequently cause the sequestration of $CD8^+$ T cells in the surrounding stroma of PC tumours, preventing the invasion of these cells peri-tumourally and therefore hampering their anti-tumour capabilities (Figure 1.2G)[23, 92]. Indeed, in a 2013 study, PDAC patients with higher $CD8^+$ T cell tumour infiltrates had longer survival times than those with lower densities.[92]

Other major factors in the immunosuppressive TME are the checkpoint molecule programmed cell death protein 1 (PD-1) and its ligand, programmed death-ligand 1 (PD-L1). PD-L1 can be expressed by tumour cells or tumour-associated immune cells, and can induce apoptosis or exhaustion of T cells via binding to the PD-1 receptor[93, 94]. High expression of PD-L1 has been correlated with poor prognosis in many cancers, including PC[95]. PD-L1 expression on tumour cells is a key mechanism of immune evasion in PC, as it can enable tumour cells to escape immune surveillance[96]. In PC, PD-L1 expression has been shown to be upregulated via NF-κB signalling, specifically tumour-infiltrating macrophage-derived TNF-α[97].

1.3.9 Genome instability and mutation

While in the original emergence of 'genome instability and mutation' in the 2011 iteration of the hallmarks of cancer, this capability of tumour cells was defined as an 'enabling characteristic', the more recent 2022 paper by Hanahan sees this feature given full hallmark status[35, 36]. It is largely recognised that the ability of tumour cells to increase their mutagenicity by compromising DNA repair pathways and increasing sensitivity to mutagenic agents is ubiquitous across all cancer types, and PC is no different. While the rate of genomic mutations in PC is much lower than cancers such as lung or melanoma, that are driven by strong environmental mutagens, four predominant mutational signatures have been identified in PDAC tumours[16, 98]. These are associated with (1) ageing; (2) apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like family of cytidine deaminases; (3) BRCA; and (4) mismatch repair mutations, the most prevalent of which are the ageing, which includes single-base substitution (SBS) signatures 1 and 5, and apolipoprotein B signatures as they are found in almost all PDAC cases[16, 99]. A 2023 study highlighted that SBS8 is also a dominant mutational signature in PDAC, however, the eitiology of this signature is unknown[100]. Aside from mutational signatures, individual gene mutations in PDAC have also been shown to be highly prevalent and have potential therapeutic utility. *KRAS*, *TP53*, *CDKN2A*, and *SMAD4* are among the many frequently mutated genes observed in PC, with activating *KRAS* mutations being the most prevalent and found in 93% of PC tumours[17]. The inactivating mutations to *TP53*, *CDKN2A* and *SMAD4* are present in 50-80% of PC cases[8]. Crucially, when *KRAS* mutations are not found, other ras pathway mutations can be seen in 60% of cases, highlighting the importance of this pathway in particular in PC[15]. Aberrant *KRAS* activation causes an increase in production of the K-ras protein, which can in turn activate intracellular signalling pathways and transcription factors that induce cell proliferation, transformation, migration and survival[37]. As such, many attempts have been made to target *KRAS* or downstream components of this pathway, but to little avail. Indeed, two major obstacles in the targeting or *KRAS* are its renowned reputation as an undruggable molecule, and the persistence of escape mechanisms that are independent of *KRAS* inactivation[37].

1.3.10 Tumour-promoting inflammation

Similar to the characteristic of 'genome instability and mutation' mentioned in section 1.3.9, tumour-promoting inflammation has also recently been promoted from an enabling characteristic of cancer, to a hallmark of cancer[36]. The corruption of anti-tumour immune cells to a pro-tumour phenotype, which subsequently secrete pro-migration and pro-survival factors into the TME, is a mechanism by which cancer cells hijack the host immune system to promote tumour growth and metastasis. In PC, the production of tumour cell-associated IL-1α induces the production of inflammatory factors such as VEGF-A and IL-6 by CAFs[101]. Indeed, IL-1α expression was shown to correlate with the clinical outcome of PDAC patients, and the inhibition of IL-1 activity diminished the production of inflammatory factors in patient-derived PDAC and CAF cell line co-cultures, suggesting that cross-talk between PDAC and CAF cells is essential for the formation of the inflammatory TME in PDAC[101].

1.3.11 Unlocking phenotypic plasticity

One of two new emerging hallmarks of cancer from the 2022 paper by Hanahan, is the characteristic of unlocking phenotypic plasticity[36]. A well-known acquired capability of tumour cells, the ability to disrupt the normal pathway of cellular differentiation in order to escape from terminal differentiation, requires a phenotypic plasticity that is critical in the process of carcinogenesis[36]. EMT, the process by which cells transition from an epithelial phenotype to a more mesenchymal-like phenotype, enabling tumour cell movement through tissues and circulatory systems as discussed in section 1.3.6, is a process which involves cellular plasticity[102]. CSCs are a subpopulation of tumour cells with unique differentiation and dedifferentiation capabilities, exhibiting phenotypic plasticity as they undergo processes of EMT, invasion and metastasis (section 1.3.6)[102]. Unique to PC, is the plasticity exhibited during the transition of PSCs from their quiescent state, where they display abundant cellular vitamin A-containing lipid droplets in their cytoplasm, to an activated states where they lose these vitamin A reserves and shift to a contractile and secretory phenotype (myofibroblast phenotype), and produce large amounts of MMPs and TIMPs[32]. Another form of phenotypic

plasticity observed within the TME, is the process of cellular senescence, discussed below.

1.3.12 Senescent cells

The second emerging hallmark of cancer that was recently introduced by Hanahan (2022) is senescent cells[36]. In the presence of cellular stress, such as nutrient deprivation or DNA damage, cells can enter a senescent state where they enter proliferative arrest and activate a senescence-associated secretory phenotype (SASP), releasing cell-dependent proteins and proteases into their environment[36]. This process is undergone by normal cells in the face of stress instigators, as well as by cancer cells as a result of oncogene-induced hyperactive signalling or proliferation, or DNA damage resulting from treatment with chemotherapy or radiotherapy. In cancer, senescence can limit tumour progression as cells enter this state of dormancy, undergoing biochemical, morphological and functional changes, despite remaining metabolically active[103, 104]. However, senescent tumour cells have also been shown to contribute to malignant progression via the activation of SASP, releasing signalling molecules into the TME that promote hallmark capabilities of surrounding tumour cells such as VEGF, interleukins, MMPs, cytokines and chemokines[104, 105]. Importantly, different mechanisms of cellular senescence have been identified in different cell types within the TME, such as fibroblasts, epithelial cells, endothelial cells and tumour cells[106-109]. Furthermore, the type of senescence induced is directly related to the stressor that induces it[106]. In PC, senescence is known to be an early event, with evidence of senescent cells being found in low-grade PCLs in both mice and humans[110]. Furthermore, the four main mutations that accumulate in PC (*KRAS*, *TP53*, *CDNK2A*, and *SMAD4*) are known to be important in the process of senescence, highlighting the substantial involvement of senescence from early PCL to invasive carcinoma[106]. Specifically, *KRAS* mutation has been shown to trigger oncogene-induced senescence (OIS) in PC cells, and this OIS is further enhanced by simultaneous *TP53* or *CDKN2A* mutation[111]. As well as PC cells, other cells within the PC TME such as PSCs, CAFs, T_{reg} cells, effector T cells and NKs have also been shown to undergo senescence and subsequently aid tumour progression via SASP[106]. Critically, while senescence was

originally thought to be an irreversible process undergone by cells, it is now known that senescence bypass via inactivation of tumour suppressor genes is possible, with such processes having been seen in PCLs whereby pancreatitis-induced inflammation inhibits OIS[112].

1.3.13 Non-mutational epigenetic reprogramming

Complimentary to the enabling characteristic-turned-hallmark of cancer known as genome instability, mentioned in section 1.3.9, is the widely recognised idea that epigenetic changes to gene expression can also reprogramme the genome and thus contribute to tumour cell acquisition of hallmark capabilities[36]. This non-mutational epigenetic reprogramming of the genome has recently emerged as one of two new enabling characteristics of the hallmarks of cancer[36]. Epigenomic changes in the tumour have been widely attributed to several common TME properties, such as hypoxia, nutrient deprivation and acquisition of an EMT-phenotype, and can take the form of DNA methylation, histone modifications or changes in chromatin structures[36, 113]. In PC, epigenetic modifications as a result of transcriptional regulators such as USP22 and FAK are thought to play a major role in the formation of the desmoplastic, hypoxia and immunologically cold TME[113]. In the context of invasion and metastasis, histone deacetylases, such as HDAC1/2, have been shown to be recruited by the transcription factor ZEB1 to the promotor of key EMT genes such as *E-cadherin*, in order to reduce E-cadherin expression and subsequently induce EMT in PC cells[114]. Furthermore, the large-scale loss of heterochromatin marks, such as methylation of histones H3K9 and H4K20, has been shown to be associated with metastatic progression[115, 116]. The epigenetic reprogramming of primary PDAC cells demonstrated that epigenetic alterations play important roles in the tumorigenicity and aggressiveness of PC[117]. Indeed, epigenetic regulation has been shown to be involved in many hallmarks of cancer processes in PC[118].

1.3.14 Polymorphic microbiomes

The second enabling characteristic that has emerged from the 2022 paper by Hanahan, is polymorphic microbiomes[36]. More explicitly, this refers to the unique collection of microbes and the complex microbial system they form within the body[119]. Increasing evidence is emerging that polymorphic variations between the microbiomes of different populations can have substantial impacts on cancer phenotype[36]. A 2020 study identified distinct microbiomes within the tumour cells and tumour-associated immune cells of bone, brain, breast, lung, melanoma, ovarian, and PC patients[120]. These microbiomes have been shown to impact host immune responses in a pro-tumorigenic manner, as well as conferring direct carcinogenic impacts affecting genome instability, tumour inflammation and resistance to anti-cancer therapies[119]. The direct mechanisms by which this unique microbiome impacts the development and progression of PC remain to be fully understood.

1.4 Pancreatic cystic lesions

Pancreatic cystic lesions (PCLs) are typically fluid-filled structures that can be found within or on the surface of the pancreas, though some may have a solid appearance[121]. While many PCLs are benign and show no malignant potential, others, such as intraductal papillary mucinous neoplasms (IPMNs) or mucinous cystic neoplasms (MCNs), possess the ability to undergo malignant transformation and can be regarded as precursor lesions of PC[121-123]. The risk factors known to be associated with PC are extensive, however, investigations into these factors are largely case-control studies and as such have notable selection and recall biases[124]. Risk factors for PC can be classified as modifiable and non-modifiable[3, 124]. Modifiable risk factors include lifestyle factors such as smoking and alcohol consumption, as well as conditions such as obesity[124]. Non-modifiable risk factors include age, gender, ethnicity, genetic risk factors, diabetes and chronic pancreatitis[3]. For patients who have a family history of PC or are predisposed to malignancy due to hereditary genetic mutation, PCLs can be identified in up to one-third of such high-risk individuals[125, 126]. Germline mutations in *BRCA1* and *BRCA2* have been shown to confer an increased risk in PC as well as breast and ovarian cancers[127]. Von Hippel-Lindau (VHL) disease, caused by a germline

mutation to the VHL tumour suppressor gene, is associated with an increased risk of pancreatic neuroendocrine tumours and non-malignant serous-type PCLs[128].

One of the most significant risk factors for PCLs is age, with patients typically being diagnosed at 50 years or older and the incidence rate rising exponentially with age thereafter[3, 124, 125, 129]. PCL size and number have been shown to increase with age[130]. Variations in PCL prevalence from country to country can be shown to correlate with population demographics. This geographic variance is further widened by differences in imaging resolution and the frequency of routine physical check-ups within the population[121, 131]. Indeed, a 2017 study showed a positive correlation between socioeconomic development (measured through Human Development Index and Gross Domestic Product) and PC incidence and mortality[132]. This observed increase in PC incidence with rising socioeconomic development is thought to be result of the western lifestyle and ageing population, which are known to be large risk factors of PC[132]. A general improvement in imaging technologies, and a growth in the ageing population, has caused the worldwide prevalence of PCLs to rise drastically over the last two decades[125, 129]. The age- and sex-adjusted prevalence of PCLs in the general population is approximately 2%, but this figure increases exponentially with age and can range up to 45% in older generations[133-135]. With the rising prevalence of PCLs globally, and poor survival rates associated with PC, there is a great need for improved characterisation of pre-malignant PCLs to allow surgery in those who need it, and avoid unnecessary surveillance and intervention in those who do not.

1.5 PCL biology – the little we know

Despite having a plethora of evidence on the factors that can increase a patients risk of developing PCLs, the etiology of these cysts is largely unknown. Indeed, while some PCLs have been classified as 'inflammatory' cysts, and typically arise in cases of pancreatic inflammation, others can develop with no prior evidence of inflammation nor other co-morbidities (Figure 1.4)[136]. Pancreatic pseudocysts, a completely benign 'inflammatory' PCL subtype, are the only PCLs for which there exists developmental knowledge. In these cases, leakage of pancreatic enzymes, often as a result of

pancreatitis or pancreatic trauma, causes damage to the pancreas and forms a subsequent collection of fluid called a pseudocyst[137, 138]. Pseudocysts are distinct from all other 'true cysts' primarily via their lack of a lining. PCLs are closed structures with an epithelial cell lining which separates the fluid within the cyst from the nearby tissues[139]. Pseudocysts do not possess this lining, but rather have a nonepithelial wall of granulation tissue that is devoid of solid debris[140]. Thus, pseudocysts can trigger a cycle of pancreatic damage and accompanying symptoms, with the potential to cause enzymatic autodigestion and subsequently erode through the pancreatic tissue into adjacent cavities or tissues[141]. Many pseudocysts resolve spontaneously with little complications, however, some patients with persistent symptoms will require some form of drainage procedure[140]. The etiology of pseudocysts is well described across pancreatic cases reports due to its frequent development of post-pancreatic trauma or infection. The origin and developmental biology of non-benign PCLs, however, is less optimally documented. The presence of a PCL is a significant risk factor for the development of PC, and as such, knowledge of the development of these PCLs is key to the early detection and management of pancreatic patients[142].

Of critical importance is elucidating the role that the fluid within PCLs, pancreatic cyst fluid (PCF), may play in the development of PCLs and their subsequent malignant transformation. While many exploratory studies have run high-throughput analyses on PCF to identify potential biomarkers among the proteins, cfDNA, miRNA and metabolites within the fluid[143-146], no studies to date have examined the potential biological activity of PCF. Interestingly, a 2017 study evaluated the microbiome of 69 PCF samples and found via next generation sequencing (NGS) that there were as many as 408 genera of bacteria within the PCF, of which 17 were uniquely abundant to PCF, suggesting a distinct bacterial profile within this fluid[147]. Despite this knowledge of factors within the PCF, researchers remain woefully ignorant of the origin of these PCLs, and how these factors come to be contained within its fluid. Furthermore, the potential interactions between these factors and the pancreatic cells and tissue surrounding the cyst are unknown, and as such our ability to manage PCLs effectively is limited by our lack of knowledge of their biology. If PCF is shown to be biologically active, and therefore

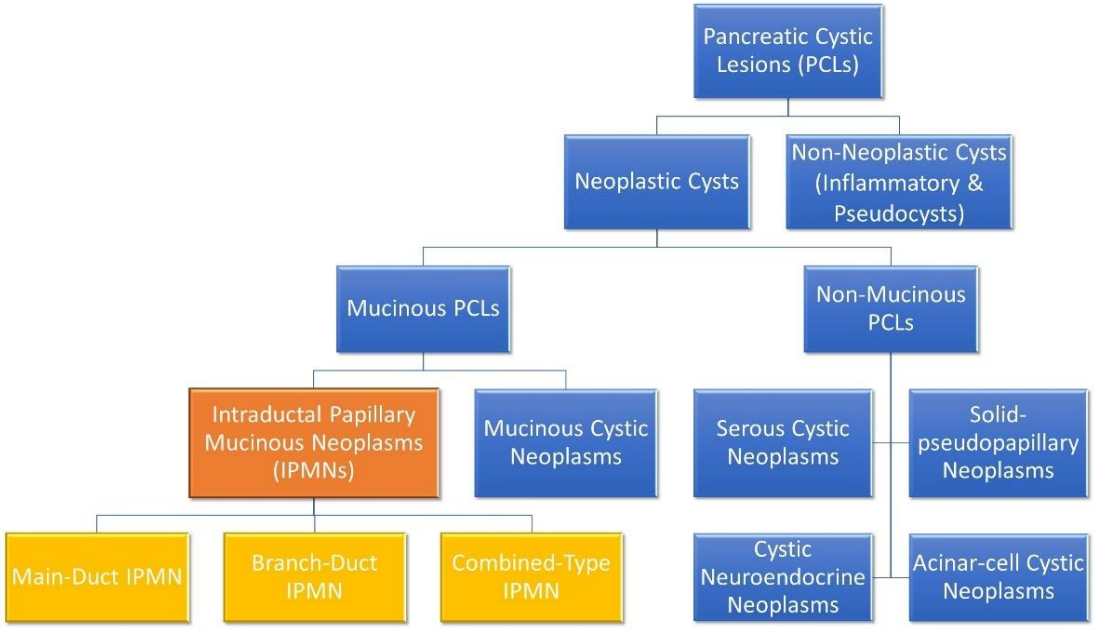


Figure 1.4 Molecular subgroups of PCLs. PCLs and their distinct subclassifications are highlighted. IPMNs are the most common subgroup and are responsible for 38% of PCLs, while mucinous cystic neoplasms, serous cystic neoplasms and cystic neuroendocrine neoplasm represent 23%, 16% and 7% of PCLs, respectively[148]. Branch-Duct IPMNs are most common (46%), followed by Combined-Type IPMNs (40%) and Main-Duct IPMNs (14%)[149].

may potentially play an active role in the malignant transformation of PCLs to PC, the aspiration of this fluid in all PCL patients could be called for. As such, determining the biological mechanisms by which PCLs form will play a key role in PCL management clinically, and will undoubtedly guide future pancreatic research and patient care.

1.6 High or low – determining which PCLs are high-risk

1.6.1 Subtypes of PCLs

When a patient with a PCL is identified, the first thing to be ascertained is the malignant potential of the PCL. Broadly speaking, PCLs can be divided into either neoplastic or non-neoplastic cysts as shown in Figure 1.4[129]. Neoplastic cysts can be either mucinous or non-mucinous, with non-mucinous PCLs rarely undergoing malignant transformation[125, 129]. Solid pseudopapillary neoplasms and cystic neuroendocrine neoplasms are notable, rare exceptions, as both are non-mucinous cystic lesions that do have some malignant potential and may require surgical resection. However, mucinous PCLs such as IPMNs and MCNs are generally regarded as precursor lesions for PC[125, 129, 150, 151]. IPMNs are the most common pre-malignant PCL, being much more common than MCNs. IPMNs are classified based on the involvement of the pancreatic ductal system as either main-duct (MD) IPMN, branch-duct (BD) IPMN or, when both main and branch ducts are involved, combined-type IPMN[121, 129]. Approximately 70% of MD-IPMNs undergo malignant transformation, whereas the rate is much lower in BD-IPMNs, ranging from 6-46%[121, 152]. The World Health Organisation describes three grades of IPMN: low-intermediate-grade dysplasia; high-grade dysplasia; and IPMNs with associated invasive carcinoma[129]. Identification of patients with high-grade dysplasia or early invasive cancer and the ability to predict those most likely to undergo malignant transformation is a key aspect of PCL patient management[153, 154].

1.6.2 Current management of PCLs: a lack of consensus

The most frequently used diagnostic tools for PCLs include computed tomography (CT), magnetic resonance imaging (MRI), and endoscopic ultrasound (EUS) +/- fine-needle aspiration (FNA), all of which have low sensitivity and specificity (SN/SP) for identifying

high- and low-risk patients[154]. As the biological behaviour of PCLs are notoriously unpredictable, there are currently a number of clinical guidelines that aim to help stratify the risk of PCLs undergoing malignant transformation[155]. At present there are three sets of guidelines in use to guide EUS and surgical referral of patients presenting with PCLs: the 2015 American Gastroenterological Association (AGA) guidelines[154], the 2017 International Association of Pancreatology Fukuoka (Fukuoka) guidelines[156], and the 2018 European evidence-based (European) guidelines[157]. The clinical indications for surveillance, EUS or surgical referral differ greatly across the three sets of guidelines. Indeed, while EUS is suggested by both the 2015 AGA and 2017 Fukuoka guidelines once certain criteria, such as cyst size or pancreatic duct dilation, are met, the 2018 European guidelines only recommend EUS when the results of this are expected to alter the clinical management of the patient (Table 1.1). In regards to surgery, both the 2015 AGA and the 2017 Fukuoka guidelines suggest surgery for any MCN, while the 2018 European guidelines require the MCN to be greater than 40mm in diameter or be accompanied by either an enhancing mural nodule, tumour-related jaundice, acute pancreatitis or new-onset diabetes mellitus (Table 1.1). As surgical resection is associated with significant morbidity and mortality, it should be reserved for those at a high risk of malignant transformation or established cancer[154]. Moreover, there is a 20% recurrence rate following surgical resection for IPMN[158], and recent studies have found multiple distinct regions of dysplasia within the pancreas, sometimes with differing mutational status of the same gene, supporting the notion of a multi-focal tumorigenic process of IPMN within the pancreas[122, 158, 159]. In terms of surveillance, the trend in these guidelines has moved towards more frequent evaluations, with the 2015 AGA guidelines suggesting surveillance once a year for IPMNs <30mm, while the 2017 Fukuoka and the 2018 European guidelines see patients with IPMNs <30mm or <40mm in diameter, respectively, receiving initial surveillance at 6 months (Table 1.2). Interestingly, the goalposts for cyst size have also been widened over the years, with IPMNs <30mm under both the 2015 AGA and 2017 Fukuoka guidelines receiving surveillance, while this cut-off has been extended to IPMNs and MCNs <40mm in the 2018 European guidelines (Table 1.2). Overall, the fact there are differing consensus guidelines in use is indicative of the imperfect state of knowledge regarding PCLs and PC, and the urgent need for improved biological characterisation of these lesions.

Table 1.1 Indications for EUS or surgery as per the current global clinical guidelines.

Guideline	Indications for EUS	Relative indications for surgery	Absolute indications for surgery
2015 AGA[154]	At least two of: • Cyst diameter >30mm • Solid nodule • PD dilation	-	• MCN • IPMN with PD ≥5mm and solid component or positive cytology
2017 Fukuoka[156]	• Growth rate ≥5mm over 2 years • Elevated serum CA19-9 • PD dilatation between 5mm and 9mm • Cyst diameter ≥30mm • Acute pancreatitis caused by IPMN • Enhancing mural nodule (<5mm) • Abrupt change in calibre of PD with distal pancreatic atrophy • Lymphadenopathy • Thickened/enhancing cyst walls	• IPMN with growth rate ≥5mm over 2 years • IPMN with elevated serum CA19-9 • IPMN with PD dilatation 5-9mm • IPMN with cyst diameter ≥30mm • IPMN with acute pancreatitis • IPMN with enhancing mural nodule (<5mm) • Abrupt change in calibre of PD with distal pancreatic atrophy • Lymphadenopathy • Thickened/enhancing cyst walls	• MCN • IPMN with suspicious or positive cytology • IPMN with tumour-related jaundice • IPMN with Enhancing mural nodule (≥5mm) • IPMN with PD dilation ≥10mm
2018 European[157]	• EUS–(FNA) should only be performed when the results are expected to change clinical management • EUS–(FNA) is recommended if the cyst has either clinical or radiological features of concern identified during the initial investigation or surveillance	• IPMN with growth rate ≥5mm per year • IPMN with elevated serum CA19-9 (>37U/mL) • IPMN with PD dilation 5-9.9mm • IPMN ≥40mm • IPMN with new-onset diabetes mellitus • IPMN with acute pancreatitis • IPMN with enhancing mural nodule (<5mm)	• MCN ≥40mm • MCN with enhancing mural nodule • MCN with tumour-related jaundice or acute pancreatitis or new-onset diabetes mellitus • IPMN with positive cytology for malignancy or high-grade dysplasia • IPMN with solid mass • IPMN with tumour-related jaundice • IPMN with enhancing mural nodule (<5mm) • IPMN with PD dilation ≥10mm

AGA=American Gastroenterological Association, European=European evidence-based, EUS=endoscopic ultrasound, Fukuoka=International Association of Pancreatology Fukuoka, IPMN=intraductal papillary mucinous neoplasm, MCN=mucinous cystic neoplasm, PD=pancreatic duct. *Adapted from Van Huijgevoort et al. (2019)*[160].

Table 1.2 Indications for surveillance as per the current global clinical guidelines.

Guideline	Cyst type	Surveillance interval	Surveillance modalities
2015 AGA[154]	IPMN <30mm	Yearly for 1 year, then every 2 years. Discontinue after 5 years if no change to cyst size or characteristics.	MRI with MRCP
2017 Fukuoka[156]	IPMN <10mm	Within 6 months, then every 2 years.	CT or MRI with MRCP
	IPMN 10-20mm	Every 6 months for 1 year, then yearly for 2 years, then every 2 years.	CT or MRI with MRCP
	IPMN 20-30mm	3-6 months, then yearly.	EUS alternating with MRI
2018 European[157]	IPMN <40mm	Every 6 months for 1 year, then yearly.	CA19-9, EUS and/or MRI
	MCN <40mm	Every 6 months for 1 year, then yearly.	CA19-9, EUS and/or MRI

AGA=American Gastroenterological Association, European=European evidence-based, CT=computerized tomography, EUS=endoscopic ultrasound, Fukuoka=International Association of Pancreatology Fukuoka, IPMN=intraductal papillary mucinous neoplasm, MCN=mucinous cystic neoplasm, MRCP=magnetic resonance cholangiopancreatography, MRI=magnetic resonance imaging. *Adapted from Van Huijgevoort et al. (2019)*[160].

EUS-guided FNA is a safe and accurate method of extracting PCF or pancreatic tissue from a patient for further analysis[152, 161]. PCF cytology has high specificity for malignancy or high grade dysplasia, but low sensitivity due to the typically low cellularity of PCF samples[152, 161]. A 12-year multi-institutional study conducted by the French Surgical Association found that 50% of patient PCF samples collected were non-diagnostic and acellular[162]. Diagnosis of PCLs by EUS requires attention to cyst morphology, including size, number of cysts present, characteristics of the wall and internal structures, calcification, positioning in relation to the main pancreatic duct and presence of lesions in the background[129]. These descriptors are considerably operator-dependent and PCL characterisation without PCF analysis is limited[121, 152]. While the low cellularity of PCF limits cytological yield, biochemical analysis of PCF has proven an important adjunct in characterising PCLs. Carcinoembryonic antigen (CEA) levels are frequently measured in EUS-aspirated PCF for the diagnosis of MCNs, with a threshold of ≥192 mg/dL[163]. CEA has been shown to have a sensitivity of between 59-67% and specificity of 83-91% for detection of MCNs, and is among the best of the biomarkers currently available for this purpose[150, 164]. Mutational profiling of patients has shown utility in the characterisation of different PCL subtypes, however, genetic evaluation of PCF is currently limited to research. *KRAS* and *GNAS* mutations in the PCF are particularly important early mutations in IPMNs as they are not found in other common types of cysts[159]. Indeed, a 2021 systematic review and meta-analysis, including 785 PCLs, reported a sensitivity of 94% and a specificity of 91% for *KRAS* + *GNAS* in the distinction of IPMNs[165]. Carbohydrate antigen 19-9 (CA19-9) is an antigen released by PC cells and is the only FDA-approved biomarker for PC diagnosis[166]. Unfortunately, the reported SN/SP values for CA19-9 are generally poor, as this marker is known to be elevated in benign conditions such as pancreatitis and diabetes, common comorbidities of pancreatic patients[167]. Importantly, CA19-9 is not elevated in premalignant PCLs, making its utility in this setting extremely limited[168]. Based on all the aforementioned limitations of current diagnostics, it is clear that there is urgent need for novel methods and markers to accurately classify and risk stratify PCLs, and the 'omics' revolution is poised to fill this void of information.

1.7 Identification of biomarkers in PCLs and PC using omics

A biological marker or 'biomarker', is defined as any characteristic that can be measured objectively and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to some therapeutic intervention[169]. Biomarkers have many clinical uses and these can be broadly classified into five distinct categories: antecedent, screening, diagnostic, staging and prognostic[170].

The term 'omics' refers to the compendium of disciplines that comprehensively assess a set of molecules[171]. The origin of omics disciplines began with the advent of genomics, the study of the entire genome of an organism, the pathways involved, the interactions between individual genes and the impact of an organism's environment with their genes[172]. This differed from the field of genetics, which preceded genomics, and studies individual variants of single genes and heredity in organisms[172]. Since the emergence of genomics, other omics-type fields have emerged, including proteomics, transcriptomics and metabolomics[173, 174]. The omics field has made huge strides in the past two decades, largely due to technological advancements, enabling the cost-effective and high-throughput analysis of biological molecules[172]. Some omics disciplines are demonstrating great potential in the search for a novel biomarker for PC (Table 1.3), but the data for PCLs are much more limited.

1.7.1 Genomics

Genomics was the first omics discipline to be established, and studies the genome in its entirety[172]. More specifically, genomics examines various characteristics of DNA and RNA such as DNA copy number, DNA modifications, single nucleotide polymorphisms (SNPs), coding RNA expression and non-coding RNA levels among others[172, 173, 175]. This discipline focuses on genetic variations or modifications, their interactions with each other and follows their impact across biological pathways and phenotypes. Technological advances have resulted in the emergence of new, high-throughput molecular techniques capable of giving vast amounts of information on the human genome such as SNP arrays, gene expression microarrays, and NGS[176]. These techniques have been employed extensively in PC research and genomic evaluation of

Table 1.3 Overview of biomarkers in PC and PCLs that have been validated in an independent cohort.

Biomarker Name	Biomarker Type	Single or Multi-study validated	Platform	Sample Type	Sample Size (total no. patients)	Sensitivity (95% Confidence Interval)	Specificity (95% Confidence Interval)	P-value	Purpose	References
KRAS &/OR GNAS	Genetic mutation panel	Multi	PCR	Cyst fluid	91	65% (52-76) 84% (70-92)	100% (83-100) 98% (86-100)	N/A N/A	MCN vs non-MCN IPMN vs non-IPMN	[177]
			PCR using NGS	Cyst fluid	197	68.5% (N/A)	95.5% (N/A)	N/A	IPMN vs non-IPMN	[178]
			NGS	Cyst fluid	595	89% (79-95)	100% (88-100)	N/A	MCN vs non-MCN	[179]
			Sanger sequencing	Cyst fluid	159	65% (N/A)	100% (N/A)	N/A	MCN vs non-MCN	
lncRNA-TFG	Long non-coding RNA	Single	Affymetrix Human Exon 1.0 ST	Tissue	28	N/A	N/A	$6.23x10^{-8}$	Positive correlation with tumorigenesis in IPMNs	[180]
CTD-2033D15.2	Long non-coding RNA	Single	Affymetrix Human Exon 1.0 ST	Tissue	28	N/A	N/A	$1.47x10^{-4}$	Negative correlation with tumorigenesis in IPMNs	[180]
HAND2-AS1	Long non-coding RNA	Single	Affymetrix Human Exon 1.0 ST	Tissue	28	N/A	N/A	$2.66x10^{-3}$	Negative correlation with tumorigenesis in IPMNs	[180]
Glucose	Metabolite	Multi	Liquid Chromatography	Cyst fluid	19	94% (N/A)	64% (N/A)	0.004	Glucose ≤ 66 mg/dL in MCNs vs non-MCNs	[181]
			Glucometer	Cyst fluid	153	92% (N/A)	87% (N/A)	N/A	Glucose ≤ 50 mg/dL in MCNs vs non-MCNs	[182]
Kynurenine	Metabolite	Single	Liquid Chromatography	Cyst fluid	19	90% (N/A)	100% (N/A)	0.002	Lower in MCNs vs non-MCNs	[181]
AcSperm & DAS & LPC(18:0) & LPC(20:3) & Indole-derivative	Metabolite panel	Single	Mass spectrometry	Blood Plasma	121	66.7% (N/A)	95% (N/A)	N/A	PDAC vs N	[183]

ADAMTS1	Methylated gene	Single	Methylation on beads	Blood cfDNA	39	87.2% (N/A)	95.8% (N/A)	N/A	PDAC vs N	[184]
BNC1	Methylated gene	Single	Methylation on beads	Blood cfDNA	39	64.1% (N/A)	93.7% (N/A)	N/A	PDAC vs N	[184]
SOX17	Methylated gene	Single	Methylation-specific ddPCR	Cyst fluid	154	78.4% (64.7–88.7)	85.6% (78.4–91.1)	N/A	High-risk PCL vs low-risk PCLs	[185]
TBX15 & BMP3	Methylated gene marker panel	Single	Whole-genome methylome discovery & qPCR	Cyst fluid	134	90% (70-99)	92% (85-96)	N/A	HGD/PC vs LGD/N	[186]
ADAMTS1 &/OR BNC1	Methylated gene panel	Single	Methylation on beads	Blood cfDNA	39	97.4% (N/A)	91.6% (N/A)	N/A	PDAC vs N	[184]
FOXE1 & SLIT2 & EYA4 & SFRP1	Methylated gene panel	Single	Methylation-specific ddPCR	Cyst fluid	154	84.3% (N/A)	89.4% (N/A)	N/A	High-risk PCL vs low-risk PCLs	[185]
miR-1290	MicroRNA	Multi	MicroRNA array analysis	Blood serum	60	88% (N/A)	84% (N/A)	N/A	PC vs N	[187]
					76	83% (N/A)	69% (N/A)	N/A	PC vs CP	
					95	83% (N/A)	78% (N/A)	N/A	PC vs CP & N	
			qRT-PCR	Blood plasma	49	N/A	N/A	0.027	PDAC vs N	[188]
			qRT-PCR	Blood serum	200	74.2% (N/A)	91.2% (N/A)	N/A	PC vs C	[189]
9-miRNA model[a]	MicroRNA panel	Single	TaqMan miRNA Array	Tissue & Cyst fluid	33 & 50	89% (N/A)	100% (N/A)	N/A	HG IPMNs, PanNETs & SPNs vs LG IPMNs & SCAs	[145]
miR-3679-5p & miR-940	MicroRNA panel	Single	qPCR	Saliva	80	72.5% (N/A)	70.0% (N/A)	N/A	PC vs N	[190]
					60	62.5% (N/A)	80.0% (N/A)	N/A	PC vs BPT	
					100	70.0% (N/A)	70.0% (N/A)	N/A	PC vs N & BPT	

CA19-9	Protein-associated	Multi	Bead-based xMAP immunoassay	Blood serum	267	57.2% (N/A)	90% (N/A)	N/A	PDAC vs N	[191]
			ELISA	Blood plasma	176	77.5% (N/A)	83.1% (N/A)	N/A	CA19-9 >20.3U/mL, PDAC vs C	[192]
			Retrospective clinical data	Blood serum	41	90% (N/A)	83.33% (N/A)	N/A	2.45 times elevated CA19-9 indicated recurrence of PC	[193]
CEA	Protein	Multi	Clinical data	Cyst fluid	31	73% (N/A)	89% (N/A)	N/A	CEA >192 ng/mL for MCN	[181]
			ELISA	Cyst fluid	149	95.5% (N/A)	81.5% (N/A)	<0.0001	CEA ≤ 10 ng/mL for SCN	[194]
			ELSA	Cyst fluid	153	58% (N/A)	96% (N/A)	N/A	CEA >192 ng/mL for MC	[182]
MUC5AC:WGA & MUC5AC:BGH & Endorepellin:WGA	Protein panel	Multi	Antibody-lectin sandwich microarray	Cyst fluid	147	92%[b](N/A)	94%[b](N/A)	N/A	Elevation in any 2 differentiates MCNs vs non-MCNs	[195]
			Antibody-lectin sandwich arrays	Cyst fluid	22	87% (N/A)	100% (N/A)	N/A	Elevation in any 2 differentiates MCNs vs non-MCNs	[196]
Thymosin- β4	Protein	Single	MALDI Imaging & Mass Spectrometry	Tissue	45	70% (N/A)	71% (N/A)	0.011	Overexpressed in IPMN with HGD	[197]
Ubiquitin	Protein	Single	MALDI Imaging & Mass Spectrometry	Tissue	45	94% (N/A)	86% (N/A)	0.04	Overexpressed in IPMN with HGD	[197]
VEGF-A	Protein	Single	ELISA	Cyst fluid	149	100% (N/A)	83.7% (N/A)	<0.0001	VEGF-A >5,000 pg/mL benign SCN	[194]
VEGF-A & CEA	Protein panel	Single	ELISA	Cyst fluid	149	95.5% (N/A)	100% (N/A)	N/A	VEGF-A >5,000 pg/mL & CEA ≤ 10 ng/mL in benign SCN	[194]

BPT=benign pancreatic tumour, C=non-cancer control, CP=chronic pancreatitis, ELISA=enzyme-linked immunosorbent assay, HG=high grade, HGD=high-grade dysplasia, IPMN=intraductal papillary mucinous neoplasm, LG= low grade, LGD=low-grade dysplasia, MALDI=matrix-assisted laser desorption ionisation, MC=mucinous cyst, MCM=mucinous cystic neoplasm, N=normal healthy, N/A=not available, NGS=next-generation sequencing, PanNET=pancreatic neuroendocrine tumour, PC=pancreatic cancer, PCL=pancreatic cystic lesion, PCR=polymerase chain reaction, PDAC=pancreatic ductal adenocarcinoma, SCN=serous cystic neoplasm, and SPN=solid-pseudopapillary neoplasm. [a]Model is intellectual property of the authors. [b]Average of three cohorts.

PCF and patient serum remains a promising field. Cancer genomics has taken great strides over recent years, illuminating the complexity of the cancer genome and more importantly, the heterogeneity present even between tumours with the same histological diagnosis[173, 176]. In the context of medical research, genomics is employed to identify genetic variants associated with disease, treatment response or prognosis[172]. The European Medicine Agency defines a "genomic biomarker" as a characteristic of DNA and/or RNA that can be measured, by its expression, function or regulation of a gene, and that is an indicator of normal biological processes, pathogenic processes and/or a biological response to therapeutic or other interventions[175, 198].

Overexpression of *epidermal growth factor receptor* (*EGFR*) is known to be associated with poor prognosis in several cancers, such as colorectal cancer (CRC), non-small cell lung cancer (NSCLC) and PC[199]. *EGFR* gene copy number and the mutational status of the *KRAS* gene, which is located downstream of *EGFR*, can predict response to *EGFR*-targeted therapy cetuximab in advanced CRC[200, 201]. *EGFR* is also used as a genomic biomarker for NSCLC, where *EGFR* mutations located in the tyrosine kinase domain are associated with a treatment response to *EGFR* tyrosine kinase inhibitors, such as gefitinib[202-204]. This receptor has also been shown to be overexpressed in PC, and this overexpression correlates with advanced disease and poor overall survival in patients[205]. Indeed, *EGFR* is known to play a key role in the pathogenesis of PC and drives its aggressive nature[206]. *EGFR*-targeted therapies have proven efficacy in the treatment of both CRC and NSCLC[199], however, the FDA has not yet approved the use of any of these *EGFR*-targeted therapies for the treatment of PC, as data regarding efficacy are inconclusive[207]. Treatment of patients with *EGFR* inhibitors in combination with standard of care regimens showed little survival benefit and increased risk of toxicities[207]. As such, *EGFR* amplification alone does not provide a sufficient biomarker for treatment response and there is an urgent need for predictive-biomarkers that could indicate those patients who would benefit most from *EGFR*-targeted therapies and what combination would prove most effective. As there are many genetic mutations known to be present in pancreatic tumours, and subsequently a great deal of crosstalk between pathways, more extensive research into the relationships between these genes and pathways could guide research towards more suitable biomarkers[207].

Genetic mutations have been shown to be hugely important in the study of many cancers and PC is no different, with *GNAS* and *KRAS* mutations representing the predominant mutations observed in this cancer[208, 209]. *KRAS* is an oncogene primarily involved in the production of protein for regulating cell division[208]. Mutations in this gene are arguably the most important in the context of PC, as they frequently occur in non-cancerous precursor lesions and are subsequently present in 90%-95% of all PC cases[210-212]. The PANDA study analysed the DNA of 113 patient PCF samples in a multicentre, prospective study[213]. Mutations in *KRAS* were shown to be indicative of a mucinous cyst, with a specificity of 96%[213]. However, there is some evidence to suggest that *KRAS* mutation alone may not be sufficient to drive a malignant phenotype, and other genetic or epigenetic events may be needed[210]. A 2011 study found that *GNAS* mutations were present in 66% of IPMNs, while mutations of either *KRAS* or *GNAS* were present in 96%[209]. The same study found *GNAS* mutations in seven out of eight cases of invasive PC that resulted from an IPMN, while it was not present in other types of pancreatic cysts or carcinoma that was not associated with an IPMN[209, 214]. Mutations in both genes are believed to occur in the early stages of IPMN carcinogenesis[215]. The reported incidence of *GNAS* and *KRAS* mutations alone in IPMNs has varied greatly between studies, but a 2016 meta-analysis revealed the prevalence of *GNAS* and *KRAS* in these cysts to be 56% and 61%, respectively[216]. Simultaneous mutations in both *GNAS* and *KRAS* have been demonstrated to occur in up to 33% of IPMNs[177, 217]. Indeed a *KRAS* and/or *GNAS* biomarker panel has been shown to have a SN/SP of up to 84%/98% for the identification of IPMNs (Table 1.3)[177, 178]. Unfortunately, these results for *GNAS* and *KRAS* mutations are not mimicked in MCNs, where two separate studies revealed a SN/SP of 65%/100% for MCNs [177, 179], while others have highlighted a distinct lack of *GNAS* mutations in all subtypes of MCNs [217].

Whole-genome sequencing of patients with PC, and subsequent RNAseq revealed the *KRAS* signalling pathway to be the most heavily impacted, however, further details elucidating passenger and driver mutations are needed[218]. It also appears that *KRAS* and *GNAS* mutational status varies with IPMN histological grade, adding further to the difficulties observed in these genetic mutations as potential biomarkers[217]. The feasibility of *KRAS* mutational status as a single marker has been evaluated in tissue, PCF,

duodenal fluid and plasma and does not appear to diagnose IPMNs or the level of cellular dysplasia consistently, being regarded as simply an early indicator of cell stress in pancreatic cells[215]. The addition of *GNAS* to PCF *KRAS* testing has been shown to increase the diagnostic accuracy of IPMN identification from 66-80.7%, though this does not achieve a statistically superior result to *KRAS* testing alone ($p > 0.05$), which has a diagnostic accuracy of 76.6%[178]. While mutational profiling of these genes may show some promise for IPMN identification, they provide no risk stratification for these cysts, and show little utility for MCNs compounding their lack of use in the clinical setting.

Importantly, large networks of genetic data have begun to emerge over the last two decades that contain genomic sequencing of patients with various cancer types. The Cancer Genome Atlas (TCGA) has executed the molecular profiling and subsequent analysis of over 11,000 tumours, spanning 32 different cancer types[219, 220]. Tumour samples are characterised using technologies that assess the sequence of the exome, copy number variation, DNA methylation, mRNA expression and sequence, microRNA expression and transcript splice variation[220]. While this network has a substantial and diverse amount of genomic data, the matching clinical data for these patients is far more limited, and is considered to be one of the major drawbacks of the database[221, 222]. Improved models have been developed in order to utilise the TCGA data to their full potential. Indeed, a deep-learning-based model on hepatocellular carcinoma has been shown to identify multi-omics features that robustly differentiate patients from six cohorts into two distinct survival subpopulations[222]. Another such platform is the International Cancer Genome Consortium, which provides data at the genomic, transcriptomic and epigenomic level, for over 50 cancer types and consolidates the data from TCGA[223]. The integration of these datasets enables the execution of genomic cancer studies on a large scale. The compilation and analysis of genetic information captured by various methods poses some analytic problems, which will be discussed later in this review. The genomic data of PC patients are accessible on a number of online platforms, and several independent studies have also been launched into the utility of these data to distinguish high- and low-risk patients with PCLs, however, these single markers and panels have not yet shown sufficient sensitivity or specificity for this purpose[152].

1.7.2 Transcriptomics

The transcriptome comprises the entire set of RNA transcripts that are produced by a single genome in a target cell or population of cells, consisting of both coding and non-coding RNAs[174]. While genomic sequencing can provide vast amounts of data and identify structural variations, it does not reveal whether these variations are carried beyond the genome and affect the transcription of the gene[224]. It follows, therefore, that the addition of transcriptomic evaluation of a tumour could help to elucidate those genetic driver mutations that have active downstream effects. Full transcriptome sequencing or RNA-seq has become a popular and widely used method of assessing genome wide expression[225, 226]. Normalisation of RNA-seq data is a key aspect of this technique as it is required to eliminate sources of variability in order to allow accurate inference of expression levels and comparisons across the dataset[225, 227]. These sources of variability, often termed 'batch effects', originate from parameters such as the quality of RNA, the equipment or techniques used for RNA extraction and the RNA-seq library preparation[225]. There are several methods of normalisation that can be used, each affecting the power and reproducibility of the results, however, a specially developed method, Remove Unwanted Variation, has been shown to overcome such pitfalls of normalisation[225, 227]. Similarly, single cell RNA sequencing, which has become popular over recent years, requires computational strategies tailored to the different noise and artefacts found amongst these data[226].

MicroRNAs (miRNAs) are small, non-coding RNA molecules that function as RNA silencers and regulators of gene expression at the post-transcriptional level[228]. These molecules have been extensively studied in the context of cancer, and many miRNAs have been identified as having differential expression levels between high- and low-risk IPMNs[228]. Indeed, a 9-miRNA model developed by Matthaei *et al.* has shown SN/SP of 89%/100% for the distinction of high- and low- risk IPMNs by both tissue and PCF (Table 1.3)[145]. Similarly, Lee *et al.* identified a 4-miRNA panel (miR-21-5p, miR-485-3p, miR-708-5p, and miR-375) that appears to distinguish IPMNs from PC, with a SN/SP of 95%/85%, though these results have not been independently validated[229]. Three long non-coding RNAs (lncRNAs) that have shown promise as risk stratification markers in the tissue of IPMN patients are CTD-2033D15.2, HAND2-AS1 and lncRNA-TFG (Table 1.3). A

recent study conducted by Ding *et al.* indicated a negative correlation with tumorigenesis in IPMNs for CTD-2033D15.2 and HAND2-AS1, while a positive correlation was observed for lncRNA-TFG[180]. These results suggest a protective role of HAND2-AS1 and CTD-2033D15.2 expression in IPMNs, while lncRNA-TFG appears as a risk factor for tumorigenesis in IPMNs[180]. MiRNAs have also been identified in the blood serum and plasma of PC which have exhibited potential for the diagnosis of PC. Multiple studies have shown elevated levels of miR-1290 in patient blood has the ability to distinguish between PC, healthy patients, and patients with chronic pancreatitis (Table 1.3)[187-189]. Wei *et al.* found that miR-1290 expression was upregulated in PC patients compared to all controls, and was decreased dramatically post tumour resection (p<0.001) indicating a potential role in tumorigenesis[189]. Vila-Navarro *et al.* used NGS to conduct genome-wide miRNA profiling and identified 30 independent miRNAs whose expression is significantly increased in PC and IPMN lesions compared to healthy individuals, and these results were validated in 2 independent sample sets[230]. Among these 30 miRNAs, 24 represent novel biomarkers that have not been reported previously in IPMNs[230]. While such results indicate great promise for the identification of a panel of miRNAs that could be used in pancreatic lesion characterisation, as this panel cannot distinguish IPMN from PC its clinical utility is greatly limited and larger, multicentre studies will be needed to further interrogate and validate these results.

One limitation of current patient sampling is that EUS-FNA is an invasive procedure for patients, with sample yields often being of low volume. Research surrounding less invasive protocols has investigated the utility of salivary properties for use as non-invasive biomarkers. Salivary miRNA has been explored as a candidate for diagnostics in PC. Xie *et al.* validated the salivary biomarkers miR-3679-5p and miR-940 for the distinction of PC from healthy individuals, and found that combining both miRNAs produced the best discriminatory power (Table 1.3)[190]. Another study identified four miRNAs (miR-21, miR-23a, miR-23b and miR-29c) in patient saliva and showed them to be significantly upregulated in the saliva of PC patients when compared to healthy controls, with a sensitivity of 71.4%, 85.7%, 85,7% and 57%, respectively, and specificity at a fixed 100%[231]. However, these same miRNAs were shown to be detected in patients with pancreatitis, while miR-23a and miR-23b were detected in patients

diagnosed with IPMNs[231]. While these miRNAs show promise in distinguishing PC from healthy controls, as patients with pancreatitis and precursor lesions have also been shown to express these markers further validation is required on a larger, external cohort to fully demonstrate the utility of these miRNAs at distinguishing various pancreatic pathologies. A 2020 systematic review and meta-analysis interrogated the potential of various salivary biomarkers in several cancer types, including transcriptomic, epigenomic (*v. inf.)* and microbiomic markers in PC, and found "good" diagnostic accuracy for such markers in PC with an area under the curve (AUC) of 0.87[232]. However, the review also highlighted the high degree of variation in the sensitivity (31-100%) and specificity (34-100%) observed in different studies of salivary biomarkers in non-oral cancers, and further interrogation of the data revealed that the probability of a patient having some malignancy is 31% if the salivary test result is negative[232]. While investigations into the use of salivary properties for diagnostic and prognostic purposes appears promising, further work is required to identify more robust biomarkers.

Transcriptomic data, as highlighted above, has profound potential for the identification of a novel biomarkers. Compilations of transcriptomic data can be found readily available in online repositories such as the Gene Expression Omnibus and EBI ArrayExpress[233, 234]. Access to vast amounts of patient-derived data has enabled huge amounts of research to be conducted that would otherwise have not been possible. Genome-wide association studies (GWAS) have become very popular, and results suggest that common variations in the human genome, such as those exhibited by SNPs with frequencies of more than 1%, are responsible for the risk observed in many genetically complex disorders[235]. Following on from such studies, a similar undertaking was carried out at the transcriptome level - the transcriptome-wide association study, where gene expression measurements were combined with summary association statistics from GWAS to identify 69 new genes associated with obesity-related traits such as BMI, lipids and height[236]. Such endeavours highlight the power of layering different biological techniques to interrogate these data and identify the most relevant points. It is evident from expansive transcriptomic data available for PC that much research has been conducted, however, to date no transcriptomic biomarker among the

many identified has been approved for use in this cancer. While vast quantities of data are often favourable, it appears that more information and progress may be gained from the integration of different data types.

1.7.3 Epigenomics

Epigenetic modifications in the context of cancer have been shown to play an important role in the initiation, progression and metastasis of malignant tumours[237]. While such alterations do not affect the sequence of the genome, and are therefore regarded as heritable changes in gene expression, they are maintained filially by cellular division[237, 238]. The epigenetic landscape of a target cell is comprised of 4 distinct mechanisms of non-genetic modification: DNA methylation and/or de-methylation; histone modification; non-coding RNA molecules; and nucleosome remodelling[239]. DNA methylation is the best studied form of epigenetic modification, and has the potential to affect the expression of genes[239]. DNA hypermethylation concerns the gain of methylation at specific DNA sites that under normal circumstances would be unmethylated[238]. Hypermethylation of gene promoter CpG islands is found in virtually every type of tumour and can cause transcriptional silencing of genes, such as tumour suppressor genes, which subsequently drives cancer formation[224, 240]. A key example of this is observed in tumours with reduced or absent *BRCA1* expression, where hypermethylation of this gene occurs in 9-37% of sporadic breast cancers[240, 241]. Conversely, hypomethylation refers to the loss of a methyl group, resulting in the overexpression of genes, which in the context of malignancy are often found to be oncogenes[224]. The modification of histone tails post-transcriptionally enables structural changes to occur to the chromatin which can either facilitate or hinder protein complex interactions with specific DNA regions[239]. Such histone modifications, therefore, have the ability to influence gene transcription in a manner which can contribute to the development and progression of cancer.

A histone methyltransferase known as enhancer of zeste homologue 2 (EZH2) is known to be overexpressed in many cancers, including PC, and has also been detected in IPMNs with moderate to severe dysplasia[242]. A 2010 study showed that high expression levels of EZH2 in PC was associated with increased node positivity and a

larger tumour size; EZH2 expression levels were also shown to relate to the degree of dysplasia in IPMNs[242]. RNA interference silencing of EZH2 sensitized PC patients to treatment with gemcitabine, resulting in significantly longer overall survival [242]. Such RNA interference silencing of EZH2 has been utilised in a PC model and caused a decrease in tumour growth and the incidence of liver metastasis[243]. More recent investigation into EZH2 has highlighted its role in the epigenetic repression of tumour suppressor gene expression. Trimethylation of *H3K27* by EZH2 allows the mediation of cell proliferation, invasion and migration[244]. Exposure of F-box and WD repeat domain-containing 7 to EZH2 causes the degradation of EXH2 in PC cells and inhibits tumour migration and invasion, indicating its role as a ligase of EZH2 that regulates EZH2 protein levels in PC and furthermore, its potential as a treatment strategy[244]. Indeed as epigenetic alterations are reversible and plastic, they can be regarded as more amenable to therapeutic intervention than non-reversible genetic mutations[245]. Hata *et al*. identified 6 methylated DNA markers in patient PCF that could distinguish high- and low-rick PCLs with accuracies from 79.8-83.6%[185]. Methylated *SOX17* was shown to be the most sensitive single marker, while a four-gene combination (*FOXE1*, *SLIT2*, *EYA4*, *SFRP1*) showed the highest accuracy at 88% (Table 1.3)[185]. Furthermore, PCF obtained from IPMNs with high-grade dysplasia had significantly higher levels of methylated DNA than other mucinous cysts[185]. A more recent study by Eissa *et al.* examined the cell-free DNA in patient blood, and found the methylated gene of *ADAMTS1* to have a SN/SP of 87.22%/95.8% for the differentiation of PC and normal samples[184]. Moreover, the same study showed that the addition of a second methylated gene, *BNC1*, such that either or both were detected in the samples, showed even better SN/SP for the same purpose (Table 1.3)[184].

For IPMNs, the epigenetic data currently available is limited. One gene whose promoter is known to be hypermethylated in almost all cancer types is the *cysteine dioxygenase 1* (*CDO1*) gene. A recent study examined this gene in pancreatic IPMN tumour tissue and found the *CDO1* promoter hypermethylation is extremely specific to IPMNs and appears to accumulate with tumour progression[246]. Among other pancreatic disease, low levels of *CDO1* promoter hypermethylation were seen in MCNs, with no other pancreatic cystic disease showing DNA hypermethylation of its promoter. A pilot

study in pancreatic juice confirmed methylation in all IPMN samples (n=6) with none detected in benign pancreatic diseases (n=6, chronic pancreatitis and autoimmune pancreatitis)[246]. Furthermore, *CDO1* hypermethylation showed utility in the differentiation of low-intermediate-grade dysplasia and high-grade dysplasia/PC[246]. While these results show promise in the search for a biomarker to stratify IPMN patients, extremely robust thresholds for *CDO1* methylation are needed to distinguish high- and low-risk patients, with little utility being seen for other PCLs. Further analyses on a large patient cohort, examining the methylation status of *CDO1* in patients with pancreatitis, pseudocysts and a variety of PCLs would be required to further validate this marker.

The establishment of epigenomic databases such as ENCODE and The International Human Epigenome Consortium and Roadmap Epigenomic Project has enabled the popularisation of epigenomics and allowed for the establishment of standardised sequencing methods[247]. Furthermore, epigenome-wide association studies in combination with GWAS and the transcriptome-wide association study data have proved to be powerful tools in pinpointing disease-relevant regulatory elements[247, 248].

1.7.4 Proteomics

The proteome consists of all proteins within a target cell, tissue or biological sample at the time of sampling[174]. The proteome can be examined at different developmental or cellular phases, and changes in the proteome can be evaluated at different time points. When moving from genomic or transcriptomic studies to those of the proteome, it is important to consider the increasing complexity of the longer nucleotide codes which can be arranged in a number of conformations and have varying chemical modification, with alternative splicing of the same protein resulting in different isoforms[174]. As cancer cells have distinct processes of replication and metabolism, the quantities and types of proteins produced can be profoundly affected[249]. Proteomics can be a qualitative and/or quantitative evaluation of the proteome and is generally conducted using mass spectroscopy (MS)[174, 249, 250].

As mentioned previously, two protein markers currently utilised in PC are CA19-9 and CEA, and these markers have been shown in many instances to be insufficient in the discernment of IPMNs and MCNs, as well as their malignant potential (Table 1.3)[251-253]. CEA levels in patient PCF can be used to distinguish between mucinous and non-mucinous cysts, but have limited sensitivity (58-73%) and specificity (89-96%)[181]. Kadayifci *et al.* evaluated the diagnostic accuracy of adding CEA to the *KRAS* and/or *GNAS* panel but found it did not provide better SN/SP (p>0.05) than the *KRAS* and/or *GNAS* panel alone[178, 182]. Indeed, a recent study showed that artificial intelligence by deep learning has better SN/SP (95.7/91.9%) for diagnosis of malignant cystic lesions than CEA levels and cytologic analyses[253]. Serum CA19-9 is the only FDA-approved marker for the identification of PC, however, it has been demonstrated that CA19-9 alone failed to detect 44.1% of cancer cases in a cohort of 34 patient samples, and added no improvement to the sensitivity of the two-gene methylation panel *ADAMTS1* and/or *BNC1*[184]. CA19-9 is widely regarded as not sufficiently sensitive to distinguish PC from healthy samples as it is frequently elevated in non-malignant conditions such as pancreatitis, and has been shown to have a SN/SP of 52.7%/90%[191]. However, the addition of CA19-9 to a marker panel (CA19-9, ICAM-1, OPG) was shown to produce better sensitivity and specificity for PC (78% and 94.1%, respectively)[191]. Indeed, Brand *et al.* identified several 3-marker panels that offered an improved ability over CA19-9 alone to distinguish PC from healthy controls[191]. Another study identified how the change in cut-off value for CA19-9 can improve the robustness of this marker, but also showed that the addition of CA19-9 to a marker panel gave the best SN/SP when compared to CA19-9 alone (p<0.05)[192]. While CA19-9 appears to have limited use clinically for diagnostic screening of patients for PC, it does have utility in predicting disease recurrence post-treatment[254].

The identification of novel protein markers in PC has been of great interest over the last two decades, as those markers in current clinical use are imperfect. The protein component of PCF has not yet been well characterised, as interrogation of the proteome is relatively new and technological advances are frequently being made[121]. Individual proteins such as thymosin-β4 and ubiquitin have been found to be significantly overexpressed in the tissue of IPMNs with high-grade dysplasia (p=0.011 and 0.0004,

respectively)[197]. Panels of proteins have also shown promise for the differentiation of mucinous and non-mucinous cysts. Elevations in any two of the 3-protein panel MUC5AC:WGA, MUC5AC:BGH and Endorepellin:WGA has shown good SN/SP for the identification of MCNs (Table 1.3)[195, 196]. Porterfield *et al.* utilised proteomic analysis by liquid chromatography-MS (LC-MS) to identify seven proteins shown to be consistently increased in the ductal fluid of PC patients compared to normal (AMYP, PRSS1, GP2-1, CCDC132, REG1A, REG1B, and REG3A), as well as one that was decreased (LIPR2), and validated these results by western blot[255]. A recent meta-analysis combined publicly available proteome and secretome data with the aim of identifying biomarkers of PC. While this analysis did not identify any protein that was shared by all of the 55 included secretome and proteome studies, by selecting proteins found in 2 or more studies an intersection between the two exposed 43 proteins common between proteome and secretome analyses[256]. Notably, 31 genes related to these secretome-proteins were shown to be upregulated in PC samples obtained from TCGA compared to control samples, while 39 such genes were revealed to be predictors of worse overall survival in PC[256].

As IPMNs are classified as mucinous cysts, it follows that their composition is partially composed of mucin proteins. Mucins, which are densely O-linked glycoproteins with a high molecular weight, play many roles in the maintenance of pancreatic health and subsequently, when altered as a result of malignancy can be important facilitators of tumorigenicity[215]. IPMNs are known to have a unique pattern of mucin expression, and this trait has been utilised in the subclassification of IPMNs[257]. Indeed, mucin proteins have been extensively investigated in the context of mucinous PCLs and evaluated as potential biomarkers but to largely no avail[215, 250, 257]. Moreover, while many studies have examined the mucin proteins of the cyst type, surprisingly, research has shown no significant pattern of RNA expression of mucin proteins identified in IPMNs[258]. The combination of mucin proteins into panels, as mentioned previously, has shown promise in the distinction of mucinous and non-mucinous cysts with good SN/SP[195, 196]. Though not a mucin, VEGF-A is also a glycoprotein and is known to be a key mediator of vascular growth[259]. Elevated levels of VEGF-A have been observed to indicate the presence of a benign serous cystic neoplasms with high SN/SP (Table 1.3)[194,

[259]. Furthermore, the addition of CEA levels to VEGF-A exhibited better still SN/SP for the identification of these cysts[194].

An important aspect of proteomic work in mucinous PCLs is the depletion of larger proteins, which in this case is not just IgG and albumin, but also the mucin proteins. The exclusion of larger, more abundant proteins by such immunodepletion steps increases assay sensitivity for smaller proteins that may not otherwise be detected. Depletion based on molecular weight is frequently employed, however the purity of the protein samples obtained by this method is generally poor. Indeed, the discovery of mucin-specific proteases that could aid depletion of these proteins appears to be making strides. A recent study by Malaker *et al.* identified a mucin-selective protease, StcE, which shows great promise in the selective digestion of human mucins from biological samples[260]. Interestingly, a recent study examined the protein component of IPMN cyst fluid supernatant and cell pellet, reporting that the cell pellet contains twice as many proteins as the supernatant and even contained over two thousand that were not identified in the supernatant[121]. This study opted to omit the immunodepletion step that is routinely used in proteomic analyses, and in doing so identified almost 4,000 proteins previously unknown in PCF[121]. This large, proteomic dataset has been deposited into the ProteomeXchange database, with the hope that it may prove a rich source of information for further IPMN studies[121]. Other online platforms of proteomic data, such as the Clinical Proteomic Tumour Analysis Consortium and PRoteomics IDEntification database allow users to upload their own data or examine those datasets submitted by others to supplement new research[261, 262].

Often overlooked in proteomic studies is the part that genomic changes play in the alteration of the proteome. While many somatic mutations have been recorded in breast cancer for instance, the role these mutations play in the alteration of the proteomic landscape remains to be elucidated [263]. A 2014 study analysed the proteomes of colon and rectal tumours, and integrated these data with that of the TCGA to perform proteogenomic analyses[264]. This investigation found that messenger RNA transcript abundance did not correspond with the difference in protein abundance observed between tumours, and that somatic variants exhibited lower protein abundance than germline variants[264]. Similar research conducted in 2016 showed how

the integration of proteomic and phosphoproteomic analyses with TCGA data for 77 genomically annotated breast cancers enabled the discovery of novel functional consequences of somatic mutations in this cancer type, and subsequently narrowed the scope of potential candidates for driver genes[263]. Such findings highlight important caveats of single-omics studies, in that the consolidation of multiple types of data allows for a better understanding of mechanisms involved and for the exclusion of those components that are deemed to play no active role and are merely resultant of the disease and not causative. In the context of IPMN stratification, these studies solidify the need for multi-omic data integration, as current individualised efforts have yet to prove fruitful.

1.7.5 Metabolomics

Metabolomics is the study of the set of metabolites present in a cell or tissue at a given timepoint and relates directly to the cellular processes taking place[265]. Metabolomic alteration of cancer cells has been regarded for nearly a century as one of the hallmarks of cancer[266]. Indeed, the switch of metabolic pathways observed in cancer cells is regarded as key for tumour growth, and is suspected to be selected for during transformation[266]. As such, it is important to consider that metabolomic studies could give novel insight into the mechanisms of cancer growth and that the detection of altered metabolite levels in PC could be used as a metabolic biomarker.

Recent studies conducted by Mayerle *et al.* identified a biomarker signature of 9 metabolites alongside CA19-9 in the blood using MS for the distinction of PC and chronic pancreatitis[267]. While not validated, this study showed the potential of this panel in both a training and test cohort, with SN/SP of 89.9%/91.3%[267]. Fahrmann *et al.* also utilised MS for the identification of a metabolite panel in the blood[183]. This panel was observed to distinguish PC from normal samples with moderate SN/SP (Table 1.3)[183]. Metabolic profiling combining MS and liquid chromatography techniques enabled the discovery of 55 metabolites that were differentially expressed in pancreatic tumours as compared to non-tumours (p<0.01)[268]. Further examination of these metabolites using weighted co-expression network analysis highlighted 8 fatty acid hubs that are highly connected and in a conserved lipid module that are decreased in PC tumours compared

to the surrounding non-tumour tissue[268]. Integration of transcriptomic data revealed 157 gene surrogates for this fatty acid set and showed that the expected lipid metabolism, particularly in the lipolytic pathway involving these gene surrogates, is significantly altered in PC[268]. These data suggest a dysregulation of the lipolytic network in PC which may play some role in tumorigenesis. Kynurenine, a metabolite known to be synthesized in response to immune activation, has shown promising ability to discern mucinous from non-mucinous PCLs with high SN/SP 90%/100%[181]. This metabolite, is surprisingly detected in lower levels in the PCF of MCNs compared to non-MCNs, suggesting some dampening of immune activation in MCNs. In that same study, Park *et al.* identified 10 metabolites that were differentially abundant in their validation cohort, 8 of which could not be matched to any known metabolite and mass spectrometry analysis was unsuccessful due to the low abundance[181]. Importantly, glucose levels in the PCF have also been observed to discriminate MCNs from non-MCNs, and a standard patient glucometer has been successful in this manner[181, 182]. If cystic glucose levels could be correlated to same in patient blood samples, this methodology could prove a less invasive manner of determining cyst type. However, such correlations would be highly unlikely given the plethora of factors which can influence blood glucose levels, such as presence of diabetes and patient fasting status.

Unfortunately, as metabolomics is a relatively new field of study, there is little research done in the context of PC that can be robust enough to evaluate its ability in a clinical setting. It is therefore of the utmost importance that the establishment of large metabolic databases, such as the Metabolomics Workbench, take place in order to enable large metabolic studies and subsequent integration of this information with other omics data to produce a robust biomarker panel[269].

1.8 Multi-omics as the key to biomarker identification

The complexity of cancer biology has been revealed through decades of research aimed at understanding this disease and attempting to treat it. Technological advances have enabled researchers to delve deeper into the pathways involved in cancer development and progression and has given new insights and inspired novel approaches for

treatment. In the case of PCL and PC characterisation, though these new techniques have been utilised to analyse the PCF, blood serum and even saliva of patients, no single methodology has proven sufficiently sensitive method for delineating these patients into defined categories. Multi-omics involves the integration of multiple layers of omics-type data to augment our understanding of disease and helps researchers to elucidate the flow of information, from the origin of the disease to the biological and functional consequences[172]. More modern hypotheses interrogate the complex and interconnected networks of disease by examining the different biological layers and their interactions with each other. This multi-layered approach has been demonstrated in the clinical setting with the rapid emergence of personalised medicine. By investigating multiple aspects of the PCF or blood serum, and treating these data as an interconnected system, rather than distinct and independent pieces, multi-omics could allow researchers to identify key pathways and players in disease stratification. CompCyst is a comprehensive test developed using machine learning techniques to guide the management of patients with PCLs[270].This test utilises selected clinical features such as symptoms, cyst size and location, as well as PCF genetic and biochemical markers, including cyst CEA levels and *KRAS* and *GNAS* mutation status[270]. Interrogation of multiple levels of patient data enabled cut offs for each marker to be determined based on the needs of the test, and the level of importance given to the sensitivity or specificity of each individual marker. The results of this study suggest that if CompCyst were applied to general PCL management, 60% of unnecessary surgeries could be avoided[270]. While these results seem promising, it is important to note that patients evaluated in this study were those most concerning for cancer and do not represent patients seen in routine clinical practice[270]. While more research is needed to examine the utility of this test in a normal clinical setting, this study shows that layering multiple levels of patient data can potentially improve management strategies for PCLs.

In terms of PCLs and PC, multi-omics opens the door to the possibility of a biomarker panel for characterisation, such that combined thresholds of several markers could prove more sensitive than a single marker alone. Indeed, there have been a vast number of studies executed with the aim of identifying biomarkers in PC, with many

genes, miRNAs and proteins having been identified by various groups and declared a potential biomarker[271]. While compilations of these data have been made, with some genes or proteins having been singled out by several papers, there remains no biomarker robust enough to distinguish different PCL subtypes and PC. It follows, therefore, that examination of these data where appropriate, and integration of these results, could provide a sufficient starting point for the identification of a biomarker panel. As mentioned previously, the addition of *GNAS* to *KRAS* testing for the diagnosis of IPMNs does not significantly increase diagnostic accuracy, however, the same study found that the combination of *GNAS* and *KRAS* mutational status with CEA testing does produce a significantly better accuracy of 86.2% (p<0.05)[178]. A 2015 multi-centre study retrospectively examined the PCF of 130 patients and identified molecular markers and clinical features that classified PCLs with a sensitivity of 90-100% and a specificity of 92-98%[272]. Using the Multivariate Organization of Combinatorial Alterations algorithm to identify composite clinical and molecular markers (subtle mutations, loss-of-heterozygosity, aneuploidy) of PCL type and grade, this study identified a panel of both clinical and molecular markers for the distinction of serous cystadenomas (SCA), solid-pseudopapillary neoplasm (SPN), MCNs and IPMNs. Furthermore, it was shown that these features could identify 67 of the 74 patients who did not require surgery, resulting in a reduction of unnecessary procedures by 91%. While the majority of the patients examined were IPMN, with 12 or less patients in each of the MCN, SCA and SPN cohorts, these results are promising indicators of what can be achieved through the integration of data. These results show great promise for the characterisation of PCLs and the stratification of patients for subsequent referral to surgery, and further studies in more robust, experimental validation cohorts will help to further elucidate the potential of this panel in the context of PC. Such research highlights the potential of biomarker panels and the combination of multiple types of markers to achieve PCL characterisation. The implication of heightened biomarker sensitivity resulting from a panel of markers compared to a single solitary biomarker remains intriguing and presents new avenues for research in this context.

A key example of the multi-omic nature of driver mutations in the context of PC is *KRAS*, which is mutated in ~90% of PC (Figure 1.5)[273]. Environmental factors, such as

smoking or alcohol consumption, can promote biochemical alterations to DNA at the epigenomic level, for example hypermethylation[274]. The addition of a methyl group to the CpG island of a DNA repair gene can cause silencing and subsequently result in reduced DNA repair proficiency, allowing a mutated *KRAS* codon to proceed from the genomic level to the transcriptomic level. The transcription of this *KRAS* mutation results in altered miRNA expression levels, and the mutated mRNA cannot be bound by the regulatory miRNA let-7, thus causing the aberrant translation of K-Ras protein[275-277]. The marked increase in K-Ras production promotes various signalling pathways, including PI3K, MAPK and the RAL-GEFs pathway[273]. GTP-bound K-Ras proteins can interact with, and influence the activity of, effector proteins causing downstream effects in many cellular pathways[273]. Moreover, *KRAS* mutant PCLs have been shown to have increased expression of the glucose transporter GLUT1 and subsequently elevated rates of glycolysis, indicating that *KRAS* mutations play a role in the metabolic switch observed in PC[278]. Indeed, the presence of *KRAS* mutations in PC have been shown to correlate with poor patient prognosis, and this can be attributed to the downstream effects seen in multiple omics layers as a result of this point mutation (Figure 1.5)[37]. This example illustrates the multi-omic nature of mutational drivers in cancer and the importance of disentangling each aspect in order to clearly observe the pathways affected and the impacts at each omics level.

As discussed above, each omics discipline has its own advantages and disadvantages, and can give information about many aspects of disease from metabolic signatures to proteomic profiles. It is only logical therefore to examine this extensive information in parallel, with the aim of revealing those attributes that can be considered robust and sensitive enough to work as a biomarker of patient risk. While this can seem a simple task, there are many statistical and analytical variables that need to be considered if such an endeavour is to be successful. The data produced by each omics discipline alone is generally vast and contains many confounding variables. When compiled, these data must be carefully handled to ensure the quality and accuracy of any results obtained. Typical analysis of a single omics data type is largely limited to correlations and tends to reflect the reactive processes of disease, rather than the causative[172]. This is primarily due to type of data collected, and an inability to infer

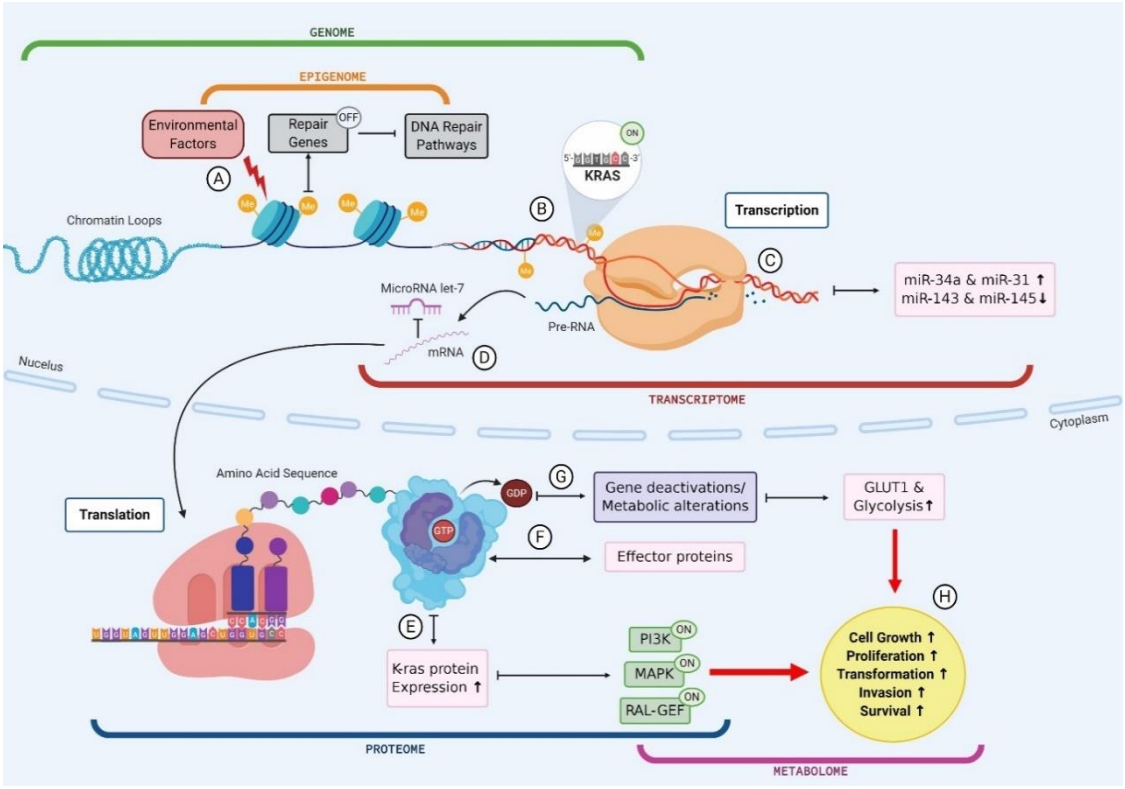


Figure 1.5 The multi-omic nature of *KRAS* mutations in PC. (**A**) Environmental factors cause biochemical alterations to the DNA such as hypermethylation. This can result in the silencing of repair genes and subsequently failure in DNA repair pathways; (**B**) Point mutations in a *KRAS* codon go unchecked as a result of DNA repair failure causing permanent activation of *KRAS* gene; (**C**) Mutant *KRAS* gene is transcribed into mRNA and subsequently results in an upregulation of miR-34a and miR-31 and a downregulation in miR-143 and miR-145; (**D**) mutant KRAS mRNA cannot be bound by regulatory miR let-7 and leaves the cell nucleus to be translated; (**E**) Mutant *KRAS* causes an increase in K-Ras protein expression, which causes activation of PI3K, MAPK and RAL-GEF pathways; (**F**) GTP bound KRAS interacts with various effector proteins and influences the localisation and activities of these effectors; (**G**) K-Ras proteins convert GTP to GDP which causes gene deactivations and metabolic alterations such as an increase in GLUT1 expression and subsequently an increase in glucose uptake via glycolysis; (**H**) Changes to cellular protein expression, gene activation and metabolic processes results in increased cell growth and proliferation, driving transformation.

whether any mutation or change in expression is a by-product of disease or the causative agent. Compilation of many data types can enable statisticians to tease out the causative and resultant factors observed in these data by enhancing the statistical depth and power of the dataset. In this way, sufficient statistical power obtained by having a large cohort is required for any successful omics study, in order to produce the most robust results[172].

There is no current gold standard for statistical modelling in multi-omics studies, and as such researchers should take great care when designing studies to ensure data analysis requirements are met[172, 221]. As large-scale studies are often confounded by technical artefacts, such as batch effects, which in most cases cannot be eliminated due to the nature of data collection, it is therefore paramount that the impact of these artefacts be taken into account before conclusions can be drawn from results[172, 227]. Exploratory data analysis is often used in the first instance to identify any potential artefacts by means of cluster analyses or dimensional reduction[279]. Dimensional reduction deconstructs the data, identifying new variables that explain the differences in observations and highlights results that are associated with technical artefacts[279]. Methods to deal with heterogeneous multi-omics datasets, such as Multi-Omics Factor Analysis (MOFA), can be employed to discover the principle source of heterogeneity in such datasets, allowing for more accurate statistical analyses to take place[280]. While the MOFA model is linear, and therefore lacks the ability to identify non-linear relationships between features, non-linear extensions of the MOFA model can be used to address this limitation[280]. An advantage of large omics databases is that such data can be reanalysed with multiple statistical approaches allowing researchers to fully interrogate their data. In order for common usage of datasets to take place, it is important that standardisation take place across all omics disciplines such that the same techniques and tools can be directly compared.

Multi-omics data can consist not just of multiple datasets examining multiple omics disciplines, but also numerous datasets where each examines the same discipline in independent cohorts[281]. Regardless, the data in each case will require an understanding of all variables involved and suitable statistical models in order to integrate these data. Appropriate statistical models will enable researchers to discern

whether certain features contribute independently to the inception of disease or are a function of one another. While graphic depiction of omic results is common among studies, as it allows interpretation of global variance structure and identification of the key features across a dataset, this type of modelling does not allow for the comparison of many datasets simultaneously[281]. Methods such as multiple co-inertia analysis (MCIA) allow for the analysis of multiple omics datasets; MCIA is an exploratory method used to investigate relationships between multiple high dimensional omics datasets by transforming various sources of data onto the same scale[281]. It enables the visualisation of several datasets containing multi-omics features on the same plane in order to identify relationships, however, MCIA is limited by its inability to reveal functional insights[281]. A reduction technique specifically designed with biomarker identification in mind, Data Integration Analysis for Biomarker discovery using Latent Components incorporates this high-dimensional multi-omics data while discriminating between phenotypic groups in order to identify those molecular factors which distinguish the groups[282]. However, with a lack of standardisation for the collection and interpretation of omics-type data, the multi-omics field faces huge barriers to data integration and analyses.

The LinkedOmics database contains multi-omics data within and across 32 cancer types for over 11,000 patients from TCGA[219]. This platform is the first of its kind, and integrates data generated by the CPTAC for select TCGA tumour samples and has therefore, over a billion data points [219]. The database allows users to apply comprehensive analyses on these data by use of three distinct modules: LinkFinder, which identifies associations between clinical and molecular attributes of interest; LinkCompare, which enables comparisons of those associations obtained via LinkFinder; and LinkInterpreter, where identified associations are further explored through pathway and network analysis. Through the use of several case studies, examining properties of individual cancers to reveal functional impacts of somatic mutations or copy number alteration on the expression of mRNA and protein, or performing pan-cancer analysis to investigate survival-associated gene expression signatures, the power of such multi-omics platforms can be seen[219, 283]. While this database only includes data from TCGA and CPTAC, the extension of its data collection for more cancer types and omics

platforms could enable the execution of robust and highly powered multi-omics studies for many cancer types. The TCGA data were utilised in a recent multi-omics study of PC, where the integration of DNA copy number variation, methylation, mRNA, and simple nucleotide variation data enabled the identification of four distinct molecular subgroups of PC (iC1, iC2, iC3, iC4)[284]. The iC1 subgroup was shown to have a better prognosis, higher immune cell infiltration and better genomic stability compared to the other groups. Furthermore, this multi-omics study identified three new genes (*GRAP2*, *ICAM3* and *A2ML1*) that were shown to correlate with prognosis in PC.

A 2019 study utilised a multi-omics approach comprising exome sequencing, transcriptomics, quantitative proteomics, karyotyping and metabolic status to evaluate the epithelial-mesenchymal plasticity of two sister breast cancer cell lines, identifying novel driver mutations, chromosomal changes, gene deletions/ amplifications, alterations in gene expression and metabolic reprogramming[285]. Indeed, this integrative multi-omics approach provided mechanistic insights into how driver mutations have potentially influenced phenotypic differences observed between the two cell lines, and subsequently identify potential targets to reverse mesenchymal-epithelial transition in these cell models[285].

With the exception of somatic mutations, the original human DNA sequence is unaltered throughout life, being unaffected by environmental or developmental factors[172]. As such, it is generally assumed that any disease-related genetic mutation observed is causal and not a result of disease. This assumption is often the reason for the level of difficulty attributed to distinguishing the causative agent of disease from those effects that are created as a result of disease[172]. Critical to the identification of genetic mutations that have the potential for use as a biomarker, is the understanding of whether that mutation is a driver mutation[286]. It is known that malignant tumours can carry multiple genetic mutations, and that a substantial proportion of these do not drive a malignant phenotype. It follows therefore, that not all mutations observed in the tumour are causative, and perhaps subsequently not found in all tumours of that cancer type owing to individual heterogeneity. The tumour profiles of clinically identical patients have been observed to share as few as one single genetic mutation[224]. In such situations, the integration of data from multiple types of omics can help to reveal the

biochemical pathways involved and ultimately those genes that are playing an active role in tumour growth. This layering of data in a multi-omic approach can help to tease out such details and begin to show the picture in its entirety, not only indicating the causative agent, but also the downstream pathways and interactions involved.

1.9 Rationale

PC is an aggressive disease with extremely poor survival rates. The discovery of precursor lesions often occurs too late, and patients are left with few treatment options. PCLs are a highly diverse group of lesions containing both non-malignant and pre-malignant subtypes, and there exists no robust method for distinguishing PCLs. Importantly, the fluid within these lesions is, as of yet, unexplored. As such, the factors within this fluid and the potential role it may play in the transformation of PCLs to PC is unknown. Multi-omics provides a more comprehensive insight into the mechanisms and pathways involved in cancer and has great potential for use in PCL risk stratification, and elucidating the role of PCF in PC progression.

1.10 Hypothesis

Low- and high-risk PCLs are biologically different, and this difference may enable the risk stratification of these patients through the use of multi-omic profiling and functional characterisation of the PCF within these PCLs.

1.11 Specific aims

1. Perform a systematic review and meta-analysis to identify potential biomarkers from the literature which have shown promise in PC diagnosis.
2. Profile the proteome and transcriptome of PCF to generate a promising novel multi-omic biomarker panel for PCL risk stratification.
3. Profile the proteome and transcriptome of the serum to generate a promising novel multi-omic biomarker panel for PCL risk stratification.
4. Integrate the PCF-based and serum-based multi-omic panels to create a cross-biofluid multi-omic panel for PCL risk stratification
5. Investigate the biological activity of PCF and the potential influence it may have on normal pancreatic cells and their acquisition of the hallmarks of cancer.

Chapter 2.

Materials and Methods

2.1 Desk-based materials and methods

2.1.1 PRISMA guidelines

A systematic review of blood-based biomarkers for the diagnosis of PDAC was conducted in accordance with PRISMA standards (Appendix 1)[287]. The review was registered with PROSPERO prior to data extraction (CRD42020207241).

2.1.2 Search strategy and inclusion criteria

Academic databases MEDLINE, Web of Science, and EMBASE were searched using individualized search strategies, containing both medical subject headings and text words, created by TCD Medical Librarian Mr. D. Mockler (Table 2.1). Publications were limited to those written in English, conducted in human participants and published on or before July 20th, 2020. No limit was placed on the date of publication prior to the date of the literature search. Studies that were identified were exported to Endnote X9 and subsequently imported to www.covidence.org for review, where covidence analytics removed any duplicates.

Human studies reporting on blood-based single biomarkers or multi-biomarker panels for the diagnosis of PDAC were included. As PDAC is considered to be synonymous with PC, where a study did not specify the subtype of PC, it was assumed that this study was referring to PDAC and it was therefore included in the review. Only studies examining primary PDAC in patients of any stage, with or without metastasis, were included. Both PDAC and control cohorts must have had a minimum of 15 patients to be included in the study. This was to ensure sufficient statistical power in each study according to power calculations[288]. Studies reporting on image-based diagnostic methods, such as EUS or CT scans, or pancreatic cyst fluid and/or pancreatic tissue-based biomarkers, were excluded as they were not deemed directly comparable with blood-based biomarkers. Studies reporting on biomarkers of any omic compartment, such as proteomics or transcriptomics, were included. Multi-biomarker panels consisting of biomarkers from different omic compartments were also included. Studies where the patient data were obtained from an online database were excluded to avoid examining biomarkers that were assessed in the same patient cohorts. Where all patients in a study, PDAC and controls, had pre-existing conditions, this study was

Table 2.1 Individualised search strategies for academic databases.

Database	Search Strategy
EMBASE	'pancreas adenocarcinoma'/exp (pancrea* NEAR/3 adenocarcinoma*):ti,ab PDAC:ti,ab #1 OR #2 OR #3 'biological marker'/exp OR 'tumor marker'/exp (Biomarker* OR 'biological indicator*'):ti,ab ((biological OR serum OR Immunologic OR diagnostic OR tumo?r OR cancer) NEAR/2 (marker* OR biomarker* OR bio-marker*)):ti,ab #5 OR #6 OR #7 'diagnosis'/exp Diagnos*:ti,ab #9 OR #10 #4 AND #8 AND #11 'editorial'/exp OR 'erratum'/de OR 'letter'/exp OR 'conference abstract':it OR 'conference review':it #12 NOT #13
Medline	Carcinoma, Pancreatic Ductal/ OR (exp Pancreatic Neoplasms/ AND exp *Adenocarcinoma/) (pancrea* adj3 adenocarcinoma*).ti,ab. PDAC.ti,ab. or/1-3 exp Biomarkers/ (Biomarker* OR biological indicator*).ti,ab. ((biological OR serum OR Immunologic OR diagnostic OR tumo?r OR cancer) adj2 (marker* OR biomarker* OR bio-marker*)).ti,ab. or/5-7 exp Diagnosis/ diagnos*.ti,ab. or/9-10 4 AND 8 AND 11
Web of Science	TS =(((pancrea* NEAR/2 adenocarcinoma*) OR PDAC) AND ((Biomarker* OR "biological indicator*") OR ((biological OR serum OR Immunologic OR diagnostic OR tumo?r OR cancer) NEAR/2 (marker* OR biomarker* OR bio-marker*))) AND Diagnos*)

excluded. Only studies which reported a significant result were included to ensure the results did not skew the downstream analyses. A full list of inclusion/exclusion criteria is given in Table 2.2. Title and abstract screening was conducted independently by two randomly assigned reviewers, and included studies were then subject to full-text screening in the same manner. Any disagreements were discussed and settled by two senior reviewers (L.E. Kane and G.S. Mellotte). Where the full-text of an article could not be located, corresponding authors were contacted to request access to the article.

2.1.3 Data extraction and risk of bias assessment

An extraction template in Excel was piloted by two reviewers for a small subset of papers before being finalised. Reviewers extracted data into their own preoptimized template in Excel, with data compilation being carried out once all studies had been extracted. Data extracted included information such as study details (title, corresponding author name and email address, country of study, dates conducted); biomarker details (biomarker name(s), biological properties, detection platform); patient cohort details (number patients per cohort, sex and age breakdown, condition); and reported statistics (analysis performed, p-value, sensitivity, specificity, AUC). A complete list of extracted data fields is included in Table 2.3. Data were extracted such that each row represented an individual biomarker or multi-biomarker panel having been assessed in one set of patient cohorts. Where a biomarker or multi-biomarker panel was assessed in multiple cohorts, for example, training and validation cohorts, these data were extracted into separate rows. As a result, larger studies which examined multiple biomarkers in multiple patient cohorts, represent more rows of data than smaller studies.

All included studies were assessed for quality and risk of bias (RoB) using the QUADAS-2 tool[289]. The QUADAS-2 tool provides an assessment for the level of bias an individual study will introduce into the systematic review based on the nature of its design. The selected questions from the four main domains of the tool were amended to align with the review, as per the QUADAS-2 guidelines. This assessment was carried out by reviewers in tandem with the data extraction. Responses were given as either “yes,” “no”, or “unclear” and domains were subsequently scored as “high,” “low”, or “unclear” RoB. Selected questions for the RoB assessment are included in Table 2.4.

Table 2.2 Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
• Human, primary case-control study. • Diagnostic cohort studies. • Primary pancreatic ductal adenocarcinoma (PDAC) of any stage. • Diagnostic biomarker for PDAC. • Must have a control cohort of healthy and/or benign patients. • Minimum 15 patients per cohort. • Must report some summary statistic for the performance of the biomarker. • Must be a whole blood/serum/plasma biomarker. • Must report a significant result.	• Narrative or systematic review. • Patient data is obtained from online database such as The Cancer Genome Atlas (TCGA). • All participants, PDAC and controls, have pre-existing conditions that may confound the index test, e.g., diabetes or pancreatitis. • Studies with a retrospective design, where investigators chose participants based on a review of case notes/ archival information. • Case reports or series. • Conference proceedings, where a full text is not available.

Table 2.3 Complete list of extracted data fields.

Category	Field to be completed
Paper Information	○ Covidence study ID number ○ Paper title ○ Corresponding author name ○ Corresponding author email address ○ Country in which study was conducted ○ Study funding source ○ Conflicts of interest ○ Ethical approval
Study Design	○ Study design (prospective/retrospective) ○ Fluid type (serum/plasma/whole blood) ○ Time of blood draw relative to treatment/FNA /surgery ○ Start date ○ End date ○ PDAC specified ○ Paper examined more than one biomarker ○ Test platform ○ Testing/training /validation cohort ○ Blinded
Biomarker details	○ Biomarker(s) name ○ Number of biomarkers ○ Biological Properties ○ Single marker or multi-marker panel ○ Number of omics compartments for each marker/panel ○ Whether CA19-9 or novel
Patient Cohorts	○ PDAC / Healthy / Benign ○ Details of how control cohort was chosen ○ Details of benign conditions and breakdown of numbers ○ Number of patients ○ Male/female breakdown ○ Age range/mean/median ○ Method of PDAC confirmation ○ Stage details for PDAC ○ Whether patients received treatment prior to blood draw ○ Treatment details
Statistical details	○ Confounding factors of note ○ Statistical analyses used ○ Sensitivity/Specificity ○ AUC/ROC ○ P-value ○ Negative predictive value/positive predictive value
Extra Notes	○ Name of reviewer who extracted data ○ Additional notes

Table 2.4 QUADAS-2 quality and risk of bias assessment questions.

Domain 1 – Patient Selection **Risk of Bias** Q1. Was a consecutive or random sample of patients enrolled? Q2. Did the study avoid inappropriate exclusions? Q3. Was a 'two-gate' design avoided? ***Could the selection of patients have introduced bias?***	 Yes/No/Unclear Yes/No/Unclear Yes/No/Unclear High/Low/Unclear
Domain 1 – Patient Selection **Concerns about applicability** ***Are there concerns that the included patients do not match the review question?***	 High/Low/Unclear
Domain 2 – Index Test **Risk of Bias** Q1. Were the index test results interpreted without knowledge of the results of the reference standard? Q2. If a threshold was used, was it pre-specified? ***Could the conduct or interpretation of the index test have introduced bias?***	 Yes/No/Unclear Yes/No/Unclear High/Low/Unclear
Domain 2 – Index Test **Concerns about applicability** ***Are there concerns that the index test, its conduct, or interpretation differ from the review question?***	 High/Low/Unclear
Domain 3 – Reference Standard **Risk of Bias** Q1. Is the reference standard likely to correctly classify the target condition? Q2. Were the reference standard results interpreted without knowledge of the results of the index tests? ***Could the reference standard, its conduct, or its interpretation have introduced bias?***	 Yes/No/Unclear Yes/No/Unclear High/Low/Unclear
Domain 3 – Reference Standard **Concerns about applicability** ***Are there concerns that the target condition as defined by the reference standard does not match the question?***	 High/Low/Unclear
Domain 4 – Flow and Timing **Risk of Bias** Q1. Was there an appropriate interval between index test (sample collection) and reference standard? Q2. Did all patients receive the same reference standard? Q3. Were all patients included in the analysis? ***Could the patient flow have introduced bias?***	 Yes/No/Unclear Yes/No/Unclear Yes/No/Unclear High/Low/Unclear

Given the high volume of included papers, studies were only extracted and RoB assessed by a single reviewer. To assess the accuracy of this process a random selection of 25 papers, to represent 10% of the total number of studies included, was generated using R. These papers were extracted and RoB assessed by a second reviewer, with both data extractions and RoB assessments subsequently checked for mistakes and/or missing information by a senior reviewer (L.E. Kane or G.S. Mellotte). Extraction accuracy was then calculated for each paper using the total number of correct datapoints as a percentage of the total number of datapoints. If there were disagreements between both senior reviewers with regards to the eligibility of a study, or a discrepancy of greater than 10% for the accuracy of the RoB or data extraction, a third reviewer was consulted (S.G. Maher) to settle disputes.

2.1.4 Data clean-up and meta-analysis

Data filtering and clean-up was conducted in Microsoft Excel. To calculate uniform 95% confidence intervals (CI) for reported sensitivity and specificity values, 2 × 2 contingency tables were constructed using the extracted values for sensitivity, specificity, number of patients with PDAC, and number of control patients. Both two-level and three-level meta-analyses were run on the data to identify the model of best fit. The three-level model had significantly lower Akaike Information Criterion or AIC, and Bayesian Information Criterion or BIC values, and was therefore deemed to be the most appropriate model (p<0.0001). A multivariate three-level meta-analysis with subgroup moderators was run in R (v 1.3.959) with the 'metaphor' package (v. 3.0-2) using reported AUC values as effect size [290, 291]. AUC values <0.5 were removed from the analysis as they are regarded as diagnostically useless. Studies where PDAC was not specified (n=25) were excluded from the primary meta-analysis. A secondary meta-analysis, including all 250 studies, was also conducted. Figures were created in GraphPad Prism (v9.2.0) and Microsoft PowerPoint (v2108).

2.2 Lab-based materials and methods

2.2.1 Reagents and materials

All chemicals and reagents used were purchased from Sigma-Aldrich (Wicklow, Ireland), unless otherwise stated, and stored as per the manufacturer's instructions. All plastic consumable materials were purchased from Sarstedt (Numbrecht, Germany) unless otherwise stated. All Eppendorf tubes were centrifuged using the Eppendorf 5417R Refrigerated Centrifuge (Eppendorf, Hamburg, Germany). Larger volume tubes (15 mL and 50 mL) and microplates were centrifuged using the Thermo Scientific™ Megafuge™ 40 Centrifuge (Fisher Scientific, Dublin, Ireland).

2.2.2 Ethical approval

Prior to commencement of this study, ethical approval was obtained from the HSE Research Ethics Committee for the collection of patient samples in three designated sites: St. James's Hospital, St. Vincent's University Hospital and Tallaght University Hospital. This study was carried out in accordance with the Declaration of Helsinki ethical guidelines for medical research involving human subjects.

2.2.3 Patient recruitment

All patients over 18 years of age who had been diagnosed with a PCL in any of the three designated sites for this study, and who had been referred for EUS-guided FNA as part of their standard medical care, as per current best practice, were approached to participate in the study. Written informed consent was obtained for each patient prior to sample collection.

2.2.4 Patient serum collection and storage

Peripheral venous blood samples were collected in Vacuette 6 mL CAT Serum Sep Clot Activator tubes (Greiner Bio-One, Stonehouse, UK) from patients prior to endoscopy. Blood samples were allowed to clot at room temperature (RT) for 30 min before being centrifuged at 2,500 × g for 15 min at 4°C. The upper serum layer was then removed and

aliquoted into 1 mL volumes. These aliquots were labelled with an anonymous patient identifier before being snap frozen in liquid nitrogen and stored at -80°C. Locations of each sample were recorded in a dedicated biobank log book.

2.2.5 Patient PCF collection and storage

Collection of PCF was carried out by EUS-guided FNA of the PCL. PCF samples were transported on wet ice before being centrifuged at 2,500 × *g* for 15 min at 4°C. The cell pellet was discarded to waste and the supernatant was transferred into a fresh 15 mL tube before being drawn 20 times through a 20-gauge needle (Becton Dickinson, NJ, USA), and subsequently a 25-gauge needle (Becton Dickinson, USA), in order to reduce sample viscosity and allow for a homogenous sample. The PCF was then aliquoted into 250 µL volumes and labelled with an anonymous patient identifier before being snap frozen in liquid nitrogen and stored at -80°C. Locations of each sample were recorded in a dedicated biobank log book.

2.2.6 Sonication of patient PCF

Sonication of PCF samples was required prior to some assays to break down large mucin proteins and reduce sample viscosity. Sonication was carried out using the Fisherbrand™ Model 120 Sonic Dismembrator (Fisher Scientific, Ireland) along with the 1.5 inch Fisherbrand™ Cup Horn (Fisher Scientific, Ireland) for this model. Cryovials containing PCF samples were placed in a holder above the cup horn. The cup horn was filled with cold (4°C) deionised H_2O (dH_2O) until the meniscus was in line with that of the sample inside the cryovial. Samples were sonicated for two × 3 sec bursts at 25% amplification, with a 5 sec pause in between bursts. The dH_2O inside the cup horn was changed between samples to ensure samples were kept cold and there was no heat build-up.

2.2.7 Proteomic profiling of PCF

2.2.7.1 Protein quantification

The protein concentration of PCF samples was assessed by Pierce™ BCA Protein Assay (Fisher Scientific, Ireland) as per the manufacturer's instructions. Briefly, a bovine serum albumin standard was serially diluted from the provided stock solution (2000 µg/mL) at the recommended concentrations to create a standard curve, with the final standard containing no bovine serum albumin (blank). A volume of 25 µL of each standard and sample dilution was plated in a clear, flat-bottomed 96-well microplate in triplicate. Working reagent was prepared by mixing 50 parts of BCA Reagent A with 1 part BCA Reagent B. A volume of 200 µL of working reagent was added to each well and the plate was mixed on a plate shaker for 30 sec. The plate was then covered and incubated at 37°C for 30 min. Once cooled to RT, the absorbance at 562 nm was measured using the GloMax Explorer microplate reader (Promega, Madison, WI, USA).

2.2.7.2 Preparation of SP3 magnetic beads

In a 1.5 mL Eppendorf tube, 15 µL each of hydrophobic and hydrophilic Speedbead Magnetic Carboxylate Modified Particles (SP3 magnetic beads) (Cytiva, Marlborough, MA, USA) per sample were combined, vortexing briefly to mix. The tube was then placed on the DynaMag™ stand (ThermoFisher, London, UK) for 2 min. Once the beads had been successfully magnetised to the side of the tube, the waste supernatant in the bottom of the tube was discarded by carefully pipetting. The tube was then removed from the stand and the beads were resuspended in 1 mL of molecular grade H_2O, vortexing briefly to wash the beads. The tube was placed in the DynaMag™ stand for a further 2 min, before discarding the supernatant and washing with 1 mL of molecular grade H_2O for a second time. The tube was placed on the DynaMag™ stand a final time for 2 min, the supernatant was discarded and the beads were resuspended in the starting volume of molecular grade H_2O. The prepared beads were stored at 4°C until required.

2.2.7.3 Sample lysis, reduction, alkylation

A uniform quantity of protein was added to a 1.5 mL Eppendorf tube for each sample. An equal volume of 2X sample buffer [300 mM NaCl, 100 mM tris (pH 8.0), 3 mM $MgCl_2$, 2% Triton-X100 (Fisher Scientific, Ireland) and 1 tablet of Complete Mini Protease Inhibitor in molecular grade H_2O] was added to each sample. Tubes were incubated on wet ice for 20 min, vortexing every 5 min. Samples were then centrifuged at 13000 RPM for 7 min at 4°C. The supernatants were transferred to a new 2 mL Eppendorf tube and topped up to 500 µL with lysis buffer [6 M urea, 2 M thiourea, and 50 mM MOPS in molecular grade H_2O]. A volume of 25 µL of 0.2 M dithiothreitol was added to each sample and allowed to incubate on a thermoshaker at 700 RPM for 15 min at 30°C. Samples were allowed to cool to RT before 25 µL of 0.4 M iodoacetamide was added to each tube. Samples were then incubated on a thermoshaker at 700 RPM for 15 min at RT in the dark. Finally, 1.2 mL of 100% acetonitrile was added to each sample and they were immediately processed via tryptic digestion.

2.2.7.4 Tryptic digestion

A volume of 20 µL of the magnetic bead emulsion prepared in section 2.2.7.2 was added to each sample prepared in section 2.2.7.3, and tubes were placed on a rotation mixer at RT for 1 h. After this time, the samples were moved to the DynaMag™ stand for 2 min before discarding the supernatant. Tubes were then removed from the stand and 200 µL of 70% (v/v) ethanol was added to each sample, vortexing briefly to mix. Tubes were placed back on the DynaMag™ stand for a further 2 min, and supernatant discarded again. Once removed from the stand, 200 µL of 100% acetonitrile was added to each sample and mixed by vortexing. Samples were placed on the DynaMag™ stand a final time, before discarding the supernatant. When the samples had been removed from the stand, 50 µL of 50 mM ammonium bicarbonate was added to each tube. A volume of 4 µL of Sequencing Grade Modified Trypsin (MyBio Ltd, Kilkenny, Ireland) was added to each tube. Samples were then incubated overnight on a thermoshaker at 500 RPM and 37°C.

2.2.7.5 SP3 peptide clean-up and elution

Following the tryptic digestion described in section 2.2.7.4, tubes were quick spun to collect the liquid in the bottom of the tube. Samples were then pipetted up and down gently to resuspend the beads and avoid creating bubbles. A further 10 µL of fresh magnetic bead emulsion created in section 2.2.7.2 was added to each sample, along with 1 mL of 100% acetonitrile. Tubes were then incubated on the rotation mixer for 18 min at RT, before being placed on the DynaMag™ stand for 2 min and the supernatant discarded. Samples were then removed from the stand and resuspended in 200 µL of 100% acetonitrile, gently pipetting up and down to mix. Tubes were placed back on the DynaMag™ stand for 2 min before removing the supernatant to waste. Samples were removed from the stand and 20 µL of molecular grade H_2O was added to elute the peptides from the beads. The magnetic beads were vortexed intermittently for 5 min at RT before being placed on the DynaMag™ stand a final time for 5 min. The eluted peptide supernatant was then carefully transferred to a fresh 1 mL Eppendorf tube to avoid the collection of any magnetic beads.

2.2.7.6 Pierce quantitative colorimetric peptide assay

Peptide concentrations of the peptide elutants were assessed by Pierce™ Quantitative Colorimetric Peptide Assay (Fisher Scientific, Ireland) as per the manufacturer's instructions. Briefly, a Peptide Digest Assay Standard was serially diluted from the provided stock solution (1000 µg/mL) at the recommended concentrations to create a standard curve, with the final standard containing no Peptide Digest Assay Standard (blank). A volume of 20 µL of each standard and sample dilution were plated in a clear, flat-bottomed 96-well microplate in triplicate. Working reagent was prepared by mixing 50 parts of Colorimetric Peptide Assay Reagent A with 48 parts of Colorimetric Peptide Assay Reagent B and 2 parts of Colorimetric Peptide Assay Reagent C. A volume of 180 µL of working reagent was added to each well and the plate was mixed on a plate shaker for 30 sec. The plate was then covered and incubated at 37°C for 15 min. Once cooled to RT, the absorbance at 480 nm was measured using the GloMax Explorer microplate reader (Promega, WI, USA).

2.2.7.7 Peptide preparation for LC-MS

A volume of 2 µL of 1% (v/v) formic acid was added to each peptide sample eluted in section 2.2.7.5. Peptide sample dilutions of 100 ng/mL in 20 µL were prepared in 0.1% (v/v) formic acid in mass-spec vials. Samples were stored at -20°C before being transported to the Conway Institute in University College Dublin on dry ice, where they were stored at -20°C before being run by one of their technicians on the ThermoFisher Q-Exactive MS (ThermoFisher, UK) coupled to a Bionex Ultimate 3000 (RSLCnano) chromatography system. The tryptic peptides were separated on a reversed-phase C18 column packed in-house (8 cm x 75 µm ID; C 18 , 3.0 µm) (ReproSil-Pur 120 Dr Maitsch GmbH.) and separated at a constant flow rate of 250 nL/min by an increasing acetonitrile gradient. Mobile phases were 0.5% (v/v) acetic acid, 2% (v/v) acetonitrile, 97.5% (v/v) water (phase A) and 0.5% (v/v) acetic acid, 2% (v/v) water, 97.5% (v/v) acetonitrile (phase B). The peptides were separated by a gradient starting from 1% of mobile phase B and increased linearly to 30% for 58 min at a flow rate of 250 nL/min. The mass spectrometer was operated in data dependent TopN 12 mode, with the following settings: mass range 320-1600 Th; resolution for MS1 scan 70,000; AGC target 3e6; resolution for MS2 scan 17,500; AGC target 5e4.

2.2.8 Proteomic profiling of patient serum

2.2.8.1 Immunodepletion of patient serum

Immunodepletion of serum samples was carried out using the Proteome Purify 12 Human Serum Protein Immunodepletion Resin kit (R&D Systems, Minneapolis, MN, USA) as per the manufacturer's instructions. Briefly, samples were allowed to thaw on wet ice before transferring 10 µL of each sample to individual 1.5 mL Eppendorf tubes. The Immunodepletion Resin was thoroughly vortexed to homogenise the reagent before use. A volume of 1 mL of the resin was then added to each tube before being placed on the rotation mixer for 1 hr at RT. After this time, 500 µL of this suspension was placed into the upper chamber of a Spin-X Filter Unit (Fisher Scientific, Ireland) and centrifuged for 2 min at 2,000 × g. The immunodepleted elutants were moved to a fresh 2 mL Eppendorf, and the final 500 µL of suspension was added into the upper chamber of the same Spin-X Filter Unit and centrifuged under the same conditions. The

immunodepleted elutants were combined for a final volume of 400-500 µL per patient, and the used Immunodepletion Resin and Spin-X Filter Units were discarded. Samples were processed immediately for acetone precipitation, and remaining volumes were stored at -80°C.

2.2.8.2 Acetone precipitation

In 2.5 mL Eppendorfs, 400 µL of immunodepleted samples generated in section 2.2.8.1 were combined with 2 mL of cold (-20°C) 100% acetone and stored at -20°C overnight. After this time, samples were centrifuged at 15,000 × g for 30 min at 4°C and the supernatant discarded to waste. A further 2 mL of cold 50% (v/v) acetone was added to each sample and thoroughly vortexed to resuspend the pellet, before being centrifuged at 15,000 × g for 30 min at 4°C. The supernatant was discarded to waste and a final 2 mL of cold 50% (v/v) acetone was added to each tube and centrifuged at 15,000 × g for 30 min at 4°C. The tubes were then left open and the pellet was allowed to air dry for 24 h at RT. Once dry, the protein pellet was immediately processed for LC-MS.

2.2.8.3 Processing and preparation of precipitated proteins for LC-MS

Protein pellets generated in section 2.2.8.2 were immediately processed using the PreOmics iST 96x kit (PreOmics GmbH, Munich, Germany) as per the manufacturer's instructions. Briefly, 50 µL of Lyse Buffer was added to each pellet and incubated at 95°C on a thermoshaker at 1,000 RPM for 10 min. The adapter plate was used to place the cartridge on top of the waste plate, and all wells were labelled. Samples were then transferred to the cartridge and cooled to RT. The Digest Enzyme was reconstituted in the Resuspend Buffer and placed on a thermoshaker at RT and 500 RPM for 10 min. Once homogenised, 50 µL of Digest Enzyme was added to each well of the cartridge and placed on a thermoshaker at 37°C and 500 RPM for 3 h. After this time, 100 µL of Stop Buffer was added to each well of the cartridge and placed on a thermoshaker at RT and 500 RPM for 1 min. The cartridge was then centrifuged at 2,250 × g for 3 min to allow the solution to flow through into the lower waste plate. A volume of 200 µL of Wash Buffer 1 was added to the cartridge and this was centrifuged again under the same

conditions. This was repeated once more with Wash Buffer 2, and the cartridge was then removed from the waste place and placed on top of the collection plate. A volume of 100 µL of Elute Buffer was added to each well of the cartridge, and the cartridge was centrifuged again at 2,250 × g for 3 min. A further 100 µL of Elute Buffer was added to the cartridge and the centrifugation step repeated. The cartridge was then discarded and the collection plate containing the elutants was sealed with a 96-well plate seal and stored at -80°C. The plate was then transferred on dry ice to the National Institute for Cellular Biotechnology in Dublin City University. Here, samples were vacuum evaporated at 45°C until completely dry. Once dry, LC-Load Buffer was added to each well to give a final peptide concentration of 1 g/L. The collection plate was then placed on a thermoshaker at RT and 500 RPM for 5 min. Peptides were then frozen at -80°C until they could be run on the MS. An UltiMate 3000 nano RSLC (ThermoFisher, UK) system interfaced with an Orbitrap Fusion Tribrid Mass Spectrometer (ThermoFisher, UK) was used for LC-MS/MS analysis. A volume of 2 µL from each sample was loaded onto a PepMap100, C18, 300 µm × 5 mm trapping column using a flow rate of 25 µL/min with 2% (v/v) acetonitrile and 0.1% (v/v) trifluoroacetic acid for 3 min. Each sample was then resolved onto an Acclaim PepMap 100, 75 µm × 50 cm, 3 µm analytical column. A binary gradient of: solvent A (0.1% (v/v) formic acid in LC-MS grade water) and solvent B (80% (v/v) acetonitrile, 0.08% (v/v) formic acid in LC-MS grade water) using 2–32% B for 50 min, 32–90% B in 5 min and holding at 90% for 5 min at a flow rate of 300 nL/min was used to elute peptides. A column temperature of 47°C and a voltage of 2.0 kV was used for peptide ionization.

2.2.9 Quantification of soluble protein levels by sandwich ELISA

Soluble protein concentrations were measured by sandwich ELISA (Assay Genie, Ireland) as per the manufacturer's instructions. Briefly, the standard was serially diluted from the provided stock solution at the recommended concentrations to create a standard curve, with the final standard containing no protein (blank). Samples were diluted with sample diluent based on the expected concentrations observed in the literature. A volume of 100 µL of each standard and sample dilution was plated in the provided, pre-coated 96-well microplate in duplicate and incubated at 37°C for 2 h. The samples and

standards were then carefully removed from the plate before adding 100 µL of Detection Reagent A to each well and incubating at 37°C for 1 h. The reagent was then discarded to waste and the plate was washed by adding 350 µL of Wash Buffer to each well, swirling and discarding to waste. This wash step was repeated twice before adding 100 µL of Detection Reagent B to each well and incubating at 37°C for 1 h. The plate was washed five times and 90 µL of Substrate Solution was added to each well. The plate was Incubated for 15-25 min in the dark at 37°C. The reaction was stopped when the colour development was sufficient by adding 50 µL of Stop Solution to each well. The absorbance at 450 nm was measured using the GloMax Explorer microplate reader (Promega, WI, USA).

2.2.10 HTG Whole Transcriptome Sequencing of PCF and serum

HTG Whole Transcriptome Sequencing of samples was conducted by HTG Molecular Diagnostics, Inc. following their established procedures in their Tucson, Arizona facility. Samples and multi-tissue control technical controls were randomised prior to placement on the HTG EdgeSeq 96-well mico-titer plate (sample plate) in order to reduce intra-plate biases. HTG EdgeSeq Lysis Buffer was added to lyse the sample making the RNA available to subsequently bind to corresponding target specific nuclease protection probes (NPPs). The lysed samples were then transferred to the sample plate. Target capture was done via HTG EdgeSeq chemistry. Briefly, an excess amount of NPPs were added to the lysed samples in the sample plate and hybridized to the target miRNA. S1 nuclease was then added in order to digest any non-hybridized RNA and excess NPPs. This produced a stoichiometric amount of target miRNA NPP duplexes. After the S1 digestion was completed, the processed samples were transferred to a new 96-well sample plate with a v-bottom, referred to as the stop plate. S1 digestion was terminated by the addition of termination solution, followed by heat denaturation of the S1 enzyme. This NGS-based tumour profiling assay examined the expression levels of 2,083 human miRNA transcripts. The assay contained 2,102 probes, including 13 housekeeper genes, five negative process controls, and one positive process controls. The full plate of samples were analysed simultaneously on an Illumina sequencing platform (Illumina, USA). The library was prepared in accordance with OP-00035, HTG EdgeSeq polymerase

chain reaction (PCR) processing. Clean-up procedures were performed according to OP-00037, HTG EdgeSeq AMPure cleanup of Illumina Sequencing Libraries. The library was quantified in accordance with OP-00079, HTG EdgeSeq KAPA Library Quantification for Illumina Sequencing. All samples and controls were quantified in triplicate. The sequencing was performed on the Illumina NextSeq sequencer in accordance with OP-00093, HTG EdgeSeq Illumina NextSeq Sequencing. Post-sequencing quality control (QC) metrics (QC0, QC1, and QC2) detect three different sample failure modes. QC0 detects degraded RNA or poor quality / quantity samples, by assessing the percentage of overall reads being allocated to the positive process control probe for each sample. QC1 detects samples with insufficient read depth; this is evaluated by setting the cut-off at the minimum number of reads that can be allocated to each sample for the data set to be repeatable. QC2 detects samples with minimal expression variability, of which is determined by the relative standard deviation of reads allocated to each probe within a sample. Data were returned from the sequencer in the form of demultiplexed FASTQ files, with four files per original well of the assay. The HTG EdgeSeq Parser was used to align the FASTQ files to the probe list to collate the data. Data were provided as data tables of raw, QC raw, $\log_2$, CPM (counts per million), and median normalized data.

2.2.11 QIAGEN qPCR Microarray

2.2.11.1 RNA isolation and quantification

RNA isolation of samples was performed using the QIAGEN miRNeasy Serum/Plasma Advanced kit (QIAGEN, Manchester, UK) as per the manufacturer's protocol. Briefly, 100 µL of sample was diluted 1:1 with 100 µL of Phosphate Buffer Saline (PBS) (ThermoFisher, London) in 2 mL tubes. The QIAGEN miRCURY LNA RNA Spike-in kit (QIAGEN, UK) was prepared as per the manufacturer's instructions prior to sample lysis. A volume of 1 µL of RNA spike-in working solution (UniSp2, UniSp4 and UniSp5) per sample prep was added to the Buffer RPL. A volume of 61 µL of RNA-spiked Buffer RPL was added to each tube and vortexed for 5 sec. Tubes were then incubated at RT for 3 min. A volume of 20 µL of Buffer RPP was added to each tube and vortexed vigorously for 20 sec. Tubes were incubated again at RT for 3 min before being centrifuged at 12,000 × g for 3 min at RT to pellet the precipitate. The supernatants were transferred

to new 2 mL tubes and 230 µL of isopropanol was added to each. Samples were mixed by vortexing before being moved to RNeasy UCP MinElute Columns and placed on the QIAGEN QIAvac vacuum system 24-manifold (QIAGEN, UK). The vacuum was turned on and samples were pulled through the column, with waste supernatants collecting in the waste bottle. The vacuum pressure was released by opening the release valve, and 700 µL of Buffer RWT was added to each tube. Samples were pulled through the column and the pressure released before adding 500 µL of Buffer RPE to each tube. The release valve was opened again before adding 500 µL of 80% (v/v) ethanol to each tube, and samples were pulled through the column again. RNeasy UCP MinElute tubes were then removed from the vacuum manifold and placed inside 2 mL collection tubes. The lids of the spin columns were opened and the tubes were centrifuged at 15,000 × g for 5 min to dry out the membrane. RNeasy UCP MinElute tubes were then placed inside fresh 1.5 mL tubes. A volume of 20 µL of RNase-free water was added directly to each membrane and allowed to incubate at RT for 1 min. The lids of the spin columns were closed and the tubes were centrifuged at 15,000 × g for 1 min to elute the RNA. Isolated RNA was then quantified using the DENovix DS-11 Spectrophotometer (DeNovix, Wilmington, DE, USA) as per the manufacturer's instructions.

2.2.11.2 cDNA synthesis

cDNA synthesis of isolated RNA was performed using the QIAGEN miRCURY LNA RT Kit (QIAGEN, UK) as per the manufacturer's instructions. Starting volumes were made up to 6 µL with RNase-free water. All reactions were set up on ice in 200 µL PCR tubes. The reverse transcription master mix was prepared as per Table 2.5. A volume of 24 µL of the reverse transcription master mix was added to each tube containing 6 µL of RNA dilution. Tubes were then placed on a Perkin Elmer PE 9700 Thermal Cycler (ThermoFisher, UK) and kept at 42°C for 60 min, 95°C for 5 min and then allowed to cool to 4°C. cDNA samples were then stored at 4°C until use.

Table 2.5 Reverse transcription reaction components.

Component	Volume per sample
5X miRCURY RT Reaction Buffer	6 µL
RNase-free water	13.5 µL
10X miRCURY RT Enzyme Mix	3 µL
Synthetic RNA spike-ins	1.5 µL
Total volume	**24 µL**

2.2.11.3 Custom quantitative RT-PCR microarray setup and analysis

A QIAGEN miRCURY LNA miRNA Custom 96-well PCR Panel (QIAGEN, USA) was created with the help of QIAGEN technicians. The miRCURY LNA SYBR GREEN PCR Kit (QIAGEN, USA) was used alongside the custom 96-well PCR panel as per the manufacturer's instructions. Low concentration ROX dye (10 µL ROX dye per 1 mL of SYBR Green) was used to comply with the needs of the instrument to be used for the RT-PCR reaction. A 3 µL volume of each template cDNA sample was diluted with 117 µL RNase-free water (1:40 dilution). The reaction master mix was prepared as per Table 2.6.

Table 2.6 Reaction master mix setup for LNA miRNA Custom PCR panel.

Component	Volume per reaction
2X miRCURY SYBR Green Master Mix	5 µL
RNase-free water	1 µL
cDNA template (1:40 dilution)	4 µL
Total volume	**10 µL**

The prepared reaction mixes were vortexed thoroughly before dispensing 10 µL per sample into the 24 designated wells. The plate was sealed and vortexed briefly to mix before being centrifuged at 3,000 × g at RT to bring all the liquid to the bottom of the wells. The plate was incubated for 5 min at RT to allow the primers to dissolve in the master mix. Plates were run on the Applied Biosystems 7500 Real-Time PCR System (ThermoFisher, UK) as per the cycling conditions in Table 2.7. Initial processing of the data was conducted using the ThermoFisher Connect™ Cloud PCR Analysis software using the Relative Quantification app (v 2021.1.1-Q1-21-build11).

Table 2.7 PCR cycling conditions for miRCURY LNA miRNA custom PCR panel.

Step	Time	Temperature	Ramp Rate
Initial heat activation	2 min	95°C	Maximal/fast mode
2-step cycling *Denaturation* *Combined annealing/extension*	 10 sec 60 sec	 95°C 56°C	 Maximal/fast mode Maximal/fast mode
Number of cycles	40 cycles		
Melting curve analysis		60-95°C	

2.2.12 Culture of cell lines

2.2.12.1 H6c7-normal cell line maintenance

A normal human pancreatic duct epithelial cell line, H6c7 cells (H6c7-normal), was obtained from Kerafast Inc. (Boston, MA, USA). These cells were maintained in Keratinocyte Serum Free Medium with L-glutamine, 5 ng/mL human recombinant EGF and 50 mg/mL bovine pituitary extract (kit) (Biosciences, Dublin, Ireland), supplemented with 50 U/mL penicillin and 50 µg/mL streptomycin (Lonza Group, Basel, Switzerland) (hereafter known as complete medium). H6c7 cells were maintained as monolayers in either 25 cm^2 or 75 cm^2 sterile vented culture flasks, in an incubator at 37°C in 95% humidified air containing 5% CO_2.

2.2.12.2 HPNE-intermediary cell line maintenance

An intermediary human pancreatic cell line formed during acinar-to-ductal metaplasia, hTERT-HPNE cells (HPNE-intermediary), were obtained from the American Type Culture Collection (ATCC) (Manassas, VA, USA). The base medium for these cells as per the ATCC website contains 75% Dulbecco's Modified Eagle's Medium (DMEM) without glucose with additional 2 mM L-glutamine (ThermoFisher, UK), 1.5 g/L sodium bicarbonate, and 25% Medium M3 Base (Incell Corp., San Antonio, TX, USA). The growth medium is created by adding Foetal Bovine Serum (FBS) (final concentration 5% (v/v)) (Lonza Group, Switzerland), 10 ng/mL human recombinant EGF (Biosciences, Ireland) and 5.5

mM D-glucose (Biosciences, Ireland) to the base medium. The complete growth medium was supplemented with 50 U/mL penicillin and 50 µg/mL streptomycin (Lonza Group, Switzerland). Once sufficient stocks of these cells were frozen and stored at -80°C, the growth of this cell line was optimised in an alternative growth medium: DMEM - high glucose supplemented with 10% (v/v) FBS, 10 ng/mL human recombinant EGF (Biosciences, Ireland) and 50 U/mL penicillin and 50 µg/mL streptomycin (Lonza Group, Switzerland) (hereafter known as complete medium). All experiments were carried out using this alternative growth medium. HPNE-intermediary cells were maintained as monolayers in either 25 cm^2 or 75 cm^2 sterile vented culture flasks, in an incubator at 37°C in 95% humidified air containing 5% CO_2.

2.2.12.3 Sub-culture

H6c7-normal cells were passaged every 3-4 days and HPNE-intermediary cells were passaged twice weekly upon reaching 70-80% confluency as determined by light microscopy. Complete growth medium, PBS and Trypsin protease Ethylene Diamine Tetra Acetic Acid (Trypsin-EDTA) (Lonza Group, Switzerland) were warmed to 37°C in a water bath prior to subculture. In the case of HPNE cells, Ca++/Mg++ free Dulbecco's PBS (ThermoFisher, London) was used in place of PBS to aid in the process of dissociation by Trypsin-EDTA. A sterile grade II laminar flow hood was thoroughly cleaned using 70% (v/v) ethanol, as well as all reagents and equipment before being placed into the hood. Spent medium was discarded from the flask to waste and the cells were rinsed in 5 mL of PBS (H6c7-normal) or Ca++/Mg++ free Dulbecco's PBS (HPNE-intermediary). Trypsin-EDTA was added to the flask to detach adherent cells (1 mL in 25 cm^2 flasks and 3 mL in 75 cm^2 flasks) and allowed to incubate for 5 min at 37°C and 5% CO_2. Light microscopy was used to ensure all cells had successfully detached from the bottom of the flask before an equal volume of complete growth medium was added to neutralise the Trypsin-EDTA. Cells were then transferred to a sterile 15 mL tube and centrifuged at 1200 RPM for 5 min at RT. The supernatant was discarded to waste and the cell pellet resuspended in complete growth medium. H6c7-normal cells were passaged at a 1:4 ratio. HPNE-intermediary cells were passaged between a 1:8 and 1:12 ratio. Cell waste was decontaminated by treating for 24 h with Haz-Tab chlorine tablets (Guest Medical,

Aylesford, UK). Passaged cells were then stored in an incubator at 37°C in 95% humidified air containing 5% CO_2.

2.2.12.4 Cell counting

Cell counting was conducted during the sub-culture process described in section 2.2.12.3, after the centrifugation step using a Bright Line Haemocytometer (Hausser Scientific, Horsham, PA, USA). The cell pellet was resuspended in 2 mL (H6c7-normal cells) or 5 mL (HPNE-intermediary cells) of complete growth medium and 20 µL of cell suspension was added to a 0.5 mL Eppendorf tube along with 180 µL of trypan blue (Biosciences, Ireland) and mixed by pipetting. A 20 µL volume of this solution was then added to the haemocytometer and covered with a glass coverslip. Live cells are impermeable and do not take up the trypan blue dye, while dead cells are stained blue. Live cells were counted in each of the four corners (Q1 + Q2 + Q3 + Q4) of the haemocytometer and the number of cells per mL of media was calculated as follows:

$$\frac{\mathrm{Q1+Q2+Q3+Q4}}{4} \times 10^4 (haemocytometer\ volume) \times 10\ (dilution\ factor) = cells/mL$$

2.2.12.5 Cryopreservation and reconstitution of cell line stocks

Frozen stocks of each cell line were generated before experimental use. During the sub-culture process described in section 2.2.12.3, and after the centrifugation step, the cell pellet obtained from one 75 cm^2 culture flask was resuspended in 1 mL of cryopreservation medium consisting of FBS with 10% (v/v) Dimethylsulfoxide. This volume was transferred to a sterile, labelled cryovial and placed in a Mr. Frosty™ Freezing Container (ThermoFisher, UK) containing isopropyl alcohol before being transferred to a -80°C freezer overnight. Cells were then removed from the Mr. Frosty™ Freezing Container the following day and kept at -80°C for short-term storage, or transferred to a liquid nitrogen container for long-term storage.

To reconstitute frozen cell stocks, cryovials were removed from storage and thawed rapidly at 37°C in a water bath. Once only a fragment of ice remained, the cells

were brought into a sterile fume hood where they were transferred to a 15 mL tube and 10 mL of pre-warmed complete growth medium was added dropwise to avoid osmotic shock. Cells were then centrifuged at 1200 RPM for 5 min and the supernatant discarded to waste. The cell pellet was resuspended in 7 mL of complete growth and transferred to a 25 cm^2 culture flask before being stored in an incubator at 37°C in 95% humidified air containing 5% CO_2.

2.2.12.6 Mycoplasma testing

2.2.12.6.1 PCR of mycoplasma unknown samples

Mycoplasma testing of cell lines was carried out every six weeks. A volume of 1 mL of supernatant was collected from a confluent 75 cm^2 culture flask and stored in the fridge at 4°C until use. Supernatants were centrifuged at 2000 RPM for 2 min to pellet any cell debris. A 25 µL PCR reaction for each sample to be tested was setup in a PCR tube containing 12.5 µL of Green GoTaq (Promega, USA), 0.5 µL of forward primer (10 uM) (ThermoFisher, UK), 0.5 µL of reverse primer (10 uM) (ThermoFisher, UK), 10.5 µL of molecular biology grade sterile water (ThermoFisher, UK), and 1 µL of cell culture supernatant. Positive and negative controls were created using a mycoplasma positive supernatant and water, respectively. PCR tubes were then placed on a thermal cycler and heated at 95°C for 5 min, before being subjected to 40 cycles of 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min. After the 40 cycles, PCR tubes were kept at 72°C for 10 min and finally allowed to cool to 4°C.

2.2.12.6.2 Agarose gel electrophoresis

While the PCR tubes were on the thermal cycler, the 2% agarose gel was prepared and allowed to set. Briefly, 10X Tris-Borate-EDTA (TBE) Buffer was made by dissolving 165 g tris, 27.5 g boric acid, and 9.3 g EDTA in 800 mL of dH_2O. The pH was adjusted to 8.3 using acetic acid and the final volume was made up to 1 L with dH_2O. 1X TBE was then prepared by diluting the 10X TBE stock 1 in 10 with dH_2O. For the gel, 4 g of agarose was added to a glass bottle containing 200 mL of 1X TBE. The agarose was heated gently in a microwave, stirring frequently, until the agarose was melted and the solution was a clear

liquid. The solution was allowed to cool for 5 min before being transferred to 4 × 50 mL tubes. A volume of 3 µL of SybrSafe (ThermoFisher, UK) was added to each 50 mL tube and inverted to mix. The gel solution was then poured into the gel electrophoresis gel rig and a 10-well comb inserted. Any bubbles present were removed with a pipette tip before allowing the gel to set for 1 h. Once set, the comb was removed and the gel was submerged in 1X TBE buffer, with the wells being positioned at the anode.

After being cooled to 4°C on the thermal cycler, samples were transferred on ice to the gel rig. A 100 bp ladder was loaded into each side of the gel at a volume of 3 µL per well. Samples and controls were then loaded into the get at a volume of 10 µL per well. Electrophoresis was then carried out using at a voltage of 80 V for 1 h.

2.2.12.6.3 ChemiDoc visualisation

After electrophoresis, the PCR products were then visualised using a BioRad ChemiDoc System (Bio-Rad Laboratories, Hercules, CA, USA) under the SYBR GREEN setting, with a band for positive mycoplasma infection being expected at 270 bp.

2.2.13 Serum starvation of cell lines

All experiments where pancreatic cell lines H6c7-normal and HPNE-intermediary were treated with PCF were run in serum-protein-free (serum-starved) conditions. To do this, cells were seeded at the required densities in serum-free medium. For H6c7-normal cells, this refers to their complete growth medium, as outlined in section 2.2.12.1, in the absence of the bovine pituitary extract. For HPNE-intermediary cells, the refers to their complete growth medium, as outlined in section 2.2.12.2, in the absence of FBS. An untreated complete medium control, as well as an untreated serum-starved control were also run as part of these experiments. Cells were seeded the night before treatment in their respective serum-free or complete medium and allowed to adhere overnight.

2.2.14 Treatment of normal pancreatic cell lines with PCF

Pancreatic cell lines H6c7-normal and HPNE-intermediary were serum-starved as described in section 2.2.13, seeded at the required density for the plate size of use in their respective serum-free medium, and left in an incubator at 37°C overnight to adhere. The next morning, treatments of 10% (v/v) PCF were prepared in half of the well volume. Triplicates were prepared together and pipetted across the 3 replicates to reduce pipetting error. PCF samples were sonicated prior to use, as described in section 2.2.6. Half of the well volume was carefully removed from each well and discarded to waste before being replaced by the same volume of either fresh medium or fresh medium containing 10% (v/v) PCF treatment to give a final concentration of 5% (v/v) PCF in culture medium. Once treated, cells were left for the required treatment time in an incubator at 37°C and 5% CO_2.

2.2.15 Proliferation assay

Pancreatic cell lines H6c7-normal and HPNE-intermediary were seeded in triplicate at 3 x 10^3 cells per well in a flat-bottomed 96-well plate in 100 µL of their respective serum-free medium, as described in section 2.2.13, and left in an incubator at 37°C overnight to adhere. The next morning, cells were treated with PCF as described in section 2.2.14. Blank wells were created by adding medium only to 3 wells for each of the four media. The plate was then left for 22 h in an incubator at 37°C. The following day, CCK-8 reagent (Sigma Aldrich, Ireland) was thawed at RT for 30 min in the dark. At 22 h post treatment, 5 µL of CCK-8 reagent was added to each well, including blanks. The plate was then incubated in the dark for 2 h in an incubator at 37°C. After 2 h (24 h post treatment), the plate was read on the GloMax Explorer microplate reader (Promega, USA) at 450 nm.

2.2.16 ApoTox Glo assay

Pancreatic cell lines H6c7-normal and HPNE-intermediary were seeded and treated PCF as outlined in section 2.2.15. Once treated, cells were left for 24 h in an incubator at 37°C. After 24 h, cells were processed using the ApoTox-Glo™ Triplex Assay (Promega, USA), as per the manufacturer's instructions, to assess viability, cytotoxicity and

apoptosis. Briefly, a volume of 20 µL of Viability/Cytotoxicity Reagent was added to all wells and mixed by orbital shaking at 500 RPM for 30 sec. The plate was then incubated in the dark for 30 min at 37°C. After 30 min, fluorescence was measured on the GloMax Explorer microplate reader (Promega, USA) at both $400_{EX}/505_{EM}$ (viability) and $485_{EX}/520_{EM}$ (cytotoxicity). A volume of 100 µL of Caspase-Glo 3/7 Reagent was added to all wells and mixed by orbital shaking at 500 RPM for 30 sec. The plate was incubated in the dark for 30 min at RT. Luminescence was then measured on the GloMax Explorer microplate reader (Promega, USA) (apoptosis / caspase activation).

2.2.17 Flow cytometry

2.2.17.1 Antibody titrations

Pancreatic cell lines H6c7-normal and HPNE-intermediary were sub-cultured as outlined in section 2.2.12.3. The cell pellet obtained from each cell line was divided across 5 × FACS tubes, resuspended in 2 mL of PBS each and centrifuged at 1300 RPM for 3 min (washed). The supernatant was discarded to waste and the cells were washed again. One tube per cell line was set aside (unstained), and a second tube per cell line was left on a heating block at 95°C for 15 min to create a dead cell control. The dead cell control and remaining 3 tubes per cell line were resuspended in 100 µL of working Zombie NIR Viability dye (BioLegend, USA) and incubated for 15 min at RT in the dark. Cells were washed in FACS buffer (2% (v/v) FBS and 0.01% (w/v) sodium azide in PBS) before resuspending all 5 tubes per cell line in 1 mL of blocking buffer (50% (v/v) FACS buffer in FBS) for 5 min at RT in the dark. Cells were then washed in FACS buffer before being resuspended in 100 µL of FACS buffer. The unstained control and dead cell control for each cell line were set aside. The remaining three tubes were stained with 1 µL, 3 µL or 5 µL extracellular antibodies for 20 min at RT in the dark. Cells were then washed with FACS buffer before being resuspended in 100 µL of FACS buffer. Fixation and permeabilization steps were carried out using the eBioscence™ Foxp3 / Transcription Factor Staining Buffer Set (ThermoFisher Scientific, USA) as per the manufacturer's instructions. Briefly, 1 mL of Fixation/Permeabilisation working solution was added to each tube and vortexed to mix. Tubes were incubated for 60 min at RT in the dark. After 60 min, 2 mL of 1X Permeabilisation buffer was added to each tube. Tubes were then

centrifuged at 600 × g for 5 min at RT. The supernatant was discarded and samples were resuspended in the residual volume of 1X Permeabilisation buffer. Again, the unstained and dead cell controls for each cell line were set aside. The remaining tubes were stained with 1 µL, 3 µL or 5 µL, as before, with intracellular antibodies for 30 min at RT in the dark. After this time, 2 mL of 1X Permeabilisation buffer was added to all tubes and they were centrifuged at 500 × g for 5 min at RT. Cells were washed twice with FACS buffer before being resuspended in 200 µL of FACS buffer. Cells were acquired using the BD FACs CANTO II Flow Cytometer and BD FACs Diva Software.

2.2.17.2 Flow cytometric staining

Pancreatic cell lines H6c7-normal and HPNE-intermediary were seeded in serum-starved conditions as described in section 2.2.13. Cells were seeding in 24-well plates at 5 x 10^4 cells per well and left overnight to adhere. Cells were treated with 5% PCF as outlined in section 2.2.14. Once treated, cells were left for either 6 h or 24 h in an incubator at 37°C. After either 6 or 24 h, cells were washed and stained as per the protocol described in section 2.2.17.1. Cells were stained with 1 µL of each intracellular and extracellular antibody as determined by the antibody titration. Cells were acquired using the BD FACs CANTO II Flow Cytometer and BD FACs Diva Software.

2.2.17.3 Processing and analysis of flow cytometric data

Data were analysed using FloJo software v10 (Becton Dickinson, USA). Gating strategies utilised for both HPNE-intermediary and H6c7-normal cell lines are outlined in Figure 2.1. The cell populations were first gated on using the side scatter area (SSC-A) and forward scatter area (FSC-A), before gating on single cells only, again using the SSC-A and FSC-A. Lastly, the live and dead cell populations were determined used the forward scatter height (FSC-H) and Zombie NIR viability dye, with the dead cell control being utilised to determine the positive and negative Zombie NIR populations. Compensation beads (Becton Dickinson, USA) were stained with antibodies of interest and used as per the manufacturer's instruction to compensate for spectral overlap of these antibody channels.

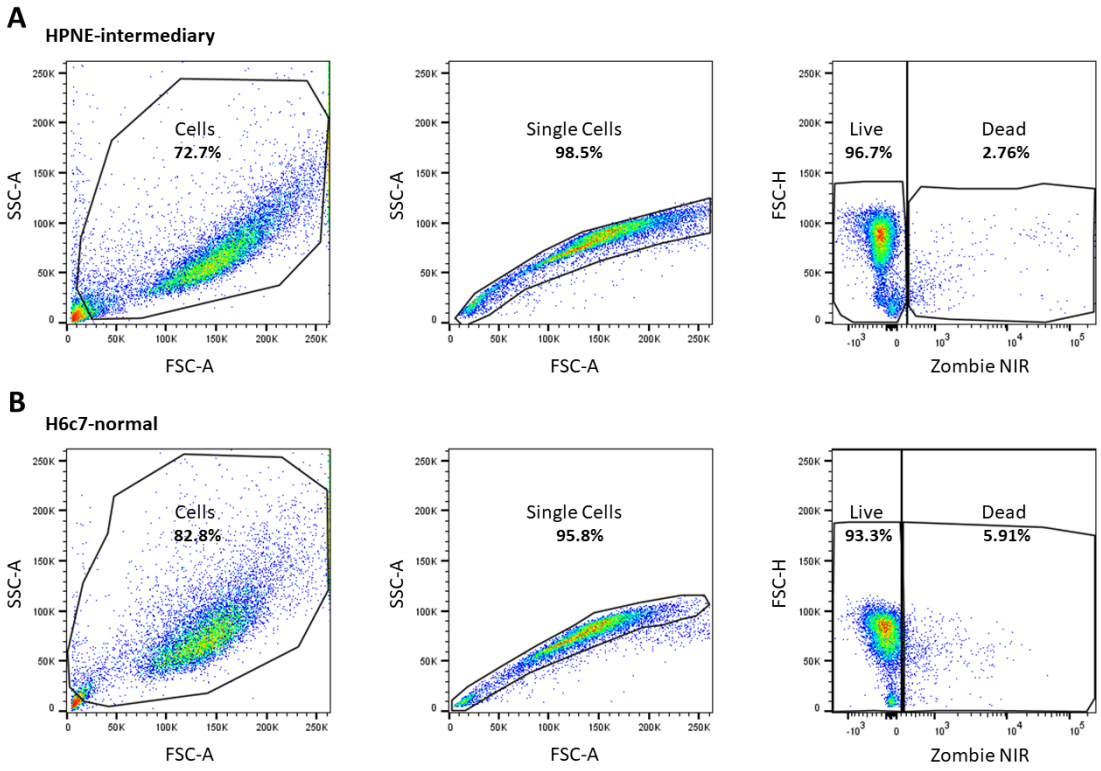


Figure 2.1 Representative dot plots showing gating strategies for cell lines. Representative dot plots from FlowJo software showing gating strategies employed for (**A**) HPNE-intermediary cells and (**B**) H6c7-normal cells. Both whole cell (left) and single cell (centre) populations were gated on using the SSC-A and FSC-A. Live and dead cell populations (right) were gated on used the FSC-H, and Zombie NIR thresholds determined via dead cell controls.

2.2.18 Seahorse XF ATP rate test

To prepare the 24-well Seahorse cell culture plate (Agilent Technologies, USA), 50 µL of 50% (v/v) Poly-D-Lysine (Sigma Aldrich, UK) in PBS was added to each well of the plate and incubated at RT for 5 min. After 5 min, the solution was discarded to waste and the wells were washed twice with 100 µL of PBS. The plate was then left to dry at RT inside a fume hood for 1 h. Once dry, cells were seeded at a density of 2×10^4 in either 100 µL of complete culture medium, or serum-protein free medium as described in section 2.2.12. The four required blank wells [A1, B4, C3 and D6] were given media only. The plate was centrifuged at 1200 RPM for 3 min to stick down the cells, and was then left overnight in an incubator at 37°C. The following morning, 50 µL of culture medium was discarded to waste from each well and replaced with 50 µL of either 5% PCF or media only, as described in section 2.2.13. PCF samples were sonicated prior to use, as described in section 2.2.5. The plate was then left for 24 h in an incubator at 37°C. At this point, 1 mL of Seahorse XF Calibrant (Agilent Technologies, USA) was added to each well of the Seahorse XF Calibration Plate beneath the Seahorse XF Sensor Cartridge (Agilent Technologies, USA). The sensor cartridge was placed back on top of the calibration plate and the completed unit was stored in a non-CO_2 incubator overnight. The following day, the assay medium (complete Seahorse medium) was prepared by supplementing 100 mL of Seahorse XF DMEM Medium, pH 7.4, with 10 mM of XF Glucose, 1 mM of XF pyruvate, and 2 mM of XF glutamine (Agilent Technologies, USA). Assay compounds were diluted in complete Seahorse medium from stock solutions to give 1.8 µM Oligomycin and 2 µM Antimycin A (Agilent Technologies, USA). Once all materials were prepared, the cell plate was removed from the incubator and the entire volume of medium from each well was transferred to a 96-well plate and stored at -20°C for later use. A volume of 500 µL of complete Seahorse medium was added to each well, before removing 400 µL of this and discarding to waste. Finally, 450 µL of complete Seahorse medium was added to each well and the cell plate was placed in the non-CO_2 incubator for at least 1 h before use. At the same time, the senor cartridge was removed from the non-CO_2 and 50 µL each of the Oligomycin and Antimycin A compounds were loaded into all ports A and B, respectively. The sensor cartridge was then placed back into the non-CO_2 incubator for 10 min. The ATP Rate Test template on the Seahorse Analyser Wave software (v2.6.3.5) was setup to include all details of the experimental

design. The sensor cartridge was placed inside the Seahorse XF instrument, without the lid, and allowed to calibrate for 15 min. The sensor cartridge was then removed from the machine, and the calibrant plate below was replaced with the cell plate, before being reinserted into the machine and the assay run for 90 min.

Once completed, the sensor cartridge was removed from the machine and the data exported. The upper, senor cartridge was discarded to waste and the cell plate was stained by crystal violet in order to normalise the data obtained to cell number. Briefly, the supernatant was discarded to waste and 50 µL of 1% (v/v) glutaraldehyde was added to each well and allowed to incubate at RT for 15 min inside a fume hood. Wells were then washed twice with 200 µL of PBS before adding 50 µL of 0.05% crystal violet (w/v) and incubating for 30 min at RT. Waste crystal violet was carefully removed, and the wells were gently washed with 200 µL of dH_2O before being left overnight to dry. Once dry, 60 µL of Triton-X was added to each well and the plate was placed on a plate shaker at 400 RPM for 1 h. The plate was then removed from the orbital shaker and 50 µL from each well was transferred by reverse-pipetting to a 96-well plate. Absorbance was then measured on the GloMax Explorer microplate reader (Promega, USA) at 595 nm.

2.2.19 Invasion assay

The invasive capacity of H6c7 and HPNE pancreatic cell lines after treatment with 5% PCF was assessed using the 96-well Collagen I Cell Invasion Assay (8 µm) (Abcam, UK) as per the manufacturer's instructions. Cell lines were serum-starved for 24 h prior to the start of the assay, using protein-serum free culture mediums described in section 2.2.12. In the top chamber of the provided assay plate, 40 µL of Collagen I was added to each well and allowed to set overnight at 4°C to create a thin film of Collagen I. Once set, the wells were washed three times with 100 µL of 1M Tris. A volume of 200 µL of FBS was added to the bottom chamber of the assay plate, with the negative control receiving PBS only, and the positive control receiving a 1:10 dilution of the provided Control Invasion Inducer in culture medium. Cells were seeded in 50 µL of serum-free medium at a density of 5×10^4 cells per well in the top chamber of the assay plate. Cells were then treated with 50 µL of either 5% PCF of media only, as described in section 2.2.14. PCF

samples were sonicated prior to use, as described in section 2.2.6. The assay plate was then placed in an incubator at 37°C for 24 h.

A standard curve was created for each cell line by serially diluting the cells in wash buffer. Standard curves were plated in 100 µL of wash buffer in triplicate on a 96-well plate (50,000; 25,000; 12,500; 6,250; 3,125; 1,562; 781; 390 cells per 100 µL). As a blank, wash buffer only was used. The cell dye was diluted 1:250 in PBS, and 50 µL of diluted cell dye was added to each well. The plate was incubated at 37°C for 1 h before reading the fluorescence on the GloMax Explorer microplate reader at $485_{Ex}/530_{Em}$. The standard curve for each cell line was created using the RFUs obtained and the data were fit with a linear trendline with zero intercept.

After 24 h, the assay plate was removed from the incubator and any remaining medium in the top chamber was carefully aspirated to waste. The top chamber was then removed, and the bottom chamber was centrifuged at 1,000 × g for 5 min at RT. The supernatant was carefully discarded to waste before washing the wells of the bottom chamber with 200 µL of wash buffer. The plate was centrifuged again at 1,000 × g for 5 min at RT and the supernatant discarded to waste. The cell dye was diluted 1:500 in cell dissociation solution and mixed well before adding 100 µL of this solution to each well. The assay plate was reassembled, with the top chamber placed on top of the bottom chamber, and incubated at 37°C for 30 min. After 30 min, the plate was gently tapped on the side to ensure dissociation of any invasive cells that were stuck to the outer side of the top chamber. The top chamber was then removed and the fluorescence was read on the GloMax Explorer microplate reader at $485_{Ex}/530_{Em}$. The number of invaded cells was calculated using the standard curves created for each cell line. Percentage invasion was calculated as follows:

$$\%\ Invasion = \frac{\text{Number of cells in the lower chamber}}{\text{Number of cells added to the top chamber}} \times 100$$

2.2.20 DNA damage competitive ELISA

The 8-Hydroxydeoxyguanosine (8-OHdG) levels in PCF, and in the supernatants from H6c7-normal and HPNE-intermediary cell lines that had been treated with PCF, were

assessed to quantify levels of DNA damage. The 8-OHdG Competitive ELISA kit (Assay Genie, Ireland) was used as per the manufacturer's instructions. Neat supernatants obtained from co-culturing H6c7-normal and HPNE-intermediary cells with 5% PCF, as described in section 2.2.13, were run alongside the matching PCF samples, diluted to 5% with relevant serum-free culture medium. PCF samples were sonicated prior to use, as described in section 2.2.6. Briefly, the standard was serially diluted from the provided stock solution (100 ng/mL) at the recommended concentrations to create a standard curve, with the final standard containing no protein (blank). A volume of 50 µL of each standard and sample was plated in the provided, pre-coated 96-well microplate in duplicate and incubated at 37°C for 45 min. The samples and standards were then carefully removed from the plate and 350 µL of Wash Buffer was added to each well (wash). This was repeated three times before adding 100 µL of SABC working solution to each well and incubating at 37°C for 30 min. The plate was then washed five times and 90 µL of TMB substrate was added to each well. The plate was Incubated for 10-20 min in the dark at 37°C. The reaction was stopped when the colour development was sufficient by adding 50 µL of Stop Solution to each well. The absorbance at 450 nm was measured using the GloMax Explorer microplate reader.

2.2.21 Data handling and statistical analysis

2.2.21.1 Processing and analysis of omics data

Both PCF and serum proteomic data obtained from LC-MS were cleaned and normalised using Perseus (v1.6.15.0)[292]. Briefly, label-free quantification (LFQ) intensity data and Fasta headers were brought into the software. The data were filtered to remove reverse sequences, potential contaminants and proteins that are only identified by peptides carrying one or more modified amino acids ("only identified by site"). The data were then annotated with the appropriate risk category (high-risk or low-risk) and filtered based on valid values. In this case, remaining proteins needed to be present in at least 70% of the total data; proteins only present in less than 30% of the data were filtered out. The data were then $\log_2$ transformed and normalised using a linear transformation. Finally, imputation was used to replace missing values from the normal distribution.

differential expression analysis with Benjamin-Hochberg corrections was conducted in Perseus using the built-in 'edgeR' package from R studio.

PCF and serum transcriptomic data were obtained from HTG Molecular in CPM normalised form. Normalised transcriptomic data were loaded into R studio (v21.09.0) and differential expression analysis was conducted using packages 'readxl' (v1.4.1), 'edgeR' (v3.32.1) and 'DESeq' (v1.30.1). Multiple comparisons for differential expression analysis was corrected using Benjamin-Hochberg corrections.

Box plots and volcano plots of significantly differentially expressed factors were created in GraphPad Prism (v9.5.0). Normalised proteomic and transcriptomic data for both PCF and serum were scaled individually before being integrated to create a single data matrix. Unsupervised hierarchical clustering with supporting heatmap and dendrograms were generated in R Studio using packages 'edgeR' (v3.32.1), 'cluster' (v2.1.4), 'purrr' (v0.3.4), 'dendextend' (v1.15.2), 'dplyr' (v1.0.9), 'ggplot2' (v3.3.5), 'ComplexHeatmap' (v2.6.2), 'RColorBrewer' (v1.1-3), 'gplots' (v3.1.1), 'pheatmap' (v1.0.12) and 'factoextra' (v1.0.7). Corrplots illustrating the correlations between patient clinical data and omics factors were created in R Studio using packages 'Hmisc' (v4.7-2), 'heatmap3' (v1.1.9), 'pheatmap' (v1.0.12), 'plot.matrix' (v1.6.2), 'RColorBrewer' (v1.1-3), 'gplots' (v3.1.1), 'corrplot' (v0.90) and 'ggcorrplot' (v0.1.3). Clinical data were converted to binary code where appropriate using the key in Appendix 2. Spearman correlations were run and p-values were corrected using Holm-Bonferroni corrections. 2-D Principle Component Analysis (PCA) was conducted in R studio using packages 'tidyverse' (v1.3.1), 'ggplot2' (v3.3.5), and 'factoextra' (v1.0.7). 3-D PCA analysis was conducted in R Studio using the same packages as the 2-D PCA analysis, but with the addition of 'rgl' (v0.108.3). Leave-one-out cross validation (LOOCV) and corresponding ROC plots were created in R Studio using packages 'tidyverse' (v1.3.1), 'dplyr' (v1.0.9), 'plyr' (v1.8.7), 'klaR' (v1.7-1) and 'caret' (v6.0-93). Assessment of optimal biomarker combinations was conducted using CombiROC software (v1.2)[293]. Data were scaled in R studio and processed using a linear transformation to ensure no negative values were present before being brought into the CombiROC software. Using the graphics function, the minimum number of biomarker features was set to 1 in order to evaluate the number of biomarkers that produced the best results. The test signal cut-

off was calculated as the mean of the control group plus the standard deviation, rounded to the nearest whole number for to be compatible with the software. For PCF the cut-off was set to 3, for serum this was set to 4. PCF sensitivity and specificity limitations were set at 83% and 25%, respectively. Serum sensitivity and specificity limitations were set at 93% and 47%, respectively.

2.2.21.2 Processing and analysis of cell line experimental data

All experimental data were processed using Microsoft Excel (v2202) and GraphPad Prism (v9.5.0). Data shown are expressed as the mean ± the standard error of the mean (SEM). Experimental data evaluating the effects of patient PCF on cell lines were assessed using non-parametric tests, such as a Mann-Whitney test or Kruskal-Wallis test, as appropriate based on the number of groups to be assessed. In cases where the data were paired across treatments, a Wilcoxon test or Friedmans test were employed. Where the analysis involved 3 or more groups, Dunn's multiple comparisons were used to determine the statistical significance between the groups. Spearman correlations and simple linear regressions were used to assess the relationships between non-parametric variables. A probability of less than 0.05 (p<0.05) was considered statistically significant.

Chapter 3.

Diagnostic accuracy of blood-based biomarkers for pancreatic cancer: A systematic review and meta-analysis

Published in full in:

Kane, L.E., Mellotte, G.S., Mylod, E., O'Brien, R.M., O'Connell, F., Buckley, C.E., Arlow, J., Nguyen, K., Mockler, D., Meade, A.D., Ryan, B.M. and Maher, S.G. (2022) Diagnostic accuracy of blood-based biomarkers for pancreatic cancer: A systematic review and meta-analysis. *Cancer Research Communications*, 2(10), pp.1229-1243.

3.1 Introduction

The search for robust biological markers for disease diagnosis and treatment has been a consistent objective within modern health care over the last few decades[294, 295]. In many instances, the use of multiple biological markers is almost intuitive when proceeding with a patient diagnosis. Symptoms, for example, are biological markers of illness used to identify the ailment at hand. As with symptoms, patients will rarely experience just one, and health care professionals will generally use the presence or absence of several symptoms to make a diagnosis[296]. In this way, clinicians have used multiple biomarkers (multi-biomarker panels) for the diagnosis of disease intuitively for thousands of years. Modern medicine often overlooks the utility to be gained from the use of multi-biomarker panels, searching for single proteins or miRNAs that are dysregulated in disease and can be used for a more streamlined diagnosis. However, recent trends in the literature favour multi-biomarker panels over single biomarkers, as single biomarkers alone fail to encompass the biological complexity of disease and therefore lack the comprehensive robustness required for a confident diagnosis[166, 295, 297]. This can be seen clearly in the variable diagnostic performance of certain single markers in patients with underlying conditions, and as such has caused a trend toward the use of "mixed" control cohorts, where both healthy volunteers and patients with benign conditions comprise the control cohort. In this way, the control cohort is arguably more clinically relevant, as it better represents the patient population in question.

PC has one of poorest prognoses of any cancer, with a 5-year survival rate of below 5%[298]. Late-stage diagnoses contribute hugely to the poor survival rates of this cancer, as the symptoms associated with PC can be vague in nature[299]. As such, earlier diagnosis is the key to improving patient prognosis in this cancer. PDAC is the most common subtype of PC and represents approximately 85% of all patients with PC[300]. Currently, the only FDA-approved biomarker for PDAC diagnosis is the blood-based biomarker CA19-9[166]. CA19-9 is a type of antigen released by PC cells, and is therefore detected at higher levels in the blood of patients with PDAC[301]. However, reported sensitivity and specificity values for CA19-9 vary greatly from study to study. While sensitivity values are generally high, being reported at roughly 80%, the specificity of this biomarker has been shown to be variable, resulting in many false positives[167].

Serum CA19-9 has been shown to be elevated in some benign conditions, such as pancreatitis, and also in other gastrointestinal malignancies contributing to the limitations of the biomarker[302]. Furthermore, the ability to express CA19-9 at all is dependent on a patient's Lewis blood group (Le), with 5%–7% of the population belonging to the Le(a−b−) group and consequently unable to express CA19-9 at any level[303]. Therefore, even though CA19-9 is widely used in current clinical practice, the results alone cannot be used to diagnose PDAC, and must always be interpreted within the clinical context of imaging and/or histopathology[304]. As such, current methods of diagnosing patients with PC at an early stage rely heavily on the presentation of symptoms, as there is no effective method for screening patients for the presence or absence of PC. A recent review from our group has highlighted the importance of novel biomarker research in pancreatic patients, and argues that the integration of multiple biomarkers to form a multi-biomarker panel could be the vital next step for the identification of more robust biomarkers[166]. Following on from this, a systematic review and meta-analysis of biomarker efficacy was conducted to truly evaluate whether more really is better in the context of biomarkers.

In this systematic review, blood-based biomarkers for the diagnosis of PDAC are evaluated. The primary aim of this review is to compare the efficacy of single biomarkers and multi-biomarker panels in the context of PDAC diagnosis to determine which biomarker type performs better. The secondary objective is to examine the current clinical standard, CA19-9, and its performance comparatively to novel biomarkers for PDAC diagnosis. The final objective of this review is to highlight promising novel biomarkers that have been examined repeatedly in the literature and may provide direction for future biomarker studies.

3.2 Hypothesis

Combining and comparing the results of previously conducted research on the identification of novel diagnostic biomarkers for PC can elucidate the most promising biomarkers that should be evaluated further in future research.

3.3 Specific aims

1. Identify promising biomarkers that have been evaluated multiple times across the literature and show potential diagnostic biomarkers either alone or as part of a multi-biomarker panel.
2. Perform a meta-analysis to determine whether single biomarkers or panels of multiple biomarkers produce the best diagnostic efficacy.
3. Evaluate the role of CA19-9 both alone and as part of a multi-biomarker panel to ascertain in what setting is has the most utility.
4. Report on the standard of experimental design seen across PC biomarker studies and highlight where these designs can be improved to guide future research.

3.4 Results

3.4.1 Identification of relevant studies

After removing duplicates, 5,885 studies were identified by our literature search as potential candidates for inclusion in this review. After two stages of screening by reviewers, 250 papers were included in this review (Figure 3.1). Most records excluded at the full-text stage were omitted due to having cohorts of less than 15 patients, no accessible full-text, or being a conference proceeding and therefore not a peer-reviewed full-text paper.

3.4.2 Accuracy and RoB assessment

Summarized results for RoB and quality assessment as conducted through the use of the QUADAS-2 tool are shown in Figure 3.2. Concerns regarding index test applicability were generally low once a blood-based biomarker was being assessed for PDAC diagnosis as per the inclusion criteria of the review. Patient selection was frequently high risk (68%) as control cohorts were often not clinically relevant, that is, contained healthy or benign patients only. RoB was low for reference standards (38.8%) and index tests (23.2%) when studies were blinded to the results, and unclear (31.6% and 20%, respectively) when no details were given. Concerns about the applicability of the index test were low in most cases (98.8%) as most biomarkers were for the diagnosis of PDAC. The accuracy of data extraction and RoB assessments were independently assessed by senior reviewers before proceeding to the data analysis stage. Data extraction was shown to have a mean accuracy of 91.47% (95% CI: 91.42–91.52), and QUADAS-2 RoB assessment was shown to have a mean accuracy of 92.63% (95% CI: 92.59–92.67).

3.4.3 Summary of extracted studies

The data extracted from the 250 papers included in the study are broadly summarized in Table 3.1. As stated previously, CA19-9 is the current FDA-approved biomarker for PDAC diagnosis. However, only 51.6% of papers included this biomarker in their study, while 96% of papers evaluated novel biomarkers. A total of 2,077 rows of data were extracted, each representing an individual biomarker entry, with 982 distinct biomarkers

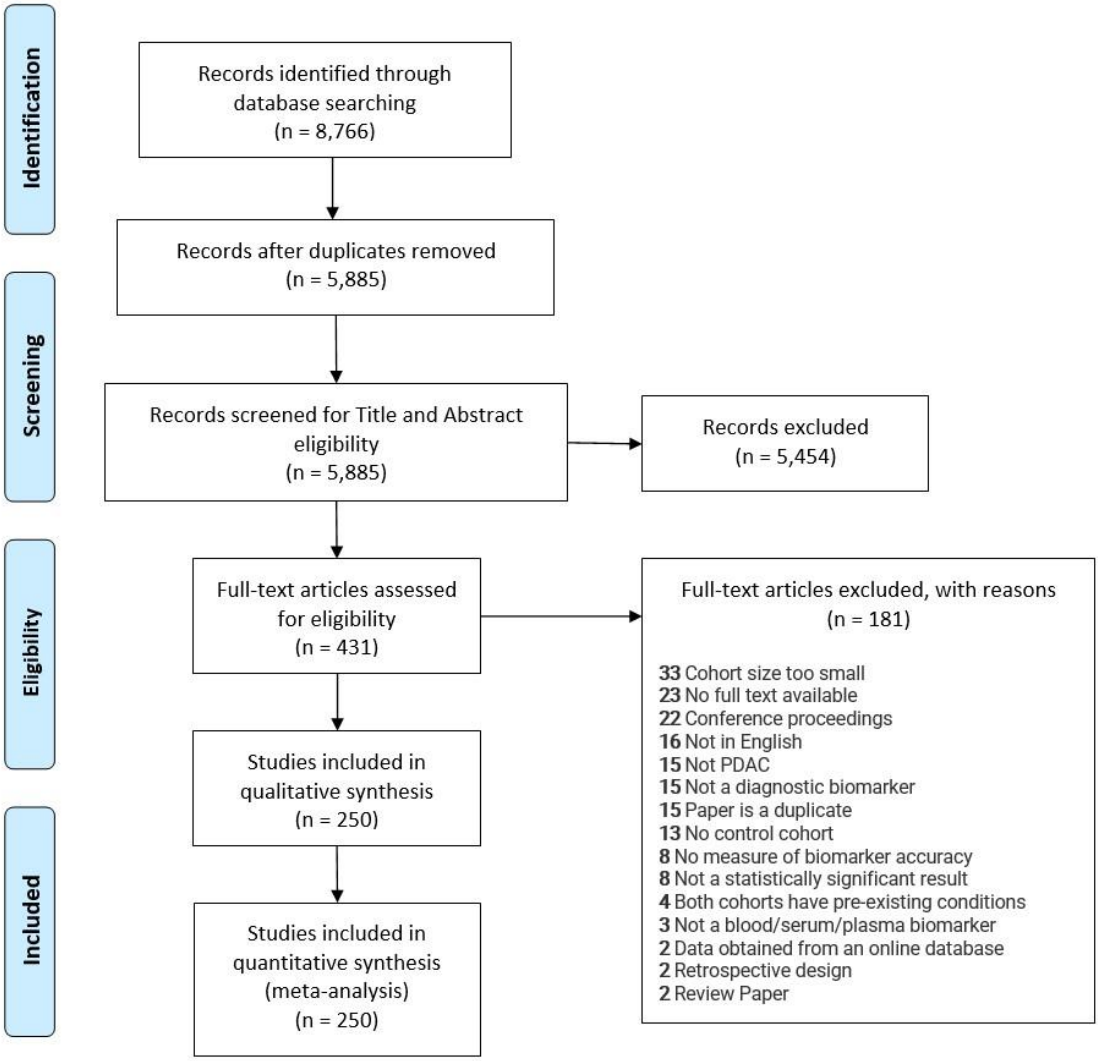


Figure 3.1. PRISMA flow diagram of record selection process. The number of studies at each point of the systematic review process are provided. Reasons for exclusions at the full-text stage are also outlined.

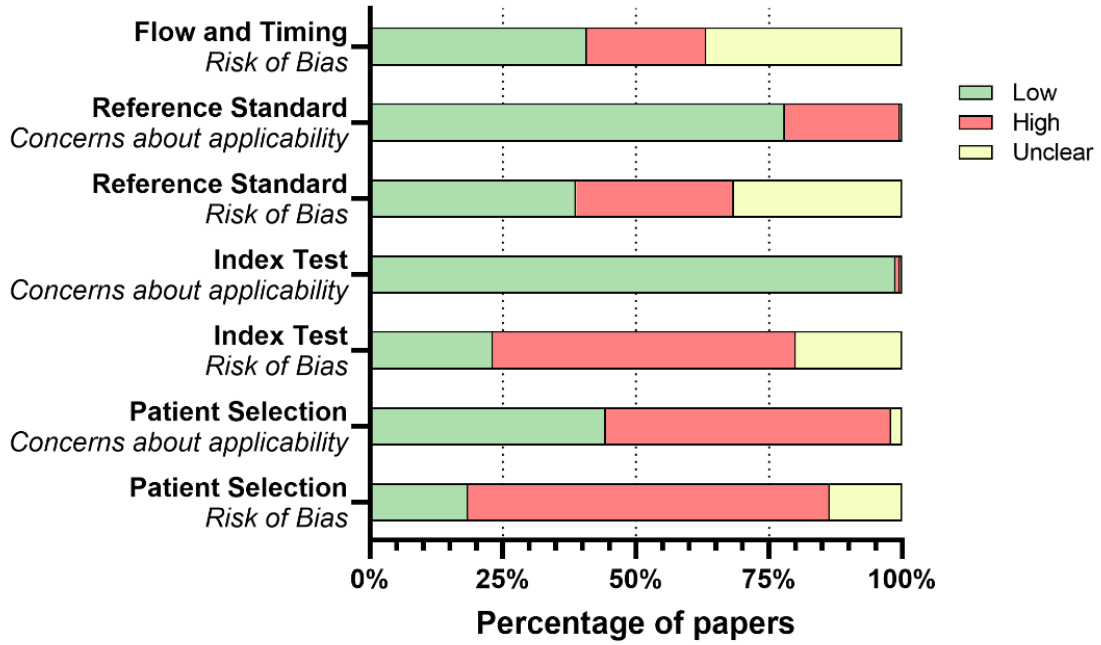


Figure 3.2. Summary of results for the QUADAS-2 RoB and study quality assessment. The percentage of included studies that were recorded as having a low (green), high (red) or unclear (yellow) RoB within each assessment category are shown.

Table 3.1 Summary of extracted data fields from included papers.

Summary of papers			
Total extracted[a]	**250**	***PDAC cohort details***	
Examining >1 biomarker	**196** (78.4%)	Mean PDAC cohort size (range)	**60.28** (15–809)
Examining CA19-9 biomarkers	**129** (51.6%)	PDAC reference standard	**1,158** (55.8%)
Examining novel biomarkers	**240** (96%)	No PDAC reference standard	**919** (44.2%)
		No PDAC stage details	**650** (31.3%)
Summary of unique biomarker entries		No sex breakdown	**611** (29.4%)
Total number of biomarker entries[b]	**2,077**	No age mean/median/range	**773** (37.2%)
Novel (Single/Multi)	**1,467** (1,228/239)	***Control cohort details***	
CA19-9 (Single/Multi)	**610** (293/317)	PDAC vs. healthy[c]	**1,084** (52.2%)
Number of unique biomarkers	**982**	*Mean cohort size (range)*	**50.7** (15–898)
Novel (Single/Multi)	**815** (675/140)	*No sex breakdown*	**356** (32.8%)
CA19-9 (Single/Multi)	**167** (1/166)	*No age mean/median/range*	**460** (42.4%)
Fluid type		PDAC vs. benign[d]	**867** (41.7%)
Serum	**1,411** (67.9%)	*Mean cohort size (range)*	**42.47** (15–786)
Plasma	**576** (27.7%)	*No sex breakdown*	**281** (32.4%)
Whole Blood	**88** (4.2%)	*No age mean/median/range*	**393** (45.3&)
Serum and Plasma	**2** (0.09%)	PDAC vs. mixed[e]	**126** (6.1%)
Study design		*Mean cohort size (range)*	**71.78** (33–199)
Prospective	**1,659** (79.9%)	*No sex breakdown*	**31** (24.6%)
Retrospective	**259** (12.5%)	*No age mean/median/range*	**30 (23.8%)**
Unclear	**159** (7.6%)	Total number of entries with:	
Cancer type		*No sex breakdown*	**668** (32.2%)
PDAC specified	**1,894** (91.2%)	*No age mean/median/range*	**883** (42.5%)
PC unspecified	**183** (8.8%)	***Statistical analyses***	
Cohort blinding		Qualitative assessment	**1,479** (71.2%)
Blinded	**740** (35.6%)	*P*-value alone	**598** (28.9%)
Unblinded	**1,012** (48.7%)	AUC	**1,206 (**58.1%)
Unclear	**325** (15.6%)	Sensitivity/Specificity	**900 (**43.3%)

Percentages for the "Summary of papers" use denominator [a]. Percentages for the "Summary of unique biomarker entries" use denominator [b], except where patient control cohorts are broken down and percentages use denominators [c-e].

included in the analysis. All studies examined blood-based biomarkers in either plasma, serum, or whole blood, with the majority of entries (67.9%) being investigated in serum. While 79.9% of studies recruited patients prospectively, 12.5% retrospectively examined clinical data of patients with PDAC from a hospital database within a certain time period (mean = 3.7 years, range = 14–0.6 years), and the remaining 7.6% were unclear about the recruitment process of patients. Study size varied greatly between papers, with PDAC cohorts ranging from 15 to 809 patients, and control cohorts ranging from 15 to 898 patients. Blinding across studies was shown to be poor, with only 35.6% of entries examined under blinded conditions, and 15.6% unclear on whether the study was blinded or not. As PDAC is synonymous with PC, studies that did not include a subtype of PC were assumed to be PDAC. Importantly, for 183 biomarker entries (8.8%) there was no specific subtype of PC given. PDAC diagnosis by use of a given reference standard (e.g., histology, cytology) was reported in most cases (55.8%). However, 44.2% of biomarker entries reported no reference standard for PDAC diagnosis. Furthermore, 29.4% provided no sex breakdown, 37.2% gave no indication of the age demographics, and 31.3% of entries had no information regarding the stage of patients with PDAC. Similarly, a substantial number of entries had no information on patient sex (32.2%) or age (42.5%) for their control cohorts. Qualitative assessments of biomarker efficacy were provided for most entries (71.2%); however, 598 entries (28.9%) contained only a p-value and did not provide any sensitivity, specificity or AUC values. More than 40% of biomarkers had no AUC value and were therefore not included in the meta-analysis.

3.4.4 Meta-analysis: full dataset

On the basis of the multivariate three-level meta-analysis with subgroup moderators, the pooled AUC value for all multi-biomarker panels (AUC = 0.898; 95% CI: 0.88–0.91) was significantly higher compared with single biomarkers (AUC = 0.803; 95% CI:0.78–0.83; p<0.0001)(Figure 3.3A). Overall, multi-biomarker panels show improved sensitivity and specificity compared with single biomarkers (Figure 3.4A). To further interrogate these data, biomarkers were subdivided into two groups: those including the current standard biomarker for pancreatic patients, CA19-9, and those without (herein known as novel biomarkers). The pooled AUC value for CA19-9–containing biomarkers (AUC =

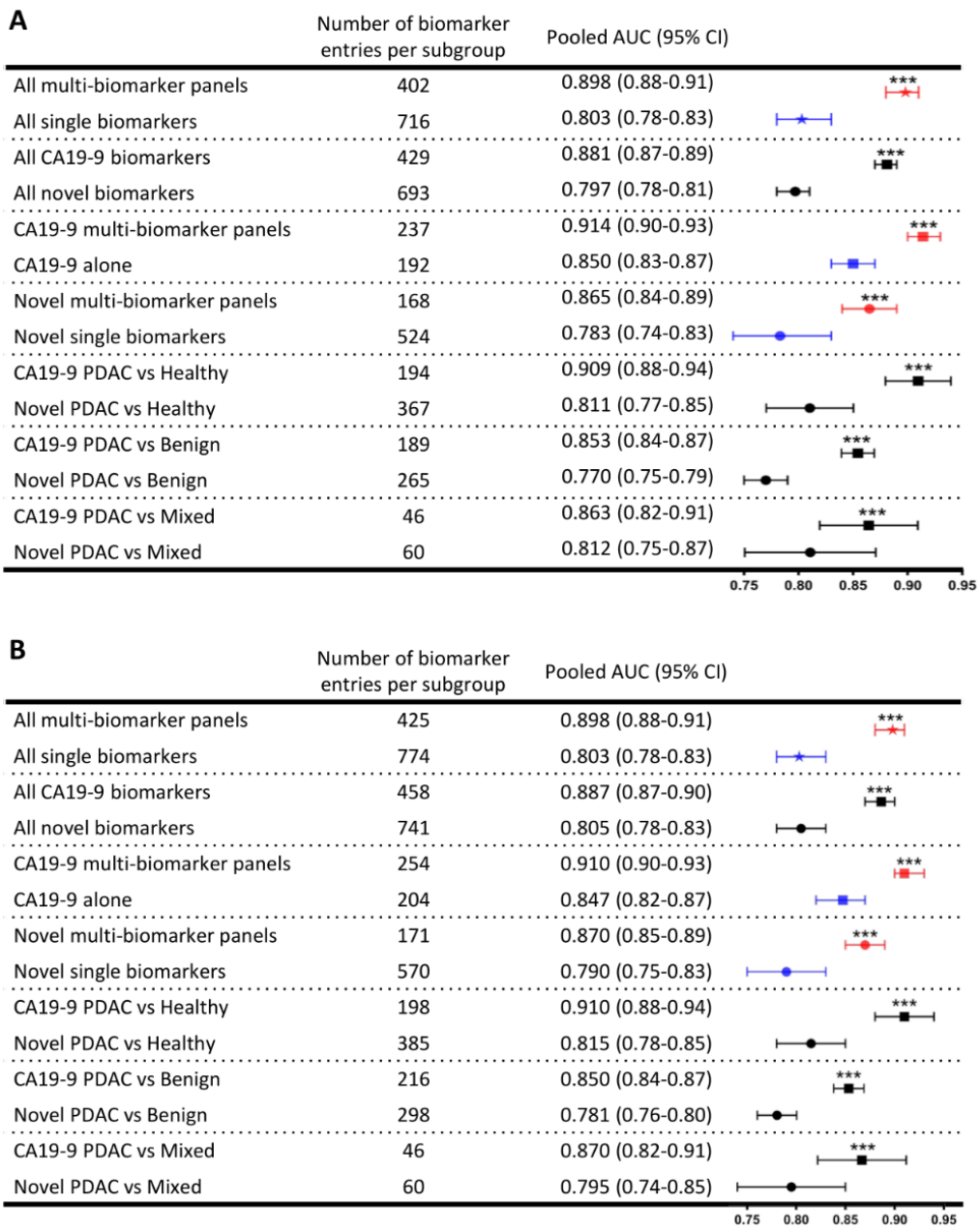


Figure 3.3 Multivariate three-level meta-analysis. Summary of multivariate three-level meta-analysis with subgroup moderators from (**A**) PDAC-specified papers only and (**B**) all extracted papers. Number of biomarker entries for each subgroup are given. The forest plot shows the pooled AUC value and 95% CIs for each biomarker subgroup from the multivariate three-level meta-analysis. Subgroups directly compared are separated by a dotted line. Symbols represent the whole dataset (★), CA19-9 subgroup (■) and novel subgroup (●). Colours represent biomarker type: all types (black), multi-biomarker panels (red), and single biomarkers (blue). Significantly higher AUC values are denoted using asterisks. ****p*<0.0001.

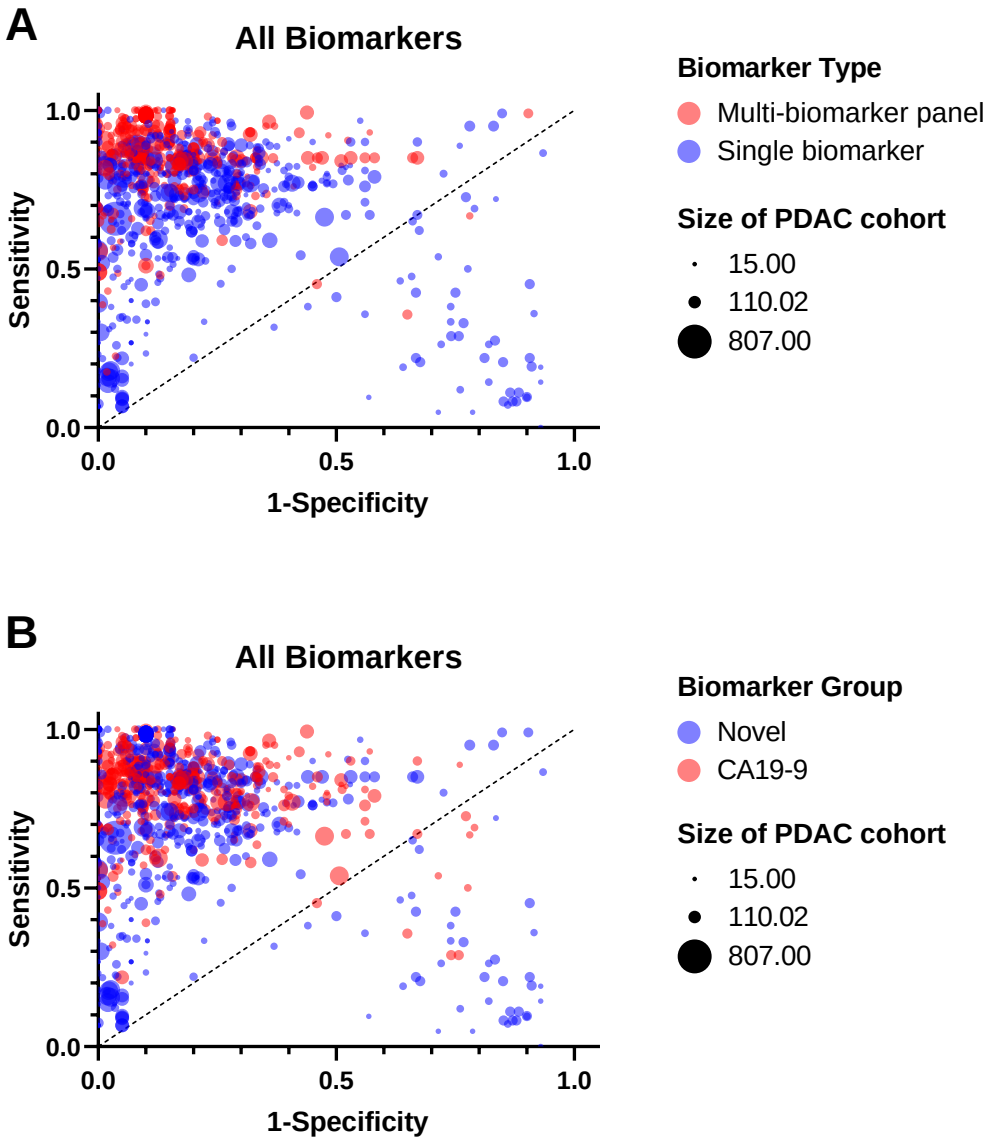


Figure 3.4 Comparison of single biomarkers and multi-biomarker panels overall and subdivided by biomarker group. Representative ROC plot showing the extracted sensitivity and 1-specificity for all biomarkers. Symbol colours represent biomarker type (**A**) and biomarker group (**B**), while symbol size represents the number of patients in the PDAC cohort for the given statistics.

0.881; 95% CI: 0.87–0.89) was significantly higher compared with novel biomarkers (AUC = 0.797; 95% CI: 78–81; p<0.0001). The sensitivity and specificity values for CA19-9 biomarkers appear improved compared with novel biomarkers (Figure 3.4B). A second meta-analysis was also conducted, including all studies, even those for which PDAC is not specified. There was no notable difference between the results of these two meta-analyses (Figure 3.3B).

3.4.5 Meta-analysis: CA19-9 and novel biomarker subgroups

Examining the CA19-9 and novel subgroups independently, the pooled AUC value for CA19-9 alone (AUC = 0.85; 95% CI: 0.83–0.87) was significantly lower compared with the multi-biomarker panels containing CA19-9 (AUC = 0.914; 95% CI: 0.90–0.93; p<0.0001)(Figure 3.3A). Multi-biomarker panels containing CA19-9 have improved sensitivity and specificity when compared with CA19-9 alone (Figure 3.5A). There is a large amount of variation in the reported sensitivity and specificity values between studies for the current standard biomarker, CA19-9, with some papers reporting values that fall below the random classifier line on the ROC plot. The estimated between-study variance in the model was $I^2_{Level\ 3}$ = 64.49%, and the within-study variance was $I^2_{Level\ 2}$ = 35.51%. This variation in CA19-9 was not a result of platform-to-platform discrepancies in CA19-9 detection throughout the studies examined in this review, as both immunoassays and mass-spectrometry-based detection of CA19-9 showed high variation within their respective platforms (Figure 3.6). As such, the differences observed are more likely to be a result of the patient populations examined rather than the platforms used. For the novel biomarkers, the pooled AUC for single biomarkers (AUC = 0.783; 95% CI: 0.74–0.83) was also significantly lower compared to novel multi-biomarker panels (AUC = 0.865; 95% CI: 0.84–0.89; p<0.0001). Novel multi-biomarker panels show improved sensitivity and specificity values over single biomarkers alone (Figure 3.5B and Figure 3.5D). Furthermore, there is less variation in the sensitivity and specificity values for multi-biomarker panels containing CA19-9 than CA19-9 alone, with a smaller interquartile range and higher mean and median values being shown for both test statistics (Figure 3.5C).

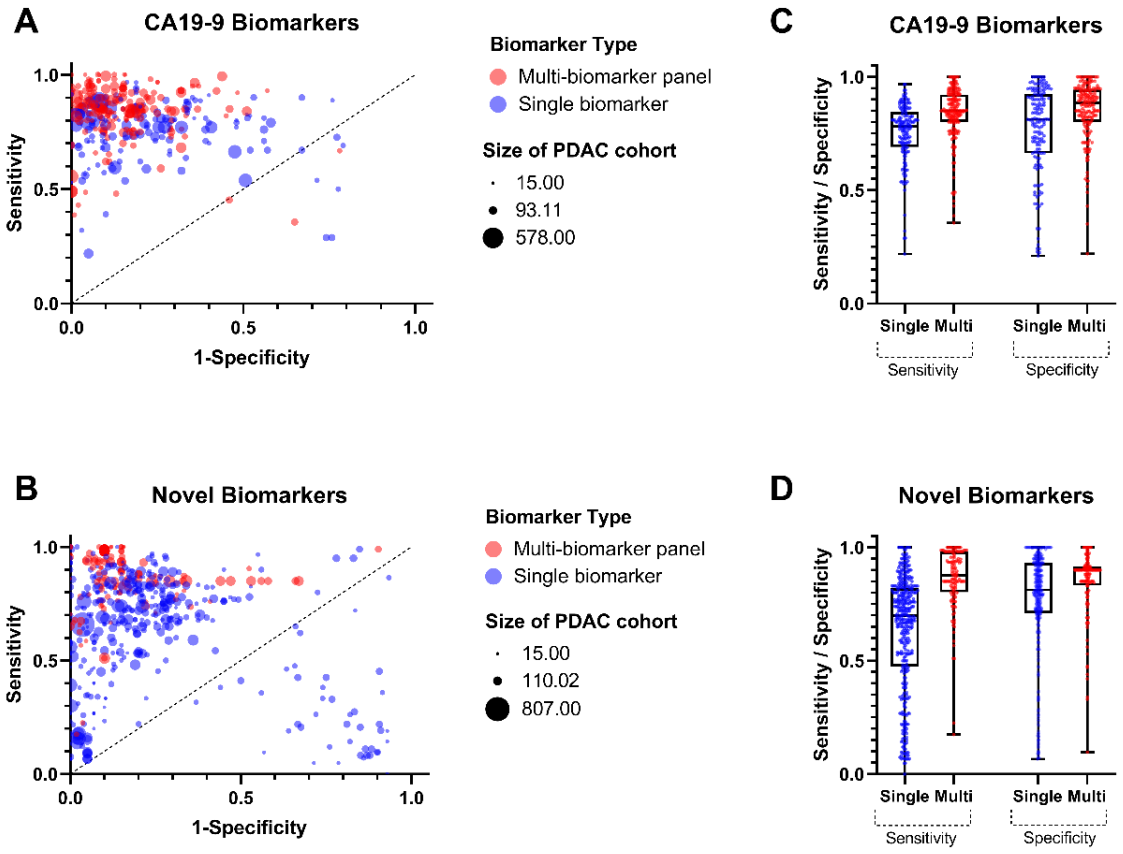


Figure 3.5 Comparison of single biomarkers with multi-biomarker panels for both CA19-9 and novel cohorts. Representative ROC plot showing the extracted sensitivity and 1-specificity for all biomarkers containing (**A**) CA19-9 and (**B**) all novel biomarkers. Symbol colours represent biomarker type while symbol size represents the number of patients in the PDAC cohort for the given statistics. Box and whisker plot showing the extracted sensitivity and specificity of all biomarkers containing (**C**) CA19-9 and (**D**) all novel biomarkers. Individual datapoints are represented by the coloured dots over the plot. Whiskers show the maximum and minimum value, boxes show the 25th and 75th percentiles, and the line within the box indicates the median.

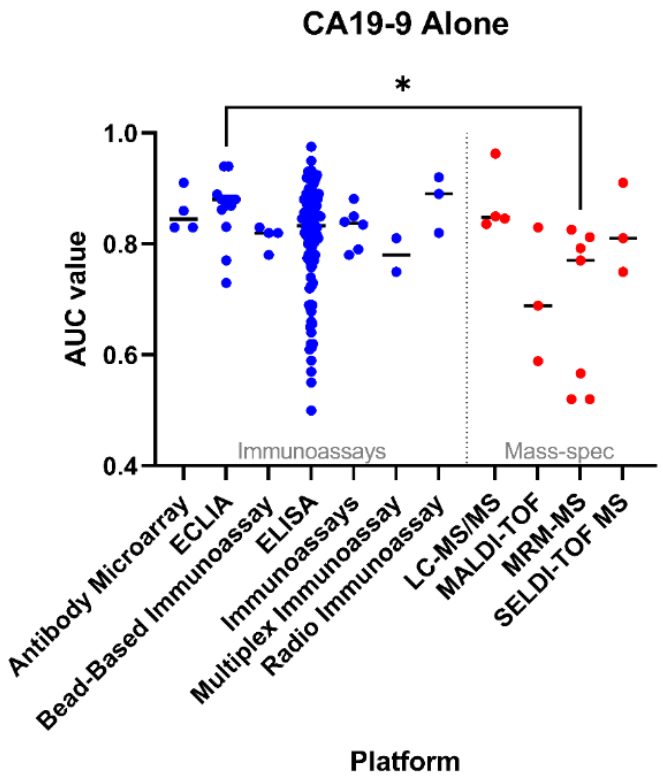


Figure 3.6 Evaluation of platform-to-platform variation in CA19-9 detection. AUC values are given for CA19-9 when evaluated using immunoassays (blue) or mass-spec (red). Kruskal-Wallis test with multiple comparisons. *p<0.05.

3.4.6 Meta-analysis: CA19-9 and novel biomarkers in different patient cohort subgroups

To further evaluate the efficacy of each biomarker and/or panel, results were subdivided on the basis of the patient cohorts involved as follows: PDAC versus healthy, PDAC versus benign, and PDAC versus mixed (healthy and benign) (Figure 3.7). Multi-biomarker panels demonstrate improved sensitivity and specificity when compared with single biomarkers across all patient cohorts. On the basis of the meta-analysis, biomarker robustness was also influenced by the patient cohort examined, with CA19-9–containing biomarkers performing best in all cohorts compared with novel biomarkers: PDAC versus healthy (AUC = 0.909; 95% CI: 0.88–0.94), PDAC versus benign (AUC = 0.853; 95% CI: 0.84–0.87), and PDAC versus mixed (AUC = 0.863; 95% CI: 0.82–0.91; $p<0.0001$) (Figure 3.3A). Furthermore, CA19-9 biomarkers examined in PDAC versus healthy cohorts have improved AUC values compared with those examined in PDAC versus mixed cohorts ($p<0.0001$).

3.4.7 Biomarker efficacy of different omic compartments

Proteomic biomarkers are the most frequently evaluated blood-based biomarkers for PC diagnosis (Figure 3.8). Proteomic biomarkers represent 77.9% of novel single biomarkers and are present in 50.3% of novel multi-biomarker panels examined. Of the panels where some other biomarker(s) was combined with the current standard, CA19-9, 76.4% opted to add another protein biomarker. Given the lack of diversity in omic compartments between studies, no distinct difference can be observed between the sensitivity and specificity values for biomarkers of different omic compartments (Figure 3.8).

Given the high prevalence of proteomic-orientated studies, biomarkers were pooled into categories that represent generalized cell compartments as follows: Genomics & Transcriptomics & Epigenomics; Metabolomics & Proteomics; and Single-cell omics & Immunomics. After pooling biomarkers into these groups, there was still no visible difference in sensitivity or specificity values between the different omic compartments (Figure 3.9).

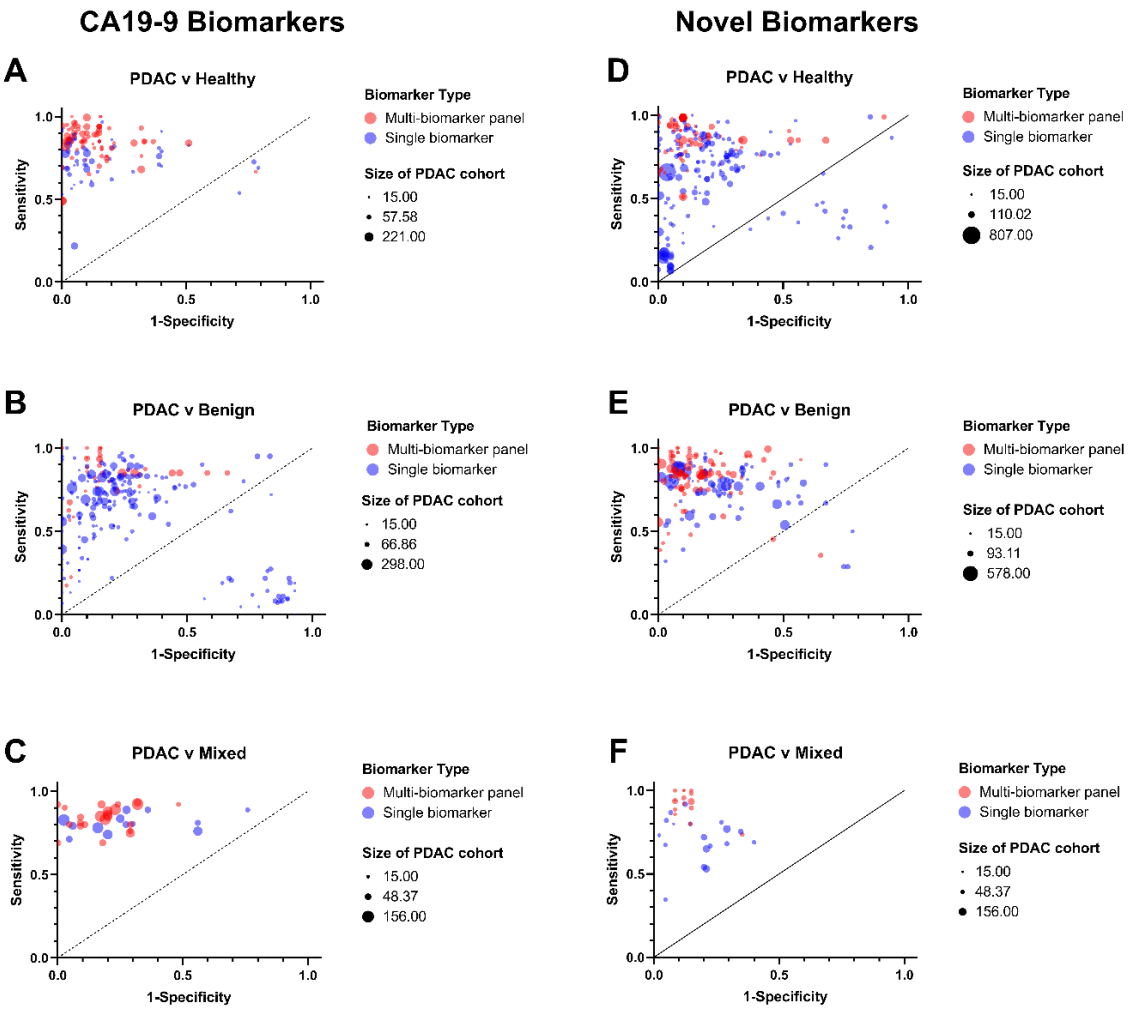


Figure 3.7 Comparison of CA19-9 and novel biomarkers subdivided by patient cohorts. Comparison of CA19-9 alone (blue) and multi-biomarker panels (red) for (**A**) PDAC versus healthy, (**B**) PDAC versus benign and, (**C**) PDAC versus mixed patient cohorts. Comparison of novel single biomarkers (blue) and novel multi-biomarker panels (red) for (**D**) PDAC versus healthy, (**E**) PDAC versus benign and, (**F**) PDAC versus mixed patient cohorts. Symbol size represents the number of patients in the PDAC cohort for the given statistics.

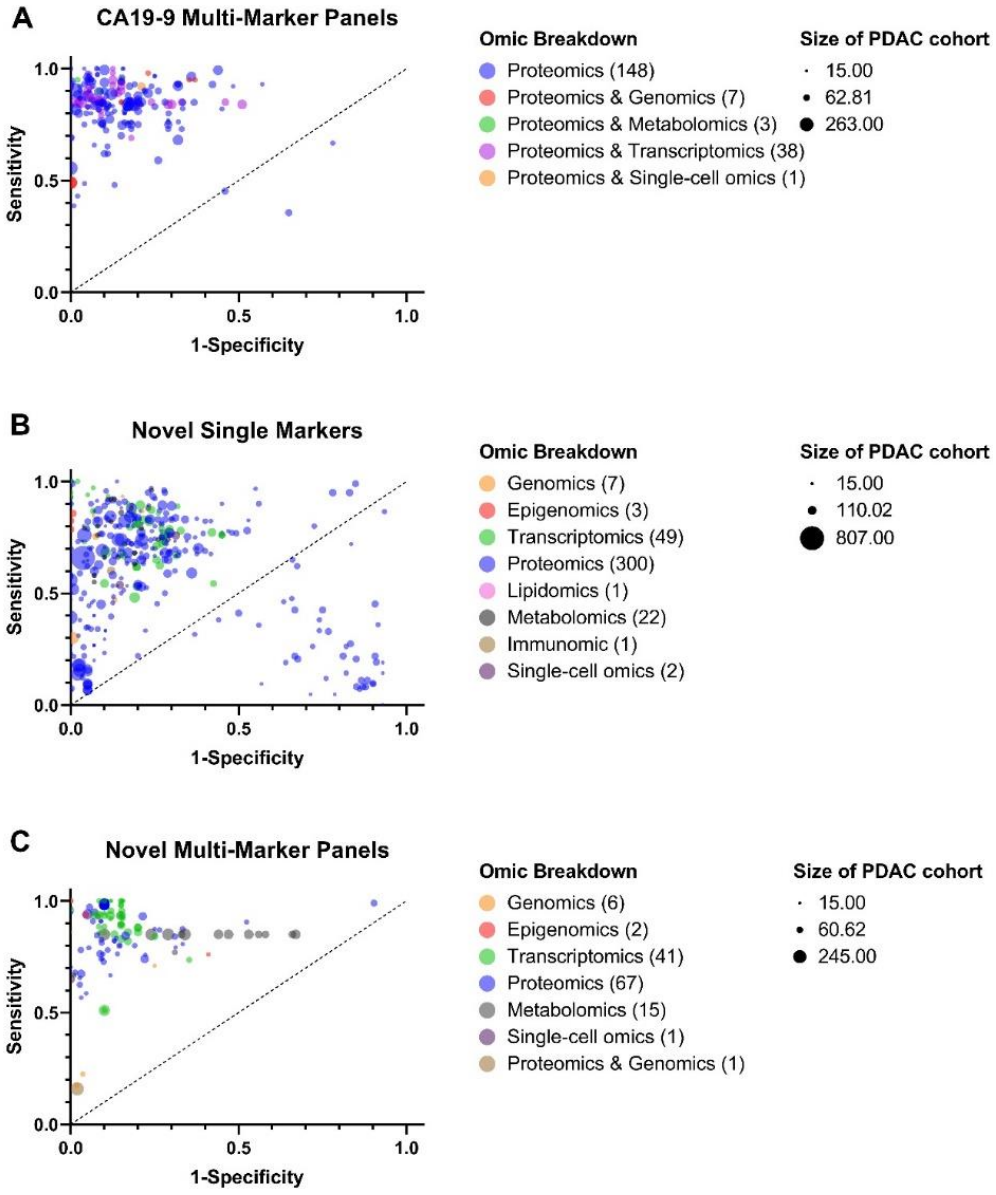


Figure 3.8 Comparison of CA19-9 and novel markers showing omic compartment breakdown. Representative ROC plot showing the extracted sensitivity and 1-specificity for (**A**) CA19-9 multi-marker panels, (**B**) novel single markers and (**C**) novel multi-marker panels. Symbol colours represent different omic compartments. Symbol size represents the number of patients in the PDAC cohort for the given statistics. Figures inside the brackets give exact numbers of markers from each omic compartment(s).

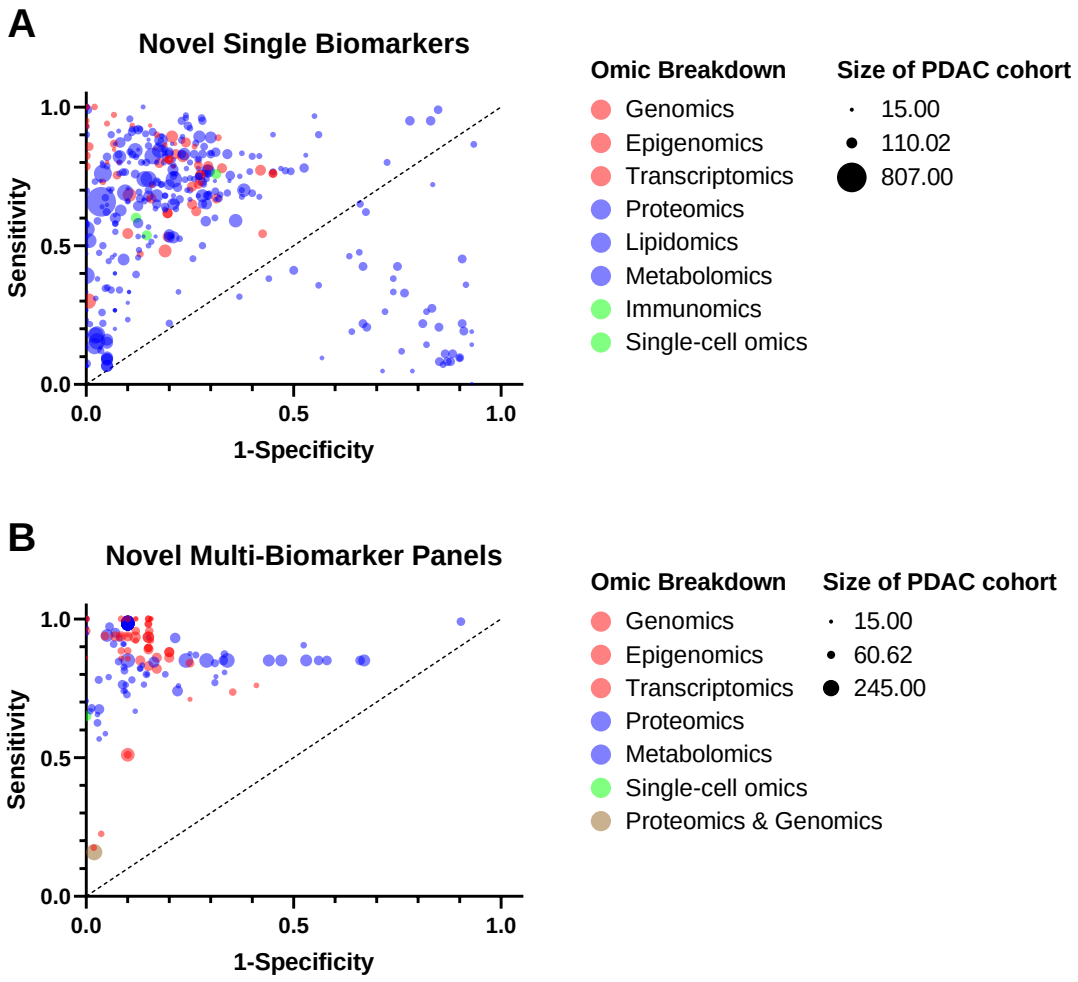


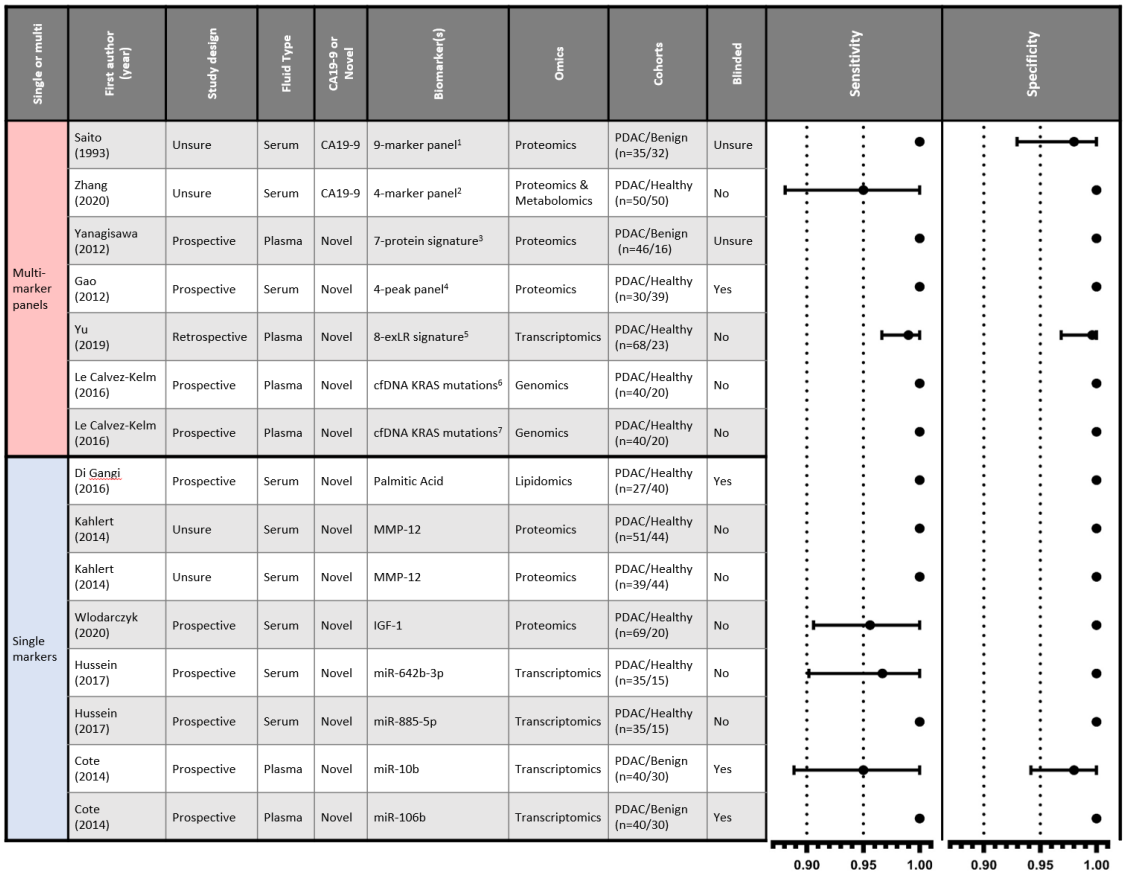
Figure 3.9 Comparison of novel markers showing pooled omic breakdown. Representative ROC plot showing the extracted sensitivity and 1-specificity for (**A**) novel single markers, and (B) novel multi-marker panels. Symbol colours represent different omic compartments, with generalized cell compartments pooled. Symbol size represents the number of patients in the PDAC cohort for the given statistics.

3.4.8 Biomarkers in the 90th percentile for sensitivity and specificity

Biomarkers in the 90th percentile of all entries for which there are sensitivity and specificity values, represented by a sensitivity value equal to or above 0.95 and a specificity value equal to or above 0.979, are summarized in Figure 3.10[305-315]. A total of 15 biomarkers comprise the 90th percentile, with seven of those being multi-biomarker panels. Most of these biomarkers were proteomic (n=6) or transcriptomic (n=5), with just one representing more than one omic compartment (proteomics and metabolomics). These biomarkers were identified from just 11 studies spanning 27 years (1993–2020), with four studies reporting on two biomarkers each. Nine of the top 15 biomarkers reported perfect sensitivity (1.00) and specificity (1.00), with a 95% CI of ± 0. Nine biomarkers are reported from studies that were not blinded, while just four biomarkers out of the 15 biomarkers were examined in a blinded study design. The most common patient cohorts for biomarker assessment were PDAC versus healthy (n = 11), with none of the top 15 biomarkers having been examined in the more clinically relevant PDAC versus mixed cohort. This is reflected in the RoB assessments for these studies, where several have high levels of bias for patient selection and index test (Figure 3.11). Only two biomarkers contained the current clinical standard biomarker, CA19-9, with 13 of 15 biomarkers being novel biomarkers. Furthermore, both biomarkers in the CA19-9 subgroup were multi-biomarker panels, with no reported sensitivity or specificity values for CA19-9 alone across all 250 included papers being within the 90th percentile of examined biomarkers.

3.4.9 Most frequently examined novel biomarkers across included studies

Novel biomarkers that have been examined most frequently across the studies included in this review are shown in Table 3.2. A total of 13 novel biomarkers were examined in more than one study and had a minimum of 21 unique entries. Tissue inhibitor matrix metalloproteinase 1 (TIMP-1) is the most examined biomarker, appearing 79 times in 10 studies, with CEA being a close second with 73 unique appearances in 34 studies. MiR-21 had the lowest number of unique appearances at 21, though these spanned across 10 studies. Of these 13 novel biomarkers, 10 were proteomic and three were transcriptomic in nature, with all 13 having been examined both alone and as part of a



Single or multi	First author (year)	Study design	Fluid Type	CA19-9 or Novel	Biomarker(s)	Omics	Cohorts	Blinded
Multi-marker panels	Saito (1993)	Unsure	Serum	CA19-9	9-marker panel[1]	Proteomics	PDAC/Benign (n=35/32)	Unsure
	Zhang (2020)	Unsure	Serum	CA19-9	4-marker panel[2]	Proteomics & Metabolomics	PDAC/Healthy (n=50/50)	No
	Yanagisawa (2012)	Prospective	Plasma	Novel	7-protein signature[3]	Proteomics	PDAC/Benign (n=46/16)	Unsure
	Gao (2012)	Prospective	Serum	Novel	4-peak panel[4]	Proteomics	PDAC/Healthy (n=30/39)	Yes
	Yu (2019)	Retrospective	Plasma	Novel	8-exLR signature[5]	Transcriptomics	PDAC/Healthy (n=68/23)	No
	Le Calvez-Kelm (2016)	Prospective	Plasma	Novel	cfDNA KRAS mutations[6]	Genomics	PDAC/Healthy (n=40/20)	No
	Le Calvez-Kelm (2016)	Prospective	Plasma	Novel	cfDNA KRAS mutations[7]	Genomics	PDAC/Healthy (n=40/20)	No
Single markers	Di Gangi (2016)	Prospective	Serum	Novel	Palmitic Acid	Lipidomics	PDAC/Healthy (n=27/40)	Yes
	Kahlert (2014)	Unsure	Serum	Novel	MMP-12	Proteomics	PDAC/Healthy (n=51/44)	No
	Kahlert (2014)	Unsure	Serum	Novel	MMP-12	Proteomics	PDAC/Healthy (n=39/44)	No
	Wlodarczyk (2020)	Prospective	Serum	Novel	IGF-1	Proteomics	PDAC/Healthy (n=69/20)	No
	Hussein (2017)	Prospective	Serum	Novel	miR-642b-3p	Transcriptomics	PDAC/Healthy (n=35/15)	No
	Hussein (2017)	Prospective	Serum	Novel	miR-885-5p	Transcriptomics	PDAC/Healthy (n=35/15)	No
	Cote (2014)	Prospective	Plasma	Novel	miR-10b	Transcriptomics	PDAC/Benign (n=40/30)	Yes
	Cote (2014)	Prospective	Plasma	Novel	miR-106b	Transcriptomics	PDAC/Benign (n=40/30)	Yes

Figure 3.10 Details of the biomarkers in the 90th percentile for sensitivity and specificity. Details of the 15 biomarkers that are in the 90th percentile of all biomarkers for both sensitivity (≥0.95) and specificity (≥0.979). Forest plots of sensitivity and specificity values with 95% CIs for each biomarker are shown. [1]CA19-9, DUPAN-2, TPA, elastase-1, lipase, amylase, gamma-glutamyl transpeptidase, alkaline phosphatase, and lactate dehydrogenase. [2]CA19-9, docosahexanoic acid, lysoPC(14:0), and histidinyl-lysine. [3]8562.3m/z, 8684.4m/z, 8765.1m/z, 9423.5m/z, 13761.5m/z, 14145.2m/z, and 17250.8m/z. [4]7,775 Da, 8,567 Da, 5,362 Da, and 5,344 Da. [5]FGA, KRT19, HIST1H2BK, ITIH2, MARCH2, CLDN1, MAL2, and TIMP1. [6]cfDNA KRAS mutations at PDAC hotspot codons (12, 13, 61). [7]cfDNA KRAS mutations at any screened codons reported in any cancer sites.

Single or multi	First author (year)	Study design	Fluid Type	CA19-9 or Novel	Biomarker(s)	Patient Selection: Risk of Bias	Patient Selection: Concerns about applicability	Index Test: Risk of Bias	Index Test: Concerns about applicability	Reference Standard: Risk of Bias	Reference Standard: Concerns about applicability	Flow and Timing: Risk of Bias
Multi-marker panels	Saito (1993)	Unsure	Serum	CA19-9	9-marker panel[1]	?	✓	?	✓	✓	✓	?
	Zhang (2020)	Unsure	Serum	CA19-9	4-marker panel[2]	?	✓	?	✓	?	✓	?
	Yanagisawa (2012)	Prospective	Plasma	Novel	7-protein signature[3]	✗	✓	?	✓	✓	✓	✓
	Gao (2012)	Prospective	Serum	Novel	4-peak panel[4]	✗	✓	✓	✓	✓	✓	✓
	Yu (2019)	Retrospective	Plasma	Novel	8-exLR signature[5]	?	✓	?	✓	?	✓	?
	Le Calvez-Kelm (2016)	Prospective	Plasma	Novel	cfDNA KRAS mutations[6]	✗	✗	✗	✓	?	✗	?
	Le Calvez-Kelm (2016)	Prospective	Plasma	Novel	cfDNA KRAS mutations[7]	✗	✗	✗	✓	?	✗	?
Single markers	Di Gangi (2016)	Prospective	Serum	Novel	Palmitic Acid	✗	✓	✓	✓	✓	✓	✓
	Kahlert (2014)	Unsure	Serum	Novel	MMP-12	?	✓	?	✓	?	✓	✗
	Kahlert (2014)	Unsure	Serum	Novel	MMP-12	?	✓	?	✓	?	✓	✗
	Wlodarczyk (2020)	Prospective	Serum	Novel	IGF-1	✗	✗	✗	✓	✗	✓	✗
	Hussein (2017)	Prospective	Serum	Novel	miR-642b-3p	✗	✗	✗	✓	✗	✓	✓
	Hussein (2017)	Prospective	Serum	Novel	miR-885-5p	✗	✗	✗	✓	✗	✓	✓
	Cote (2014)	Prospective	Plasma	Novel	miR-10b	✗	✗	✗	✓	?	✗	?
	Cote (2014)	Prospective	Plasma	Novel	miR-106b	✗	✗	✗	✓	?	✗	?

Figure 3.11 Details of the risk of bias assessment for biomarkers in the 90th percentile for sensitivity and specificity. Details of the 15 biomarkers that are in the 90th percentile of all biomarkers for both sensitivity (≥0.95) and specificity (≥0.979). Risk of bias assessment results are given for each biomarker as high (✗), low (✓) and unclear (?). [1]CA19-9, DUPAN-2, TPA, elastase-1, lipase, amylase, gamma-glutamyl transpeptidase, alkaline phosphatase, and lactate dehydrogenase. [2]CA19-9, docosahexanoic acid, lysoPC(14:0), and histidinyl-lysine. [3]8562.3m/z, 8684.4m/z, 8765.1m/z, 9423.5m/z, 13761.5m/z, 14145.2m/z, and 17250.8m/z. [4]7,775 Da, 8,567 Da, 5,362 Da, and 5,344 Da. [5]FGA, KRT19, HIST1H2BK, ITIH2, MARCH2, CLDN1, MAL2, and TIMP1. [6]cfDNA KRAS mutations at PDAC hotspot codons (12, 13, 61). [7]cfDNA KRAS mutations at any screened codons reported in any cancer sites.

Table 3.2 Performance of the most frequently examined novel biomarkers.

Novel biomarker	Number of papers	Omic compartment	Number of unique appearances		Sensitivity range (mean)	Specificity range (mean)	AUC range (mean)
TIMP-1	**10**	**Transcriptomics**	**79**	*Alone: 30*	0.1–0.5 (0.28)	0.09–0.96 (0.4)	0.61–0.95 (0.78)
				Part of panel: 49	0.36–1 (0.89)	0.22–1 (0.81)	0.83–0.99 (0.94)
CEA	**34**	**Proteomics**	**73**	*Alone: 42*	0.06–0.8 (0.41)	0.17–0.99 (0.68)	0.53–0.82 (0.67)
				Part of panel: 31	0.16–0.98 (0.77)	0.67–1 (0.9)	0.78–0.99 (0.92)
CA242	**5**	**Proteomics**	**45**	*Alone: 19*	0.31–0.83 (0.62)	0.51–1 (0.78)	0.62–0.89 (0.75)
				Part of panel: 26	0.57–0.97 (0.77)	0.75–1 (0.92)	0.8–0.98 (0.9)
CA125	**11**	**Proteomics**	**38**	*Alone: 20*	0.07–0.89 (0.42)	0.69–1 (0.93)	0.57–0.79 (0.7)
				Part of panel: 18	0.79–0.99 (0.96)	0.69–096 (0.89)	0.8–0.93 (0.87)
miR-483	**2**	**Transcriptomics**	**37**	*Alone: 8*	N/A	N/A	0.7–0.75 (0.73)
				Part of panel: 29	N/A	N/A	0.63–0.99 (0.82)
MUC5AC	**3**	**Proteomics**	**31**	*Alone: 27*	0.65–0.95 (0.78)	0.7–0.9 (0.77)	0.68–0.94 (0.82)
				Part of panel: 4	N/A	N/A	0.69–0.93 (0.81)
THBS2	**5**	**Proteomics**	**28**	*Alone: 15*	0.07–0.52 (0.19)	0.974–1 (0.99)	0.61–0.89 (0.79)
				Part of panel: 13	0.62–0.9 (0.75)	0.9–1 (0.95)	0.76–0.98 (0.9)
IL-8	**7**	**Proteomics**	**24**	*Alone: 6*	0.15–0.72 (0.33)	0.72–0.95 (0.83)	0.6–0.71 (0.65)
				Part of panel: 18	0.75–0.99 (0.87)	0.58–1 (0.85)	0.81–1 (0.92)
CRP	**5**	**Proteomics**	**24**	*Alone: 7*	0.2–0.77 (0.48)	0.55–0.93 (0.71)	N/A
				Part of panel: 17	0.83–0.99 (0.96)	0.9–1 (0.91)	0.91–0.98 (0.96)
ALB	**4**	**Proteomics**	**24**	*Alone: 4*	0.79–0.79 (0.79)	N/A	0.18–0.87 (0.4)
				Part of panel: 20	0.92–0.99 (0.97)	0.9–1 (0.91)	0.95–0.98 (0.97)
LAMC2	**2**	**Proteomics**	**23**	*Alone: 11*	N/A	N/A	0.65–0.87 (0.81)
				Part of panel: 12	N/A	N/A	0.8–0.96 (0.88)
LRG1	**3**	**Proteomics**	**22**	*Alone: 10*	0.5–0.46 (0.17)	0.1–0.37 (0.19)	0.64–0.94 (0.77)
				Part of panel: 12	0.36–0.92 (0.72)	0.22–0.83 (0.56)	0.82–0.96 (0.89)
miR-21	**10**	**Transcriptomics**	**21**	*Alone: 17*	N/A	N/A	0.49–0.99 (0.7)
				Part of panel: 4	0.85–0.9 (0.87)	0.85–0.87 (0.86)	0.82–0.95 (0.88)

Sensitivity, specificity, and AUC value breakdowns are given for novel biomarkers that were examined in more than one study and had a minimum of 21 unique entries. N/A indicates that there are no extracted data for this field for a minimum of two biomarker entries.

panel. The mean sensitivity of all novel biomarkers is higher when examined as part of a multi-biomarker panel. This holds true for mean specificity in most biomarkers also, with just CA125 and thrombospondin-2 (THBS2) showing improved mean specificity alone compared with when part of a panel. Mean AUC values are higher when examined as part of a panel for all biomarkers except Mucin 5AC (MUC5AC), with superior mean AUC values for this biomarker being found when examined alone. Albumin (ALB) had the highest mean sensitivity values both alone and as part of a panel, with LRG1 having the lowest in both cases. THBS1 had the highest mean specificity both alone and as part of a panel, with LRG1 again performing the worst in both categories. MUC5AC had the highest mean AUC value alone; however, it had the lowest mean AUC value as part of a panel. Conversely, ALB has the highest mean AUC value as part of a panel, and the lowest mean AUC value when examined alone. TIMP-1 is the only of these biomarkers that also appears among the biomarkers in the 90th percentile in Figure 3.10, where it appears as part of an 8-biomarker panel.

3.5 Discussion

Currently, there is no biomarker that can effectively and consistently discriminate patients with PDAC from those without. The aim of this systematic review and meta-analysis was to examine the performance of all published blood-based biomarkers used for the diagnosis of PDAC. Specifically, papers that evaluated some blood-based biomarker(s) for the diagnosis of PDAC were examined, with no limit placed on the publication date of the paper or the "omic" compartment of the biomarker(s) assessed. The performance of all biomarkers was examined in order to evaluate whether single biomarkers or multi-biomarker panels generally have the best efficacy for PDAC diagnosis. This evaluation was carried out through the use of a multivariate three-level meta-analysis, with AUC values as effect sizes, and by comparing sensitivity and specificity values.

3.5.1 Multi-biomarker panels - the better choice

Overall, multi-biomarker panels are significantly more robust than single biomarkers alone, and in the context of PDAC, this holds true for both CA19-9 and novel biomarker subgroups, as well as across different patient control cohorts. This review shows extensive evidence, both graphically and statistically, that panels of more than one biomarker tend to perform better than single biomarkers alone for the diagnosis of PDAC. Furthermore, it was evident when eliminating confounding variables by subdividing the data into different groups, using variables such as biomarker type (CA19-9 or novel) or patient control cohort examined (healthy, benign, or mixed), that this result is robust and prevails throughout multiple subgroup analyses. Importantly, the inclusion of studies that do not specify PDAC did not greatly alter the results of the meta-analysis, suggesting that the cohorts within these studies are similar to those included in the original analysis. While there were many single biomarkers reported in the included studies with impressive efficacy, the results of the meta-analysis indicate that on the whole, multi-biomarker panels produce the most robust diagnostic performance. This information is crucial for future studies, as it suggests that researchers should focus their efforts on the identification of multiple biomarkers, rather than attempting to isolate one single biomarker. It is important to note also, that the creation of a multi-

biomarker panel is not as straightforward as it may seem, and when dealing with multiple levels of patient data and consequently different cut-offs for individual biomarkers, care must be taken to ensure the desired sensitivity and specificity of the panel as a whole. Integration of these data will allow researchers the flexibility to tailor their panel to certain conditions, and determine individual cut-offs based on the needs of the test[166]. Computational approaches, such as machine learning, have shown utility in this context and could provide future research with a more streamlined approach to multi-biomarker panel generation[316]. One caveat to this, however, is the integration of multi-omic data, which would require cautious consideration of the various unit measurements involved and, in each case, careful control of confounding variables and potential artefacts of experimental design[172]. In any case, whether multi-omic or single-omic, the thoughtful generation of highly sensitive and specific multi-biomarker panels should arguably be the primary aim of future studies hoping to identify novel diagnostic biomarkers for PDAC.

3.5.2 CA19-9 and its role as the current clinical standard biomarker

As CA19-9 is regarded as the current standard biomarker for PC diagnosis, the data were separated into two groups, those including CA19-9 and those without (novel biomarkers), to evaluate the performance of CA19-9 across all studies. For both subgroups of data, multi-biomarkers were shown to perform significantly better than single biomarkers alone. Indeed, it was found that while CA19-9 is the "gold standard" for pancreatic diagnosis, the addition of some other biomarker to create a multi-biomarker panel with CA19-9 resulted in improved biomarker efficacy compared with CA19-9 alone. Furthermore, there was a substantial amount of variation between studies in the reported sensitivity and specificity values of CA19-9, and this is in keeping with current literature[191, 192, 317, 318].

The results of the meta-analysis showed that the addition of CA19-9 to a multi-biomarker panel provided a clear improvement over novel biomarker panels that did not contain CA19-9. In addition, CA19-9 alone appears to have consistently outperformed novel single biomarkers. While CA19-9 may be the most commonly used biomarker for diagnosis of PDAC in patients with PC, elevated expression has been

shown in various benign conditions such as pancreatitis, which contributes to its non-specificity for PDAC[318]. Given the similarities that are often observed between benign pancreatic patient blood and PC patient blood, it follows therefore that a biomarker may have a diminished ability to distinguish benign cohorts from those with cancer when compared with healthy cohorts. To ensure the differences being observed were not a result of the patient cohorts being evaluated by different studies, the data were further separated based on the patient cohorts distinguished from PDAC: PDAC versus healthy, PDAC versus benign, and PDAC versus mixed. When separated based on patient control cohorts, the improved ability of CA19-9 biomarkers over novel biomarkers to diagnose PDAC is clear.

It is also demonstrated here that the efficacy of a biomarker or biomarker panel to diagnose PDAC, whether including CA19-9 or a novel biomarker, is dependent on the reference cohort in question. Both CA19-9 and novel biomarkers exhibit improved ability to distinguish patients with PDAC from healthy controls when compared with both benign and mixed cohorts, with CA19-9 multi-biomarker panels producing the best diagnostic performance across all cohorts. There is considerable variation across pooled AUC values for CA19-9 biomarkers and novel biomarkers in the mixed patient cohort setting, demonstrating the increased difficulty faced in this control cohort. As a result of the recognized poor specificity of CA19-9, the lack of a standardized detection method[167], and the variation in cut-off levels being used for PDAC diagnosis across the studies[319-321], in clinical practice CA19-9 is rarely relied upon for diagnostic purposes, despite being FDA approved for this indication. More often it is used to support a diagnosis, based on appropriate imaging and/ or biopsy, or in staging with Immuno-PET imaging[322], as a biomarker of recurrence[301, 303], or as a biomarker of tumour resectability[323, 324]. In fact, the results of this meta-analysis provide a strong argument in favour of the inclusion of CA19-9 when evaluating a new biomarker panel, while exercising caution given the variation in results obtained across different studies, and the reduced potential of CA19-9 in certain benign conditions. The role of CA19-9 in PC diagnosis, therefore, seems reliant on the identification of a robust multi-biomarker panel that can adequately control for the inherent defects of the biomarker.

3.5.3 Multi-omics in biomarker identification

While it is evident from the results of this review that multi-biomarker panels are the most robust biomarker type, it remains to be seen which biological factors produce the most robust biomarkers. A major limitation of this review results from the lack of diversity seen in the "omic" compartments (genomics, proteomics, etc.) of biomarkers. Indeed, proteomic biomarkers make up the vast majority of biomarkers evaluated across all studies in this review, making comparisons between proteomic biomarkers and other omic compartments difficult. It is evident, however, that the combination of different omic compartments with CA19-9 (proteomics) did result in high sensitivity and specificity values, though the number of studies examining multi-omic biomarker panels is too low to see any distinct difference. By examining the biomarkers that fall into the 90th percentile for both sensitivity and specificity, it is evident that while proteomic biomarkers may represent the majority, they do not solely comprise the top biomarkers. Indeed, nearly as many of these "top" biomarkers were transcriptomic in nature, highlighting the importance of examining different biological compartments for the discovery of robust biomarkers. Furthermore, while instances where multiple omic compartments were integrated to form a panel were uncommon among the papers included in this review, one multi-omic biomarker panel was among the 90th percentile biomarkers. A 4-biomarker panel containing CA19-9 and three metabolites was among those with the highest sensitivity and specificity, demonstrating the potential for such multi-omic biomarker panels in this context. While it is not within the scope of this review to evaluate whether multi-omic panels produce better results than single-omic panels, current trends in biomarker discovery are leaning toward multi-omic data integration[166, 325, 326]. The evaluation of multiple biological compartments to give a comprehensive overview of disease, and subsequently the generation of a robust panel that encompasses the complexity of that disease is an appealing concept that has much potential in this context and requires further elucidation.

3.5.4 Promising novel biomarkers for PC diagnosis

Systematic reviews are uniquely poised to identify trends in the literature that may otherwise go unnoticed. Importantly, this systematic review allowed for the

identification of 13 novel biomarkers that have been repeatedly examined as blood-based biomarkers for PDAC diagnosis across multiple studies, and show promise both alone and as part of a multi-biomarker panel. While again, the majority of these biomarkers are proteins, the transcriptomic biomarker TIMP-1 emerged as the most frequently assessed novel biomarker. Though it showed poor mean sensitivity and specificity alone, TIMP-1 performed well as part of a panel with improved mean sensitivity, specificity, and AUC values. This is further evident from its appearance among the 90th percentile biomarkers, where it achieved high sensitivity and specificity as part of an 8-biomarker panel of extracellular vesicle long RNAs. Because of its association with cell survival, cell growth, and tumorigenesis, the TIMP-1 protein has been investigated as a potential biomarker, both alone and as part of a panel, in several other cancer types such as gastric[327, 328], colorectal[329-331], and breast[332, 333]. This association, however, could be the reason for TIMP-1s poor utility alone in PDAC diagnosis. Indeed, TIMP-1 protein performance as a blood-based biomarker in PDAC has been shown to be impaired in patients with jaundice[334], though not in patients with chronic pancreatitis[335, 336]. Furthermore, TIMP-1 expression is known to be increased in patients with, and at an increased risk of developing, type 2 diabetes[337, 338], as well as in obese patients[339]. This evidence suggests that the utility of TIMP-1 in PDAC diagnosis is promising, and may lie in its addition to a biomarker panel rather than its use alone due to its impairment in patients with several benign conditions. Interestingly, TIMP-1 has been examined alongside another promising novel biomarker, inflammatory protein leucine-rich-alpha-2-glycoprotein 1 (LRG1). LRG1 has been shown to promote angiogenesis and regulate tumorigenesis, and is a promising biomarker candidate for several other cancer types[340]. Indeed, a plasma-based panel of TIMP-1, LRG1, and CA19-9 discriminated PDAC from healthy controls with improved accuracy compared with CA19-9 alone[341]. LRG1 has also been evaluated in plasma alongside TTR and CA19-9, where this panel exceeded the accuracy of CA19-9 alone by over 10% in its ability to discriminate PDAC from benign controls and other cancers[342]. While LRG1 shows promise as part of a panel for PDAC diagnosis, it has poor mean sensitivity and specificity alone, and there is a lack of research into its performance in control cohorts with various benign conditions.

Conversely, cancer antigen 125 (CA125), CEA, and carbohydrate antigen 242 (CA242) have been evaluated extensively in PC and as such there is a plethora of research on these biomarkers. CEA is an established and widely used tumour biomarker that is known to be increased in several cancers such as colorectal[343], breast[344], and lung[345]. While CEA levels are currently measured for PDAC diagnosis in some clinical settings, it is not FDA approved for PDAC diagnosis and its utility and accuracy remains limited, with a 2018 systematic review and meta-analysis reporting CEA to be inferior to CA19-9[346]. CEA levels are also known to be elevated in patients with chronic pancreatitis, with serum CEA being unable to distinguish patients with PDAC from those with chronic pancreatitis[347]. Indeed, in this review, it was shown that CEA alone exhibits poor diagnostic performance across included studies, with improved results being obtained when CEA is examined as part of a panel. CA125 is a known biomarker for ovarian cancer, which has been shown to have superior performance to CEA for PDAC diagnosis[348, 349]. It has also produced higher mean sensitivity, specificity, and AUC values than CEA across the studies included here. Furthermore, a 2017 systematic review and meta-analysis showed that a CA125-based diagnostic panel for PDAC was superior to CA125 or CA19-9 alone[350]. Similar to CEA, CA242 has also been extensively evaluated for PDAC diagnosis, with serum CA242 levels having been shown to positively correlate with CA19-9 levels[351]. CA242 has also been demonstrated to have better diagnostic performance than CEA, with mean sensitivity, specificity and AUC values in this study being higher for CA242 than CEA[352]. Unfortunately, CA242 is known to be elevated in the blood of patients with type 2 diabetes, and as such, has limited utility alone for PDAC diagnosis, despite exhibiting higher specificity than CA19-9, CEA, and CA125[348, 353]. While none of these biomarkers have stood out on their own as having utility across all patient cohorts, they are frequently examined as part of biomarker panels with other novel biomarkers. Laminin subunit gamma-2 (LAMC2), for example, is a promising new biomarker which was examined in a large-scale study of over 400 patients across three continents, where it was elevated in PC serum compared with controls and demonstrated a sensitivity that was comparable with CA19-9[354, 355]. Furthermore, a serum-based panel of LAMC2 with both CA19-9 and CA125 has been shown to produce accurate discrimination of PDAC from benign controls[192]. These studies provide promising results for LAMC2 as a potential diagnostic biomarker both

alone and as part of a panel, though the breadth of research is limited at this time, with no results for mean sensitivity or specificity being obtained in this review. Another promising biomarker that has produced results similar to CA19-9 is MUC5AC. Serum MUC5AC levels have been shown to be increased in patients with PDAC compared with both benign and chronic pancreatitis cohorts, with MUC5AC performing on par with CA19-9, though again, the combination of the two produced the best results[356, 357]. Interestingly, the measurement of the CA19-9 antigen on circulating MUC5AC proteins showed promise in a study comprising over 500 patients from three different institutions, where both the sensitivity and specificity of the biomarker were improved by this method compared with measuring just CA19-9 alone[358]. While the initial research on MUC5AC shows favourable results, there are few papers examining the capability of MUC5AC as part a multi-biomarker panel.

The addition of CA19-9 to some novel biomarker is a trend across most studies aimed at identifying diagnostic biomarkers for PDAC, with results generally reporting an improved result from the panel compared to individual biomarkers alone. Studies examining the potential of ALB in this setting are no different, with the vast majority of entries for ALB in this review originating from multi-biomarker panels. Indeed, a 5-biomarker panel containing ALB and CA19-9 produced improved diagnostic capabilities compared with CA19-9 alone[359]. Similarly, the combination of ALB with CA19-9 and IGF-1 also performed better than CA19-9 at distinguishing PDAC from chronic pancreatitis[347]. Interleukin-8 (IL-8), a proinflammatory cytokine, has also shown limited utility alone but appears to achieve reasonable results when included in a panel. Serum IL-8 has been shown to be higher in patients with PC than controls; however, the mean AUC value for IL-8 alone is poor and improved when included in a panel with other biomarkers[360-362]. This is also the case for THBS2, which produces modest discrimination alone, and is significantly improved when examined alongside CA19-9[363, 364]. Conversely, c-reactive protein (CRP), a biomarker of inflammation, has shown limited utility as part of a panel, where the panel showed no improvement with the addition of CRP[347]. A 2020 study also showed a significant difference in CRP levels between PDAC and normal controls; however, after running extensive statistical tests on all candidate biomarkers it was not included in the final panel of six biomarkers[365].

While CRP is increased in patients with PDAC compared with controls, it is also elevated in patients with moderate and severe pancreatitis[366]. Moreover, as CRP is derived from the liver, it is substantially influenced by the presence of jaundice making it unreliable in patients with this comorbidity[347]. Interestingly, several studied have examined these more "unreliable" candidates together, and obtained promising results. A 4-biomarker panel with CA19-9, CRP, and IL-8 demonstrated good discrimination of PDAC from controls[367]. While a 2014 study showed that a panel consisting of ALB, CA19-9, CRP, and IL-8 had the highest diagnostic value for distinguishing PDAC from controls, with this panel proving to be effective in identifying other cancers, such as breast, cervical, colorectal, prostate, and lung[367]. These studies highlight the utility of all of these biomarkers together, rather than independently.

Finally, two miRNA emerged as the most frequently examined across the studies included in this review, miR-21 and miR-483. MiR-21 levels in the circulation have been shown to be higher in PDAC compared with healthy controls, and are also associated with advanced stage, metastasis, and shorter survival[368, 369]. However, miR-21 shows poor discriminatory ability between IPMN and PDAC, suggesting the involvement of miR-21 in an early step of pancreatic tumorigenesis[368]. Indeed, the ability of miR-21 to distinguish PDAC from controls was overshadowed by several other miRNA in a 2019 study, such as miR-33a and miR-320a, which outperformed miR-21 in combination, thus excluding miR-21 from the final panel[370]. Overexpression of miR-483 is also thought to be an early event in PDAC progression, having been shown to be present in premalignant PCL and early-stage disease[371]. A 2016 large-scale miRNA study with over 400 patients with PDAC showed that serum miR-483 expression was significantly increased in patients with PDAC compared with both benign and healthy controls together[372]. Unfortunately, there is a lack a research into the diagnostic potential of these individual miRNAs, as most studies focus on the large-scale screening of miRNAs and utilize complex modelling to narrow down their validation cohort to the most statistically relevant biomarkers.

On closer examination of the literature around these frequently examined biomarkers, it is clear that no one biomarker produces highly accurate diagnostic results alone. Indeed, the evidence would suggest that the primary utility of all of these

biomarkers can be found in their use within multi-biomarker panels. While individually each of these biomarkers has their limitations, it is evident that when put together they can account for the weaknesses of the others to improve the end results. Furthermore, the addition of CA19-9 stands out as a clear prerequisite for the design of future multi-biomarker panels. These novel candidates provide a glimpse into the promising future of PDAC diagnostic biomarker discovery, though they remain to be examined within cohorts of patients with various underlying conditions and comorbidities that may influence their performance. Importantly, while blood-based biomarkers in the PDAC setting are likely to be used primarily as companion diagnostics, several of these biomarkers may also prove useful in the risk stratification of pancreatic patients with underlying conditions, given their dysregulation across certain control cohorts as outlined in this study.

3.5.5 The state of current PC research

This review highlights the variability in data quality and study design across PC research. Here, studies which employ biomarkers for the diagnosis of PDAC have been interrogated, identifying many studies that fail to provide sufficient information regarding their patient cohorts, their experimental design or their index test of interest. A substantial number of papers fail to report on the subtype of PC examined, simply conflating all subtypes as PC. For the purposes of the meta-analysis, papers that do not specify PDAC as the subtype of interest were excluded so as to reduce confounding variables. However, this lack of detail is a major flaw within many PC studies, where the specific subtype examined should always be clearly indicated.

Furthermore, almost half of the included biomarker entries did not have information regarding the reference standard used to diagnose patients with PDAC. In these cases, it was unclear whether all patients in this cohort had been diagnosed using the same reference standard or not, resulting in high levels of bias amongst these papers. A third of the biomarkers examined had no details attributed to them regarding the stage details of the PDAC cohort, with reporting of sex and age breakdowns in this cohort also poor. Control cohorts had similar issues, with high numbers of biomarkers also lacking sex and age information.

Unfortunately, the number of studies examining arguably the most clinically relevant control cohort (mixed) is extremely low compared with healthy alone and benign alone. While this review has identified many studies evaluating various types of biomarkers for PDAC diagnosis, the lack of studies conducted in clinically relevant cohorts may be the reason for the unfortunate lack of biomarkers currently in clinical use. Blinding of studies was also extremely poor, with very few opting to adopt this strategy for biomarker identification. This has further contributed to the high levels of bias observed across the studies included in this review.

Finally, evaluation of biomarker efficacy was extremely flawed in some cases, with a substantial number of biomarkers being attributed only with a p-value and no qualitative assessment (e.g., AUC or sensitivity) of the biomarker. Overall, huge flaws exist in current PC research in the context of identification of biomarkers for PDAC diagnosis. High levels of bias can be seen in many studies, with missing or unclear information regarding key study design points further compounding these issues. These are major flaws which recur again and again in the literature and could be contributing to the lack of repeated examination of high performing biomarkers in follow-up studies and could subsequently be responsible for the poor progress seen in this field in recent years.

3.5.6 Limitations of this systematic review and meta-analysis

As modern vernaculars regard PC and its PDAC subtype to be synonymous, any paper that did not specify an alternative subtype of PC was included in the extraction stage of this study and assumed to be PDAC. While a small minority of the total included studies make up this population, it is important to note that the inclusion of these data may not be appropriate in some cases as PDAC may not have been the subtype of PC examined. Furthermore, it is important to note that recent studies have highlighted the issue of potential sample contamination by endocrine, exocrine, acinar or even normal cells, which have been shown to affect the subtype classification[373]. This is an issue across all pancreatic literature and as such would persist throughout this review.

CA19-9 is highlighted in this review as the current FDA-approved biomarker for PDAC diagnosis; however, CA19-9 cut-off values were not standardized across all studies included in the review. As such, all CA19-9 entries may not have resulted from the same cut-off value and it may not be appropriate to compare them directly, as changes in CA19-9 cut-off values have been demonstrated to improve biomarker robustness[192]. A major caveat of this review, which results from the nature of the data extraction, is that certain biomarkers or biomarker panels may arise several times from a single study, having been examined in multiple patient cohorts within that study, for example, in the context of model training and validation. Unfortunately, as in many studies, there can be overlap between the patients recruited for the training and validation cohorts, resulting in repeated sampling from the same patients. The within-study variance has been controlled for in the multivariate meta-analysis; however, repeated sampling from the same patients was not accounted for and may introduce a level of bias toward some biomarkers. Furthermore, in many instances, studies have opted to evaluate several single biomarkers and subsequently combine these biomarkers to form a multi-biomarker panel. Some multi-biomarker panels have also been examined in some studies both alone and with the addition of CA19-9 to the panel. Possible bias due to repeated entries is an important limitation of this study, which could not be avoided due to the nature of current research papers and study designs. Importantly, while the QUADAS-2 tool that was used to assess study quality and RoB has been used previously for similar systematic reviews of diagnostic biomarkers[374], there may be other forms of bias introduced by these studies that were not accounted for in this assessment.

In summary, blood-based multi-biomarker panels for the diagnosis of PDAC exhibit superior performance in comparison with single biomarkers, in both CA19-9–containing biomarkers and novel biomarkers, and across all patient control cohorts. CA19-9 shows little utility alone, as it is less effective in mixed control cohorts, though when used in combination with a panel of multiple biomarkers these CA19-9–containing panels produce a better diagnostic performance than novel multi-biomarker panels. These results suggest that future biomarker studies for PDAC diagnosis should focus on the identification of a multi-biomarker panel which includes CA19-9, while drawing from the

pool of promising novel biomarkers that have been identified and examined across several different studies.

Chapter 4.

Multi-omic profiling of pancreatic cyst fluid for the identification of novel biomarkers of patient pancreatic cancer risk

4.1 Introduction

According to the American Cancer Society, PC will have the worst 5-year survival rate of any cancer in 2023, at just 12%[1]. Indeed, in 2023, the estimated number of new cases of PC in the United States is 64,050, with 50,550 expected deaths. While the 5-year survival rate has increased since the 11% announced in 2022, this minor improvement shows the lack of progress in PC research, and the urgent need for drastic changes in PC detection and management[375]. The poor rates of survival seen in PC are attributed mainly to the late detection of the disease, with vague symptoms such as weight loss or abdominal pain, causing patients to present to their GP only when their symptoms become severe, and therefore when the disease is at a late stage of development[299]. The late-stage diagnosis of these patients limits their treatment options, further compounding the problem. Early detection of PC, therefore, is the primary concern of most PC research, as it has the potential to make a substantial difference to the treatment and survival of these patients.

PCLs are fluid-filled sacs, on or inside the pancreas, that have the potential to become premalignant[121]. While some PCLs are completely benign, others have been shown to have malignant potential and could therefore play a role in the progression to PC[121]. While this is known, the issue arises in distinguishing which PCLs are benign and which are premalignant and should, as such, be monitored and/or treated accordingly. At present, there are several sets of clinical guidelines worldwide for the stratification of PCLs into risk groups based on their clinical presentation[376, 377]. Factors such as size of the PCL, presence or absence of a solid nodule, and location in/on the pancreas are assessed in order to classify PCLs as low- or high-risk for PC development. Unfortunately, the presence of several sets of guidelines worldwide indicates the lack of consensus among clinicians as to the cutoffs or defined parameters for these factors. PCLs can generally be grouped into two categories, mucinous cystic neoplasms, which have malignant potential, or non-mucinous cystic neoplasms, which are most often benign[378]. As such, detection of mucinous cysts is often among the first steps in the diagnostic cascade. Presently, the measurement of CEA as part of the aspiration of PCLs via endoscopic ultrasound is routine, with levels above the threshold of ≥192 mg/dL indicating mucinous cysts[163]. Unfortunately, a 2015 multicenter study evaluating CEA

demonstrated a sensitivity of 61% and a specificity of 77% for mucinous cyst identification, indicating that 39% of mucinous cysts would be misdiagnosed at this CEA threshold[378]. More recently, a 2021 systematic review including data from 609 PCLs reported a sensitivity of 56% and a specificity of 96% for CEA, with a substantial 44% of mucinous cysts being misdiagnosed[163]. CA19-9 is the only FDA-approved biomarker for the diagnosis of PC, however, its utility in this setting is extremely poor, particularly in patients with underlying conditions such as diabetes or pancreatitis[167]. As such, the risk stratification of these patients is inaccurate, and could therefore be contributing to the overall problem of early detection. Furthermore, the rate of PCL incidence is increasing each year as the advent of higher resolution imaging allows the incidental capture of more PCLs[163]. The identification of novel, robust biomarkers for the early detection of PC risk is urgently needed for these patients, and could provide a much needed change to the way in which PCLs are managed, enabling the discovery of high-risk patients at earlier stages of PC development.

In this study, biomarkers that were identified in Chapter 3 through a review of over 40 years of PC literature as being promising potential diagnostic biomarkers for PC, were examined in an independent patient cohort. These blood-based biomarkers were assessed for their utility in PCF as potential biomarkers of PC risk stratification. Building on the results of the previous chapter, where multi-biomarker panels were shown to produce the best results, and proteomic- and transcriptomic-based biomarkers made up the majority of 'promising biomarkers' highlighted across the literature, the proteome and transcriptome of patient PCF were profiled to identify differentially expressed proteins and miRNA. These factors were then examined alone, and as a multi-omic panel, to assess whether panels that encompassed multiple cellular levels of complexity could better control for the dysregulation of one factor, facilitating improved performance of a biomarker panel.

4.2 Hypothesis

There are factors within PCF which are differentially expressed between low- and high-risk patients, and these factors can be used to stratify patients based on their risk of PC.

4.3 Specific Aims

1. Examine the performance of biomarkers that were previously identified in the literature in a novel patient cohort.
2. Profile the proteome of PCF and identify proteins that are differentially expressed between low- and high-risk patients.
3. Profile the transcriptome of PCF and identify miRNA that are differentially expressed between low- and high-risk patients.
4. Examine the utility of differentially expressed proteins and miRNA as biomarkers of PC risk, both alone and integrated into a multi-omic panel.
5. Validate these results using different techniques in the same patient cohort.

4.4 Experimental design

4.4.1 Patient demographic information for PCF cohort

Three different patient cohorts were examined as part of this study. Demographic and clinical information for patients in the proteomic cohort, the transcriptomic cohort and the multi-omic cohort are given in Table 4.1. Risk classifications for these patients were assigned as per the 2018 European evidence-based guidelines for pancreatic cystic neoplasms [157]. Details on specific clinical criteria used to determine these classifications for each patient are unavailable, and as such, this lack of information should be noted as a limitation to the risk classifications of these patients.

4.4.2 Proteomic profiling of PCF

PCF samples were sonicated prior to use, as described in section 2.2.6. Proteomic examination of patient PCF was carried out as per section 2.2.7. The volume of sample required to obtain 50 µg of protein for each individual PCF sample was determined using the BCA assay results from section 2.2.7.1, and this volume was the starting point for each sample in the workflow.

4.4.3 Transcriptomic profiling of PCF

Transcriptomic evaluation of patient PCF was carried out as per section 2.2.10. HTG EdgeSeq Biofluid Lysis Buffer was used to lyse PCF samples prior to sequencing.

4.4.4 Analysis of proteomic and transcriptomic PCF data

Proteomic and transcriptomic data obtained for PCF were processed, analysed and subsequently scaled and integrated using the methodologies outlined in section 2.2.21.1.

Table 4.1 Patient demographic information for PCF cohorts.

	Proteomics		Transcriptomics		Multi-omic	
	Low Risk	**High Risk**	**Low Risk**	**High Risk**	**Low Risk**	**High Risk**
No of patients (M/F)	17 (7/10)	15 (6/9)	15 (5/10)	15 (8/8)	12 (5/7)	12 (5/7)
Mean age (range)	56 (22-82)	71 (39-85)	56 (22-80)	71 (39-85)	53 (22-80)	71 (39-85)
Smoking habits						
None	11	7	10	8	8	6
Active	3	1	2	1	2	1
Ex-smoker	2	5	3	5	2	4
Not known	1	2	0	1	0	1
Alcohol consumption						
None	8	9	8	9	6	8
Active	6	3	6	3	5	3
Heavy	0	0	0	1	0	0
Abstinent (ex-heavy)	2	0	1	0	1	0
Not known	1	2	0	2	0	1
Diabetic	1	1	1	1	1	1
Pancreatitis	1	0	1	0	1	0
Von Hippel-Lindau	0	1	0	1	0	1

Cohorts are divided into low- and high-risk groups. Mean age is rounded to the nearest whole number.

4.4.5 Quantification of CA19-9 in patient serum by sandwich ELISA

Concentrations of soluble CA19-9 in patient serum samples were measured via sandwich ELISA as per section 2.2.9. Samples were diluted both 1:2 with sample diluent to ensure capture within the standard curve.

4.4.6 Quantification of CEA in patient PCF

Concentrations of soluble CEA in patient PCF samples were measured clinically by the cytology team on the day of sample collection.

4.4.7 Quantification of soluble LCN2, REG1A, PIGR and PRSS8 by sandwich ELISA

Concentrations of soluble LCN2, REG1A, PIGR and PRSS8 in patient PCF and matched serum samples were measured via sandwich ELISA as per section 2.2.9. PCF was sonicated prior to use as per section 2.2.6. Samples were diluted both 1:20 and 1:40 with sample diluent to ensure capture within the standard curve.

4.4.8 QIAGEN qPCR custom microarray for PCF

The validation of transcriptomic results was carried out using QIAGEN RT-qPCR custom microarrays. Assay setup was carried out as per section 2.2.11. PCF was sonicated prior to RNA isolation as per section 2.2.6. Custom QIAGEN pre-coated microarray plates were used, evaluating 18 targets and 6 quality controls as listed in Table 4.2. Targets of interest within the PCF were SNORA66, miR-216a-5p and miR-216b-5p. Plates were prepared fresh and run on the same day. A sample layout and plating map for this workflow can be seen in (Appendix 3).

Table 4.2 QIAGEN 24-array custom plate targets and quality controls.

miRNA targets			miRNA Quality controls
SNORA66*	miR-6741-5p	miR-6727-5p	UniSp2
miR-216a-5p*	miR-3180-3p	miR-2861	UniSp3
miR-216b-5p*	miR-3180	miR-375-3p	UniSp4
miR-3197	miR-6782-5p	miR-500b-3p	UniSp6
miR-1237-5p	miR-1207-5p	miR-532-5p	miR-451a
miR-197-5p	miR-1908-5p	miR-130a-5p	miR-23a-3p

Targets of interest for this chapter are denoted with *.

4.5 Results

4.5.1 Four promising biomarkers are significantly increased in high-risk PCF

Promising biomarkers that were identified in Chapter 3 as being repeatedly examined in the literature, and that demonstrated good sensitivity and specificity for PC diagnosis, were interrogated in this cohort within the PCF. From this list, five proteins were identified via label-free proteomics (Figure 4.1A-E), and one miRNA was identified via whole transcriptome sequencing of patient PCF (Figure 4.1F). TIMP-1, MUC5AC, CRP and LAMC2 were found to be present at significantly higher levels in high-risk PCF compared to low-risk PCF (p<0.05)(Figure 4.1A-D). The levels of LRG1 and miR-21-3p were not significantly different between the low- and high-risk groups (p>0.05)(Figure 4.1E-F).

The four proteins that were significantly different between the low- and high-risk groups (TIMP-1, MUC5AC, CRP and LAMC2) were then used to stratify patients based on their expression of these proteins. Unsupervised hierarchical clustering of patients into risk groups using these proteins was performed with an accuracy of 58.3% (Figure 4.2). TIMP-1 and LAMC2 expression levels were most related, with MUC5AC showing little relatedness to the levels of the other three proteins. This is not unexpected, given the large amount of spread within the MUC5AC data in both the low- and high-risk patients (Figure 4.1B). Overall, these four blood-based biomarkers for the diagnosis of PC performed poorly in the risk stratification of patients based on their expression levels within the PCF.

4.5.2 Eight novel proteins were identified as being significantly increased in high-risk PCF

Label-free proteomics identified 465 proteins present in the PCF samples after data clean-up. Differential expression analysis revealed eight proteins [PIGR, S100A8, REG1A, LGALS3, TCN1, LCN2, PRSS8 and MUC6] to be significantly upregulated in high-risk PCF compared to low-risk PCF (p<0.002, FDR=0.05, s0=0.1)(Figure 4.3A). Despite being differentially expressed, the distribution of the expression levels from patient-to-patient within both the low- and high-risk groups was quite large (Figure 4.3B).

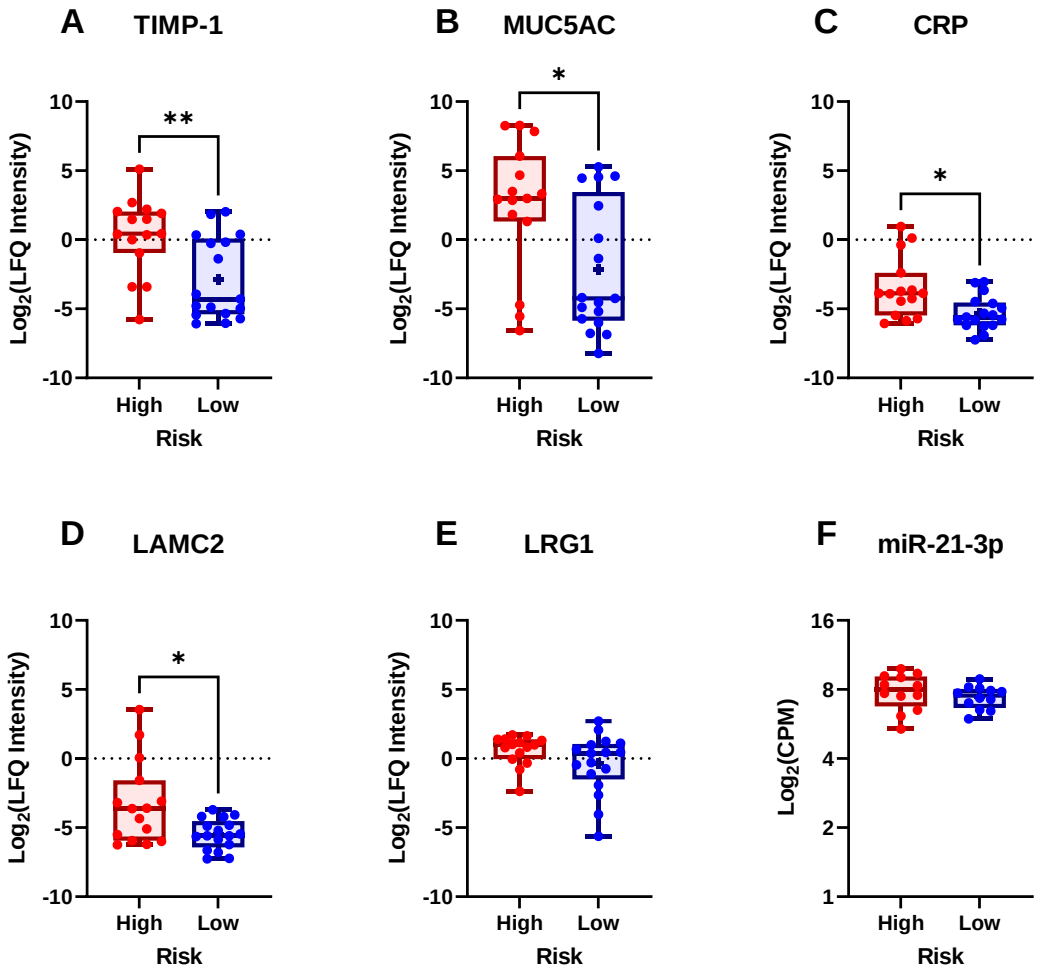


Figure 4.1 Four promising biomarkers are significantly increased in high-risk PCF compared to low-risk. (**A-E**) Boxplots showing the expression level of promising protein biomarkers, in Log_2(LFQ intensity). Data are presented as mean ± SEM for n=15 high-risk patients and n=17 low-risk patients. (**F**) Boxplot showing the expression level of a promising miRNA biomarker, in Log_2(counts per million). Data are presented as mean ± SEM for n=12 low-risk and n=12 high-risk patients. Mann-Whitney test. **p*<0.05, ***p*<0.01.

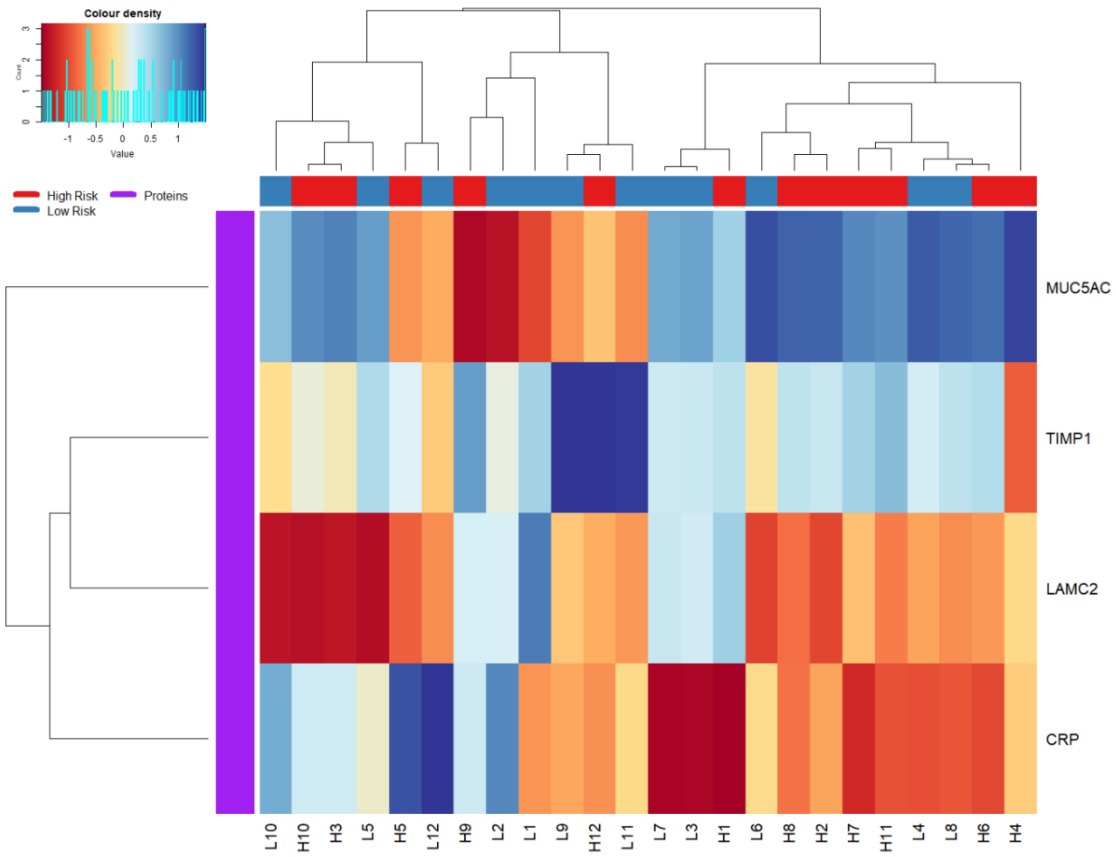


Figure 4.2 Expression of the four top occurring biomarkers in PCF clusters patients into risk groups with 58.3% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the four significant promising biomarkers. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins (purple). Data are presented for n=12 low-risk and n=12 high-risk patients.

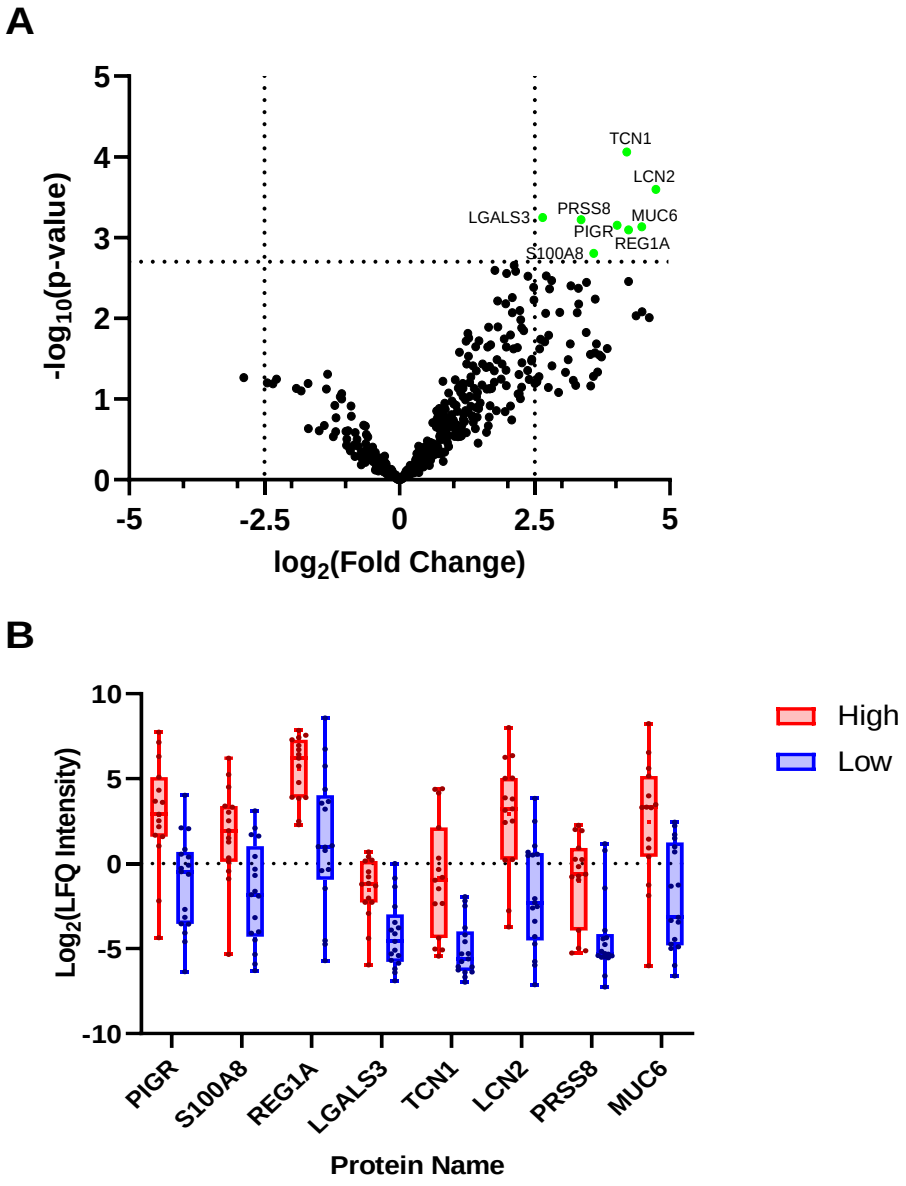


Figure 4.3 Eight proteins were identified as being significantly upregulated in high-risk patient PCF compared to low-risk. (**A**) Differential expression analysis identified eight proteins that were differentially expressed between low- and high-risk PCF samples (p<0.002, FDR=0.05, s0=0.1). Proteins identified in green were considered significantly upregulated in high-risk PCF compared to low-risk PCF. Dotted lines indicate the fold change and significance cut-offs. (**B**) Box plots showing the distribution of patient expression levels, in Log_2(LFQ intensity), for each of the eight significant proteins. High-risk patients (n=15) are shown in red; low-risk patients (n=17) are shown in blue.

As expected, spearman correlations showed that expression levels of each of the eight proteins correlated with patient risk (p<0.01)(Figure 4.4). While none of these proteins correlated with the presence of diabetes or pancreatitis, nor with smoking habits or sex, five proteins had significant positive correlations with age (p<0.05). Indeed, PIGR, REG1a, TCN1, LCN2 and MUC6 expression levels were shown to increase with increasing age. Furthermore, TCN1 levels had the strongest positive correlation out of the five (p<0.001). PIGR was the only protein to correlate with alcohol consumption, with PIGR levels significantly decreasing with increased alcohol consumption in this cohort (p<0.01). While LCN2, PRSS8 and MUC6 also had negative correlations with alcohol consumption, these were not significant (p>0.05).

Several clinical factors were also shown to correlate with each other in this patient cohort. Increased risk significantly correlated with increased age (p<0.01); increased smoking significantly correlated with increased risk of pancreatitis (p<0.01); increased smoking significantly correlated with increased alcohol consumption (p<0.05); and increased smoking significantly correlated with sex, or given the binary coding for this, increased smoking significantly correlated with being male (p<0.01).

4.5.3 8-protein panel in PCF stratifies patients with modest accuracy

These eight proteins were then scaled and integrated to create an 8-protein biomarker panel. This 8-protein panel was then used to stratify patients into distinct groups based on their expression of these proteins. Unsupervised hierarchical clustering of patients into risk groups using this 8-protein panel was performed with an accuracy of 81.25%, with just n=3 low-risk and n=3 high-risk patients being grouped incorrectly (Figure 4.5). Expression levels of most proteins are highly related, with S100A8 being the most unrelated.

PCA was conducted using two components as this was found to account for 72.6% of the variance (61.8% in component 1 and 10.8% in component 2) (Appendix 4) (Figure 4.6). LCN2, MUC6 and TCN1 were the most important contributors to the first component, with S100A8 being the most important contributor to the second component. REG1A and PRSS8, and well as TCN1 and MUC6 were highly positively

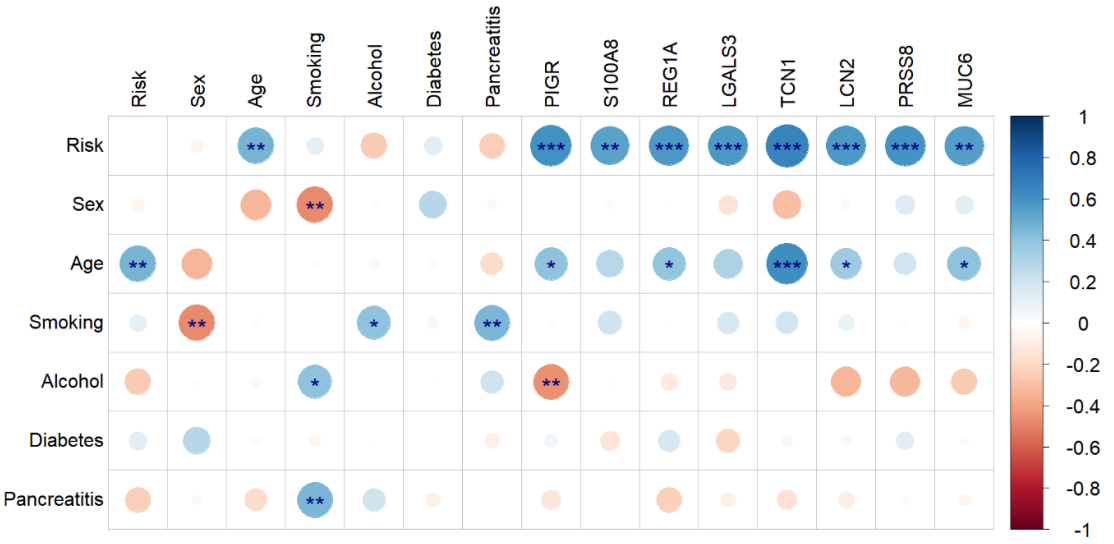


Figure 4.4 Expression of significant proteins in PCF significantly correlates with several clinical factors. Spearman correlations between patient clinical data and the eight differentially expressed proteins are given as a corrplot. Colour intensity relates to R^2 value; circle size relates to the p-value. Data are presented for n=17 low-risk and n=15 high-risk patients. **p*<0.05, ***p*<0.01, ****p*<0.001.

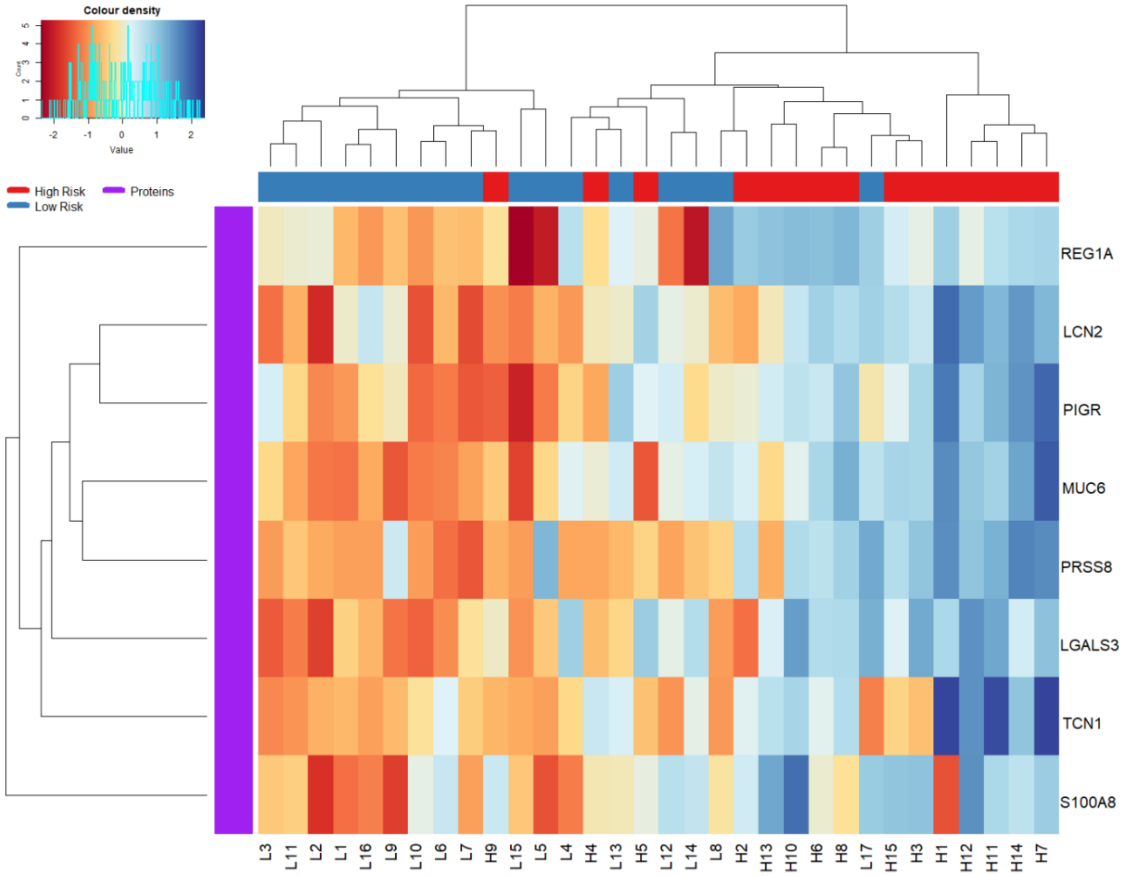


Figure 4.5 8-protein panel in PCF clusters patients into risk groups with 81.25% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the eight significant proteins. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins (purple). Data are presented for n=17 low-risk and n=15 high-risk patients.

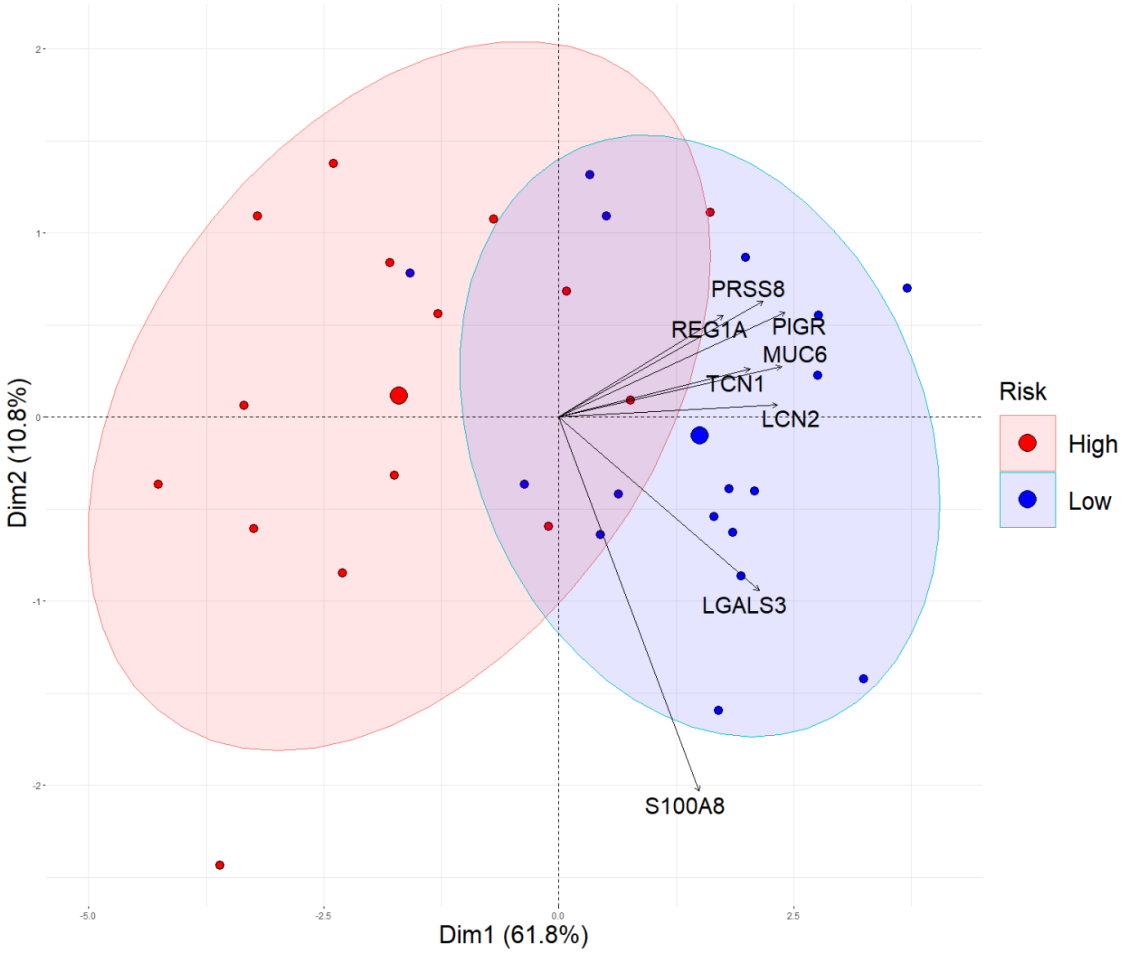


Figure 4.6 PCA analysis of 8-protein panel in PCF shows moderate distinction of the risk groups. 2-D PCA using the eight differentially expressed proteins, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding protein. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. Data are presented for n=17 low-risk (blue) and n=15 high-risk (red) patients.

correlated. S100A8 was the most uncorrelated with the other proteins, as was shown to be the case in the unsupervised hierarchical clustering (Figure 4.5), and it also had the largest amount of variance corresponding to it. REG1A, however, had the smallest amount of variance associated with it. The ellipse encapsulating 80% of the low-risk classification was much smaller than that of the high-risk classification, indicating larger variance in the high-risk population.

LOOCV was then carried out, where the model was trained and then tested using the same patient cohort. Here, the 8-protein panel produced an AUC value of 0.608, with a sensitivity of 70.6% and a specificity of 60%, indicating that this panel performs well in its ability to assign a high-risk classification to those who are high-risk (Figure 4.7).

4.5.4 Three miRNA were identified as being significantly increased in high-risk PCF

Whole transcriptome sequencing identified 2,096 miRNA present in the PCF samples after data clean-up. Differential expression analysis revealed three miRNA [SNORA66, miR-216a-5p and miR-216b-5p] to be significantly upregulated in high-risk PCF compared to low-risk PCF (adj-p<0.05, FDR=0.05, s0=0.1)(Figure 4.8A). The distribution of the expression levels from patient-to-patient within both the low- and high-risk groups is modest, with higher variability being seen in the high-risk group compared to the low-risk (Figure 4.8B).

Importantly, spearman correlations showed that only expression levels of miR-216a-5p significantly correlated with patient risk (p<0.05), with both SNORA66 and miR-216b-5p having positive correlation also, though not significant (p>0.05)(Figure 4.9). None of these miRNA significantly correlated with any clinical parameters (p>0.05). Indeed, while all three miRNA had negative correlation with diabetes status, and positive correlations with pancreatitis status, these were not significant (p>0.05).

Several clinical factors were shown to correlate with each other in this patient cohort also. Increased risk significantly correlated with increased age (p<0.01), and increased smoking significantly correlated with increased alcohol consumption (p<0.001).

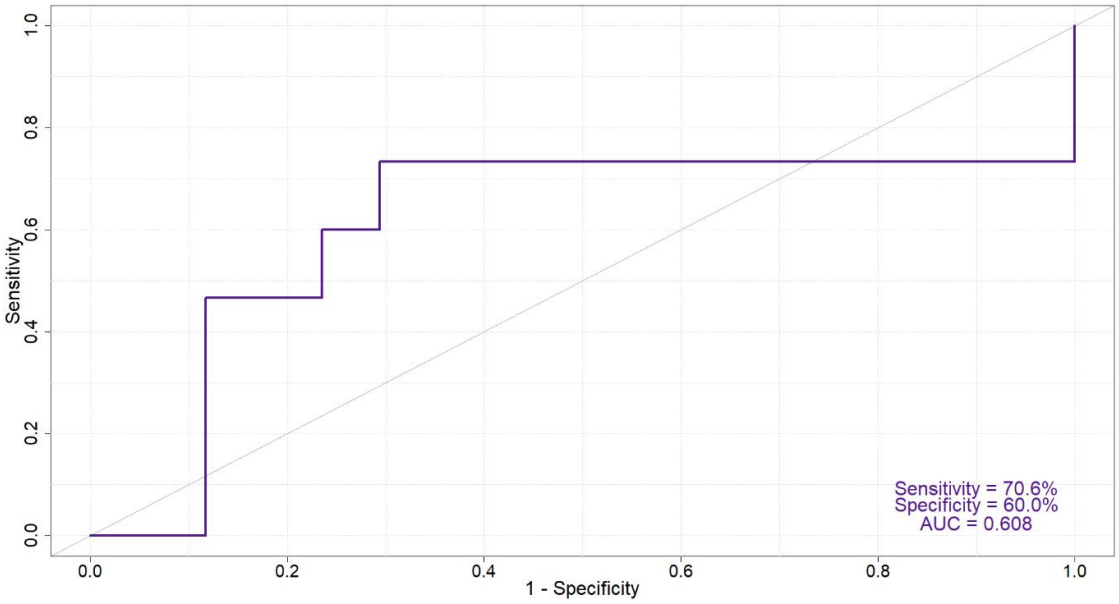


Figure 4.7 8-protein panel in PCF classifies patients based on risk with AUC of 0.608. LOOCV of patients using the eight differentially expressed proteins. Data are presented for n=17 low-risk patients and n=15 high-risk patients.

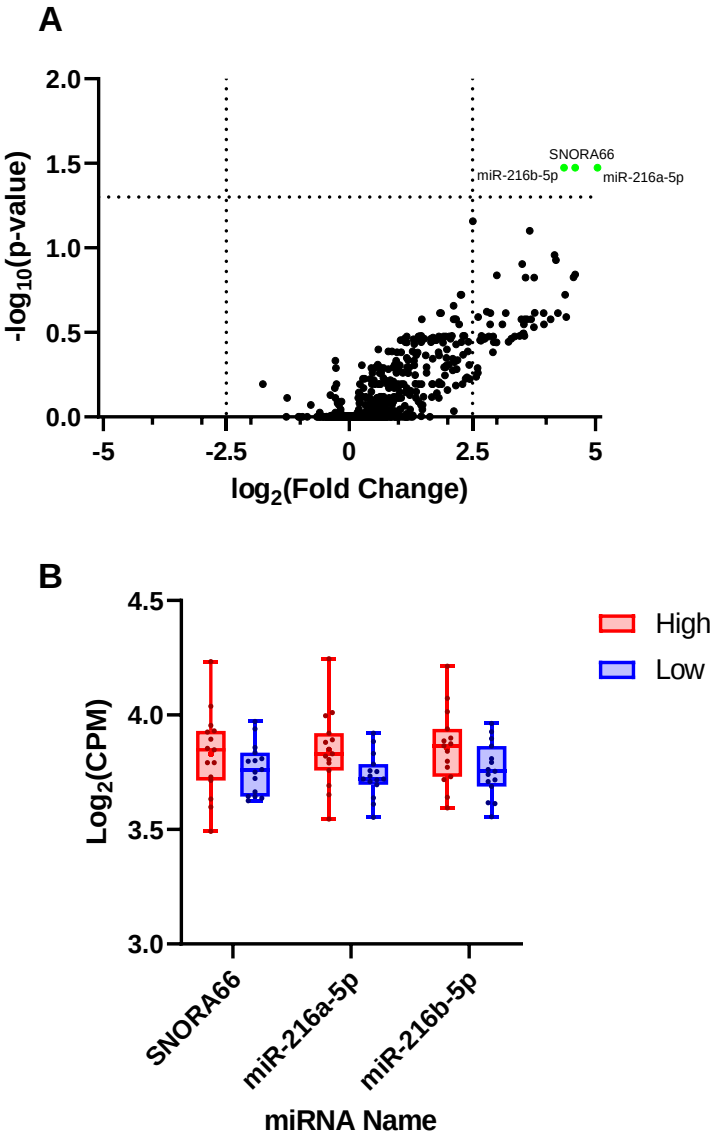


Figure 4.8 Three miRNA were identified as being significantly upregulated in high-risk patient PCF compared to low-risk. (**A**) Differential expression analysis identified three miRNA that were differentially expressed between low- and high-risk PCF samples (p<0.05, FDR=0.05, s0=0.1). MiRNA identified in green were considered significantly upregulated in high-risk PCF compared to low-risk PCF. Dotted lines indicate the fold change and significance cut-offs. (**B**) Box plots showing the distribution of patient expression levels, in Log_2(CPM), for each of the three significant miRNA. High-risk patients (n=15) are shown in red; low-risk patients (n=15) are shown in blue.

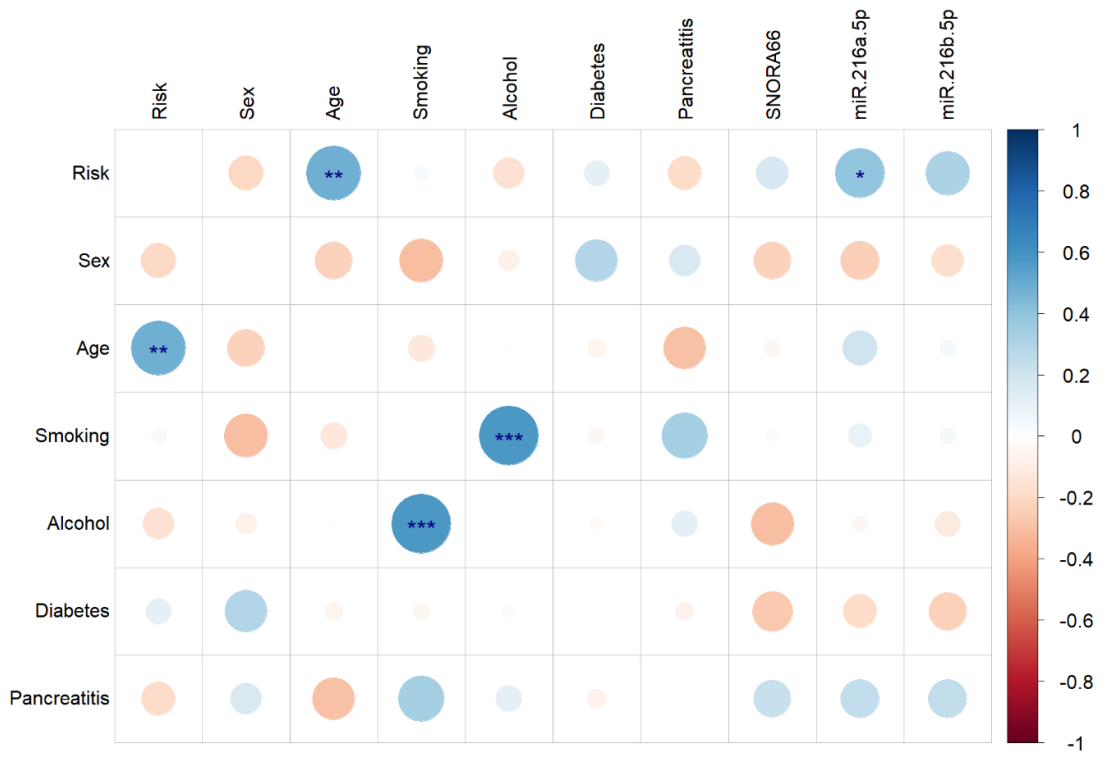


Figure 4.9 Expression of significant miRNA in PCF does not significantly correlate with clinical factors. Spearman correlations between patient clinical data and the three differentially expressed miRNA are given as a corrplot. Colour intensity relates to R^2 value; circle size relates to the p-value. Data are presented for n=15 low-risk and n=15 high-risk patients.*p<0.05, **p<0.01, ***p<0.001.

4.5.5 3-miRNA panel in PCF stratifies patients with poor accuracy

These three miRNA were then scaled and integrated to create a 3-miRNA biomarker panel. This 3-miRNA panel was then used to stratify patients into distinct groups based on their expression of these miRNA. Unsupervised hierarchical clustering of patients into risk groups using this 3-miRNA panel was performed with an accuracy of 60%, with n=6 low-risk and n=6 high-risk patients being grouped incorrectly (Figure 4.10). Expression levels of miR-216a-5p and miR-216b-5p were most related to each other, with SNORA66 being the most unrelated.

PCA was conducted using two components as this was found to account for 98.9% of the variance (93.2% in component 1 and 5.7% in component 2) (Appendix 5)(Figure 4.11). MiR-216a-5p and miR-216b-5p were the most important contributors to the first component, with SNORA66 being the most important contributor to the second component. MiR-216a-5p and miR-216b-5p were shown to be highly positively correlated, with SNORA66 having no correlation with either miR-216a-5p or miR-216b-5p, as was shown to be the case in the unsupervised hierarchical clustering (Figure 4.10). All three miRNA had a large amount of variance associated with them. The ellipse encapsulating 80% of the low-risk classification is much smaller than that of the high-risk classification, indicating larger variance in the high-risk population. Importantly, the entire low-risk ellipse sits inside that of the high-risk ellipse, showing the poor separation of the two groups. Furthermore, as miR-216a-5p and miR-216b-5p represent the first component which accounts for 93.2% of the variance of the dataset, in this analysis SNORA66 appears to be a source of variance in this panel.

LOOCV was then carried out, where the model was trained and then tested using the same patient cohort. Here, despite showing poor separation of the groups in both hierarchical clustering and PCA analysis, the 3-miRNA panel produced an AUC value of 0.658, with a sensitivity of 60% and a specificity of 53.3% (Figure 4.12).

4.5.6 11-feature multi-omic panel in PCF stratifies patients with high accuracy

The eight proteins and three miRNA that were identified as significantly differentially expressed were then scaled and integrated to create an 11-feature multi-omic

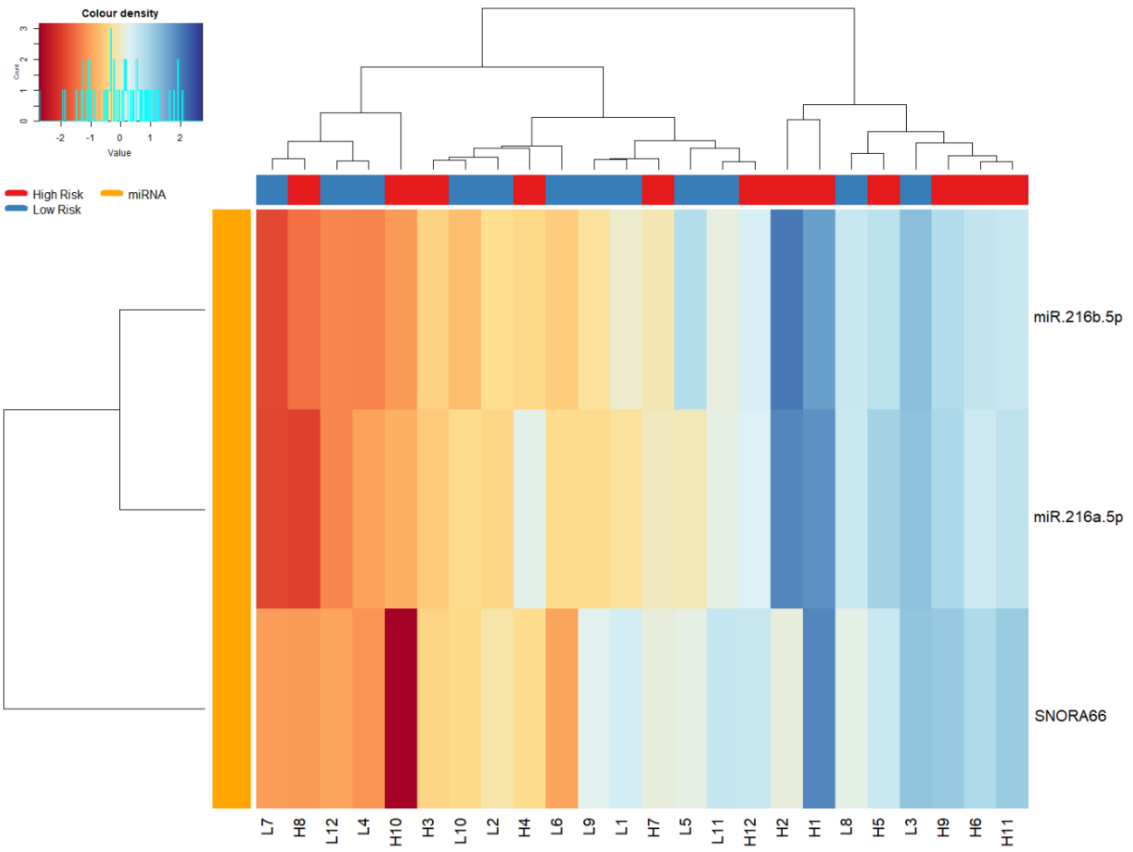


Figure 4.10 3-miRNA panel in PCF clusters patients into risk groups with 60% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the three significant miRNA. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the miRNA (orange). Data are presented for n=15 low-risk and n=15 high-risk patients.

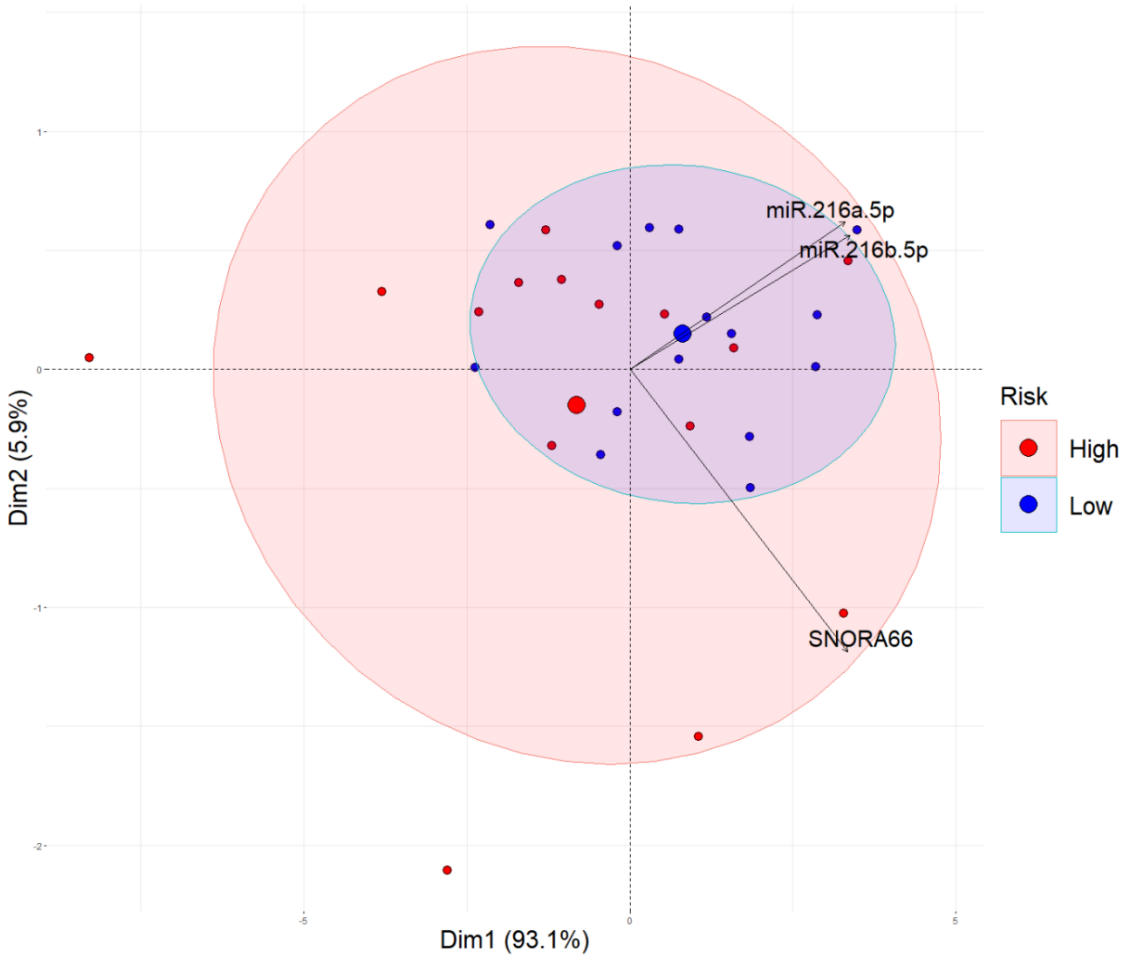


Figure 4.11 PCA analysis of 3-miRNA panel in PCF shows poor distinction of the risk groups. 2-D PCA using the three differentially expressed miRNA, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding miRNA. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. Data are presented for n=15 low-risk (blue) and n=15 high-risk (red) patients.

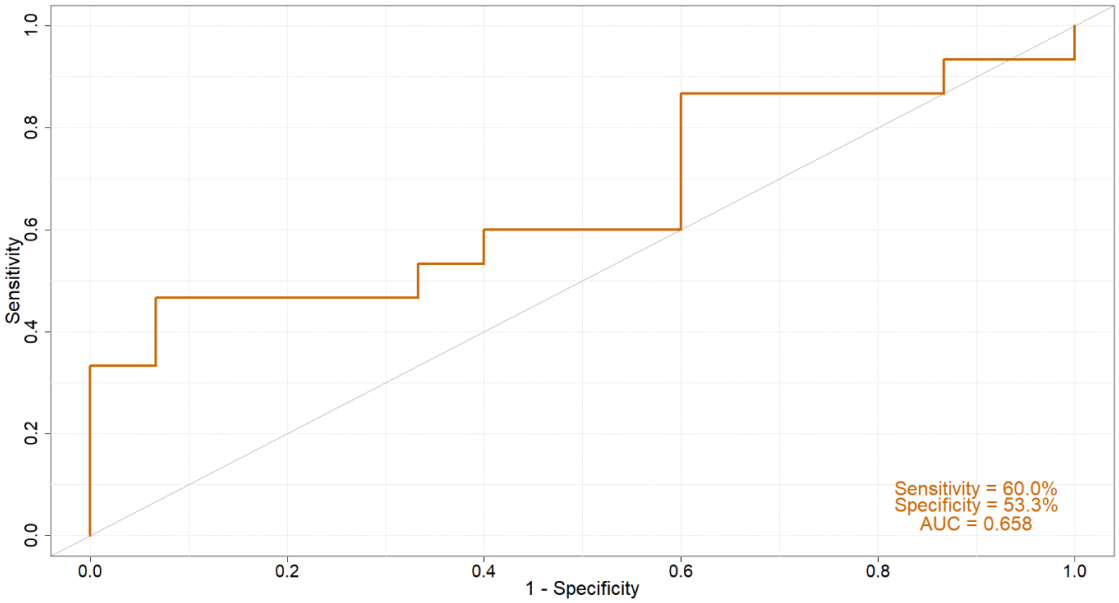


Figure 4.12 3-miRNA panel in PCF classifies patients based on risk with AUC of 0.658. LOOCV of patients using the three differentially expressed miRNA. Data are presented for n=15 low-risk patients and n=15 high-risk patients.

biomarker panel. This 11-feature panel was then used to stratify patients into distinct groups based on their expression of these eleven biomarkers. Unsupervised hierarchical clustering of patients into risk groups using this 11-feature panel was performed with an accuracy of 95.8%, with just n=1 high-risk patient being grouped incorrectly (Figure 4.13). Importantly, this outlier patient has a diagnosis of VHL disease, which genetically pre-disposes one to the development of PC. As such, this patient is classified as high-risk based on the presence of this geneticcondition. In fact, when this factor was removed from the classification process, and the patient was reclassified based on the rest of their clinical factors, this patient was reclassified as low-risk. In this case, the 11-feature panel produced 100% clustering accuracy (Figure 4.14). As expected, the clustering of the proteins and miRNA in this panel were separate, and formed the relationships that were observed in each panel individually. Examining the correlations between the eleven biomarkers within the panel further demonstrated the low correlations between the proteins and miRNA (Figure 4.15). Indeed, there were no significant correlations between any of the three miRNA and eight proteins (p>0.05). All three miRNA correlated significantly and strongly with each other (p<0.05). The correlations between S100A8 expression and the other biomarkers were shown to be poor, with S100A8 representing a smaller segment than the other biomarkers and only correlating significantly with REG1A, LGALS3 and TCN1 (p<0.05). REG1A, LGALS3 and TCN1 therefore, were the only proteins that correlated with all seven other proteins.

PCA was conducted using two components as this was found to account for 71.1% of the variance (47.8% in component 1 and 23.3% in component 2) (Appendix 6)(Figure 4.16). The eight proteins were the most important contributors to the first component, with the three miRNA being the most important contributors to the second component. TCN1 and LCN2 contributed the most to the first component, with miR-216b-5p contributing most to the second component. In this cohort, miR-216a-5p and SNORA66 were shown to be highly positively correlated, which is contrary to what was seen in the PCA analysis of the miRNA alone, highlighting the differences a minor change in patient cohort can have (Figure 4.16). All three miRNA had large amounts of variance associated with them, with SNORA66 having the least variance of the three miRNA (Figure 4.16). The proteins had less variance compared to the miRNA, with S100A8

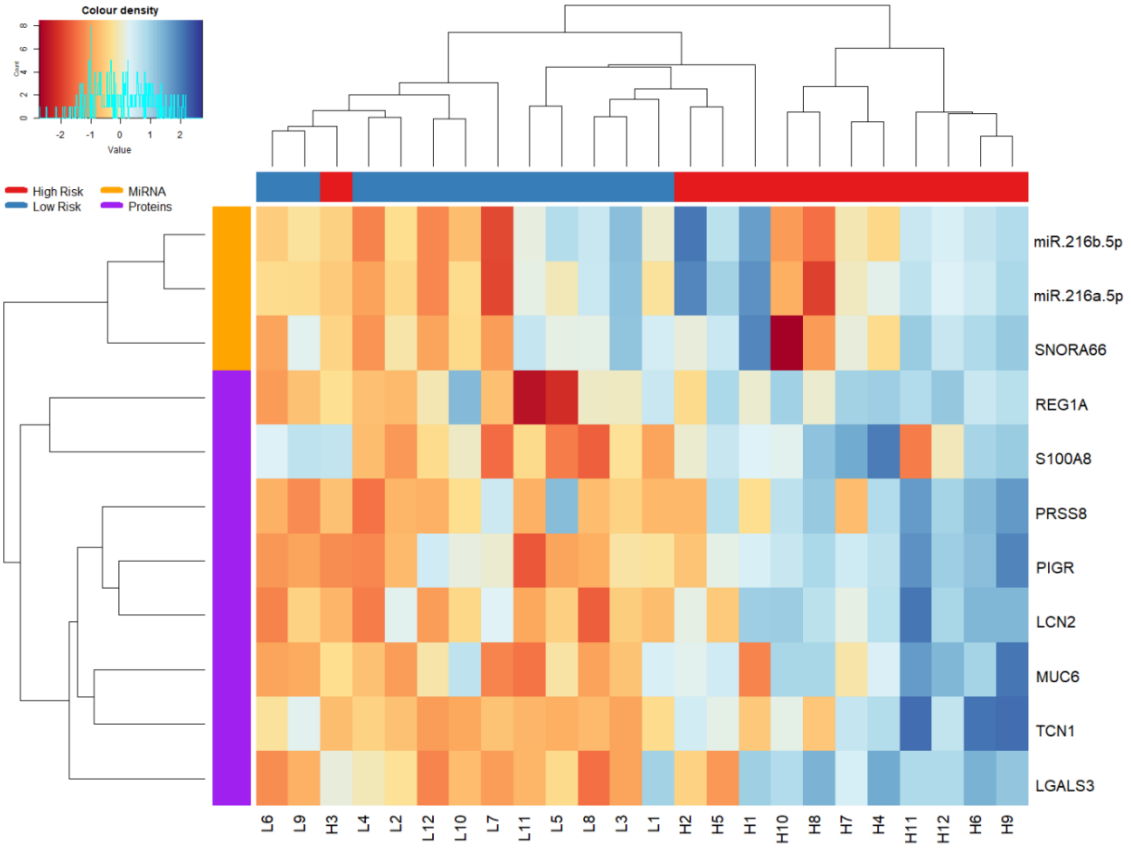


Figure 4.13 11-feature multi-omic panel in PCF clusters patients into risk groups with 95.8% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the 11-feature multi-omic panel. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins and miRNA. Data are presented for n=12 low-risk and n=12 high-risk patients.

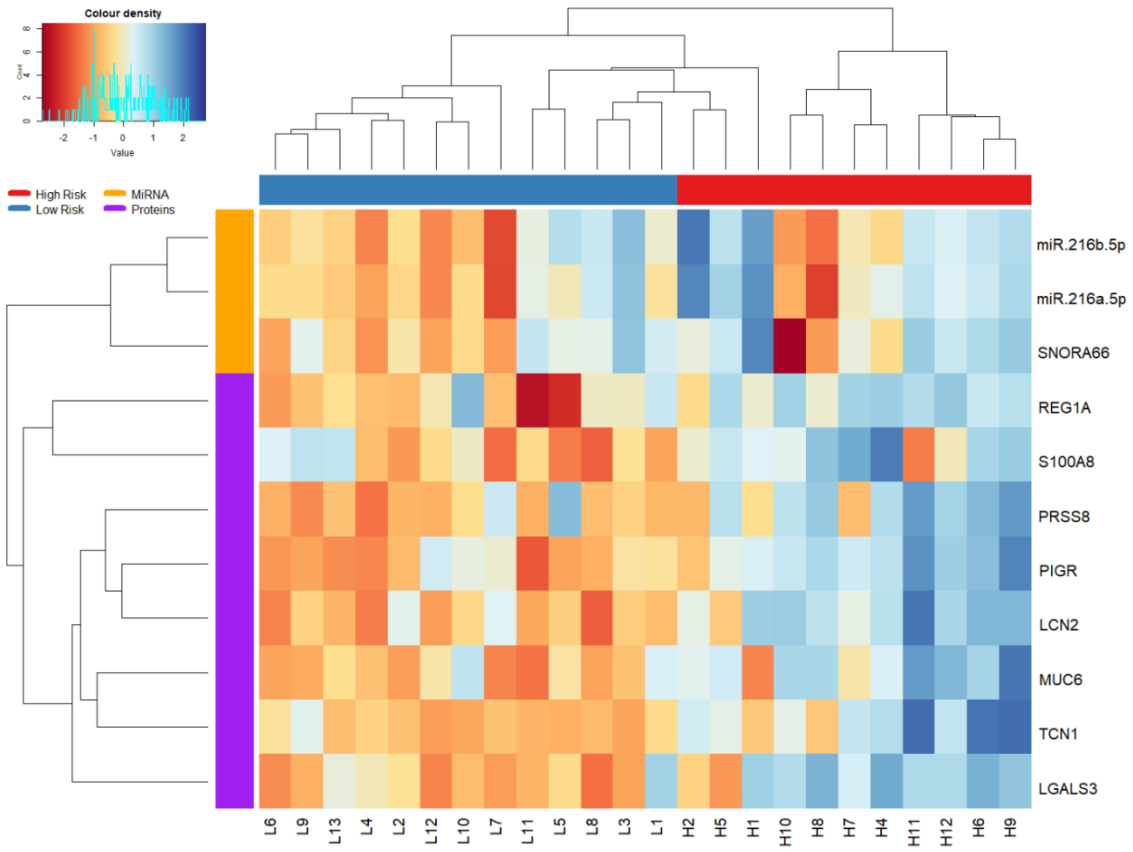


Figure 4.14 11-feature multi-omic panel in PCF clusters patients into risk groups with 100% accuracy when VHL patient is reclassified. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the 11-feature multi-omic panel. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins and miRNA. VHL outlier patient has been reclassified from high-risk to low-risk. Data are presented for n=13 low-risk and n=11 high-risk patients.

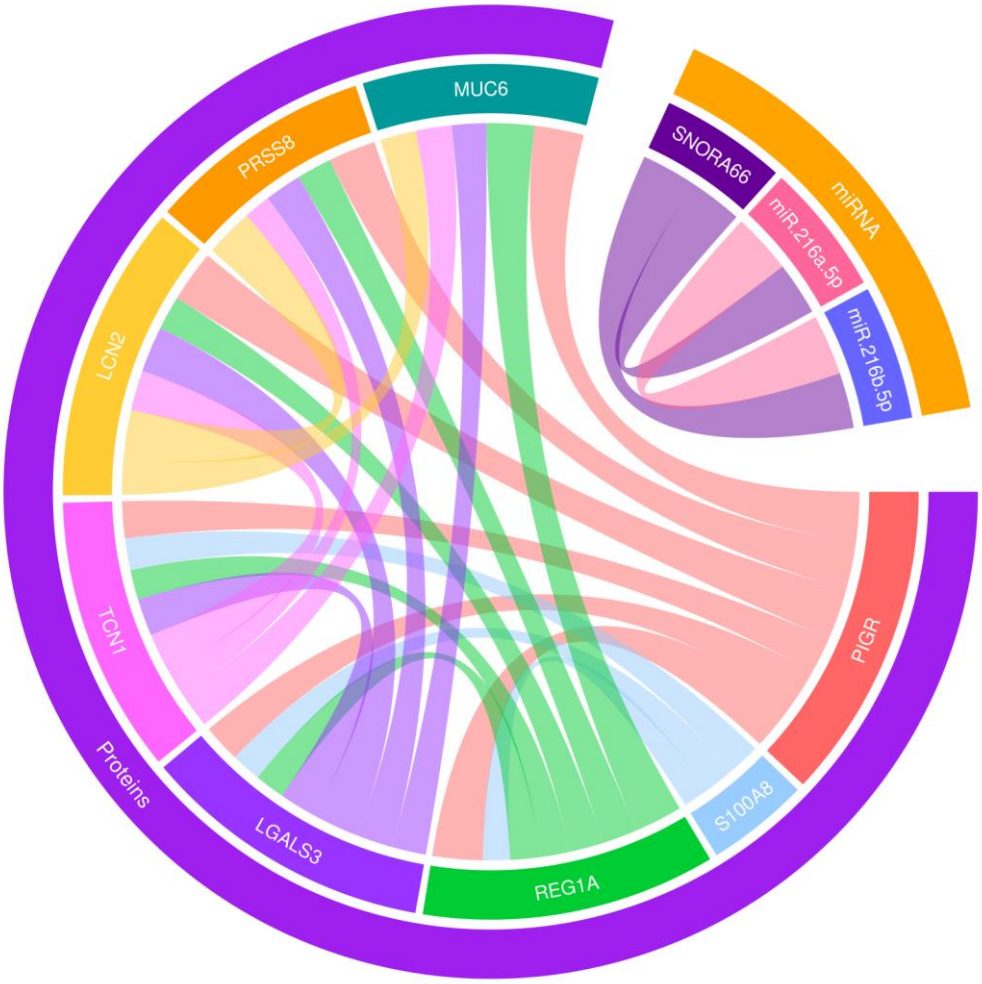


Figure 4.15 Biomarkers within the 11-feature multi-omic panel in PCF significantly positively correlate with the expression of each other. Chord diagram showing significant spearman correlations (p<0.05) of between eight proteins (purple sector) and three miRNA (orange sector) that make up the 11-feature panel. Inner chords reflect correlations between the biomarkers. Chord thickness is directly related to the strength of the correlation, with thicker chords indicating stronger correlations.

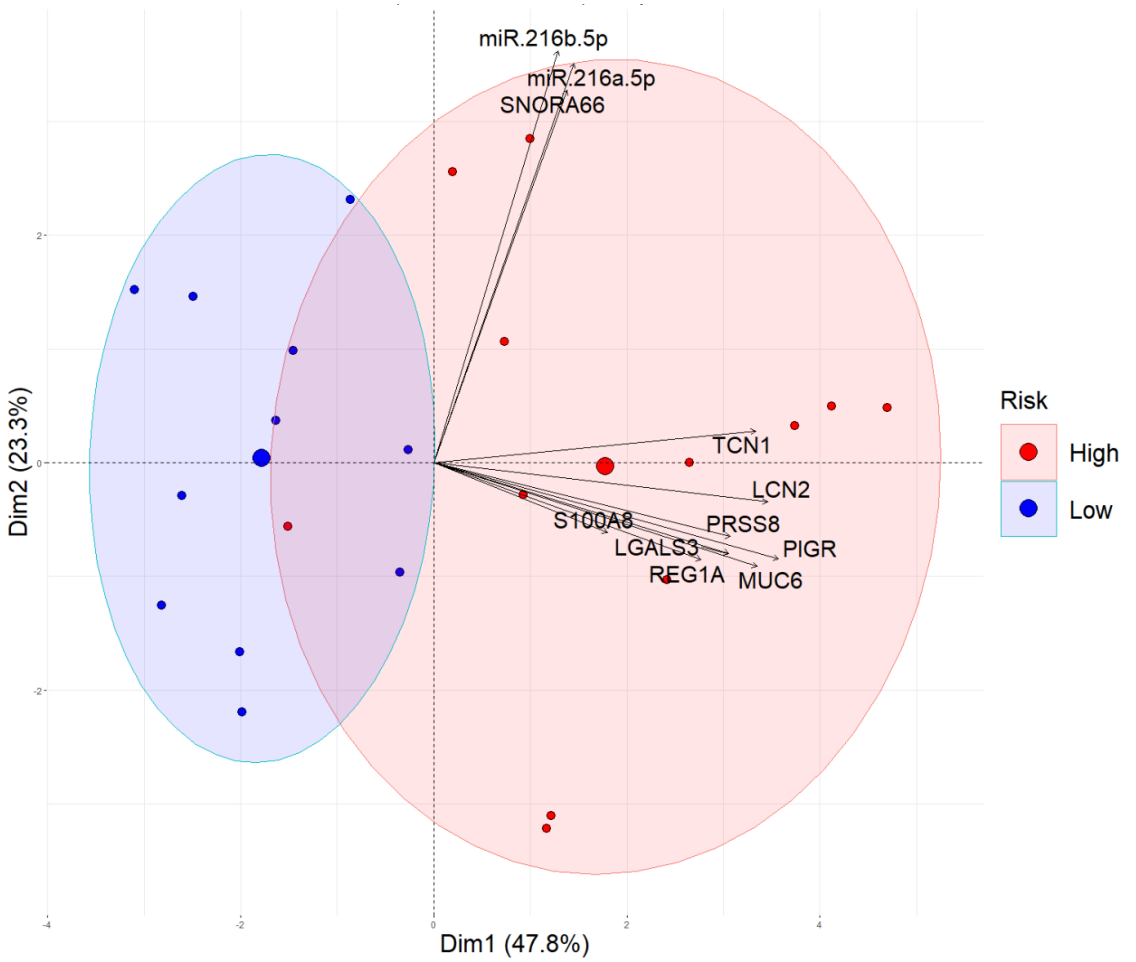


Figure 4.16 PCA analysis of 11-feature multi-omic panel in PCF shows adequate distinction of the risk groups. 2-D PCA using the 11-feature multi-omic panel, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding biomarker. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. Data are presented for n=12 low-risk (blue) and n=12 high-risk (red) patients.

having the least variance of all in the 11-feature panel. The ellipse encapsulating 80% of the low-risk classification was much smaller than that of the high-risk classification, indicating larger variance in the high-risk population again. Importantly, the single high-risk datapoint that caused the high-risk ellipse to overlap substantially with the low-risk ellipse was the VHL patient mentioned previously. When this patient was reclassified as low-risk, the separation of the low- and high-risk ellipses was greatly improved (Figure 4.17).

LOOCV was then carried out, where the model was trained and then tested using the same patient cohort. Here, the 11-feature multi-omic panel produced an AUC value of 0.806, with a sensitivity of 66.7% and a specificity of 75% (Figure 4.18). Despite the modest performances of the 3-miRNA panel and 8-protein panel alone, the integration of these two panels together improved the overall performance. However, while both the 8-protein and 3-miRNA panels had improved sensitivity compared to specificity, the 11-feature multi-omic panel had the reverse, with higher specificity compared to sensitivity. Importantly, when the VHL outlier patient was reclassified to low-risk as before, the performance of all panels improved (Figure 4.19). Indeed, the 8-protein panel improved its AUC value from 0.608 to 0.679, the 3-miRNA panel improved its AUC value from 0.658 to 0.728, and the 11-feature multi-omic panel improved its AUC value from 0.806 to 0.867. Furthermore, in the reclassified cohort, the 11-feature multi-omic panel had improved sensitivity (92.3%) compared to specificity (63.6%), further emphasising how the classification of this patient affects the performance of the panel.

4.5.7 CA19-9 does not improve the accuracy of the 11-feature multi-omic panel

The utility of CA19-9 in a multi-biomarker panel was demonstrated in Chapter 3, where it was shown to have poor diagnostic performance on its own, but when added to a multi-biomarker panel it improved the performance of the panel. As such, the CA19-9 levels of these patients were measured in the matched serum by sandwich ELISA. CA19-9 levels were not significantly different in low- or high-risk patient serum (p>0.05)(Figure 4.20A). Indeed, the concentrations of CA19-9 in all patient serum samples clustered very closely, suggesting CA19-9 has no utility in the risk stratification setting. Importantly, two high-risk patient CA19-9 levels could not be included in these data as

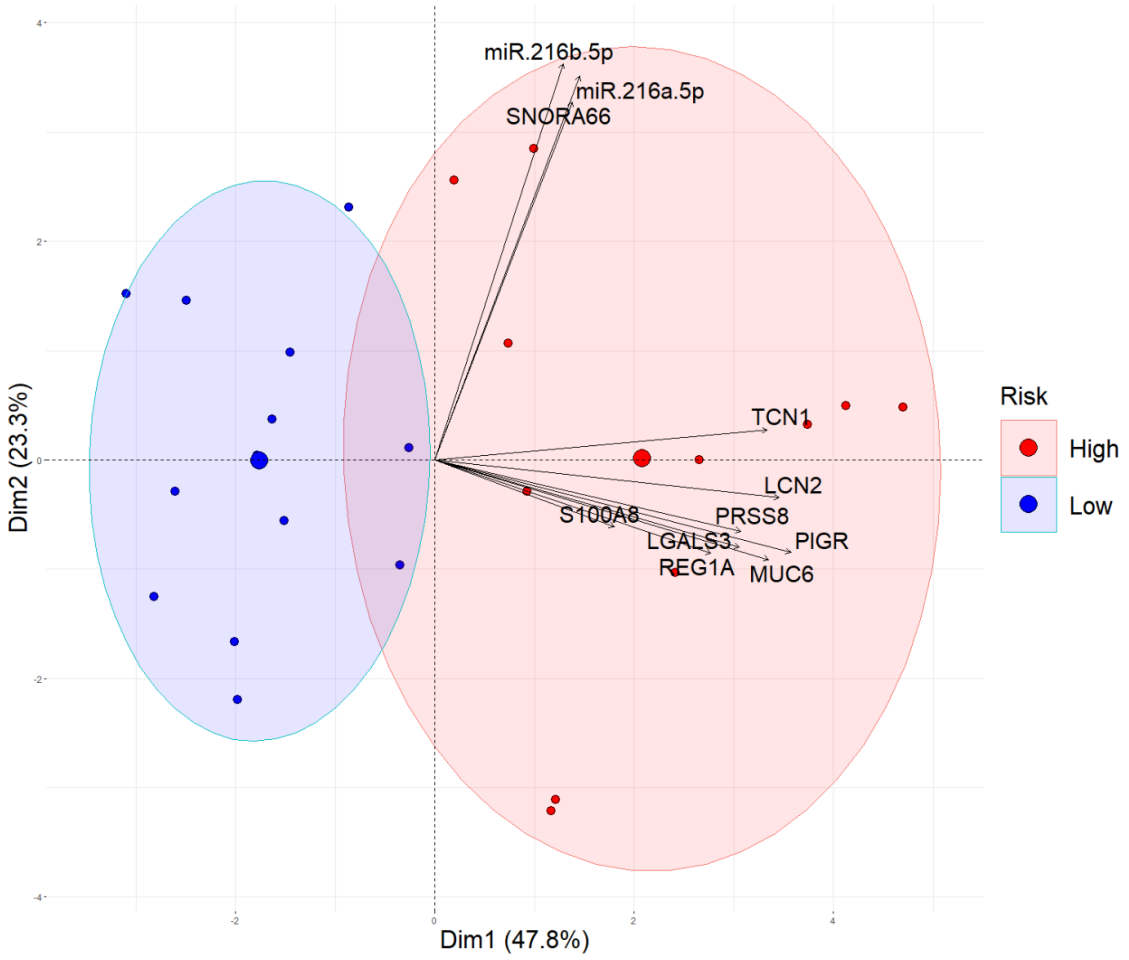


Figure 4.17 PCA analysis of 11-feature multi-omic panel in PCF shows improved distinction of the risk groups when VHL patient is reclassified. 2-D PCA using the 11-feature multi-omic panel, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding biomarker. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. VHL outlier patient has been reclassified from high-risk to low-risk. Data are presented for n=13 low-risk (blue) and n=11 high-risk (red) patients.

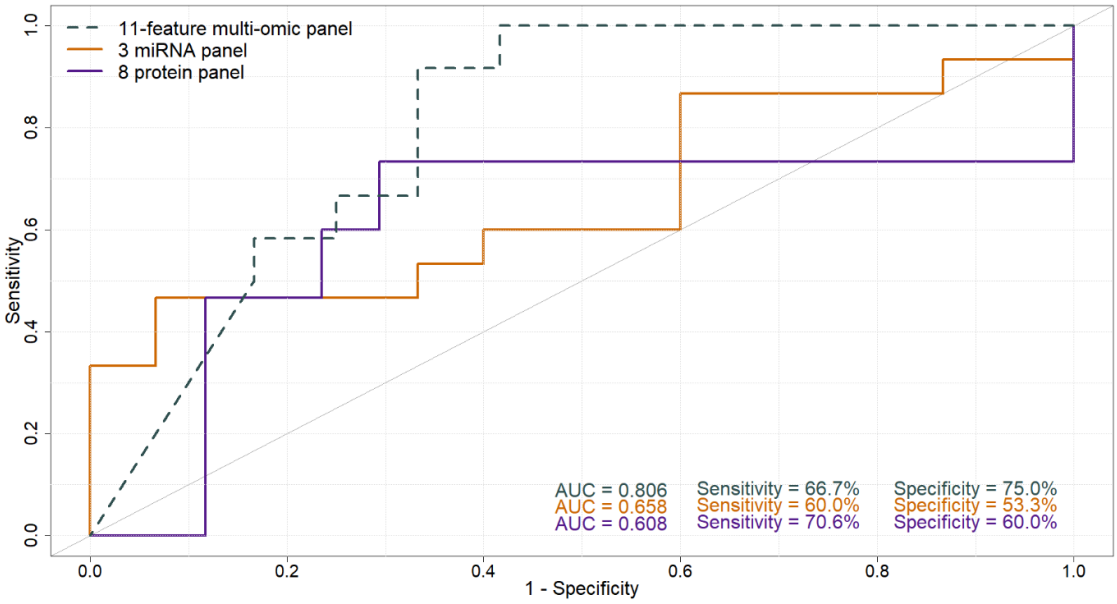


Figure 4.18 11-feature multi-omic panel in PCF classifies patients based on risk with AUC of 0.806. LOOCV of patients using the (dashed grey line) 11-feature multi-omic panel, (orange line) 3-miRNA panel, and (purple line) 8-protein panel. 11-feature panel data are presented for n=12 low-risk patients and n=12 high-risk patients; 3-miRNA panel data are presented for n=15 low-risk patients and n=15 high-risk patients; 8-protein panel data are presented for n=17 low-risk patients and n=15 high-risk patients.

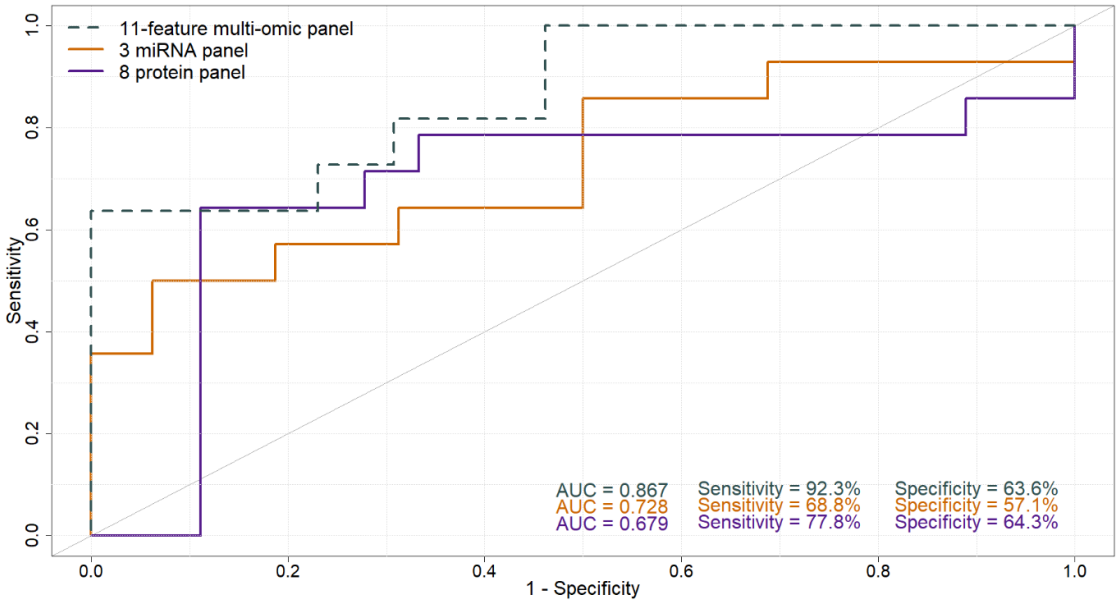


Figure 4.19 11-feature multi-omic panel in PCF classifies patients based on risk with AUC of 0.867 when VHL patient is reclassified. LOOCV of patients with VHL patient reclassified to low-risk, using the (dashed grey line) 11-feature multi-omic panel, (orange line) 3-miRNA panel, and (purple line) 8-protein panel. 11-feature panel data are presented for n=13 low-risk patients and n=11 high-risk patients; 3-miRNA panel data are presented for n=16 low-risk patients and n=14 high-risk patients; 8-protein panel data are presented for n=18 low-risk patients and n=14 high-risk patients.

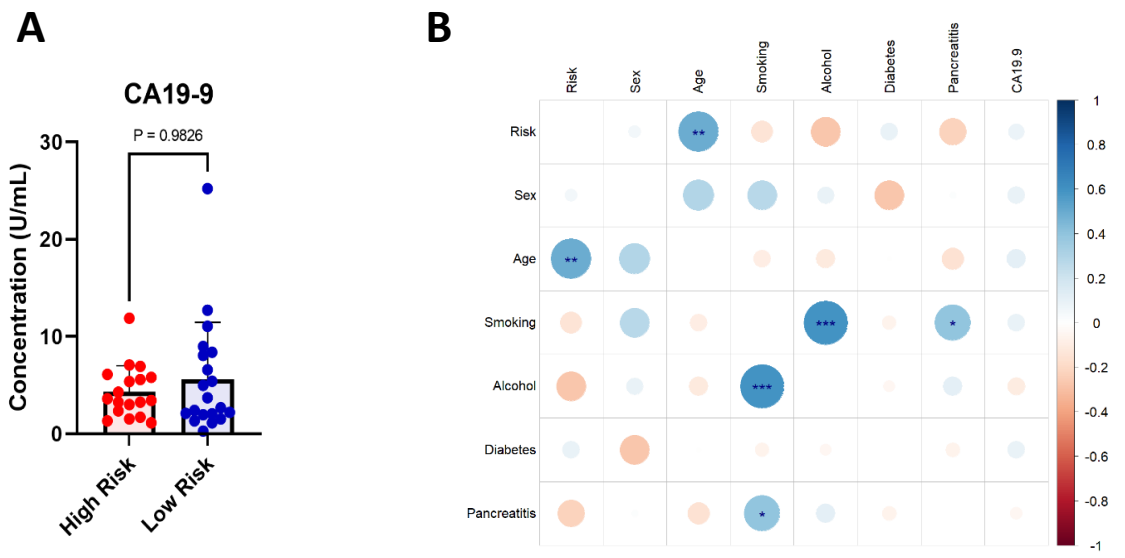


Figure 4.20 CA19-9 levels are not significantly different in patients with low- or high-risk PCLs. (**A**) Serum concentration of CA19-9 (U/mL) in high-risk (red) and low-risk (blue) patients. Mann-Whitney test. Data are presented as mean ± SEM. (**B**) Spearman correlations between patient clinical data and the serum CA19-9 levels are given as a corrplot. Colour intensity relates to R^2 value; circle size relates to the p-value. Data are presented for n=20 low-risk and n=18 high-risk patients. *p<0.05, **p<0.01, ***p<0.001.

they were above the detection range of the ELISA. These concentrations were determined to be 262.48 U/mL and 727.66 U/mL by using the equation of the curve. When included in the analysis, however, there remains no significant difference between low- and high-risk CA19-9 serum levels (p=0.6156)(data not shown).

CA19-9 levels were also shown to have no significant correlation with any clinical parameters (p>0.05)(Figure 4.20B). In this cohort, age is significantly increased with risk (p<0.01); increased alcohol consumption is also shown to increased alongside increased smoking (p<0.001); and smoking is shown to be increased in patients with pancreatitis (p<0.05).

The CA19-9 expression levels obtained were then integrated into the 11-feature multi-omic panel to investigate whether its addition would improve the performance of the panel. Given that CA19-9 levels are not altered between low- and high-risk patients, it is unsurprising that the 12-feature panel with CA19-9 has diminished ability to cluster the patients into distinct groups, demonstrating an accuracy of 58.3% (Figure 4.21).

4.5.8 CEA does not improve the accuracy of the 11-feature multi-omic panel

The potential utility of CEA in pancreatic disease was demonstrated in Chapter 3, where it was identified as a promising blood-based biomarker for PC diagnosis. As such, the CEA levels measured clinically for these patients were obtained. CEA levels were significantly increased in the PCF of high-risk patients compared to low-risk (p<0.001)(Figure 4.22A). Indeed, while the levels of CEA in high-risk PCF were generally tightly clustered, two high-risk patients had substantially higher levels of CEA than the rest of the group. CEA levels were also shown to have significant positive correlations with risk and age (p<0.001)(Figure 4.22B). In this cohort, age is significantly increased with risk (p<0.05); increased alcohol consumption is also shown to increased alongside increased smoking (p<0.05); and age is shown to be increased in male patients compared to female patients (p<0.05).

The CEA expression levels obtained were then integrated into the 11-feature multi-omic panel to investigate whether its addition would improve the performance of the panel. The addition of CEA to the 11-feature multi-omic panel resulted in no change

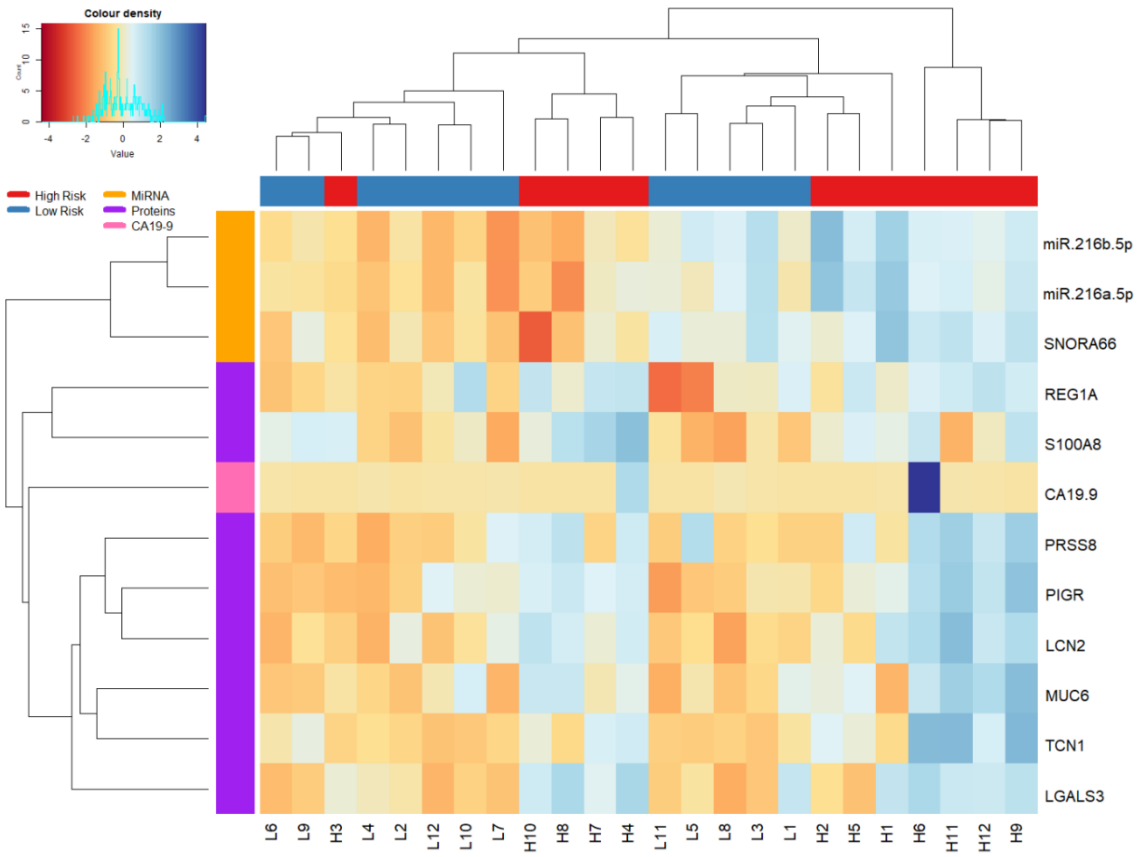


Figure 4.21 11-feature multi-omic panel in PCF plus serum CA19-9 clusters patients into risk groups with 58.3% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the 11-feature multi-omic panel in PCF plus serum CA19-9. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the biomarkers. Data are presented for n=12 low-risk and n=12 high-risk patients.

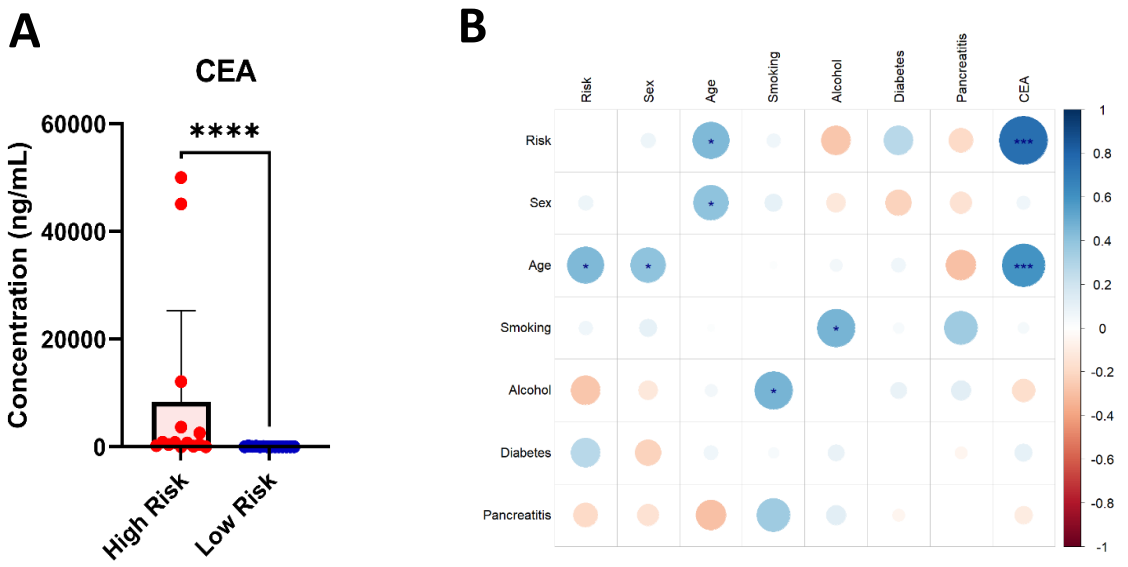


Figure 4.22 PCF CEA levels are significantly increased in high-risk patients compared to low-risk. (**A**) PCF concentration of CEA (ng/mL) in high-risk (red) and low-risk (blue) patients. Mann-Whitney test. Data are presented as mean ± SEM. (**B**) Spearman correlations between patient clinical data and the PCF CEA levels are given as a corrplot. Colour intensity relates to R^2 value; circle size relates to the p-value. Data are presented for n=14 low-risk and n=14 high-risk patients. *p<0.05, ***p<0.001.

in the ability of the panel to cluster patients into distinct groups, demonstrating an accuracy of 95% (Figure 4.23). However, LOOCV of this 11-feature multi-omic panel in PCF with the addition of PCF CEA generated an AUC value of just 0.690 (Figure 4.24). Reclassification of the VHL patient in this case caused the performance of the panel to worsen, producing an AUC value of 0.636 (Figure 4.25).

4.5.9 Proteomic and transcriptomic results could not be validated using other techniques

Sandwich ELISA was used in an attempt to validate the results seen in the proteomics, in the same patient cohort. The concentrations of four of the eight significant proteins that were identified in the PCF were examined via sandwich ELISA (LCN2, REG1A, PIGR, PRSS8). These data were normalised using two approaches, protein normalisation and dilution factor or volume normalisation. Given the variation in cyst size and protein content between patients, it was necessary to evaluate both normalisation methods to ensure no potential effects were missed. When normalised to the amount of protein, as determined by BCA assay, none of the four proteins were significantly differentially expressed between low- or high-risk patients (p>0.05)(Figure 4.26A-D). While an increase in concentration can be seen in the high-risk group compared to the low-risk for LCN2, REG1A and PRSS8, these increases were not significant (p>0.05). These data were then normalised to the dilution factor, and again, none of the four proteins were significantly differentially expressed between low- or high-risk patients (p>0.05)(Figure 4.26E-H). When correlating the LFQ intensities obtained from the proteomics with the protein concentrations obtained via sandwich ELISA, PRSS8, LCN2 and REG1A had significant correlations, but the R^2 values indicate that these relationships are not strong (p<0.05)(Figure 4.27).

qPCR microarrays were used in an attempt to validate the results obtained in the transcriptomics, in the same patient cohort. No significant difference was found in the expression levels of miR-216a-50, miR-216b-5p or SNORA66 between the low- and high-risk patient PCF (p>0.05)(Figure 4.28). MiR-216a-5p levels were close to statistical significant (p=0.0571), and as such may reach significance with more power (Figure 4.28A). The number of patients for which data was obtained was extremely poor across

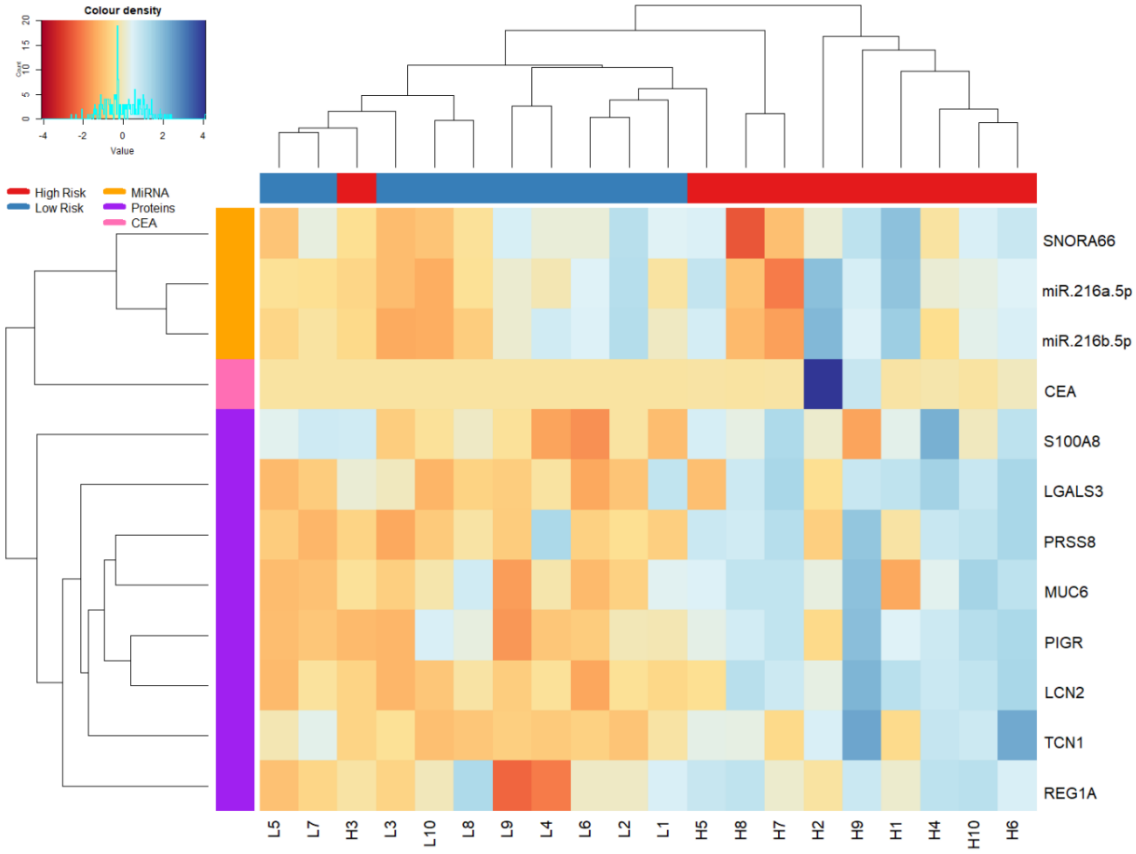


Figure 4.23 11-feature multi-omic panel in PCF plus PCF CEA clusters patients into risk groups with 95% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the 11-feature multi-omic panel in PCF plus PCF CEA. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the biomarkers. Data are presented for n=10 low-risk and n=10 high-risk patients.

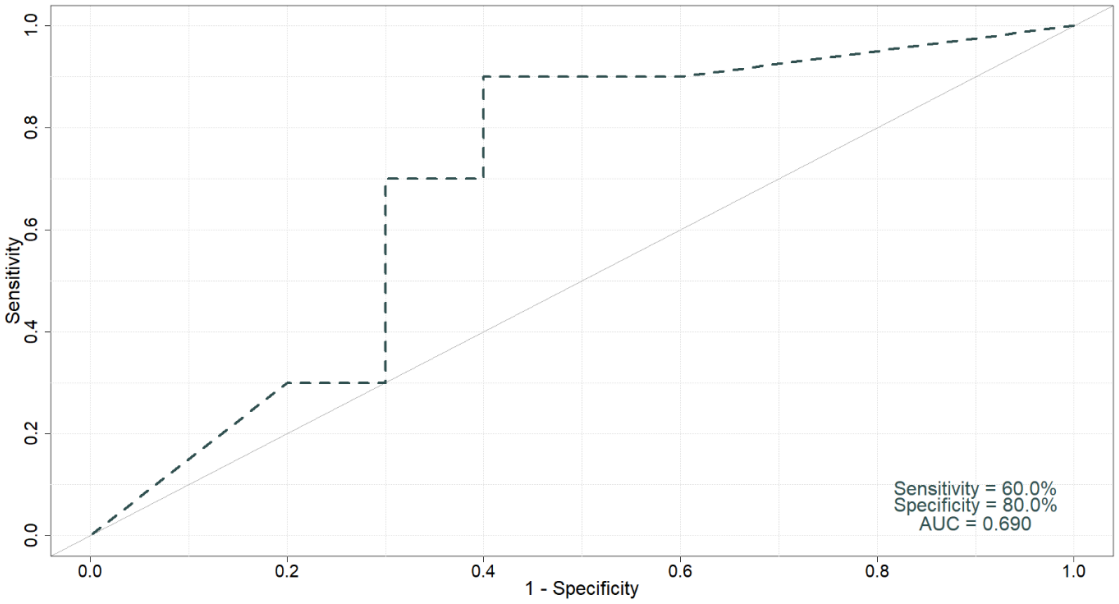


Figure 4.24 11-feature multi-omic panel in PCF plus PCF CEA classifies patients based on risk with AUC of 0.690. LOOCV of patients using the 11-feature multi-omic panel in PCF plus PCF CEA. Data are presented for n=10 low-risk patients and n=10 high-risk patients.

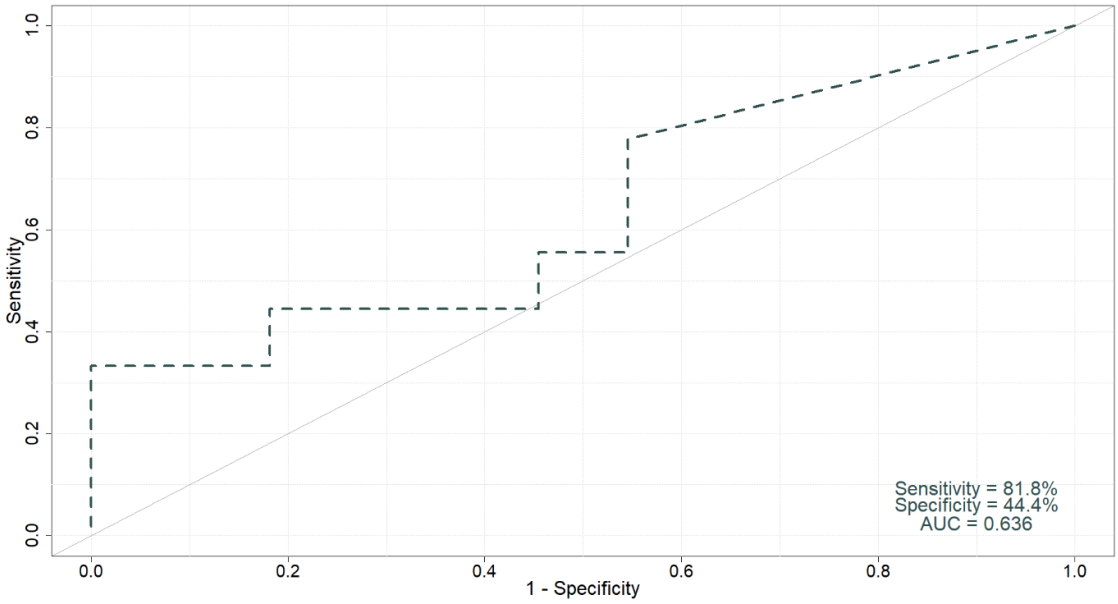


Figure 4.25 11-feature multi-omic panel in PCF plus PCF CEA classifies patients based on risk with AUC of 0.636 when VHL patient is reclassified. LOOCV of patients using the 11-feature multi-omic panel in PCF plus PCF CEA. VHL patient has been reclassified from high-risk to low-risk. Data are presented for n=10 low-risk patients and n=10 high-risk patients.

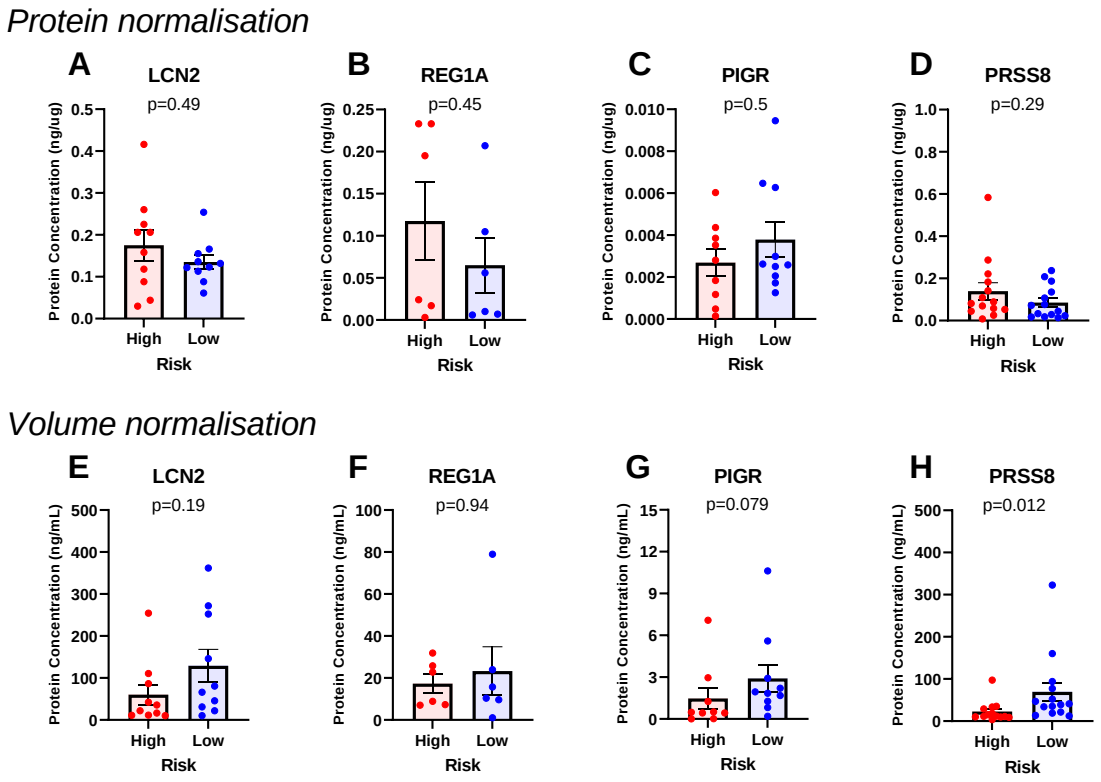


Figure 4.26 Sandwich ELISA does not validate proteomic results regardless of the normalisation method. Protein concentrations of four proteins identified as significant via proteomics (**A-D**) normalised to protein quantity (ng/µg) and (**E-H**) normalised to volume (ng/mL). LCN2 data are presented as mean ± SEM for n=10 low-risk and n=10 high-risk patients. REG1A data are presented as mean ± SEM for n=6 low-risk and n=6 high-risk patients. PIGR data are presented as mean ± SEM for n=10 low-risk and n=9 high-risk patients. PRSS8 data are presented as mean ± SEM for n=14 low-risk and n=14 high-risk patients. Mann-Whitney test.

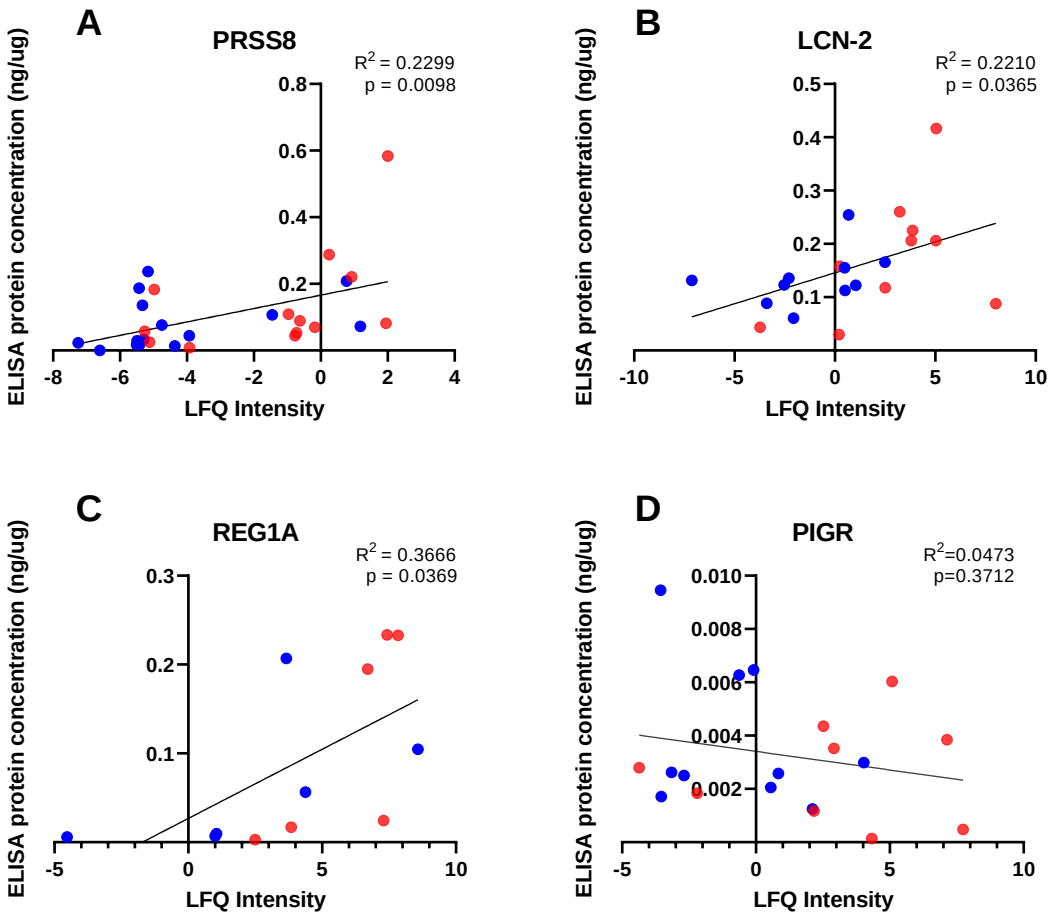


Figure 4.27 There is no correlation between LFQ Intensity values obtained via proteomics and sandwich ELISA protein concentration. Simple linear regression of LFQ intensity and sandwich ELISA protein concentrations of four proteins identified as significant via proteomics. (**A**) PRSS8 data are presented for n=14 low-risk and n=14 high-risk patients. (**B**) LCN-2 data are presented for n=10 low-risk and n=10 high-risk patients. (**C**) REG1A data are presented for n=6 low-risk and n=6 high-risk patients. (**D**) PIGR data are presented for n=10 low-risk and n=9 high-risk patients. Low-risk patients are shown in blue; high-risk patients are shown in red.

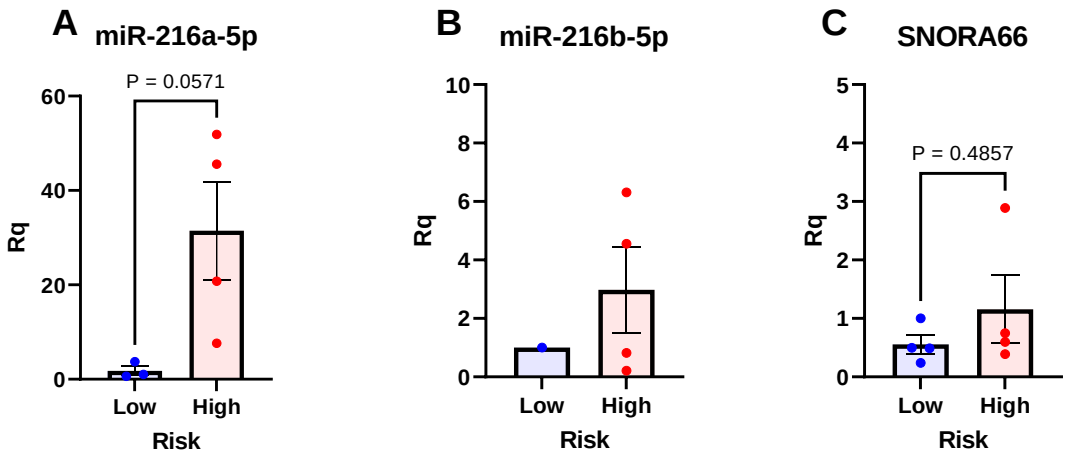


Figure 4.28 qPCR microarray does not validate transcriptomic results. Relative quantification (Rq) of 3-miRNA panel identified via transcriptomics. (**A**) MiR-216a-5p data are presented as mean ± SEM for n=3 low-risk and n=4 high-risk patients. (**B**) MiR-216b-5p data are presented as mean ± SEM for n=1 low-risk and n=4 high-risk patients. (**C**) SNORA66 data are presented as mean ± SEM for n=4 low-risk and n=4 high-risk patients. Mann-Whitney test.

all three miRNA, with miR-216b-5p in particular having usable results for just one low-risk patient, as the other six patients tested did not amplify (Figure 4.28B).

4.6 Discussion

The integration of multiple omic levels to create an 11-feature multi-omic panel produced the most robust results for patient PC risk classification. Indeed, while the 8-protein and 3-miRNA panels alone demonstrated a modest ability to classify PCL patients based on risk, the performance of these two sets of biomarkers combined into one panel delivered the best results overall, providing promising preliminary data as a novel biomarker panel in this setting.

Interestingly, it was shown that the 3-miRNA panel performed better than the 8-protein panel for risk classification by LOOCV, despite only one of the three miRNA significantly correlating with risk status, and demonstrating poor stratification in both the unsupervised hierarchical clustering and the PCA. Indeed, given the results of the systematic review and meta-analysis performed in Chapter 3, it would be expected that the 8-protein panel would be superior, as panels of more biomarkers were shown to perform better[379]. These contrasting results for the 3-miRNA highlight the importance of the method of evaluation used for biomarker efficacy. Unsupervised hierarchical clustering, for example, allows the datapoints to cluster based on patient expression of certain variables, and in this way it groups like-with-like to enable the visualisation of patterns[380]. Here, this analysis was used to investigate whether the patient cohorts would separate into groups based on their expression of these factors. While separation of patient risk groups was poor using the 3-miRNA panel, it is important to note that such methods can be greatly influenced by poor performing variables, especially when there are so few to begin with. Indeed, it was shown that when using this 3-miRNA panel to train and test a LOOCV model, this model performs modestly. Similarly, PCA analysis allows the visualisation of large datasets in smaller components or dimensions via dimensionality reduction, clustering similar samples together and aligning highly correlated variables[381]. In this way, PCA finds the majority of its utility in 'long' datasets with dimensionality issues, where there are more variables than the number of samples and as such, it can be difficult to discern the individual effects of a single variable[382]. In this study, three miRNA are dimensionally reduced to two components using PCA, which is not a common practice for such small datasets. However, this approach highlighted the outlier among the three biomarkers, SNORA66, and illustrated that the majority of

variance could be accounted for by miR-216a-5p and miR-216b-5p, suggesting that SNORA66 is perhaps the least important component of this 3-miRNA panel.

Interestingly, miR-216a-5p and miR-216b-5p are pancreas specific miRNAs, and as such, they have been studied extensively in the context of PC[383]. MiR-216b-5p has been demonstrated to function as a tumour suppressive RNA in pancreatic tissues by repressing PC cell proliferation, inducing apoptosis and cell cycle arrest, and supressing invasive and migratory capabilities[384, 385]. As such, its expression is generally reduced in PC tissues, and this is associated with poor prognosis[384, 386]. While circulating miR-216a-5p levels have been shown to be elevated in PC patients compared to controls[387], tissue expression levels are also known to be downregulated in PC compared to healthy controls[386].MiR-216a-5p is also believed to function in a tumour suppressive capacity in PC by inhibiting tumour growth via the JAK/STAT pathway[388]. Interestingly, despite the downregulation of these miRNA in PC, both miR-216a-5p and miR-216b-5p have been shown to be increased in high-risk IPMNs compared to low-risk previously[389]. While no research appears to have been conducted to date examining the mechanisms involved in this shift in expression from PCL to PC, and none of the three miRNA significantly correlated with any clinical factors, in this setting these miRNA remain promising potential biomarkers of PC risk. SNORA66, on the other hand, remains largely unstudied in PC or PCLs. Indeed, SNORA66 is one of many small nucleolar noncoding RNAs that has cellular housekeeping functions and is therefore commonly utilised as a housekeeping RNA[390]. Despite this, several snoRNAs have been reported to be dysregulated in some cancer types, with SNORA23, for example, being overexpressed in PDAC[391]. As such, while there is no previous evidence of SNORA66 dysregulation in PCLs or PC, it is not unusual for housekeeping genes to become dysregulated in disease, and this may be the first instance of such a report for SNORA66 in pancreatic disease. Overall, the various approaches used to analyse these data highlight the strengths and weaknesses of the panel, and demonstrate that while hierarchical clustering and PCA are useful for interrogating datasets, training and testing models which are developed for the examination of biomarkers, such as LOOCV, gives the best sense of biomarker performance. Metrics such as AUC value, sensitivity and specificity are the most important in this context, and should therefore be given the most weight.

Conversely, the 8-protein panel performed well in all analyses, likely due to the fact that there are more variables to help distinguish and restructure the data according to their expression. In this way, larger panels allow for better handling of outlier patients, as when one variable becomes dysregulated, the others within the panel can compensate for this. Despite the large patient-to-patient variation in the expression of these proteins, all eight significantly correlated with risk. Interestingly, six of these proteins [LCN2, REG1A, PIGR, S100A8, LGALS3 and MUC6] have been highlighted previously in PC literature. LCN2 is a small glycoprotein that is secreted by various cell types such as macrophages, neutrophils and epithelial cells[392]. LCN2 has been demonstrated to be highly expressed in PDAC tissues compared to normal tissues[393], and has shown utility as a circulating biomarker of familial PC risk[336, 394]. Importantly, LCN2 was previously found to be highly expressed in early dysplastic PCLs, indicating its potential role as a biomarker of PC risk and early PC detection[395]. Similar levels of LCN2 expression have been observed in normal and chronic pancreatitis tissue, but these levels are elevated 4.3-fold in PC [396]. Conversely, the levels of LCN2 detected in the serum were significantly reduced in PC patients compared to controls, demonstrating the variability of this marker between detection modalities[396]. Importantly, LCN2 expression is also associated with good prognosis in PC, having been shown to inhibit PC stemness, reverse EMT, and supress invasion and angiogenesis[393, 397, 398]. Indeed, poorly differentiated PC tumours have exhibited little to no expression of LCN2[395, 398]. Similarly, increased REG1A levels in PC patient serum has been shown to correlate with improved survival compared to those with low REG1A levels[399]. REG1A has been shown to be significantly increased in the urine, serum and tissue of PC patients compared to healthy controls[399-402]. An upregulation of REG1A in PC cells has also been demonstrated to accelerate cell proliferation and tumour growth both *in vivo* and *in vitro*[403]. Indeed, REG1A expression in tissue was shown to increase from benign ductal epithelium to PCL to PC[399]. These results suggest the upregulation of REG1A during PDAC development, and that this upregulation may have protective properties.

Increased LGALS3 expression has also been observed as an early PC event. LGALS3 expression has been shown to be 1.5-fold higher in chronic pancreatitis tissues compared to healthy controls, but up to 6.5-fold higher in PC tissue, increasing

incrementally as the disease progresses[396, 404]. Indeed, both LGALS3 mRNA and protein levels have been shown to be significantly increased in PC tissue compared to healthy controls[405]. Interestingly, LCN2, REG1A, LGALS3 and S100A8 have all been previously detected in patient PCF, with S100A8 and LCN2 also being detected in PCF cell pellets, though no indication as to the level of expression or differential expression between controls were reported[121]. Importantly, many S100 proteins have been identified as overexpressed in numerous cancer types, highlighting a potential role for these proteins in cancer development[406, 407]. High S100A8 expression in pancreatic ductal fluid predicted worse disease-free and overall survival in late stage PC patients[408]. Furthermore, S100A8 was shown to be overexpressed in PC tumours compared to normal and pancreatitis tissues[402]. Stimulation of pancreatic cell lines with S100A8 has been shown to increase cell motility, proliferation and pro-inflammatory cytokine secretion *in vitro*, further highlighting the potential role of S100A8 in PC progression[409, 410]. High PIGR expression in PC patient tissue was also shown to be an indicator of poor prognosis, with PIGR levels being significantly higher in PDAC patients preoperatively compared to postoperatively[411, 412]. PIGR serum levels have also been observed at significantly higher levels in PC patients compared to healthy controls[413]. Lastly, MUC6 or mucin 6 is a secreted protein whose levels have been shown to be downregulated in PC tissues compared to normal tissues, and is therefore a negative prognostic indicator[414, 415]. MUC6 expression has also been shown to be higher in well-differentiated tumours compared to moderately or poorly differentiated tumours[416, 417]. Furthermore, MUC6 expression has been found to be increased in 'uninvolved' pancreatic ducts of PC patients compared to normal ducts[416]. Greater expression of MUC6 proteins was observed in PCLs compared to normal pancreatic tissue, with MUC6 levels shown to decrease with tumour progression suggesting MUC6 loss is an early event in PC[257, 417, 418]. These data overall suggest the dysregulation of these six proteins in pancreatic disease, and therefore strengthens the potential for their use as biomarkers in this context.

Unfortunately, there is no evidence in the literature to-date to suggest the dysregulation of PRSS8 or TCN1 in PC. While PRSS8 has been shown to be expressed at levels more than 100-fold higher in ovarian cancer patient tissue compared to normal

healthy donors, with evidence suggesting that PRSS8 is upregulated in the early stages of this disease, no data could be found on PRSS8 expression in PC[419]. TCN1 or transcobalamin is a vitamin B12-binding protein which has been shown to be a predictor of poor prognosis in colon cancer patients, but again, there is no indication in literature as to the expression of TCN1 in pancreatic disease[420]. While no confirmation of the dysregulation of PRSS8 or TCN1 could be identified in the literature, here it is reported for the first time for all eight of these proteins, that the expression levels of these proteins are higher in high-risk PCF compared to low-risk PCF, and as such all eight have good potential in this context to be biomarkers of patient PC risk. Importantly, when examined together as an 8-protein panel, these proteins stratify patients with modest accuracy.

The utility of both the 3-miRNA and 8-proteins panels alone is limited, with extremely similar AUC values being obtained for both via LOOCV. However, the integration of these two panels to form a single multi-omic biomarker panel is where the true potential of these biomarkers can be seen. Indeed, current trends in biomarker identification lean towards the creation of multi-omic panels that can better control for the complexity of the disease[166]. By encompassing factors from multiple biological levels, multi-omic biomarker panels can better compensate for the dysregulation of individual biological levels. In fact, substantial improvements in results from all analyses can be seen through the use of the 11-feature multi-omic panel. Importantly, the identification of one outlier patient further compounds the strength of this panel. VHL syndrome is a familial neoplastic condition, caused by a germline mutation to the VHL tumour suppressor gene, which can increase a patients risk of PCLs and PC[128, 421]. In this case, the patient with VHL syndrome presented with a PCL but was classified as high-risk based on their genetic predisposition to PC. When re-examined with the presence of a VHL mutation excluded, this PCL was reclassified as low-risk. As such, these data were examined with this patient classified as both low- and high-risk, and notable changes in panel performance can be observed. Indeed, in the initial model with this patient classified as high-risk, this datapoint can be frequently seen as an outlier. In the unsupervised hierarchical clustering specifically, it is the only datapoint that is incorrectly clustered. When reclassified, the model performance improves for all

analyses, with perfect hierarchical clustering, improved separation of the two groups in the PCA, and increased AUC for the LOOCV. While it is impressive that the model could identify this outlier, it is important to highlight the substantial alteration to model performance that just one patient could make. Furthermore, while in this case the presence of a VHL mutation stood out clinically as a potential confounding factor, it may not be appropriate to reclassify or remove this patient from this analysis as their original classification as high-risk was based on the same guidelines as all other patients. These data emphasise the need for validation of these results in a larger, independent patient cohorts where longitudinal progression data can confirm whether high-risk patients progressed to PC.

An attempt was made to validate the differential expression of these proteins and miRNA in alterative platforms within the same patient cohort. Unfortunately, neither of these attempts managed to do so. For the proteins, while ELISA has been used to validate proteomic results previously[422-425], other studies have had similar issues using ELISA as a validation platform[426]. Here, no correlation between LFQ Intensity values and protein concentrations for four of the eight proteins could be demonstrated. As sandwich ELISAs are highly sensitive assays where diluted sample is placed on a pre-coated plate, it is possible that the viscous and mucinous nature of the PCF could have affected the efficacy of the assay. Indeed, the workflow for the proteomic profiling involved extensive pre-processing of the PCF, while preparation for the ELISA only included diluting PCF in sample diluent. Further validation efforts should consider alternative platforms, such as western blot, or pre-processing of the PCF samples prior to incubation on the ELISA plate, as well as increased power of the sample numbers. Indeed, a lack of power also appeared to be an issue in the validation of the miRNA, as many samples did not amplify and this left few datapoints to power any statistical analyses. While miR-216a-5p trended towards significance, these validations should be re-examined in a larger cohort of patient PCF to ensure sufficient power is acquired.

The addition of CA19-9 to the 11-feature multi-omic panel was expected to improve the performance of the panel, as was the result seen across a the literature in Chapter 3. Unfortunately, as no variation was seen in CA19-9 serum levels between low- or high-risk patients, the addition of CA19-9 to the panel, in this case, worsened its

performance. These data highlight two important points: firstly, that CA19-9 appears to have limited utility in a risk stratification setting; and secondly, that the addition of a poor performing biomarker to a strong panel can have substantial consequences to panel performance. A similar result was seen with the addition of CEA to the panel, where although the unsupervised hierarchical clustering was unaffected, the LOOCV performance of the panel was substantially reduced following the addition of CEA to the 11-feature multi-omic panel. As such, it is of the utmost importance, when examining biomarker panels, to ensure that all features within the panel are robust and contribute positively to its performance. While promising potential blood-based biomarkers where identified in Chapter 3 and re-examined here in an independent cohort, their performance as a panel was demonstrated to be poor. Indeed, while four of these proteins were significantly increased in high-risk PCF compared to low-risk via Mann Whitney test, it is important to remember that when stringent differential expression analysis with FDR was conducted, these proteins were not significant. As such, while many factors may be increased or decreased from one group to the next, only those with the most robust performance should be brought forward.

The results of this study provide a novel multi-omic biomarker panel which shows potential for the risk stratification of PCL patients. Importantly, these data also highlight potential caveats to biomarker panel design and analysis, and as such demonstrate the importance of careful and extensive validation of results in novel patient cohorts. However, it is important to note that a PCF-based panel, while potentially robust, would also be quite invasive to utilise in a clinical setting. As such, current research trends more towards liquid biopsies in the context of the peripheral blood, and the utility of blood-based biomarkers as a less invasive alternative.

Chapter 5.

Multi-omic profiling of serum for the identification of novel biomarkers of patient pancreatic cancer risk

5.1 Introduction

PC has the worst survival rate of any cancer, being reported at just 12% in 2023 by the American Cancer Society[1]. This low survival rate is primarily attributed to the late stage of presentation of most patients. Indeed, the symptoms associated with PC are extremely vague, such as abdominal pain or weight loss, causing patients to ignore these issues until they are more severe, and the disease is therefore at an advanced stage of development[299]. The treatment options for patients at late stage disease are limited, contributing further to the poor survival of PC patients. As such, early identification of PC or those who are at an increased risk of developing PC is required to expand treatment options for these patients and subsequently improve survival rates.

PCLs are fluid-filled sacs found on or inside the pancreas which can either be benign or pre-malignant[121]. Unfortunately, current methods of distinguishing between the two are imperfect. At present, clinical guidelines evaluating factors such as cyst size or location on the pancreas are used to stratify patients with PCLs into risk categories. However, there are numerous different iterations of these guidelines worldwide, with conflicting cut-offs and metrics, highlighting the lack of consensus among clinical staff as to the most appropriate or best performing guidelines[376, 377]. This means that patients with high-risk PCLs are being missed at this important junction due to the lack of standardized risk stratification.

In Chapter 4, two commonly used biomarkers for PCL and PC diagnosis and management, CEA and CA19-9, were evaluated in a novel patient cohort and shown to have poor PCL risk stratification capabilities. CEA is measured in the PCF aspirate post-endoscopic ultrasound and used to identify mucinous cysts, the subtype of PCLs with the most malignant potential[163]. CA19-9 is an FDA-approved blood-based biomarker for PC diagnosis, however, it has been shown to be dysregulated in patients with underlying conditions such as diabetes or pancreatitis, making its performance in this cohort highly inaccurate[167]. As these two commonly utilised biomarkers were demonstrated to add no diagnostic accuracy to a panel of novel PCF biomarkers, but rather worsen the performance of the panel, the need for robust biomarkers that can accurately distinguish high- and low-risk PCLs is evident.

In this study, promising biomarkers that had been previously examined across PC literature, as identified in Chapter 3, were evaluated in the serum of a novel patient cohort. These blood-based diagnostic biomarkers for PC were interrogated to assess their potential as biomarkers of PCL cancer risk. Furthermore, multi-omic profiling of the serum of PCL patients was conducted with the aim of identifying novel biomarkers of patient risk. Proteomic and transcriptomic analysis of PCL patient serum was carried out to identify differentially expressed proteins and miRNA. These features were assessed both alone, and as part of a multi-omic panel, to determine whether the examination of biomarkers across multiple omic levels could better account for the complexity of pancreatic disease, and therefore compensate appropriately for those within the panel that might become dysregulated. Lastly, to build on the work that was carried out in Chapter 4, where a promising multi-omic PCF biomarker panel was identified, the top performing markers in both the PCF and serum were combined to generate a cross-biofluid multi-omic. This panel was then assessed in a matched patient cohort, to evaluate whether, in the same manner as multi-omics, cross-biofluid panels may encapsulate both pancreas-specific and circulating biological processes, allowing for improved robustness in patient risk stratification.

5.2 Hypotheses

1. There are factors within the serum which are differentially expressed between low- and high-risk patients, and these factors can be used to stratify patients based on their risk of PC.
2. The dimensional reduction and combination of a PCF-based biomarker panel and a serum-based biomarker panel, to create a cross-biofluid biomarker panel, will produce the most robust risk stratification accuracy.

5.3 Specific Aims (Part One)

1. Examine the performance of biomarkers that were previously identified in the literature in a novel patient cohort.
2. Profile the proteome of the serum and identify proteins that are differentially expressed between low- and high-risk patients.
3. Profile the transcriptome of the serum and identify miRNA that are differentially expressed between low- and high-risk patients.
4. Examine the utility of differentially expressed proteins and miRNA in the serum as biomarkers of PC risk, both alone and integrated into a multi-omic panel.

5.4 Specific Aims (Part Two)

1. Integrate the top performing biomarkers from the PCF and serum to generate a cross-biofluid multi-omic panel of patient PC risk.

5.5 Experimental design

5.5.1 Patient demographic information for serum cohort

Three different patient cohorts were examined as part of this study. Demographic and clinical information for patients in the proteomic cohort, the transcriptomic cohort and the multi-omic cohort are given in Table 5.1. Risk classifications for these patients were assigned as per the 2018 European evidence-based guidelines for pancreatic cystic neoplasms [157]. Details on specific clinical criteria used to determine these classifications for each patient are unavailable, and as such, this lack of information should be noted as a limitation to the risk classifications of these patients.

5.5.2 Proteomic profiling of patient serum

Proteomic examination of patient serum was carried out as per section 2.2.8. The estimated starting material for sample processing and LC-MS preparation as per section 2.2.8.3 is 40 µg per sample. Peptide elutants were therefore resuspended in 40 µL of LC-Load Buffer to give a final peptide concentration of 1 g/L prior to MS analysis.

5.5.3 Transcriptomic profiling of patient serum

Transcriptomic evaluation of patient serum was carried out as per section 2.2.10. HTG EdgeSeq Plasma Lysis Buffer was used to lyse serum samples, and the subsequent lysate was diluted 1:2 with sample diluent prior to sequencing.

5.5.4 Analysis of proteomic and transcriptomic serum data

Proteomic and transcriptomic data obtained from the serum were processed, analysed and subsequently scaled and integrated using the methodologies outlined in section 2.2.21.1.

5.5.5 QIAGEN qPCR custom microarray for serum

The validation of transcriptomic results was carried out using QIAGEN RT-qPCR custom microarrays. Assay setup was carried out as per section 2.2.11. Custom QIAGEN pre-

Table 5.1 Patient demographic information for serum cohorts.

	Proteomics		Transcriptomics		Multi-omic	
	Low Risk	**High Risk**	**Low Risk**	**High Risk**	**Low Risk**	**High Risk**
No of patients (M/F)	45 (21/24)	23 (10/13)	15 (5/10)	15 (8/8)	14 (5/9)	15 (8/8)
Mean age (range)	60 (22-84)	69 (39-85)	56 (22-80)	71 (39-85)	56 (22-80)	71 (39-85)
Smoking habits						
None	24	14	10	8	9	8
Active	7	1	2	1	2	1
Ex-smoker	10	5	3	5	3	5
Not known	4	3	0	1	0	1
Alcohol consumption						
None	23	12	8	9	8	9
Active	12	6	6	3	5	3
Heavy	2	1	0	1	0	1
Abstinent (ex-heavy)	4	0	1	0	1	0
Not known	4	4	0	2	0	2
Diabetic	5	3	1	1	1	1
Pancreatitis	2	0	1	0	1	0
Von Hippel-Lindau	0	1	0	1	0	1

Cohorts are divided into low- and high-risk groups. Mean age is rounded to the nearest whole number.

coated microarray plates were used, evaluating 18 targets and 6 quality controls as listed in Table 5.2. Targets of interest within the serum were miR-197-5p, miR-6741-5p, miR-3180-3p, miR-3180 and miR-6782-5p. Plates were prepared fresh and run on the same day. A sample layout and plating map for this workflow can be seen in (Appendix 3).

Table 5.2 QIAGEN 24-array custom plate targets and quality controls.

miRNA targets			miRNA Quality controls
SNORA66	miR-6741-5p*	miR-6727-5p	UniSp2
miR-216a-5p	miR-3180-3p*	miR-2861	UniSp3
miR-216b-5p	miR-3180*	miR-375-3p	UniSp4
miR-3197	miR-6782-5p*	miR-500b-3p	UniSp6
miR-1237-5p	miR-1207-5p	miR-532-5p	miR-451a
miR-197-5p*	miR-1908-5p	miR-130a-5p	miR-23a-3p

Targets of interest for this chapter are denoted with *.

5.5.6 Panel reduction and integration using CombiROC software

The PCF-based panel generated in Chapter 4 and the serum-based panel generated in this chapter were dimensionally reduced and integrated as per the methods outlined in section 2.2.21.1.

5.6 Results: Part One

5.6.1 Three promising biomarkers are not differentially expressed in serum

Promising biomarkers that were identified in Chapter 3 as being repeatedly examined in the literature, and that demonstrated good sensitivity and specificity for PC diagnosis, were interrogated in this cohort within the serum. From this list, two proteins were identified via label-free proteomics (Figure 5.1A-B), and one miRNA was identified via whole transcriptome sequencing of patient serum (Figure 5.1C). CRP, LRG1 and miR-21-3p levels were not significantly different in high-risk serum compared to low-risk serum (p>0.05)(Figure 5.1).

The expression levels of these three biomarkers within the serum were then used to stratify patients based on their expression of these proteins and miRNA. Unsupervised hierarchical clustering of patients into risk groups using these biomarkers was performed with an accuracy of 65.5% (Figure 5.2). CRP and LRG1 expression levels were most related, with miR-21-3p being the least related. Overall, these three blood-based biomarkers for the diagnosis of PC performed poorly in the risk stratification of PCL patients based on their expression levels within the serum.

5.6.2 Eight proteins were identified as being decreased in high-risk serum

Label-free proteomics identified 145 proteins present in the serum samples after data clean-up. Two proteins [SHROOM3 and IGHV3-72] were found to be significantly downregulated in high-risk serum compared to low-risk serum (p<0.05)(Figure 5.3A). Despite not being significantly differentially expressed, a further six proteins with the lowest p-values were taken forward for biomarker analysis [IGJ, IGHA1, PPBP, APOD, SFN, IGHG1], as panels of more biomarkers were shown to produce better results in Chapter 3. A total of eight proteins were examined, as the 8-protein panel in PCF examined in Chapter 4 was shown to have good accuracy. The distribution of the expression levels from patient-to-patient within both the low- and high-risk groups was quite large (Figure 5.3B).

SHROOM3 was the only one of the eight proteins to significantly correlate with patient risk, having a negative correlation (p<0.01)(Figure 5.4). While none of these

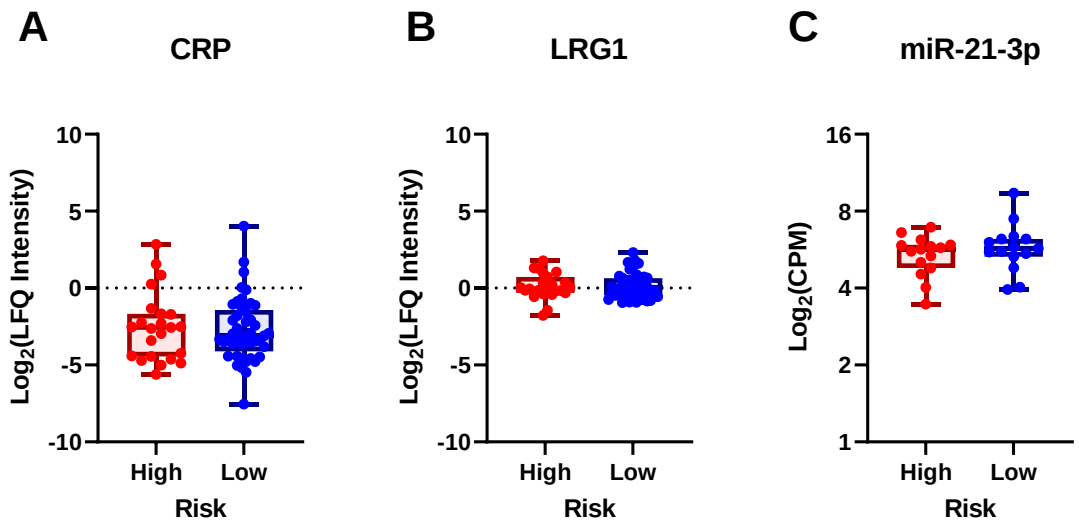


Figure 5.1 Three promising biomarkers are not significantly differentially expressed high-risk serum compared to low-risk. (**A-B**) Boxplots showing the expression level of promising protein biomarkers, in Log_2(LFQ intensity). Data are presented as mean ± SEM for n=23 high-risk patients and n=45 low-risk patients. (**C**) Boxplot showing the expression level of a promising miRNA biomarker, in Log_2(CPM). Data are presented as mean ± SEM for n=15 low-risk and n=15 high-risk patients. Mann-Whitney test.

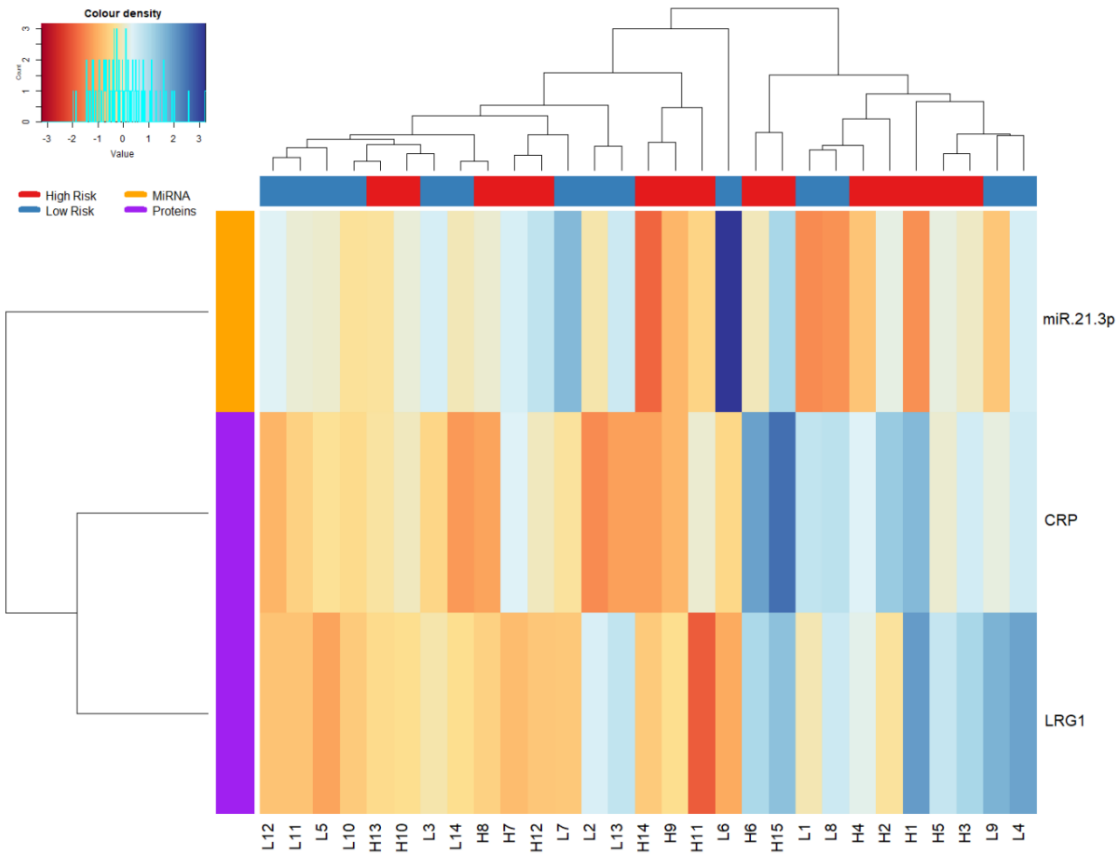


Figure 5.2 Expression of the three top occurring biomarkers in serum clusters patients into risk groups with 65.5% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the three promising biomarkers in the serum. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins (purple) and miRNA (orange). Data are presented for n=14 low-risk and n=15 high-risk patients.

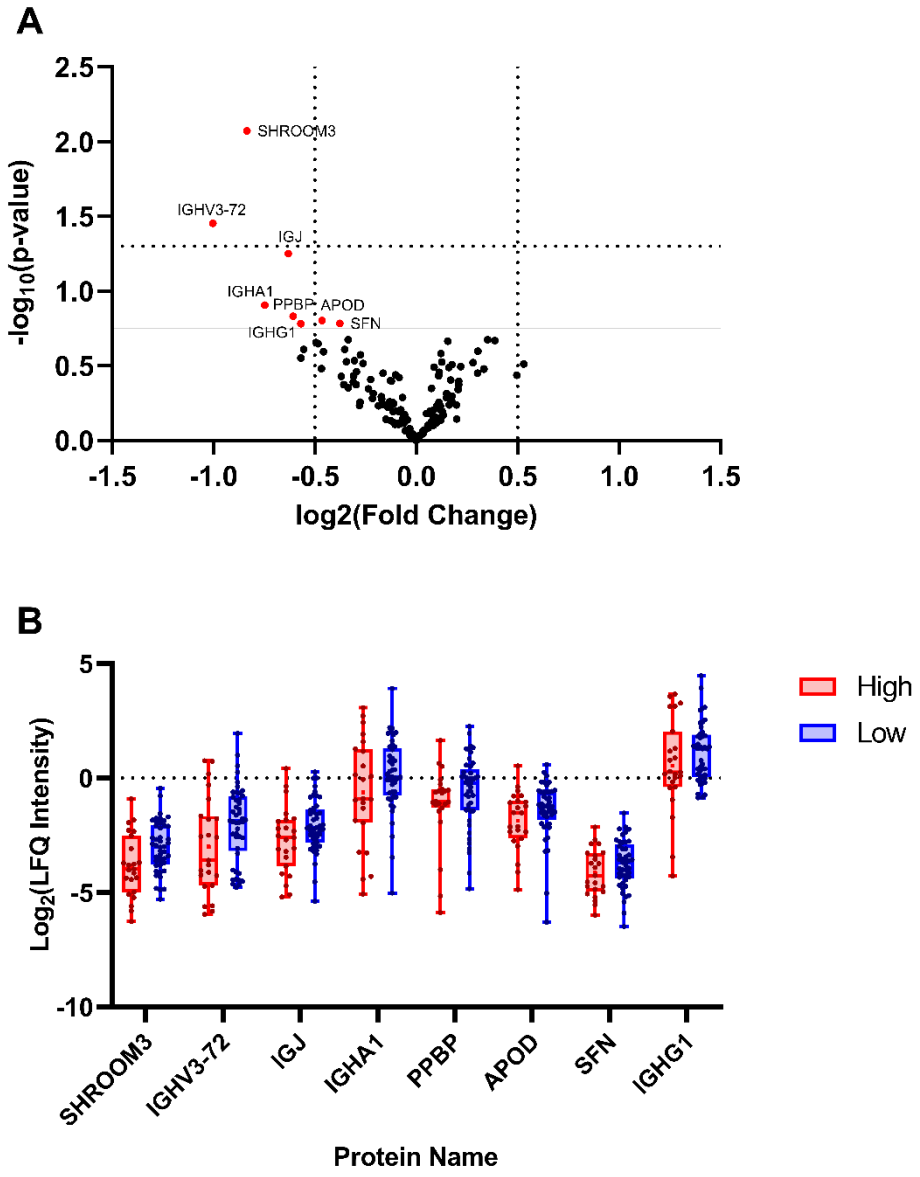


Figure 5.3 Eight proteins were identified as being downregulated in high-risk patient serum compared to low-risk. (**A**) Volcano plot showing expression patterns of proteins between low- and high-risk serum. Proteins identified in red were considered to be downregulated in high-risk serum compared to low-risk serum. Dotted lines indicate the fold change and significance cut-offs. (**B**) Box plots showing the distribution of patient expression levels, in Log_2(LFQ intensity), for each of the eight proteins. High-risk patients (n=23) are shown in red; low-risk patients (n=45) are shown in blue.

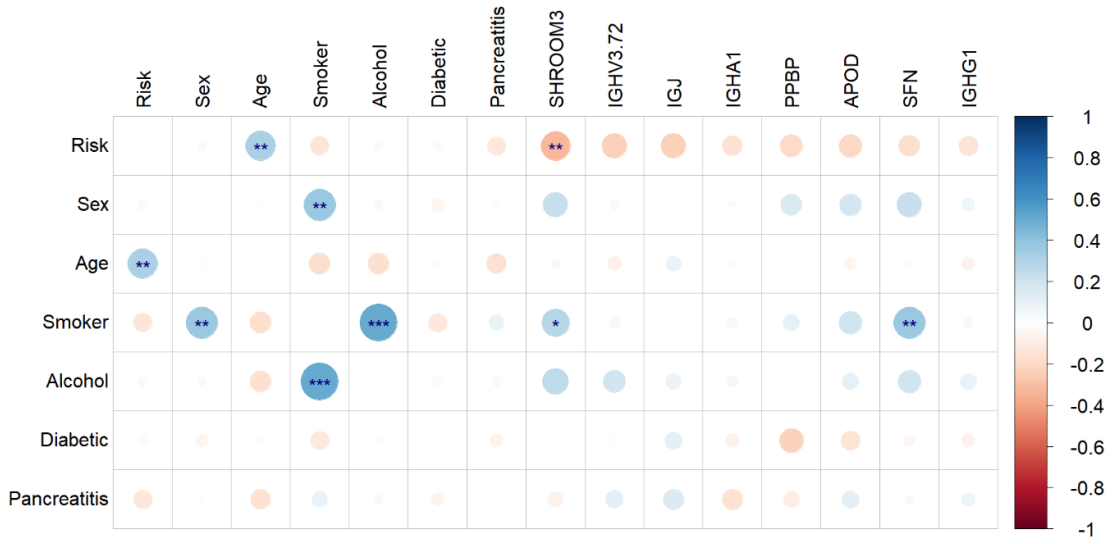


Figure 5.4 Expression of SHROOM3 in the serum significantly correlates with patient cancer risk. Spearman correlations between patient clinical data and the eight differentially expressed proteins are given as a corrplot. Colour intensity relates to R^2 value; circle size relates to the p-value. Data are presented for n=45 low-risk and n=23 high-risk patients. **p*<0.05, ***p*<0.01, ****p*<0.001.

proteins correlated with the presence of diabetes or pancreatitis, nor with alcohol consumption, age or sex, both SHROOM3 and SFN positively correlated with smoking habits (p<0.05).

Several clinical factors were also shown to correlate with each other in this patient cohort. Increased risk significantly correlated with increased age (p<0.01); increased smoking significantly correlated with increased alcohol consumption (p<0.001); and increased smoking significantly correlated with sex, or given the binary coding for this, increased smoking significantly correlated with being male (p<0.01).

5.6.3 8-protein panel in serum classifies patients with modest accuracy

These eight proteins were then scaled and integrated to create an 8-protein biomarker panel. This 8-protein panel was then used to stratify patients into distinct groups based on their expression of these proteins. Unsupervised hierarchical clustering of patients into risk groups using this 8-protein panel was performed with an accuracy of 76.5%, with n=8 low-risk and n=8 high-risk patients being grouped incorrectly out of the 78 patients (Figure 5.5). The relationships between these eight proteins show a split into 2 groups, with IGHG1, IGHA1, IGHV3-72 and IGJ forming one group and PPBP, APOD, SHROOM3 and SFN forming the second. The expression levels of the only two significantly differentially expressed proteins, SHROOM3 and IGHV3-72, were not closely related.

PCA was conducted using two components, however, these two components only accounted for 56.7% of the variance (36.2% in component 1 and 20.5% in component 2) (Appendix 7) (Figure 5.6). Here, the separation of these eight proteins was seen again, as IGHG1, IGHA1, IGHV3-72 and IGJ were the most important contributors to the first component, with PPBP, APOD, SHROOM3 and SFN being the most important contributors to the second component. The proteins in these two groups were highly positively correlated with each other. APOD had the least amount of variance attributed to it out of all eight proteins. The ellipse encapsulating 80% of each classification is quite large, with substantial overlap between the two, indicating poor separation of the two groups based on these two components. The third component in the PCA accounted for

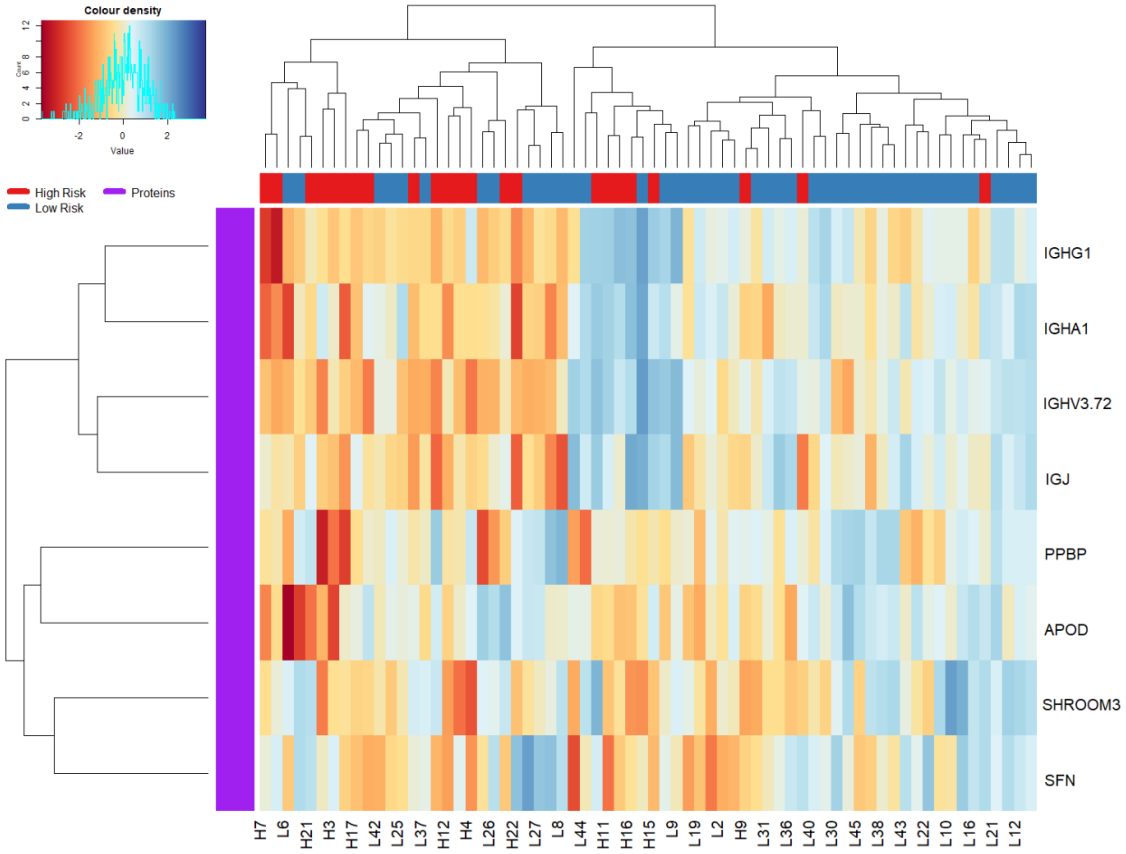


Figure 5.5 8-protein panel in serum clusters patients into risk groups with 76.5% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the eight proteins. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins (purple). Data are presented for n=45 low-risk and n=23 high-risk patients.

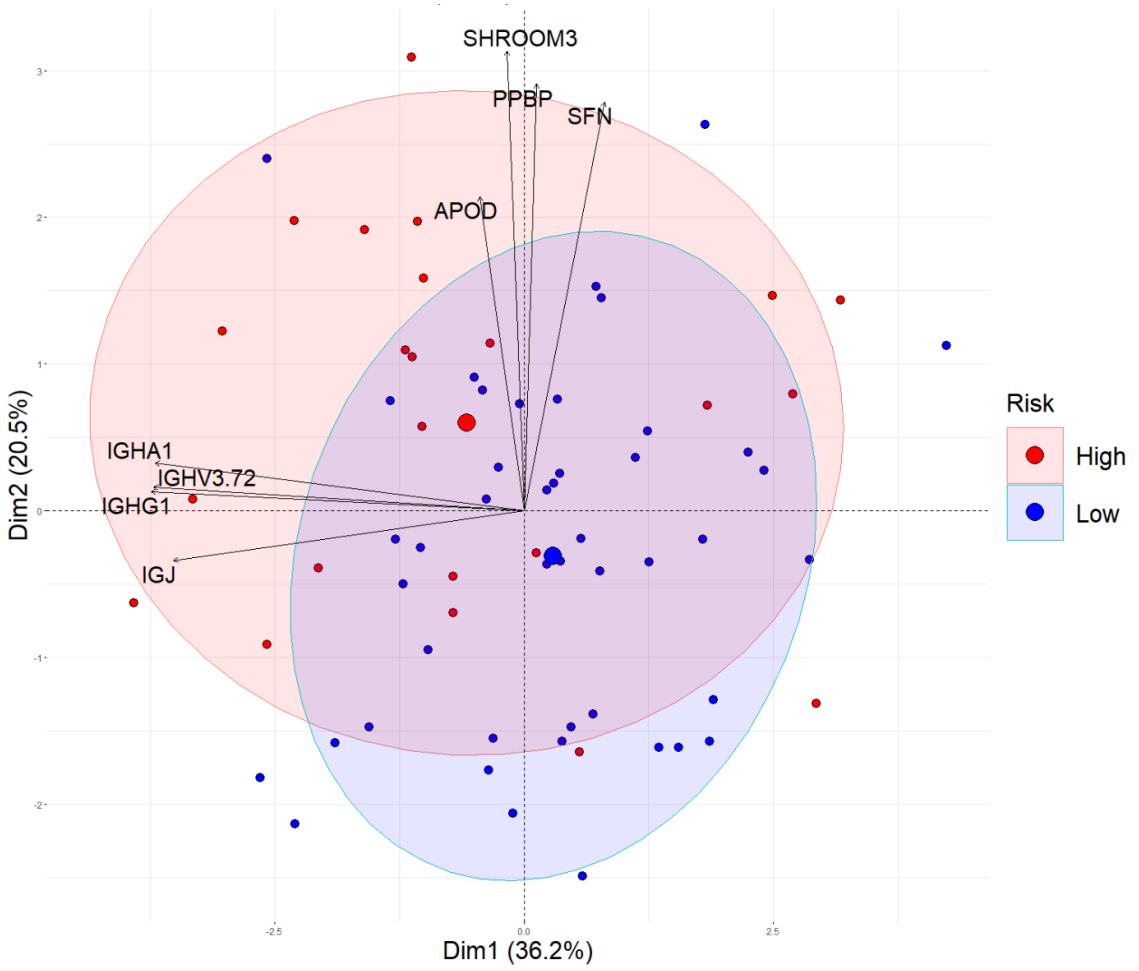


Figure 5.6 PCA analysis of 8-protein panel in serum shows poor distinction of the risk groups. 2-D PCA using the eight proteins, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding protein. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. Data are presented for n=23 low-risk (blue) and n=45 high-risk (red) patients.

12.7% of the variance. As such, when examined with a third component in 3-dimensions, the total variance accounted for is 69.4% (Figure 5.7). With three components, separation of groups is still poor, with any patterns being difficult to discern (Figure 5.7A). When examining only the first two components, again the separation of the eight proteins into two distinct groups is evident (Figure 5.7B). However, when looking at just component 1 and component 3 the proteins that were shown to account for the variance in component 2 are shown to also align greatly with component 3 (Figure 5.7C). Finally, examining components 2 and 3 together shows that SHROOM3, SFN and PPBP account for the variance in component 2, while APOD accounts for component 3 (Figure 5.7D).

LOOCV was then carried out, where the model was trained and then tested using the same patient cohort. Here, the 8-protein panel produced an AUC value of 0.608, with a sensitivity of 82.2% and a specificity of 34.8%, indicating that this panel performs well in its ability to assign a high-risk classification to those who are high-risk (Figure 5.8).

5.6.4 Five miRNA were identified as significantly increased in high-risk serum

Whole transcriptome sequencing identified 2,096 miRNA present in the serum samples after data clean-up. Differential expression analysis revealed five miRNA [miR-197-5p, miR-6741-5p, miR-3180, miR-3180-3p and miR-6782-5p] to be significantly upregulated in high-risk serum compared to low-risk serum (adj-p<0.05, FDR=0.05, s0=0.1)(Figure 5.9A). The distribution of the expression levels from patient-to-patient within both the low- and high-risk groups was modest, with higher variability being seen in the high-risk group compared to the low-risk (Figure 5.9B).

Spearman correlations showed that expression levels of the five miRNA did not significantly correlate with patient risk (p>0.05)(Figure 5.10). Only one of these miRNA significantly correlated with a clinical parameter, with miR-6741-5p expression positively correlating with increased smoking (p<0.05). While miR-6782-5p had a negative correlation with age, and miR-3180 and miR-3180-3p negatively correlated with alcohol consumption, these were not significant (p>0.05).

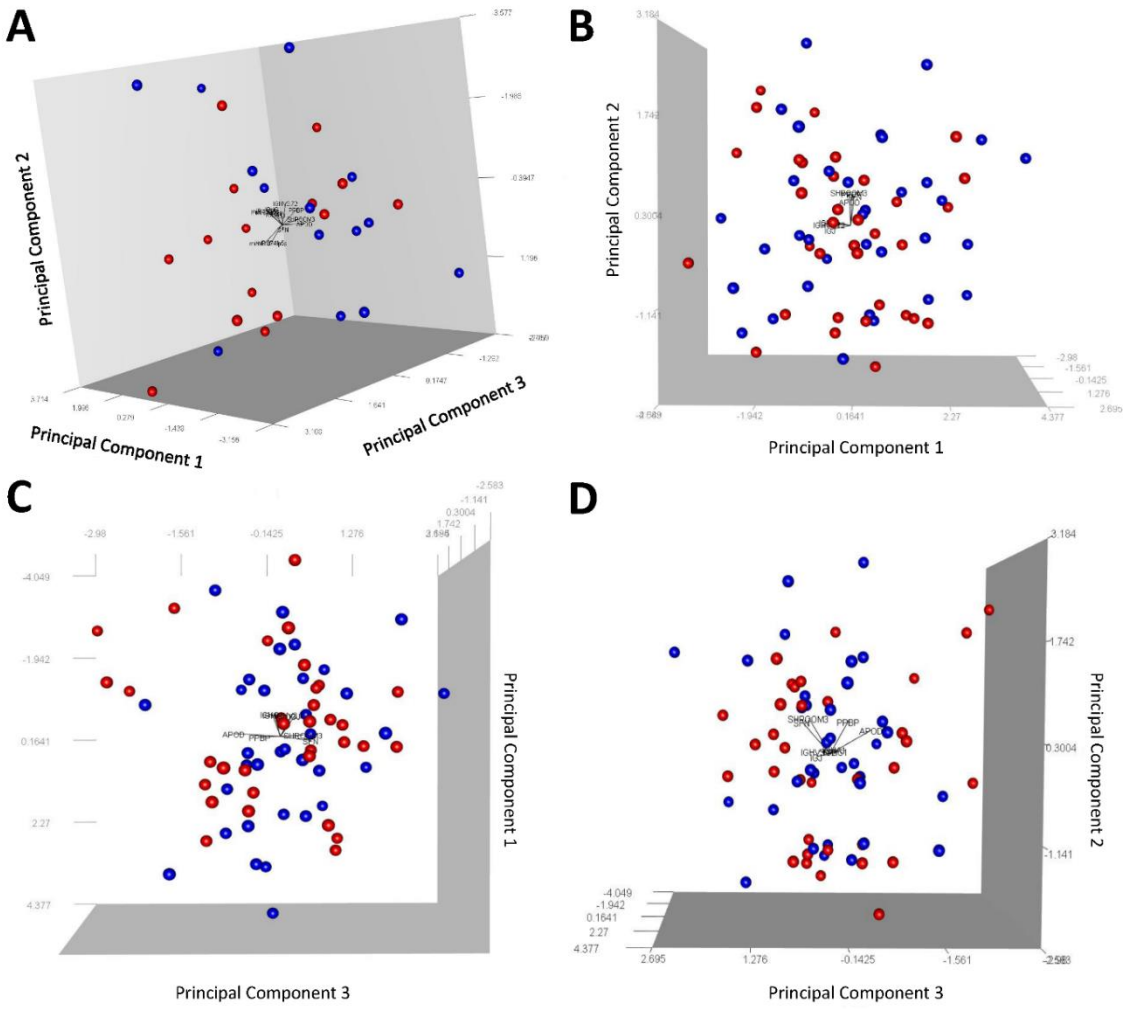


Figure 5.7 PCA analysis in 3-dimensions of 8-protein panel in serum shows poor distinction of risk groups. 3-D PCA using the 18-protein panel in serum, with biplot overlayed showing (**A**) principal components 1, 2 and 3, (**B**) principal components 1 and 2, (**C**) principal components 1 and 3, and (**D**) principal components 2 and 3. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding biomarker. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Data are presented for n=45 low-risk (blue) and n=23 high-risk (red) patients.

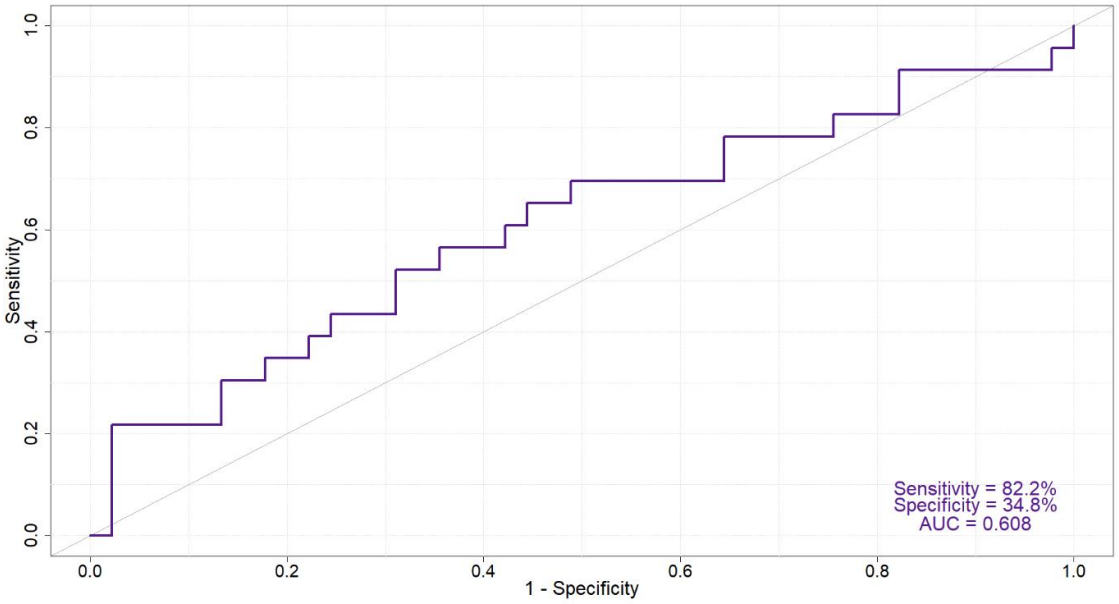


Figure 5.8 8-protein panel in serum classifies patients based on risk with AUC of 0.608. LOOCV of patients using the eight differentially expressed proteins. Data are presented for n=45 low-risk patients and n=23 high-risk patients.

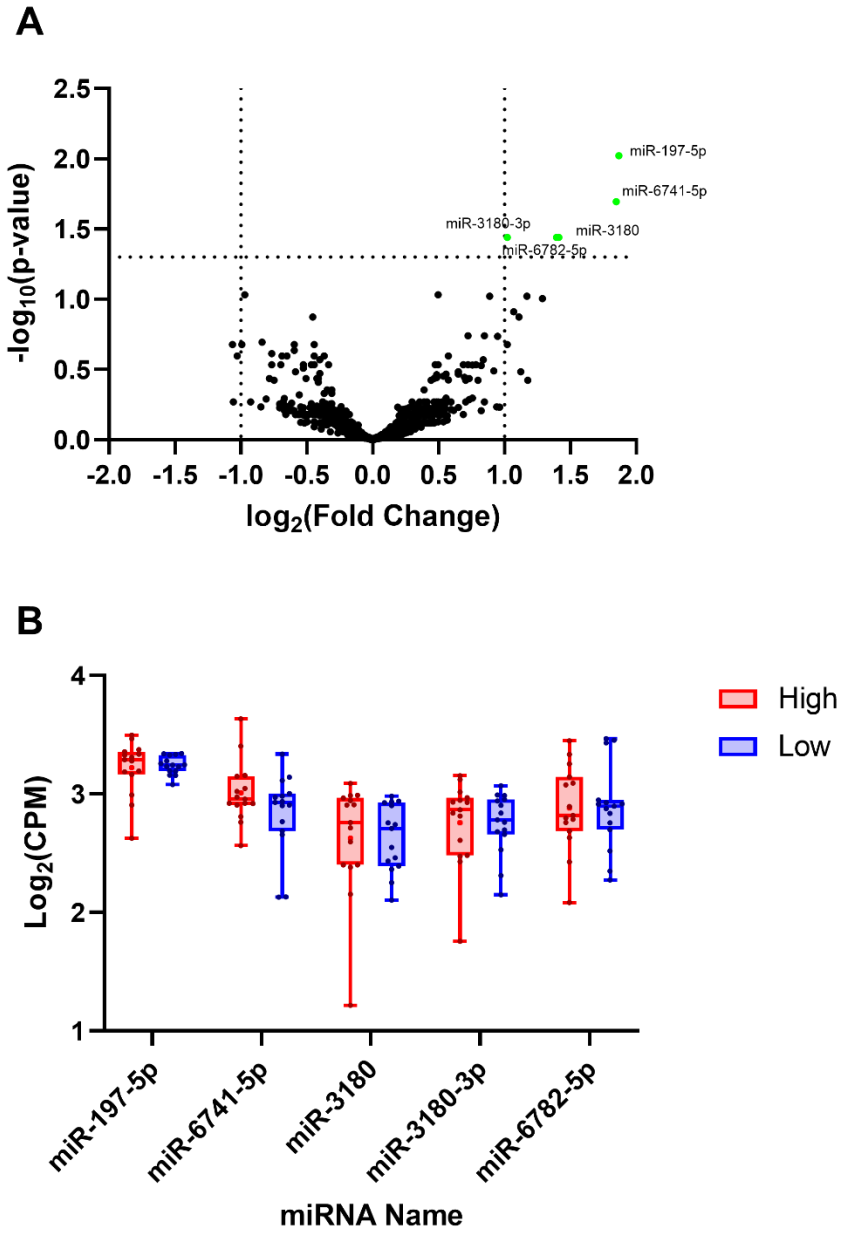


Figure 5.9 Five miRNA were identified as being significantly upregulated in high-risk patient serum compared to low-risk. (**A**) Differential expression analysis identified five miRNA that were differentially expressed between low- and high-risk serum samples (p<0.05, FDR=0.05, s0=0.1). MiRNA identified in green were considered significantly upregulated in high-risk serum compared to low-risk serum. Dotted lines indicate the fold change and significance cut-offs. (**B**) Box plots showing the distribution of patient expression levels, in Log2(counts per million), for each of the five significant miRNA. High-risk patients (n=15) are shown in red; low-risk patients (n=15) are shown in blue.

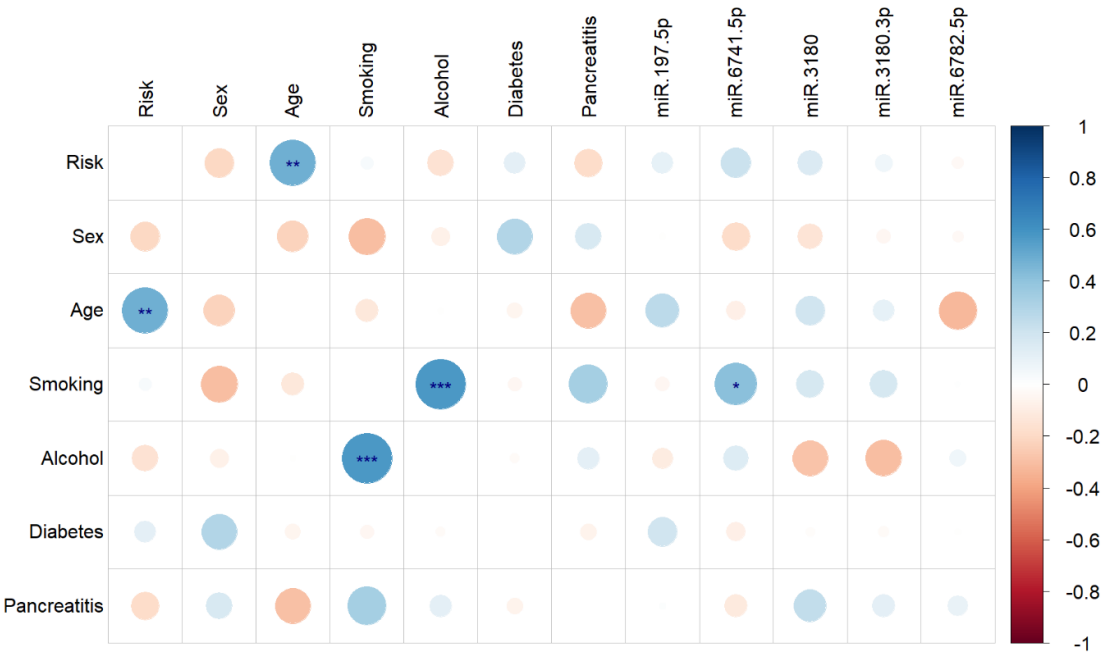


Figure 5.10 Expression of significant miRNA in the serum does not significantly correlate with clinical factors. Spearman correlations between patient clinical data and the five differentially expressed miRNA are given as a corrplot. Colour intensity relates to R^2 value; circle size relates to the p-value. Data are presented for n=15 low-risk and n=15 high-risk patients.*p<0.05, **p<0.01, ***p<0.001.

Several clinical factors were shown to correlate with each other in this patient cohort also. Increased risk significantly correlated with increased age (p<0.01), and increased smoking significantly correlated with increased alcohol consumption (p<0.001).

5.6.5 5-miRNA panel in serum stratifies patients with poor accuracy

These five miRNA were then integrated to create a 5-miRNA biomarker panel. This 5-miRNA panel was then used to stratify patients into distinct groups based on their expression of these miRNA. Unsupervised hierarchical clustering of patients into risk groups using this 5-miRNA panel was performed with an accuracy of 60%, with n=6 low-risk and n=6 high-risk patients being grouped incorrectly (Figure 5.11). Expression levels of miR-3180-3p and miR-3180 were highly related, as were the levels of miR-6741-5p and miR-6782-5p.

PCA was conducted using two components as this was found to account for 85.2% of the variance (45.2% in component 1 and 40% in component 2) (Appendix 8)(Figure 5.12). MiR-6741-5p, miR-3180-3p and miR-3180 were the most important contributors to the first component, with miR-6782-5p and miR197-5p being the most important contributors to the second component. MiR-3180-3p and miR-3180 were shown to be the most highly positively correlated among the five miRNA. MiR-197-5p was shown to have the least amount of variance associated with it, with the remaining four miRNA having similar variance. The ellipse encapsulating 80% of the low-risk classification sits almost entirely inside that of the high-risk classification, indicating poor separation of the two groups.

LOOCV was then carried out, where the model was trained and then tested using the same patient cohort. Here, the 5-miRNA panel produced an AUC value of 0.427, with a sensitivity of 46.7% and a specificity of 40.0%, demonstrating extremely poor risk classification in this cohort (Figure 5.13).

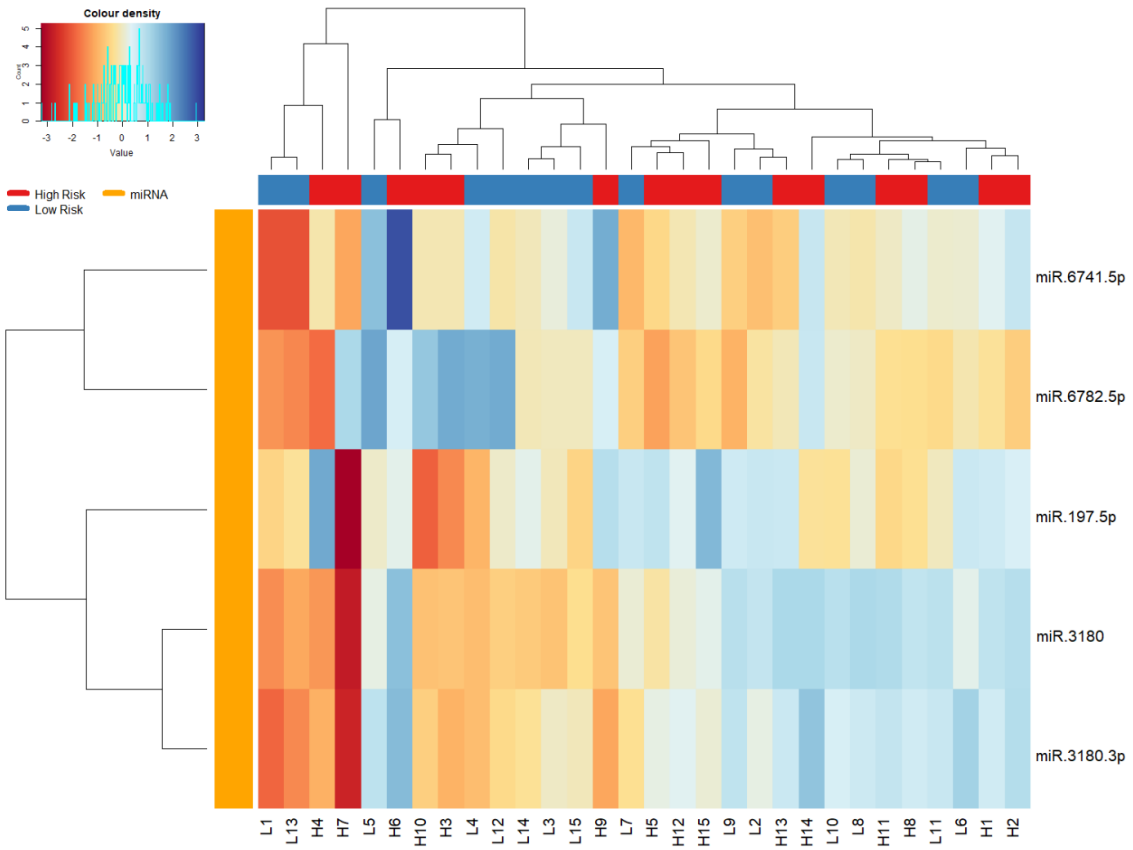


Figure 5.11 5-miRNA panel in serum clusters patients into risk groups with 60% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the five significant miRNA. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the miRNA (orange). Data are presented for n=15 low-risk and n=15 high-risk patients.

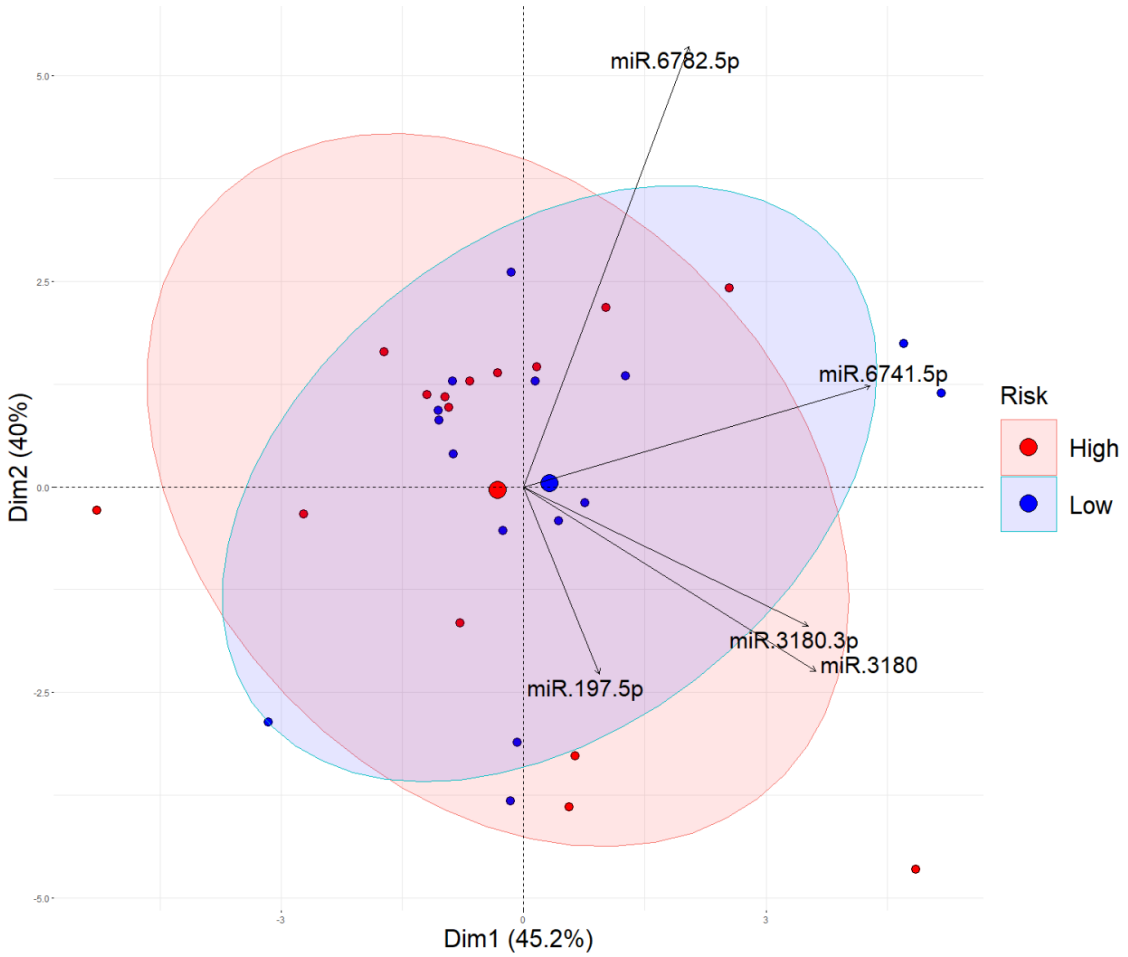


Figure 5.12 PCA analysis of 5-miRNA panel in serum shows poor distinction of the risk groups. 2-D PCA using the five differentially expressed miRNA, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding miRNA. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. Data are presented for n=15 low-risk (blue) and n=15 high-risk (red) patients.

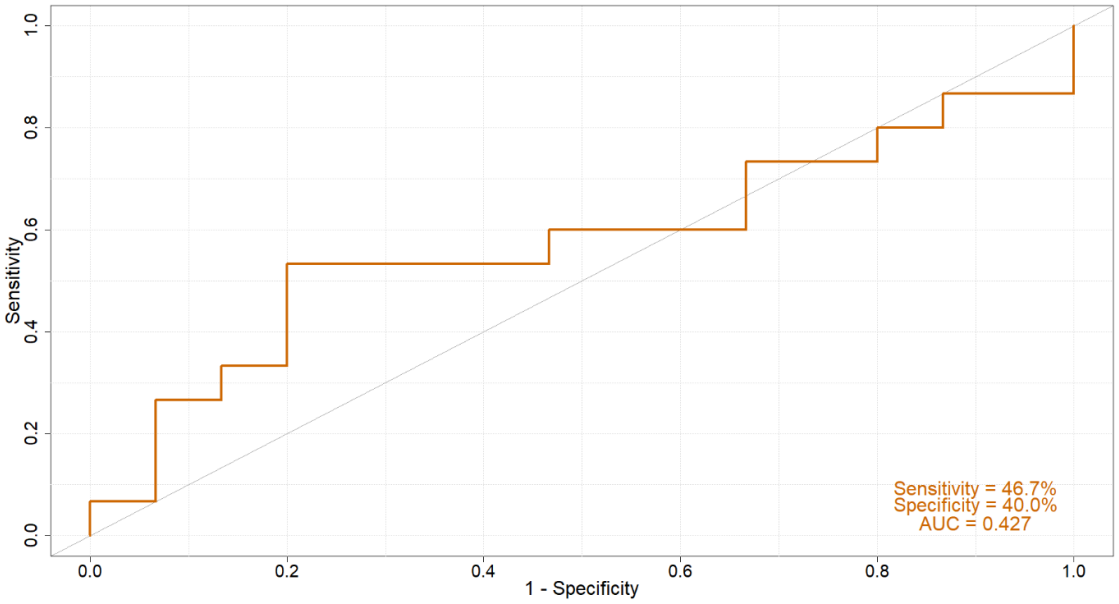


Figure 5.13 5-miRNA panel in serum classifies patients based on risk with AUC of 0.427. LOOCV of patients using the five differentially expressed miRNA. Data are presented for n=15 low-risk patients and n=15 high-risk patients.

5.6.6 13-feature multi-omic panel in serum stratifies patients with high accuracy

The eight proteins and five miRNA that were identified as differentially expressed were then scaled and integrated to create a 13-feature multi-omic biomarker panel. This 13-feature panel was then used to stratify patients into distinct groups based on their expression of these thirteen biomarkers. Unsupervised hierarchical clustering of patients into risk groups using this 13-feature panel was performed with an accuracy of 79.3% (Figure 5.14). The biomarker relatedness in this panel shows separation of the proteins and miRNA, however, the dendrogram indicates that the expression levels of SFN are more closely related to that of the miRNA than the other seven proteins, despite the proteins being decreased in high-risk serum and the miRNA being increased. Importantly, one patient in this cohort can be considered an outlier patient, as they have a diagnosis of VHL disease, which genetically pre-disposes one to the development of PC. Re-examining this panel with this patient reclassified as low-risk, as carried out previously in Chapter 4, the 13-feature panel produced a worse clustering accuracy of 75.9% (Figure 5.15).

Examining the correlations between the thirteen biomarkers within the panel further demonstrated the lack of correlations between these biomarkers (Figure 5.16). Indeed, there were no significant correlations between SFN, PPBP and SHROOM3 and any of the other ten biomarkers ($p>0.05$). As such, no individual biomarker significantly correlated with all others in the 13-feature panel, or with the others in their miRNA or protein sector. IGHV3-72 had the most correlations within the panel, significantly correlating with IGJ, IGHA1 and IGHG1 expression ($p<0.05$). IGHA1, IGHG1, miR-197-5p, miR-6741-5p and miR-6782-5p all significantly correlated with two other biomarkers within the panel, with IGJ, APOD, miR-3180 and miR3180-3p all significantly correlating with only one other biomarker ($p<0.05$). Two negative correlations between proteins and miRNA were observed, with IGHV3-72 significantly correlating with miR-6741-5p and APOD significantly correlating with miR-6782-5p ($p<0.05$). The strongest correlations were found between miR-197-5p and both miR-3180 and miR-3180-3p, as indicated by the thicker chords.

PCA was first conducted using two components, however, this was found to account for only 46.6% of the variance (25.1% in component 1 and 21.5% in component

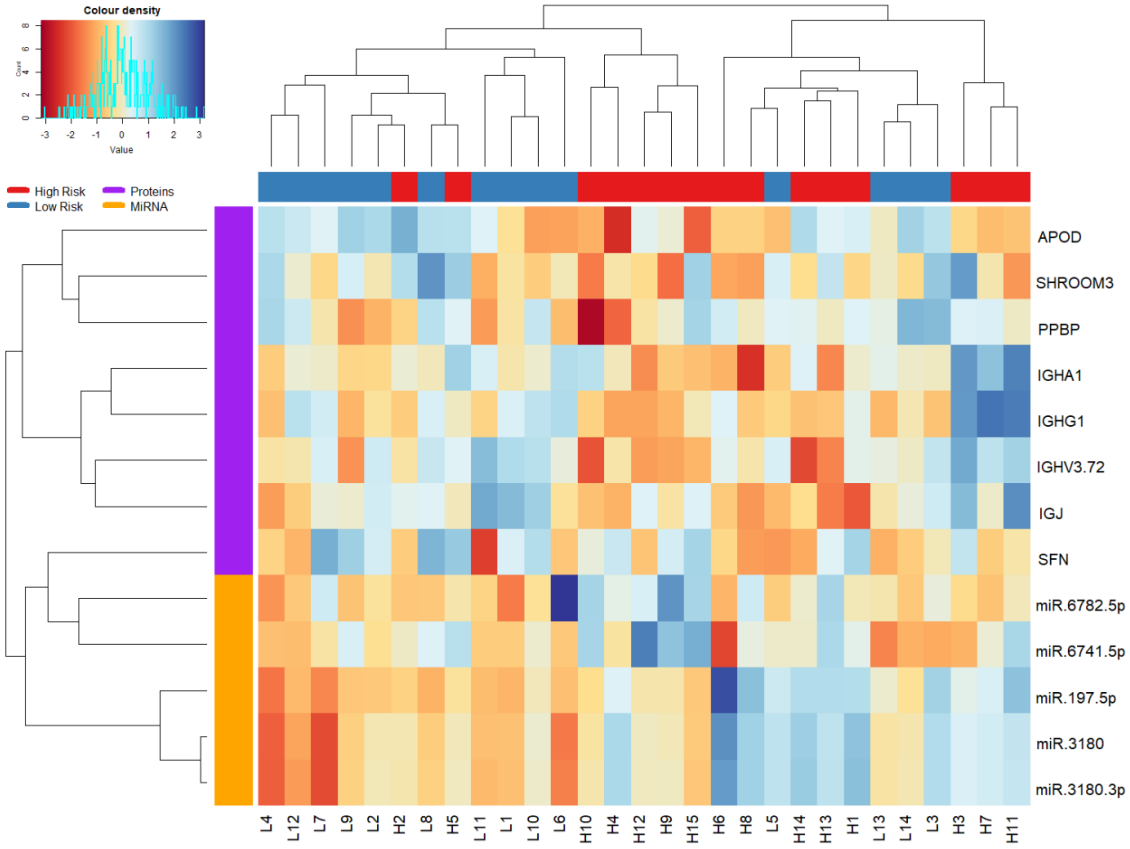


Figure 5.14 13-feature multi-omic panel in serum clusters patients into risk groups with 79.3% accuracy. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the 13-feature multi-omic panel. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins (purple) and miRNA (orange). Data are presented for n=14 low-risk and n=15 high-risk patients.

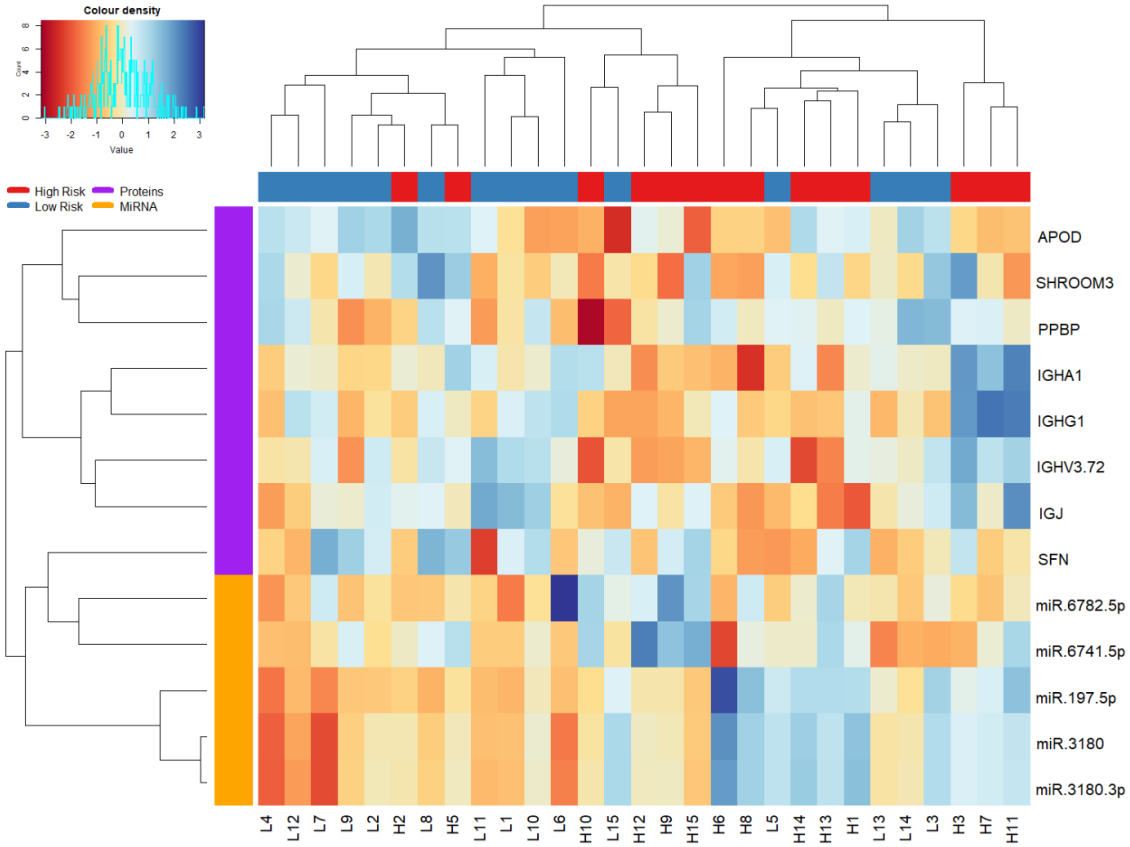


Figure 5.15 13-feature multi-omic panel in serum clusters patients into risk groups with 75.9% accuracy when VHL patient is reclassified. Unsupervised hierarchical clustering of patients into high-risk (red) and low-risk (blue) groups based on their expression of the 13-feature multi-omic panel. Dendrograms show (top) the relatedness of the patients, and (left) the relatedness of the proteins (purple) and miRNA (orange). VHL outlier patient has been reclassified from high-risk to low-risk. Data are presented for n=15 low-risk and n=14 high-risk patients.

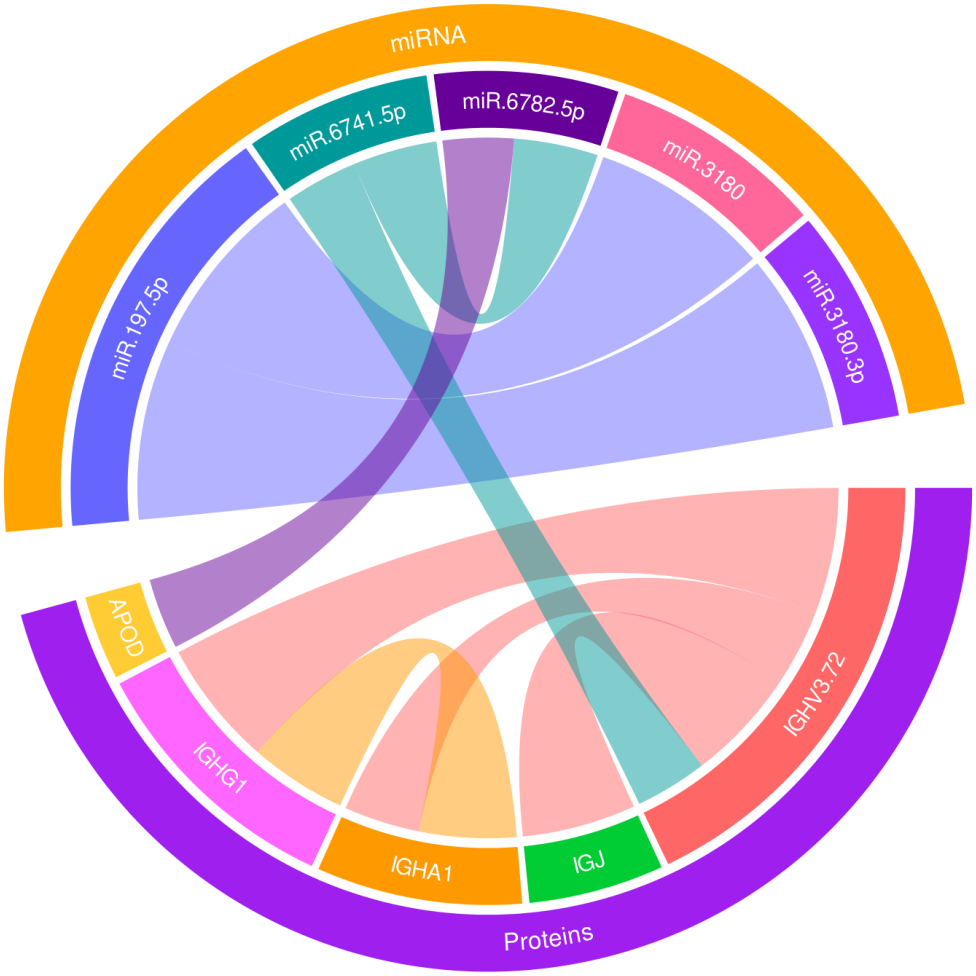


Figure 5.16 Ten biomarkers within the 13-feature biomarker panel in serum significantly correlate with expression of each other. Chord diagram showing significant spearman correlations (p<0.05) between five proteins (purple sector) and five miRNA (orange sector) within the 13-feature panel. Inner chords reflect correlations between the biomarkers. Chord thickness is directly related to the strength of the correlation, with thicker chords indicating stronger correlations. Chords within the proteins or miRNA segments indicate positive correlations. Chords between the proteins and miRNA segments indicate negative correlations.

2) (Appendix 9)(Figure 5.17). MiR-3180-3p, miR-3180, miR-197-5p, APOD, SFN, SHROOM3 and IGJ were the most important contributors to the first component, with miR-6741-5p, miR6782-5p, PPBP, IGHG1, IGHV3-72 and IGHA1 being the most important contributors to the second component. MiR-3180-3p, miR-3180 and miR-197-5p were shown to be highly positively correlated, as were miR-6741-5p and miR-6782-5p, SFN and APOD, and IGHG1 and IGHV3-72. MiR-3180-3p, miR-3180 and miR-197-5p had the largest amount of variance associated with them, with SFN, APOD and PPBP having the least. The ellipse encapsulating 80% of the low-risk classification was much smaller than that of the high-risk classification, indicating larger variance in the high-risk population. Importantly, when the VHL patient mentioned previously was reclassified, the separation and distribution of the datapoints does not change (Figure 5.18).

The third component in the PCA accounted for 14.8% of the variance. As such, when examined with a third component in 3-dimensions, the total variance accounted for is just 61.4% (Figure 5.19). With three components, the separation of the two groups is modest, though not distinct (Figure 5.19A). The view of components 1 and 2 provides a mirror image of the 2-D PCA plot, showing the same trends in the proteins and miRNA (Figure 5.19B). Examining components 1 and 3 shows that APOD, SHROOM3, PPBP, miR-6782-5p and miR-6741-5p account for the majority of variance in the third component (Figure 5.19C). This can be seen again when comparing components 2 and 3 (Figure 5.19D).

LOOCV was then carried out, where the model was trained and then tested using the same patient cohort. Here, the 13-feature multi-omic panel produced an AUC value of 0.824, with a sensitivity of 71.4% and a specificity of 80.0% (Figure 5.20). Despite the poor performance of the 5-miRNA panel, and modest performance of the 8-protein panel, the integration of these two panels together improved the overall performance. However, while both the 8-protein and 5-miRNA panels had improved sensitivity compared to specificity, the 13-feature multi-omic panel had the reverse, with higher specificity compared to sensitivity. Importantly, when the VHL outlier patient was reclassified to low-risk as before, the performance of the 5-miRNA panel and the 13-feature multi-omic panel improved, though the 8-protein panel performance was worse (Figure 5.21). Indeed, the 8-protein panels AUC value decreased from 0.608 to 0.561,

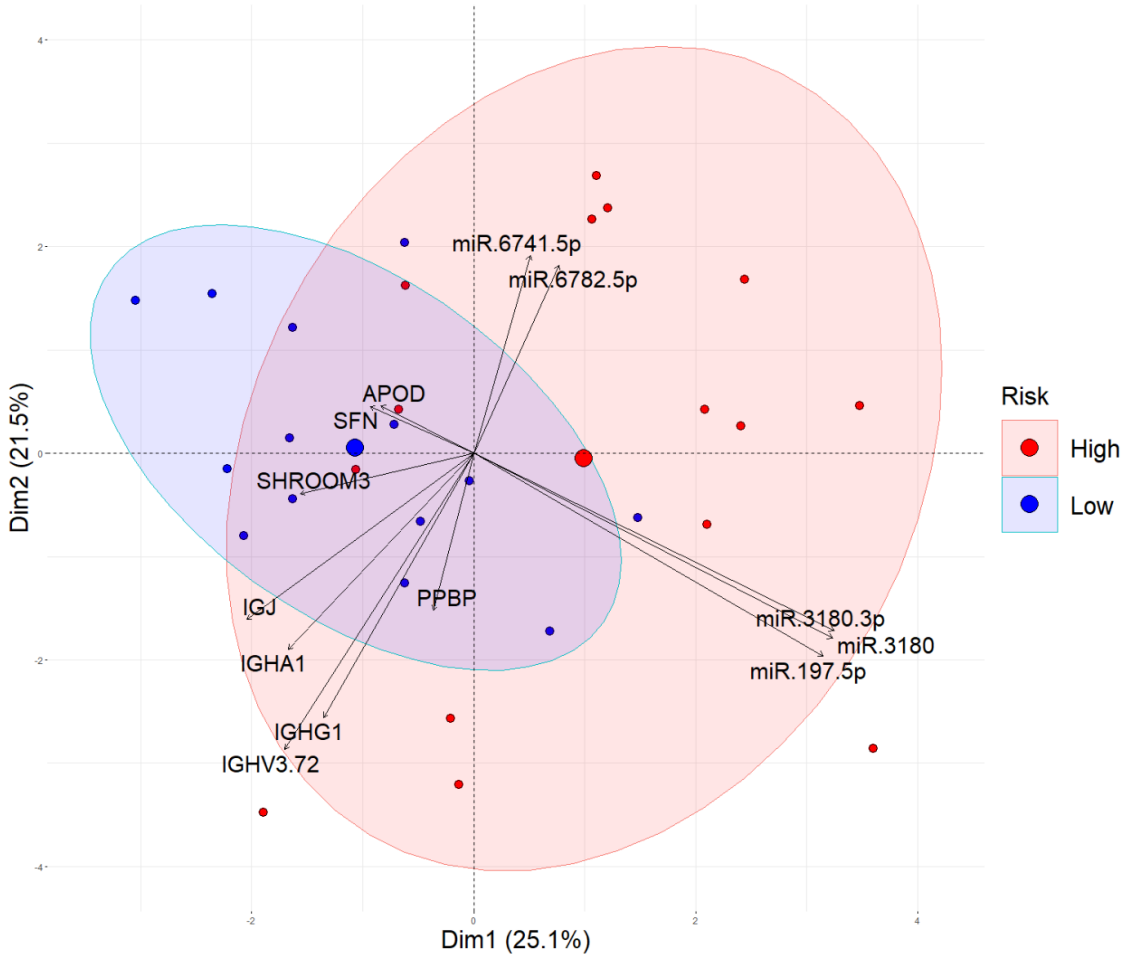


Figure 5.17 PCA analysis of 13-feature multi-omic panel in serum shows modest distinction of the risk groups. 2-D PCA using the 13-feature multi-omic panel in serum, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding biomarker. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. Data are presented for n=14 low-risk (blue) and n=15 high-risk (red) patients.

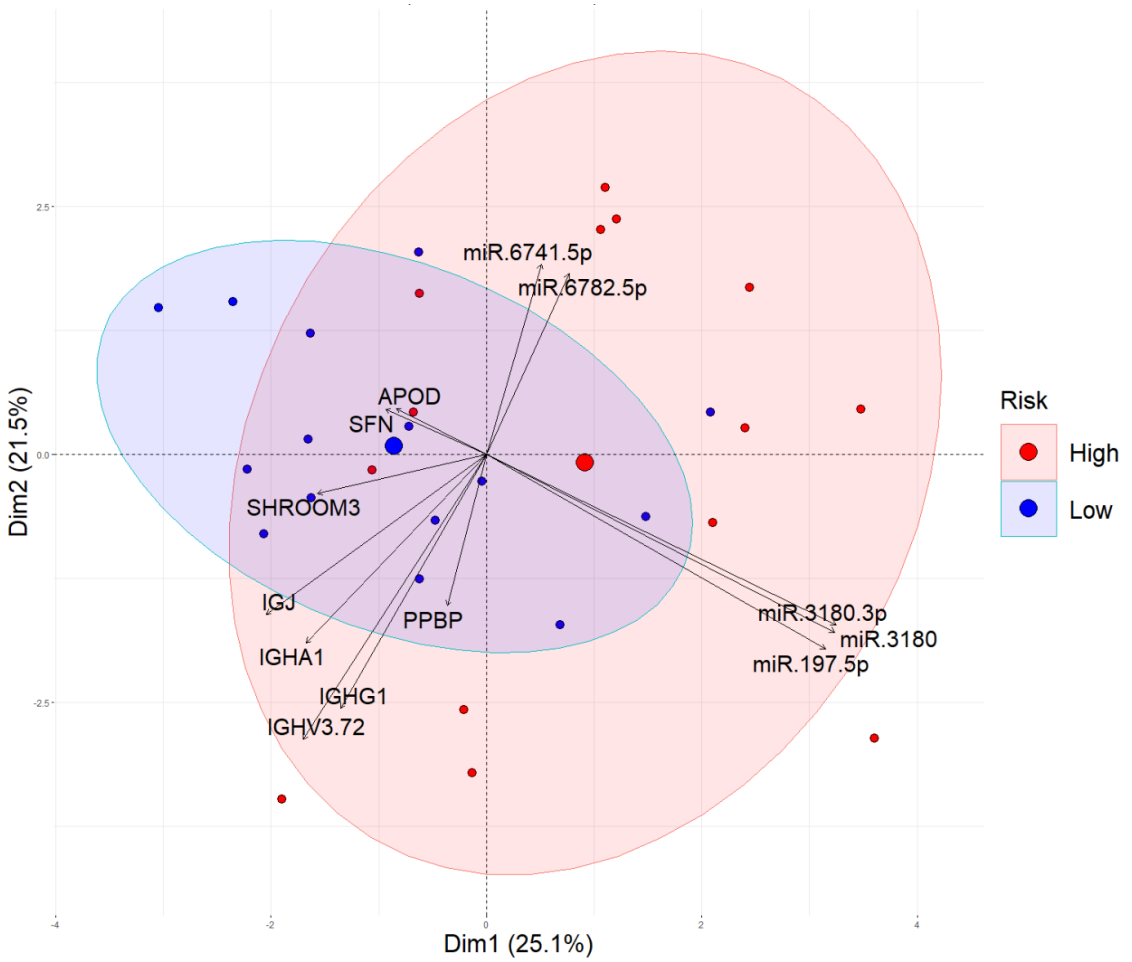


Figure 5.18 PCA analysis of 13-feature multi-omic panel in serum shows no change in the distinction of the risk groups when VHL patient is reclassified. 2-D PCA using the 13-feature multi-omic panel in serum, with biplot overlayed. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding biomarker. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Ellipses represent 80% of the data captured within the risk classifications. VHL outlier patient has been reclassified from high-risk to low-risk. Data are presented for n=15 low-risk (blue) and n=14 high-risk (red) patients.

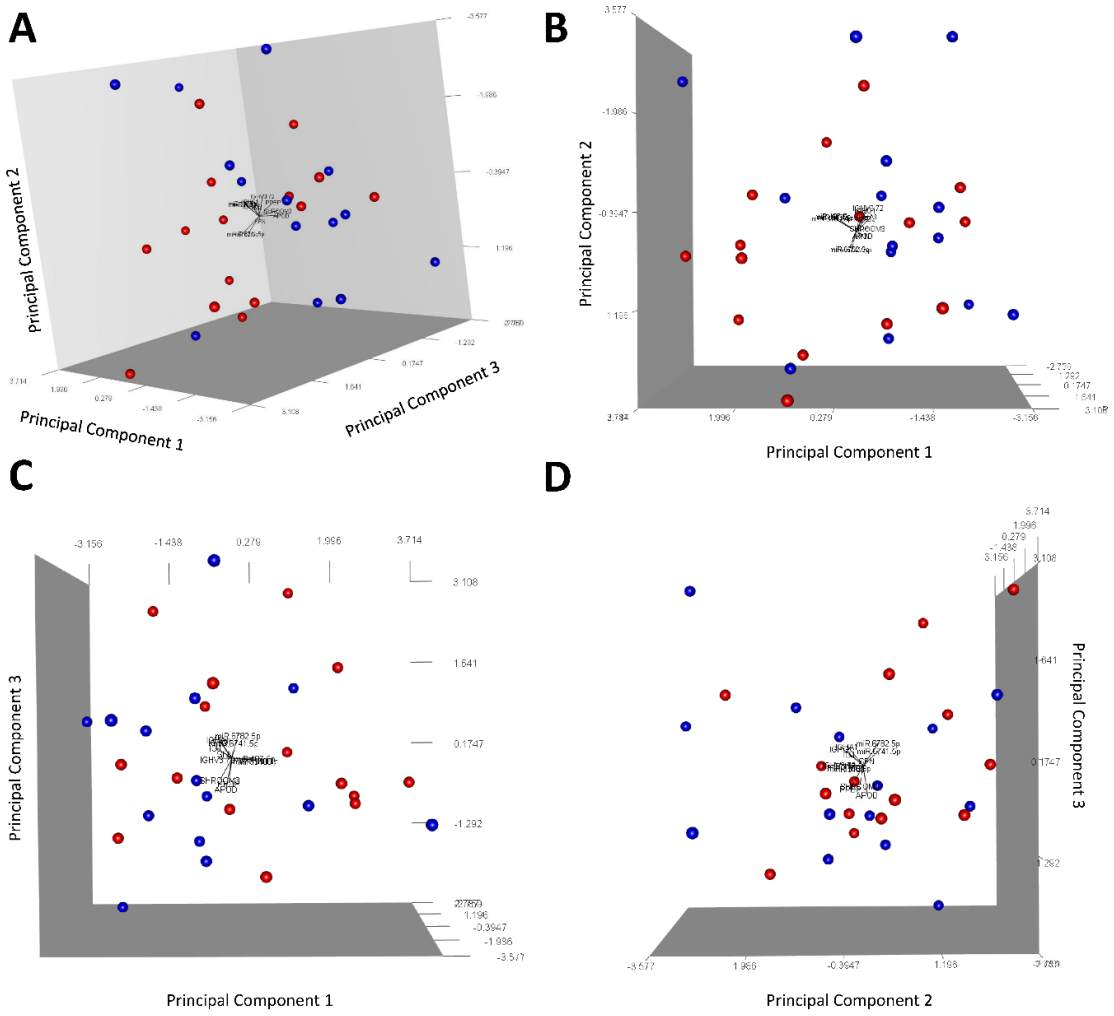


Figure 5.19 PCA analysis in 3-dimensions of 13-feature multi-omic panel in serum shows modest distinction of risk groups. 3-D PCA using the 13-feature multi-omic panel in serum, with biplot overlayed showing (**A**) principal components 1, 2 and 3, (**B**) principal components 1 and 2, (**C**) principal components 1 and 3, and (**D**) principal components 2 and 3. Biplot scale is set to zero to ensure vectors (arrows) are scaled to represent their respective loadings. The length of each vector is proportional to the variance of the corresponding biomarker. The smaller the angle between a vector and a principal component axis, i.e. the more parallel they are, the more it contributes to that component. Small angles between vectors indicate high positive correlations; right angles represent no correlation; opposite angles indicate high negative correlations. Data are presented for n=15 low-risk (blue) and n=14 high-risk (red) patients.

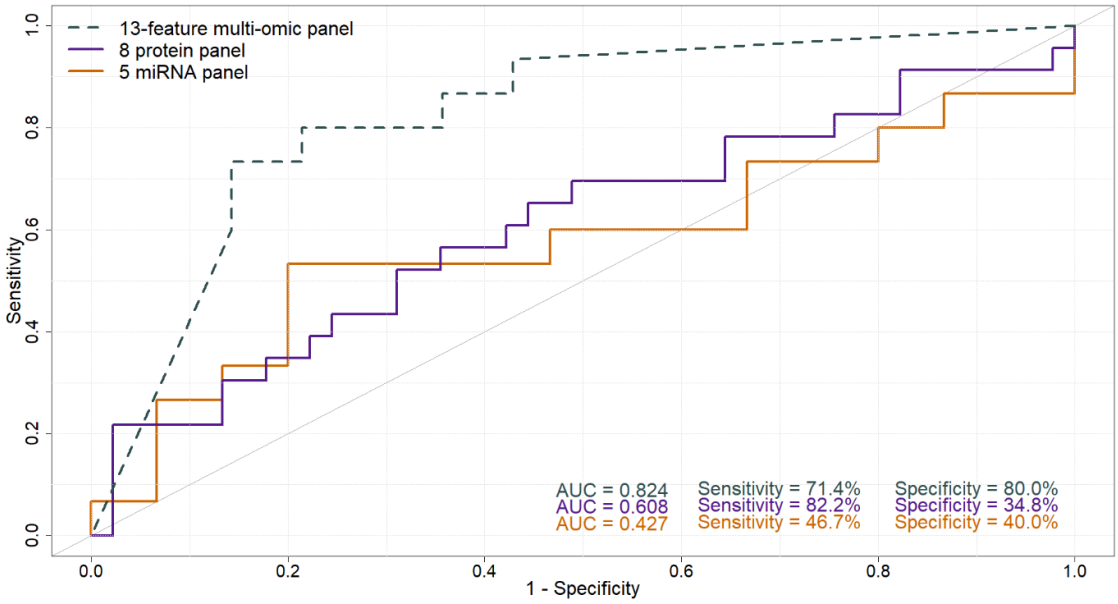


Figure 5.20 13-feature multi-omic panel in serum classifies patients based on risk with AUC of 0.824. LOOCV of patients using the (dashed grey line) 13-feature multi-omic panel, (orange line) 5-miRNA panel, and (purple line) 8-protein panel. 13-feature panel data are presented for n=14 low-risk patients and n=15 high-risk patients; 5-miRNA panel data are presented for n=15 low-risk patients and n=15 high-risk patients; 8-protein panel data are presented for n=45 low-risk patients and n=23 high-risk patients.

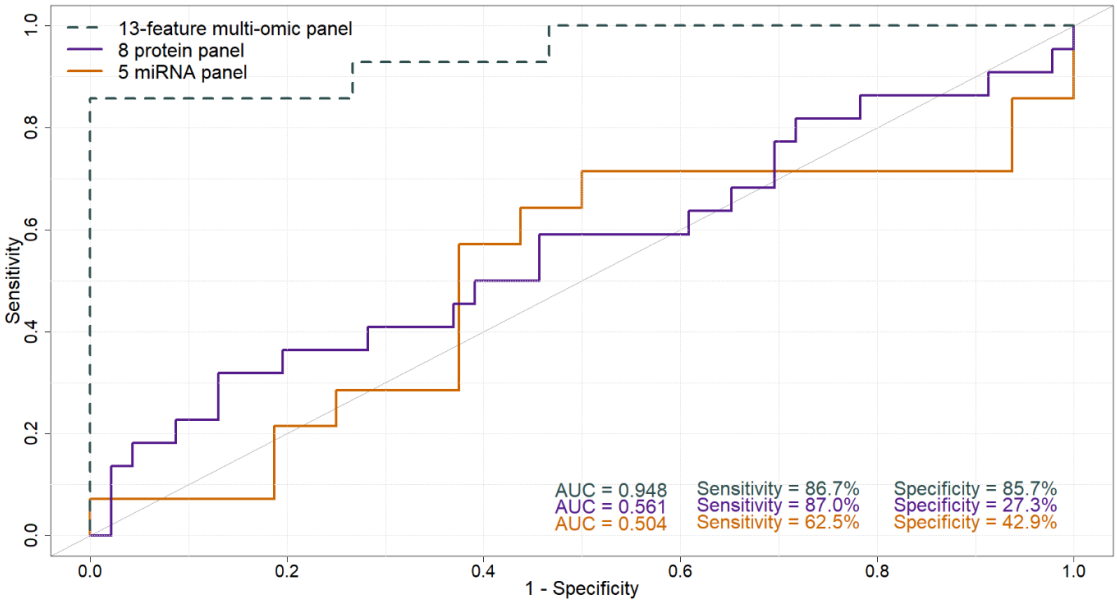


Figure 5.21 13-feature multi-omic panel in serum classifies patients based on risk with AUC of 0.948 when VHL patient is reclassified. LOOCV of patients with VHL patient reclassified to low-risk, using the (dashed grey line) 13-feature multi-omic panel, (orange line) 5-miRNA panel, and (purple line) 8-protein panel. 13-feature panel data are presented for n=15 low-risk patients and n=14 high-risk patients; 5-miRNA panel data are presented for n=16 low-risk patients and n=14 high-risk patients; 8-protein panel data are presented for n=46 low-risk patients and n=22 high-risk patients.

while the 3-miRNA panel improved its AUC value from 0.427 to 0.504, and the 13-feature multi-omic panel improved its AUC value from 0.824 to 0.948. Furthermore, in the reclassified cohort, the 13-feature multi-omic panel had improved sensitivity (86.7%) compared to specificity (85.7%), further emphasising how the classification of this patient affects the performance of the panel.

qPCR microarrays were used in an attempt to validate the results obtained from the transcriptomic analysis in the serum, however, while 10 samples (n=5 high-risk and n=5 low-risk) were run on this array most did not amplify and as such no statistical analysis was possible (Appendix 10).

5.7 Results: Part Two

5.7.1 10-feature multi-omic cross-biofluid panel stratifies patients with high accuracy

In order to utilise all potential biomarkers, the 11-feature multi-omic panel in PCF that was identified in Chapter 4 was revisited. Here, the performance of every possible combination of these 11 features was examined using CombiROC software with sensitivity and specificity cut-offs of 83% and 25%, respectively (Table 5.3). Twenty combinations of these features were observed to meet these cut-offs, with four being the minimum number of features required to do so. The same analysis was run on the 13-feature multi-omic serum panel, with sensitivity and specificity cut-offs of 93% and 47%, respectively. Thirty-six combinations were identified that met these criteria, with six being the minimum number of features required to do so. The top performing combination with the minimum number of features was determined using CombiROC software for both the PCF and the serum. A 4-feature PCF panel consisting of S100A8, LGALS3, SNORA66 and miR-216b-5p, and a 6-feature serum panel consisting of IGHV3-72, IGJ, IGHA1, PPBP, miR-3180 and miR-3180-3p were identified. These two panels were then scaled and integrated to create a 10-feature multi-omic cross-biofluid panel, and examined in a matched patient cohort. LOOCV was then carried out, where the model was trained and then tested using the same patient cohort. Here, the 10-feature multi-omic cross-biofluid panel produced an AUC value of 0.927, with a sensitivity of 83.3% and a specificity of 91.2% (Figure 5.22). The 4-feature multi-omic PCF panel

Table 5.3 CombiROC reduction creates a 10-feature multi-omic cross-biofluid panel.

	PCF panel	Serum panel
Software cut-offs		
Test signal	3	4
Sensitivity	83%	93%
Specificity	25%	47%
No. of combinations at cut-offs	20 combinations	36 combinations
No. of combinations with:		
4 features	4	0
5 features	9	0
6 features	6	4
7 features	1	12
8 features	0	13
9 features	0	6
10 features	0	1
10+ features	0	0
Top performing combination with minimum number of features	S100A8, LGALS3, SNORA66, miR-216b-5p	IGHV3-72, IGJ, IGHA1, PPBP, miR-3180, miR-3180-3p

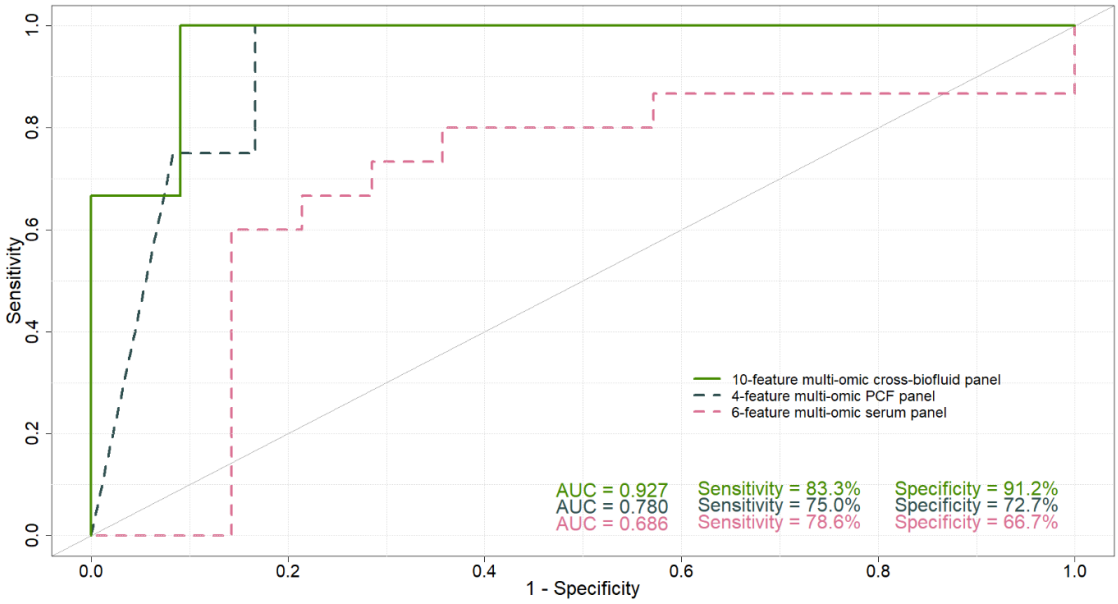


Figure 5.22 10-feature multi-omic cross-biofluid panel classifies patients based on risk with AUC of 0.927. LOOCV of patients using the (green line) 10-feature multi-omic cross-biofluid panel, (dashed grey line) 4-feature multi-omic PCF panel, and (dashed pink line) 6-feature multi-omic serum panel. 10-feature multi-omic cross-biofluid panel data are presented for n=11 low-risk patients and n=12 high-risk patients; 4-feature multi-omic PCF panel data are presented for n=12 low-risk patients and n=12 high-risk patients; 6-feature multi-omic serum panel data are presented for n=14 low-risk patients and n=15 high-risk patients.

performed better than the 6-feature multi-omic serum panel, with an AUC of 0.780, a sensitivity of 75% and a specificity of 72.7%. The 6-feature multi-omic serum panel had the worst performance, with an AUC value of 0.686, a sensitivity of 78.6% and a specificity of 66.7%. Again, while both the 4-feature multi-omic PCF panel and the 6-feature multi-omic serum panels had improved sensitivity compared to specificity, the 10-feature multi-omic cross-biofluid panel had the reverse, with higher specificity compared to sensitivity. Importantly, when the VHL outlier patient was reclassified to low-risk as before, the performances of both the 4-feature multi-omic PCF panel and the 6-feature multi-omic serum panel improved, while the 10-feature multi-omic cross-biofluid panel performed worse (Figure 5.23). Indeed, the 4-feature multi-omic PCF panels AUC value increased from 0.780 to 0.846, and the 6-feature multi-omic serum panels AUC value increased from 0.686 to 0.695, while the 10-feature multi-omic cross-biofluid panels AUC value dropped from 0.927 to 0.909. Furthermore, in the reclassified cohort, the 10-feature multi-omic cross-biofluid panel had improved sensitivity (83.3%) compared to specificity (63.6%).

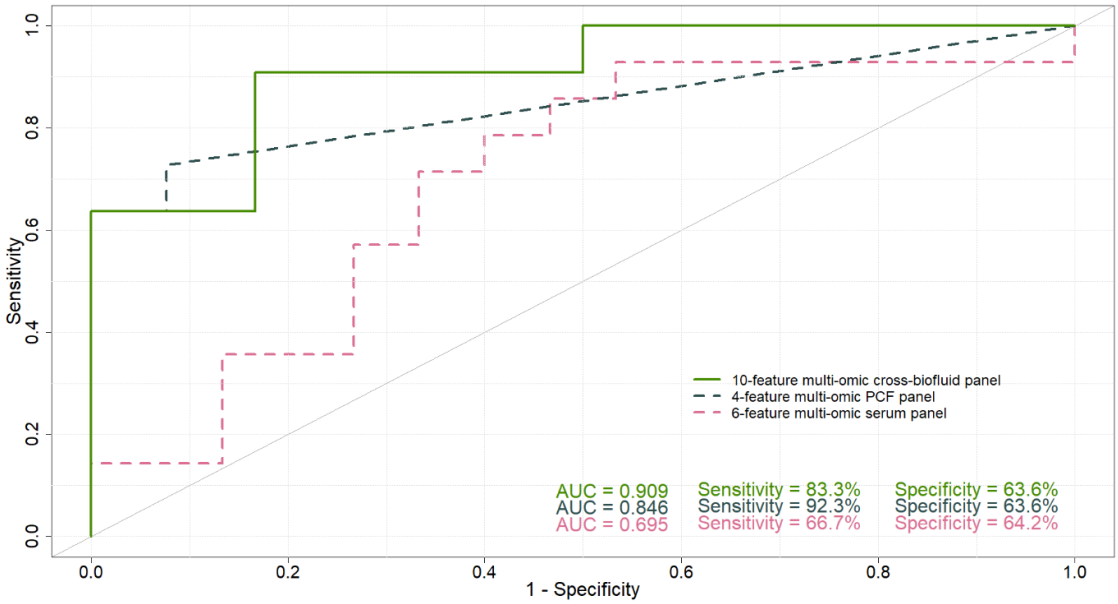


Figure 5.23 10-feature multi-omic cross-biofluid panel classifies patients based on risk with AUC of 0.909 when VHL patient is reclassified. LOOCV of patients with VHL patient reclassified to low-risk, using the (green line) 10-feature multi-omic cross-biofluid panel, (dashed grey line) 4-feature multi-omic PCF panel, and (dashed pink line) 6-feature multi-omic serum panel. 10-feature multi-omic cross-biofluid panel data are presented for n=11 low-risk patients and n=12 high-risk patients; 4-feature multi-omic PCF panel data are presented for n=12 low-risk patients and n=12 high-risk patients; 6-feature multi-omic serum panel data are presented for n=14 low-risk patients and n=15 high-risk patients.

5.8 Discussion

The cross-biofluid integration of both multi-omic panels from PCF and serum produced the highest classification accuracy of any panel examined in Chapters 4 and 5, without reclassifying the outlier VHL patient. Indeed, while the integration of proteomic and transcriptomic biomarkers to create a multi-omic panel in the serum produced substantially better risk stratification than either omic level alone, it is clear from these results that the layering of data from multiple biological levels could be the key to the generation of more robust biomarkers.

While from a numeric perspective, it may be unsurprising that the 8-protein panel performed better than the 5-miRNA panel in the unsupervised hierarchical clustering and LOOCV of these patient cohorts, it is important to remember that not all of the eight proteins brought forward were in fact significantly differentially expressed. Indeed, only two of the eight proteins were significant, which is in contrast to the miRNA, all five of which were found to be significantly differentially expressed. Also important to note, is the starting amount of proteins and miRNA. Differential expression analysis was run on over two thousand miRNA, but only 145 proteins. It may be reasonable, therefore, to assume that the significant proteins could be more strongly differentially expressed than the miRNA. Indeed, when looking at the correlations of these biomarkers with clinical factors, one protein (SHROOM3) significantly correlated with risk, while no miRNA correlated with this factor. Looking into the literature, it is interesting to see that there is no evidence of the dysregulation or functionality of any of the five significant miRNA in pancreatic disease. MiR-197-5p, for example, is known to be significantly dysregulated in a number of diseases and many cancer types, such as ovarian cancer, lung cancer, colorectal cancer, prostate cancer and thyroid cancer, and it has been shown to play a role in cell proliferation, differentiation, apoptosis, metastasis and drug resistance[427]. Similarly, miR-6741-5p has been shown to be increased in the serum of prostate cancer patients compared to patients with negative prostate biopsies, and as part of an 18-miRNA panel could distinguish prostate cancer with a sensitivity and specificity of 90%[428]. MiR-6741-5p expression levels in plasma have also been indicated as a promising prognostic biomarker in COVID-19 patients[429]. MiR-3180 has been suggested as a potential biomarker of hepatitis B virus infection persistence[430], while

miR-3180-3p was shown to be significantly upregulated in the serum of chemotherapy (cisplatin) resistant gastric cancer patients compared to chemotherapy sensitive patients, and significantly correlated with high TNM stage[431]. MiR-3180-3p has also been suggested as a potential plasma-based biomarker for the distinction of primary versus recurrent glioblastoma, demonstrating an AUC value of 98.5%[432]. Lastly, miR-6782-5p has been shown to be present at levels 2.3-fold higher in the seminal plasma of men with testicular germ cell tumours compared to healthy controls[433], and has also been found to be downregulated in the serum exosomes of gastric cancer patients compared to healthy controls[434]. Interestingly, miR-6782-5p levels were also detected at significantly lower levels in the serum of women with a high-risk of developing ovarian cancer, compared to those at a lower risk[435]. While all five of these miRNA have shown utility as biomarkers in other cancer types or diseases, none have been previously reported as being differentially expressed in pancreatic patients, further adding to the idea that their utility in this setting may be limited.

The differentially expressed proteins, on the other hand, appear more frequently in pancreatic literature. Unfortunately, however, a review of this literature highlights potential conflicting results with this study. While here, all eight proteins were found to be downregulated in high-risk serum compared to low-risk, just one of the eight proteins has been previously demonstrated to be decreased in PC compared to controls. IGHA1 plays a key role in immunoglobulin receptor binding activity, and has been measured previously in the PCF of both chronic pancreatitis and non-pancreatitis patients[436]. Importantly, IGHA1 has also been measured in normal pancreatic FFPE tissue samples, but was not detected in chronic pancreatitis or PC FFPE tissue specimens[437]. Of the remaining seven proteins, only one other was also found to be decreased in PC, however, the literature surrounding this protein is inconsistent. PPBP, also known as CXCL7, is a neutrophil chemoattractant that has been demonstrated by Matsubara *et al.* (2011) to be significantly decreased in the plasma of PC patients compared to healthy controls[438]. However, Pan *et al.* (2011) found PPBP levels to be elevated in chronic pancreatitis and PDAC plasma compared to healthy controls[439]. In a 2021 study, Kim *et al.* generated a plasma-based multi-biomarker panel consisting of 14 proteins, including PPBP, that could distinguish PDAC from controls with AUC values of up to 0.977[440].

Here, Kim *et al.* found PPBP levels to be increased in PDAC patients compared to controls[440]. Overall these three studies highlight the importance of validation of biomarkers across independent patient cohorts, and the vast differences that can be seen in biomarker expression profiles across different patient cohorts. IGJ, for example, was identified previously via proteomic evaluation of pancreatic patient plasma as being upregulated in PDAC plasma compared to healthy controls[439]. However, in this chapter IGJ was shown to be downregulated in high-risk serum compared to low-risk. Three more of the eight proteins were also found to be elevated in PC compared to controls. APOD, for one, is a glycoprotein that was found to be upregulated in PC tissue compared to normal controls[441]. It has also been demonstrated to be present at higher levels in the serum of PDAC patients with shorter survival[442]. IGHG1 is an important functional isoform of immunoglobulins that plays a role in immune effector cell cytolytic activities, which has also been shown to be increased in PC tissue compared to adjacent non-cancerous tissue[443], as well as in the tissue of chronic pancreatitis patients compared to normal controls[444]. Li *et al.* (2021) demonstrated that IGHG1 expression is increased in gastric cancer tissues compared to normal controls, and that IGHG1 expression promoted proliferation, migration and invasion in gastric cancer cell lines[445]. Lastly, SFN is a cell cycle checkpoint protein whose expression levels in tissue have been shown to correlate positively with poor prognosis in ovarian cancer[446]. Robin *et al.* (2020) demonstrated that SFN was significantly increased in the stroma of PDAC tissue compared to the healthy adjacent fibrous tissue[447]. A computational study, combining information from multiple online platforms, found SFN expression to be higher in PC tissues compared to normal tissues, with this study also revealing a negative correlation between SFN expression and survival probability in PC patients[448]. While evidence in the literature suggests the dysregulation of these six proteins in pancreatic disease, whether in line with the results of this chapter or not, it is interesting to note that the two proteins that were significantly differentially expressed, are the only two among the eight that do not appear in this literature. SHROOM3 is an actin-associated protein that is known to regulate epithelial cell shape and tissue morphogenesis[449]. While the *SHROOM3* gene has been shown to be strongly associated with renal disease, no evidence could be found in the literature for the dysregulation of this protein in pancreatic disease[450]. Similarly, while there have been no studies to-date reporting on

IGHV3-72 in pancreatic disease, the levels of this protein in plasma exosomes have been shown to have utility in distinguishing lung adenocarcinoma from lung squamous cell carcinoma[451]. As such, data are presented here for the first time indicating the dysregulation of SHROOM3, IGHV3-72, miR-197-5p, miR-6741-5p, miR-3180, miR-3180-3p and miR-6782-5p in pancreatic disease. Although all biomarkers need to emerge into the literature at some point, it is surprising to have so many identified in one patient cohort, with seven out of thirteen biomarkers brought forward in this chapter having not been studied in this context previously. However, given the thousands of 'unique' novel biomarkers seen in Chapter 3 that had been identified as blood-based biomarkers for PC diagnosis, it is not unusual for many of these features to be forgotten and/or not rediscovered as research progresses. Indeed, three of those 'promising' biomarkers from Chapter 3 that were detected in the serum of this patient cohort were not significantly different between low- and high-risk patients and showed limited utility in this setting. It should be, therefore, the primary concern of biomarker discovery literature, to narrow the field of view to just those biomarkers that add robustness to a panel, while simultaneously excluding those that worsen its performance.

Here, CombiROC software was used to interrogate both the 11-feature multi-omic panel in PCF identified in Chapter 4, and the 13-feature multi-omic panel in the serum discussed in this chapter. Using appropriate cut-offs, these panels were reduced down to the least number of biomarkers that would still produce highly sensitive results in order to allow the integration of these biomarkers to create a cross-biofluid panel. In the same way that multi-omic panels have the potential to encapsulate the complexity of disease, and have the unique ability to control for the dysregulation of one omic level or factor via compensation of other biomarkers within the panel[166], cross-biofluid panels present a new and exciting progression from this. By selectively reducing the PCF and serum panels, and generating a new panel that consists of two omic levels, as well as two distinct biofluids from the same patient, high sensitivity, specificity and AUC metrics were achieved. Furthermore, the performance of this panel is not improved by the reclassification of the VHL patient, further emphasising its utility in this setting. Interestingly, while the sensitivity cut-off was higher in the serum panel (93%) compared to the PCF panel (83%), more features were required to achieve this high performance

in the serum than in the PCF, perhaps indicating that PCF-based biomarkers perform better than serum-based biomarkers in this cohort, which would not be unexpected given the direct proximity of PCF to the PCL in question. Also of note, is the ratio of proteins to miRNA in these two reduced panels, with two proteins (S100A8 and LGALS3) and two miRNA (SNORA66 and miR-216b-5p) making up the PCF panel, while four proteins (IGHV3-72, IGJ, IGHA1 and PPBP) and two miRNA (miR-3180 and miR-3180-3p) making up the serum panel. Again, it is evident that the proteins in the serum have outperformed the miRNA, despite three of the four included proteins being non-significantly differentially expressed. Indeed, just one of the two significant proteins were included in the final panel of biomarkers, with SHROOM3, the only significant protein in the serum that also significantly correlated with patient risk, not being included in the reduced panel. Importantly, when examining the results from the PCA and correlations, both in the proteins and miRNA alone, and as a multi-omic panel, no trends emerge that align with those biomarkers that were brought forward in the cross-biofluid panel. Within the proteins, two distinct groups could be seen both in the unsupervised hierarchical clustering and within the PCA, however, of the four proteins brought forward into the cross-biofluid panel, three are from one group (IGHV3-72, IGJ and IGHA1) and one is from the other (PPBP). Furthermore, while IGHV3-72 expression was shown to correlate with IGJ and IGHA1 expression levels, PPBP expression levels did not correlate with any other biomarker. Indeed, in the 3-D PCA of the 13-feature serum panel, PPBP was shown to be among those biomarkers most associated with the third component. Finally, while IGHV3-72 expression significantly correlated with one miRNA (miR-6741-5p), this miRNA was not among those brought forward. MiR-3180 and miR-3180-3p were shown to be closed related, however, they both had stronger correlations with miR-197-5p than with each other. Again, these results highlight the importance of evaluating candidate biomarkers in the most appropriate ways. While unsupervised hierarchical clustering and PCA can illustrate patterns and relationships within the data, biomarker performance is better judged by training and testing models, as in the LOOCV.

While the generation of a cross-biofluid panel using CombiROC as a reduction tool produced high sensitivity, specificity and AUC values after LOOCV, one major caveat to this that must be addressed is the number of reduced panels in the PCF and the serum

that would have also met these sensitivity and specificity cut-offs. There were four PCF panels and four serum panels that met these cut-offs, each of which had a different combination of biomarkers. As such, it is possible that any and/or all of these cross-biofluid combinations could achieve these results, or even better. While these combinations were chosen as they were ranked first in the CombiROC software, demonstrating the best sensitivity when generating a ROC curve within the software itself, other combinations cannot be discounted on this basis and should therefore be explored. Unfortunately, as CombiROC does not have the capacity to evaluate more than 14 features at any one time due to the large number of permutations required, an alternative software with stronger processing power would be necessary to do this. Nevertheless, the results obtained here demonstrate the power that can be achieved through this process, and highlight a new promising avenue for biomarker panel generation.

The results reported in this chapter, not only describe the dysregulation of proteins and miRNA in pancreatic disease that have not previously been seen, but also demonstrate their potential utility as biomarkers of patients PC risk in this PCL cohort. An 8-protein panel and a 5-miRNA panel were integrated to create a 13-feature multi-omic panel in patient serum that could classify patients based on their risk of PC with high accuracy. The results reported in Chapter 4 were built on, when the 11-feature multi-omic panel in PCF was revaluated. Using novel CombiROC software, these two multi-omic panels were reduced and integrated, to create a cross-biofluid multi-omic panel that could stratify patients with improved accuracy compared to either multi-omic panel alone. This research not only highlights promising novel biomarkers of patient PC risk stratification, but provides a unique methodology for the generation of biomarker panels across biological samples. While these data remain to be further validated in an independent patient cohort, the outputs reported here give hope not just for the establishment of robust biomarkers in pancreatic disease, but for biomarker research as a whole.

Lastly, this work showcases the vast diversity of dysregulated components to be found within the PCF and serum of PCL patients. Given the expansive research conducted to date demonstrating the various factors within the PCF, and the results of

this thesis so far, it is important to understand how these factors become dysregulated and what role they may have in the progression of PCL patients to PC.

Chapter 6.

Functional characterisation of the role of pancreatic cyst fluid in the development of pancreatic cancer

6.1 Introduction

The hallmarks of cancer and enabling characteristics are widely regarded as the key features which best define malignant disease. In a procedure as intricate as the development of cancer, a simple definition or stepwise process does little to encapsulate the insurmountable complexity that is this disease. Indeed, over the course of the last 23 years alone, there have been three iterations of the "Hallmarks of Cancer", each attempting to encapsulate the mechanisms involved in the development of cancer by providing a list of strict criteria that must be met in order to define the disease[34-36]. With each of these three sets of principles, modern research has caused the ensuing recapitulation to be revised, and without exception, to expand. In fact, what originally began as six hallmarks of cancer, has become a list of parameters that is fourteen-strong, and which aim to summarize the functional capabilities that, when acquired by human cells, provides them with the ability to form a malignant tumour[36]. While at present, it seems impractical to assume that all such existing properties are recorded amongst this list of fourteen, it is widely accepted that, for now, these such characteristics are among the many attributed to cancer cells and should therefore be regarded as the building blocks that are required for the development of cancer. In this sense, it is of the upmost importance to characterise the roles that potential carcinogens may play in the acquisition of these traits by normal human cells in order to prevent the development of cancer, where possible. One such potential carcinogen that may contribute to the development of PC, PCF, remains to be fully understood.

PC is one of the world's worst prognosis cancers, with the 5-year survival rate for all stages combined in 2023 being just 12%[1]. This dismal outlook of PC patients is mostly due to the vague nature of the early symptoms, which go unnoticed and results in a late stage of disease at initial diagnosis[166]. PCLs are universally regarded as potential precursor lesions to PC, many subtypes of which are known to be pre-malignant[166]. Patients with PCLs are at a significantly higher risk of developing PC than those without PCLs, making these lesions an important stage in the early diagnosis of PC[452]. However, in a disease such as PC, where survival is notoriously poor, it is surprising that there is one major aspect of these lesions that has largely gone unchecked. To date, no studies have reported on the biological activity of the fluid

within these PCLs, or its potential role in PCL progression to PC. Indeed, the impact of cystic fluid on cell biology and activity has been largely limited to a handful of studies examining the effect of ovarian ascites fluid on ovarian cancer (OC) cell lines. A 2007 study conducted by Puiffe *et al.* explored the effect of patient-derived OC ascites on the OV-90 OC cell line. Interestingly, they found that the effects of ascites fluid on these cells could be categorized as either stimulatory or inhibitory[453]. Stimulatory ascites were shown to effect the expression of several genes (*ISGF3G, TRIB1, MKP1, RGS4, PLEC1*, and *MOSPD1*) and cause an increase in invasive potential, while inhibitory ascites caused a reduction in invasive potential. Interestingly, they found that when exposing cells to both inhibitory and stimulatory ascites, the inhibitory effect was dominant. A 2014 study cultured three OC cell lines in either culture medium or OC patient ascites, and found that exposure to ascites fluid differentially triggered a dissemination phenotype that was dependent on whether the cells were initially more epithelial- or mesenchymal-like[454]. Indeed, those with a more epithelial phenotype were driven towards colony and spheroid formation, while predominantly mesenchymal-like cells became more migratory, indicating that the initial features of the cells exposed to this fluid will dictate the effects that exposure confers. Given the impact that OC ascites fluid can have on OC cell line biology, this research endeavoured to uncover the effect that PCF could be having on the biology of those cells that make-up and surround the cyst, and the potential role that PCF could have in the progression of PCLs to PC.

In order to interrogate this, both normal and intermediary pancreatic cell lines were treated with PCF from patients that were classified as having either a low- or high-risk of PC development, as defined by the 2018 European evidence-based guidelines on pancreatic cystic neoplasms, and examined the effects of this fluid on normal cell biology and any potential changes that would align with the hallmarks of cancer[157]. Experiments were run in serum-protein free or “serum-starved” conditions. Serum-starvation of cells is often carried out in order to synchronize cells to the same cell cycle phase prior to exposing them to a treatment[455]. This removes the variable of cell cycle phase from the response of the cells to the treatment . Here, serum-starvation of normal cells was carried out to sensitize the cells prior to PCF exposure, such that effects could still be seen at low-volume treatments. *In vivo*, normal pancreatic cells can be

exposed to neat PCF for long periods of time. This type of experimental design is not possible *in vitro* as the volumes of PCF obtained from patients is generally extremely low. As such, serum-starvation was used in order to better prime the cells for exposure to PCF, and provide less subtle changes in the biological processes assessed.

As this work is largely preliminary, and mostly aims to interrogate whether PCF exposure can cause biological or functional changes to the cells, here, the baseline effects of PCF on the ability of normal pancreatic cells to resist cell death, invade tissue and metastasize, reprogram their cellular metabolism, and avoid immune destruction were examined. The cytotoxic potential of PCF was also profiled, and whether it is harmful to normal pancreatic cells or causes DNA damage. The answers to these important questions may shed new light on the development of PC, and provide new clinical avenues for the prevention of PC development in PCL patients.

6.2 Hypothesis

PCF is biologically active and has functional effects on the normal cells surrounding the cyst, potentially contributing to the eventual development of PC in some patients.

6.3 Specific aims

1. Elucidate whether exposure to PCF is cytotoxic to normal pancreatic cells, causing some measurable changes to cell death, viability, proliferation or DNA damage.
2. Examine whether exposure to PCF triggers a change in the metabolic profile of normal pancreatic cells.
3. Investigate whether exposure to PCF affects the expression of phenotypic or functional markers in normal pancreatic cells.
4. Determine whether exposure to PCF causes functional changes to normal pancreatic cells.

6.4 Experimental design

6.4.1 Dose-response curve for normal pancreatic cell line treatment with PCF

Pancreatic cell lines H6c7-normal and HPNE-intermediary were serum-starved as described in section 2.2.12, and seeded in triplicate at 3 x 10^3 cells per well in a flat-bottomed 96-well plate in 100 µL of their respective serum-free medium, and left in an incubator at 37°C overnight to adhere. The next morning, treatments were prepared in 50 µL volumes as in Table 6.1. Triplicates were prepared together and pipetted across the 3 replicates to reduce pipetting error. PCF samples were sonicated prior to use, as described in section 2.2.5, and a sample from one patient was used for all treatments to produce one replicate for the dose-response curve. A volume of 50 µL was carefully removed from each well of the 96-well plate and discarded to waste before being replaced by 50 µL of either fresh medium or fresh medium containing PCF as in Table 6.1. Final treatment concentrations of PCF in 100 µL were 0% (Untreated), 5%, 10%, 15% and 20% (v/v). Blank wells were created by adding medium only to 3 wells for each of the four media. Once treated, cells were left for 22 h in an incubator at 37°C and 5% CO_2. After 22 h, a proliferation assay was performed as described in section 2.2.15.

Table 6.1 PCF treatments for dose response curve.

Conditions	Volume of media per well	Volume of cyst fluid per well
cM blank	50 µL cM	-
SFM blank	50 µL SFM	-
Untreated cM control	50 µL cM	-
Untreated SFM control (0% PCF)	50 µL SFM	-
5% PCF treatment	45 µL SFM	5 µL PCF
10% PCF treatment	40 µL SFM	10 µL PCF
15% PCF treatment	35 µL SFM	15 µL PCF
20% PCF treatment	30 µL SFM	20 µL PCF

SFM = serum-free medium; cM = complete medium; PCF = pancreatic cyst fluid.

6.4.2 Flow cytometric staining of PCF treated cells

Flow cytometric staining was carried out as per section 2.2.17.2. Analysis of flow cytometric data was carried out as per section 2.2.17.3. The panel of extracellular and intracellular antibodies used can be found in Table 6.2. All antibodies were used at a volume of 1uL per sample tube. Zombie NIR Viability dye was used at a dilution of 1:1000 in PBS to assess live vs dead cells. Zombie NIR dye working solution was also used at a volume of 1µL per sample.

6.4.3 Analysis of experimental data

Experimental data were handled and analysed as outlined in section 2.2.21.2. Given the high level of patient-to-patient variability among samples, outliers were excluded where appropriate using outlier analysis in GraphPad Prism (v9.5.0). The ROUT method of outlier analysis was used to identify outliers, with Q (the chance of falsely identifying one or more outliers) set to 1. Outliers identified using this method where excluded from the dataset they were identified in.

Table 6.2 Flow cytometry panel of fluorochrome-conjugated anti-human antibodies.

Antibody	Channel	Location	Company	Clone	Product code
Vimentin	FITC	Intracellular	BioLegend, USA	O91D3	677809
Slug	PE	Intracellular	Miltenyi Biotec, UK	REA404	130-131-007
N-cadherin	PE-Cy5.5	Extracellular	BioLegend, USA	8C11	350813
E-cadherin	PE-Cy7	Extracellular	BioLegend, USA	67A4	324115
EGFR	APC	Extracellular	BioLegend, USA	AY13	352905
CD133	BV421	Extracellular	BioLegend, USA	S16015F	393907
PD-L1	BV510	Extracellular	BioLegend, USA	29E.2A3	329733

6.5 Results

6.5.1 PCF can be cytotoxic and significantly affects apoptosis, proliferation and viability in normal pancreatic cell lines

In order to determine if PCF could influence the biology of normal pancreatic cell lines, serum-starved H6c7-normal and HPNE-intermediary cell lines were treated with low concentrations of PCF and their proliferation post 24 h was examined. PCF treatment resulted in a reduction in proliferation for both H6c7-normal and HPNE-intermediary cell lines compared to the untreated control (UTC)(Figure 6.1). Both cell lines had reduced proliferation when treated with 5%, 10%, 15% and 20% (v/v) PCF, though this effect is more pronounced in the H6c7-normal cells. As the biological effect of PCF on normal pancreatic cell biology can be seen at treatment with just 5% PCF, and there is little variability observed in this effect across higher treatment concentrations, all subsequent experiments are carried out with the 5% PCF concentration to best utilise these limited-volume patient samples.

Further interrogation into the influence of PCF on normal cell biology was conducted by examining cell viability, cell death, and cytotoxicity of PCF. Treatment of H6c7-normal cells with PCF resulted in a significant increase in the number of viable cells ($p<0.001$), along with a significant decrease in apoptosis ($p<0.05$)(Figure 6.2). While no change was seen in the proliferation of these cells, the PCF was also shown to demonstrate no significant cytotoxic effect on H6c7-normal cells (Figure 6.2). A large amount of variability can be observed in the effects of different patient PCF on the level of apoptosis seen in H6c7-normal cells (Figure 6.2C). Indeed, while the overall trend shows a decrease in apoptosis when the cells are treated with PCF, the substantial spread of these data indicates a considerable amount of patient-to-patient variability which causes the general result to be difficult to determine. Conversely, treatment with PCF significantly increased the level of apoptosis seen in HPNE-intermediary cells ($p<0.05$), though again patient-to-patient variability is quite large (Figure 6.3C). PCF treatment did not significantly alter HPNE-intermediary cell viability, despite being significantly cytotoxic towards these cells ($p<0.01$)(Figure 6.3A-B). It is important to note that these results are further compounded by the exclusion of results from three patients (one low-risk and two high-risk) from this analysis as outliers, having fold

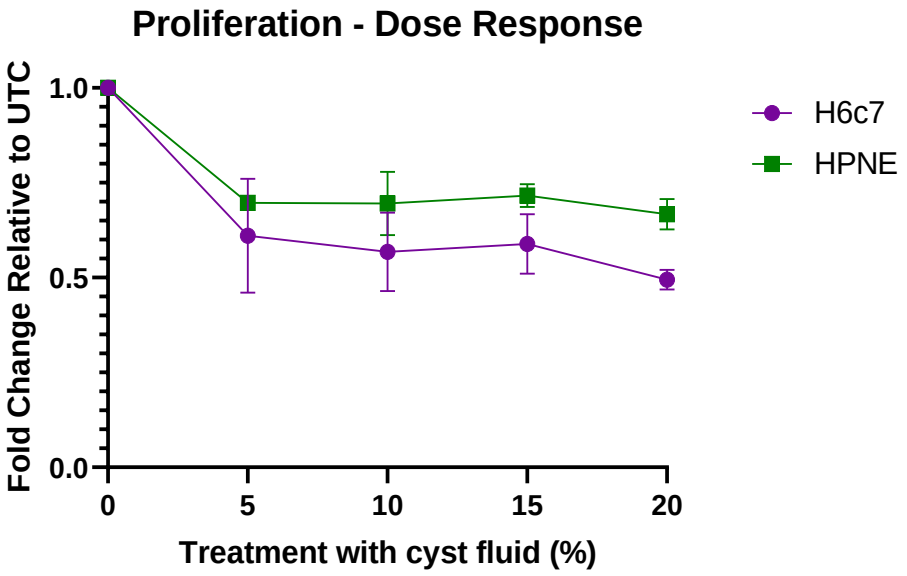


Figure 6.1 Treatment with PCF affects the proliferation of normal pancreatic cell lines. Dose Response curve showing the effect of varying concentrations [0%, 5%, 10%, 15%, 20% (v/v)] of PCF on serum-starved H6c7-normal (purple) and HPNE-intermediary (green) cell line proliferation after 24 h. Data are presented as mean ± SEM for 2 patients.

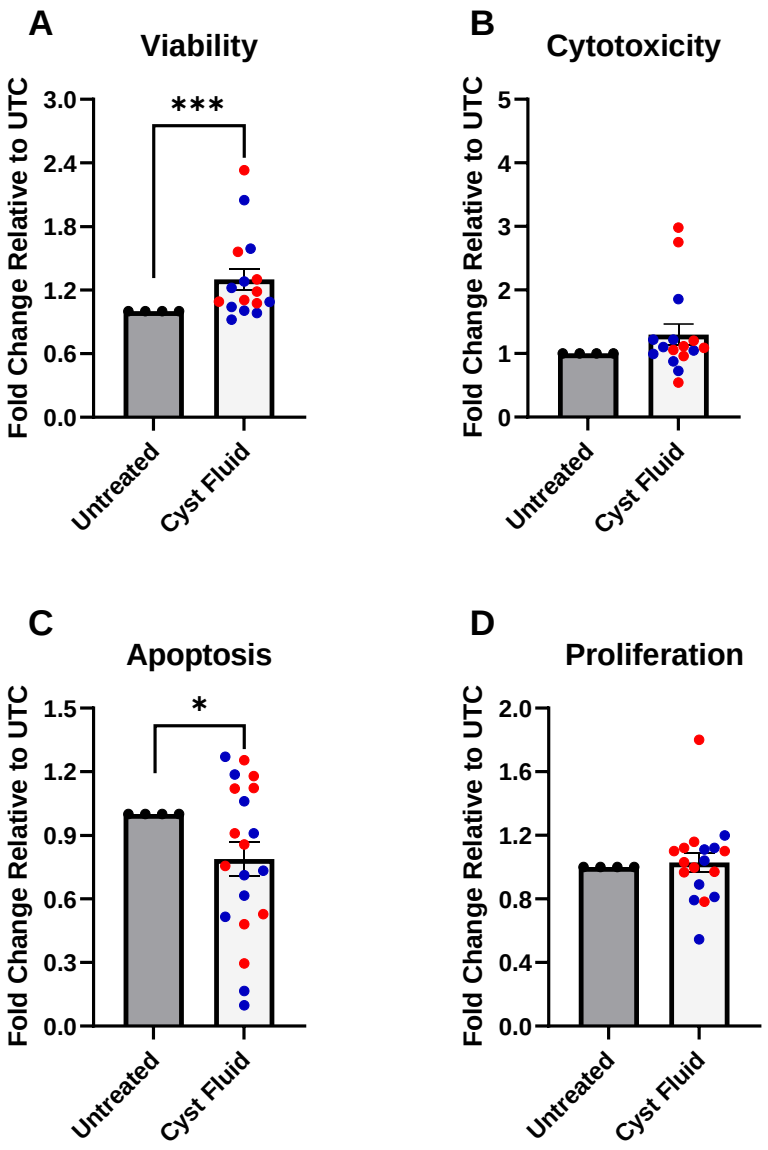


Figure 6.2 Treatment with PCF for 24 h significantly increases the viability and decreases the apoptosis of H6c7-normal cells. Fold change relative to the UTC of H6c7-normal cells treated for 24 h with 5% (v/v) PCF on (**A**) viability, (**B**) cytotoxicity, (**C**) apoptosis, and (**D**) proliferation. Data are presented as mean ± SEM for 20 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. One-sample Wilcoxon-signed rank test. *p<0.05, ***p<0.001.

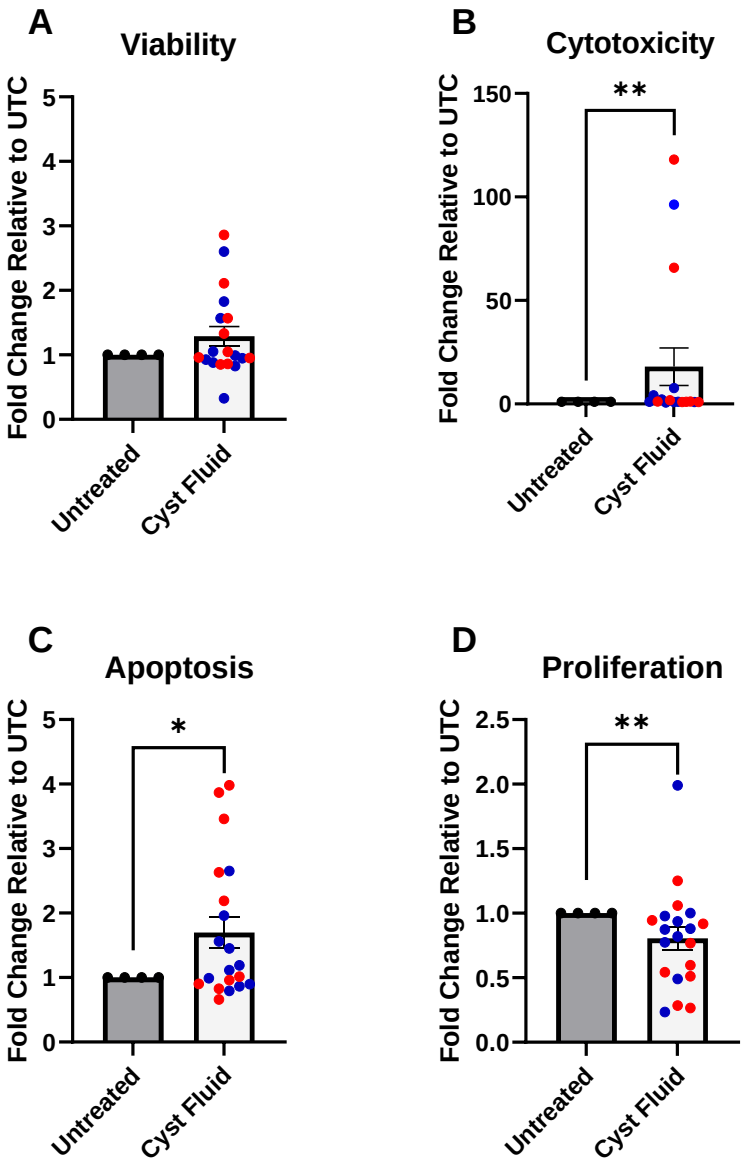


Figure 6.3 Treatment with PCF for 24 h is cytotoxic to HPNE-intermediary cells, significantly increasing apoptosis and decreasing proliferation. Fold change relative to the UTC of HPNE-intermediary cells treated for 24 h with 5% (v/v) PCF on (**A**) viability, (**B**) cytotoxicity, (**C**) apoptosis, and (**D**) proliferation. Data are presented as mean ± SEM for 20 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. One-sample Wilcoxon-signed rank test. *p<0.05, **p<0.01.

change values of more than 190-fold greater than the UTC. PCF also caused a significant decrease in HPNE-intermediary cell proliferation, though again, the spread of these data is sizeable indicating a degree of patient-to-patient variability in this effect (p<0.01)(Figure 6.3D). Overall, treatment with PCF proved to be significantly cytotoxic to HPNE-intermediary cells, and has demonstrated the potential to alter the viability, apoptosis and proliferation of both normal cell lines. These effects, however, appear to vary greatly from patient-to-patient, and between the two cell lines, making an overall pattern difficult to discern.

When broken down into low- and high-risk patient PCF treatments, both low- (p<0.05) and high-risk (p<0.01) PCF are shown to be responsible for the significant increase in H6c7-normal cell viability (Figure 6.4A). However, while there is a general trend towards a decrease in apoptosis, neither low- nor high-risk PCF alone causes a significant change to H6c7-normal cell apoptosis, with the large patient-to-patient variability being evident in both groups (Figure 6.4C). Both low- and high-risk PCF remain non-cytotoxic to H6c7-normal cells, and have no significant effect on cell proliferation (Figure 6.4B and 6.4D). In the HPNE-intermediary cells, again while there is a trend towards an increase in apoptosis, this is non-significant in both low- and high-risk PCF treatments, with the spread of the data being substantial here again (Figure 6.5C). Although neither low- nor high-risk PCF had a significant cytotoxic effect on HPNE-intermediary cells individually, these results also exclude data from three outlier patients as above (Figure 6.5B). When included in the analysis, both low- and high-risk PCF have significant cytotoxic effects on HPNE-intermediary cells (p<0.05)(data not shown). In terms of proliferation, while there is a decrease in this when treated with low- or high-risk PCF, the high-risk PCF caused a significant decrease (p<0.05)(Figure 6.5D). When divided into risk categories, it remains evident that the effects seen on normal pancreatic cell biology are patient dependent and risk does not appear to be the defining factor.

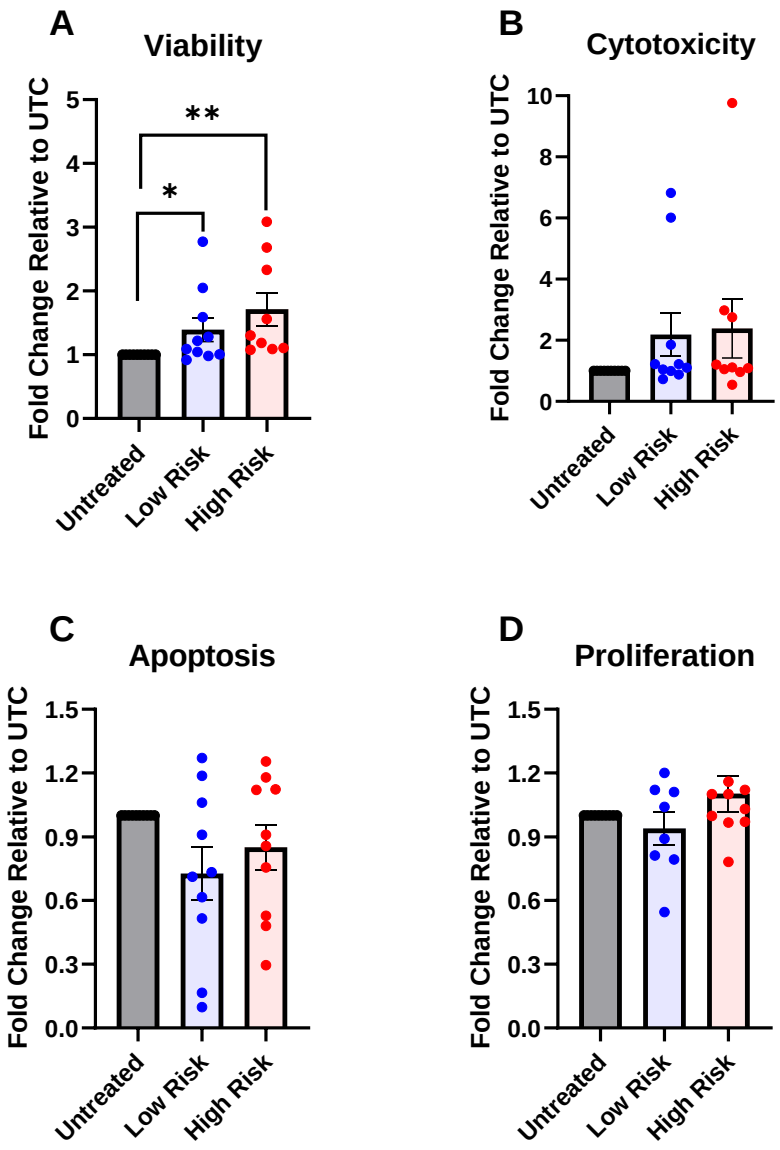


Figure 6.4 Treatment with high-risk PCF for 24 h significantly increases the viability of H6c7-normal cells. Fold change relative to the UTC of H6c7-normal cells treated for 24 h with 5% (v/v) low-risk or high-risk PCF on (**A**) viability, (**B**) cytotoxicity, (**C**) apoptosis, and (**D**) proliferation. Data are presented as mean ± SEM for n=10 low-risk and n=10 high-risk patients. One-sample Wilcoxon-signed rank test coupled with a Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05, **p<0.01.

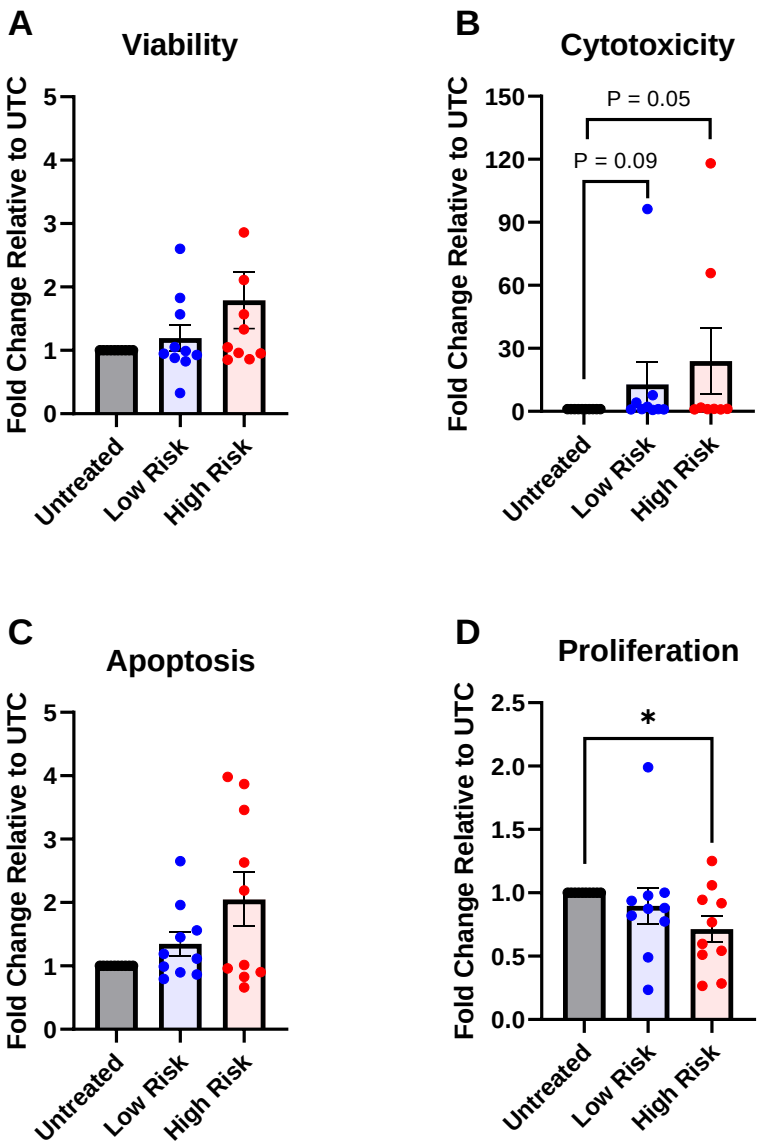


Figure 6.5 Treatment with high-risk PCF for 24 h significantly decreases the proliferation of HPNE-intermediary cells. Fold change relative to the UTC of HPNE-intermediary cells treated for 24 h with 5% (v/v) low-risk or high-risk PCF on (**A**) viability, (**B**) cytotoxicity, (**C**) apoptosis, and (**D**) proliferation. Data are presented as mean ± SEM for n=10 low-risk and n=10 high-risk patients. One-sample Wilcoxon-signed rank test coupled with a Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05.

6.5.2 Serum starvation significantly alters the basal respiration of normal pancreatic cell lines

Before examining the effect of PCF on cell line metabolism, the influence of serum starvation on the basal metabolic profile of normal pancreatic cell lines was examined. To assess the basal respiration of H6c7-normal and HPNE-intermediary cells, and whether this could be challenged within a 24 h time period, Seahorse XF ATP rate tests were run on both cell lines after culturing in normal and serum-starved conditions for 24 h. Serum-starvation did not significantly alter the total amount of ATP produced by H6c7 cells, or the amount of ATP produced by glycolysis [glycoATP] or oxidative phosphorylation [mitoATP] (Figure 6.6A-C). Despite this, serum-starvation caused a significant shift in the primary metabolic pathway used by the cells, with oxidative phosphorylation being responsible for significantly more ATP production than glycolysis (p<0.01)(Figure 6.6D). H6c7-normal cells grown in normal culture medium demonstrate a minor dependency on oxidative phosphorylation, though this is not significant (Figure 6.6D).

For HPNE-intermediary cells, there is a significant increase in the amount of glycoATP produced in serum-starved cells compared to those cultured under normal conditions (p<0.05)(Figure 6.7A). Though not significant, this trend continues in both mitoATP and total ATP production, with increases in both seen in serum-starved HPNE-intermediary cells (Figure 6.7B-C). HPNE-intermediary cells rely significantly more on oxidative phosphorylation than glycolysis as their primary metabolic pathway (p<0.01)(Figure 6.7D). Serum-starved HPNE-intermediary cells also rely significantly more on oxidative phosphorylation than glycolysis, despite the increases in both forms of ATP (p<0.05)(Figure 6.7D). This reliance is slightly reduced in the serum-starved HPNE-intermediary cells, though not significantly.

6.5.3 PCF does not significantly alter the metabolic profile of normal pancreatic cell lines

Having shown that the metabolic profiles of these cells could be altered within a 24 h time period, the effect of PCF on the metabolic profile of normal pancreatic cell lines, H6c7-normal and HPNE-intermediary, was assessed via Seahorse XF ATP rate test.

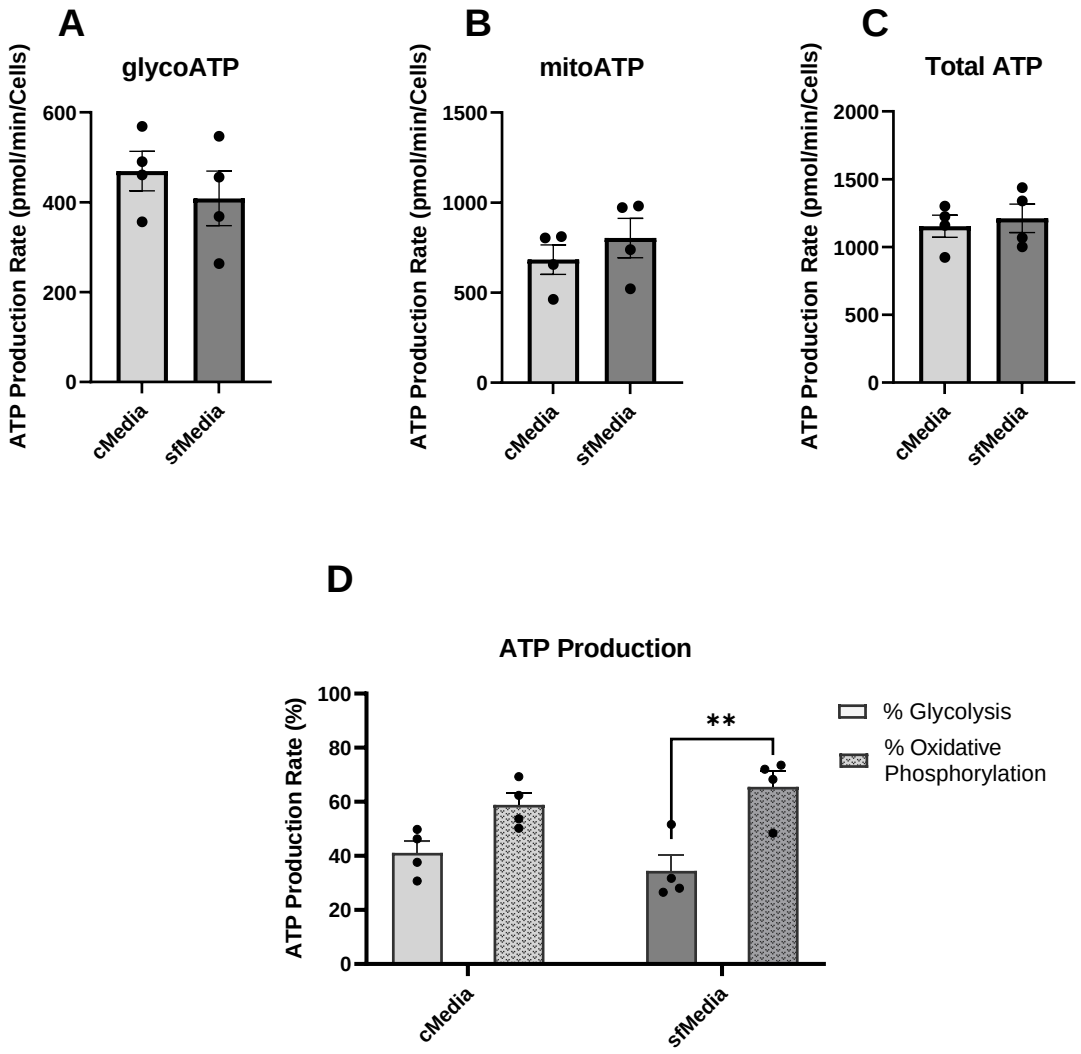


Figure 6.6 Serum starvation for 24 h significantly increases the reliance of H6c7-normal cells on oxidative phosphorylation compared to glycolysis. Metabolic profile of H6c7-normal cells grown in either complete media (cMedia) or serum-free media (sfMedia) for 24 h. Unpaired t-test comparing (**A**) ATP produced from glycolysis (glycoATP), (**B**) ATP produced from mitochondrial oxidative phosphorylation (mitoATP), and (**C**) Total ATP (glycoATP + mitoATP). (**D**) 2-way ANOVA with Tukey's multiple comparisons test comparing the percentage ATP produced by glycolysis versus oxidative phosphorylation. Data are presented as mean ± SEM for n=4 experimental replicates. ** p<0.01.

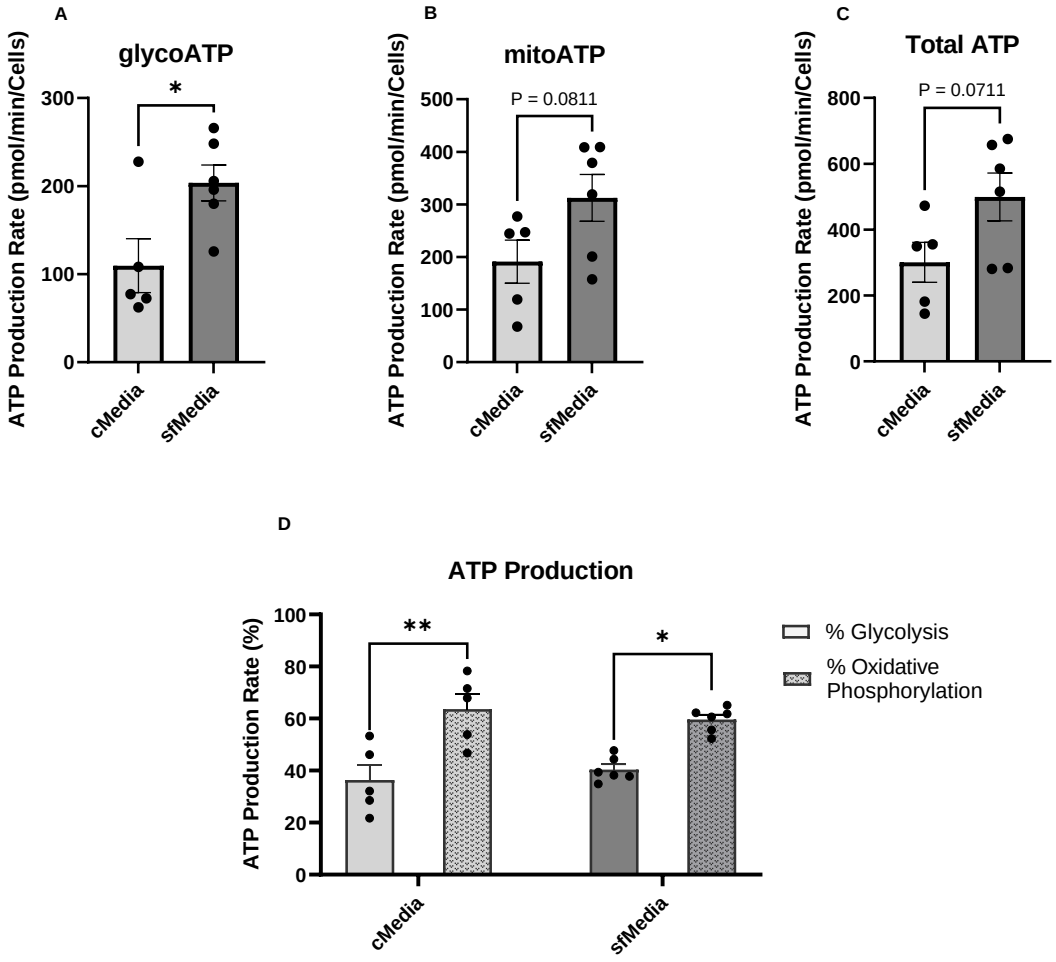


Figure 6.7 Serum starvation for 24 h significantly increases the production of glycoATP in HPNE-intermediary cells. Metabolic profile of HPNE-intermediary cells grown in either complete media (cMedia) or serum-free media (sfMedia) for 24 h. Unpaired t-test comparing (**A**) ATP produced from glycolysis (glycoATP), (**B**) ATP produced from mitochondrial oxidative phosphorylation (mitoATP), and (**C**) Total ATP (glycoATP + mitoATP). (**D**) 2-way ANOVA with Tukey's multiple comparisons test comparing the percentage ATP produced by glycolysis versus oxidative phosphorylation. Data are presented as mean ± SEM for n=6 experimental replicates. **p*<0.05, ***p*<0.01.

Treatment with PCF significantly increased the production of glycoATP in H6c7-normal cells (p<0.05), but did not alter the production of mitoATP or the total amount of ATP produced (Figure 6.8A-C). While not significant, a slight shift towards the reliance of H6c7-normal cells on glycolysis after treatment with PCF can be observed (Figure 6.8D). However, there is a substantial amount of patient-to-patient variability in both the %glycolysis and %oxidative phosphorylation, resulting in an overall even distribution of ATP production by glycolysis and oxidative phosphorylation for H6c7-normal cells when treated with PCF.

PCF treatment did not significantly alter the production of glycoATP, mitoATP, or the total amount of ATP produced by HPNE-intermediary cells (Figure 6.9A-C). HPNE-intermediary cells rely significantly on oxidative phosphorylation as their primary metabolic pathway, and treatment with PCF does not alter this (p<0.0001)(Figure 6.9D). Importantly, the patient-to-patient variability seen in the H6c7-normal cells in this context is not reflected in the HPNE-intermediary cells, with the distribution of data points being more homogenous, despite the patient samples being utilised across both cell lines being matched.

When broken down into low- and high-risk patient PCF treatments, high-risk PCF is shown to be responsible for the significant increase in glycoATP production in H6c7-normal cells (p<0.05), with low-risk PCF demonstrating no significant effect (Figure 6.10A). Both low- and high-risk PCF had no significant effect on the amount of mitoATP or the total ATP produced by H6c7-normal cells (Figure 6.10B-C). Importantly, while low-risk PCF demonstrates a similar breakdown of ATP production by glycolysis versus oxidative phosphorylation, high-risk PCF appears to cause a shift towards glycolysis in H6c7-normal cells, though not significant (Figure 6.10D). The glycoATP, mitoATP and amount of total ATP produced by HPNE-intermediary cells were not significantly altered by low- or high-risk PCF (Figure 6.11A-C). The significant reliance of these cells on oxidative phosphorylation over glycolysis (p<0.01) was also not affected by low- or high-risk PCF treatment, though it was slightly more pronounced in cells treated with low-risk PCF (p<0.0001)(Figure 6.11D). Overall, it is evident that PCF has the potential to alter normal pancreatic cell line metabolism but the sizeable amount of patient-to-patient variability observed causes there to be no overarching trend in these effects. It

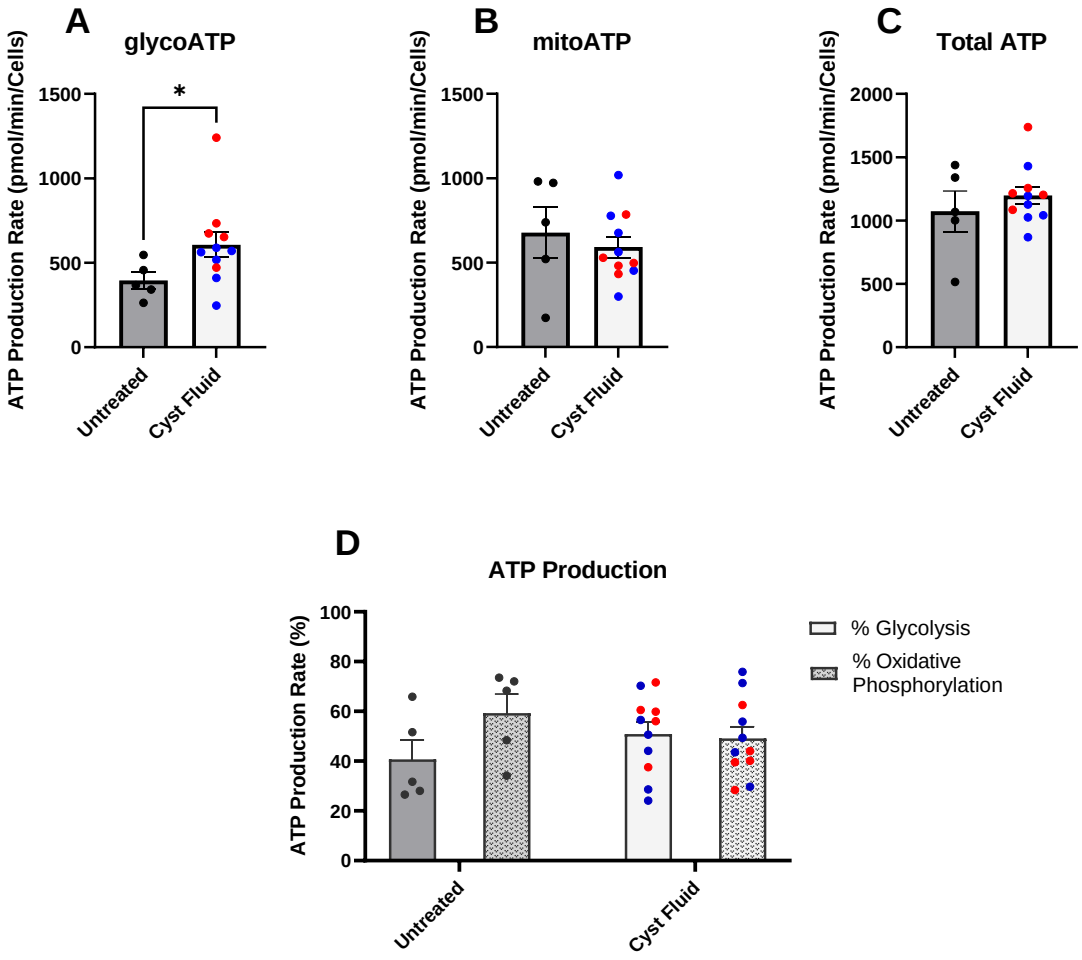


Figure 6.8 Treatment with PCF for 24 h significantly increases the production of glycoATP in H6c7-normal cells. Metabolic profile of H6c7-normal cells treated with 5% (v/v) PCF for 24 h. Mann-Whitney test comparing (**A**) ATP produced from glycolysis (glycoATP), (**B**) ATP produced from mitochondrial oxidative phosphorylation (mitoATP), and (**C**) Total ATP (glycoATP + mitoATP). (**D**) 2-way ANOVA with Tukey's multiple comparisons test comparing the percentage ATP produced by glycolysis versus oxidative phosphorylation. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Data are presented as mean ± SEM for n=11 patients. *p<0.05.

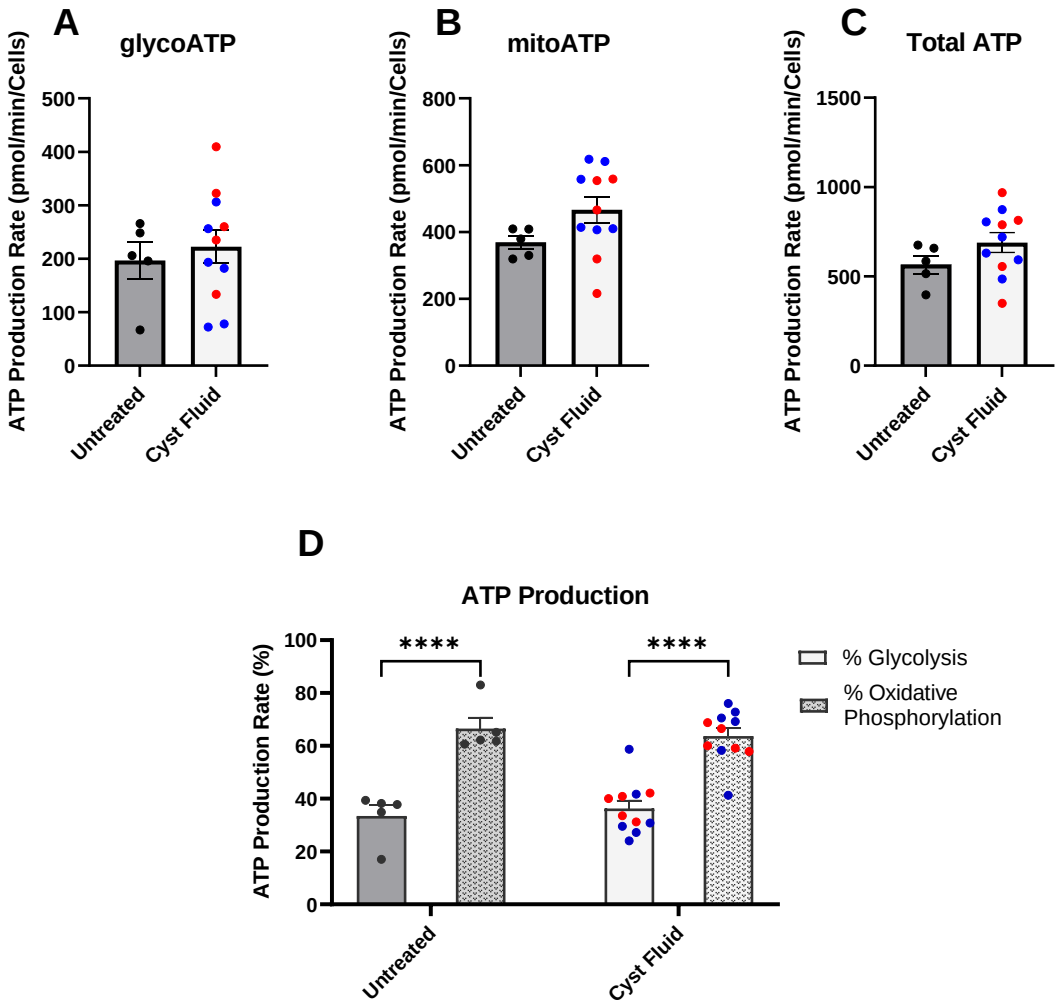


Figure 6.9 Treatment with PCF for 24 h does not significantly alter the metabolic profile of HPNE-intermediary cells. Metabolic profile of HPNE-intermediary cells treated with 5% (v/v) PCF for 24 h. Mann-Whitney test comparing (**A**) ATP produced from glycolysis (glycoATP), (**B**) ATP produced from mitochondrial oxidative phosphorylation (mitoATP), and (**C**) Total ATP (glycoATP + mitoATP). (**D**) 2-way ANOVA with Tukey's multiple comparisons test comparing the percentage ATP produced by glycolysis versus oxidative phosphorylation. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Data are presented as mean ± SEM for n=11 patients. ****p<0.0001.

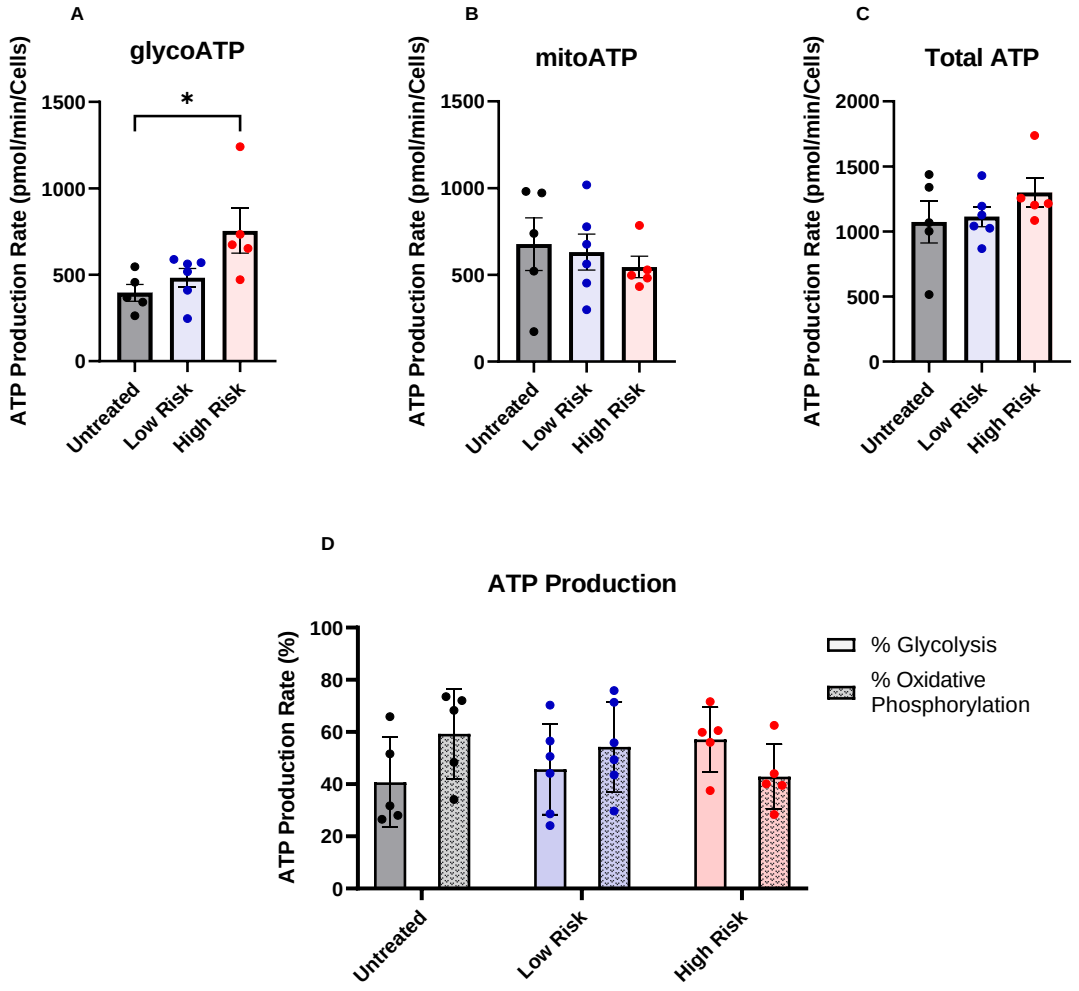


Figure 6.10 Treatment high-risk PCF for 24 h significantly increases the production of glycoATP in H6c7-normal cells. Metabolic profile of H6c7-normal cells treated with 5% (v/v) low-risk or high-risk PCF for 24 h. Kruskal-Wallis test with Dunn's multiple comparisons comparing (**A**) ATP produced from glycolysis (glycoATP), (**B**) ATP produced from mitochondrial oxidative phosphorylation (mitoATP), and (**C**) Total ATP (glycoATP + mitoATP). (**D**) 2-way ANOVA with Tukey's multiple comparisons test comparing the percentage ATP produced by glycolysis versus oxidative phosphorylation. Data are presented as mean ± SEM for n=6 low-risk patients and n=5 high-risk patients. *p<0.05.

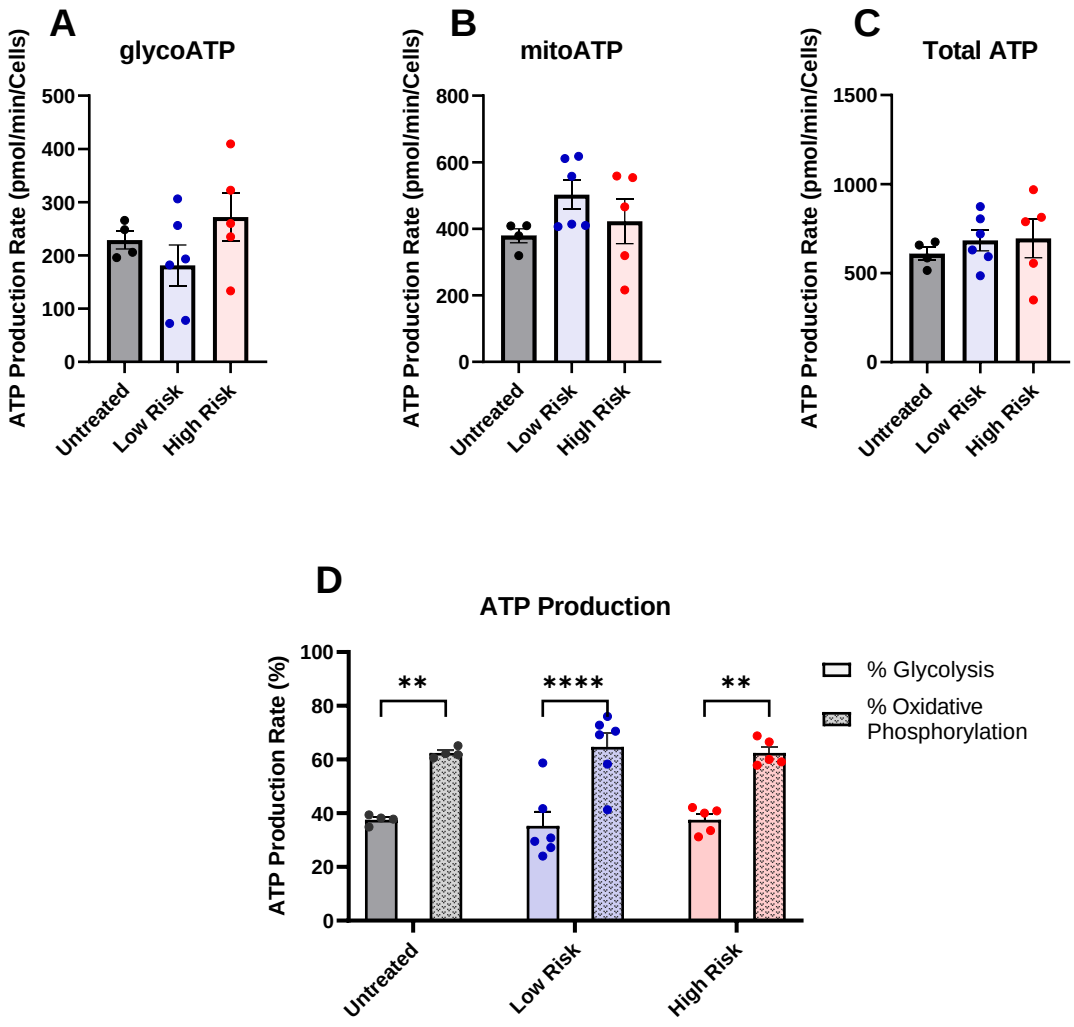


Figure 6.11 Treatment with low- or high-risk PCF for 24 h does not significantly alter the metabolic profile of HPNE-intermediary cells. Metabolic profile of HPNE-intermediary cells treated with 5% (v/v) low-risk or high-risk PCF for 24 h. Kruskal-Wallis test with Dunn's multiple comparisons comparing (**A**) ATP produced from glycolysis (glycoATP), (**B**) ATP produced from mitochondrial oxidative phosphorylation (mitoATP), and (**C**) Total ATP (glycoATP + mitoATP). (**D**) 2-way ANOVA with Tukey's multiple comparisons test comparing the percentage ATP produced by glycolysis versus oxidative phosphorylation. Data are presented as mean ± SEM for n=6 low-risk patients and n=5 high-risk patients. **p<0.01, ****p<0.0001.

is also important to note that these experiments were run using a total of 20 patient PCFs, but nine (n=4 low-risk and n=5 high-risk) of these killed the both the H6c7-normal and HPNE-intermediary cells, so metabolic read outs could not be obtained. Again, risk does not appear to be the defining characteristic.

6.5.4 PCF does not alter the expression of phenotypic and functional markers on normal pancreatic cell lines after 6 h

Given the subtle changes to metabolic profiles observed after 24 h, the expression of key phenotypic and functional markers were assessed flow cytometrically following normal pancreatic cell line, H6c7-normal and HPNE-intermediary, treatment with 5% (v/v) PCF for just 6 h. Treatment with PCF for 6 h did not significantly alter the percentage of Vimentin$^+$, N-cadherin$^+$, E-cadherin, EGFR$^+$, Slug$^+$, CD133$^+$ or PD-L1$^+$ H6c7-normal cells (Figure 6.12). Serum-starvation caused a noticeable increase in the percentage of N-cadherin$^+$, CD133$^+$ and PD-L1$^+$ H6c7-normal cells compared to the basal levels, and this was not affected by PCF treatment. Treatment with PCF for 6 h also did not significantly alter the percentage of Vimentin$^+$, N-cadherin$^+$, E-cadherin$^+$, EGFR$^+$, Slug$^+$, CD133$^+$ or PD-L1$^+$ HPNE-intermediary cells (Figure 6.13). In this case, serum-starvation caused a visible increase in the percentage of E-cadherin$^+$ cells, and a decrease in the percentage of EGFR$^+$ cells, which again was not affected by PCF treatment.

Separating the PCF treatments into those from low- or high-risk patients shows no significant differences in the percentage of Vimentin$^+$, N-cadherin$^+$, E-cadherin, EGFR$^+$, Slug$^+$, CD133$^+$ or PD-L1$^+$ H6c7-normal cells after low- or high-risk PCF treatment for 6 h (Figure 6.14). Similarly, there is no significant different in the percentage of HPNE-intermediary cells after treatment with low- or high-risk PCF for 6 h (Figure 6.15). Importantly, these data do not have the large spread that were observed in previous experiments, perhaps indicating that a 6 h treatment is too short a timeframe to evaluate these effects.

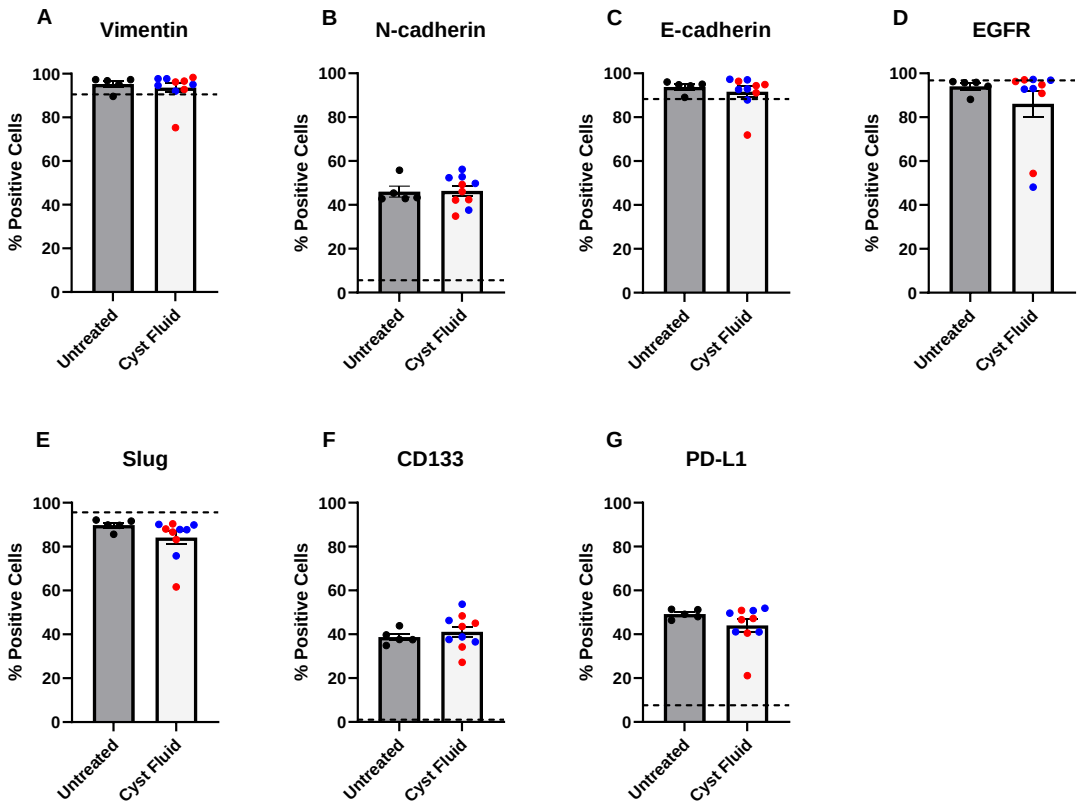


Figure 6.12 Treatment with PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on H6c7-normal cells. Percentage of H6c7-normal cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) PCF for 6 h. Dotted lines indicate the basal expression level of H6c7-normal cells for each marker. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test.

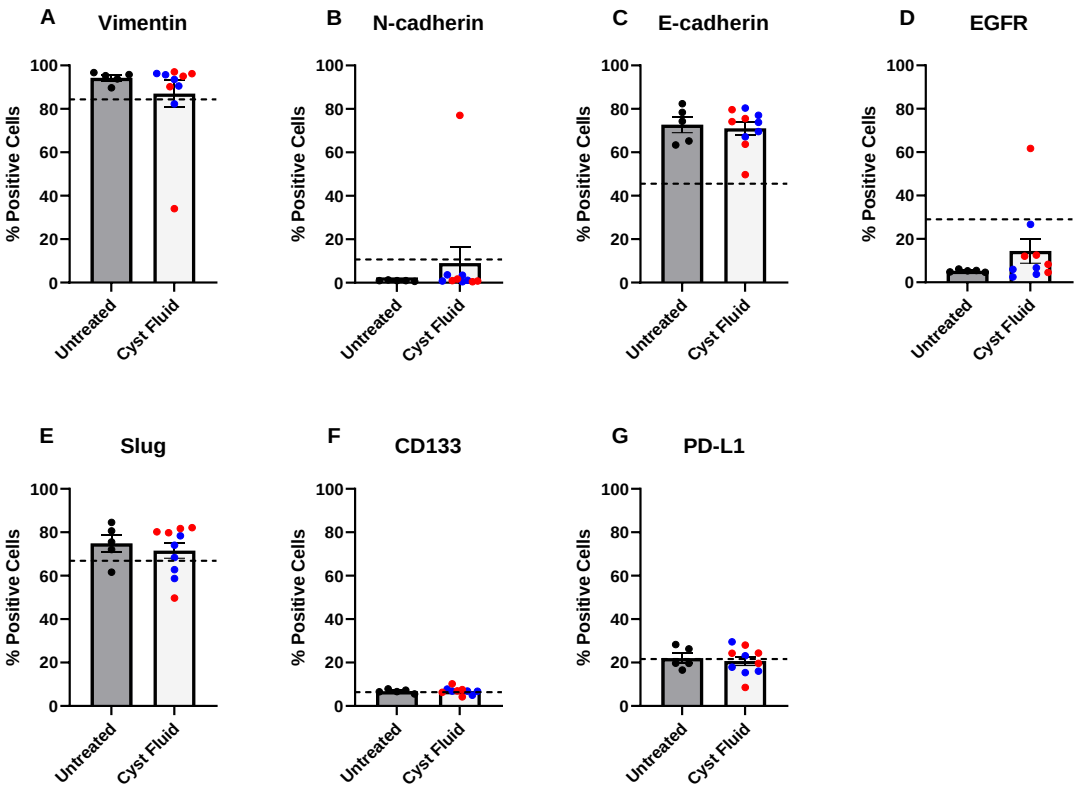


Figure 6.13 Treatment with PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on HPNE-intermediary cells. Percentage of HPNE-intermediary cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) PCF for 6 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each marker. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test.

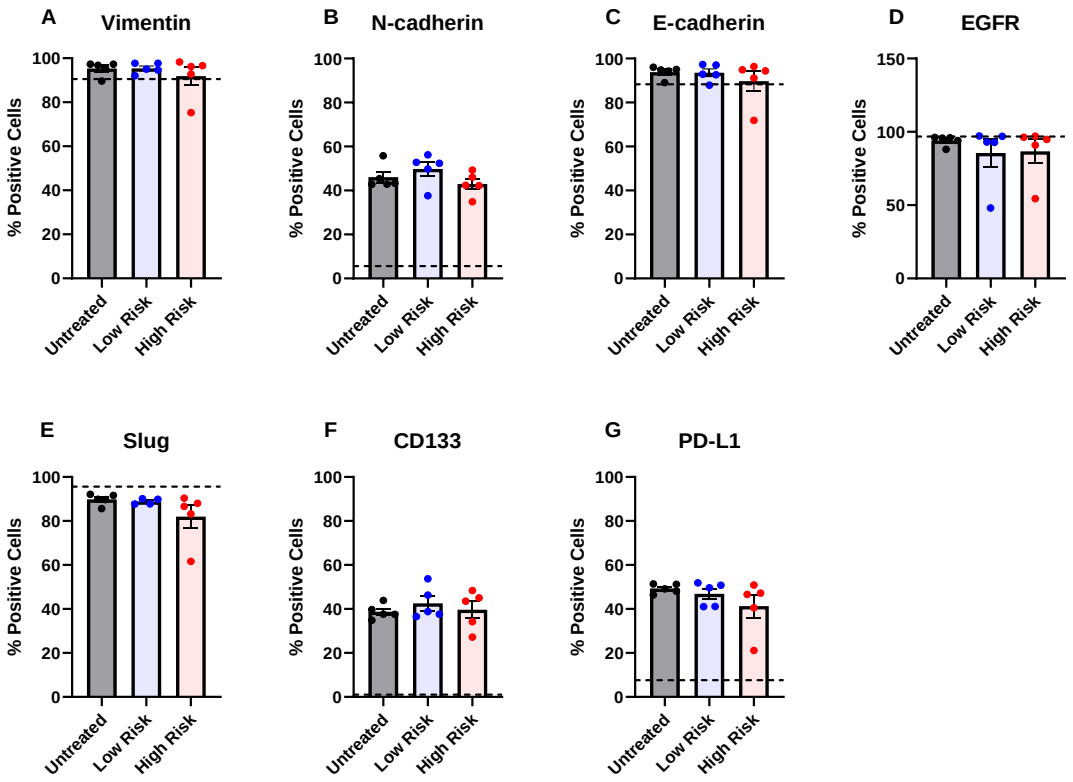


Figure 6.14 Treatment with low- or high-risk PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on H6c7-normal cells. Percentage of H6c7-normal cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) low- or high-risk PCF for 6 h. Dotted lines indicate the basal expression level of H6c7-normal cells for each marker. Data are presented as mean ± SEM for n=5 low-risk patients and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons.

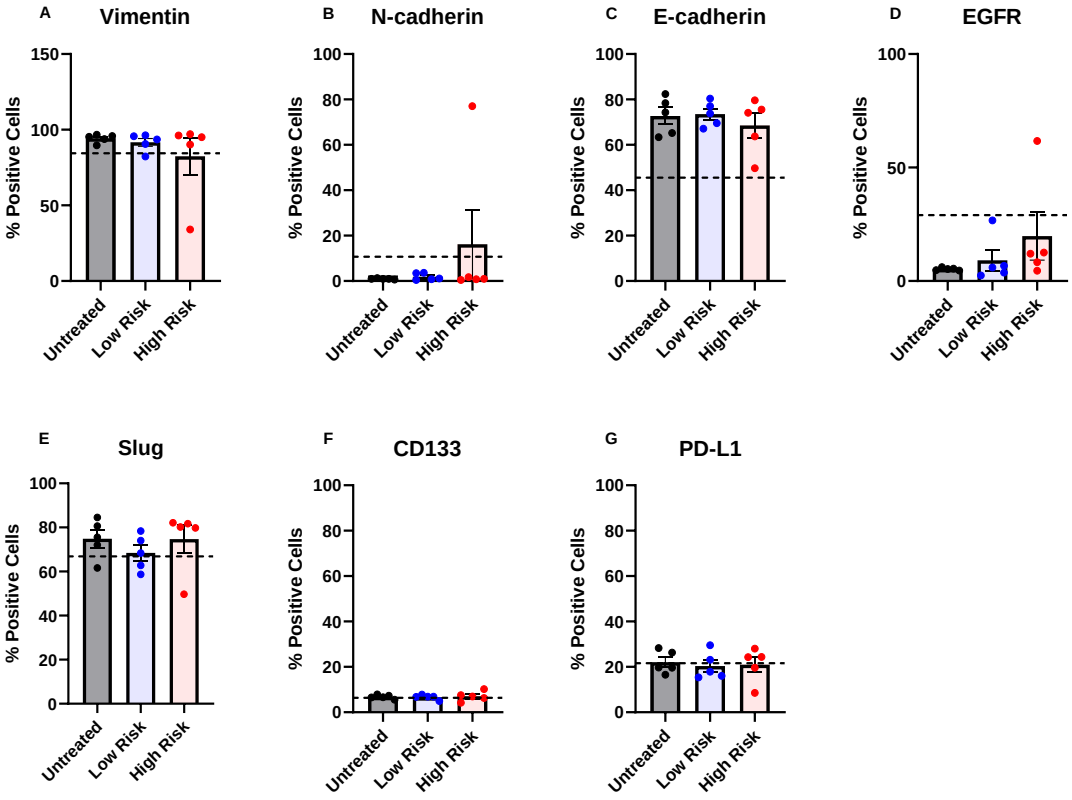


Figure 6.15 Treatment with low- or high risk PCF for 6 h does not significantly alter the expression of phenotypic and functional markers on HPNE-intermediary cells. Percentage of HPNE-intermediary cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) low- or high-risk PCF for 6 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each marker. Data are presented as mean ± SEM for n=5 low-risk patients and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons.

6.5.5 PCF significantly decreases the percentage of Vimentin$^+$ and PD-L1$^+$ H6c7-normal cells after 24 h

As 6 h was shown to be too short a time frame to observe changes in marker expression, this timeframe was extended to 24 h of treatment. Seven phenotypic and functional markers of interest were assessed flow cytometrically following H6c7-cell line treatment with 5% (v/v) PCF for 24 h. While serum-starvation for 24 h did not cause a noticeable change to the percentage of Vimentin$^+$ H6c7-normal cells compared to the basal levels, PCF treatment significantly decreased the percentage of Vimentin$^+$ H6c7-normal cells (p<0.05)(Figure 6.16A). Conversely, serum-starvation for 24 h caused a visible increase in the percentage of PD-L1$^+$ H6c7-normal cells compared to the basal levels, however, PCF treatment significantly decreased the percentage of PD-L1$^+$ H6c7-normal cells towards these basal levels (p<0.01)(Figure 6.16G). Treatment with PCF for 24 h did not significantly alter the percentage of N-cadherin$^+$, E-cadherin$^+$, EGFR$^+$, Slug$^+$ or CD133$^+$ H6c7-normal cells (Figure 6.16).

When separated based on risk, the decrease in the percentage of Vimentin$^+$ H6c7-normal cells can be seen in both low- and high-risk PCF treatments, though neither on their own is significantly different compared to the untreated control (Figure 6.17A). For the PD-L1$^+$ H6c7-normal cells, again both low- and high-risk PCF decrease the percentage pf PD-L1$^+$ cells, with high-risk PCF causing a significant decrease (p<0.05)(Figure 6.17G). The percentage of Vimentin$^+$, N-cadherin$^+$, E-cadherin$^+$, EGFR$^+$, Slug$^+$ and CD133$^+$ H6c7-normal cells was not significantly altered following treatment with low- or high-risk PCF for 24 h (Figure 6.17). Importantly, after 24 h PCF treatment the spread of data is again quite large, showing the variability of the effects seen between patient-to-patient PCF with risk separation having little improvement on this.

When comparing the effects of PCF after 6 h or 24 h, the percentage of Vimentin$^+$ H6c7-normal cells was significantly decreased at 24 h compared to 6 h, indicating that these changes become more pronounced over time (p<0.001)(Figure 6.18A). Interestingly, the percentage of PD-L1$^+$ H6c7-normal cells increased over time following serum starvation, but decreased over time in the presence of PCF, though not significantly (p>0.05)(Figure 6.18B). While not significantly altered compared to serum starved controls, the percentage of EGFR$^+$ H6c7-normal cells was significantly reduced

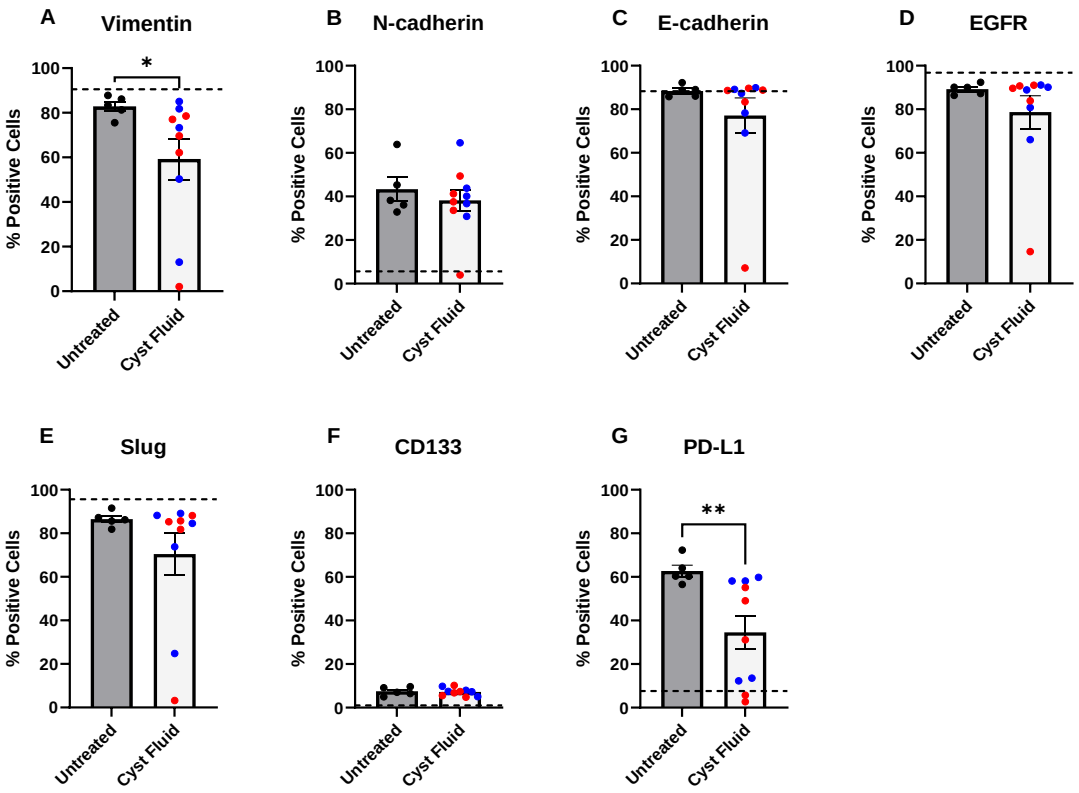


Figure 6.16 Treatment with PCF for 24 h significantly decreases the expression of vimentin and PD-L1 on H6c7-normal cells. Percentage of H6c7-normal cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) PCF for 24 h. Dotted lines indicate the basal expression level of H6c7-normal cells for each marker. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test. *p<0.05, **p<0.01.

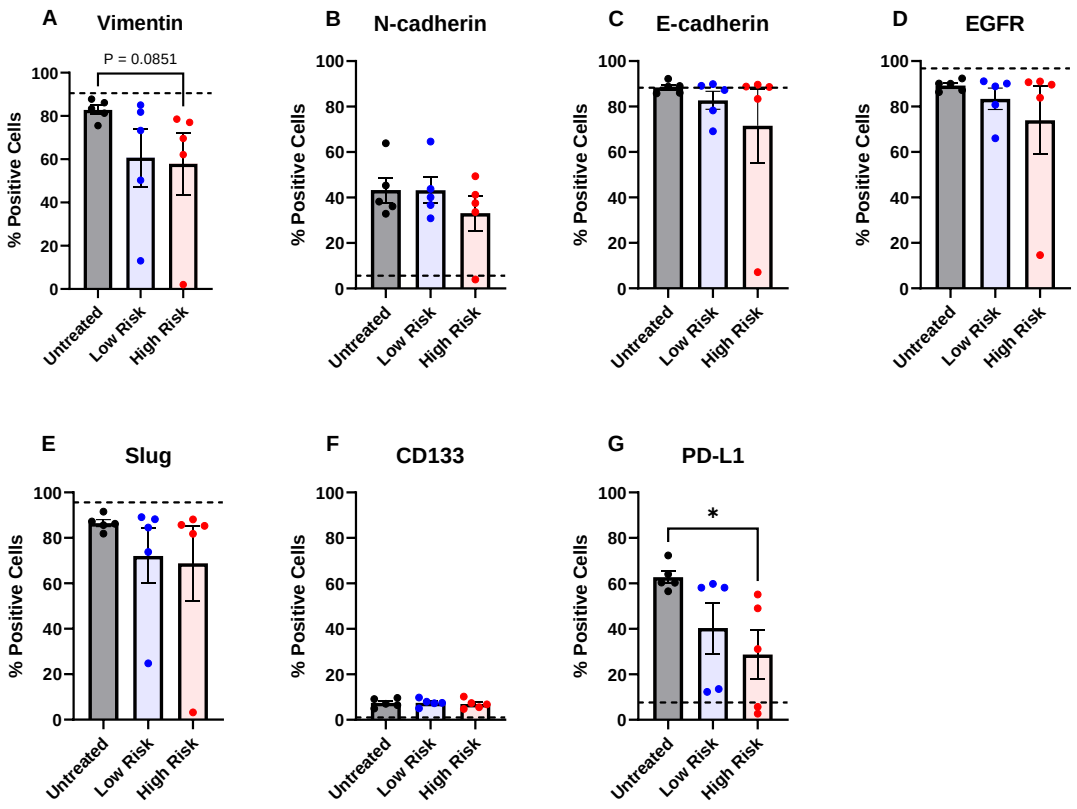


Figure 6.17 Treatment with high-risk PCF for 24 h significantly decreases the expression of PD-L1 on H6c7-normal cells. Percentage of H6c7-normal cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) low- or high-risk PCF for 24 h. Dotted lines indicate the basal expression level of H6c7-normal cells for each marker. Data are presented as mean ± SEM for n=5 low-risk patients and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05.

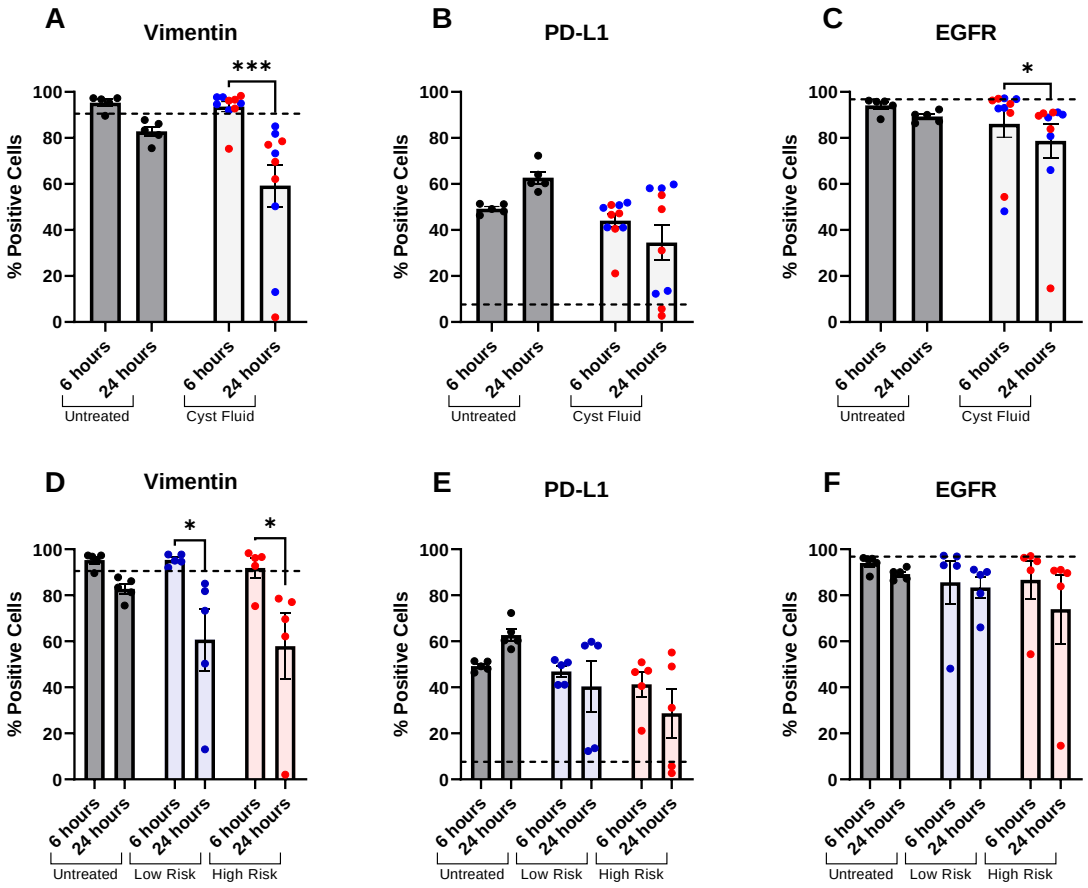


Figure 6.18 PCF significantly decreases the percentage of Vimentin⁺ and EGFR⁺ H6c7-normal cells at 24 h compared to 6 h. Percentage of H6c7-normal cells expressing phenotypic and functional markers of interest at 6 h and 24 h (**A-C**) post-treatment with 5% (v/v) PCF and (**D-F**) post-treatment with 5% (v/v) low- or high-risk PCF. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05, ***p<0.001.

after 24 h compared to 6 h (p<0.05)(Figure 6.18C). When examined based on risk, both low- and high-risk PCF significantly decreased the percentage of Vimentin$^+$ H6c7-normal cells at 24 h compared to 6 h (p<0.05)(Figure 6.18D). Neither low- nor high-risk PCF caused significant changes to the percentage of PD-L1$^+$ or EGFR$^+$ H6c7-normal cells between the 6 h and 24 h treatments (p>0.05)(Figure 6.18E-F).

6.5.6 Low-risk PCF significantly increases the percentage of EGFR$^+$ HPNE-intermediary cells after 24 h

As changes in marker expression could be seen after 24 h in the H6c7-normal cells, these seven phenotypic and functional markers of interest were assessed flow cytometrically following HPNE-intermediary cell line treatment with 5% (v/v) PCF for 24 h. Serum-starvation for 24 h did not cause a noticeable change in the percentage of Vimentin$^+$, N-cadherin$^+$, Slug$^+$, CD133$^+$ or PD-L1$^+$ HPNE-intermediary cells compared to basal levels, and treatment with PCF also did not cause a significant difference (Figure 6.19). Serum-starvation for 24 h caused a noticeable increase in the percentage of E-cadherin$^+$ cells compared to basal levels, with PCF treatment having no effect on this (Figure 6.19C). Conversely, serum-starvation for 24 h caused a visible decrease in the percentage of EGFR$^+$ HPNE-intermediary cells compared to the basal levels, however, PCF treatment significantly increased the percentage of EGFR$^+$ cells towards these basal levels (p<0.01)(Figure 6.19D). When separated based on risk, the increase in the percentage of EGFR$^+$ HPNE-intermediary cells can be seen in both low- and high-risk PCF treatments, though only low-risk PCF caused a statistically significant increase (p<0.05)(Figure 6.20D). The percentage of Vimentin$^+$, N-cadherin$^+$, E-cadherin$^+$, Slug$^+$, CD133$^+$ and PD-L1$^+$ HPNE-intermediary cells was not significantly altered following treatment with low- or high-risk PCF for 24 h (Figure 6.20). Similarly to the H6c7-normal cells, at 24 h PCF treatment the data can be seen to spread out again, showing the patient-to-patient variability. In the HPNE-intermediary cells also, the separation of risk groups does not correct this.

Examining the change in these markers between the 6 h and 24 h treatments demonstrates that, while there was no change in the percentage of Vimentin$^+$ HPNE-intermediary cells between the untreated control and PCF treatment at 24 h, these cells

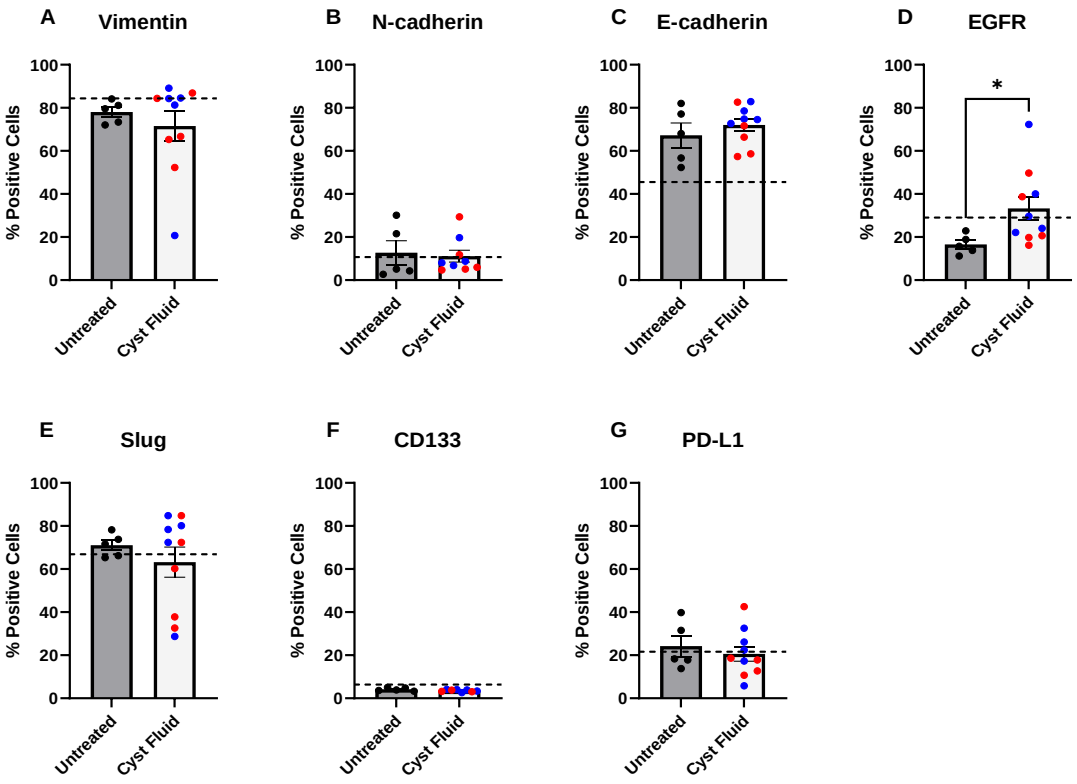


Figure 6.19 Treatment with PCF for 24 h significantly increases the expression of EGFR on HPNE-intermediary cells. Percentage of HPNE-intermediary cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) PCF for 24 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each marker. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test. *p<0.05.

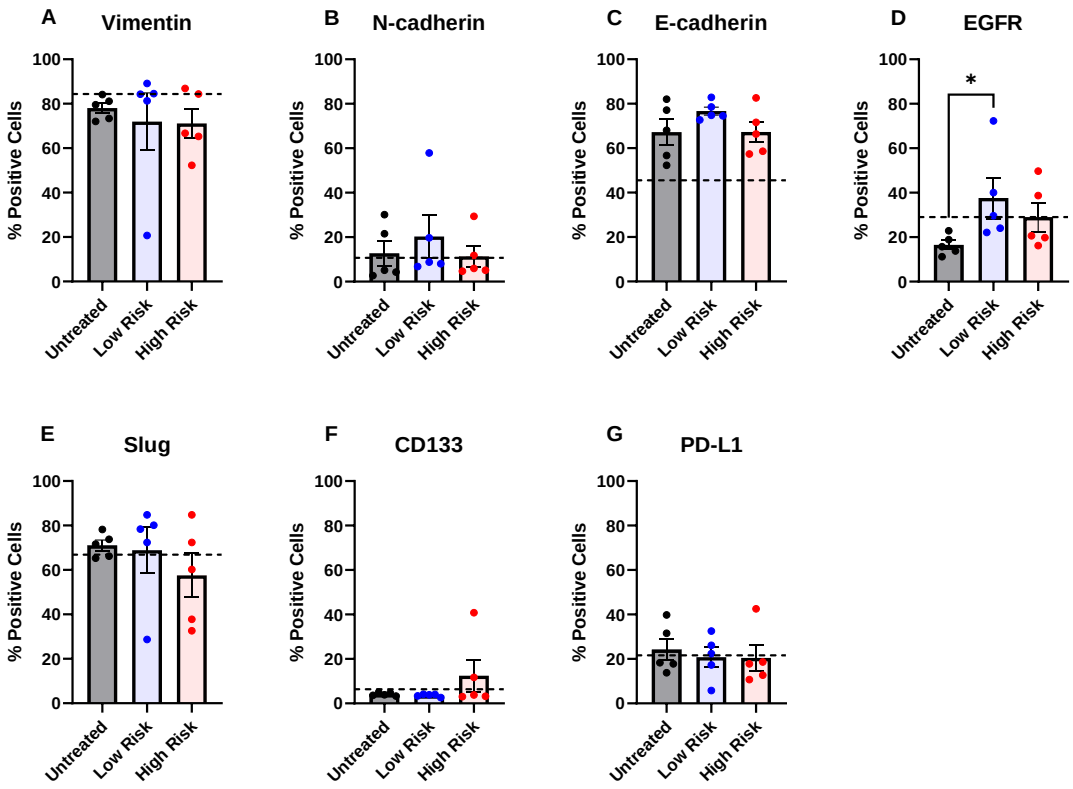


Figure 6.20 Treatment with low-risk PCF for 24 h significantly increases the expression of EGFR on HPNE-intermediary cells. Percentage of HPNE-intermediary cells expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) low- or high-risk PCF for 24 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each marker. Data are presented as mean ± SEM for n=5 low-risk patients and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05.

were significantly decreased at 24 h in both untreated and PCF treated conditions compared to 6 h, indicating that this change is a result of the serum-starvation rather than the PCF treatment (p<0.01)(Figure 6.21A). This trend continues when the treatments are split into low- and high-risk PCF, however, only the shift in the untreated controls remains significant (p<0.05)(Figure 6.21D). There was no significant change in the percentage of PD-L1$^+$ HPNE-intermediary cells treated with PCF (p>0.05)(Figure 6.21B), or low- or high-risk PCF (p>0.05)(Figure 6.21E). There was a significant increase in the percentage of EGFR$^+$ HPNE-intermediary cells after 24 h treatments with PCF compared to 6 h (p<0.001)(Figure 6.21C), with low-risk PCF causing a significant increase on its own also (p<0.05)(Figure 6.21F).

6.5.7 PCF does not significantly alter the co-expression of phenotypic and functional markers on normal cells after 6 h

Following on from the alterations seen in the percentage of Vimentin$^+$, PD-L1$^+$ and EGFR$^+$ cells following treatment with PCF, co-expression of these three phenotypic and functional markers was assessed flow cytometrically following cell line treatment with 5% (v/v) PCF for 6 h. No significant difference in the percentage of any co-expressing cell population was observed in H6c7-normal cells following treatment with PCF for 6 h (Figure 6.22A-O). Of note, the percentage of Vimentin$^+$/CD133$^+$, Vimentin$^+$/N-cadherin$^+$, Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/CD133$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/N-cadherin$^+$, EGFR$^+$/CD133$^+$ and EGFR$^+$/N-cadherin$^+$ H6c7-normal cells were visibly increased following 6 h of serum-starvation compared to basal levels, and PCF treatment did not significantly alter this (Figure 6.22). The percentage of HPNE-intermediary cells co-expressing these markers was not significantly affected by PCF treatment for 6 h (Figure 6.23A-O). The percentage of Vimentin$^+$/CD133$^+$, Vimentin$^+$/N-cadherin$^+$, Vimentin$^+$/EGFR$^+$, Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$, EGFR$^+$/N-cadherin$^+$, and EGFR$^+$/E-cadherin$^+$ HPNE-intermediary cells is noticeable decreased following serum-starvation for 6 h compared to basal levels. Conversely, Vimentin$^+$/E-cadherin$^+$ and PD-L1$^+$/CD133$^+$ HPNE-intermediary cells were noticeably increased following 6 h of serum-starvation compared to basal levels (Figure 6.23).

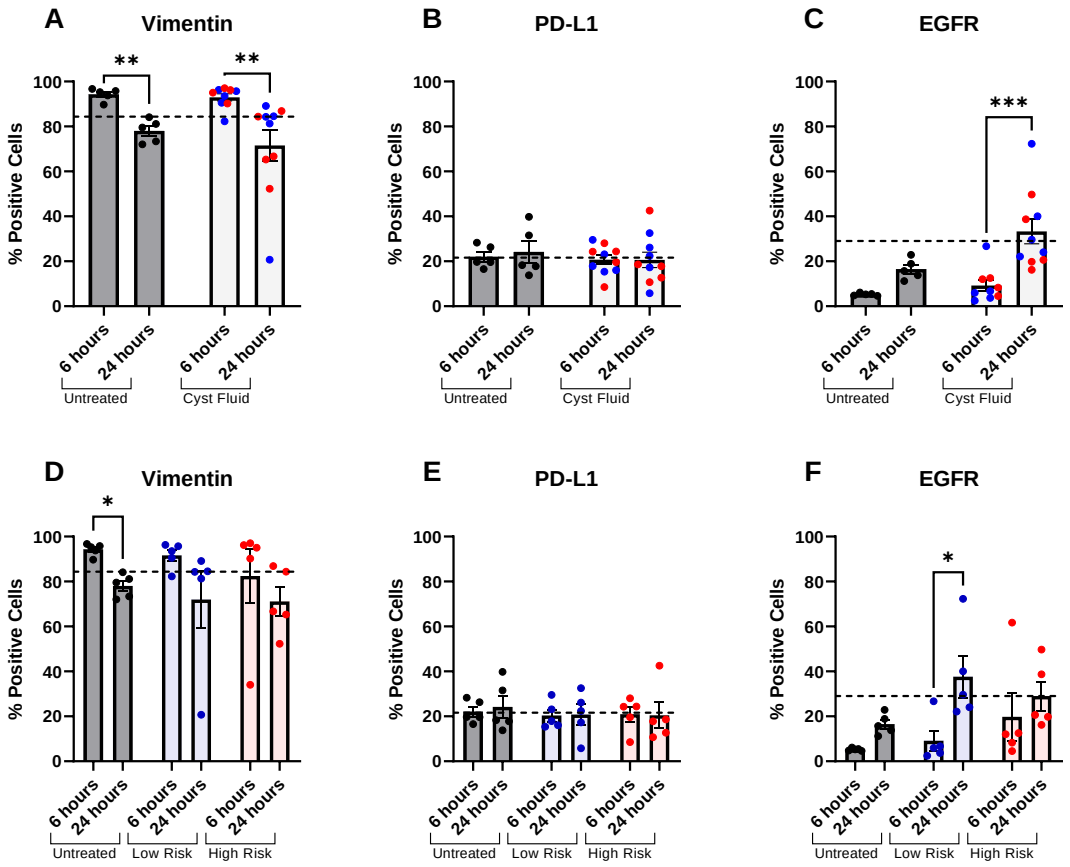


Figure 6.21 PCF significantly increases the percentage of EGFR⁺ HPNE-intermediary cells at 24 h compared to 6 h. Percentage of HPNE-intermediary cells expressing phenotypic and functional markers of interest at 6 h and 24 h (**A-C**) post-treatment with 5% (v/v) PCF and (**D-F**) post-treatment with 5% (v/v) low- or high-risk PCF. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05, **p<0.01, ***p<0.001.

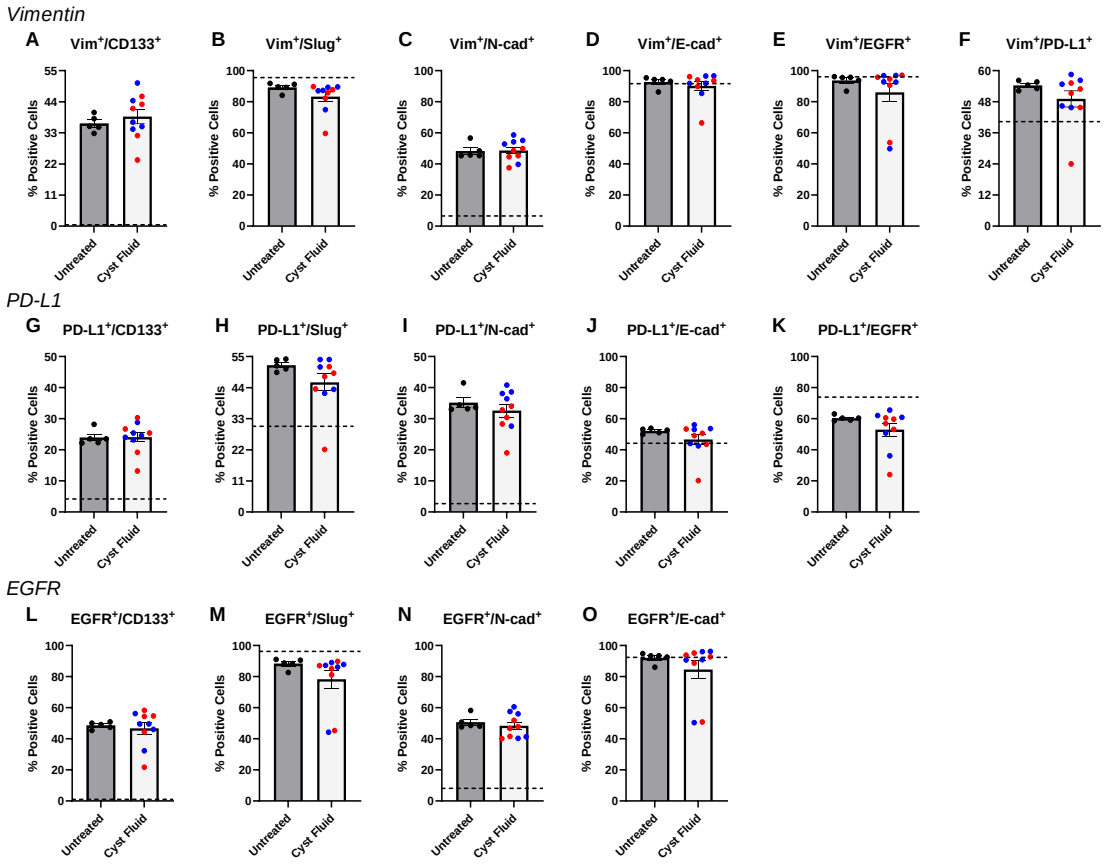


Figure 6.22 Treatment with PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on H6c7-normal cells. Percentage of H6c7-normal cells expressing co-expressing phenotypic and functional markers of interest (**A-O**) post-treatment with 5% (v/v) PCF for 6 h. Dotted lines indicate the basal expression level of H6c7-normal cells for each co-expression. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

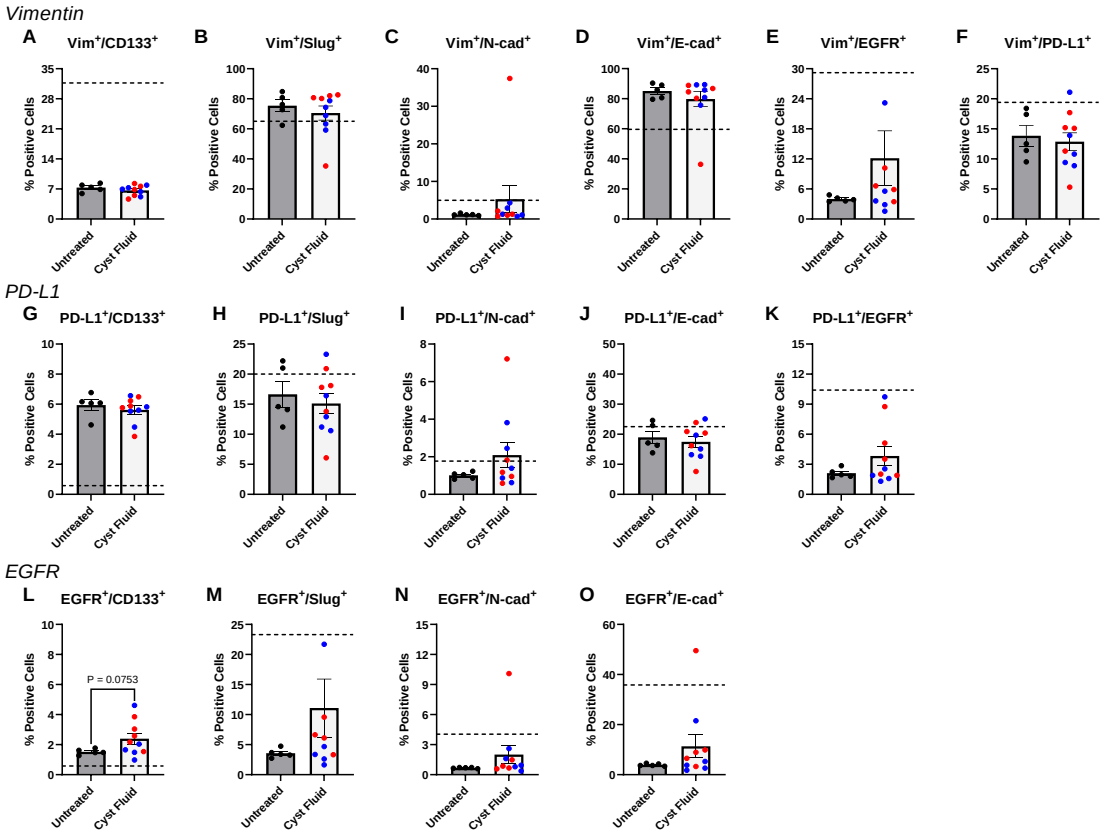


Figure 6.23 Treatment with PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on HPNE-intermediary cells. Percentage of HPNE-intermediary cells expressing co-expressing phenotypic and functional markers of interest (**A-O**) post-treatment with 5% (v/v) PCF for 6 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each co-expression. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

Separation of these data into low- and high-risk treatment groups does not produce any significant alterations in the percentage of H6c7-normal cells or HPNE-intermediary co-expressing these markers of interest (Figure 6.24, Figure 6.25). Again, at the 6 h PCF treatment these data can be seen to clump together, and the patient-to-patient variability is not visible at this time point.

6.5.8 PCF significantly alters the percentage of Vimentin$^+$/Slug$^+$, Vimentin$^+$/E-cadherin$^+$, Vimentin$^+$/EGFR$^+$, Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cadherin$^+$, and PD-L1$^+$/EGFR$^+$ H6c7-normal cells after 24 h

Given the significant changes seen in the percentage of Vimentin$^+$, PD-L1$^+$ and EGFR$^+$ cells following treatment with PCF, co-expression of these three phenotypic and functional markers was assessed flow cytometrically following cell line treatment with 5% (v/v) PCF for 24 h. Treatment with PCF for 24 h significantly decreased the percentage Vimentin$^+$/Slug$^+$, Vimentin$^+$/E-cadherin$^+$, Vimentin$^+$/EGFR$^+$, Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cadherin$^+$, and PD-L1$^+$/EGFR$^+$ H6c7-normal cells (p<0.05)(Figure 6.26). Serum-starvation did not change the percentage of Vimentin$^+$/E-cadherin$^+$ or PD-L1$^+$/EGFR$^+$ H6c7-normal cells compared to basal levels, however, treatment with PCF significantly decreased the percentage of these cells (p<0.05). Conversely, serum-starvation noticeably increased the percentage of Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$ and PD-L1$^+$/E-cadherin$^+$ H6c7-normal cells, with PCF treatment significantly decreasing the percentage of these cells back towards basal levels (p<0.05). The percentage of PD-L1$^+$/N-cadherin$^+$ and EGFR$^+$/CD133$^+$ H6c7-normal cells was also greatly reduced following treatment with PCF, though not significantly (p=0.0509 and p=0.0699, respectively). No significant difference was observed in the percentage of Vimentin$^+$/CD133$^+$, Vimentin$^+$/N-cadherin$^+$, PD-L1$^+$/CD133$^+$, PD-L1$^+$/N-cadherin$^+$, EGFR$^+$/CD133$^+$, EGFR$^+$/Slug$^+$, EGFR$^+$/N-cadherin$^+$ or EGFR$^+$/E-cadherin$^+$ H6c7-normal cells following treatment with PCF for 24 h (Figure 6.26). The percentage of Vimentin$^+$/CD133$^+$, Vimentin$^+$/N-cadherin$^+$, PD-L1$^+$/N-cadherin$^+$, EGFR$^+$/CD133$^+$ and EGFR$^+$/N-cadherin$^+$ H6c7-normal cells were also visibly increased following 24 h of serum-starvation compared to basal levels, and PCF treatment did not significantly alter this. When separated into low- and high-risk PCF treatments, high-risk PCF significantly

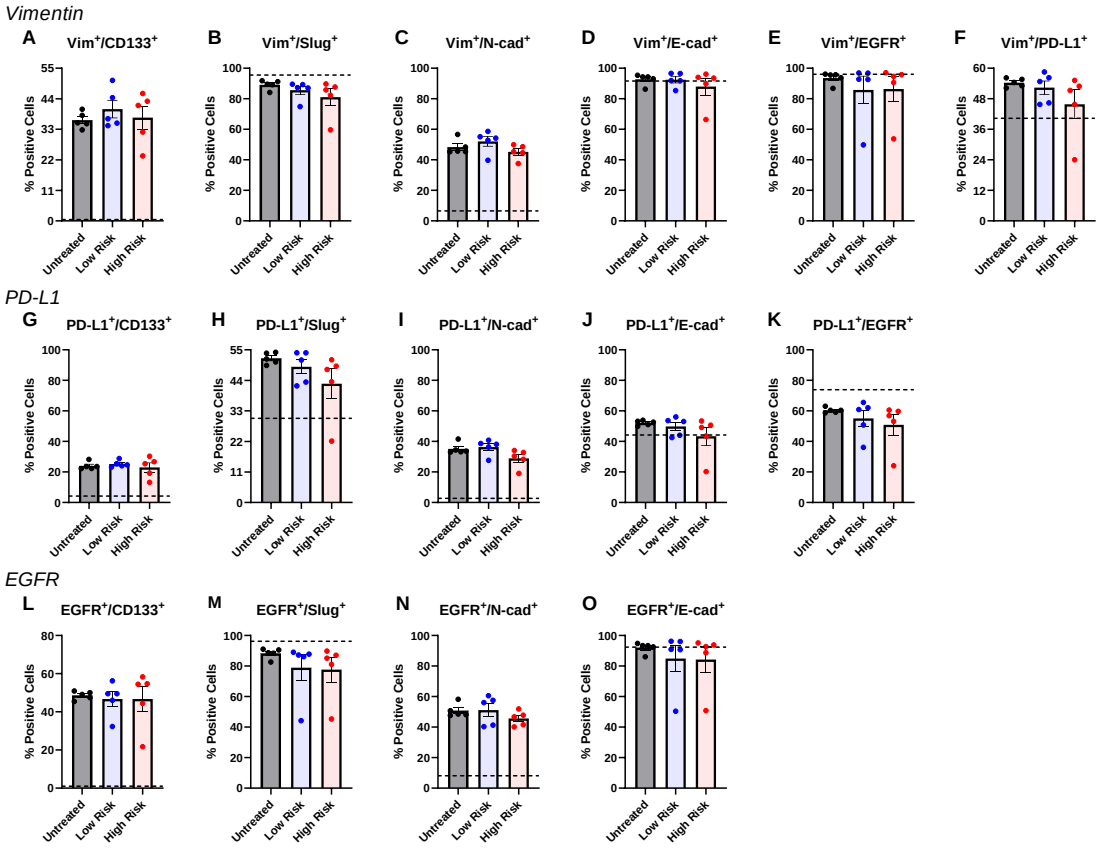


Figure 6.24 Treatment with low- of high-risk PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on H6c7-normal cells. Percentage of H6c7-normal cells co-expressing phenotypic and functional markers of interest (**A-O**) post-treatment with 5% (v/v) low- or high-risk PCF for 6 h. Dotted lines indicate the basal expression level of H6c7-normal cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

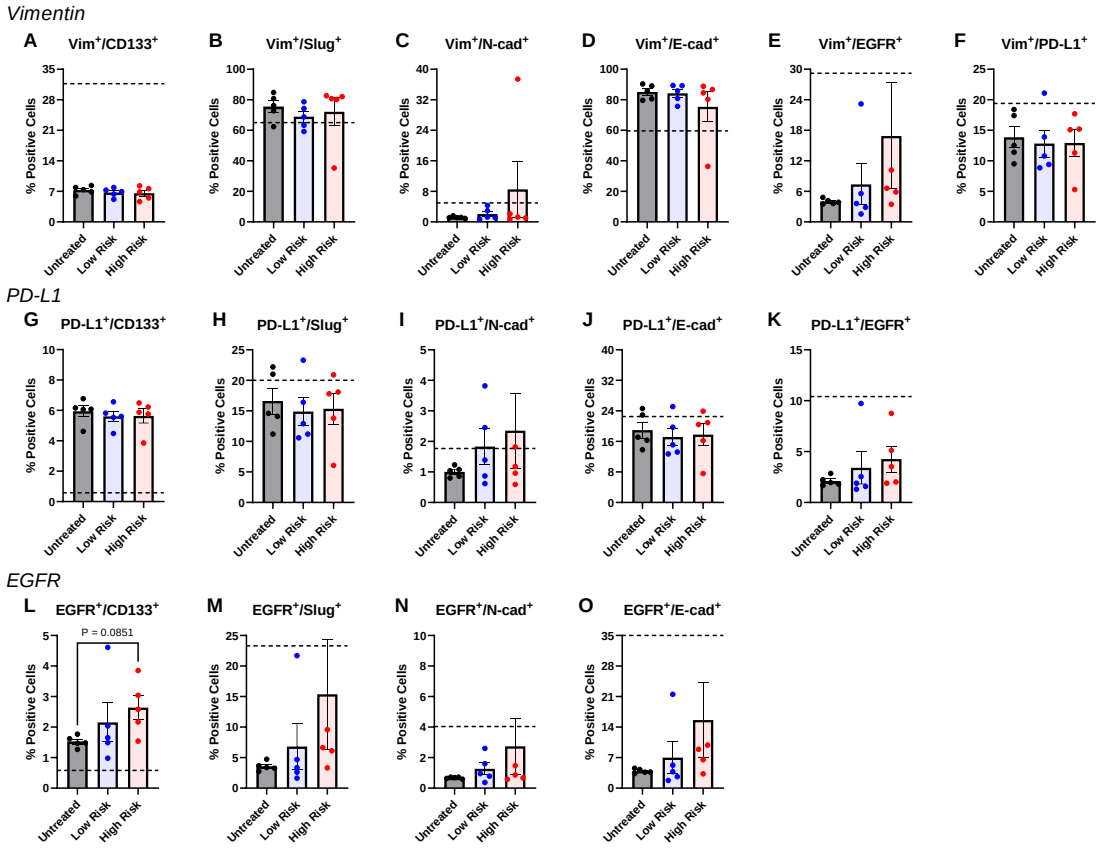


Figure 6.25 Treatment with low- or high-risk PCF for 6 h does not significantly alter the co-expression of phenotypic and functional markers on HPNE-intermediary cells. Percentage of HPNE-intermediary cells co-expressing phenotypic and functional markers of interest (**A-G**) post-treatment with 5% (v/v) low- or high-risk PCF for 6 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

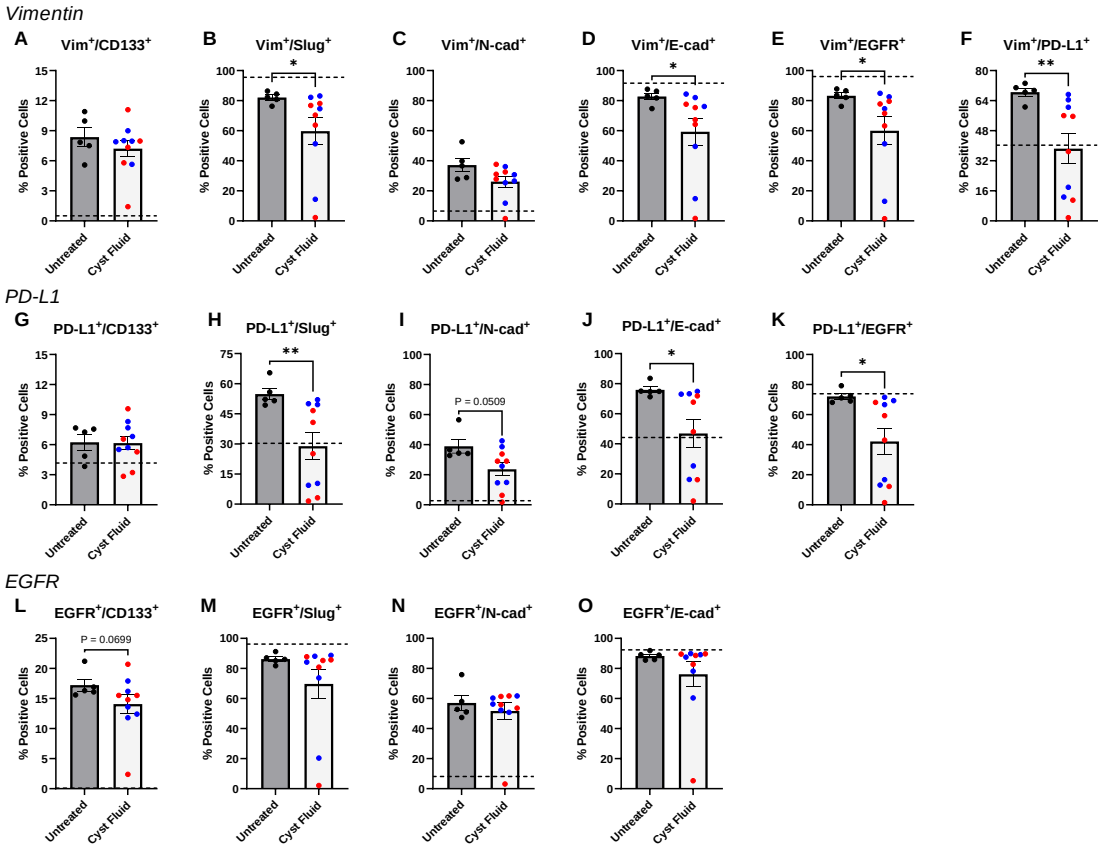


Figure 6.26 Treatment with PCF for 24 h significantly decreases the percentage of Vim$^+$/Slug$^+$, Vim$^+$/E-cad$^+$, Vim$^+$/EGFR$^+$, Vim$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cad$^+$ and PD-L1$^+$/EGFR$^+$ H6c7-normal cells. Percentage of H6c7-normal cells expressing co-expressing phenotypic and functional markers of interest (**A-O**) post-treatment with 5% (v/v) PCF for 24 h. Dotted lines indicate the basal expression level of H6c7-normals cells for each co-expression. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test. *p<0.05, **p<0.01. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

decreases the percentage of Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cadherin$^+$ and PD-L1$^+$/EGFR$^+$ H6c7-normal cells after 24 h of treatment (p<0.05)(Figure 6.27). Importantly, when the PCF treatments are separated out, neither low- nor high-risk PCF on their own retains the significant decrease in the percentage of Vimentin$^+$/Slug$^+$, Vimentin$^+$/E-cadherin$^+$ or Vimentin$^+$/EGFR$^+$ H6c7-normal cells following 24 h PCF treatment (Figure 6.27). Again, at the 24 h timepoint the spread amongst the data is quite large, possibly contributing to the loss of statistical power between low- and high-risk treatments when separated out. Indeed, there is a decrease in the percentage of Vimentin$^+$/Slug$^+$, Vimentin$^+$/E-cadherin$^+$, Vimentin$^+$/EGFR$^+$ and PD-L1$^+$/N-cadherin$^+$ H6c7-normal cells when treated with high-risk PCF, though not significant.

Interestingly, when comparing the percentage of those significant cell subsets after 6 h and 24 h of PCF treatment, only the percentages of Vimentin$^+$/Slug$^+$, Vimentin$^+$/E-cadherin$^+$ and Vimentin$^+$/EGFR$^+$ H6c7-normal cells are decreased at 24 h compared to 6 h post-treatment with PCF (p<0.01, p<0.0001 and p<0.01, respectively)(Figure 6.28A-C). When divided into low- and high-risk treatments, neither alone significantly altered the percentage of Vimentin$^+$/Slug$^+$ H6c7-normal cells (p=0.0814)(Figure 6.28E). Both low- and high-risk PCF significantly decreased the percentage of Vimentin$^+$/E-cadherin$^+$ H6c7-normal cells (p<0.05)(Figure 6.28F), however, only high-risk PCF caused a significant decreased in the percentage of Vimentin$^+$/EGFR$^+$ cells (p<0.05)(Figure 6.28F). Serum-starvation significantly increased the percentage of Vimentin$^+$/PD-L1$^+$ H6c7-normal cells (p<0.05)(Figure 6.28D), but PCF alone or separated into low- and high-risk treatments did not significantly alter these cells (p>0.05)(Figure 6.28D and 6.28H).There was no significant difference in the percentage of PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cadherin$^+$ or PD-L1$^+$/EGFR$^+$ H6c7-normal cells at 6 hr compared to 24 h post-treatment with PCF (p>0.05)(Figure 6.28I-K), and separating into low- or high-risk treatment did not change this (Figure 6.28L-N).Interestingly, while both serum-starvation and PCF treatment appear to cause a decrease in the percentage of Vimentin$^+$/Slug$^+$, Vimentin$^+$/E-cadherin$^+$ and Vimentin$^+$/EGFR$^+$ H6c7-normal cells, an opposing effect can be seen in the percentage of Vimentin$^+$/PD-L1$^+$, PD-L1$^+$/E-cadherin$^+$ and PD-L1$^+$/EGFR$^+$ H6c7-normal cells, where serum-starvation appears to increase the

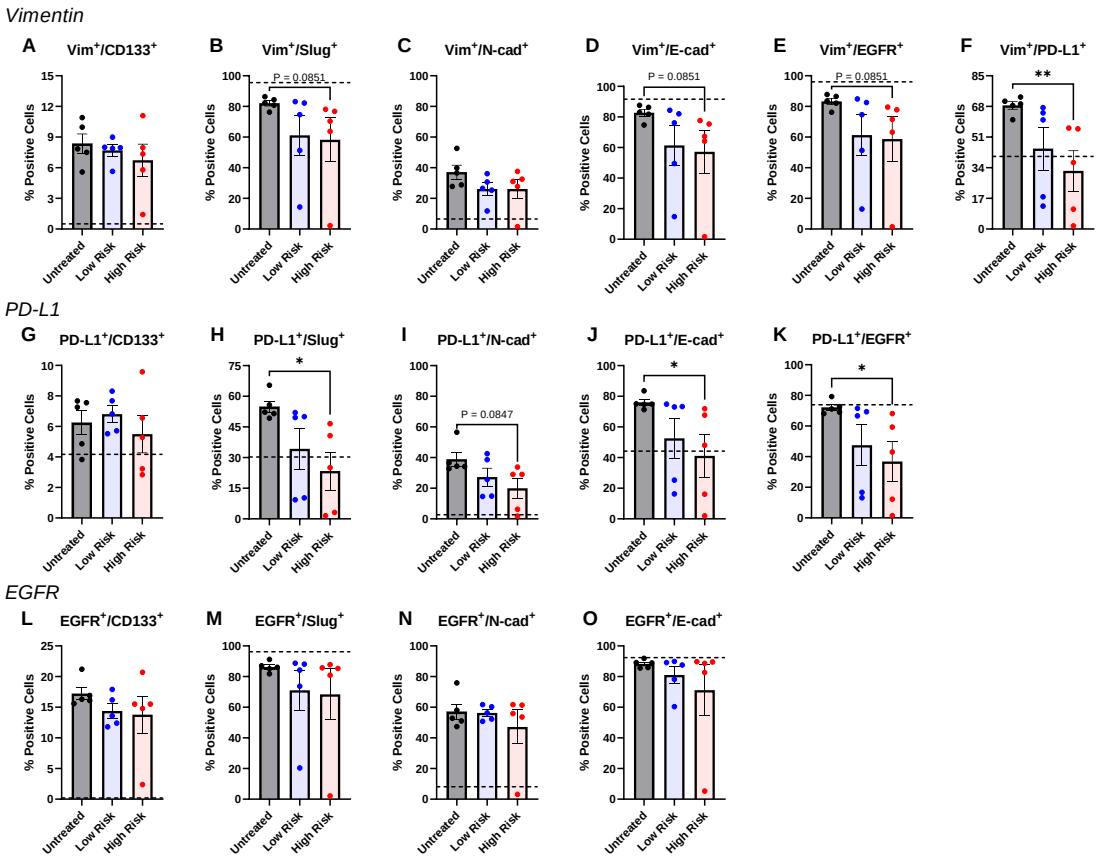


Figure 6.27 Treatment with high-risk PCF for 24 h significantly decreases the percentage of Vim$^+$/PD-L1$^+$, PD-L1$^+$/Slug$^+$, PD-L1$^+$/E-cad$^+$ and PD-L1$^+$/EGFR$^+$ H6c7-normal cells. Percentage of H6c7-normal cells co-expressing phenotypic and functional markers of interest (**A-O**) post-treatment with 5% (v/v) low- or high-risk PCF for 24 h. Dotted lines indicate the basal expression level of H6c7-normal cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05, **p<0.01. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

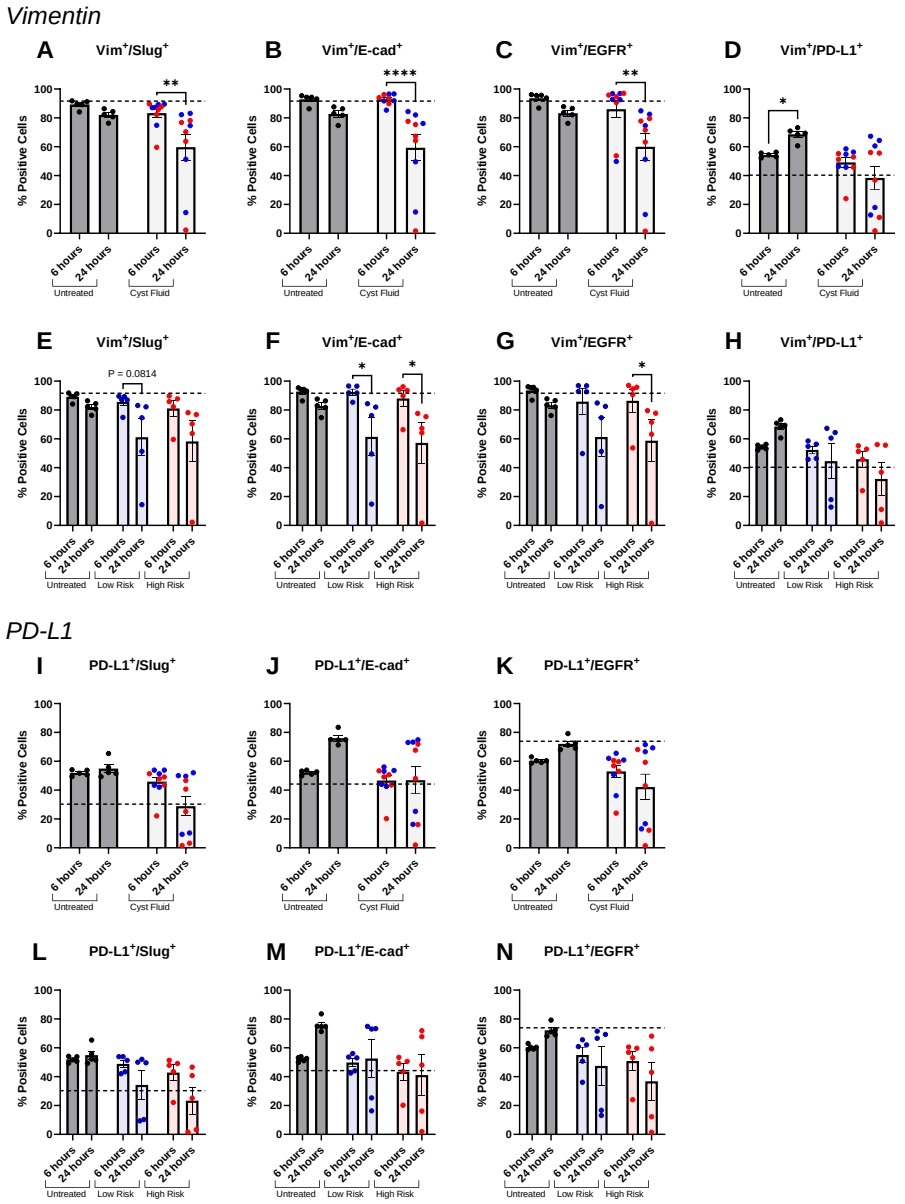


Figure 6.28 Low- and/or high-risk PCF significantly decreases the percentage of Vim$^+$/Slug$^+$, Vim$^+$/E-cad$^+$ and Vim$^+$/EGFR$^+$ H6c7-normal cells at 24 h compared to 6 h. Percentage of H6c7-normal cells co-expressing phenotypic and functional markers of interest at 6 h and 24 h (**A-D** and **I-K**) post-treatment with 5% (v/v) PCF and (**E-H** and **L-N**) post-treatment with 5% (v/v) low- or high-risk PCF. Dotted lines indicate the basal expression level of H6c7-normal cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05, **p<0.01, ****p<0.0001. [E-cad = E-cadherin, Vim = Vimentin]

percentage of cells co-expressing these markers, while PCF treatment decreases them (Figure 6.28). Serum-starvation caused no notable change in the percentage of PD-L1$^+$/Slug$^+$ H6c7-normal cells, with PCF treatment causing a visible decrease (Figure 6.28I). The effects of PCF on the percentage of H6c7-cells co-expressing phenotypic and functional markers of interest is more pronounced at 24 h than 6 h, overall.

6.5.9 PCF significantly increases the percentage of Vimentin$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cadherin$^+$ HPNE-intermediary cells after 24 h

Co-expression of these phenotypic and functional markers was also assessed for HPNE-intermediary cells following treatment with 5% (v/v) PCF for 24 h. Treatment with PCF for 24 h significantly increased the percentage of Vimentin$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cadherin$^+$ HPNE-intermediary cells (p<0.05)(Figure 6.29). Serum-starvation noticeably decreased the percentage of these cells compared to basal levels, with PCF treatment causing them to significantly increase back towards basal levels (p<0.05). No change was observed following serum-starvation or PCF treatment in the percentage of Vimentin$^+$/Slug$^+$, Vimentin$^+$/N-cadherin$^+$, Vimentin$^+$/E-cadherin$^+$, PD-L1$^+$/E-cadherin$^+$ or EGFR$^+$/N-cadherin$^+$ HPNE-intermediary cells. Serum-starvation for 24 h resulted in a visible decrease in the percentage of Vimentin$^+$/CD133$^+$, Vimentin$^+$/PD-L1$^+$ and PD-L1$^+$/Slug$^+$ HPNE-intermediary cells compared to basal levels, and PCF treatment did not significantly alter this. Conversely, serum-starvation noticeably increased the percentage of PD-L1$^+$/CD133$^+$, PD-L1$^+$/N-cadherin$^+$ and EGFR$^+$/CD133$^+$ HPNE-intermediary cells, with PCF treatment causing no significant change to these levels (Figure 6.29).

When separated into low- and high-risk PCF treatments, low-risk PCF significantly increases the percentage of Vimentin$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cadherin$^+$ HPNE-intermediary cells following 24 h treatment (p<0.05)(Figure 6.30). Importantly, when the PCF treatments are separated out, neither low nor high-risk PCF on their own retains the significant increase in the percentage of PD-L1$^+$/EGFR$^+$ HPNE-intermediary cells. Again, at the 24 h timepoint the spread amongst the data is quite large, with many datapoints being quite far from the mean. This issue is still apparent

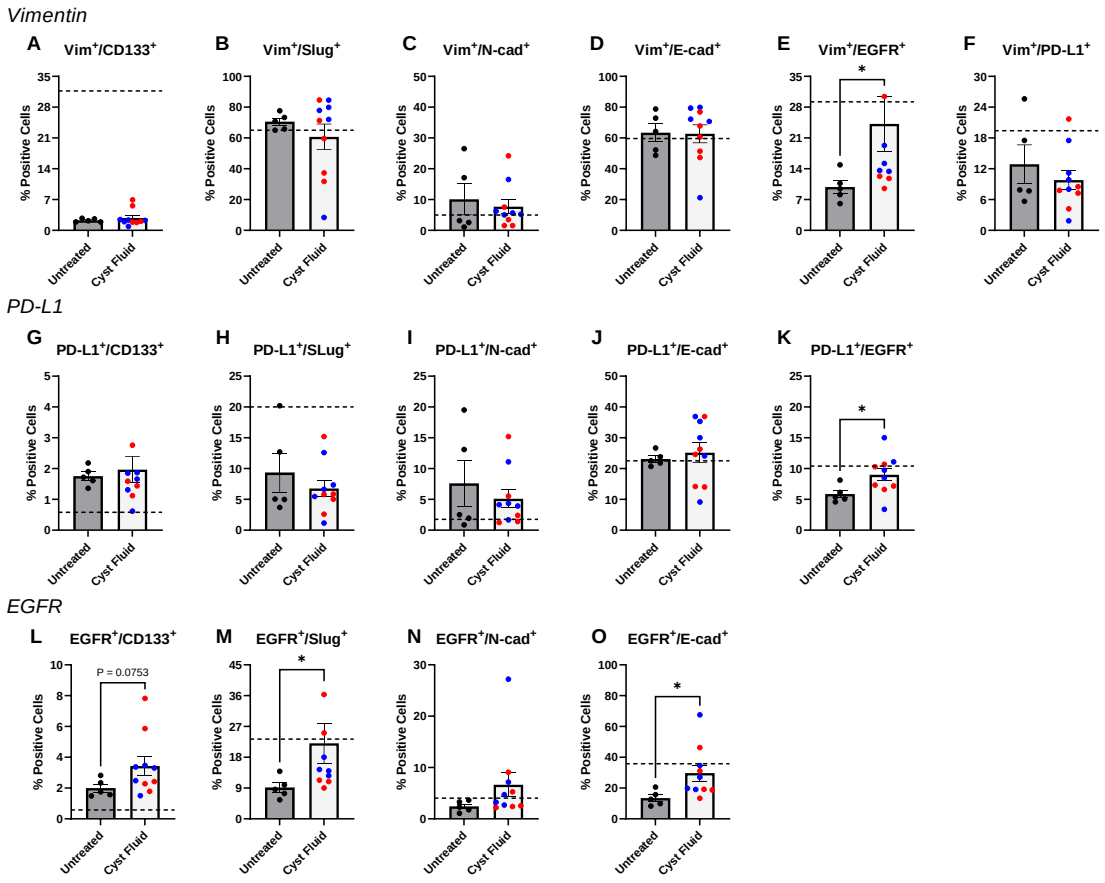


Figure 6.29 Treatment with PCF for 24 h significantly increases the percentage of Vim$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cad$^+$ HPNE-intermediary cells. Percentage of HPNE-intermediary cells expressing co-expressing phenotypic and functional markers of interest (**A-O**) post-treatment with 5% (v/v) PCF for 24 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each co-expression. Data are presented as mean ± SEM for 10 patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Mann-Whitney test. *p<0.05. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

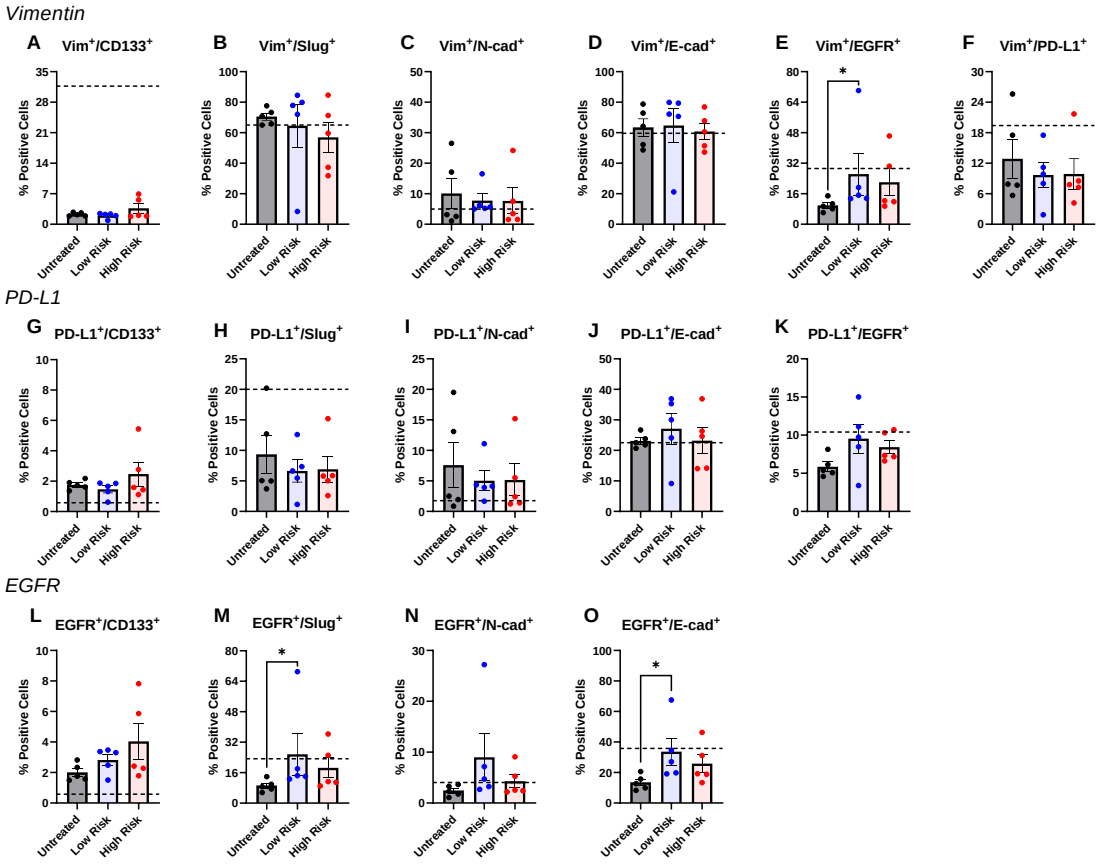


Figure 6.30 Treatment with low-risk PCF for 24 h significantly increases the percentage of Vim$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cad$^+$ HPNE-intermediary cells. Percentage of HPNE-intermediary cells co-expressing phenotypic and functional markers of interest (**A-O**) post-treatment with 5% (v/v) low- or high-risk PCF for 24 h. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. *p<0.05. [E-cad = E-cadherin, N-cad = N-cadherin, Vim = Vimentin]

when the data are separated into low- and high-risk, indicating again that the patient-to-patient variation on PCF effect is quite high, and that risk is not the dependent factor.

Examining those subsets of cells that were significantly altered by PCF at 24 h post-treatment and comparing this to the effects seen at 6 h, these changes are significantly more pronounced after 24 h (Figure 6.31). PCF treatment significantly increased the percentage of Vimentin$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cadherin$^+$ HPNE-intermediary cells after 24 h compared to 6 h (p<0.01)(Figure 6.31A-D). For each of these subsets, low-risk PCF significantly increased the percentage of these cells (p<0.05), with high-risk PCF causing no significant change (p>0.05)(Figure 6.31A, C and D). Serum-starvation also caused a visible increase in the percentage of Vimentin$^+$/EGFR$^+$, PD-L1$^+$/EGFR$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cadherin$^+$ HPNE-intermediary cells, though not significant (p>0.05)(Figure 6.31E-H).

6.5.10 PCF significantly increases the invasive capacity of normal pancreatic cell lines

Given the changes in the expression of several markers seen in both normal pancreatic cell lines, the functional consequences of these changes were subsequently assessed. To examine whether treatment with PCF effects the functionality of normal pancreatic cell lines, H6c7-normal and HPNE-intermediary, the invasive potential of these cells was examined after 24 h of treatment with 5% (v/v) PCF. The percentage of H6c7-normal cells that invaded through a Collagen I coated membrane after treatment with PCF was significantly increased compared to serum-starved, untreated controls (p<0.001)(Figure 6.32A). When separated into low- and high-risk PCF treatments, while both caused a slight increase in the percentage of invasion, high-risk PCF had a significant increase compared to the untreated control (p<0.001)(Figure 6.32B). Importantly, though this is a 24 h timepoint and the total patient number here is quite large, with eight low-risk and nine high-risk patients, the spread of these data is very small. All datapoints cluster together nicely, with no defined outliers.

For HPNE-intermediary cells, PCF treatment did not significantly alter the percentage of cells that invaded through the Collagen I coated membrane after 24 h (Figure 6.33A). However, when divided into low- and high-risk PCF treatment, there was

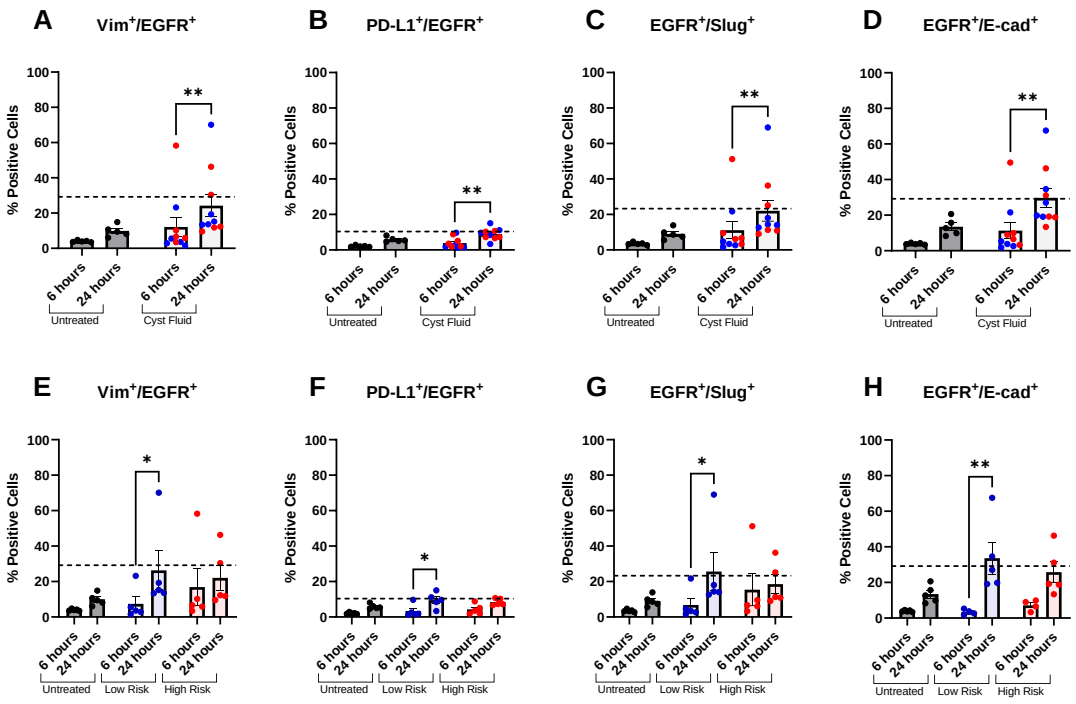


Figure 6.31 PCF significantly increases the percentage of Vim⁺/EGFR⁺, PD-L1⁺/EGFR⁺, EGFR⁺/Slug⁺ and EGFR⁺/E-cad⁺ HPNE-intermediary cells at 24 h compared to 6 h. Percentage of HPNE-intermediary cells co-expressing phenotypic and functional markers of interest at 6 h and 24 h (**A-D**) post-treatment with 5% (v/v) PCF and (**E-H**) post-treatment with 5% (v/v) low- or high-risk PCF. Dotted lines indicate the basal expression level of HPNE-intermediary cells for each co-expression. Data are presented as mean ± SEM for n=5 low-risk and n=5 high-risk patients. PCF from low-risk patients is shown in blue; PCF from high-risk patients is shown in red. Kruskal-Wallis test with Dunn's multiple comparisons. **p*<0.05, ***p*<0.01. [E-cad = E-cadherin, Vim = Vimentin]

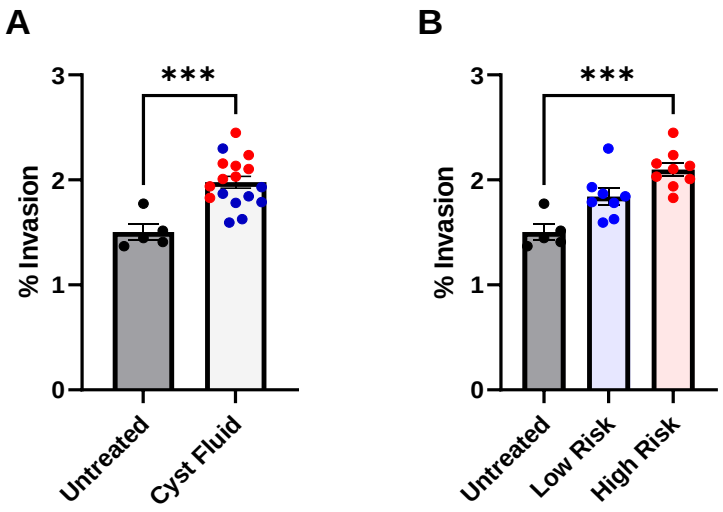


Figure 6.32 Treatment with high-risk PCF for 24 h significantly increases the invasive capacity of H6c7-normal cells. (**A**) Percentage invasion through a Collagen I coated matrix of H6c7-normal cells treated with 5% (v/v) PCF for 24 h. Data are presented as mean ± SEM for 17 patients. Mann Whitney test. (**B**) Percentage invasion through a Collagen I coated matrix of H6c7-normal cells treated with 5% (v/v) low-risk or high-risk PCF for 24 h. Data are presented as mean ± SEM for n=8 low-risk patients and n=9 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. ***p<0.001.

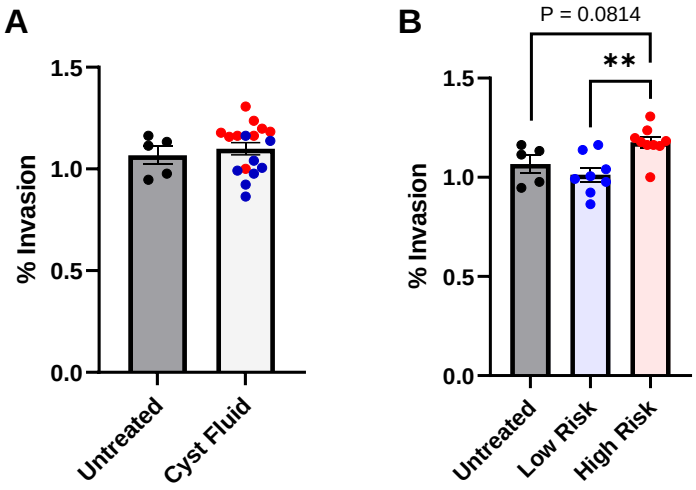


Figure 6.33 Treatment with high-risk PCF for 24 h significantly increases the invasive capacity of HPNE-intermediary cells. (**A**) Percentage invasion through a Collagen I coated matrix of HPNE-intermediary cells treated with 5% (v/v) PCF for 24 h. Data are presented as mean ± SEM for 17 patients. Mann Whitney test. (**B**) Percentage invasion through a Collagen I coated matrix of HPNE-intermediary cells treated with 5% (v/v) low-risk or high-risk PCF for 24 h. Data are presented as mean ± SEM for n=8 low-risk patients and n=9 high-risk patients. Kruskal-Wallis test with Dunn's multiple comparisons. ***p*<0.01.

a significant increase in the percentage of HPNE-intermediary cells that invaded through the membrane when treated with high-risk PCF compared to those treated with low-risk PCF (p<0.01)(Figure 6.33B). Furthermore, there is a visible increase in the percentage of cells that invaded after treatment with high-risk PCF compared to the untreated control, though not significant (p=0.0814). Similarly to the H6c7-normal cells, the variation between datapoints here is very small, with low- and high-risk PCF clustering well into distinct groups, possibly indicating that the functional effects of PCF are more homogenous across patient samples.

6.5.11 PCF does not significantly alter the production of 8-OHdG by normal pancreatic cell lines

Finally, to assess whether treatment with PCF was harmful and induced DNA damage in normal pancreatic cell lines, H6c7-normal and HPNE-intermediary, the supernatants from these treatments were examined for 8-OHdG levels (DNA damage marker) via competitive ELISA. The supernatants from H6c7-normal cells treated with PCF had significantly higher levels of 8-OHdG than their matched PCF samples, when diluted to the same concentration of 5% (v/v) PCF (p<0.001)(Figure 6.34A). When separated out into low- and high-risk, the levels of 8-OHdG in the supernatants are still significantly higher than their respective matched PCF samples (p<0.05)(Figure 6.34B-C). For HPNE-intermediary cells, overall the concentrations of 8-OHdG is also higher in the supernatants of cells treated with PCF than in the matched PCF samples (p<0.01)(Figure 6.35A). While this remains significantly increased in low-risk supernatants compared to matched PCF samples (p<0.05), there is no significant difference in the levels of 8-OHdG in high-risk supernatants compared to the matched PCF samples (Figure 6.35B-C).

In order to examine whether the levels of 8-OHdG were increased in the supernatants of cells treated with PCF for 24 h compared to untreated control cell supernatants, the concentrations of 8-OHdG found in the cyst samples were subtracted from their respective supernatants. One caveat to this, is that the concentration of 8-OHdG found in PCF samples differed greatly depending on whether the sample was diluted in the serum-free medium of H6c7-normal cells or HPNE-intermediary cells. Unfortunately, there was also no correlation between the values obtained when diluting

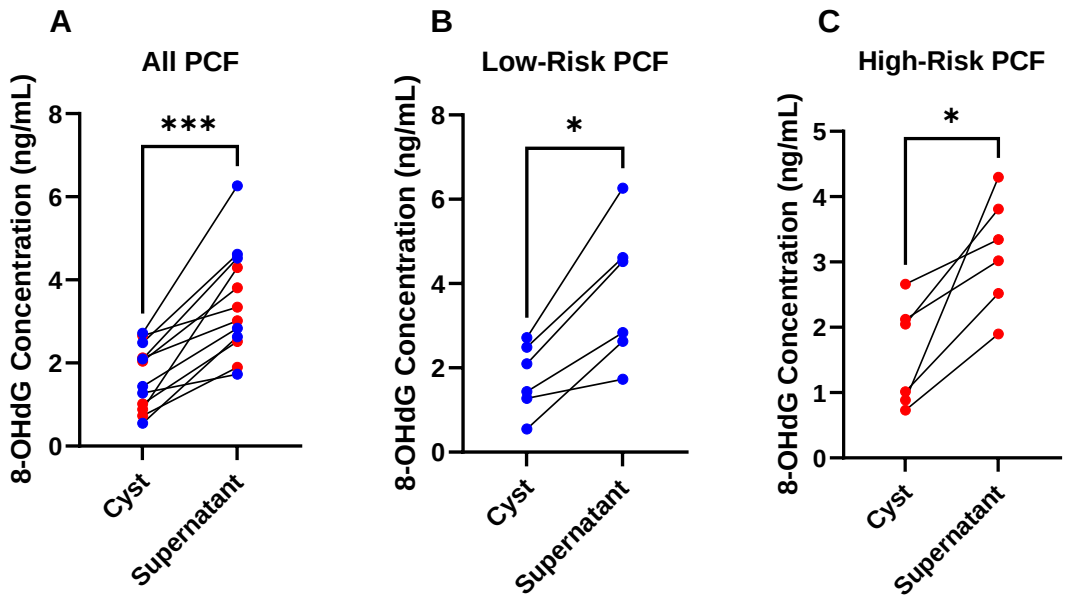


Figure 6.34 8-OHdG levels are significantly increased in the supernatants of H6c7-normal cells treated with PCF for 24 h compared to the matched cyst samples. Line graphs showing 8-OHdG concentrations in the supernatant of H6c7-normal cells treated with 5% (v/v) PCF for 24 h, and the corresponding levels in the matched cyst fluid samples diluted to 5% (v/v) in culture medium. (**A**) 8-OHdG concentrations in all PCF cyst samples and matched supernatants. (**B**) 8-OHdG concentrations in low-risk PCF cyst samples and matched supernatants. (**C**) 8-OHdG concentrations in high-risk PCF cyst samples and matched supernatants. Matched data are joined by a line. Data are presented for n=6 low-risk patients and n=6 high-risk patients. Wilcoxon matched-pairs signed rank test. *p<0.05, ***p<0.001.

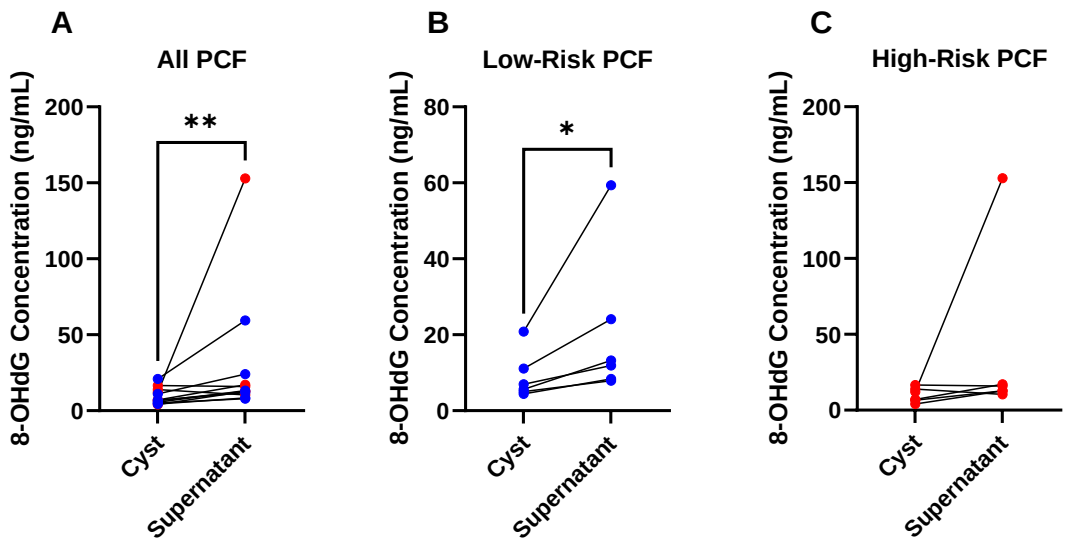


Figure 6.35 8-OHdG levels are significantly increased in the supernatants of HPNE-intermediary cells treated with PCF for 24 h compared to the matched cyst samples. Line graphs showing 8-OHdG concentrations in the supernatant of HPNE-intermediary cells treated with 5% (v/v) PCF for 24 h, and the corresponding levels in the matched cyst fluid samples diluted to 5% (v/v) in culture medium. (**A**) 8-OHdG concentrations in all PCF cyst samples and matched supernatants. (**B**) 8-OHdG concentrations in low-risk PCF cyst samples and matched supernatants. (**C**) 8-OHdG concentrations in high-risk PCF cyst samples and matched supernatants. Matched data are joined by a line. Data are presented for n=6 low-risk patients and n=6 high-risk patients. Wilcoxon matched-pairs signed rank test. *p<0.05, **p<0.01.

PCF in these two mediums, indicating some interference of the cell line mediums with the assay.

There was no significant difference in the concentrations of 8-OHdG in the supernatants of H6c7-normal cells treated with PCF for 24 h (Figure 6.36A), or when separated into low- and high-risk treatments (Figure 6.36B). Similarly, there was no significant change in the concentrations of 8-OHdG in the supernatants of HPNE-intermediary cells treated with PCF for 24 h (Figure 6.37A), nor when treated with low- or high-risk PCF (Figure 6.37B). There is a noticeably large spread of 8-OHdG concentrations in both H6c7-normal and HPNE-intermediary untreated controls, as well as across the PCF treatments, possibly as a result of the interference of the cell culture medium with the performance of the assay.

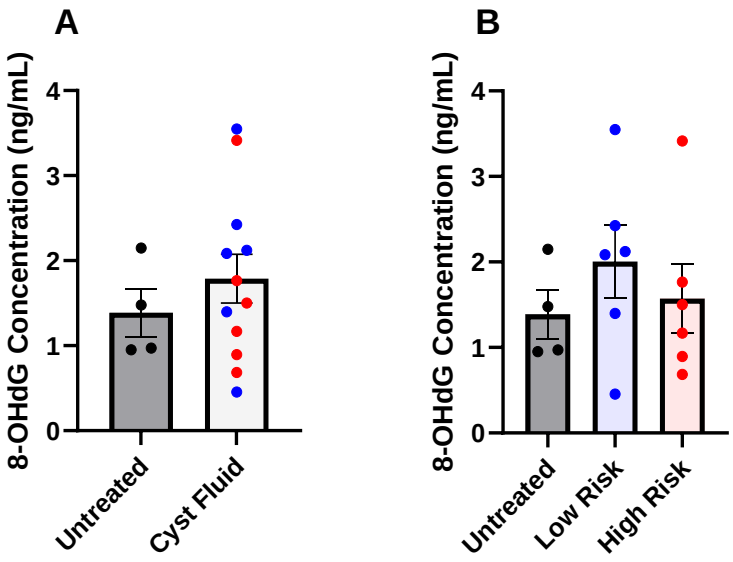


Figure 6.36 8-OHdG levels in H6c7-normal cell supernatants are not significantly altered following 24 h treatment with PCF. 8-OHdG concentrations in the supernatant of H6c7-normal cells untreated or treated with 5% (v/v) PCF for 24 h (where the concentration of 8-OHdG in the PCF has been subtracted from the treatment values). (**A**) Mann-Whitney test comparing 8-OHdG concentrations post-treatment with PCF. (**B**) Kruskal-Wallis test with Dunn's multiple comparisons comparing 8-OHdG concentrations post-treatment with low- or high-risk PCF. Data are presented as mean ± SEM for n=6 low-risk patients and n=6 high-risk patients.

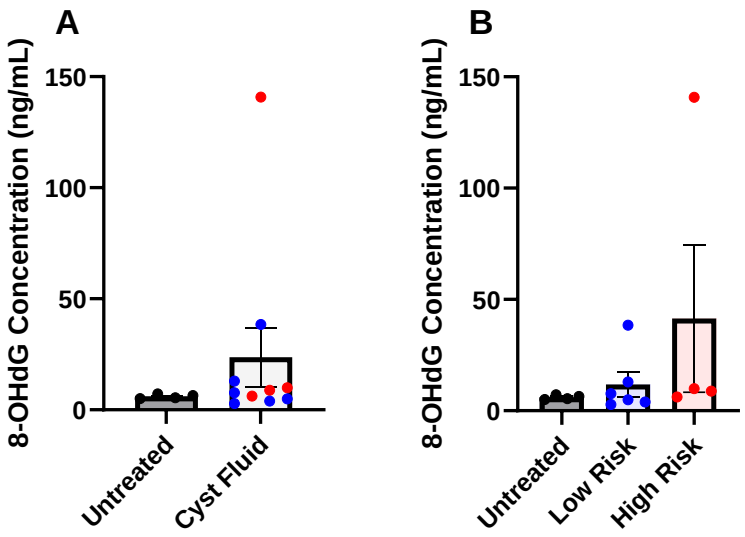


Figure 6.37 8-OHdG levels in HPNE-intermediary cell supernatants are not significantly altered following 24 h treatment with PCF. 8-OHdG concentrations in the supernatant of HPNE-intermediary cells untreated or treated with 5% (v/v) PCF for 24 h (where the concentration of 8-OHdG in the PCF has been subtracted from the treatment values). (**A**) Mann-Whitney test comparing 8-OHdG concentrations post-treatment with PCF. (**B**) Kruskal-Wallis test with Dunn's multiple comparisons comparing 8-OHdG concentrations post-treatment with low- or high-risk PCF. Data are presented as mean ± SEM for n=6 low-risk patients and n=6 high-risk patients.

6.6 Discussion

PCF can cause changes to normal pancreatic cell biology, affecting viability, apoptosis, proliferation, cell phenotypic and functional markers, metabolism and invasive potential. Importantly, the data presented here are for those PCF samples that did not kill the majority of cells that were exposed to the fluid for just 24 h. In some instances, treatment with PCF killed almost all exposed cells, and as such readings for these samples could not be recorded or included in the data shown. As such, the data shown should be regarded as the results of the less cytotoxic PCF samples, and not as a representation of the effects of all PCF samples on normal cell biology. Overall, the effects elicited by PCF were largely distinct between the H6c7-normal and HPNE-intermediary cells, illustrating what are, potentially, two different stages in the process of malignant transformation.

H6c7-normal cells, are immortalized epithelial cells derived from the duct of a normal pancreas. IPMNS are PCLs that occur in the main or branch ducts of the pancreas and represent approximately half of all PCLs, making H6c7-normal cells a good model of the normal cells that would precede IPMN development[456]. Exposure to just 5% (v/v) PCF significantly altered H6c7-normal cell biology, resulting in increased viability, decreased apoptosis, an increase in glycolytic metabolism, increased invasive potential and decreases in a range of phenotypic and functional markers. Indeed, H6c7-normal cells appeared to thrive in the presence of this low concentration of PCF. Vimentin is often referred to as a key regulator of EMT, being increased in cells that lose adhesion molecules and become more invasive[457]. In PC, vimentin expression has been shown to be three-fold higher than in other tumours, and to increase the invasive potential of tumour cells[458, 459]. Indeed, the presence of Vimentin$^+$ circulating tumour cells in PC patients has been shown to correlate with tumour burden and prognosis[460], with high vimentin expression in PC tumours predicting poor prognosis[461, 462]. Serum-starvation and environmental stress were also demonstrated to increase vimentin expression on PC cells[463]. Here, it was shown that exposure to PCF causes a significant reduction in the percentage of Vimentin$^+$ H6c7-normal cells, with several subsets of Vimentin$^+$ cells also being significantly reduced. Exposure to PCF, and more specifically high-risk PCF, also caused a decrease in PD-L1 expression on H6c7-normal cells. The binding of PD-L1

to its PD-1 receptor on T cells can lead to cell exhaustion or apoptosis[96]. Given that pancreatic tumours are notoriously immunologically cold, with limited T cell infiltration, the role of T cells and PD-L1 in the pancreas and PC is not very well understood[464]. However, a meta-analysis compiling data from nine studies found that increased PD-L1 expression in PC tumours positively correlated with poor prognosis[95]. This is in line with many other cancers, such as lung[465], ovarian[466], breast[467] and renal cancer[468]. Interestingly, while serum-starvation appeared to cause an increase in the percentage of PD-L1$^+$ H6c7-normal cells, PCF exposure appears to dampen this, keeping PD-L1 expression closer to the lower levels that were recorded basally. Interestingly, the same subsets of vimentin and PD-L1 expressing cells were significantly reduced following PCF treatment, those expressing vimentin or PD-L1 and either slug, EGFR or e-cadherin. These decreases in PD-L1$^+$/EGFR$^+$ and Vimentin$^+$/EGFR$^+$, PD-L1$^+$/Slug$^+$ and Vimentin$^+$/Slug$^+$, and PD-L1$^+$/E-cadherin$^+$ and Vimentin$^+$/E-cadherin$^+$ Hc67-normal cells following PCF treatment indicates a decrease in a subset of EGFR$^+$, Slug$^+$ and E-cadherin$^+$ cells. Similarly to vimentin and PD-L1, high EGFR expression is found in up to 90% of PC tumours and is associated with a poor prognosis[469, 470]. Slug is an EMT-inducing transcription factor that is generally upregulated during EMT and whose high expression is also an indicator of poor prognosis in PC[471, 472]. Reductions in these subsets of cells expressing indicators of poor prognosis, therefore, would be favourable and indicate a potential anti-cancer environment. Distinctly, low e-cadherin expression is linked to poor prognosis in breast cancer[473]. However, an overall decrease in e-cadherin is not evident, despite these decreases in small subsets of E-cadherin$^+$ cells. Serum-starvation, a notoriously stressful environment for cell lines, and particularly normal cell lines, caused an increase in PD-L1$^+$/Vimentin$^+$, PD-L1$^+$/EGFR$^+$ and PD-L1$^+$/E-cadherin$^+$ cell subsets, while PCF had the opposite effect and significantly decreased the percentage of these cells.

As well as effecting the expression of phenotypic and functional markers on these normal cells, PCF also conferred a functional change onto the H6c7-normal cells. Indeed, PCF exposure caused a significant increase in the invasive potential of H6c7-normal cells through a type I collagen membrane. Conversely, while these cells have been shown to have increased invasive potential following treatment with both low-

and high-risk PCF, a significant reduction in vimentin was seen in these cells, as mentioned above. Crucially, a 2011 study showed that premalignant pancreatic HPDE cells that were plated in culture dishes coated with type I collagen had a 3-fold increase in snail expression, another well-known regulator of EMT, compared to cells plated in uncoated dishes[474]. Indeed, Shields *et al.* revealed that PC cells upregulate snail expression in order to promote membrane type 1-matrix metalloproteinase-dependent collagen I invasion[474]. Similarly, a 2014 study showed an increase of vimentin and n-cadherin on PC cells exposed to type I collagen, which was linked to a subsequent increase in cell migration and invasion[475]. In this way, the observed decrease in vimentin expression of H6c7-normal cells following PCF treatment via flow cytometry is independent of the increase in invasion caused by exposure to PCF in the presence of type I collagen. This decrease in vimentin expression, coupled with no change in e-cadherin or n-cadherin, of H6c7-normals cells treated with PCF is a unique phenomenon. While there was a significant decrease in the percentage of cells that were Vimentin$^+$/E-cadherin$^+$, there was no change in Vimentin$^+$/N-cadherin$^+$ cells, suggesting that vimentin expressing cells may have downregulated e-cadherin expression, but there was no change in their expression of n-cadherin. As no change was observed in any of these markers at 6 h, perhaps 24 h is also too short of a time point to see the full effect of PCF on EMT markers. Indeed, the percentage of Vimentin+ H6c7-normal cells was shown to be significantly decreased at 24 h post-PCF treatment compared to 6 h, indicating that further powering of these data, both in the presence and absence of type I collagen, and at longer incubation times, is required to fully elucidate the effects of PCF on EMT.

PCF treatment also produced a slight metabolic shift from oxidative phosphorylation to glycolysis in H6c7-normal cells, though not a significant change. Indeed, serum-starvation was shown to have more of an impact on H6c7-normal cell metabolism than PCF, causing a shift towards a more glycolytic phenotype. A shift towards an upregulation of glycolysis is observed in many cancer types, and can be the result of various factors such as environment, mutations, activation of oncogenes or loss of tumour suppressors[476]. Serum-starvation results in the inadequate uptake of nutrients by cells which leads to oxidative stress[477]. Importantly, nutrient deprivation

can drive cancer cells to utilize glycolysis to produce ATP, as is seen here in the H6c7-normal cells[478]. The extensive effects of environmental stress and nutrient deprivation on cellular metabolism are still being understood, and as such the condition of serum-starvation in this setting may be overshadowing those of PCF exposure, making them difficult to discern. Future experiments aiming to fully elucidate alterations to cellular metabolism caused by PCF should consider doing so in normal growth conditions also.

Overall, it appears that H6c7-normal cells thrive under PCF exposure. The downregulation of apoptosis and increased viability of the cells, coupled with the decreased in vimentin and PD-L1 expression, suggests that serum-starvation is more stressful then exposure to PCF. However, it is important to note that despite the trends seen in these data, the overall distribution of datapoints was extremely large in most instances. The effects of PCF on H6c7-normal cell biology varied greatly from patient to patient, with risk stratification doing little to solve this. While the overall theme from the data shown suggests that H6c7-normal cells respond well to PCF, in fact this is the response of the cells to PCF that did not kill the majority of the population after just 24 h of exposure. Indeed, while the data shown indicate no significant increase in DNA damage following treatment with PCF, there is a visible increase that may become more pronounced when issues with culture medium interference are eliminated. Moreover, a starkly different, and less favourable, reaction to PCF exposure can be seen in the HPNE-intermediary cells.

HPNE-intermediary pancreatic cells represent an intermediary stage during acinar-to-ductal metaplasia of the pancreas[479]. While defined as ‘normal’, given that this is a non-malignant cell line, it is clear that these cells are an intermediate phase between normal and malignant pancreatic cells. As such, they represent a step forwards on the ladder from H6c7-normal cells towards PC. Interestingly, the effects of PCF exposure on HPNE-intermediary cells are much more cytotoxic in nature. Exposure to 5% (v/v) PCF significantly increased apoptosis and decreased proliferation, as well as altering the expression of certain phenotypic and functional markers in HPNE-intermediary cells. PCF, overall, had significant cytotoxic effects on these cells when all PCF samples were included, which is distinctly different from the outcomes shown in H6c7-normal cells. The percentage of $EGFR^+$ HPNE-intermediary cells was significantly

increased following PCF treatment. Several subsets of EGFR$^+$ HPNE-intermediary cells, namely EGFR$^+$/Vimentin$^+$, EGFR$^+$/PD-L1$^+$, EGFR$^+$/Slug$^+$ and EGFR$^+$/E-cadherin$^+$, were also significantly increased following treatment with PCF for 24 h. These effects were significantly more pronounced at 24 h post-treatment compared to 6 h, suggesting again that 6 h is not a sufficient timeframe to evaluate these effects. Importantly, unregulated EGFR activity can promote an oncogenic phenotype by facilitating increased proliferation, migration and invasion, and a resistance to apoptotic signals[470]. Here, the opposite effect is observed, where increased EGFR expression is coupled with increased apoptosis and decreased proliferation. While this would appear to be counterintuitive, it is important to note that negative regulators of the EGFR pathway have been documented in PC. Indeed, carboxyl terminus of heat shock protein 70-interacting protein (CHIP) is a ubiquitin ligase that has been shown to ubiquitinate EGFR in PC, subsequently suppressing cancer cell growth, invasion and migration, as well as increasing apoptosis in cells exposed to erlotinib[480]. CHIP has been shown to be present in the cytoplasm of cells, and is found at significantly lower levels in PC tissues compared to adjacent normal tissues[481]. Indeed, Wang *et al.* demonstrated that high CHIP expression is associated with better overall survival in PC patients, and that serum levels of CHIP are significantly higher in normal controls and patients with chronic pancreatitis compared to PC patients[481]. As such, while the percentage of EGFR$^+$ HPNE-intermediary cells are shown to be significantly increased following exposure to PCF, the potential role of CHIP in suppressing the effects often seen following increased EGFR expression in this setting remains to be fully understood. Serum-starvation, in the case of the HPNE-intermediary cells, also had profound influences on the expression of multiple phenotypic and function markers, especially the subsets of cells co-expressing certain markers. Interestingly, serum-starvation caused a decrease in EGFR expression on HPNE-intermediary cells, which is contrary to the alterations triggered by PCF exposure. It has been shown that serum-starvation of multicellular aggregates of cancer cells can alter the activation pathway of EGFR in these cells as they overcome nutrient deprivation[482]. Factors such as Rab7 have been shown to be crucial in maintaining EGFR levels, especially in serum-starved conditions, as it is believed to play a protective role against proteosome-mediated degradation[483]. As such, serum-starvation has a unique

impact on EGFR levels that, while cancer cells can evolve to overcome, normal cells may find to be detrimental to their survival.

PCF exposure for 24 h did not significantly alter the metabolic profile of HPNE-intermediary cells, likely due to the reasons outlined above in relation to serum-starvation. It did, however, cause an increase in the percentage of cells that invaded through a type I collagen matrix. Here, while no significant change was seen in the expression of EMT markers by flow cytometry, this increase is possibly due to the combined presence of the collagen I and PCF, as seen in the H6c7-normal cells. Lastly, there was no significant increase in the production of the DNA damage marker 8-OHdG by HPNE-intermediary cells exposed to PCF, but again, these data are known to be flawed and should therefore be taken with uncertainty.

Overall, these results indicate that PCF is biologically active, and has differential effects on H6c7-normal and HPNE-intermediary cells. These findings are in-line with some of the research conducted in OC ascites, where different effects are seen depending on the patient sample used, and also on the phenotypic profile of the cell line exposed. These results, however, represent a small snapshot into the effects of PCF on the cells surrounding PCLs *in vivo*. Prolonged exposure to PCF, given the impacts shown here on cell death, proliferation, viability, EMT, invasion, metabolism and DNA damage, could be extremely harmful to normal pancreatic cells. Indeed, the significant alterations shown here on several of the hallmarks of cancer, were obtained after exposure to just 5% (v/v) PCF for 24 h. The longitudinal effects of PCF on normal cell biology, as well as the potential impacts of increased concentrations or neat PCF on normal cells remain to be understood. More importantly, the effects of those PCF samples that could not be included in the data shown, as exposure to these PCFs killed the majority of cells, resulting in a lack of usable data, remain to be elucidated. Incubation of these normal cells with neat PCF would be a 20-fold increase in treatment concentration. Given the cytotoxic effects of the PCF on HPNE-intermediary cells at just 5%, examining the effects of such increases in concentration, and whether these cytotoxic effects are exacerbated, are of the utmost importance. Indeed, Mo *et al.* demonstrated that a murine OC cell model that was exposed to 50% (v/v) ascites from OC-bearing mice for 7 days became less sensitive to paclitaxel, a first-line chemotherapy

for OC [484]. In this way, PCF may not only play a role in the progression of PCLs to PC, but could also be contributing to the development of treatment resistance in PC tumours. PCF could play a key role in the development and progression of PC, and as such, understanding the mechanisms involved are vital for the successful advancement of PC research, and also the clinical handling of PC patients. If PCF is shown to play a role in PCL transformation, this could have a profound impact on the current treatment guidelines for PCL patients. Indeed, this could warrant the establishment of regular screening procedures for patients with a family history of PCLs or PC, and a push for PCL aspiration, even in those of a small size or low-risk. Furthermore, if PCF is shown to encourage treatment resistance in PC tumours, aspiration of PCLs in patients with PC may become a necessary procedure in order to counteract this, despite the potential for seeding. While the results shown here are not definitive, the biological activity of PCF has been demonstrated, and its impact on several hallmarks of cancer have been revealed. Further research is now needed to fully understand the pathways and processes involved, and this research is urgently required.

Chapter 7.

General Discussion

7.1 PC – the fight to improve survival

PC has the worst survival rate of any cancer, with a 5-year survival rate in 2023 of just 12%[1]. The abysmal rates of survival seen in PC are largely attributed to the notoriously vague symptoms associated with early stage disease[7]. Common symptoms, such as weight loss or lower back pain, go unnoticed by patients until they have severely worsened, causing patients to therefore present to their GP at a late-stage of development[8, 9]. Patients with neoplastic PCLs are generally at an increased risk of PC development, however, our ability to distinguish PCLs at a low- or high-risk of malignant transformation are limited[121-123]. As such, there is urgent need for novel biomarkers to detect these high-risk patients at an early stage in order to improve the survival rates of this cancer. Furthermore, the mechanisms involved in the malignant transformation of PCLs to PC are not well understood. The role of PCF in this process, if any, remains to be elucidated in the literature. This thesis has provided important advancements on both fronts. Firstly, by identifying promising PCF-based and serum-based multi-omic biomarker panels, as well as a unique and high performing cross-biofluid multi-omic panel, for the distinction of low- and high-risk PCLs. Secondly, by demonstrating the biological activity of PCF, and highlighting the influence that treatment with just a small volume of the patient-derived PCF can have on normal cell biology in the context of facilitating the hallmarks of cancer. This multi-faceted research has not only provided insights into past literature and highlighted the strengths and weaknesses of previous PC studies (Chapter 3), but has also curated novel multi-omic biomarker panels that show promise in the early detection of high-risk patients (Chapters 4 and 5), as well as examining, for the first time, the potential role that PCF may play in the development of PC in PCL patients (Chapter 6). Finally, the overlaps between these facets, and how they can come together to guide future pancreatic research and clinical management strategies, are discussed.

7.2 Using the past to guide the future

This thesis has shown the advantages to be gained by examining past literature before developing new studies. Here, it was demonstrated through a thorough examination of over 40 years of PC literature, that important lessons can be learned from past studies,

and that promising aspects, such as high performing biomarkers, can be discovered. Indeed, the many weaknesses in study design and data handling of previous PC studies have been highlighted in Chapter 3. Reporting of key information with regards to patient demographic data, cancer subtype of interest, patient cancer stage, treatments received by patients, reference standards used for PC diagnosis, and the recruitment process used for the study, is severely lacking across the literature. This absence of information can not only allow for the reporting of results that may be skewed due to some patient demographic-based bias that is unknown to the reader, thus inflating the potential clinical utility of the biomarker, but also hampers the dissemination and integration of PC research, as it becomes unclear if studies are comparable[485]. In a field where every small advancement has the capacity to make a huge difference, such as PC, these errors and omissions in data reporting can be costly, and as such, future studies should aim to be more transparent and forthcoming with such important information.

The systematic review and subsequent meta-analysis performed in Chapter 3 has also highlighted the path that future diagnostic biomarker studies in PC should take, while also providing vital information with regards to biomarker performance for all biomarker studies, not just PC. Indeed, here it was shown for the first time that panels of multiple biomarkers are superior to single biomarkers alone. As such, future work in biomarker discovery and selection are advised to curate a biomarker panel, rather than focus on identifying a single 'one-size-fits-all' biomarker. Furthermore, the data shown here have emphasized the importance of appropriate control cohorts that provide a more clinically relevant representation of those that could be tested in a clinical setting. As a result, it is important to recognise that while some studies may report excellent biomarker performances where the control cohort consists of just healthy or benign individuals, these biomarkers may not perform well in a clinical setting. As such, examining biomarkers in the first instance within control cohorts of clinical relevance is key to robust biomarker identification, and should be the aim of all biomarker studies, not just those in PC.

As a guiding hand to future work, Chapter 3 highlighted biomarkers and biomarker panels across the literature that were either repeatedly examined, or that performed within the 90th percentile of biomarkers that were assessed as part of this

work. These biomarkers were largely proteomic and transcriptomic in nature, and demonstrated great potential for PC diagnosis. Building on this, Chapters 4 and 5 delved into the proteome and transcriptome of PCL patient PCF and serum, interrogating the performance of those 'promising' biomarkers found within these biofluids. Here, their ability to distinguish low- and high-risk PCLs was assessed. Given the step-wise development of PC from pancreatitis to PCL to invasive carcinoma discussed in Chapter 1, and the many factors which are likewise either upregulated or downregulated step-wise in this process, such as *KRAS* and *TP53* mutations, it is reasonable, therefore, to assume that diagnostic biomarkers may be up or downregulated in high-risk PCL patients compared to low-risk[15-17]. Unfortunately, only a small fraction of these promising biomarkers could be found in the PCF and serum of the patients examined in this study, and despite some significant differences on an individual level in the PCF, their utility as a biomarker panel appear limited. As such, this thesis delved further into the proteome and transcriptome of these biofluids, examining independent of past research, as is usually the case in biomarker research, the factors that were differentially expressed between low- and high-risk patients. Given the results of Chapter 3, instead of examining these factors individually, they were integrated to form multi-biomarker panels. Alone, the performance of the proteomic and transcriptomic panels were limited, however, when scaled and integrated, the multi-biomarker panels, both within the PCF and serum, had improved performances over either omic panel alone. As such, it was demonstrated that multi-omic panels may perform superiorly to single-omic panels, a theory which could not be successfully examined in the systematic review given the lack of previous research into such panels. Branching out further, into what could be considered an intuitive avenue, these PCF-based and serum-based multi-omic panels were interrogated further to create a novel cross-biofluid multi-omic panel. While in the clinical setting, various biological factors can be utilised to form a diagnosis, including simultaneous measurement of both blood-based and PCF-based biomarkers individually, the use of cross-biofluid panels in mainstream medicine is unconventional. However, the novel cross-biofluid panel introduced here has demonstrated high performance in the risk stratification of PCLs, and managed the inclusion of a VHL outlier better than any other panel shown here, further compounding its utility in this setting.

However, it is important to note that the quality of the results obtained depend on the methods of data handling that are employed.

7.3 Data handling - a double-edged sword

The processes of data clean-up, handling, analysis, and interpretation, are integral to any study, but in the context of large omics datasets, the methodologies involved are critical. While transcriptomic data, in terms of counts per million or otherwise, are straightforward in terms of their identification and handling, proteomic data leave much room for error. Indeed, from the initial identification of proteins from MS/Ms data, to the removal of various contaminants, to filtering and imputation steps, and finally to analysis and visual representation, there are many points at which individual researchers may differ in their approach, and as such, there is large scope for differences and error. Even excluding the initial variables, which are known sources of error in MS protocols (such as sample handling prior to MS, the choice of MS instrument, the length of separation time and the choice of templates subsequently used to identify proteins), the clean-up of LFQ Intensity data is a lengthy process, with many decision points. It is important to highlight this, as even just small changes here with regards to data filtering can change not only the number of proteins one ends up with at the end of the process, but also the individual proteins that are highlighted as significant after the differential expression analysis[486]. Importantly, studies have shown that there is a lack of reproducibility across MS outputs, even when utilising the same sample set across independent laboratories, and that feature selection is a major part of this[486-488].

Feature selection, in the context of biomarker discovery, is a critical step and as such there have been many methods developed for this purpose[486]. The inclusion of only the strongest biomarkers, while excluding those that worsen the panel, has been shown to have a huge impact on the performance of the panel (Chapter 5). Indeed, when creating the cross-biofluid panel, feature reduction was carried out using the CombiROC software. Ideally, the full PCF and serum panels would be examined together (a total of 24 features) in order to choose the best combination across both biofluids. However, the CombiROC software does not allow for the examination of panels greater than 14

biomarkers as the number of permutations becomes too much for the software to analyse, making this type of approach impossible with this software[293]. Indeed, the integration of layers of data is generally better conducted gradually, where the most important biomarkers within each layer are identified first (i.e. feature selection). This is due to the orientation of the data matrix. When there are more biomarkers than there are samples, and the data frame is therefore wider than it is long, also known as high-dimensional low sample size (HDLSS), issues with statistical power and false discovery applications become more prominent, and the accuracy of these variables can become positively biased[489]. This is when overfitting of the data becomes an issue, where the accuracy of the test in a training cohort is high, but this accuracy is substantially lower in a separate test cohort. As such, the approach taken in this study was to first reduce the dimensionality of the individual data layers by selecting the top candidate biomarkers, and then integrate these layers of data. It is important to note, however, that feature selection in this context could be improved by the inclusion of longitudinal patient data. Indeed, knowledge of whether the patients included in this study have progressed to cancer, would give much needed definition to the dataset, providing confidence in the stratification of these patients and therefore, in the performance of this panel.

Another important step was the use of LOOCV, which is better suited to smaller sample sizes than similar approaches, such as *k*-fold cross validation, and is known to give more reliable accuracy estimates for such datasets[490]. In fact, the choice of biomarker assessment was shown to be vital in this study (Chapter 5), as analyses such as PCA and unsupervised hierarchical clustering, while useful when interrogating the data, have limited utility in providing biomarker efficacy metrics. Chapter 3 highlighted the vast number of PC studies failing to report qualitative metrics of biomarker performance such as AUC value, sensitivity or specificity, but instead reported just a p-value. In Chapters 4 and 5, it was demonstrated that while significantly differentially expressed factors, as indicated by a $p<0.05$, may be found to be increased or decreased in one cohort compared to another, their utility as a biomarker cannot be determined by this metric alone. While p-values provide information on whether the expression of a factor is statistically significantly associated with an outcome, it is greatly affected by

the strength of the association and the sample size, and as such, while it is a good starting point in this setting, its utility in biomarker performance assessment is limited[491]. As such, in the context of assessing biomarker performance, ROC curves producing AUC values as well as sensitivity and specificity values are the recommended method of evaluation, and have been demonstrated to be superior metrics in this thesis (Chapters 3, 4 and 5)[492].

When handling patient data, which is naturally extremely heterogeneous, outliers are a normal yet important aspect of data interpretation. In this study, a clinical outlier was encountered that was classified as high-risk based on the presence of a VHL genetic mutation, despite their PCL receiving a low-risk classification when this feature was excluded. In Chapters 4 and 5, the performance of both multi-omic biomarker panels could be improved by the reclassification of the patient from high- to low-risk. However, this may not be an appropriate course of action in this context. Indeed, mutations in oncogenes and tumour suppressor genes such as *KRAS* and *TP53* are known to occur at extremely high rates in the early stages of PC development[15-17]. Furthermore, early mutations in these genes have been shown to be associated with PC development. As such, patients in this setting with genetic mutations are not uncommon, and given the known association of these mutations with PC development, as with the VHL mutation, it is more clinically appropriate to classify this patient, as per current guidelines, as high-risk. Importantly, the cross-biofluid multi-omic panel curated in Chapter 5, is the only panel shown here whose performance was not improved by the reclassification of this VHL patient. In fact, the performance of the panel worsened with the reclassification of the VHL outlier from high-risk to low-risk. For this reason, the cross-biofluid panel produced in Chapter 5 shows a huge amount of promise for the effective risk stratification of PCLs, and should be further interrogated in an independent cohort of clinically relevant patients in order to assess its full potential.

7.4 Omics and Hallmarks of Cancer – the bigger picture

The mechanisms involved in the malignant transformation of PCLs to PC are not well understood. While many studies have shown a step-wise increase or decrease in certain

factors, such as REG1A or LGALS3, throughout the transformation process, this vague and disconnected information does little to elucidate the pathways and key mechanisms involved[396, 399, 404]. In this thesis, it was demonstrated for the first time, that the fluid within PCLs is biologically active and may have a role to play in this process (Chapter 6). By providing a preliminary overview of how exposure to low concentrations of PCF can influence the development of several hallmarks of cancer processes in normal cells, this thesis put forward the theory that the presence of this fluid may be detrimental to patients with PCLs, and should therefore be aspirated at the earliest opportunity. Indeed, substantial alterations in normal cell biology and function could be seen at low PCF concentrations, and even low experimental replicates, highlighting the potent activity of this biofluid. Furthermore, this research has shown that both low- and high-risk PCF have the ability to alter normal cell biology, with both fluids in some cases, causing mass cell death. In this way, it was demonstrated that while certain clinical factors may be helpful in determining low- and high-risk patients, the effects of the PCF within these cysts is not well discriminated using this risk classification. Similarly, studies examining the effects of ovarian ascites on ovarian cancer cell biology also found that no clinical factors could be correlated to the distinct profiles of differential influences demonstrated[453]. In this way, the discriminating factor in the differential effects of PCF on normal cell biology remains to be determined. Importantly, a 2022 study by Fraunhoffer *et al.* utilised multi-omic data integration and modelling to interrogate the mechanisms involved in PDAC[493]. Using a Master Regulator-Gradient model developed in-house, transcriptional networks were integrated with epigenomic states and metabolic pathways to reveal that tumour epithelial cell phenotype and the corresponding immunological component strongly correlated with PDAC prognosis[493]. Furthermore, oxidative phosphorylation independent metabolism, centred on the Warburg effect and glutaminolysis, was shown to be associated with an unfavourable phenotype, while the presence of enzymes implicated in complex lipid biosynthesis indicated a favourable phenotype[493]. In such multi-omic modelling, mechanisms and processes can be teased out and associated with clinical outcomes such as prognosis. Here, the generation of multi-omic data from PCF samples (Chapters 4 and 5), coupled with matched data from PCF treated cell lines (Chapter 6) could be the key to elucidating the mechanisms involved in PCL progression. Furthermore, the acquisition of

longitudinal patient data with regards to progression/outcomes would add further, invaluable dimension to this dataset. Importantly, several of the differentially expressed factors within the PCF, such as miR-216b[384, 385], miR-216a[388], LCN2[393, 397, 398], REG1A[403] and S100A8[409, 410], have been previously linked to certain hallmarks of cancer, as discussed in Chapter 6. S100A8, in particular, was previously shown to stimulate motility, proliferation and pro-inflammatory cytokine secretion in pancreatic cell lines, demonstrating a potential role in PC progression[409, 410]. However, at this point, while it can be said that individual proteins and miRNA identified via multi-omic profiling of PCF may contribute to the process of PCL to PC transformation, there remains a plethora of factors within the PCF from other omic levels that are yet to be profiled. Furthermore, the intricacies of the interconnected interactions of these factors, and the pathways involved, makes elucidating the responsible factor or pathway difficult. While the modelling of this work would be complex and likely require a personalised approach to its development, it provides a novel avenue by which many of these mechanisms could be teased out simultaneously. Furthermore, it could provide a wealth of knowledge for biomarker studies. Indeed, Chapters 4 and 5 have shown that integrating levels of omic data can result in more robust biomarkers than a single level alone. By expanding this dataset to comprise the metabolome and the genome at a minimum, the options for biomarker discovery become expansive. Importantly, these opportunities would extend not just to risk stratification, but when coupled with longitudinal patient clinical data, also to diagnostic and even prognostic biomarkers. Furthermore, in depth knowledge of the factors contained within the PCF would provide a better understanding of the mechanisms involved when normal cells are exposed to PCF, and enable the thorough and informed interrogation of these pathways when these datasets are combined.

7.5 Limitations

This thesis has provided an expanse of results from both high- and low-throughput methodologies, each with their own strengths and weaknesses. The primary caveat to the majority of this work is a lack of validation studies. In terms of the omics work performed in Chapters 4 and 5, while robust analyses demonstrated promising results,

attempts to validate these using other laboratory techniques were unsuccessful. Furthermore, in most instances the total number of patients in both the low- and high-risk cohorts was less than 30 individuals. As power calculations have demonstrated that a minimum of 15 patients per category would be required for sufficient power (Chapter 3), this is arguably too small a cohort for such studies. More importantly, however, is the lack of an independent validation cohort to reassess the performance of the panels in a blinded and unbiased setting. As such, while these data can be classified as promising, they remain to be thoroughly examined and should therefore be interpreted with caution. Furthermore, as discussed at length in Chapter 1, the clinical guidelines used to stratify patients with PCLs based on PC risk differ globally[154-157]. In this thesis, the patient cohorts examined were stratified using the 2018 European evidence-based guidelines. However, as there is a lack of consensus among clinicians as to the most appropriate stratification cut-offs and parameters, these classifications may not be accurate[157]. As such, longitudinal patient data with information regarding malignant transformation is needed. Moreover, in respect to the low-risk patients, no clear cut-off exists for the interpretation of whether this was the correct classification. Indeed, regardless of whether five or ten years have passed without this patient progressing to PC, there is no definitive time-specific cut-off to determine whether this was the correct classification. Indeed, the 2015 American Gastroenterological Association guidelines ceases surveillance on PCLs after 5 years without a change in cyst size or characteristics, while both the 2017 International Association of Pancreatology Fukuoka guidelines, and the 2018 European evidence-based guidelines suggest continued surveillance every 1 or 2 years, depending on cyst size (Table 1.2)[154, 156, 157]. As a result, a definitive classification of low-risk appears difficult to ascertain, and will be a major limitation of any risk stratification studies in PCLs.

The various points at which the handling of omics data can introduce bias are discussed above in section 7.3, however, the methodologies related to the acquisition of one omics dataset in particular has defined limitations. Indeed, the serum proteomic data discussed in Chapter 5 consisted of just 145 proteins after data clean-up, compared to the 465 identified in the PCF in Chapter 4. Despite immunodepletion steps which removed the 12 most abundant proteins from the serum samples, it appears the depth

of proteome coverage within the serum may have been poor. As such, these data may not represent the full proteome of these serum samples, and should be re-examined with more stringent depletion steps, possibly including fractionation in order to capture lower molecular weight proteins and less abundant proteins more efficiently[494]. This re-evaluation will be essential for the modelling work suggested above (section 7.4).

Further to the limitations of the CombiROC software used in Chapter 5 as mentioned above in section 7.3, a major final step in this process provides vast avenues for different outcomes. Indeed, through the panel reduction process for both the serum and PCF panels, several reduced panels met the cut-offs for sensitivity and specificity within the software. As such, those brought forward in this work may not have been the best, or only, set of biomarkers that could be chosen for the creation of the cross-biofluid panel. It is important, therefore, to identify a software or model with sufficient processing power to handle the inclusions of the entire 24-biomarker set across both panels, in order to perform the necessary permutations for feature selection. In this way, a more robust selection can be carried out with less room for variation and bias.

Lastly, the treatment of normal cell lines with PCF in order to assess its impact on normal cell biology was conducted under serum starved conditions. Discussed at length in Chapter 6, is the possible impact this starvation may have had on the results obtained in this study. It is unclear whether the serum starvation of the cells has interfered with, clouded, or exacerbated, the results demonstrated here. While serum deprivation is used in many studies for cell synchronisation, adverse effects of this culture technique can be numerous, such as enriching the cancer stem cell population, altering protein expression, increasing sensitivity to radiotherapy, increasing apoptosis and decreasing proliferation[495-498]. Indeed, a 2019 study demonstrated that serum-reduced media had negative impacts on lung epithelial cell viability, morphology and protein expression[499]. As such, while it is clear that PCF has altered the biology of normal cells in many aspects, the exact influences, and directions of these influences, cannot yet be known due to the unknown interference of serum-starvation, and as such interpretation of these data is limited[500]. Further research is therefore required, under different culture conditions, to fully elucidate the effects of PCF on normal cell biology.

7.6 Future directions

The results reported in this thesis present promising avenues for future research to build, and expand on. Firstly, the re-evaluation of patient serum via fractionation and subsequent proteomics is key to acquire a more in depth read of the serum proteome[494]. Furthermore, expansion of the omics dataset curated in this thesis, to include at least genomic and metabolic data for both the serum and PCF, will add further dimension and enable the generation of an even more robust multi-omic biomarker panel. These omics data will then need to be evaluated in a sufficiently powered, blinded, independent validation cohort in order to appropriately assess biomarker efficacy. Longitudinal data is needed to examine the progression of these patients, and give more certainty to the risk classifications used in this study, while novel and study-specific modelling should be considered in order to utilise this dataset to its full potential. Furthermore, the expansion of the control cohort to include benign PCF, and therefore represent a more clinically relevant control cohort, should be considered.

Secondly, further examination of the effects of PCF on normal cell biology is needed, specifically targeting the hallmarks of cancer and the role PCF plays in the acquisition of these traits. Furthermore, the effects of PCF on other key cell types within the pancreatic microenvironment, such as stromal and immune cells, as well as PC cells, will be important to understand the bigger picture of this progression. Indeed, examining the upregulation of MMPs and TIMPs in the presence of PCF may elucidate whether PCF is involved in the desmoplastic reaction seen in PC[64, 66, 72]. Moreover, the differential responses of normal cells to PCF exposure, specifically with reference to the mass cell death observed in some cases, will need to be explored further. A more expansive clinical dataset combined with the additional omics suggested could provide further insight into these differential influences. In fact, the presence of a gradient should be considered, if a distinct categorical factor cannot be found to correlate with these effects.

Lastly, the interrogation of the role of PCF on cell biology will need to be carried out under a number of different conditions. It has been discussed that serum starvation may have interfered with the results demonstrated in this thesis (section 7.5). As such, future work should aim to repeat and expand on those experiments conducted, in both

normal growth conditions, and under a range of other stress conditions observed in the pancreatic microenvironment, such as hypoxia and desmoplasia[20, 501]. The use of 3D pancreatic organoid models should also be considered in order to capture the unique and multi-faceted pancreatic microenvironment[502]. Furthermore, the use of higher concentrations of PCF should be a key aspect of future work. In this thesis, PCF treatment experiments were conducted at a low concentration of PCF. However, clinically, this may not be relevant. Future work should explore the potentially drastic changes to be seen in the cellular responses to higher concentrations of PCF, given that *in vivo* the epithelial cells that make up the cyst lining, and the cellular component within the PCF, would be in contact with neat PCF[139].

7.7 Conclusion

This thesis has not only highlighted important flaws and knowledge gaps in pancreatic literature, but has also highlighted ways in which future research can, and should, be designed in order to improve reporting and dissemination of results. This research has shown extensive evidence of potential biomarkers and biomarker panels with great promise for the risk stratification of patients with PCLs, a facet of pancreatic research which requires urgent attention in order to improve PC patient survival. This thesis reports on a novel, cross-biofluid multi-omic biomarker panel, with great promise for the accurate risk stratification of these patients. However, most importantly, this research has shown, for the first time, that PCF is biologically active, and can influence the proliferation, cell death, viability, metabolic profile, invasive capacity and phenotypic and functional marker expression of normal and intermediary pancreatic cell lines. These results will have a meaningful impact on the surveillance and management strategies of pancreatic patients, and while further research is required to fully elucidate these impacts, the results reported here cannot be taken lightly. PCF contains many factors that have the potential to influence those cells which come into contact with it, and as such, may play a role in the progression of patients with PCLs to PC. Taken together, these findings provide not only a novel and promising biomarker panel for the early detection of patients at a high-risk of PC, but also suggests a new course of action in the management of PCLs in favour of cyst aspiration. While this action could be

considered unnecessarily invasive for patients classified as low-risk, this thesis has provided evidence that all PCF can alter cell biology, and should therefore be considered potentially carcinogenic. Although much research is still required to expand on the knowledge generated in this thesis, this work provides important contributions to the positive progression of pancreatic research, and patient management.

References

1. Siegel, R.L., K.D. Miller, N.S. Wagle, and A. Jemal, *Cancer statistics, 2023.* CA: a cancer journal for clinicians, 2023. **73**(1): p. 17-48.
2. Klein, A.P., *Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors.* Nature reviews Gastroenterology & hepatology, 2021. **18**(7): p. 493-502.
3. Midha, S., S. Chawla, and P.K. Garg, *Modifiable and non-modifiable risk factors for pancreatic cancer: A review.* Cancer letters, 2016. **381**(1): p. 269-277.
4. Mishra, N.K., S. Southekal, and C. Guda, *Survival analysis of multi-omics data identifies potential prognostic markers of pancreatic ductal adenocarcinoma.* Frontiers in genetics, 2019. **10**: p. 624.
5. Rawla, P., T. Sunkara, and V. Gaduputi, *Epidemiology of pancreatic cancer: global trends, etiology and risk factors.* World journal of oncology, 2019. **10**(1): p. 10.
6. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019.* CA: a cancer journal for clinicians, 2019. **69**(1): p. 7-34.
7. Ullman, N.A., P.R. Burchard, R.F. Dunne, and D.C. Linehan, *Immunologic strategies in pancreatic cancer: Making cold tumors hot.* Journal of Clinical Oncology, 2022. **40**(24): p. 2789.
8. Kleeff, J., M. Korc, M. Apte, C. La Vecchia, C.D. Johnson, A.V. Biankin, R.E. Neale, M. Tempero, D.A. Tuveson, and R.H. Hruban, *Pancreatic cancer.* Nature reviews Disease primers, 2016. **2**(1): p. 1-22.
9. Mizrahi, J.D., R. Surana, J.W. Valle, and R.T. Shroff, *Pancreatic cancer.* The Lancet, 2020. **395**(10242): p. 2008-2020.
10. Hruban, R.H., K. Takaori, D.S. Klimstra, N.V. Adsay, J. Albores-Saavedra, A.V. Biankin, S.A. Biankin, C. Compton, N. Fukushima, and T. Furukawa, *An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms.* The American journal of surgical pathology, 2004. **28**(8): p. 977-987.
11. Hruban, R.H., N.V. Adsay, J. Albores–Saavedra, C. Compton, E.S. Garrett, S.N. Goodman, S.E. Kern, D.S. Klimstra, G. Klöppel, and D.S. Longnecker, *Pancreatic intraepithelial neoplasia: a new nomenclature and classification system for*

pancreatic duct lesions. The American journal of surgical pathology, 2001. **25**(5): p. 579-586.

12. Matsuda, Y., T. Furukawa, S. Yachida, M. Nishimura, A. Seki, K. Nonaka, J. Aida, K. Takubo, T. Ishiwata, and W. Kimura, *The prevalence and clinicopathological characteristics of high-grade pancreatic intraepithelial neoplasia: autopsy study evaluating the entire pancreatic parenchyma.* Pancreas, 2017. **46**(5): p. 658-664.
13. Distler, M., D. Aust, J. Weitz, C. Pilarsky, and R. Grützmann, *Precursor lesions for sporadic pancreatic cancer: PanIN, IPMN, and MCN.* BioMed research international, 2014. **2014**.
14. Peters, M.L.B., A. Eckel, P.P. Mueller, A.C. Tramontano, D.T. Weaver, A. Lietz, C. Hur, C.Y. Kong, and P.V. Pandharipande, *Progression to pancreatic ductal adenocarcinoma from pancreatic intraepithelial neoplasia: Results of a simulation model.* Pancreatology, 2018. **18**(8): p. 928-934.
15. Raphael, B.J., R.H. Hruban, A.J. Aguirre, R.A. Moffitt, J.J. Yeh, C. Stewart, A.G. Robertson, A.D. Cherniack, M. Gupta, and G. Getz, *Integrated genomic characterization of pancreatic ductal adenocarcinoma.* Cancer cell, 2017. **32**(2): p. 185-203. e13.
16. Pihlak, R., J.M. Weaver, J.W. Valle, and M.G. McNamara, *Advances in molecular profiling and categorisation of pancreatic adenocarcinoma and the implications for therapy.* Cancers, 2018. **10**(1): p. 17.
17. Biankin, A.V., N. Waddell, K.S. Kassahn, M.-C. Gingras, L.B. Muthuswamy, A.L. Johns, D.K. Miller, P.J. Wilson, A.-M. Patch, and J. Wu, *Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes.* Nature, 2012. **491**(7424): p. 399-405.
18. Ho, W.J., E.M. Jaffee, and L. Zheng, *The tumour microenvironment in pancreatic cancer—clinical challenges and opportunities.* Nature reviews Clinical oncology, 2020. **17**(9): p. 527-540.
19. Maitra, A. and R.H. Hruban, *Pancreatic cancer.* Annu. Rev. Pathol. Mech. Dis., 2008. **3**: p. 157-188.
20. Whatcott, C.J., C.H. Diep, P. Jiang, A. Watanabe, J. LoBello, C. Sima, G. Hostetter, H.M. Shepard, D.D. Von Hoff, and H. Han, *Desmoplasia in primary tumors and*

metastatic lesions of pancreatic cancer. Clinical cancer research, 2015. **21**(15): p. 3561-3568.

21. Mekapogu, A., S. Pothula, R. Pirola, J. Wilson, and M. Apte, *Pancreatic Stellate Cells in Health and Disease.* The Pancreapedia: Exocrine Pancreas Knowledge Base, 2020.
22. Murakami, T., Y. Hiroshima, R. Matsuyama, Y. Homma, R.M. Hoffman, and I. Endo, *Role of the tumor microenvironment in pancreatic cancer.* Annals of gastroenterological surgery, 2019. **3**(2): p. 130-137.
23. Apte, M., Z. Xu, S. Pothula, D. Goldstein, R. Pirola, and J. Wilson, *Pancreatic cancer: The microenvironment needs attention too!* Pancreatology, 2015. **15**(4): p. S32-S38.
24. Channon, L.M., V.M. Tyma, Z. Xu, D.W. Greening, J.S. Wilson, C.J. Perera, and M.V. Apte, *Small extracellular vesicles (exosomes) and their cargo in pancreatic cancer: key roles in the hallmarks of cancer.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2022: p. 188728.
25. Pothula, S.P., Z. Xu, D. Goldstein, R.C. Pirola, J.S. Wilson, and M.V. Apte, *Key role of pancreatic stellate cells in pancreatic cancer.* Cancer letters, 2016. **381**(1): p. 194-200.
26. Apte, M.V., J.S. Wilson, A. Lugea, and S.J. Pandol, *A starring role for stellate cells in the pancreatic cancer microenvironment.* Gastroenterology, 2013. **144**(6): p. 1210-1219.
27. Hosein, A.N., R.A. Brekken, and A. Maitra, *Pancreatic cancer stroma: an update on therapeutic targeting strategies.* Nature reviews Gastroenterology & hepatology, 2020. **17**(8): p. 487-505.
28. Shimizu, K., M. Kobayashi, J. Tahara, and K. Shiratori, *Cytokines and peroxisome proliferator-activated receptor γ ligand regulate phagocytosis by pancreatic stellate cells.* Gastroenterology, 2005. **128**(7): p. 2105-2118.
29. Binkley, C.E., L. Zhang, J.K. Greenson, T.J. Giordano, R. Kuick, D. Misek, S. Hanash, C.D. Logsdon, and D.M. Simeone, *The molecular basis of pancreatic fibrosis: common stromal gene expression in chronic pancreatitis and pancreatic adenocarcinoma.* Pancreas, 2004. **29**(4): p. 254-263.

30. Kakizaki, Y., N. Makino, T. Tozawa, T. Honda, A. Matsuda, Y. Ikeda, M. Ito, Y. Saito, W. Kimura, and Y. Ueno, *Stromal fibrosis and expression of matricellular proteins correlate with histological grade of intraductal papillary mucinous neoplasm of the pancreas.* Pancreas, 2016. **45**(8): p. 1145.
31. Apte, M., R.C. Pirola, and J.S. Wilson, *Pancreatic stellate cell: physiologic role, role in fibrosis and cancer.* Current opinion in gastroenterology, 2015. **31**(5): p. 416-423.
32. Xu, Z., A. Vonlaufen, P.A. Phillips, E. Fiala-Beer, X. Zhang, L. Yang, A.V. Biankin, D. Goldstein, R.C. Pirola, and J.S. Wilson, *Role of pancreatic stellate cells in pancreatic cancer metastasis.* The American journal of pathology, 2010. **177**(5): p. 2585-2596.
33. Öhlund, D., A. Handly-Santana, G. Biffi, E. Elyada, A.S. Almeida, M. Ponz-Sarvise, V. Corbo, T.E. Oni, S.A. Hearn, and E.J. Lee, *Distinct populations of inflammatory fibroblasts and myofibroblasts in pancreatic cancer.* Journal of Experimental Medicine, 2017. **214**(3): p. 579-596.
34. Hanahan, D. and R.A. Weinberg, *The hallmarks of cancer.* cell, 2000. **100**(1): p. 57-70.
35. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation.* Cell, 2011. **144**(5): p. 646-674.
36. Hanahan, D., *Hallmarks of cancer: new dimensions.* Cancer discovery, 2022. **12**(1): p. 31-46.
37. Buscail, L., B. Bournet, and P. Cordelier, *Role of oncogenic KRAS in the diagnosis, prognosis and treatment of pancreatic cancer.* Nature Reviews Gastroenterology & Hepatology, 2020. **17**(3): p. 153-168.
38. Schlieman, M., B. Fahy, R. Ramsamooj, L. Beckett, and R. Bold, *Incidence, mechanism and prognostic value of activated AKT in pancreas cancer.* British journal of cancer, 2003. **89**(11): p. 2110-2115.
39. Schlitter, A.M., A. Segler, K. Steiger, C.W. Michalski, C. Jäger, B. Konukiewitz, N. Pfarr, V. Endris, M. Bettstetter, and B. Kong, *Molecular, morphological and survival analysis of 177 resected pancreatic ductal adenocarcinomas (PDACs): Identification of prognostic subtypes.* Scientific reports, 2017. **7**(1): p. 1-12.

40. Ou, A., X. Zhao, and Z. Lu, *The potential roles of p53 signaling reactivation in pancreatic cancer therapy.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2022. **1877**(1): p. 188662.

41. Hwang, R.F., E.M. Gordon, W.F. Anderson, and D. Parekh, *Gene therapy for primary and metastatic pancreatic cancer with intraperitoneal retroviral vector bearing the wild-type p53 gene.* Surgery, 1998. **124**(2): p. 143-151.

42. Takei, Y., S. Okamoto, K. Kawamura, Y. Jiang, T. Morinaga, M. Shingyoji, I. Sekine, S. Kubo, Y. Tada, and K. Tatsumi, *Expression of p53 synergistically augments caspases-mediated apoptosis induced by replication-competent adenoviruses in pancreatic carcinoma cells.* Cancer Gene Therapy, 2015. **22**(9): p. 445-453.

43. Velasco, R.M., A.G. García, P.J. Sánchez, I.M. Sellart, and V.J. Sanchez-Arevalo Lobo, *Tumour microenvironment and heterotypic interactions in pancreatic cancer.* Journal of Physiology and Biochemistry, 2023. **79**(1): p. 179-192.

44. Chang, D.Z., Y. Ma, B. Ji, H. Wang, D. Deng, Y. Liu, J.L. Abbruzzese, Y.-j. Liu, C.D. Logsdon, and P. Hwu, *Mast Cells in Tumor Microenvironment Promotes the In Vivo Growth of Pancreatic Ductal Adenocarcinoma.* Clinical Cancer Research, 2011. **17**(22): p. 7015-7023.

45. Ma, Y., R.F. Hwang, C.D. Logsdon, and S.E. Ullrich, *Dynamic Mast Cell–Stromal Cell Interactions Promote Growth of Pancreatic Cancer.* Cancer research, 2013. **73**(13): p. 3927-3937.

46. Apte, M., R. Pirola, and J. Wilson, *New insights into alcoholic pancreatitis and pancreatic cancer.* Journal of gastroenterology and hepatology, 2009. **24**: p. S51-S56.

47. Gore, A.J., S.L. Deitz, L.R. Palam, K.E. Craven, and M. Korc, *Pancreatic cancer–associated retinoblastoma 1 dysfunction enables TGF-β to promote proliferation.* The Journal of clinical investigation, 2014. **124**(1): p. 338-352.

48. Cave, D.D., M. Di Guida, V. Costa, M. Sevillano, L. Ferrante, C. Heeschen, M. Corona, A. Cucciardi, and E. Lonardo, *TGF-β1 secreted by pancreatic stellate cells promotes stemness and tumourigenicity in pancreatic cancer cells through L1CAM downregulation.* Oncogene, 2020. **39**(21): p. 4271-4285.

49. Principe, D.R., B. DeCant, E. Mascariñas, E.A. Wayne, A.M. Diaz, N. Akagi, R. Hwang, B. Pasche, D.W. Dawson, and D. Fang, *TGFβ Signaling in the Pancreatic*

Tumor Microenvironment Promotes Fibrosis and Immune Evasion to Facilitate Tumorigenesis. Cancer research, 2016. **76**(9): p. 2525-2539.

50. Drabsch, Y. and P. Ten Dijke, *TGF-β signalling and its role in cancer progression and metastasis.* Cancer and Metastasis Reviews, 2012. **31**: p. 553-568.
51. Cicenas, J., K. Kvederaviciute, I. Meskinyte, E. Meskinyte-Kausiliene, A. Skeberdyte, and J. Cicenas Jr, *KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 mutations in pancreatic cancer.* Cancers, 2017. **9**(5): p. 42.
52. Fiorini, C., M. Cordani, C. Padroni, G. Blandino, S. Di Agostino, and M. Donadelli, *Mutant p53 stimulates chemoresistance of pancreatic adenocarcinoma cells to gemcitabine.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2015. **1853**(1): p. 89-100.
53. Bryant, K.L., J.D. Mancias, A.C. Kimmelman, and C.J. Der, *KRAS: feeding pancreatic cancer proliferation.* Trends in biochemical sciences, 2014. **39**(2): p. 91-100.
54. Greten, F.R., C.K. Weber, T.F. Greten, G. Schneider, M. Wagner, G. Adler, and R.M. Schmid, *Stat3 and NF-κB activation prevents apoptosis in pancreatic carcinogenesis.* Gastroenterology, 2002. **123**(6): p. 2052-2063.
55. Preis, M. and M. Korc, *Signaling pathways in pancreatic cancer.* Critical Reviews in Eukaryotic Gene Expression, 2011. **21**(2): p. 115-129.
56. Skinner, H.G., R.E. Gangnon, K. Litzelman, R.A. Johnson, S.T. Chari, G.M. Petersen, and L.A. Boardman, *Telomere length and pancreatic cancer: a case–control study.* Cancer epidemiology, biomarkers & prevention, 2012. **21**(11): p. 2095-2100.
57. Van Heek, N.T., A.K. Meeker, S.E. Kern, C.J. Yeo, K.D. Lillemoe, J.L. Cameron, G.J.A. Offerhaus, J.L. Hicks, R.E. Wilentz, and M.G. Goggins, *Telomere shortening is nearly universal in pancreatic intraepithelial neoplasia.* The American journal of pathology, 2002. **161**(5): p. 1541-1547.
58. Campa, D., M. Matarazzi, W. Greenhalf, M. Bijlsma, K.U. Saum, C. Pasquali, H. van Laarhoven, A. Szentesi, F. Federici, and P. Vodicka, *Genetic determinants of telomere length and risk of pancreatic cancer: a PANDoRA study.* International journal of cancer, 2019. **144**(6): p. 1275-1283.

59. Luu, H.N., J.Y. Huang, R. Wang, J. Adams-Haduch, A. Jin, W.-P. Koh, and J.-M. Yuan, *Association between leukocyte telomere length and the risk of pancreatic cancer: Findings from a prospective study.* PLoS One, 2019. **14**(8): p. e0221697.

60. Li, S., H.-X. Xu, C.-T. Wu, W.-Q. Wang, W. Jin, H.-L. Gao, H. Li, S.-R. Zhang, J.-Z. Xu, and Z.-H. Qi, *Angiogenesis in pancreatic cancer: current research status and clinical implications.* Angiogenesis, 2019. **22**: p. 15-36.

61. Annese, T., R. Tamma, S. Ruggieri, and D. Ribatti, *Angiogenesis in pancreatic cancer: Pre-clinical and clinical studies.* Cancers, 2019. **11**(3): p. 381.

62. Masamune, A., K. Kikuta, T. Watanabe, K. Satoh, M. Hirota, and T. Shimosegawa, *Hypoxia stimulates pancreatic stellate cells to induce fibrosis and angiogenesis in pancreatic cancer.* American Journal of Physiology-Gastrointestinal and Liver Physiology, 2008. **295**(4): p. G709-G717.

63. Saiyin, H., C.M. Ardito-Abraham, Y. Wu, Y. Wei, Y. Fang, X. Han, J. Li, P. Zhou, Q. Yi, and A. Maitra, *Identification of novel vascular projections with cellular trafficking abilities on the microvasculature of pancreatic ductal adenocarcinoma.* The Journal of pathology, 2015. **236**(2): p. 142-154.

64. Bramhall, S.R., J.P. Neoptolemos, G.W. Stamp, and N.R. Lemoine, *Imbalance of expression of matrix metalloproteinases (MMPs) and tissue inhibitors of the matrix metalloproteinases (TIMPs) in human pancreatic carcinoma.* The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland, 1997. **182**(3): p. 347-355.

65. Curran, S. and G.I. Murray, *Matrix metalloproteinases in tumour invasion and metastasis.* The Journal of pathology, 1999. **189**(3): p. 300-308.

66. Jones, L.E., M.J. Humphreys, F. Campbell, J.P. Neoptolemos, and M.T. Boyd, *Comprehensive analysis of matrix metalloproteinase and tissue inhibitor expression in pancreatic cancer: increased expression of matrix metalloproteinase-7 predicts poor survival.* Clinical Cancer Research, 2004. **10**(8): p. 2832-2845.

67. Matsuyama, Y., S. Takao, and T. Aikou, *Comparison of matrix metalloproteinase expression between primary tumors with or without liver metastasis in pancreatic and colorectal carcinomas.* Journal of surgical oncology, 2002. **80**(2): p. 105-110.

68. Okada, Y., G. Eibl, S. Guha, J.P. Duffy, H.A. Reber, and O.J. Hines, *Nerve growth factor stimulates MMP-2 expression and activity and increases invasion by human pancreatic cancer cells.* Clinical & experimental metastasis, 2004. **21**: p. 285-292.

69. Hirai, I., W. Kimura, K. Ozawa, S. Kudo, K. Suto, H. Kuzu, and A. Fuse, *Perineural invasion in pancreatic cancer.* Pancreas, 2002. **24**(1): p. 15-25.

70. Schneiderhan, W., F. Diaz, M. Fundel, S. Zhou, M. Siech, C. Hasel, P. Möller, J.r.E. Gschwend, T. Seufferlein, and T. Gress, *Pancreatic stellate cells are an important source of MMP-2 in human pancreatic cancer and accelerate tumor progression in a murine xenograft model and CAM assay.* Journal of cell science, 2007. **120**(3): p. 512-519.

71. Phillips, P., J. McCarroll, S. Park, M. Wu, R. Pirola, M. Korsten, J. Wilson, and M. Apte, *Rat pancreatic stellate cells secrete matrix metalloproteinases: implications for extracellular matrix turnover.* Gut, 2003. **52**(2): p. 275-282.

72. Gress, T., F. Müller-Pillasch, M. Lerch, H. Friess, M. Büchler, and G. Adler, *Expression and in-situ localization of genes coding for extracellular matrix proteins and extracellular matrix degrading proteases in pancreatic cancer.* International journal of cancer, 1995. **62**(4): p. 407-413.

73. Ellenrieder, V., B. Alber, U. Lacher, S.F. Hendler, A. Menke, W. Boeck, M. Wagner, M. Wilda, H. Friess, and M. Büchler, *Role of MT-MMPs and MMP-2 in pancreatic cancer progression.* International journal of cancer, 2000. **85**(1): p. 14-20.

74. Knapinska, A.M., C.-A. Estrada, and G.B. Fields, *The roles of matrix metalloproteinases in pancreatic cancer.* Progress in molecular biology and translational science, 2017. **148**: p. 339-354.

75. Du, L., S. Yamamoto, B.L. Burnette, D. Huang, K. Gao, N. Jamshidi, and M.D. Kuo, *Transcriptome profiling reveals novel gene expression signatures and regulating transcription factors of TGF β-induced epithelial-to-mesenchymal transition.* Cancer medicine, 2016. **5**(8): p. 1962-1972.

76. Nakajima, S., R. Doi, E. Toyoda, S. Tsuji, M. Wada, M. Koizumi, S.S. Tulachan, D. Ito, K. Kami, and T. Mori, *N-cadherin expression and epithelial-mesenchymal transition in pancreatic carcinoma.* Clinical cancer research, 2004. **10**(12): p. 4125-4133.

77. Hotz, B., M. Arndt, S. Dullat, S. Bhargava, H.-J. Buhr, and H.G. Hotz, *Epithelial to mesenchymal transition: expression of the regulators snail, slug, and twist in pancreatic cancer.* Clinical cancer research, 2007. **13**(16): p. 4769-4776.

78. Lee, H.J., D. Do You, D.W. Choi, Y.S. Choi, S.J. Kim, Y.S. Won, and H.J. Moon, *Significance of CD133 as a cancer stem cell markers focusing on the tumorigenicity of pancreatic cancer cell lines.* Journal of the Korean Surgical Society, 2011. **81**(4): p. 263-270.

79. Li, C., D.G. Heidt, P. Dalerba, C.F. Burant, L. Zhang, V. Adsay, M. Wicha, M.F. Clarke, and D.M. Simeone, *Identification of pancreatic cancer stem cells.* Cancer research, 2007. **67**(3): p. 1030-1037.

80. Hermann, P.C., S.L. Huber, T. Herrler, A. Aicher, J.W. Ellwart, M. Guba, C.J. Bruns, and C. Heeschen, *Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer.* Cell stem cell, 2007. **1**(3): p. 313-323.

81. Li, X., H. Zhao, J. Gu, and L. Zheng, *Prognostic value of cancer stem cell marker CD133 expression in pancreatic ductal adenocarcinoma (PDAC): a systematic review and meta-analysis.* International journal of clinical and experimental pathology, 2015. **8**(10): p. 12084.

82. DeBerardinis, R.J. and N.S. Chandel, *We need to talk about the Warburg effect.* Nature metabolism, 2020. **2**(2): p. 127-129.

83. Ying, H., A.C. Kimmelman, C.A. Lyssiotis, S. Hua, G.C. Chu, E. Fletcher-Sananikone, J.W. Locasale, J. Son, H. Zhang, and J.L. Coloff, *Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism.* Cell, 2012. **149**(3): p. 656-670.

84. Suzuki, T., M. Otsuka, T. Seimiya, T. Iwata, T. Kishikawa, and K. Koike, *The biological role of metabolic reprogramming in pancreatic cancer.* MedComm, 2020. **1**(3): p. 302-310.

85. Amrutkar, M. and I.P. Gladhaug, *Stellate cells aid growth-permissive metabolic reprogramming and promote gemcitabine chemoresistance in pancreatic cancer.* Cancers, 2021. **13**(4): p. 601.

86. Fu, Y., S. Liu, S. Yin, W. Niu, W. Xiong, M. Tan, G. Li, and M. Zhou, *The reverse Warburg effect is likely to be an Achilles' heel of cancer that can be exploited for cancer therapy.* Oncotarget, 2017. **8**(34): p. 57813.

87. Mace, T.A., Z. Ameen, A. Collins, S. Wojcik, M. Mair, G.S. Young, J.R. Fuchs, T.D. Eubank, W.L. Frankel, and T. Bekaii-Saab, *Pancreatic cancer-associated stellate cells promote differentiation of myeloid-derived suppressor cells in a STAT3-dependent manner.* Cancer research, 2013. **73**(10): p. 3007-3018.

88. Corzo, C.A., T. Condamine, L. Lu, M.J. Cotter, J.-I. Youn, P. Cheng, H.-I. Cho, E. Celis, D.G. Quiceno, and T. Padhya, *HIF-1α regulates function and differentiation of myeloid-derived suppressor cells in the tumor microenvironment.* Journal of Experimental Medicine, 2010. **207**(11): p. 2439-2453.

89. Gabrilovich, D.I., S. Ostrand-Rosenberg, and V. Bronte, *Coordinated regulation of myeloid cells by tumours.* Nature Reviews Immunology, 2012. **12**(4): p. 253-268.

90. Clark, C.E., S.R. Hingorani, R. Mick, C. Combs, D.A. Tuveson, and R.H. Vonderheide, *Dynamics of the immune reaction to pancreatic cancer from inception to invasion.* Cancer research, 2007. **67**(19): p. 9518-9527.

91. Mace, T.A., M. Bloomston, and G.B. Lesinski, *Pancreatic cancer-associated stellate cells: A viable target for reducing immunosuppression in the tumor microenvironment.* Oncoimmunology, 2013. **2**(7): p. e24891.

92. Ene–Obong, A., A.J. Clear, J. Watt, J. Wang, R. Fatah, J.C. Riches, J.F. Marshall, J. Chin–Aleong, C. Chelala, and J.G. Gribben, *Activated pancreatic stellate cells sequester CD8+ T cells to reduce their infiltration of the juxtatumoral compartment of pancreatic ductal adenocarcinoma.* Gastroenterology, 2013. **145**(5): p. 1121-1132.

93. Nagaraju, G.P., R.R. Malla, R. Basha, and I.G. Motofei. *Contemporary clinical trials in pancreatic cancer immunotherapy targeting PD-1 and PD-L1*. in *Seminars in Cancer Biology*. 2022. Elsevier.

94. Feng, M., G. Xiong, Z. Cao, G. Yang, S. Zheng, X. Song, L. You, L. Zheng, T. Zhang, and Y. Zhao, *PD-1/PD-L1 and immunotherapy for pancreatic cancer.* Cancer letters, 2017. **407**: p. 57-65.

95. Gao, H.-L., L. Liu, Z.-H. Qi, H.-X. Xu, W.-Q. Wang, C.-T. Wu, S.-R. Zhang, J.-Z. Xu, Q.-X. Ni, and X.-J. Yu, *The clinicopathological and prognostic significance of PD-*

L1 expression in pancreatic cancer: a meta-analysis. Hepatobiliary & Pancreatic Diseases International, 2018. **17**(2): p. 95-100.

96. Sanmamed, M.F. and L. Chen, *Inducible expression of B7-H1 (PD-L1) and its selective role in tumor site immune modulation.* Cancer journal (Sudbury, Mass.), 2014. **20**(4): p. 256.
97. Tsukamoto, M., K. Imai, T. Ishimoto, Y. Komohara, Y.i. Yamashita, S. Nakagawa, N. Umezaki, T. Yamao, Y. Kitano, and T. Miyata, *PD-L1 expression enhancement by infiltrating macrophage-derived tumor necrosis factor-α leads to poor pancreatic cancer prognosis.* Cancer science, 2019. **110**(1): p. 310-320.
98. Alexandrov, L.B., S. Nik-Zainal, D.C. Wedge, S.A. Aparicio, S. Behjati, A.V. Biankin, G.R. Bignell, N. Bolli, A. Borg, and A.-L. Børresen-Dale, *Signatures of mutational processes in human cancer.* Nature, 2013. **500**(7463): p. 415-421.
99. Alexandrov, L.B., P.H. Jones, D.C. Wedge, J.E. Sale, P.J. Campbell, S. Nik-Zainal, and M.R. Stratton, *Clock-like mutational processes in human somatic cells.* Nature Genetics, 2015. **47**(12): p. 1402-1407.
100. Allen, M.J., A. Zhang, P. Bavi, J.C. Kim, G.H. Jang, D. Kelly, S. Perera, R.E. Denroche, F. Notta, J.M. Wilson, A. Dodd, S. Ramotar, S. Hutchinson, S.E. Fischer, R.C. Grant, S. Gallinger, J.J. Knox, and G.M. O'Kane, *Molecular characterisation of pancreatic ductal adenocarcinoma with NTRK fusions and review of the literature.* Journal of Clinical Pathology, 2023. **76**(3): p. 158-165.
101. Tjomsland, V., A. Spångeus, J. Välilä, P. Sandström, K. Borch, H. Druid, S. Falkmer, U. Falkmer, D. Messmer, and M. Larsson, *Interleukin 1α sustains the expression of inflammatory factors in human pancreatic cancer microenvironment by targeting cancer-associated fibroblasts.* Neoplasia, 2011. **13**(8): p. 664-IN3.
102. Bhatia, S., P. Wang, A. Toh, and E.W. Thompson, *New insights into the role of phenotypic plasticity and EMT in driving cancer progression.* Frontiers in molecular biosciences, 2020. **7**: p. 71.
103. Bernadotte, A., V.M. Mikhelson, and I.M. Spivak, *Markers of cellular senescence. Telomere shortening as a marker of cellular senescence.* Aging (Albany NY), 2016. **8**(1): p. 3.

104. Lau, L. and G. David, *Pro-and anti-tumorigenic functions of the senescence-associated secretory phenotype.* Expert opinion on therapeutic targets, 2019. **23**(12): p. 1041-1051.

105. Basisty, N., A. Kale, O.H. Jeon, C. Kuehnemann, T. Payne, C. Rao, A. Holtz, S. Shah, V. Sharma, and L. Ferrucci, *A proteomic atlas of senescence-associated secretomes for aging biomarker development.* PLoS biology, 2020. **18**(1): p. e3000599.

106. Cortesi, M., M. Zanoni, F. Pirini, M.M. Tumedei, S. Ravaioli, I.G. Rapposelli, G.L. Frassineti, and S. Bravaccini, *Pancreatic cancer and cellular senescence: Tumor microenvironment under the spotlight.* International journal of molecular sciences, 2022. **23**(1): p. 254.

107. Krtolica, A., S. Parrinello, S. Lockett, P.-Y. Desprez, and J. Campisi, *Senescent fibroblasts promote epithelial cell growth and tumorigenesis: a link between cancer and aging.* Proceedings of the National Academy of Sciences, 2001. **98**(21): p. 12072-12077.

108. Abbadie, C., O. Pluquet, and A. Pourtier, *Epithelial cell senescence: an adaptive response to pre-carcinogenic stresses?* Cellular and Molecular Life Sciences, 2017. **74**(24): p. 4471-4509.

109. Erusalimsky, J. and D. Kurz, *Endothelial cell senescence.* The Vascular Endothelium II, 2006: p. 213-248.

110. Caldwell, M., G. DeNicola, C. Martins, M. Jacobetz, A. Maitra, R. Hruban, and D. Tuveson, *Cellular features of senescence during the evolution of human and murine ductal pancreatic cancer.* Oncogene, 2012. **31**(12): p. 1599-1608.

111. Frey, N., S. Venturelli, L. Zender, and M. Bitzer, *Cellular senescence in gastrointestinal diseases: from pathogenesis to therapeutics.* Nature Reviews Gastroenterology & Hepatology, 2018. **15**(2): p. 81-95.

112. Guerra, C., M. Collado, C. Navas, A.J. Schuhmacher, I. Hernández-Porras, M. Cañamero, M. Rodriguez-Justo, M. Serrano, and M. Barbacid, *Pancreatitis-induced inflammation contributes to pancreatic cancer by inhibiting oncogene-induced senescence.* Cancer cell, 2011. **19**(6): p. 728-739.

113. Jacks, B.E., C.U. Ekpemiro, A.A. Adeosun, U.O. Ogbonna, F.T. Ogundiran, F. Babalola, N.P. Onyechi, O.O. Ajayi, M.G. Boms, and A.N. Nwanguma, *Molecular*

Markers of Pancreatic Cancer: A 10-Year Retrospective Review of Molecular Advances. Cureus, 2022. **14**(9).

114. Aghdassi, A., M. Sendler, A. Guenther, J. Mayerle, C.-O. Behn, C.-D. Heidecke, H. Friess, M. Büchler, M. Evert, and M.M. Lerch, *Recruitment of histone deacetylases HDAC1 and HDAC2 by the transcriptional repressor ZEB1 downregulates E-cadherin expression in pancreatic cancer.* Gut, 2012. **61**(3): p. 439-448.
115. McDonald, O.G., X. Li, T. Saunders, R. Tryggvadottir, S.J. Mentch, M.O. Warmoes, A.E. Word, A. Carrer, T.H. Salz, and S. Natsume, *Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis.* Nature genetics, 2017. **49**(3): p. 367-376.
116. Vakoc, C.R. and D.A. Tuveson, *Untangling the genetics from the epigenetics in pancreatic cancer metastasis.* Nature genetics, 2017. **49**(3): p. 323-324.
117. Khoshchehreh, R., M. Totonchi, J. Carlos Ramirez, R. Torres, H. Baharvand, A. Aicher, M. Ebrahimi, and C. Heeschen, *Epigenetic reprogramming of primary pancreatic cancer cells counteracts their in vivo tumourigenicity.* Oncogene, 2019. **38**(34): p. 6226-6239.
118. Neureiter, D., T. Jäger, M. Ocker, and T. Kiesslich, *Epigenetics and pancreatic cancer: pathophysiology and novel treatment aspects.* World Journal of Gastroenterology: WJG, 2014. **20**(24): p. 7830.
119. Lythgoe, M.P., B.H. Mullish, A.E. Frampton, and J. Krell, *Polymorphic microbes: a new emerging hallmark of cancer.* Trends in Microbiology, 2022.
120. Nejman, D., I. Livyatan, G. Fuks, N. Gavert, Y. Zwang, L.T. Geller, A. Rotter-Maskowitz, R. Weiser, G. Mallel, and E. Gigi, *The human tumor microbiome is composed of tumor type–specific intracellular bacteria.* Science, 2020. **368**(6494): p. 973-980.
121. Park, J., D. Han, M. Do, J. Woo, J.I. Wang, Y. Han, W. Kwon, S.W. Kim, J.Y. Jang, and Y. Kim, *Proteome characterization of human pancreatic cyst fluid from intraductal papillary mucinous neoplasm by liquid chromatography/tandem mass spectrometry.* Rapid Communications in Mass Spectrometry, 2017. **31**(20): p. 1761-1772.

122. Soufi, M., M.T. Yip-Schneider, R.A. Carr, A.M. Roch, H.H. Wu, and C.M. Schmidt, *Multifocal High-Grade Pancreatic Precursor Lesions: A Case Series and Management Recommendations.* Journal of pancreatic cancer, 2019. **5**(1): p. 8-11.

123. Matthaei, H., R.D. Schulick, R.H. Hruban, and A. Maitra, *Cystic precursors to invasive pancreatic cancer.* Nature reviews Gastroenterology & hepatology, 2011. **8**(3): p. 141.

124. McGuigan, A., P. Kelly, R.C. Turkington, C. Jones, H.G. Coleman, and R.S. McCain, *Pancreatic cancer: A review of clinical diagnosis, epidemiology, treatment and outcomes.* World journal of gastroenterology, 2018. **24**(43): p. 4846.

125. Stark, A., T.R. Donahue, H.A. Reber, and O.J. Hines, *Pancreatic cyst disease: a review.* Jama, 2016. **315**(17): p. 1882-1893.

126. Tersmette, A.C., G.M. Petersen, G.J.A. Offerhaus, F.C. Falatko, K.A. Brune, M. Goggins, E. Rozenblum, R.E. Wilentz, C.J. Yeo, and J.L. Cameron, *Increased risk of incident pancreatic cancer among first-degree relatives of patients with familial pancreatic cancer.* Clinical Cancer Research, 2001. **7**(3): p. 738-744.

127. Greer, J.B. and D.C. Whitcomb, *Role of BRCA1 and BRCA2 mutations in pancreatic cancer.* Gut, 2007. **56**(5): p. 601-605.

128. Tirosh, A., S.M. Sadowski, W.M. Linehan, S.K. Libutti, D. Patel, N. Nilubol, and E. Kebebew, *Association of VHL genotype with pancreatic neuroendocrine tumor phenotype in patients with von Hippel–Lindau disease.* JAMA oncology, 2018. **4**(1): p. 124-126.

129. Brugge, W.R., *Diagnosis and management of cystic lesions of the pancreas.* Journal of Gastrointestinal Oncology, 2015. **6**(4): p. 375.

130. Kromrey, M.-L., R. Bülow, J. Hübner, C. Paperlein, M.M. Lerch, T. Ittermann, H. Völzke, J. Mayerle, and J.-P. Kühn, *Prospective study on the incidence, prevalence and 5-year pancreatic-related mortality of pancreatic cysts in a population-based study.* Gut, 2018. **67**(1): p. 138-145.

131. Torisu, Y., K. Takakura, Y. Kinoshita, Y. Tomita, M. Nakano, and M. Saruta, *Pancreatic cancer screening in patients with presumed branch-duct intraductal papillary mucinous neoplasms.* World journal of clinical oncology, 2019. **10**(2): p. 67.

132. Wong, M.C., J.Y. Jiang, M. Liang, Y. Fang, M.S. Yeung, and J.J. Sung, *Global temporal patterns of pancreatic cancer and association with socioeconomic development.* Scientific reports, 2017. **7**(1): p. 1-9.

133. Maggi, G., G. Guarneri, G. Gasparini, A. Fogliati, S. Partelli, M. Falconi, and S. Crippa, *Pancreatic cystic neoplasms: What is the most cost-effective follow-up strategy?* Endoscopic Ultrasound, 2018. **7**(5): p. 319.

134. de Jong, K., C.Y. Nio, J.J. Hermans, M.G. Dijkgraaf, D.J. Gouma, C.H. Van Eijck, E. Van Heel, G. Klass, P. Fockens, and M.J. Bruno, *High prevalence of pancreatic cysts detected by screening magnetic resonance imaging examinations.* Clinical Gastroenterology and Hepatology, 2010. **8**(9): p. 806-811.

135. Chang, Y.R., J.K. Park, J.-Y. Jang, W. Kwon, J.H. Yoon, and S.-W. Kim, *Incidental pancreatic cystic neoplasms in an asymptomatic healthy population of 21,745 individuals: large-scale, single-center cohort study.* Medicine, 2016. **95**(51): p. e5535.

136. Sethi, V., B. Giri, A. Saluja, and V. Dudeja, *Insights into the pathogenesis of pancreatic cystic neoplasms.* Digestive diseases and sciences, 2017. **62**: p. 1778-1786.

137. Naik-Mathuria, B.J., E.H. Rosenfeld, A. Gosain, R. Burd, R.A. Falcone Jr, R. Thakkar, B. Gaines, D. Mooney, M. Escobar, and M. Jafri, *Proposed clinical pathway for nonoperative management of high-grade pediatric pancreatic injuries based on a multicenter analysis: A pediatric trauma society collaborative.* Journal of Trauma and Acute Care Surgery, 2017. **83**(4): p. 589-596.

138. Parekh, P.J., D. Howerton, and D.A. Johnson, *Not your everyday case of acute pancreatitis: a rare complication of a common diagnosis.* ACG Case Reports Journal, 2013. **1**(1): p. 40.

139. Tacelli, M., C. Celsa, B. Magro, M. Barchiesi, L. Barresi, G. Capurso, P.G. Arcidiacono, C. Camma, and S.F. Crino, *Diagnostic performance of endoscopic ultrasound through-the-needle microforceps biopsy of pancreatic cystic lesions: Systematic review with meta-analysis.* Digestive Endoscopy, 2020. **32**(7): p. 1018-1030.

140. Law, N.-M. and M.L. Freeman, *Emergency complications of acute and chronic pancreatitis.* Gastroenterology Clinics, 2003. **32**(4): p. 1169-1194.

141. Bhasin, D.K., S.S. Rana, V. Sharma, C. Rao, V. Gupta, R. Gupta, M. Kang, and K. Singh, *Non-surgical management of pancreatic pseudocysts associated with arterial pseudoaneurysm.* Pancreatology, 2013. **13**(3): p. 250-253.

142. Matsubara, S., M. Tada, M. Akahane, H. Yagioka, H. Kogure, T. Sasaki, T. Arizumi, O. Togawa, Y. Nakai, and N. Sasahira, *Incidental pancreatic cysts found by magnetic resonance imaging and their relationship with pancreatic cancer.* Pancreas, 2012. **41**(8): p. 1241-1246.

143. Lee, L.S., P.A. Banks, A.M. Bellizzi, N.I. Sainani, V. Kadiyala, S. Suleiman, D.L. Conwell, and J.A. Paulo, *Inflammatory protein profiling of pancreatic cyst fluid using EUS-FNA in tandem with cytokine microarray differentiates between branch duct IPMN and inflammatory cysts.* Journal of immunological methods, 2012. **382**(1-2): p. 142-149.

144. Paziewska, A., M. Polkowski, K. Goryca, J. Karczmarski, A. Wiechowska-Kozlowska, M. Dabrowska, M. Mikula, and J. Ostrowski, *Mutational mosaics of cell-free DNA from pancreatic cyst fluids.* Digestive Diseases and Sciences, 2020. **65**: p. 2294-2301.

145. Matthaei, H., D. Wylie, M.B. Lloyd, M. Dal Molin, J. Kemppainen, S.C. Mayo, C.L. Wolfgang, R.D. Schulick, L. Langfield, and B.F. Andruss, *miRNA biomarkers in cyst fluid augment the diagnosis and management of pancreatic cysts.* Clinical cancer research, 2012. **18**(17): p. 4713-4724.

146. Shi, J., Z. Yi, L. Jin, L. Zhao, A. Raskind, L. Yeomans, Z.C. Nwosu, D.M. Simeone, C.A. Lyssiotis, and K.A. Stringer, *Cyst fluid metabolites distinguish malignant from benign pancreatic cysts.* Neoplasia, 2021. **23**(11): p. 1078-1088.

147. Li, S., G.M. Fuhler, N. Bn, T. Jose, M.J. Bruno, M.P. Peppelenbosch, and S.R. Konstantinov, *Pancreatic cyst fluid harbors a unique microbiome.* Microbiome, 2017. **5**(1): p. 1-13.

148. Valsangkar, N.P., V. Morales-Oyarvide, S.P. Thayer, C.R. Ferrone, J.A. Wargo, A.L. Warshaw, and C. Fernández-del Castillo, *851 resected cystic tumors of the pancreas: a 33-year experience at the Massachusetts General Hospital.* Surgery, 2012. **152**(3): p. S4-S12.

149. Fritz, S., M. Klauss, F. Bergmann, O. Strobel, L. Schneider, J. Werner, T. Hackert, and M.W. Büchler, *Pancreatic main-duct involvement in branch-duct IPMNs: an underestimated risk.* Annals of surgery, 2014. **260**(5): p. 848-856.

150. Sakhdari, A., P.A. Moghaddam, C.Y. Ok, O. Walter, K. Tomaszewicz, M.-L. Caporelli, X. Meng, J. LaFemina, G. Whalen, and E. Belkin, *Somatic molecular analysis augments cytologic evaluation of pancreatic cyst fluids as a diagnostic tool.* Oncotarget, 2019. **10**(40): p. 4026.

151. Tanaka, M., C. Fernández-del Castillo, V. Adsay, S. Chari, M. Falconi, J.-Y. Jang, W. Kimura, P. Levy, M.B. Pitman, and C.M. Schmidt, *International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas.* Pancreatology, 2012. **12**(3): p. 183-197.

152. Maker, A.V., S. Carrara, N.B. Jamieson, M. Pelaez-Luna, A.M. Lennon, M. Dal Molin, A. Scarpa, L. Frulloni, and W.R. Brugge, *Cyst fluid biomarkers for intraductal papillary mucinous neoplasms of the pancreas: a critical review from the international expert meeting on pancreatic branch-duct-intraductal papillary mucinous neoplasms.* Journal of the American College of Surgeons, 2015. **220**(2): p. 243-253.

153. Lee, A., V. Kadiyala, and L.S. Lee, *Evaluation of AGA and Fukuoka Guidelines for EUS and surgical resection of incidental pancreatic cysts.* Endoscopy International Open, 2017. **5**(02): p. E116-E122.

154. Vege, S.S., B. Ziring, R. Jain, P. Moayyedi, M.A. Adams, S.D. Dorn, S.L. Dudley-Brown, S.L. Flamm, Z.F. Gellad, and C.B. Gruss, *American gastroenterological association institute guideline on the diagnosis and management of asymptomatic neoplastic pancreatic cysts.* Gastroenterology, 2015. **148**(4): p. 819-822.

155. Hanania, A.N., L.E. Bantis, Z. Feng, H. Wang, E.P. Tamm, M.H. Katz, A. Maitra, and E.J. Koay, *Quantitative imaging to evaluate malignant potential of IPMNs.* Oncotarget, 2016. **7**(52): p. 85776.

156. Tanaka, M., C. Fernández-del Castillo, T. Kamisawa, J.Y. Jang, P. Levy, T. Ohtsuka, R. Salvia, Y. Shimizu, M. Tada, and C.L. Wolfgang, *Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas.* Pancreatology, 2017. **17**(5): p. 738-753.

157. European Study Group on Cystic Tumours of the Pancreas, *European evidence-based guidelines on pancreatic cystic neoplasms.* Gut, 2018. **67**(5): p. 789-804.

158. Miller, J.R., J.E. Meyer, J.A. Waters, M. Al-Haddad, J. DeWitt, S. Sherman, K.D. Lillemoe, and C.M. Schmidt, *Outcome of the pancreatic remnant following segmental pancreatectomy for non-invasive intraductal papillary mucinous neoplasm.* Hpb, 2011. **13**(11): p. 759-766.

159. Tan, M.C., O. Basturk, A.R. Brannon, U. Bhanot, S.N. Scott, N. Bouvier, J. LaFemina, W.R. Jarnagin, M.F. Berger, and D. Klimstra, *GNAS and KRAS mutations define separate progression pathways in intraductal papillary mucinous neoplasm-associated carcinoma.* Journal of the American College of Surgeons, 2015. **220**(5): p. 845-854.

160. van Huijgevoort, N.C., M. Del Chiaro, C.L. Wolfgang, J.E. van Hooft, and M.G. Besselink, *Diagnosis and management of pancreatic cystic neoplasms: current evidence and guidelines.* Nature reviews Gastroenterology & hepatology, 2019. **16**(11): p. 676-689.

161. Maker, A.V., L.S. Lee, C.P. Raut, T.E. Clancy, and R.S. Swanson, *Cytology from pancreatic cysts has marginal utility in surgical decision-making.* Annals of surgical oncology, 2008. **15**(11): p. 3187-3192.

162. Le Borgne, J., L. de Calan, C. Partensky, and F.S. Association, *Cystadenomas and cystadenocarcinomas of the pancreas: a multiinstitutional retrospective study of 398 cases.* Annals of surgery, 1999. **230**(2): p. 152.

163. McCarty, T.R., R. Garg, and T. Rustagi, *Pancreatic cyst fluid glucose in differentiating mucinous from nonmucinous pancreatic cysts: a systematic review and meta-analysis.* Gastrointestinal Endoscopy, 2021. **94**(4): p. 698-712. e6.

164. Thornton, G., M. McPhail, S. Nayagam, M. Hewitt, P. Vlavianos, and K. Monahan, *Endoscopic ultrasound guided fine needle aspiration for the diagnosis of pancreatic cystic neoplasms: a meta-analysis.* Pancreatology, 2013. **13**(1): p. 48-57.

165. McCarty, T.R., S. Paleti, and T. Rustagi, *Molecular analysis of EUS-acquired pancreatic cyst fluid for KRAS and GNAS mutations for diagnosis of intraductal*

papillary mucinous neoplasia and mucinous cystic lesions: a systematic review and meta-analysis. Gastrointestinal Endoscopy, 2021. **93**(5): p. 1019-1033. e5.

166. Kane, L.E., G.S. Mellotte, K.C. Conlon, B.M. Ryan, and S.G. Maher, *Multi-Omic Biomarkers as Potential Tools for the Characterisation of Pancreatic Cystic Lesions and Cancer: Innovative Patient Data Integration.* Cancers, 2021. **13**(4): p. 769.
167. Luo, G., K. Jin, S. Deng, H. Cheng, Z. Fan, Y. Gong, Y. Qian, Q. Huang, Q. Ni, and C. Liu, *Roles of CA19-9 in pancreatic cancer: Biomarker, predictor and promoter.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2020: p. 188409.
168. Hasan, S., R. Jacob, U. Manne, and R. Paluri, *Advances in pancreatic cancer biomarkers.* Oncology reviews, 2019. **13**(1): p. 410.
169. Biomarkers Definitions Working Group, A.J. Atkinson Jr, W.A. Colburn, V.G. DeGruttola, D.L. DeMets, G.J. Downing, D.F. Hoth, J.A. Oates, C.C. Peck, and R.T. Schooley, *Biomarkers and surrogate endpoints: preferred definitions and conceptual framework.* Clinical pharmacology & therapeutics, 2001. **69**(3): p. 89-95.
170. Chen, X.-H., S. Huang, and D. Kerr, *Biomarkers in clinical medicine.* IARC Sci Publ, 2011. **163**: p. 303-322.
171. Fiocchi, C., *Integrating omics: the future of IBD?* Digestive Diseases, 2014. **32**(Suppl. 1): p. 96-102.
172. Hasin, Y., M. Seldin, and A. Lusis, *Multi-omics approaches to disease.* Genome biology, 2017. **18**(1): p. 83.
173. Vucic, E.A., K.L. Thu, K. Robison, L.A. Rybaczyk, R. Chari, C.E. Alvarez, and W.L. Lam, *Translating cancer 'omics' to improved outcomes.* Genome research, 2012. **22**(2): p. 188-195.
174. Manzoni, C., D.A. Kia, J. Vandrovcova, J. Hardy, N.W. Wood, P.A. Lewis, and R. Ferrari, *Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences.* Briefings in bioinformatics, 2018. **19**(2): p. 286-302.
175. Novelli, G., C. Ciccacci, P. Borgiani, M.P. Amati, and E. Abadie, *Genetic tests and genomic biomarkers: regulation, qualification and validation.* Clinical cases in mineral and bone metabolism, 2008. **5**(2): p. 149.

176. Matsui, S., *Genomic biomarkers for personalized medicine: development and validation in clinical studies.* Computational and mathematical methods in medicine, 2013. **2013**.

177. Singhi, A.D., M.N. Nikiforova, K.E. Fasanella, K.M. McGrath, R.K. Pai, N.P. Ohori, T.L. Bartholow, R.E. Brand, J.S. Chennat, and X. Lu, *Preoperative GNAS and KRAS testing in the diagnosis of pancreatic mucinous cysts.* Clinical Cancer Research, 2014. **20**(16): p. 4381-4389.

178. Kadayifci, A., M. Atar, J.L. Wang, D.G. Forcione, B.W. Casey, M.B. Pitman, and W.R. Brugge, *Value of adding GNAS testing to pancreatic cyst fluid KRAS and carcinoembryonic antigen analysis for the diagnosis of intraductal papillary mucinous neoplasms.* Digestive Endoscopy, 2017. **29**(1): p. 111-117.

179. Singhi, A.D., K. McGrath, R.E. Brand, A. Khalid, H.J. Zeh, J.S. Chennat, K.E. Fasanella, G.I. Papachristou, A. Slivka, and D.L. Bartlett, *Preoperative next-generation sequencing of pancreatic cyst fluid is highly accurate in cyst classification and detection of advanced neoplasia.* Gut, 2018. **67**(12): p. 2131-2141.

180. Ding, J., Y. Li, Y. Zhang, B. Fan, Q. Li, J. Zhang, and J. Zhang, *Identification of key lncRNAs in the tumorigenesis of intraductal pancreatic mucinous neoplasm by coexpression network analysis.* Cancer Medicine, 2020. **9**(11): p. 3840-3851.

181. Park, W.G., M. Wu, R. Bowen, M. Zheng, W.L. Fitch, R.K. Pai, D. Wodziak, B.C. Visser, G.A. Poultsides, and J.A. Norton, *Metabolomic-derived novel cyst fluid biomarkers for pancreatic cysts: glucose and kynurenine.* Gastrointestinal Endoscopy, 2013. **78**(2): p. 295-302.

182. Carr, R.A., M.T. Yip-Schneider, R.E. Simpson, S. Dolejs, J.G. Schneider, H. Wu, E.P. Ceppa, W. Park, and C.M. Schmidt, *Pancreatic cyst fluid glucose: rapid, inexpensive, and accurate diagnosis of mucinous pancreatic cysts.* Surgery, 2018. **163**(3): p. 600-605.

183. Fahrmann, J.F., L.E. Bantis, M. Capello, G. Scelo, J.B. Dennison, N. Patel, E. Murage, J. Vykoukal, D.L. Kundnani, and L. Foretova, *A plasma-derived protein-metabolite multiplexed panel for early-stage pancreatic cancer.* JNCI: Journal of the National Cancer Institute, 2019. **111**(4): p. 372-379.

184. Eissa, M.A., L. Lerner, E. Abdelfatah, N. Shankar, J.K. Canner, N.M. Hasan, V. Yaghoobi, B. Huang, Z. Kerner, and F. Takaesu, *Promoter methylation of ADAMTS1 and BNC1 as potential biomarkers for early detection of pancreatic cancer in blood.* Clinical epigenetics, 2019. **11**(1): p. 59.

185. Hata, T., M. Dal Molin, S.-M. Hong, K. Tamura, M. Suenaga, J. Yu, H. Sedogawa, M.J. Weiss, C.L. Wolfgang, and A.M. Lennon, *Predicting the grade of dysplasia of pancreatic cystic neoplasms using cyst fluid DNA methylation markers.* Clinical Cancer Research, 2017. **23**(14): p. 3935-3944.

186. Majumder, S., W.R. Taylor, T.C. Yab, C.K. Berger, B.A. Dukek, X. Cao, P.H. Foote, C.W. Wu, D.W. Mahoney, and H.R. Aslanian, *Novel Methylated DNA Markers Discriminate Advanced Neoplasia in Pancreatic Cysts: Marker Discovery, Tissue Validation, and Cyst Fluid Testing.* American Journal of Gastroenterology, 2019. **114**(9): p. 1539-1549.

187. Li, A., J. Yu, H. Kim, C.L. Wolfgang, M.I. Canto, R.H. Hruban, and M. Goggins, *MicroRNA array analysis finds elevated serum miR-1290 accurately distinguishes patients with low-stage pancreatic cancer from healthy and disease controls.* Clinical cancer research, 2013. **19**(13): p. 3600-3610.

188. Karasek, P., N. Gablo, J. Hlavsa, I. Kiss, P. Vychytilova-Faltejskova, M. Hermanova, Z. Kala, O. Slaby, and V. Prochazka, *Pre-operative plasma miR-21-5p is a sensitive biomarker and independent prognostic factor in patients with pancreatic ductal adenocarcinoma undergoing surgical resection.* Cancer Genomics-Proteomics, 2018. **15**(4): p. 321-327.

189. Wei, J., L. Yang, Y.-n. Wu, and J. Xu, *Serum miR-1290 and miR-1246 as Potential Diagnostic Biomarkers of Human Pancreatic Cancer.* Journal of Cancer, 2020. **11**(6): p. 1325-1333.

190. Xie, Z., X. Yin, B. Gong, W. Nie, B. Wu, X. Zhang, J. Huang, P. Zhang, Z. Zhou, and Z. Li, *Salivary microRNAs show potential as a noninvasive biomarker for detecting resectable pancreatic cancer.* Cancer prevention research, 2015. **8**(2): p. 165-173.

191. Brand, R.E., B.M. Nolen, H.J. Zeh, P.J. Allen, M.A. Eloubeidi, M. Goldberg, E. Elton, J.P. Arnoletti, J.D. Christein, and S.M. Vickers, *Serum biomarker panels for the detection of pancreatic cancer.* Clinical cancer research, 2011. **17**(4): p. 805-816.

192. Chan, A., I. Prassas, A. Dimitromanolakis, R.E. Brand, S. Serra, E.P. Diamandis, and I.M. Blasutig, *Validation of biomarkers that complement CA19. 9 in detecting early pancreatic cancer.* Clinical Cancer Research, 2014. **20**(22): p. 5787-5795.

193. Azadeh, A., R. Felix, T. Krause, M. Bernhardt, P. Jo, A. König, K. Mathias, L. Andreas, M. Ghadimi, and G. Jochen, *CA19-9 for detecting recurrence of pancreatic cancer.* Scientific Reports (Nature Publisher Group), 2020. **10**(1): p. 1-10.

194. Carr, R.A., M.T. Yip-Schneider, S. Dolejs, B.A. Hancock, H. Wu, M. Radovich, and C.M. Schmidt, *Pancreatic cyst fluid vascular endothelial growth factor A and carcinoembryonic antigen: a highly accurate test for the diagnosis of serous cystic neoplasm.* Journal of the American College of Surgeons, 2017. **225**(1): p. 93-100.

195. Sinha, J., Z. Cao, J. Dai, H. Tang, K. Partyka, G. Hostetter, D.M. Simeone, Z. Feng, P.J. Allen, and R.E. Brand, *A gastric glycoform of MUC5AC is a biomarker of mucinous cysts of the pancreas.* PloS one, 2016. **11**(12): p. e0167070.

196. Cao, Z., K. Maupin, B. Curnutte, B. Fallon, C.L. Feasley, E. Brouhard, R. Kwon, C.M. West, J. Cunningham, and R. Brand, *Specific glycoforms of MUC5AC and endorepellin accurately distinguish mucinous from nonmucinous pancreatic cysts.* Molecular & Cellular Proteomics, 2013. **12**(10): p. 2724-2734.

197. Rebours, V., J. Le Faouder, S. Laouirem, M. Mebarki, M. Albuquerque, J.-M. Camadro, T. Léger, P. Ruszniewski, P. Lévy, and V. Paradis, *In situ proteomic analysis by MALDI imaging identifies ubiquitin and thymosin-β4 as markers of malignant intraductal pancreatic mucinous neoplasms.* Pancreatology, 2014. **14**(2): p. 117-124.

198. European Medicines Agency, *ICH Topic E15: Definitions for genomic biomarkers, pharmacogenomics, pharmacogenetics, genomic data and sample coding categories*. 2007: ICH Harmonised Tripartite Guideline. p. 4-5.

199. Tos, A.D., *The biology of epidermal growth factor receptor and its value as a prognostic/predictive factor.* The International journal of biological markers, 2007. **22**(1_suppl4): p. 3-9.

200. Khan, S.A., Z. Zeng, J. Shia, and P.B. Paty, *EGFR gene amplification and KRAS mutation predict response to combination targeted therapy in metastatic colorectal cancer.* Pathology & Oncology Research, 2017. **23**(3): p. 673-677.

201. Lievre, A., J.-B. Bachet, D. Le Corre, V. Boige, B. Landi, J.-F. Emile, J.-F. Côté, G. Tomasic, C. Penna, and M. Ducreux, *KRAS mutation status is predictive of response to cetuximab therapy in colorectal cancer.* Cancer research, 2006. **66**(8): p. 3992-3995.

202. Lynch, T.J., D.W. Bell, R. Sordella, S. Gurubhagavatula, R.A. Okimoto, B.W. Brannigan, P.L. Harris, S.M. Haserlat, J.G. Supko, and F.G. Haluska, *Activating mutations in the epidermal growth factor receptor underlying responsiveness of non–small-cell lung cancer to gefitinib.* New England Journal of Medicine, 2004. **350**(21): p. 2129-2139.

203. Schmid, K., N. Oehl, F. Wrba, R. Pirker, C. Pirker, and M. Filipits, *EGFR/KRAS/BRAF mutations in primary lung adenocarcinomas and corresponding locoregional lymph node metastases.* Clinical Cancer Research, 2009. **15**(14): p. 4554-4560.

204. Sorscher, S., *EGFR mutations and sensitivity to gefitinib.* The New England journal of medicine, 2004. **351**(12): p. 1260.

205. Oliveira-Cunha, M., W.G. Newman, and A.K. Siriwardena, *Epidermal growth factor receptor in pancreatic cancer.* Cancers, 2011. **3**(2): p. 1513-1526.

206. Friess, H., J. Kleeff, M. Korc, and M.W. Büchler, *Molecular aspects of pancreatic cancer and future perspectives.* Digestive surgery, 1999. **16**(4): p. 281-290.

207. Nedaeinia, R., A. Avan, M. Manian, R. Salehi, and M. Ghayour-Mobarhan, *EGFR as a potential target for the treatment of pancreatic cancer: dilemma and controversies.* Current drug targets, 2014. **15**(14): p. 1293-1301.

208. Kung, J.S., O.A. Lopez, E.E. McCoy, S. Reicher, and V.E. Eysselein, *Fluid genetic analyses predict the biological behavior of pancreatic cysts: three-year experience.* JOP. Journal of the Pancreas, 2014. **15**(5): p. 427-432.

209. Wu, J., H. Matthaei, A. Maitra, M. Dal Molin, L.D. Wood, J.R. Eshleman, M. Goggins, M.I. Canto, R.D. Schulick, and B.H. Edil, *Recurrent GNAS mutations define an unexpected pathway for pancreatic cyst development.* Science translational medicine, 2011. **3**(92): p. 92ra66-92ra66.

210. Kanda, M., H. Matthaei, J. Wu, S.M. Hong, J. Yu, M. Borges, R.H. Hruban, A. Maitra, K. Kinzler, and B. Vogelstein, *Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia.* Gastroenterology, 2012. **142**(4): p. 730-733.

211. Lu, S., T. Ahmed, P. Du, and Y. Wang, *Genomic variations in pancreatic cancer and potential opportunities for development of new approaches for diagnosis and treatment.* International journal of molecular sciences, 2017. **18**(6): p. 1201.

212. Macgregor-Das, A.M. and C.A. Iacobuzio-Donahue, *Molecular pathways in pancreatic carcinogenesis.* Journal of surgical oncology, 2013. **107**(1): p. 8-14.

213. Khalid, A., M. Zahid, S.D. Finkelstein, J.K. LeBlanc, N. Kaushik, N. Ahmad, W.R. Brugge, S.A. Edmundowicz, R.H. Hawes, and K.M. McGrath, *Pancreatic cyst fluid DNA analysis in evaluating pancreatic cysts: a report of the PANDA study.* Gastrointestinal Endoscopy, 2009. **69**(6): p. 1095-1102.

214. Reid, M.D., B. Saka, S. Balci, A.S. Goldblum, and N.V. Adsay, *Molecular genetics of pancreatic neoplasms and their morphologic correlates: an update on recent advances and potential diagnostic applications.* American journal of clinical pathology, 2014. **141**(2): p. 168-180.

215. Tulla, K.A. and A.V. Maker, *Can we better predict the biologic behavior of incidental IPMN? A comprehensive analysis of molecular diagnostics and biomarkers in intraductal papillary mucinous neoplasms of the pancreas.* Langenbeck's archives of surgery, 2018. **403**(2): p. 151-194.

216. Lee, J.-H., Y. Kim, J.-W. Choi, and Y.-S. Kim, *KRAS, GNAS, and RNF43 mutations in intraductal papillary mucinous neoplasm of the pancreas: a meta-analysis.* Springerplus, 2016. **5**(1): p. 1172.

217. Hosoda, W., E. Sasaki, Y. Murakami, K. Yamao, Y. Shimizu, and Y. Yatabe, *GNAS mutation is a frequent event in pancreatic intraductal papillary mucinous neoplasms and associated adenocarcinomas.* Virchows Archiv, 2015. **466**(6): p. 665-674.

218. Liang, W.S., D.W. Craig, J. Carpten, M.J. Borad, M.J. Demeure, G.J. Weiss, T. Izatt, S. Sinari, A. Christoforides, and J. Aldrich, *Genome-wide characterization of pancreatic adenocarcinoma patients using next generation sequencing.* PloS one, 2012. **7**(10): p. e43192.

219. Vasaikar, S.V., P. Straub, J. Wang, and B. Zhang, *LinkedOmics: analyzing multi-omics data within and across 32 cancer types.* Nucleic acids research, 2018. **46**(D1): p. D956-D963.

220. Weinstein, J.N., E.A. Collisson, G.B. Mills, K.R.M. Shaw, B.A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J.M. Stuart, and C.G.A.R. Network, *The cancer genome atlas pan-cancer analysis project.* Nature genetics, 2013. **45**(10): p. 1113.

221. Huang, S., K. Chaudhary, and L.X. Garmire, *More is better: recent progress in multi-omics data integration methods.* Frontiers in genetics, 2017. **8**: p. 84.

222. Chaudhary, K., O.B. Poirion, L. Lu, and L.X. Garmire, *Deep learning–based multi-omics integration robustly predicts survival in liver cancer.* Clinical Cancer Research, 2018. **24**(6): p. 1248-1259.

223. Zhang, J., J. Baran, A. Cros, J.M. Guberman, S. Haider, J. Hsu, Y. Liang, E. Rivkin, J. Wang, and B. Whitty, *International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data.* Database, 2011. **2011**.

224. Du, Y., B. Zhao, Z. Liu, X. Ren, W. Zhao, Z. Li, L. You, and Y. Zhao, *Molecular subtyping of pancreatic cancer: translating genomics and transcriptomics into the clinic.* Journal of Cancer, 2017. **8**(4): p. 513.

225. Peixoto, L., D. Risso, S.G. Poplawski, M.E. Wimmer, T.P. Speed, M.A. Wood, and T. Abel, *How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets.* Nucleic acids research, 2015. **43**(16): p. 7664-7674.

226. Colomé-Tatché, M. and F.J. Theis, *Statistical single cell multi-omics integration.* Current Opinion in Systems Biology, 2018. **7**: p. 54-59.

227. Risso, D., J. Ngai, T.P. Speed, and S. Dudoit, *Normalization of RNA-seq data using factor analysis of control genes or samples.* Nature biotechnology, 2014. **32**(9): p. 896.

228. Hernandez, Y.G. and A.L. Lucas, *MicroRNA in pancreatic ductal adenocarcinoma and its precursor lesions.* World journal of gastrointestinal oncology, 2016. **8**(1): p. 18.

229. Lee, L.S., A.E. Szafranska-Schwarzbach, D. Wylie, L.A. Doyle, A.M. Bellizzi, V. Kadiyala, S. Suleiman, P.A. Banks, B.F. Andruss, and D.L. Conwell, *Investigating microRNA expression profiles in pancreatic cystic neoplasms.* Clinical and translational gastroenterology, 2014. **5**(1): p. e47.

230. Vila-Navarro, E., M. Vila-Casadesús, L. Moreira, S. Duran-Sanchon, R. Sinha, À. Ginés, G. Fernández-Esparrach, R. Miquel, M. Cuatrecasas, and A. Castells, *MicroRNAs for detection of pancreatic neoplasia: biomarker discovery by next-generation sequencing and validation in 2 independent cohorts.* Annals of surgery, 2017. **265**(6): p. 1226.

231. Humeau, M., A. Vignolle-Vidoni, F. Sicard, F. Martins, B. Bournet, L. Buscail, J. Torrisani, and P. Cordelier, *Salivary microRNA in pancreatic cancer patients.* PloS one, 2015. **10**(6): p. e0130996.

232. Rapado-González, Ó., C. Martínez-Reglero, Á. Salgado-Barreira, B. Takkouche, R. López-López, M.M. Suárez-Cunqueiro, and L. Muinelo-Romay, *Salivary biomarkers for cancer diagnosis: a meta-analysis.* Annals of medicine, 2020. **52**(3-4): p. 131-144.

233. Davis, S. and P.S. Meltzer, *GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor.* Bioinformatics, 2007. **23**(14): p. 1846-1847.

234. Wang, Z., C.D. Monteiro, K.M. Jagodnik, N.F. Fernandez, G.W. Gundersen, A.D. Rouillard, S.L. Jenkins, A.S. Feldmann, K.S. Hu, and M.G. McDermott, *Extraction and analysis of signatures from the Gene Expression Omnibus by the crowd.* Nature communications, 2016. **7**(1): p. 1-11.

235. Cantor, R.M., K. Lange, and J.S. Sinsheimer, *Prioritizing GWAS results: a review of statistical methods and recommendations for their application.* The American Journal of Human Genetics, 2010. **86**(1): p. 6-22.

236. Gusev, A., A. Ko, H. Shi, G. Bhatia, W. Chung, B.W. Penninx, R. Jansen, E.J. De Geus, D.I. Boomsma, and F.A. Wright, *Integrative approaches for large-scale transcriptome-wide association studies.* Nature genetics, 2016. **48**(3): p. 245.

237. Silverman, B.R. and J. Shi, *Alterations of epigenetic regulators in pancreatic cancer and their clinical implications.* International journal of molecular sciences, 2016. **17**(12): p. 2138.

238. Sandoval, J. and M. Esteller, *Cancer epigenomics: beyond genomics.* Current opinion in genetics & development, 2012. **22**(1): p. 50-55.

239. Zagorac, S., L. Garcia-Bermejo, and B. Sainz, *The Epigenetic Landscape of Pancreatic Cancer Stem Cells.* Epigenomes, 2018. **2**(2): p. 10.

240. Jones, P.A. and S.B. Baylin, *The fundamental role of epigenetic events in cancer.* Nature reviews genetics, 2002. **3**(6): p. 415-428.

241. Jacot, W., E. Lopez-Crapez, C. Mollevi, F. Boissière-Michot, J. Simony-Lafontaine, A. Ho-Pun-Cheung, E. Chartron, C. Theillet, A. Lemoine, and R. Saffroy, *BRCA1 Promoter Hypermethylation is Associated with Good Prognosis and Chemosensitivity in Triple-Negative Breast Cancer.* Cancers, 2020. **12**(4): p. 828.

242. Toll, A.D., A. Dasgupta, M. Potoczek, C.J. Yeo, C.G. Kleer, J.R. Brody, and A.K. Witkiewicz, *Implications of enhancer of zeste homologue 2 expression in pancreatic ductal adenocarcinoma.* Human pathology, 2010. **41**(9): p. 1205-1209.

243. Chen, Y., D. Xie, W.Y. Li, C.M. Cheung, H. Yao, C.Y. Chan, C.-y. Chan, F.-P. Xu, Y.-H. Liu, and J.J. Sung, *RNAi targeting EZH2 inhibits tumor growth and liver metastasis of pancreatic cancer in vivo.* Cancer letters, 2010. **297**(1): p. 109-116.

244. Jin, X., C. Yang, P. Fan, J. Xiao, W. Zhang, S. Zhan, T. Liu, D. Wang, and H. Wu, *CDK5/FBW7-dependent ubiquitination and degradation of EZH2 inhibits pancreatic cancer cell migration and invasion.* Journal of Biological Chemistry, 2017. **292**(15): p. 6269-6280.

245. Rahman, M.M., A.C. Brane, and T.O. Tollefsbol, *MicroRNAs and Epigenetics Strategies to Reverse Breast Cancer.* Cells, 2019. **8**(10): p. 1214.

246. Fujiyama, Y., Y. Kumamoto, N. Nishizawa, S. Nakamoto, H. Harada, K. Yokota, Y. Tanaka, K. Igarashi, H. Oiki, and K. Okuwaki, *Promoter DNA Hypermethylation of the Cysteine Dioxygenase 1 (CDO1) Gene in Intraductal Papillary Mucinous Neoplasm (IPMN).* Annals of Surgical Oncology, 2020. **27**(10): p. 4007-4016.

247. Meissner, A., *What can epigenomics do for you?* Genome Biology, 2012. **13**(10): p. 420.

248. Rakyan, V.K., T.A. Down, D.J. Balding, and S. Beck, *Epigenome-wide association studies for common human diseases.* Nature Reviews Genetics, 2011. **12**(8): p. 529-541.

249. Yu, K.-H. and M. Snyder, *Omics profiling in precision oncology.* Molecular & Cellular Proteomics, 2016. **15**(8): p. 2525-2536.

250. Kwon, R.S. and D.M. Simeone, *The use of protein-based biomarkers for the diagnosis of cystic tumors of the pancreas.* International journal of proteomics, 2011. **2011**.

251. Ngamruengphong, S., M.J. Bartel, and M. Raimondo, *Cyst carcinoembryonic antigen in differentiating pancreatic cysts: a meta-analysis.* Digestive and Liver Disease, 2013. **45**(11): p. 920-926.

252. Al-Haddad, M., J. DeWitt, S. Sherman, C.M. Schmidt, J.K. LeBlanc, L. McHenry, G. Coté, A.H. El Chafic, L. Luz, and J.S. Stuart, *Performance characteristics of molecular (DNA) analysis for the diagnosis of mucinous pancreatic cysts.* Gastrointestinal endoscopy, 2014. **79**(1): p. 79-87.

253. Kurita, Y., T. Kuwahara, K. Hara, N. Mizuno, N. Okuno, S. Matsumoto, M. Obata, H. Koda, M. Tajika, and Y. Shimizu, *Diagnostic ability of artificial intelligence using deep learning analysis of cyst fluid in differentiating malignant from benign pancreatic cystic lesions.* Scientific reports, 2019. **9**(1): p. 1-9.

254. Azadeh, A., R. Felix, T. Krause, M. Bernhardt, P. Jo, A. König, K. Mathias, L. Andreas, M. Ghadimi, and G. Jochen, *CA19-9 for detecting recurrence of pancreatic cancer.* Scientific Reports (Nature Publisher Group), 2020. **10**(1).

255. Porterfield, M., P. Zhao, H. Han, J. Cunningham, K. Aoki, D.D. Von Hoff, M.J. Demeure, J.M. Pierce, M. Tiemeyer, and L. Wells, *Discrimination between adenocarcinoma and normal pancreatic ductal fluid by proteomic and glycomic analysis.* Journal of proteome research, 2014. **13**(2): p. 395-407.

256. de Oliveira, G., P. Paccielli Freire, S. Santiloni Cury, D. de Moraes, J. Santos Oliveira, M. Dal-Pai-Silva, P.P.d. Reis, and R. Francisco Carvalho, *An Integrated Meta-Analysis of Secretome and Proteome Identify Potential Biomarkers of Pancreatic Ductal Adenocarcinoma.* Cancers, 2020. **12**(3): p. 716.

257. Nagata, K., M. Horinouchi, M. Saitou, M. Higashi, M. Nomoto, M. Goto, and S. Yonezawa, *Mucin expression profile in pancreatic cancer and the precursor lesions.* Journal of hepato-biliary-pancreatic surgery, 2007. **14**(3): p. 243-254.

258. Carrara, S., M.G. Cangi, P.G. Arcidiacono, F. Perri, M.C. Petrone, G. Mezzi, C. Boemo, A. Talarico, E. Dal Cin, and G. Grassini, *Mucin expression pattern in pancreatic diseases: findings from EUS-guided fine-needle aspiration biopsies.* American Journal of Gastroenterology, 2011. **106**(7): p. 1359-1363.

259. Yip-Schneider, M.T., H. Wu, R.P. Dumas, B.A. Hancock, N. Agaram, M. Radovich, and C.M. Schmidt, *Vascular endothelial growth factor, a novel and highly accurate pancreatic fluid biomarker for serous pancreatic cysts.* Journal of the American College of Surgeons, 2014. **218**(4): p. 608-617.

260. Malaker, S.A., K. Pedram, M.J. Ferracane, B.A. Bensing, V. Krishnan, C. Pett, J. Yu, E.C. Woods, J.R. Kramer, and U. Westerlind, *The mucin-selective protease StcE enables molecular and functional analysis of human cancer-associated mucins.* Proceedings of the National Academy of Sciences of the United States of America, 2019. **116**(15): p. 7278-7287.

261. Rudnick, P.A., S.P. Markey, J. Roth, Y. Mirokhin, X. Yan, D.V. Tchekhovskoi, N.J. Edwards, R.R. Thangudu, K.A. Ketchum, and C.R. Kinsinger, *A description of the clinical proteomic tumor analysis consortium (CPTAC) common data analysis pipeline.* Journal of proteome research, 2016. **15**(3): p. 1023-1032.

262. Côté, R.G., J. Griss, J.A. Dianes, R. Wang, J.C. Wright, H.W. van den Toorn, B. van Breukelen, A.J. Heck, N. Hulstaert, and L. Martens, *The PRoteomics IDEntification (PRIDE) Converter 2 framework: an improved suite of tools to facilitate data submission to the PRIDE database and the ProteomeXchange consortium.* Molecular & Cellular Proteomics, 2012. **11**(12): p. 1682-1689.

263. Mertins, P., D. Mani, K.V. Ruggles, M.A. Gillette, K.R. Clauser, P. Wang, X. Wang, J.W. Qiao, S. Cao, and F. Petralia, *Proteogenomics connects somatic mutations to signalling in breast cancer.* Nature, 2016. **534**(7605): p. 55-62.

264. Zhang, B., J. Wang, X. Wang, J. Zhu, Q. Liu, Z. Shi, M.C. Chambers, L.J. Zimmerman, K.F. Shaddox, and S. Kim, *Proteogenomic characterization of human colon and rectal cancer.* Nature, 2014. **513**(7518): p. 382-387.

265. Costello, E., *A metabolomics-based biomarker signature discriminates pancreatic cancer from chronic pancreatitis.* Gut, 2018. **67**(1): p. 2-3.

266. Cantor, J.R. and D.M. Sabatini, *Cancer cell metabolism: one hallmark, many faces.* Cancer discovery, 2012. **2**(10): p. 881-898.

267. Mayerle, J., H. Kalthoff, R. Reszka, B. Kamlage, E. Peter, B. Schniewind, S.G. Maldonado, C. Pilarsky, C.-D. Heidecke, and P. Schatz, *Metabolic biomarker signature to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis.* Gut, 2018. **67**(1): p. 128-137.

268. Zhang, G., P. He, H. Tan, A. Budhu, J. Gaedcke, B.M. Ghadimi, T. Ried, H.G. Yfantis, D.H. Lee, and A. Maitra, *Integration of metabolomics and transcriptomics revealed a fatty acid network exerting growth inhibitory effects in human pancreatic cancer.* Clinical cancer research, 2013. **19**(18): p. 4983-4993.

269. Sud, M., E. Fahy, D. Cotter, K. Azam, I. Vadivelu, C. Burant, A. Edison, O. Fiehn, R. Higashi, and K.S. Nair, *Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools.* Nucleic acids research, 2016. **44**(D1): p. D463-D470.

270. Springer, S., D.L. Masica, M. Dal Molin, C. Douville, C.J. Thoburn, B. Afsari, L. Li, J.D. Cohen, E. Thompson, and P.J. Allen, *A multimodality test to guide the management of patients with a pancreatic cyst.* Science translational medicine, 2019. **11**(501): p. eaav4772.

271. Harsha, H., K. Kandasamy, P. Ranganathan, S. Rani, S. Ramabadran, S. Gollapudi, L. Balakrishnan, S.B. Dwivedi, D. Telikicherla, and L.D.N. Selvan, *A compendium of potential biomarkers of pancreatic cancer.* PLoS medicine, 2009. **6**(4): p. e1000046.

272. Springer, S., Y. Wang, M. Dal Molin, D.L. Masica, Y. Jiao, I. Kinde, A. Blackford, S.P. Raman, C.L. Wolfgang, and T. Tomita, *A combination of molecular markers and clinical features improve the classification of pancreatic cysts.* Gastroenterology, 2015. **149**(6): p. 1501-1510.

273. Liu, P., Y. Wang, and X. Li, *Targeting the untargetable KRAS in cancer therapy.* Acta Pharmaceutica Sinica B, 2019. **9**(5): p. 871-879.

274. Natale, F., M. Vivo, G. Falco, and T. Angrisano, *Deciphering DNA methylation signatures of pancreatic cancer and pancreatitis.* Clinical epigenetics, 2019. **11**(1): p. 132.

275. Kent, O.A., R.R. Chivukula, M. Mullendore, E.A. Wentzel, G. Feldmann, K.H. Lee, S. Liu, S.D. Leach, A. Maitra, and J.T. Mendell, *Repression of the miR-143/145 cluster by oncogenic Ras initiates a tumor-promoting feed-forward pathway.* Genes & development, 2010. **24**(24): p. 2754-2759.

276. Kent, O.A., J.T. Mendell, and R. Rottapel, *Transcriptional regulation of miR-31 by oncogenic KRAS mediates metastatic phenotypes by repressing RASA1.* Molecular Cancer Research, 2016. **14**(3): p. 267-277.

277. Rachagani, S., M.A. Macha, N. Heimann, P. Seshacharyulu, D. Haridas, S. Chugh, and S.K. Batra, *Clinical implications of miRNAs in the pathogenesis, diagnosis and therapy of pancreatic cancer.* Advanced drug delivery reviews, 2015. **81**: p. 16-33.

278. Pupo, E., D. Avanzato, E. Middonti, F. Bussolino, and L. Lanzetti, *KRAS-driven metabolic rewiring reveals novel actionable targets in cancer.* Frontiers in oncology, 2019. **9**: p. 848.

279. Meng, C., O.A. Zeleznik, G.G. Thallinger, B. Kuster, A.M. Gholami, and A.C. Culhane, *Dimension reduction techniques for the integrative analysis of multi-omics data.* Briefings in bioinformatics, 2016. **17**(4): p. 628-641.

280. Argelaguet, R., B. Velten, D. Arnol, S. Dietrich, T. Zenz, J.C. Marioni, F. Buettner, W. Huber, and O. Stegle, *Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets.* Molecular systems biology, 2018. **14**(6).

281. Meng, C., B. Kuster, A.C. Culhane, and A.M. Gholami, *A multivariate approach to the integration of multi-omics datasets.* BMC bioinformatics, 2014. **15**(1): p. 162.

282. Singh, A., C.P. Shannon, B. Gautier, F. Rohart, M. Vacher, S.J. Tebbutt, and K.-A. Lê Cao, *DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays.* Bioinformatics, 2019. **35**(17): p. 3055-3062.

283. Vasaikar, S., P. Straub, J. Wang, and B. Zhang, *LinkedOmics: Analyzing multi-omics data within and across 32 cancer types.* 2019. **79**(13): p. 5112.

284. Kong, L., P. Liu, M. Zheng, B. Xue, K. Liang, and X. Tan, *Multi-omics analysis based on integrated genomics, epigenomics and transcriptomics in pancreatic cancer.* Epigenomics, 2020. **12**(6): p. 507-524.

285. Bhatia, S., J. Monkman, T. Blick, P.H. Duijf, S.H. Nagaraj, and E.W. Thompson, *Multi-omics characterization of the spontaneous mesenchymal–epithelial transition in the PMC42 breast cancer cell lines.* Journal of clinical medicine, 2019. **8**(8): p. 1253.

286. Karczewski, K.J. and M.P. Snyder, *Integrative omics for health and disease.* Nature Reviews Genetics, 2018. **19**(5): p. 299.

287. Page, M.J., J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, and S.E. Brennan, *The PRISMA 2020*

statement: an updated guideline for reporting systematic reviews. Systematic reviews, 2021. **10**(1): p. 1-11.

288. Hickey, G.L., S.W. Grant, J. Dunning, and M. Siepe, *Statistical primer: sample size and power calculations—why, when and how?* European journal of cardio-thoracic surgery, 2018. **54**(1): p. 4-9.
289. Whiting, P.F., A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, M.M. Leeflang, J.A. Sterne, P.M. Bossuyt, and Q.-. Group*, *QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies.* Annals of internal medicine, 2011. **155**(8): p. 529-536.
290. Harrer, M., P. Cuijpers, T.A. Furukawa, and D.D. Ebert, *Doing meta-analysis with R: A hands-on guide*. 2021: Chapman and Hall/CRC.
291. Viechtbauer, W., *Conducting meta-analyses in R with the metafor package.* Journal of statistical software, 2010. **36**(3): p. 1-48.
292. Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M.Y. Hein, T. Geiger, M. Mann, and J. Cox, *The Perseus computational platform for comprehensive analysis of (prote) omics data.* Nature methods, 2016. **13**(9): p. 731-740.
293. Mazzara, S., R.L. Rossi, R. Grifantini, S. Donizetti, S. Abrignani, and M. Bombaci, *CombiROC: an interactive web tool for selecting accurate marker combinations of omics data.* Scientific reports, 2017. **7**(1): p. 1-11.
294. Mueller, C., B. Muller, and A.P. Perruchoud, *Biomarkers: past, present, and future.* Swiss Medical Weekly, 2008. **138**(15/16): p. 225.
295. Chatterjee, S.K. and B.R. Zetter, *Cancer biomarkers: knowing the present and predicting the future.* 2005. **1**(1): p. 37-50.
296. Simon, A.E., J. Waller, K. Robb, and J. Wardle, *Patient delay in presentation of possible cancer symptoms: the contribution of knowledge and attitudes in a population sample from the United Kingdom.* Cancer Epidemiology and Prevention Biomarkers, 2010. **19**(9): p. 2272-2277.
297. Williams, F.M., *Biomarkers: in combination they may do better.* Arthritis research & therapy, 2009. **11**(5): p. 1-2.
298. Bengtsson, A., R. Andersson, and D. Ansari, *The actual 5-year survivors of pancreatic ductal adenocarcinoma based on real-world data.* Scientific Reports, 2020. **10**(1): p. 1-9.

299. Mills, K., L. Birt, J.D. Emery, N. Hall, J. Banks, M. Johnson, J. Lancaster, W. Hamilton, G.P. Rubin, and F.M. Walter, *Understanding symptom appraisal and help-seeking in people with symptoms suggestive of pancreatic cancer: a qualitative study.* BMJ open, 2017. **7**(9): p. e015682.

300. Jentzsch, V., J.A. Davis, and M. Djamgoz, *Pancreatic Cancer (PDAC): introduction of evidence-based complementary measures into integrative clinical management.* Cancers, 2020. **12**(11): p. 3096.

301. van Manen, L., J.V. Groen, H. Putter, M. Pichler, A.L. Vahrmeijer, B.A. Bonsing, and J.S.D. Mieog, *Stage-Specific Value of Carbohydrate Antigen 19-9 and Carcinoembryonic Antigen Serum Levels on Survival and Recurrence in Pancreatic Cancer: A Single Center Study and Meta-Analysis.* Cancers, 2020. **12**(10): p. 2970.

302. Luo, G., Z. Fan, H. Cheng, K. Jin, M. Guo, Y. Lu, C. Yang, K. Fan, Q. Huang, and J. Long, *New observations on the utility of CA19-9 as a biomarker in Lewis negative patients with pancreatic cancer.* Pancreatology, 2018. **18**(8): p. 971-976.

303. Azizian, A., F. Rühlmann, T. Krause, M. Bernhardt, P. Jo, A. König, M. Kleiß, A. Leha, M. Ghadimi, and J. Gaedcke, *CA19-9 for detecting recurrence of pancreatic cancer.* Scientific reports, 2020. **10**(1): p. 1-10.

304. Zhang, L., S. Sanagapalli, and A. Stoita, *Challenges in diagnosis of pancreatic cancer.* World journal of gastroenterology, 2018. **24**(19): p. 2047.

305. Saito, S., K. Taguchi, N. Nishimura, A. Watanabe, K. Ogoshi, M. Niwa, T. Furukawa, and M. Takahashi, *Clinical usefulness of computer-assisted diagnosis using combination assay of tumor markers for pancreatic carcinoma.* Cancer, 1993. **72**(2): p. 381-388.

306. Zhang, X., X. Shi, X. Lu, Y. Li, C. Zhan, M.L. Akhtar, L. Yang, Y. Bai, J. Zhao, and Y. Wang, *Novel metabolomics serum biomarkers for pancreatic ductal adenocarcinoma by the comparison of pre-, postoperative and normal samples.* Journal of Cancer, 2020. **11**(16): p. 4641.

307. Yanagisawa, K., S. Tomida, K. Matsuo, C. Arima, M. Kusumegi, Y. Yokoyama, S.B. Ko, N. Mizuno, T. Kawahara, and Y. Kuroyanagi, *Seven-signal proteomic signature for detection of operable pancreatic ductal adenocarcinoma and their discrimination from autoimmune pancreatitis.* International journal of proteomics, 2012. **2012**.

308. Gao, H., Z. Zheng, Z. Yue, F. Liu, L. Zhou, and X. Zhao, *Evaluation of serum diagnosis of pancreatic cancer by using surface-enhanced laser desorption/ionization time-of-flight mass spectrometry.* International journal of molecular medicine, 2012. **30**(5): p. 1061-1068.

309. Yu, S., Y. Li, Z. Liao, Z. Wang, Z. Wang, Y. Li, L. Qian, J. Zhao, H. Zong, and B. Kang, *Plasma extracellular vesicle long RNA profiling identifies a diagnostic signature for the detection of pancreatic ductal adenocarcinoma.* Gut, 2020. **69**(3): p. 540-550.

310. Le Calvez-Kelm, F., M. Foll, M.B. Wozniak, T.M. Delhomme, G. Durand, P. Chopard, M. Pertesi, E. Fabianova, Z. Adamcakova, and I. Holcatova, *KRAS mutations in blood circulating cell-free DNA: a pancreatic cancer case-control.* Oncotarget, 2016. **7**(48): p. 78827.

311. Di Gangi, I.M., T. Mazza, A. Fontana, M. Copetti, C. Fusilli, A. Ippolito, F. Mattivi, A. Latiano, A. Andriulli, and U. Vrhovsek, *Metabolomic profile in pancreatic cancer patients: a consensus-based approach to identify highly discriminating metabolites.* Oncotarget, 2016. **7**(5): p. 5815.

312. Kahlert, C., M. Fiala, G. Musso, N. Halama, S. Keim, M. Mazzone, F. Lasitschka, M. Pecqueux, F. Klupp, and T. Schmidt, *Prognostic impact of a compartment-specific angiogenic marker profile in patients with pancreatic cancer.* Oncotarget, 2014. **5**(24): p. 12978.

313. Wlodarczyk, B., A. Borkowska, P. Wlodarczyk, E. Malecka-Panas, and A. Gasiorowska, *Serum Levels of Insulin-like Growth Factor 1 and Insulin-like Growth Factor–binding Protein 2 as a Novel Biomarker in the Detection of Pancreatic Adenocarcinoma.* Journal of clinical gastroenterology, 2020. **54**(9): p. e83-e88.

314. Hussein, N.A.E.M., Z.A. El Kholy, M.M. Anwar, M.A. Ahmad, and S.M. Ahmad, *Plasma miR-22-3p, miR-642b-3p and miR-885-5p as diagnostic biomarkers for pancreatic cancer.* Journal of cancer research and clinical oncology, 2017. **143**(1): p. 83-93.

315. Cote, G.A., A.J. Gore, S.D. McElyea, L.E. Heathers, H. Xu, S. Sherman, and M. Korc, *A pilot study to develop a diagnostic test for pancreatic ductal adenocarcinoma based on differential expression of select miRNA in plasma and bile.* The American journal of gastroenterology, 2014. **109**(12): p. 1942.

316. Savareh, B.A., H.A. Aghdaie, A. Behmanesh, A. Bashiri, A. Sadeghi, M. Zali, and R. Shams, *A machine learning approach identified a diagnostic model for pancreatic cancer through using circulating microRNA signatures.* Pancreatology, 2020. **20**(6): p. 1195-1204.

317. Hasan, S., R. Jacob, U. Manne, and R. Paluri, *Advances in pancreatic cancer biomarkers.* Oncology reviews, 2019. **13**(1).

318. Kaur, S., M.J. Baine, M. Jain, A.R. Sasson, and S.K. Batra, *Early diagnosis of pancreatic cancer: challenges and new developments.* Biomarkers in medicine, 2012. **6**(5): p. 597-612.

319. O'Brien, D.P., N.S. Sandanayake, C. Jenkinson, A. Gentry-Maharaj, S. Apostolidou, E.-O. Fourkala, S. Camuzeaux, O. Blyuss, R. Gunu, and A. Dawnay, *Serum CA19-9 is significantly upregulated up to 2 years before diagnosis with pancreatic cancer: implications for early disease detection.* Clinical Cancer Research, 2015. **21**(3): p. 622-631.

320. Peng, H.-Y., M.-C. Chang, C.-M. Hu, H.-I. Yang, W.-H. Lee, and Y.-T. Chang, *Thrombospondin-2 is a highly specific diagnostic marker and is associated with prognosis in pancreatic cancer.* Annals of surgical oncology, 2019. **26**(3): p. 807-814.

321. Taniuchi, K., M. Tsuboi, M. Sakaguchi, and T. Saibara, *Measurement of serum PODXL concentration for detection of pancreatic cancer.* OncoTargets and therapy, 2018. **11**: p. 1433.

322. Lohrmann, C., E.M. O'Reilly, J.A. O'Donoghue, N. Pandit-Taskar, J.A. Carrasquillo, S.K. Lyashchenko, S. Ruan, R. Teng, W. Scholz, and P.W. Maffuid, *Retooling a blood-based biomarker: phase I assessment of the high-affinity CA19-9 antibody HuMab-5B1 for immuno-PET imaging of pancreatic cancer.* Clinical Cancer Research, 2019. **25**(23): p. 7014-7023.

323. van Veldhuisen, E., J.A. Vogel, S. Klompmaker, O.R. Busch, H.W. van Laarhoven, K.P. van Lienden, J.W. Wilmink, H.A. Marsman, and M.G. Besselink, *Added value of CA19-9 response in predicting resectability of locally advanced pancreatic cancer following induction chemotherapy.* HPB, 2018. **20**(7): p. 605-611.

324. Herreros-Villanueva, M., L. Ruiz-Rebollo, M. Montes, M. Rodriguez-Lopez, M. Francisco, J. Cubiella, E. Iyo, E. Garabitos, E.M. Moneo, and M. Martos, *CA19-9*

capability as predictor of pancreatic cancer resectability in a Spanish cohort. Molecular biology reports, 2020. **47**(3): p. 1583-1588.

325. Turanli, B., E. Yildirim, G. Gulfidan, K.Y. Arga, and R. Sinha, *Current State of "Omics" Biomarkers in Pancreatic Cancer.* Journal of Personalized Medicine, 2021. **11**(2): p. 127.

326. Long, N.P., K.H. Jung, N.H. Anh, H.H. Yan, T.D. Nghi, S. Park, S.J. Yoon, J.E. Min, H.M. Kim, and J.H. Lim, *An integrative data mining and omics-based translational model for the identification and validation of oncogenic biomarkers of pancreatic cancer.* Cancers, 2019. **11**(2): p. 155.

327. Grunnet, M., M. Mau-Sørensen, and N. Brünner, *Tissue inhibitor of metalloproteinase 1 (TIMP-1) as a biomarker in gastric cancer: a review.* Scandinavian journal of gastroenterology, 2013. **48**(8): p. 899-905.

328. Wang, C.-S., T.-L. Wu, K.-C. Tsao, and C.-F. Sun, *Serum TIMP-1 in gastric cancer patients: a potential prognostic biomarker.* Annals of Clinical & Laboratory Science, 2006. **36**(1): p. 23-30.

329. Vočka, M., D. Langer, V. Fryba, J. Petrtyl, T. Hanus, M. Kalousova, T. Zima, and L. Petruzelka, *Serum levels of TIMP-1 and MMP-7 as potential biomarkers in patients with metastatic colorectal cancer.* The International journal of biological markers, 2019. **34**(3): p. 292-301.

330. Meng, C., X. Yin, J. Liu, K. Tang, H. Tang, and J. Liao, *TIMP-1 is a novel serum biomarker for the diagnosis of colorectal cancer: A meta-analysis.* PLoS ONE, 2018. **13**(11): p. e0207039.

331. Huang, X., Y. Lan, E. Li, J. Li, Q. Deng, and X. Deng, *Diagnostic values of MMP-7, MMP-9, MMP-11, TIMP-1, TIMP-2, CEA, and CA19-9 in patients with colorectal cancer.* Journal of International Medical Research, 2021. **49**(5): p. 03000605211012570.

332. Cheng, G., X. Fan, M. Hao, J. Wang, X. Zhou, and X. Sun, *Higher levels of TIMP-1 expression are associated with a poor prognosis in triple-negative breast cancer.* Molecular cancer, 2016. **15**(1): p. 1-13.

333. Wu, Z.S., Q. Wu, J.H. Yang, H.Q. Wang, X.D. Ding, F. Yang, and X.C. Xu, *Prognostic significance of MMP-9 and TIMP-1 serum and tissue expression in breast cancer.* International journal of cancer, 2008. **122**(9): p. 2050-2056.

334. Prokopchuk, O., B. Grünwald, U. Nitsche, C. Jäger, O.L. Prokopchuk, E.C. Schubert, H. Friess, M.E. Martignoni, and A. Krüger, *Elevated systemic levels of the matrix metalloproteinase inhibitor TIMP-1 correlate with clinical markers of cachexia in patients with chronic pancreatitis and pancreatic cancer.* BMC cancer, 2018. **18**(1): p. 1-12.

335. Ilies, M., P.K. Sappa, C.A. Iuga, F. Loghin, M.G. Salazar, F.U. Weiss, G. Beyer, M.M. Lerch, U. Völker, and J. Mayerle, *Plasma protein profiling of patients with intraductal papillary mucinous neoplasm of the pancreas as potential precursor lesions of pancreatic cancer.* Clinica Chimica Acta, 2018. **477**: p. 127-134.

336. Slater, E.P., V. Fendrich, K. Strauch, S. Rospleszcz, A. Ramaswamy, E. Mätthai, B. Chaloupka, T.M. Gress, P. Langer, and D.K. Bartsch, *LCN2 and TIMP1 as potential serum markers for the early detection of familial pancreatic cancer.* Translational oncology, 2013. **6**(2): p. 99-103.

337. Wang, Y., J.-M. Yuan, A. Pan, and W.-P. Koh, *Tissue inhibitor matrix metalloproteinase 1 and risk of type 2 diabetes in a Chinese population.* BMJ Open Diabetes Research and Care, 2020. **8**(1): p. e001051.

338. Lee, S.W., K.E. Song, D.S. Shin, S.M. Ahn, E.S. Ha, D.J. Kim, M.S. Nam, and K.-W. Lee, *Alterations in peripheral blood levels of TIMP-1, MMP-2, and MMP-9 in patients with type-2 diabetes.* Diabetes research and clinical practice, 2005. **69**(2): p. 175-179.

339. Papazoglou, D., K. Papatheodorou, N. Papanas, T. Papadopoulos, T. Gioka, G. Kabouromiti, S. Kotsiou, and E. Maltezos, *Matrix metalloproteinase-1 and tissue inhibitor of metalloproteinases-1 levels in severely obese patients: what is the effect of weight loss?* Experimental and clinical endocrinology & diabetes, 2010. **118**(10): p. 730-734.

340. Xie, Z.-B., Y.-F. Zhang, C. Jin, Y.-S. Mao, and D.-L. Fu, *LRG-1 promotes pancreatic cancer growth and metastasis via modulation of the EGFR/p38 signaling.* Journal of Experimental & Clinical Cancer Research, 2019. **38**(1): p. 1-12.

341. Capello, M., L.E. Bantis, G. Scelo, Y. Zhao, P. Li, D.S. Dhillon, N.J. Patel, D.L. Kundnani, H. Wang, and J.L. Abbruzzese, *Sequential validation of blood-based protein biomarker candidates for early-stage pancreatic cancer.* JNCI: Journal of the National Cancer Institute, 2017. **109**(4).

342. Park, J., Y. Choi, J. Namkung, S.G. Yi, H. Kim, J. Yu, Y. Kim, M.-S. Kwon, W. Kwon, and D.-Y. Oh, *Diagnostic performance enhancement of pancreatic cancer using proteomic multimarker panel.* Oncotarget, 2017. **8**(54): p. 93117.

343. Gao, Y., J. Wang, Y. Zhou, S. Sheng, S.Y. Qian, and X. Huo, *Evaluation of serum CEA, CA19-9, CA72-4, CA125 and ferritin as diagnostic markers and factors of clinical parameters for colorectal cancer.* Scientific reports, 2018. **8**(1): p. 1-9.

344. Hing, J., C. Mok, P. Tan, S. Sudhakar, C. Seah, W. Lee, and S. Tan, *Clinical utility of tumour marker velocity of cancer antigen 15–3 (CA 15–3) and carcinoembryonic antigen (CEA) in breast cancer surveillance.* The Breast, 2020. **52**: p. 95-101.

345. Cheng, C., Y. Yang, W. Yang, D. Wang, and C. Yao, *The diagnostic value of CEA for lung cancer-related malignant pleural effusion in China: a meta-analysis.* Expert review of respiratory medicine, 2022. **16**(1): p. 99-108.

346. Xing, H., J. Wang, Y. Wang, M. Tong, H. Hu, C. Huang, and D. Li, *Diagnostic value of CA 19-9 and carcinoembryonic antigen for pancreatic cancer: a meta-analysis.* Gastroenterology research and practice, 2018.

347. Ferri, M.J., M. Saez, J. Figueras, E. Fort, M. Sabat, S. López-Ben, R. de Llorens, R.N. Aleixandre, and R. Peracaula, *Improved pancreatic adenocarcinoma diagnosis in jaundiced and non-jaundiced pancreatic adenocarcinoma patients through the combination of routine clinical markers associated to pancreatic adenocarcinoma pathophysiology.* PLoS One, 2016. **11**(1): p. e0147214.

348. Gu, Y.-L., C. Lan, H. Pei, S.-N. Yang, Y.-F. Liu, and L.-L. Xiao, *Applicative value of serum CA19-9, CEA, CA125 and CA242 in diagnosis and prognosis for patients with pancreatic cancer treated by concurrent chemoradiotherapy.* Asian Pacific journal of cancer prevention, 2015. **16**(15): p. 6569-6573.

349. Charkhchi, P., C. Cybulski, J. Gronwald, F.O. Wong, S.A. Narod, and M.R. Akbari, *CA125 and ovarian cancer: a comprehensive review.* Cancers, 2020. **12**(12): p. 3730.

350. Meng, Q., S. Shi, C. Liang, J. Xiang, D. Liang, B. Zhang, Y. Qin, S. Ji, W. Xu, and J. Xu, *Diagnostic accuracy of a CA125-based biomarker panel in patients with pancreatic cancer: a systematic review and meta-analysis.* Journal of Cancer, 2017. **8**(17): p. 3615.

351. Ozkan, H., M. Kaya, and A. Cengiz, *Comparison of tumor marker CA 242 with CA 19-9 and carcinoembryonic antigen (CEA) in pancreatic cancer.* Hepato-gastroenterology, 2003. **50**(53): p. 1669-1674.

352. Zhang, Y., J. Yang, H. Li, Y. Wu, H. Zhang, and W. Chen, *Tumor markers CA19-9, CA242 and CEA in the diagnosis of pancreatic cancer: a meta-analysis.* International journal of clinical and experimental medicine, 2015. **8**(7): p. 11683.

353. Dou, H., G. Sun, and L. Zhang, *CA242 as a biomarker for pancreatic cancer and other diseases*, in *Progress in Molecular Biology and Translational Science*. 2019, Elsevier. p. 229-239.

354. Kosanam, H., I. Prassas, C.C. Chrystoja, I. Soleas, A. Chan, A. Dimitromanolakis, I.M. Blasutig, F. Rückert, R. Gruetzmann, and C. Pilarsky, *Laminin, gamma 2 (LAMC2): a promising new putative pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues.* Molecular & Cellular Proteomics, 2013. **12**(10): p. 2820-2832.

355. Jin, G., Q. Ruan, F. Shangguan, and L. Lan, *RUNX2 and LAMC2: promising pancreatic cancer biomarkers identified by an integrative data mining of pancreatic adenocarcinoma tissues.* Aging (Albany NY), 2021. **13**(19): p. 22963.

356. Zhang, J., Y. Wang, T. Zhao, Y. Li, L. Tian, J. Zhao, and J. Zhang, *Evaluation of serum MUC5AC in combination with CA19-9 for the diagnosis of pancreatic cancer.* World Journal of Surgical Oncology, 2020. **18**(1): p. 1-7.

357. Kaur, S., L.M. Smith, A. Patel, M. Menning, D.C. Watley, S.S. Malik, S.R. Krishn, K. Mallya, A. Aithal, and A.R. Sasson, *A combination of MUC5AC and CA19-9 improves the diagnosis of pancreatic cancer: a multicenter study.* The American journal of gastroenterology, 2017. **112**(1): p. 172.

358. Yue, T., K.A. Maupin, B. Fallon, L. Li, K. Partyka, M.A. Anderson, D.E. Brenner, K. Kaul, H. Zeh, and A.J. Moser, *Enhanced discrimination of malignant from benign pancreatic disease by measuring the CA 19-9 antigen on specific protein carriers.* PloS one, 2011. **6**(12): p. e29180.

359. Mattila, N., H. Seppänen, H. Mustonen, B. Przybyla, C. Haglund, and R. Lassila, *Preoperative biomarker panel, including fibrinogen and FVIII, improves diagnostic accuracy for pancreatic ductal adenocarcinoma.* Clinical and Applied Thrombosis/Hemostasis, 2018. **24**(8): p. 1267-1275.

360. Bellone, G., C. Smirne, F.A. Mauri, E. Tonel, A. Carbone, A. Buffolino, L. Dughera, A. Robecchi, M. Pirisi, and G. Emanuelli, *Cytokine expression profile in human pancreatic carcinoma cells and in surgical specimens: implications for survival.* Cancer Immunology, Immunotherapy, 2006. **55**(6): p. 684-698.

361. Nolen, B.M., R.E. Brand, D. Prosser, L. Velikokhatnaya, P.J. Allen, H.J. Zeh, W.E. Grizzle, A. Lomakin, and A.E. Lokshin, *Prediagnostic serum biomarkers as early detection tools for pancreatic cancer in a large prospective cohort study.* PloS one, 2014. **9**(4): p. e94928.

362. Wingren, C., A. Sandström, R. Segersvärd, A. Carlsson, R. Andersson, M. Löhr, and C.A. Borrebaeck, *Identification of serum biomarker signatures associated with pancreatic cancer.* Cancer research, 2012. **72**(10): p. 2481-2490.

363. Kim, J., W.R. Bamlet, A.L. Oberg, K.G. Chaffee, G. Donahue, X.-J. Cao, S. Chari, B.A. Garcia, G.M. Petersen, and K.S. Zaret, *Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers.* Science translational medicine, 2017. **9**(398): p. eaah5583.

364. Le Large, T.Y., L.L. Meijer, R. Paleckyte, L.N. Boyd, B. Kok, T. Wurdinger, T. Schelfhorst, S.R. Piersma, T.V. Pham, and N.C. van Grieken, *Combined Expression of Plasma Thrombospondin-2 and CA19-9 for Diagnosis of Pancreatic Cancer and Distal Cholangiocarcinoma: A Proteome Approach.* The oncologist, 2020. **25**(4): p. e634-e643.

365. Kim, H., K.N. Kang, Y.S. Shin, Y. Byun, Y. Han, W. Kwon, C.W. Kim, and J.-Y. Jang, *Biomarker panel for the diagnosis of pancreatic ductal adenocarcinoma.* Cancers, 2020. **12**(6): p. 1443.

366. Gluszek, S., J. Matykiewicz, U. Grabowska, M. Chrapek, L. Nawacki, I. Wawrzycka, M. Gluszek-Osuch, and D. Kozieł, *Clinical usefulness of pentraxin 3 (PTX3) as a biomarker of acute pancreatitis and pancreatic cancer.* Medical Studies/Studia Medyczne, 2020. **36**(1): p. 6-13.

367. Zhang, P., M. Zou, X. Wen, F. Gu, J. Li, G. Liu, J. Dong, X. Deng, J. Gao, and X. Li, *Development of serum parameters panels for the early detection of pancreatic cancer.* International journal of cancer, 2014. **134**(11): p. 2646-2655.

368. Abue, M., M. Yokoyama, R. Shibuya, K. Tamai, K. Yamaguchi, I. Sato, N. Tanaka, S. Hamada, T. Shimosegawa, and K. Sugamura, *Circulating miR-483-3p and miR-*

21 is highly expressed in plasma of pancreatic cancer. International journal of oncology, 2015. **46**(2): p. 539-547.

369. Stroese, A.J., H. Ullerich, G. Koehler, V. Raetzel, N. Senninger, and S.A. Dhayat, *Circulating microRNA-99 family as liquid biopsy marker in pancreatic adenocarcinoma.* Journal of cancer research and clinical oncology, 2018. **144**(12): p. 2377-2390.
370. Vila-Navarro, E., S. Duran-Sanchon, M. Vila-Casadesús, L. Moreira, À. Ginès, M. Cuatrecasas, J.J. Lozano, L. Bujanda, A. Castells, and M. Gironella, *Novel circulating miRNA signatures for early detection of pancreatic neoplasia.* Clinical and translational gastroenterology, 2019. **10**(4).
371. Shao, H., Y. Zhang, J. Yan, X. Ban, X. Fan, X. Chang, Z. Lu, Y. Wu, L. Zong, and S. Mo, *Upregulated microRNA-483-3p is an early event in pancreatic ductal adenocarcinoma (PDAC) and as a powerful liquid biopsy biomarker in PDAC.* OncoTargets and therapy, 2021. **14**: p. 2163.
372. Johansen, J.S., D. Calatayud, V. Albieri, N.A. Schultz, C. Dehlendorff, J. Werner, B.V. Jensen, P. Pfeiffer, S.E. Bojesen, and N. Giese, *The potential diagnostic value of serum microRNA signature in patients with pancreatic cancer.* International journal of cancer, 2016. **139**(10): p. 2312-2324.
373. Martens, S., P. Lefesvre, R. Nicolle, A.V. Biankin, F. Puleo, J.-L. Van Laethem, and I. Rooman, *Different shades of pancreatic ductal adenocarcinoma, different paths towards precision therapeutic applications.* Annals of Oncology, 2019. **30**(9): p. 1428-1436.
374. MacLean, E., T. Broger, S. Yerlikaya, B.L. Fernandez-Carballo, M. Pai, and C.M. Denkinger, *A systematic review of biomarkers to detect active tuberculosis.* Nature microbiology, 2019. **4**(5): p. 748-758.
375. Siegel, R.L., K.D. Miller, H.E. Fuchs, and A. Jemal, *Cancer statistics, 2022.* CA: a cancer journal for clinicians, 2022. **72**(1): p. 7-33.
376. Hasan, A., K. Visrodia, J.J. Farrell, and T.A. Gonda, *Overview and comparison of guidelines for management of pancreatic cystic neoplasms.* World Journal of Gastroenterology, 2019. **25**(31): p. 4405.
377. Marchegiani, G., R. Salvia, A. Stefano, B. Alberto, P. Tommaso, C. Andrea, M. Laura, Z.C. Costanza, B. Claudio, and A.H. Mohammed, *Guidelines on pancreatic*

cystic neoplasms: major inconsistencies with available evidence and clinical practice—results from an international survey. Gastroenterology, 2021. **160**(7): p. 2234-2238.

378. Gaddam, S., S.G. Phillip, J.W. Keach, D. Mullady, N. Fukami, S.A. Edmundowicz, R.R. Azar, R.J. Shah, F.M. Murad, and V.M. Kushnir, *Suboptimal accuracy of carcinoembryonic antigen in differentiation of mucinous and nonmucinous pancreatic cysts: results of a large multicenter study.* Gastrointestinal endoscopy, 2015. **82**(6): p. 1060-1069.
379. Kane, L.E., G.S. Mellotte, E. Mylod, R.M. O'Brien, F. O'Connell, C.E. Buckley, J. Arlow, K. Nguyen, D. Mockler, and A.D. Meade, *Diagnostic accuracy of blood-based biomarkers for pancreatic cancer: A systematic review and meta-analysis.* Cancer Research Communications, 2022. **2**(10): p. 1229-1243.
380. Alashwal, H., M. El Halaby, J.J. Crouse, A. Abdalla, and A.A. Moustafa, *The application of unsupervised clustering methods to Alzheimer's disease.* Frontiers in computational neuroscience, 2019. **13**: p. 31.
381. Ringnér, M., *What is principal component analysis?* Nature biotechnology, 2008. **26**(3): p. 303-304.
382. Yao, F., J. Coquery, and K.-A. Lê Cao, *Independent principal component analysis for biologically meaningful dimension reduction of large biological data sets.* BMC bioinformatics, 2012. **13**: p. 1-15.
383. Endo, K., H. Weng, N. Kito, Y. Fukushima, and N. Iwai, *MiR-216a and miR-216b as markers for acute phased pancreatic injury.* Biomedical research, 2013. **34**(4): p. 179-188.
384. You, Y., J. Tan, Y. Gong, H. Dai, H. Chen, X. Xu, A. Yang, Y. Zhang, and P. Bie, *MicroRNA-216b-5p functions as a tumor-suppressive RNA by targeting TPT1 in pancreatic cancer cells.* Journal of Cancer, 2017. **8**(14): p. 2854.
385. Wu, X., W. Chen, H. Cai, J. Hu, B. Wu, Y. Jiang, X. Chen, D. Sun, and Y. An, *MiR-216b inhibits pancreatic cancer cell progression and promotes apoptosis by down-regulating KRAS.* Archives of Medical Science, 2018. **14**(6): p. 1321-1332.
386. Felix, T.F., R.M. Lopez Lapa, M. De Carvalho, N. Bertoni, T. Tokar, R.A. Oliveira, M.A. M. Rodrigues, C.N. Hasimoto, W.K. Oliveira, and L. Pelafsky, *MicroRNA*

modulated networks of adaptive and innate immune response in pancreatic ductal adenocarcinoma. PloS one, 2019. **14**(5): p. e0217421.

387. Erener, S., C.E. Ellis, A. Ramzy, M.M. Glavas, S. O'Dwyer, S. Pereira, T. Wang, J. Pang, J.E. Bruin, and M.J. Riedel, *Deletion of pancreas-specific miR-216a reduces beta-cell mass and inhibits pancreatic cancer progression in mice.* Cell Reports Medicine, 2021. **2**(11): p. 100434.

388. Hou, B.-h., Z.-x. Jian, P. Cui, S.-j. Li, R.-q. Tian, and J.-r. Ou, *miR-216a may inhibit pancreatic tumor growth by targeting JAK2.* FEBS letters, 2015. **589**(17): p. 2224-2232.

389. Saha, B., B. Chhatriya, S. Pramanick, and S. Goswami, *Bioinformatic analysis and integration of transcriptome and proteome results identify key coding and noncoding genes predicting malignancy in intraductal papillary mucinous neoplasms of the pancreas.* BioMed Research International, 2021. **2021**.

390. Doz, M., C. Chouaid, L. Com-Ruelle, E. Calvo, M. Brosa, J. Robert, L. Decuypère, C. Pribil, A. Huerta, and B. Detournay, *The association between asthma control, health care costs, and quality of life in France and Spain.* BMC pulmonary medicine, 2013. **13**: p. 1-10.

391. Cui, L., K. Nakano, S. Obchoei, K. Setoguchi, M. Matsumoto, T. Yamamoto, S. Obika, K. Shimada, and N. Hiraoka, *Small nucleolar noncoding RNA SNORA23, up-regulated in human pancreatic ductal adenocarcinoma, regulates expression of spectrin repeat-containing nuclear envelope 2 to promote growth and metastasis of xenograft tumors in mice.* Gastroenterology, 2017. **153**(1): p. 292-306. e2.

392. Xiao, X., B.S. Yeoh, and M. Vijay-Kumar, *Lipocalin 2: an emerging player in iron homeostasis and inflammation.* Annual review of nutrition, 2017. **37**: p. 103-130.

393. Hao, P., J. Zhang, S. Fang, M. Jia, X. Xian, S. Yan, Y. Wang, Q. Ren, F. Yue, and H. Cui, *Lipocalin-2 inhibits pancreatic cancer stemness via the AKT/c-Jun pathway.* Human Cell, 2022. **35**(5): p. 1475-1486.

394. Bartsch, D.K., N. Gercke, K. Strauch, R. Wieboldt, E. Matthäi, V. Wagner, S. Rospleszcz, A. Schäfer, F.S. Franke, and I. Mintziras, *The combination of MiRNA-196b, LCN2, and TIMP1 is a potential set of circulating biomarkers for screening individuals at risk for familial pancreatic cancer.* Journal of clinical medicine, 2018. **7**(10): p. 295.

395. Moniaux, N., S. Chakraborty, M. Yalniz, J. Gonzalez, V.K. Shostrom, J. Standop, S.M. Lele, M. Ouellette, P.M. Pour, and A. Sasson, *Early diagnosis of pancreatic cancer: neutrophil gelatinase-associated lipocalin as a marker of pancreatic intraepithelial neoplasia.* British journal of cancer, 2008. **98**(9): p. 1540-1547.

396. Chen, R., D. Crispin, S. Pan, S. Hawley, M.W. McIntosh, D. May, H. Anton-Culver, A. Ziogas, M.P. Bronner, and T.A. Brentnall, *Pilot study of blood biomarker candidates for detection of pancreatic cancer.* Pancreas, 2010. **39**(7): p. 981.

397. Xu, B., D.-Y. Jin, W.-H. Lou, and D.-S. Wang, *Lipocalin-2 is associated with a good prognosis and reversing epithelial-to-mesenchymal transition in pancreatic cancer.* World journal of surgery, 2013. **37**: p. 1892-1900.

398. Tong, Z., A.B. Kunnumakkara, H. Wang, Y. Matsuo, P. Diagaradjane, K.B. Harikumar, V. Ramachandran, B. Sung, A. Chakraborty, and R.S. Bresalier, *Neutrophil gelatinase–associated lipocalin: a novel suppressor of invasion and angiogenesis in pancreatic cancer.* Cancer research, 2008. **68**(15): p. 6100-6108.

399. Li, Q., H. Wang, G. Zogopoulos, Q. Shao, K. Dong, F. Lv, K. Nwilati, X.-y. Gui, A. Cuggia, and J.-L. Liu, *Reg proteins promote acinar-to-ductal metaplasia and act as novel diagnostic and prognostic markers in pancreatic ductal adenocarcinoma.* Oncotarget, 2016. **7**(47): p. 77838.

400. Radon, T.P., N.J. Massat, R. Jones, W. Alrawashdeh, L. Dumartin, D. Ennis, S.W. Duffy, H.M. Kocher, S.P. Pereira, and L. Guarner, *Identification of a three-biomarker panel in urine for early detection of pancreatic adenocarcinoma.* Clinical Cancer Research, 2015. **21**(15): p. 3512-3521.

401. Zhang, Y., X. Yuan, X. Zhu, Q. Wang, X. Yu, Q. Wei, and L. Li, *Serum REG Iα as a potential novel biomarker in cancer: an observational study.* Medicine, 2020. **99**(38).

402. Shen, J., M.D. Person, J. Zhu, J.L. Abbruzzese, and D. Li, *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 research, 2004. **64**(24): p. 9018-9026.

403. Zhou, L., R. Zhang, L. Wang, S. Shen, H. Okamoto, A. Sugawara, L. Xia, X. Wang, N. Noguchi, and T. Yoshikawa, *Upregulation of REG Iα accelerates tumor*

progression in pancreatic cancer with diabetes. International journal of cancer, 2010. **127**(8): p. 1795-1803.

404. Pan, S., R. Chen, Y. Tamura, D.A. Crispin, L.A. Lai, D.H. May, M.W. McIntosh, D.R. Goodlett, and T.A. Brentnall, *Quantitative glycoproteomics analysis reveals changes in N-glycosylation level associated with pancreatic ductal adenocarcinoma.* Journal of proteome research, 2014. **13**(3): p. 1293-1306.
405. Berberat, P.O., H. Friess, L. Wang, Z. Zhu, T. Bley, L. Frigeri, A. Zimmermann, and M.W. Büchler, *Comparative analysis of galectins in primary tumors and tumor metastasis in human pancreatic cancer.* Journal of Histochemistry & Cytochemistry, 2001. **49**(4): p. 539-549.
406. Emberley, E.D., L.C. Murphy, and P.H. Watson, *S100 proteins and their influence on pro-survival pathways in cancer.* Biochemistry and Cell Biology, 2004. **82**(4): p. 508-515.
407. Cross, S.S., F.C. Hamdy, J.-C. Deloulme, and I. Rehman, *Expression of S100 proteins in normal human tissues and common cancers using tissue microarrays: S100A6, S100A8, S100A9 and S100A11 are all overexpressed in common cancers.* Histopathology, 2005. **46**(3): p. 256-269.
408. Chen, K.T., P.D. Kim, K.A. Jones, K. Devarajan, B.B. Patel, J.P. Hoffman, H. Ehya, M. Huang, J.C. Watson, and J.L. Tokar, *Potential prognostic biomarkers of pancreatic cancer.* Pancreas, 2014. **43**(1).
409. Sheikh, A., C. Ang, S. Tonack, E. Tweedle, I. Schwarte-Waldhoff, J. Neoptolemos, and E. Costello, *S100A8 and S100A9 increase pancreatic and colorectal cancer cell motility and proliferation.* Pancreas, 2008. **37**(4): p. 496.
410. Nedjadi, T., A. Evans, A. Sheikh, L. Barerra, S. Al-Ghamdi, L. Oldfield, W. Greenhalf, J.P. Neoptolemos, and E. Costello, *S100A8 and S100A9 proteins form part of a paracrine feedback loop between pancreatic cancer cells and monocytes.* Bmc Cancer, 2018. **18**: p. 1-8.
411. Sogawa, K., S. Takano, F. Iida, M. Satoh, S. Tsuchida, Y. Kawashima, H. Yoshitomi, A. Sanda, Y. Kodera, and H. Takizawa, *Identification of a novel serum biomarker for pancreatic cancer, C4b-binding protein α-chain (C4BPA) by quantitative proteomic analysis using tandem mass tags.* British journal of cancer, 2016. **115**(8): p. 949-956.

412. Ohkuma, R., E. Yada, S. Ishikawa, D. Komura, Y. Kubota, K. Hamada, A. Horiike, T. Ishiguro, Y. Hirasawa, and H. Ariizumi, *High expression levels of polymeric immunoglobulin receptor are correlated with chemoresistance and poor prognosis in pancreatic cancer.* Oncology Reports, 2020. **44**(1): p. 252-262.

413. Makawita, S., C. Smith, I. Batruch, Y. Zheng, F. Rückert, R. Grützmann, C. Pilarsky, S. Gallinger, and E.P. Diamandis, *Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers.* Molecular & cellular proteomics, 2011. **10**(10).

414. Xu, D.F., L.S. Wang, and J.H. Zhou, *Long non-coding RNA CASC2 suppresses pancreatic cancer cell growth and progression by regulating the miR-24/MUC6 axis.* International journal of oncology, 2020. **56**(2): p. 494-507.

415. Sierzega, M., D. Młynarski, R. Tomaszewska, and J. Kulig, *Semiquantitative immunohistochemistry for mucin (MUC1, MUC2, MUC3, MUC4, MUC5AC, and MUC6) profiling of pancreatic ductal cell adenocarcinoma improves diagnostic and prognostic performance.* Histopathology, 2016. **69**(4): p. 582-591.

416. Remmers, N., J.M. Anderson, E.M. Linde, D.J. DiMaio, A.J. Lazenby, H.H. Wandall, U. Mandel, H. Clausen, F. Yu, and M.A. Hollingsworth, *Aberrant Expression of Mucin Core Proteins and O-Linked Glycans Associated with Progression of Pancreatic CancerMucin and O-Linked Glycosylation in Pancreatic Cancer Progression.* Clinical Cancer Research, 2013. **19**(8): p. 1981-1993.

417. Kim, G.E., H.I. Bae, H.U. Park, S.F. Kuan, S.C. Crawley, J.J. Ho, and Y.S. Kim, *Aberrant expression of MUC5AC and MUC6 gastric mucins and sialyl Tn antigen in intraepithelial neoplasms of the pancreas.* Gastroenterology, 2002. **123**(4): p. 1052-1060.

418. Ohya, A., K. Yamanoi, H. Shimojo, C. Fujii, and J. Nakayama, *Gastric gland mucin-specific O-glycan expression decreases with tumor progression from precursor lesions to pancreatic cancer.* Cancer science, 2017. **108**(9): p. 1897-1902.

419. Tamir, A., A. Gangadharan, S. Balwani, T. Tanaka, U. Patel, A. Hassan, S. Benke, A. Agas, J. D'Agostino, and D. Shin, *The serine protease prostasin (PRSS8) is a potential biomarker for early detection of ovarian cancer.* Journal of ovarian research, 2016. **9**(1): p. 1-13.

420. Liu, G.-j., Y.-j. Wang, M. Yue, L.-m. Zhao, Y.-D. Guo, Y.-p. Liu, H.-c. Yang, F. Liu, X. Zhang, and L.-h. Zhi, *High expression of TCN1 is a negative prognostic biomarker and can predict neoadjuvant chemosensitivity of colon cancer.* Scientific reports, 2020. **10**(1): p. 11951.

421. Varshney, N., A.A. Kebede, H. Owusu-Dapaah, J. Lather, M. Kaushik, and J.S. Bhullar, *A review of Von Hippel-Lindau syndrome.* Journal of Kidney Cancer and VHL, 2017. **4**(3): p. 20.

422. Shan, J., Z. Sun, J. Yang, J. Xu, W. Shi, Y. Wu, Y. Fan, and H. Li, *Discovery and preclinical validation of proteomic biomarkers in saliva for early detection of oral squamous cell carcinomas.* Oral diseases, 2019. **25**(1): p. 97-107.

423. Do, M., H. Kim, D. Shin, J. Park, H. Kim, Y. Han, J.-Y. Jang, and Y. Kim, *Marker identification of the grade of dysplasia of intraductal papillary mucinous neoplasm in pancreatic cyst fluid by quantitative proteomic profiling.* Cancers, 2020. **12**(9): p. 2383.

424. Teng, P.-n., N.W. Bateman, B.L. Hood, and T.P. Conrads, *Advances in proximal fluid proteomics for disease biomarker discovery.* Journal of proteome research, 2010. **9**(12): p. 6091-6100.

425. Kristjansdottir, B., K. Partheen, E.T. Fung, J. Marcickiewicz, C. Yip, M. Brännström, and K. Sundfeldt, *Ovarian cyst fluid is a rich proteome resource for detection of new tumor biomarkers.* Clinical proteomics, 2012. **9**(1): p. 1-9.

426. Garcia-Obregon, S., M. Azkargorta, I. Seijas, J. Pilar-Orive, F. Borrego, F. Elortza, M.D. Boyano, and I. Astigarraga, *Identification of a panel of serum protein markers in early stage of sepsis and its validation in a cohort of patients.* Journal of Microbiology, Immunology and Infection, 2018. **51**(4): p. 465-472.

427. Wang, D.-d., X. Chen, D.-d. Yu, S.-j. Yang, H.-Y. Shen, S.-l. Zhong, J.-h. Zhao, and J.-h. Tang, *miR-197: A novel biomarker for cancers.* Gene, 2016. **591**(2): p. 313-319.

428. Urabe, F., J. Matsuzaki, Y. Yamamoto, T. Kimura, T. Hara, M. Ichikawa, S. Takizawa, Y. Aoki, S. Niida, and H. Sakamoto, *Large-scale Circulating microRNA Profiling for the Liquid Biopsy of Prostate CancerSerum miRNA Biomarker for Prostate Cancer.* Clinical Cancer Research, 2019. **25**(10): p. 3016-3025.

429. Akula, S.M., J.F. Williams, L.R. Pokhrel, A.N. Bauer, S. Rajput, and P.P. Cook, *Cellular miR-6741-5p as a Prognostic Biomarker Predicting Length of Hospital Stay among COVID-19 Patients.* Viruses, 2022. **14**(12): p. 2681.

430. Chen, J., X. Wu, B. Cai, Z. Su, L. Li, Y. An, and L. Wang, *Circulating microRNAs as potential biomarkers of HBV infection persistence.* Infection, Genetics and Evolution, 2017. **54**: p. 152-157.

431. Jin, L. and Z. Zhang, *Serum miR-3180-3p and miR-124-3p may Function as Noninvasive Biomarkers of Cisplatin Resistance in Gastric Cancer.* Clinical Laboratory, 2020. **66**(12).

432. Bustos, M.A., N. Rahimzadeh, S. Ryu, R. Gross, L.T. Tran, V.M. Renteria-Lopez, R.I. Ramos, A. Eisenberg, P. Hothi, and S. Kesari, *Cell-free plasma microRNAs that identify patients with glioblastoma.* Laboratory Investigation, 2022. **102**(7): p. 711-721.

433. Mørup, N., R. Stakaitis, I. Golubickaite, M. Riera, M.D. Dalgaard, M.H. Schierup, N. Jørgensen, G. Daugaard, A. Juul, and K. Almstrup, *Small RNAs in seminal plasma as novel biomarkers for germ cell tumors.* Cancers, 2021. **13**(10): p. 2346.

434. Ren, Z.-j., Y. Zhao, G. Wang, Z.-c. Zhang, L. Ma, M.-z. Teng, and Y.-m. Li, *Identification of differentially expressed miRNAs derived from serum exosomes associated with gastric cancer by microarray analysis.* Clinica Chimica Acta, 2022. **531**: p. 25-35.

435. Pandey, R., H.-H. Woo, F. Varghese, M. Zhou, and S.K. Chambers, *Circulating miRNA profiling of women at high risk for ovarian cancer.* Translational Oncology, 2019. **12**(5): p. 714-725.

436. Paulo, J.A., V. Kadiyala, L.S. Lee, P.A. Banks, D.L. Conwell, and H. Steen, *Proteomic analysis (GeLC–MS/MS) of ePFT-collected pancreatic fluid in chronic pancreatitis.* Journal of proteome research, 2012. **11**(3): p. 1897-1912.

437. Paulo, J.A., L.S. Lee, P.A. Banks, H. Steen, and D.L. Conwell, *Proteomic analysis of formalin-fixed paraffin-embedded pancreatic tissue using liquid chromatography tandem mass spectrometry (LC-MS/MS).* Pancreas, 2012. **41**(2): p. 175.

438. Matsubara, J., K. Honda, M. Ono, Y. Tanaka, M. Kobayashi, G. Jung, K. Yanagisawa, T. Sakuma, S. Nakamori, and N. Sata, *Reduced Plasma Level of CXC*

Chemokine Ligand 7 in Patients with Pancreatic Cancer. Cancer epidemiology, biomarkers & prevention, 2011. **20**(1): p. 160-171.

439. Pan, S., R. Chen, D.A. Crispin, D. May, T. Stevens, M.W. McIntosh, M.P. Bronner, A. Ziogas, H. Anton-Culver, and T.A. Brentnall, *Protein alterations associated with pancreatic cancer and chronic pancreatitis found in human plasma using global quantitative proteomics profiling.* Journal of proteome research, 2011. **10**(5): p. 2359-2376.

440. Kim, Y., I. Yeo, I. Huh, J. Kim, D. Han, J.-Y. Jang, and Y. Kim, *Development and multiple validation of the protein multi-marker panel for diagnosis of pancreatic Cancer.* Clinical Cancer Research, 2021. **27**(8): p. 2236-2245.

441. Chen, R., C.Y. Eugene, S. Donohoe, S. Pan, J. Eng, K. Cooke, D.A. Crispin, Z. Lane, D.R. Goodlett, and M.P. Bronner, *Pancreatic cancer proteome: the proteins that underlie invasion, metastasis, and immunologic escape.* Gastroenterology, 2005. **129**(4): p. 1187-1197.

442. Holm, M., M. Saraswat, S. Joenväärä, H. Seppänen, R. Renkonen, and C. Haglund, *Label-free proteomics reveals serum proteins whose levels differ between pancreatic ductal adenocarcinoma patients with short or long survival.* Tumor Biology, 2020. **42**(6): p. 1010428320936410.

443. Li, X., R. Ni, J. Chen, Z. Liu, M. Xiao, F. Jiang, and C. Lu, *The presence of IGHG1 in human pancreatic carcinomas is associated with immune evasion mechanisms.* Pancreas, 2011. **40**(5): p. 753-761.

444. Chen, R., T.A. Brentnall, S. Pan, K. Cooke, K.W. Moyes, Z. Lane, D.A. Crispin, D.R. Goodlett, R. Aebersold, and M.P. Bronner, *Quantitative proteomics analysis reveals that proteins differentially expressed in chronic pancreatitis are also frequently involved in pancreatic cancer.* Molecular & Cellular Proteomics, 2007. **6**(8): p. 1331-1342.

445. Li, Y., P. Wang, D. Ye, X. Bai, X. Zeng, Q. Zhao, and Z. Zhang, *IGHG1 induces EMT in gastric cancer cells by regulating TGF-β/SMAD3 signaling pathway.* Journal of Cancer, 2021. **12**(12): p. 3458.

446. Hu, Y., Q. Zeng, C. Li, and Y. Xie, *Expression profile and prognostic value of SFN in human ovarian cancer.* Bioscience reports, 2019. **39**(5).

447. Robin, F., G. Angenard, L. Cano, L. Courtin-Tanguy, E. Gaignard, Z.-E. Khene, D. Bergeat, B. Clément, K. Boudjema, and C. Coulouarn, *Molecular profiling of stroma highlights stratifin as a novel biomarker of poor prognosis in pancreatic ductal adenocarcinoma.* British Journal of Cancer, 2020. **123**(1): p. 72-80.

448. Mogal, M.R., A. Junayed, M.R. Mahmod, S.A. Sompa, S.A. Lima, N. Kar, M. Khatun, M.A. Zubair, and M.A. Sikder, *A computational approach to justifying stratifin as a candidate diagnostic and prognostic biomarker for Pancreatic Cancer.* BioMed Research International, 2022. **2022**.

449. Khalili, H., A. Sull, S. Sarin, F.J. Boivin, R. Halabi, B. Svajger, A. Li, V.W. Cui, T. Drysdale, and D. Bridgewater, *Developmental origins for kidney disease due to Shroom3 deficiency.* Journal of the American Society of Nephrology, 2016. **27**(10): p. 2965-2973.

450. Cañadas-Garre, M., K. Anderson, J. McGoldrick, A. Maxwell, and A. McKnight, *Genomic approaches in the search for molecular biomarkers in chronic kidney disease.* Journal of translational medicine, 2018. **16**(1): p. 1-44.

451. Bao, M., Y. Huang, Z. Lang, H. Zhao, Y. Saito, T. Nagano, I. Kawagoe, D. Divisi, X. Hu, and G. Jiang, *Proteomic analysis of plasma exosomes in patients with non-small cell lung cancer.* Translational Lung Cancer Research, 2022. **11**(7): p. 1434.

452. Munigala, S., A. Gelrud, and B. Agarwal, *Risk of pancreatic cancer in patients with pancreatic cyst.* Gastrointestinal endoscopy, 2016. **84**(1): p. 81-86.

453. Puiffe, M.-L., C. Le Page, A. Filali-Mouhim, M. Zietarska, V. Ouellet, P.N. Tonin, M. Chevrette, D.M. Provencher, and A.-M. Mes-Masson, *Characterization of ovarian cancer ascites on cell invasion, proliferation, spheroid formation, gene expression in an in vitro model of epithelial ovarian cancer.* Neoplasia, 2007. **9**(10): p. 820-IN8.

454. Carduner, L., J. Leroy-Dudal, C. Picot, O. Gallet, F. Carreiras, and S. Kellouche, *Ascites-induced shift along epithelial-mesenchymal spectrum in ovarian cancer cells: Enhancement of their invasive behavior partly dependant on αv integrins.* Clinical & experimental metastasis, 2014. **31**(6): p. 675-688.

455. Chen, M., J. Huang, X. Yang, B. Liu, W. Zhang, L. Huang, F. Deng, J. Ma, Y. Bai, and R. Lu, *Serum starvation induced cell cycle synchronization facilitates human somatic cells reprogramming.* PloS one, 2012. **7**(4): p. e28203.

456. Del Chiaro, M., Z. Ateeb, M.R. Hansson, E. Rangelova, R. Segersvärd, N. Kartalis, C. Ansorge, M.J. Löhr, U. Arnelo, and C. Verbeke, *Survival analysis and risk for progression of intraductal papillary mucinous neoplasia of the pancreas (IPMN) under surveillance: a single-institution experience.* Annals of surgical oncology, 2017. **24**(4): p. 1120-1126.

457. Chen, Z., Z. Fang, and J. Ma, *Regulatory mechanisms and clinical significance of vimentin in breast cancer.* Biomedicine & Pharmacotherapy, 2021. **133**: p. 111068.

458. Walsh, N., N. O'Donovan, S. Kennedy, M. Henry, P. Meleady, M. Clynes, and P. Dowling, *Identification of pancreatic cancer invasion-related proteins by proteomic analysis.* Proteome science, 2009. **7**(1): p. 1-14.

459. Satelli, A. and S. Li, *Vimentin in cancer and its potential as a molecular target for cancer therapy.* Cellular and molecular life sciences, 2011. **68**(18): p. 3033-3046.

460. Wei, T., X. Zhang, Q. Zhang, J. Yang, Q. Chen, J. Wang, X. Li, J. Chen, T. Ma, and G. Li, *Vimentin-positive circulating tumor cells as a biomarker for diagnosis and treatment monitoring in patients with pancreatic cancer.* Cancer Letters, 2019. **452**: p. 237-243.

461. Handra-Luca, A., S. Hong, K. Walter, C. Wolfgang, R. Hruban, and M. Goggins, *Tumour epithelial vimentin expression and outcome of pancreatic ductal adenocarcinomas.* British journal of cancer, 2011. **104**(8): p. 1296-1302.

462. Zhou, B., J. Xiang, M. Jin, X. Zheng, G. Li, and S. Yan, *High vimentin expression with E-cadherin expression loss predicts a poor prognosis after resection of grade 1 and 2 pancreatic neuroendocrine tumors.* BMC cancer, 2021. **21**(1): p. 1-10.

463. Wang, W., L. Dong, B. Zhao, J. Lu, and Y. Zhao, *E-cadherin is downregulated by microenvironmental changes in pancreatic cancer and induces EMT.* Oncology Reports, 2018. **40**(3): p. 1641-1649.

464. Weisberg, S.P., D.J. Carpenter, M. Chait, P. Dogra, R.D. Gartrell-Corrado, A.X. Chen, S. Campbell, W. Liu, P. Saraf, and M.E. Snyder, *Tissue-resident memory T cells mediate immune homeostasis in the human pancreas through the PD-1/PD-L1 pathway.* Cell reports, 2019. **29**(12): p. 3916-3932. e5.

465. Zhao, Y., F. Shi, Q. Zhou, Y. Li, J. Wu, R. Wang, and Q. Song, *Prognostic significance of PD-L1 in advanced non-small cell lung carcinoma.* Medicine, 2020. **99**(45).

466. Wang, L., *Prognostic effect of programmed death-ligand 1 (PD-L1) in ovarian cancer: a systematic review, meta-analysis and bioinformatics study.* Journal of ovarian research, 2019. **12**(1): p. 1-10.

467. Li, S., L. Chen, and J. Jiang, *Role of programmed cell death ligand-1 expression on prognostic and overall survival of breast cancer: a systematic review and meta-analysis.* Medicine, 2019. **98**(16).

468. Carlsson, J., P. Sundqvist, V. Kosuta, A. Fält, F. Giunchi, M. Fiorentino, and S. Davidsson, *PD-L1 expression is associated with poor prognosis in renal cell carcinoma.* Applied Immunohistochemistry & Molecular Morphology, 2020. **28**(3): p. 213-220.

469. Troiani, T., E. Martinelli, A. Capasso, F. Morgillo, M. Orditura, F. De Vita, and F. Ciardiello, *Targeting EGFR in pancreatic cancer treatment.* Current drug targets, 2012. **13**(6): p. 802-810.

470. Cohenuram, M. and M.W. Saif, *Epidermal growth factor receptor inhibition strategies in pancreatic cancer: past, present and the future.* Jop, 2007. **8**(1): p. 4-15.

471. Anastassiou, D., V. Rumjantseva, W. Cheng, J. Huang, P.D. Canoll, D.J. Yamashiro, and J.J. Kandel, *Human cancer cells express Slug-based epithelial-mesenchymal transition gene expression signature obtained in vivo.* BMC cancer, 2011. **11**(1): p. 1-9.

472. Bilal, F., E.J. Arenas, K. Pedersen, A. Martínez-Sabadell, B. Nabet, E. Guruceaga, S. Vicent, J. Tabernero, T. Macarulla, and J. Arribas, *The Transcription Factor SLUG Uncouples Pancreatic Cancer Progression from the RAF–MEK1/2–ERK1/2 Pathway.* Cancer Research, 2021. **81**(14): p. 3849-3861.

473. Liu, T., X. Zhang, M. Shang, Y. Zhang, B. Xia, M. Niu, Y. Liu, and D. Pang, *Dysregulated expression of Slug, vimentin, and E-cadherin correlates with poor clinical outcome in patients with basal-like breast cancer.* Journal of surgical oncology, 2013. **107**(2): p. 188-194.

474. Shields, M.A., S. Dangi-Garimella, S.B. Krantz, D.J. Bentrem, and H.G. Munshi, *Pancreatic cancer cells respond to type I collagen by inducing snail expression to promote membrane type 1 matrix metalloproteinase-dependent collagen invasion.* Journal of Biological Chemistry, 2011. **286**(12): p. 10495-10504.

475. Duan, W., J. Ma, Q. Ma, Q. Xu, J. Lei, L. Han, X. Li, Z. Wang, Z. Wu, and S. Lv, *The activation of β1-integrin by type I collagen coupling with the hedgehog pathway promotes the epithelial-mesenchymal transition in pancreatic cancer.* Current cancer drug targets, 2014. **14**(5): p. 446-457.

476. Zheng, J., *Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation.* Oncology letters, 2012. **4**(6): p. 1151-1157.

477. White, E.Z., N.M. Pennant, J.R. Carter, O. Hawsawi, V. Odero-Marah, and C.V. Hinton, *Serum deprivation initiates adaptation and survival to oxidative stress in prostate cancer cells.* Scientific Reports, 2020. **10**(1): p. 1-18.

478. Wu, C.-A., Y. Chao, S.-G. Shiah, and W.-W. Lin, *Nutrient deprivation induces the Warburg effect through ROS/AMPK-dependent activation of pyruvate dehydrogenase kinase.* Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2013. **1833**(5): p. 1147-1156.

479. Sabanovic, B., M. Giulietti, M. Cecati, G. Spolverato, C. Benna, S. Pucciarelli, and F. Piva, *Effects of the Exposure of Human Non-Tumour Cells to Sera of Pancreatic Cancer Patients.* Biomedicines, 2022. **10**(10): p. 2588.

480. Fitzgerald, T.L., K. Lertpiriyapong, L. Cocco, A.M. Martelli, M. Libra, S. Candido, G. Montalto, M. Cervello, L. Steelman, and S.L. Abrams, *Roles of EGFR and KRAS and their downstream signaling pathways in pancreatic cancer and pancreatic cancer stem cells.* Advances in biological regulation, 2015. **59**: p. 65-81.

481. Wang, T., J. Yang, J. Xu, J. Li, Z. Cao, L. Zhou, L. You, H. Shu, Z. Lu, and H. Li, *CHIP is a novel tumor suppressor in pancreatic cancer and inhibits tumor growth through targeting EGFR.* Oncotarget, 2014. **5**(7): p. 1969.

482. Humtsoe, J.O. and R.H. Kramer, *Differential epidermal growth factor receptor signaling regulates anchorage-independent growth by modulation of the PI3K/AKT pathway.* Oncogene, 2010. **29**(8): p. 1214-1226.

483. Wang, T., M. Zhang, Z. Ma, K. Guo, V. Tergaonkar, Q. Zeng, and W. Hong, *A role of Rab7 in stabilizing EGFR-Her2 and in sustaining Akt survival signal.* Journal of cellular physiology, 2012. **227**(6): p. 2788-2797.

484. Mo, L., V. Pospichalova, Z. Huang, S.K. Murphy, S. Payne, F. Wang, M. Kennedy, G.J. Cianciolo, V. Bryja, and S.V. Pizzo, *Ascites increases expression/function of*

multidrug resistance proteins in ovarian cancer cells. PloS one, 2015. **10**(7): p. e0131579.

485. Andre, F., L.M. McShane, S. Michiels, D.F. Ransohoff, D.G. Altman, J.S. Reis-Filho, D.F. Hayes, and L. Pusztai, *Biomarker studies: a call for a comprehensive biomarker study registry.* Nature reviews Clinical oncology, 2011. **8**(3): p. 171-176.

486. Li, F., Y. Zhou, Y. Zhang, J. Yin, Y. Qiu, J. Gao, and F. Zhu, *POSREG: proteomic signature discovered by simultaneously optimizing its reproducibility and generalizability.* Briefings in Bioinformatics, 2022. **23**(2).

487. West-Nielsen, M., E.V. Høgdall, E. Marchiori, C.K. Høgdall, C. Schou, and N.H. Heegaard, *Sample handling for mass spectrometric proteomic investigations of human sera.* Analytical chemistry, 2005. **77**(16): p. 5114-5123.

488. Diamandis, E.P., *Mass spectrometry as a diagnostic and a cancer biomarker discovery tool: opportunities and potential limitations.* Molecular & Cellular Proteomics, 2004. **3**(4): p. 367-378.

489. Subramanian, J. and R. Simon, *Overfitting in prediction models–is it a problem only in high dimensions?* Contemporary clinical trials, 2013. **36**(2): p. 636-641.

490. Wong, T.-T., *Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation.* Pattern Recognition, 2015. **48**(9): p. 2839-2846.

491. Pencina, M.J., R.B. D'Agostino, and R.S. Vasan, *Statistical methods for assessment of added usefulness of new biomarkers.* Clinical chemistry and laboratory medicine, 2010. **48**(12): p. 1703-1711.

492. Kampfrath, T. and S.S. Levinson, *Brief critical review: statistical assessment of biomarker performance.* Clinica Chimica Acta, 2013. **419**: p. 102-107.

493. Fraunhoffer, N.A., A.M. Abuelafia, M. Bigonnet, O. Gayet, J. Roques, R. Nicolle, G. Lomberk, R. Urrutia, N. Dusetti, and J. Iovanna, *Multi-omics data integration and modeling unravels new mechanisms for pancreatic cancer and improves prognostic prediction.* NPJ precision oncology, 2022. **6**(1): p. 57.

494. Govaert, E., K. Van Steendam, S. Willems, L. Vossaert, M. Dhaenens, and D. Deforce, *Comparison of fractionation proteomics for local SWATH library building.* Proteomics, 2017. **17**(15-16): p. 1700052.

495. Langan, T.J., K.R. Rodgers, and R.C. Chou, *Synchronization of mammalian cell cultures by serum deprivation.* Cell Cycle Synchronization: Methods and Protocols, 2017: p. 97-105.

496. Tavaluc, R.T., L.S. Hart, D.T. Dicker, and W.S. El-Deiry, *Effects of low confluency, serum starvation and hypoxia on the side population of cancer cell lines.* Cell cycle, 2007. **6**(20): p. 2554-2562.

497. Levin, V.A., S.C. Panchabhai, L. Shen, S.M. Kornblau, Y. Qiu, and K.A. Baggerly, *Different changes in protein and phosphoprotein levels result from serum starvation of high-grade glioma and adenocarcinoma cell lines.* Journal of proteome research, 2010. **9**(1): p. 179-191.

498. Oya, N., F. Zölzer, F. Werner, and C. Streffer, *Effects of serum starvation on radiosensitivity, proliferation and apoptosis in four human tumor cell lines with different p53 status.* Strahlentherapie und Onkologie, 2003. **179**(2): p. 99.

499. Rashid, M.u. and K.M. Coombs, *Serum-reduced media impacts on cell viability and protein expression in human lung epithelial cells.* Journal of cellular physiology, 2019. **234**(6): p. 7718-7724.

500. Pirkmajer, S. and A.V. Chibalin, *Serum starvation: caveat emptor.* American Journal of Physiology-Cell Physiology, 2011. **301**(2): p. C272-C279.

501. Mortezaee, K., *Enriched cancer stem cells, dense stroma, and cold immunity: Interrelated events in pancreatic cancer.* Journal of Biochemical and Molecular Toxicology, 2021. **35**(4): p. e22708.

502. Boj, S.F., C.-I. Hwang, L.A. Baker, I.I.C. Chio, D.D. Engle, V. Corbo, M. Jager, M. Ponz-Sarvise, H. Tiriac, and M.S. Spector, *Organoid models of human and mouse ductal pancreatic cancer.* Cell, 2015. **160**(1-2): p. 324-338.

Appendices

Appendix 1. PRISMA 2020 Checklist.

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Title
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Methods
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Methods
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Methods

Section and Topic	Item #	Checklist item	Location where item is reported
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Methods
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Methods
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Methods
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Methods
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Methods
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Methods
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Methods
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Methods
	16b	Cite studies that might appear to meet the inclusion criteria, but	Methods

Section and Topic	Item #	Checklist item	Location where item is reported
		which were excluded, and explain why they were excluded.	
Study characteristics	17	Cite each included study and present its characteristics.	*
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	**
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Results
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Results
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Results
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Results
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Results
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion
	23b	Discuss any limitations of the evidence included in the review.	Discussion
	23c	Discuss any limitations of the review processes used.	Discussion
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Methods
	24b	Indicate where the review protocol can be accessed, or state that	Methods

Section and Topic	Item #	Checklist item	Location where item is reported
		a protocol was not prepared.	
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Methods
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Funding
Competing interests	26	Declare any competing interests of review authors.	Competing Interests
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	***

Items reported using *, ** and *** are publicly available online in the supplementary materials of the published manuscript.

Appendix 2. Clinical data binary code key for correlations.

Clinical factors	0	1	2	3
Risk	Low Risk	High Risk		
Sex	Female	Male		
Smoking habits	Non-smoker	Ex-smoker	Active	
Alcohol consumption	None	Previous heavy, now abstinent	Active	Heavy
Diabetes status	No	Yes		
Pancreatitis status	No	Yes		

Categorical clinical data are coded in a logical order such that increasing numbers correspond to increasing risk, where appropriate.

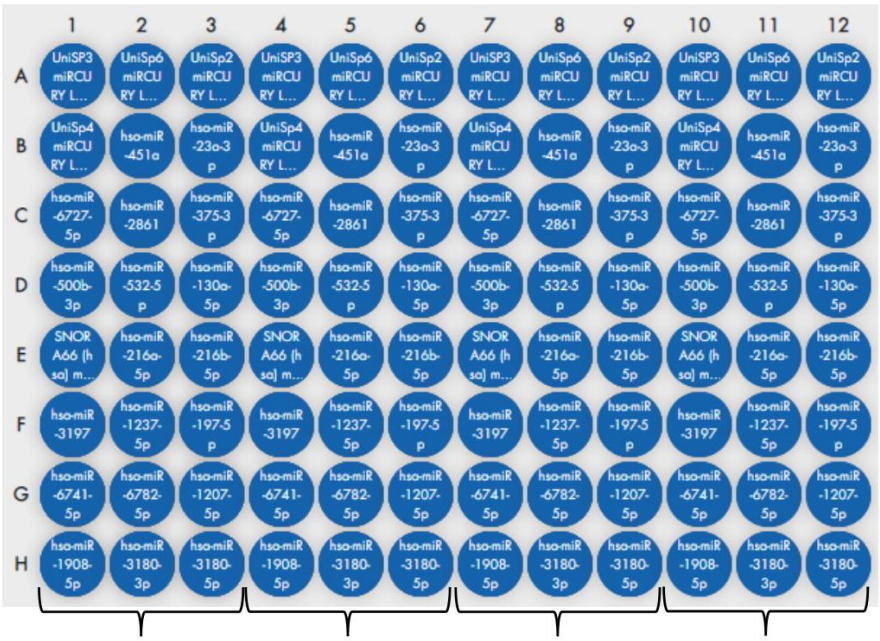
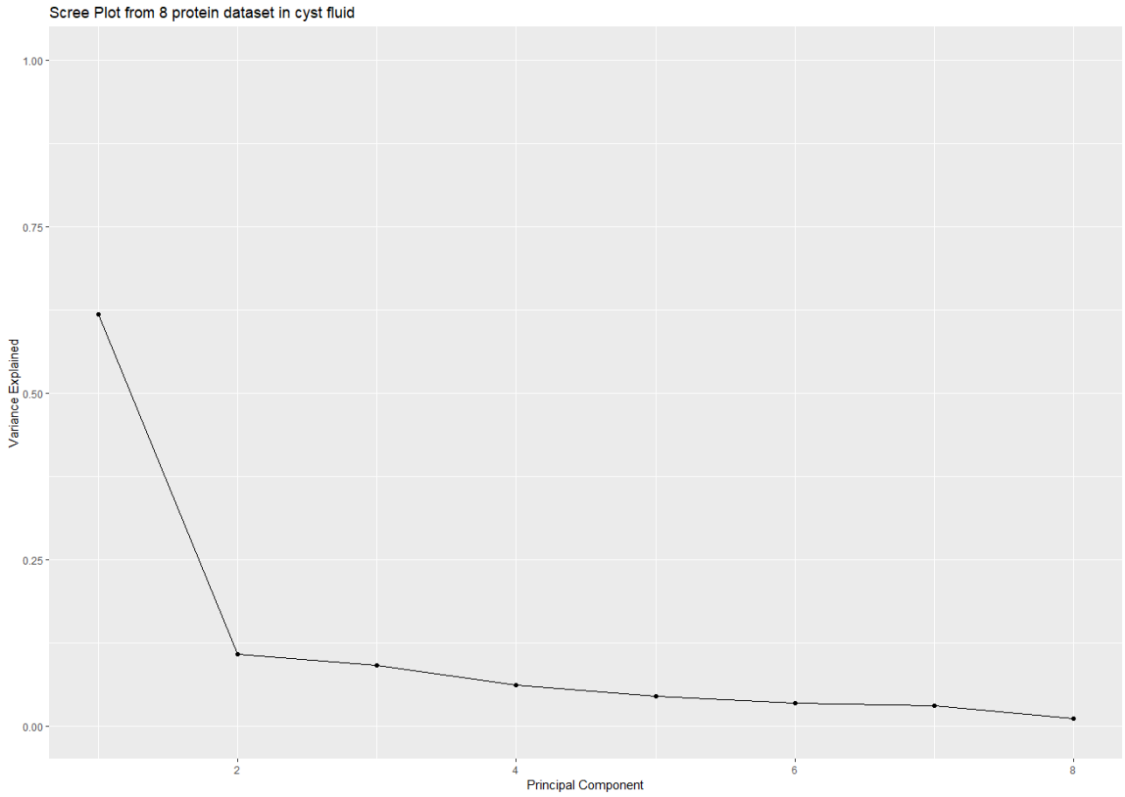
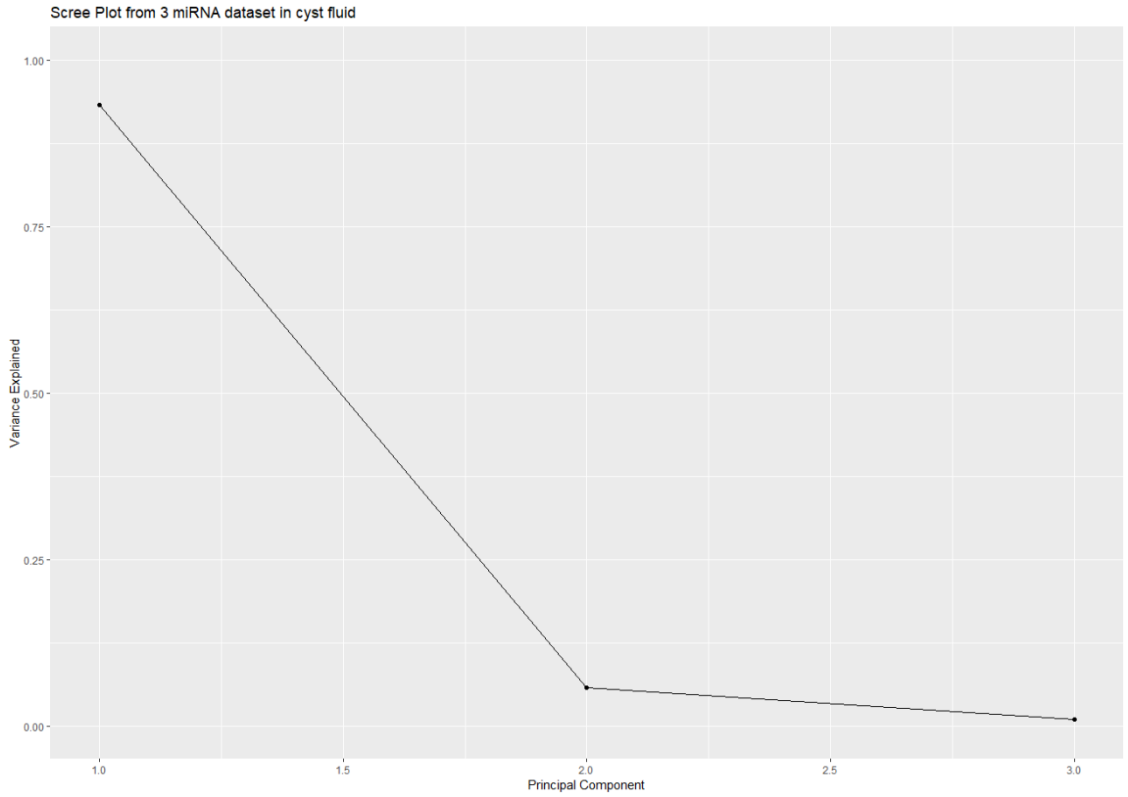


Plate 1	Cyst sample 1	Cyst sample 2	Cyst sample 3	Cyst sample 4
Plate 2	Cyst sample 5	Cyst sample 6	Cyst sample 7	Cyst sample 8
Plate 3	Cyst sample 9	Cyst sample 10	Serum Sample 1	Serum Sample 2
Plate 4	Serum Sample 3	Serum Sample 4	Serum Sample 5	Serum Sample 6
Plate 5	Serum Sample 7	Serum Sample 8	Serum Sample 9	Serum Sample 10

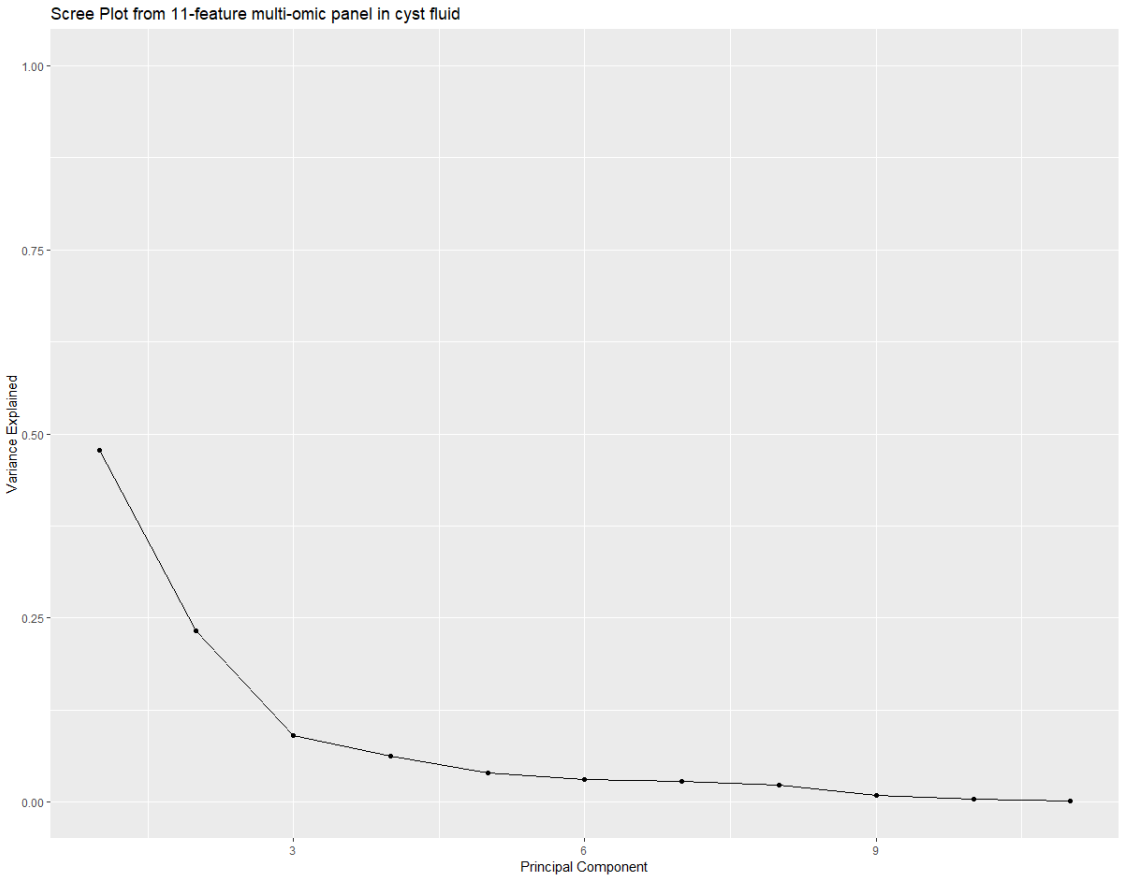
Appendix 3. QIAGEN custom 24-array plate layouts.



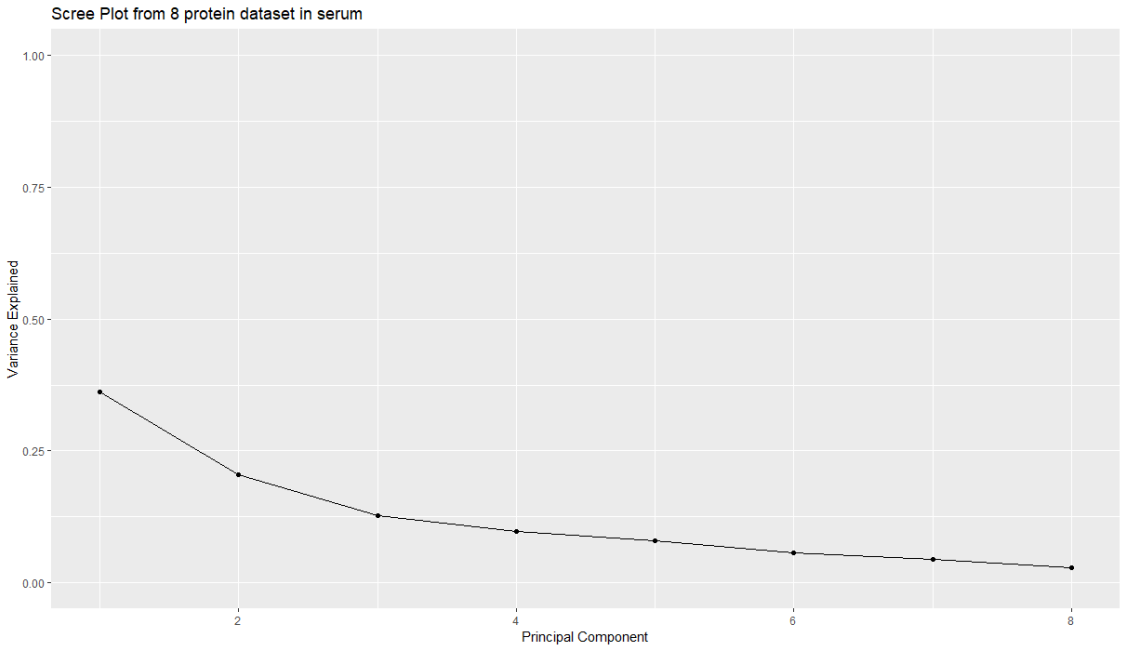
Appendix 4. Scree plot showing the variance explained by each principal component for the 8-protein panel in PCF.



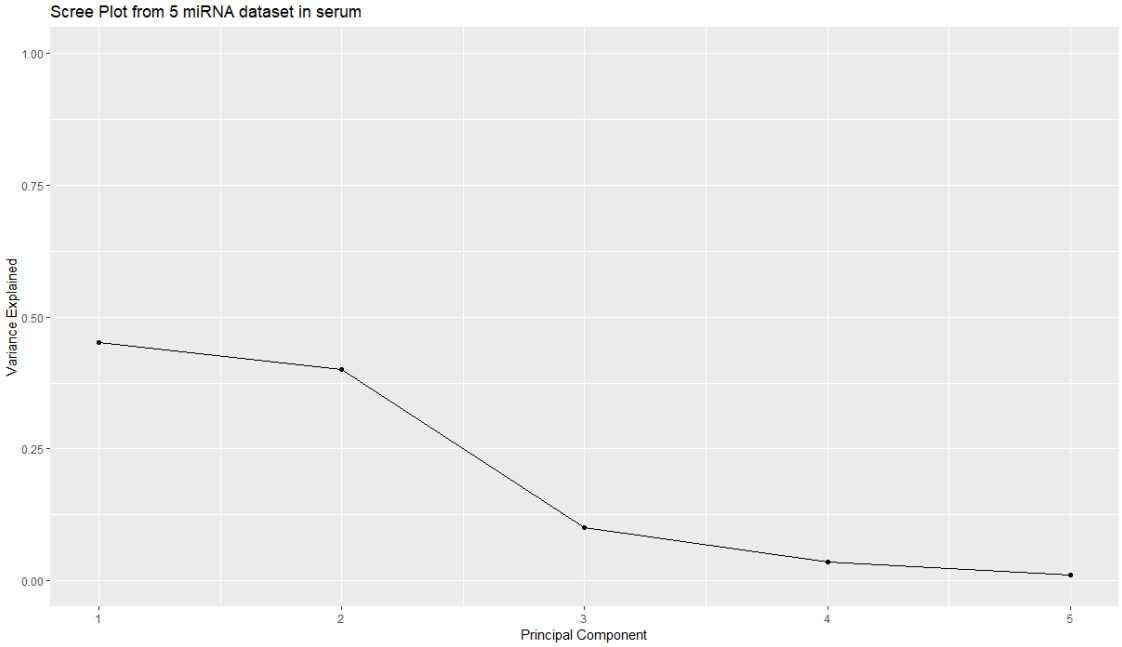
Appendix 5. Scree plot showing the variance explained by each principal component for the 3-miRNA panel in PCF.



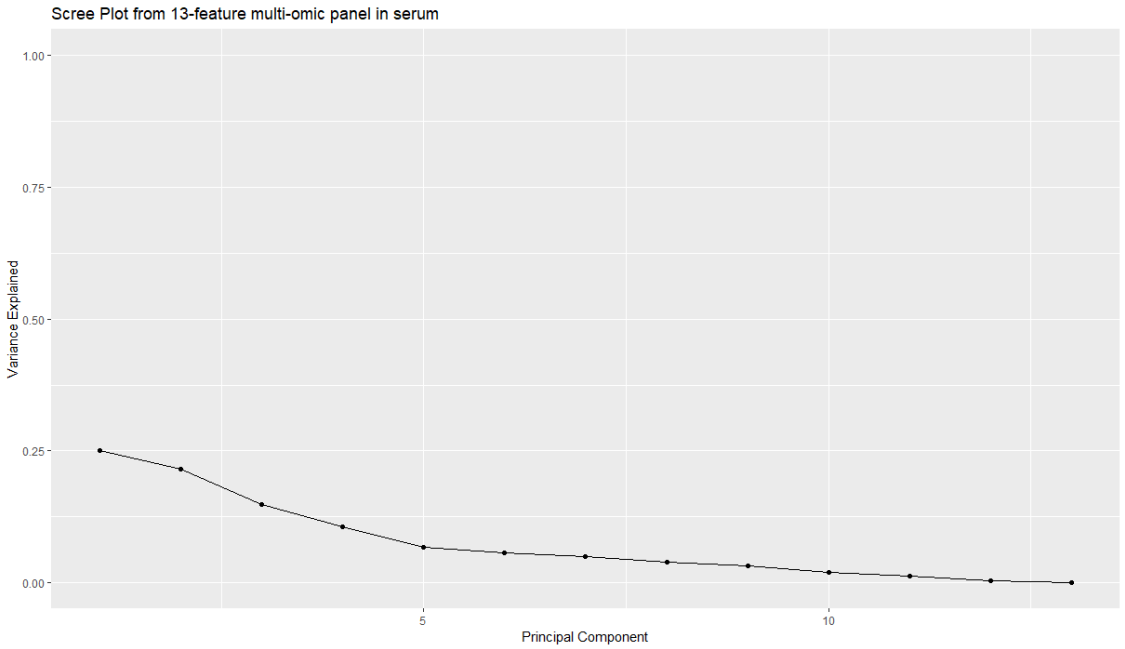
Appendix 6. Scree plot showing the variance explained by each principal component for the 11-feature multi-omic panel in PCF.



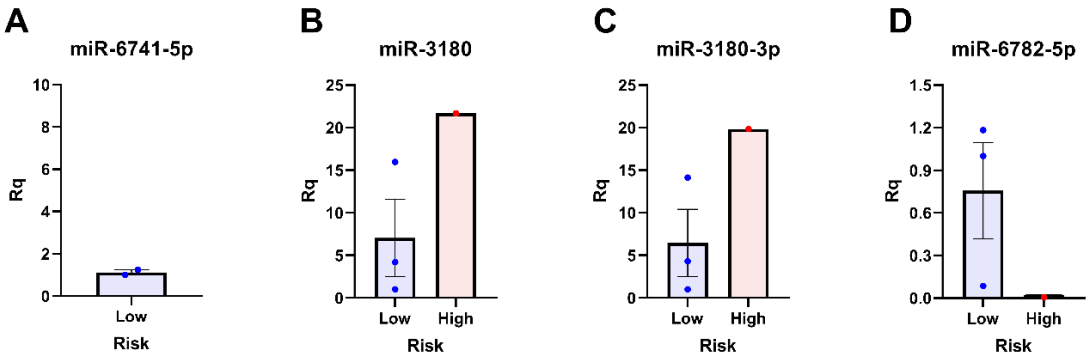
Appendix 7. Scree plot showing the variance explained by each principal component for the 8-protein panel in the serum.



Appendix 8. Scree plot showing the variance explained by each principal component for the 5-miRNA panel in the serum.



Appendix 9. Scree plot showing the variance explained by each principal component for the 13-feature multi-omic panel in the serum.



Appendix 10. qPCR microarray does not validate transcriptomic results in the serum. Relative quantification (Rq) of four miRNAs identified in the serum via transcriptomics. (**A**) MiR-6741-5p data are presented as mean ± SEM for n=2 low-risk patients. (**B**) MiR-3180 data are presented as mean ± SEM for n=3 low-risk and n=1 high-risk patients. (**C**) MiR-3180-3p data are presented as mean ± SEM for n=3 low-risk and n=1 high-risk patients. (**D**) MiR-6782-5p data are presented as mean ± SEM for n=3 low-risk and n=1 high-risk patients.